

# ERCIM NEWS

European Research Consortium for Informatics and Mathematics  
[www.ercim.org](http://www.ercim.org)

Number 62, July 2005



**Special:**  
**Multimedia Informatics**

## KEYNOTE

- 3 by Viviane Reding,  
Member of the European Commission,  
responsible for Information Society  
and Media

## JOINT ERCIM ACTIONS

- 4 Jérôme Chailloux nominated  
ERCIM Manager
- 4 Image and Video Understanding  
wins the ERCIM Working Group  
Award
- 5 Grid@Asia — Advanced Grid  
Research Workshops through  
European and Asian Co-operation
- 6 Strategic Workshops organized  
by ERCIM

## NEWS FROM W3C

- 6 W3C Celebrated Ten Years Leading  
the Web in Europe
- 6 Launch of the W3C Mobile Web  
Initiative
- 7 New Fee Structure for W3C  
Membership
- 7 W3C Internationalization Activity  
looks towards Africa
- 7 Latest W3C Recommendations

## SPECIAL THEME: MULTIMEDIA INFORMATICS

### Introduction

- 8 Multimedia Informatics  
by Joachim Köhler
- Multimedia Indexing and Retrieval**  
*Invited article:*
- 10 Managing Digital Photo Collections  
by Lynn Wilcox
- 11 Multimedia Indexing: The Multimedia  
Challenge  
by Patrick Gros, Manolis Delakis,  
and Guillaume Gravier
- 12 MediAssist: Managing Personal Digital  
Photo Archives  
by Noel Murphy, Cathal Gurrin  
and Gareth J. F. Jones
- 14 The MUSCLE Benchmarking Initiative  
by Allan Hanbury and Michael Nölle
- 15 New Testbed of One Million Images  
by Gregory Grefenstette, Pierre-Alain  
Moëllic, Patrick Hède, Christophe Millet  
and Christian Fluhr
- 16 Managing the Growth of Multimedia  
Digital Content  
by David Bainbridge, Paul Browne, Paul  
Cairns, Stefan Rüger and Li-Qun Xu
- 18 Maps of Music  
by Andreas Rauber, Thomas Lidy and  
Robert Neumayer
- 19 Structuring Multimedia Archives  
with Static Documents  
by Denis Lalanne and Rolf Ingold
- 20 MultimediaN: Personalized Information  
Delivery  
by Marcel Worrning and Nellie Schipper

- 21 SPIEGLE: A Multimedia Search Engine  
Generator  
by Arjen de Vries

- 23 Towards a 'Smart Content Factory'  
by Georg Güntner

- 24 Fischlär-News: Multimedia Access  
to Broadcast TV News  
by Alan F. Smeaton, Noel E. O'Connor  
and Hyowon Lee

- 25 Interactive Multimedia-Enabled  
Learning and Training  
by Claire Kenny, Declan McMullen,  
Mark Melia and Claus Pahl

### Multimedia Networking

- 27 Representation and Communication  
of Multimedia Data and Metadata  
by Sara Colantonio, Maria Grazia Di Bono,  
Massimo Martinelli, Gabriele Pieri  
and Ovidio Salvetti

- 28 A Cognitive Architecture  
for Semantically Based Medical  
Image Retrieval  
by John Moustakas, Socrates Dimitriadis  
and Kostas Marias

- 29 Personalized and Adaptive Multimedia  
Retrieval  
by Joemon M. Jose and Jana Urban

- 31 ADMITS: Adaptation  
in Distributed Multimedia IT Systems  
by Laszlo Böszörményi

- 32 Solutions for an Interpreter-Enabled  
Multimedia Conferencing System  
by Ferenc Sárközy and Géza Haidegger

- 34 Host Recommendation in the Adaptive  
Distributed Multimedia Server  
by Ottó Hutter, Tibor Szkaliczki and  
Balázs Goldschmidt

- 35 Scalable Audio Streaming  
to Mobile Devices  
by Jonathan Sherwin  
and Cormac J. Sreenan

- 36 A Networked Approach  
to TV Content Distribution  
by Adrian Cahill, John Roche  
and Cormac J. Sreenan

- 38 Presentation, Control  
and Collaboration  
in the Networked Classroom  
by Leandro Navarro-Moldes  
and Manuel Oneto

- 39 Video Transcoding Architectures  
for Multimedia Real Time Services  
by Maurizio A. Bonuccelli, Francesca  
Lonetti and Francesca Martelli

### Interactive Multimedia Applications

- 40 Designing Multi-Modal  
Multi-Device Interfaces  
by Silvia Berti and Fabio Paternò

- 41 LimSee2: A Cross-Platform SMIL  
Authoring Tool  
by Romain Deltour, Nabil Layaïda  
and Daniel Weck

- 43 Merging Virtual Reality and Television  
through Interactive Virtual Actors  
Marilyn — Multimodal Avatar  
Responsive Live Newscaster  
by Sepideh Chakaveh

- 44 Interactive Multimedia for Supporting  
the Quality of the Production  
by George L. Kovács, Géza Haidegger  
and János Nacsá

- 45 The CINEMA Project: A Video-Based  
Human-Computer Interaction System  
for Audio-Visual Immersion  
by Renaud Dardenne, Jean-Jacques  
Embrechts, Marc Van Droogenbroeck,  
and Nicolas Werner

- 46 Personalised Enriched Broadcast  
Experience  
by Mounia Lalmas, Nick Bryan-Kinns  
and Alan Pearmain

## R&D AND TECHNOLOGY TRANSFER

- 48 LIGHT: XML-Innovative Generation  
for Home Networking Technologies  
by Luca Tarrini and Vittorio Miori

- 49 Modelling of Authentic Reflectance  
Behaviour in Virtual Environments  
by Michal Haindl and Jiří Filip

- 51 Haptic Training Systems in Virtual  
Surgery  
by José San Martín, David Miraut,  
Carolina Gómez and Sofía Bayona

- 52 Computer Recognizes Whale Tails  
by Annette Kik, Eric Pauwels  
and Elena Rangelova

- 53 Text Document Classification  
by Jana Novovičová

- 54 Advancing Black-Box Reuse in a  
Multimedia Application Framework  
by Bernhard Wagner

- 55 CASSEM: Vibration Control in the  
Smart Way  
by Salim Belouettar

- 57 Point6: The IPv6 Skill Centre —  
Moving to the Next-Generation  
Internet Protocol  
by César Viho and Annie Floch

- 58 Working Slower with More Powerful  
Computers  
by Lorenz M. Hilty, Andreas Köhler,  
Fabian van Schéele and Rainer Zah

- 59 Coordinating IST Research  
across Europe  
by Simon Lambert

- 60 Software Automation meets  
Interactive Media Development  
by Dirk Deridder, Thomas Cleenewerck,  
Johan Brichau and Theo D'Hondt

## EVENTS

- 62 Workshop on Challenges in Software  
Evolution  
by Tom Mens

- 62 WWV 2005 — First Workshop  
on Automated Specification  
and Verification of Web Sites  
by María Alpuente, Santiago Escobar  
and Moreno Falaschi

- 63 Announcements

- 66 EURO-LEGAL

- 67 IN BRIEF



**D**igital convergence between audiovisual media, high-speed networks and smart devices is a reality. Multimedia informatics is a key enabling technology in this process. It helps make media content more directly manageable by computers, in an age where new information stored on paper, film, magnetic and optical media is reckoned to double every three years.

What will be the benefits? Digital convergence presents new opportunities for businesses to unlock new markets, and for public bodies to improve the way they work. For example, national newspapers can reach customers through print, online or mobile services. Hospitals will be able to make X-Rays available instantly at the patient's bedside.

Convergence also presents new challenges, especially in the interoperability between the network, device and content levels. Media content needs to be reformatted, restructured and re-indexed continually for multi-channel distribution – and this can best be done with automated solutions.

What is the EU's role? On the research front, the EU's sixth Framework Programme for Research and Development contributes to digital content, cognition and interface development as well as to eLearning and culture.

Digital content research combines semantic methods with audiovisual and Web technologies, allowing providers to create new forms of attractive and meaningful content for the consumer. In one instance, commercial producers and distributors will aim to deliver media content to end-users through different channels including interactive TV, personal computer, kiosks, mobile and handheld devices. In another case, a network of excellence - led by ERCIM - will bring together research groups in data mining and machine learning, to automate semantic-based multimedia retrieval from media content.

Cognition research breaks new ground in modeling machine perception and understanding more closely the human brain. One goal of this approach is to arrive at a new kind of computer that can describe what it sees, in real life situations. Imagine the benefits of a system which can report a traffic jam that it sees over a video camera, or which can translate sign language into words.

We also need easy-to-use interfaces, as far as possible using our own language and preferences. Research here aims, for example, at a new generation of machine translation, based on automated speech recognition and spoken language translation of broadcast news and speeches, starting with English and Spanish. Research will also explore multimodality, for example smart meeting rooms and electronic assistants which can collect, annotate and distribute different kinds of meeting materials on the fly – spoken, written or visual.

What does the future hold? My new i2010 initiative, adopted by the Commission on 1 June, sets an objective of increasing our investment in ICT research by 80% over the coming years. But i2010 does not stop there. It addresses key policy and regulatory matters, such as digital rights management. It also aims at a more inclusive information society. A recent eLearning conference explored the need to reform education and training systems, and to promote digital literacy, e-skills and lifelong learning in an ageing population. In the area of digital culture, we intend to strengthen EU policies concerning the preservation and exploitation of Europe's written and audiovisual heritage, with the help of all the public players concerned, making easy access by citizens to Europe's valuable resources an everyday reality.



**Viviane Reding,**  
Member of the European  
Commission responsible for  
Information Society and Media

*Viviane Reding*

## Jérôme Chailloux nominated ERCIM Manager

ERCIM's board of directors has nominated Jérôme Chailloux as manager of ERCIM during their meeting in Helsinki on 29th May. Jérôme Chailloux was proposed by INRIA, the host of the ERCIM office.

The ERCIM manager is the valid representative of ERCIM vis-à-vis third parties. He is responsible for ensuring that the implementation of ERCIM's general policy is within the framework specified by the membership. Jérôme will be in charge of managing both W3C and ERCIM offices. An important task will be the joint coordination of the ERCIM office and W3C Europe. There is a strong demand on behalf of the ERCIM consortium for an efficiently managed W3C Europe since the ERCIM EEIG members are liable for its activities.



From 1980 to 1987, Jérôme Chailloux was a researcher at INRIA, where he became a research director whilst occupying a number of teaching positions (École Polytechnique, CERICS). He worked in the areas of automatic VLSI circuits, software engineering and knowledge-based systems, and is the main inventor and developer of the programming language Le-Lisp, specializing in artificial intelligence.

From 1987 to 1995, Jérôme Chailloux co-founded the company ILOG, the second subsidiary of INRIA, taking on the positions of Chief Scientific Officer and director. ILOG is a world leader in the production of software components in the fields of optimization, decision aid and visualization.

From 1995 to 2001, he was Chief Information Officer of the genomics company GENSET, which is listed on the Nouveau Marché and NASDAQ stock exchanges. In this role, he decided upon strategy, implemented IT and bioinformatics resources and led one of Europe's largest teams of bioinformaticians.

Until 2000, he was a member of the Coordination Committee for Science and Information Technology and Communication of the French National Ministry for Education, Research and Technology.

He is president of the Sophia Complexity Association, partner in the investment fund of The Hyper Company, scientific adviser for the companies Genclis, Chiasma and the Thrombosis Research Institute, and manager of the consultancy company CERTICS.

### Please contact:

Jérôme Chailloux, ERCIM office  
Tel: +33 4 92385010  
E-mail: [jerome.chailloux@ercim.org](mailto:jerome.chailloux@ercim.org)

## Image and Video Understanding wins the ERCIM Working Group Award

The ERCIM Working Group "Image and Video Understanding" is the winner of the 2005 Working Group Award. The award consists of the right to spend up to 20,000 Euro for its activities. It was presented to the Working Group chair Eric Pauwels, CWI, by the ERCIM president Keith Jeffery during the ERCIM meetings in Helsinki on 30 May 2005.

Shortly after it became an official ERCIM Working Group, the Image and Video Understanding Group successfully submitted its proposal for the MUSCLE Network of Excellence (NOE) which covered the same scientific topics. The consortium was enlarged by approximately 20 additional scientific members in the formation of the new Network. The high level of collaborative research that arose from this network resulted in the group winning the 2005 ERCIM Working Group Award.



Keith Jeffery (left) presents the Working Group Award to Eric Pauwels.

The 2005 award was the last award in the current format. ERCIM is currently developing a new scheme to support the Working Groups in their attempts to create new project proposals.

### Links:

MUSCLE Network of Excellence:  
<http://www.muscle-noe.org>

ERCIM Working Groups:  
<http://www.ercim.org/activity/workgroup.html>



## Grid@Asia — Advanced Grid Research Workshops through European and Asian Co-operation

The Grid@Asia project will foster collaboration in Grid research and technologies between the European Union and Asian countries with a particular focus on China and South Korea. The project is coordinated by ERCIM and supported by the European Commission.

Grid@Asia will define a joint research agenda to address international Grid priorities relying on a core of leading European Grid research institutes. This initiative will be supported by Asian partners to ensure on-site organisation, enhanced visibility and the participation of high-profile industrial and scientific delegations. Grid@Asia is expected to provide Europe with a clear picture of the Grid community in those two Asian countries and to prepare a reliable ground for sustainable and long-term collaboration. The project is implemented through three principle steps:

- identification of Chinese and South Korea key players in Grid research and technologies,
- organisation of focused workshops around EU/Asia research and industrial agendas
- establishment of sustainable cooperation and dissemination activities.

Grid@Asia will support long term international cooperation by weaving additional links with leading Asian Grid research communities, in particular through the integration of Asian expertise with leading European Grid initiatives in the 6th Framework Programme of the European Union (such as Networks of Excellence, Integrated Projects, STREPS, etc.) and later on within the forthcoming 7th Framework Programme (FP). This will position the European Grid community as a leading centre of excellence, enrich European expertise in the field and support the adoption of common Grid standards worldwide.

To reach these goals it was decided to organise a series of three workshops in China and South Korea (Beijing, Shanghai and Seoul). The first event



Workshop participants.

took place from 21 to 23 June 2005 in Beijing, hosted by the Beihang University.

The workshop focussed on three main scientific themes — GRID Middleware, GRID Applications, Tools and Programming Environments — and allowed the European and Asian Grid communities to discuss and identify their common areas of interest. Two series of presentations were given. The first on current European projects supported by the Commission, such as AKOGRIMO, NEXTGRID, SIMDAT, DILIGENT, DEISA, CoreGRID and GRIDLAB and similar Chinese projects. The second featured project ideas which could lead to the submission of common proposals. National and multilateral programmes including their funding mechanisms were also presented with the goal to initiate collaborations between promising European and Asian research teams within the 6th and 7th FP.

The next workshop, to be organised before the end of 2005, will focus on Grid applications.

**Link:**  
<http://www.gridatasia.net>

**Please contact:**  
Bruno Le Dantec, ERCIM office  
Tel: +33 4 92 38 50 10  
E-mail: [office@ercim.org](mailto:office@ercim.org)

## Strategic Workshops organized by ERCIM

by Jessica Michel

Beyond the Horizon, a coordination action funded by the Future and Emerging Technologies (FET) activity of the EU's IST-FP6 Programme, has launched its opening workshops in June.

Six workshops will take place between June and September uniting leading European experts in IST-related research areas requiring support. The goal is to create critical scientific mass in these areas and ensure that European research stays one step ahead of general trends. The scheduled workshops include the following:

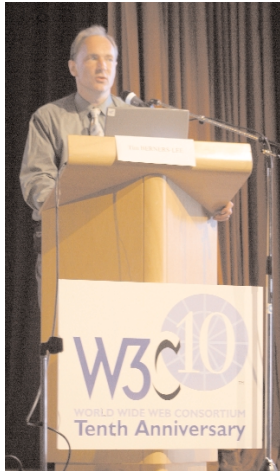
- Intelligent and Cognitive Systems, Zurich, Switzerland, 11-13 June
- Security, Trust and Dependability, Paris, France, 20-23 June
- Bio-ICT Synergies, Sophia Antipolis, France, 28-29 June
- Pervasive Computing & Communications, Vienna, Austria, 28 July
- Software Intensive Systems, Koblenz, Germany, 9-10 September
- Nanotechnologies and Nanoelectronics, Brussels, Belgium, 11-12 October
- Plenary Workshop, Brussels, Belgium, 12-13 December.

ERCIM must mobilize its extensive network of researchers to contribute maximally to the success of this project. If you would like to learn more about participating in any one of these thematic areas, or attend the Plenary Workshop during which inter-disciplinary research directions will be explored, please contact Jessica Michel.

**Link:**  
<http://www.beyond-the-horizon.net>

**Please contact:**  
Jessica Michel, ERCIM office  
Tel: +33 4 9238 5089  
E-mail: [jessica.michel@ercim.org](mailto:jessica.michel@ercim.org)

## W3C Celebrated Ten Years Leading the Web in Europe



Tim Berners-Lee.

W3C held a celebration of its ten years in Europe on Friday, 3 June 2005, in Sophia Antipolis, France. This half-day celebration afforded W3C Members and invited guests the opportunity to reflect on the progress and the role of the Web in Europe.

W3C10 Europe was part two of the W3C Tenth Anniversary celebration and follows on the W3C10 celebration last December, which marked the anniversary of W3C's founding at the Massachusetts Institute of Technology.

W3C10 Europe speakers and panelists discussed the importance of the Web in Europe, W3C's central role in the development of the Web, and visions of the future of the Web. The program emphasized two themes:

- The Web as Unifying Force in Europe: The European Union experience provides a compelling backdrop for considering how to expand the frontiers of the Web to enable participation by new communities. Speakers addressed the integration



Tim Berners-Lee  
and Robert Cailliau.



Keith Jeffery.

of the Web into the lives of Europeans, some challenges of internationalization, and advances in sharing data across communities.

- Policies Shaping the Web in Europe: Speakers discussed how Europeans are tackling important Web policy issues such as privacy, ensuring access by people with disabilities, and use of the Web in the public sector.

Tim Berners-Lee, W3C Director and inventor of the Web, reunited with his former CERN colleague Robert Cailliau to share personal reflections and stories about how the Web got started at CERN. Berners-Lee also delivered a keynote stressing the importance of Web standards and addressing current challenges in the European industry and research communities.

### Links:

W3C10 Europe: <http://www.w3.org/2005/06/W3C10.html>

Tim Berners-Lee keynote on "The Economic Importance of Standards": <http://www.w3.org/2005/Talks/w3c10-TimBernersLee/>

All presentations available from:

<http://www.w3.org/2005/Talks/w3c10-Overview/>

## Launch of the W3C Mobile Web Initiative

Today, mobile Web access suffers from interoperability and usability problems. Browsing the Web from a mobile device is not as convenient as expected. Users often find that their favorite Web sites are not accessible or not as easy to use on their mobile phone as on their desktop computer. Content providers have difficulties building Web sites that work well on all types and configurations of mobile phones offering Web access.

On 11 May 2005, W3C launched the Mobile Web Initiative (W3C-MWI) to make browsing the Web from mobile devices a reality. "Mobile access to the Web has been a second class experience for far too long," explained Tim Berners-Lee, W3C Director. "MWI recognizes the mobile device as a first class participant, and will produce materials to help developers make the mobile Web experience worthwhile."

Mobile Web Initiative participants will initially focus on the two areas 'best practices' and 'mobile device descriptions'. The Mobile Web Best Practices Working Group (MWBP WG) is chartered to develop authoring guidelines, checklists and best practices to help content providers to develop Web content that works well on mobile devices. The Device Description Working Group (DD WG) is chartered to address the development of improved device description solutions, that is, a

database of descriptions that can be used by content authors to adapt their content to a particular device.

W3C is already active in the mobile Web space, developing Web standards for multimodal interaction and device-independent design, as well as profiles for mobile devices; related standards include XHTML, SVG Mobile Profiles, and the SMIL Basic Profile. MWI work will complement these current efforts. MWI is also chartered to establish cooperative ties with related groups, including the Open Mobile Alliance (OMA) and 3GPP. These ties will help ensure that the needs of users are well-defined and that the efforts of the MWI and these related groups are complementary.

W3C is pleased to welcome the Founding Sponsors of the Mobile Web Initiative: Afiliat, Bango.net, Drutt Corporation, Ericsson, France Telecom, HP, Jataayu Software, MobileAware, Nokia, NTT DoCoMo, Opera Software, TIM Italia, Segala M Test, Sevenval, RuleSpace, V-Enable, Vodafone and Volantis. Participation in MWI-sponsored W3C Working Groups is open to all organizations.

### Links:

Mobile Web Initiative: <http://www.w3.org/2005/MWI/>

Mobile Web Initiative Sponsorship Program:

<http://www.w3.org/2005/MWI/Sponsoring.html>



## New Fee Structure for W3C Membership

In keeping with its international mission to lead the Web to its full potential, W3C announced a new fee structure designed to reduce the barrier of entry for organizations in developing countries. The goal is to make it easier for small companies and not-for-profit organizations to become W3C Members and become engaged in the development of foundation technologies for the World Wide Web.

Standardized technologies built in a flexible manner, with attention to internationalization needs (languages and/or infrastructure) can have dramatic impact on life, education and commerce in a given region.

However, one of the greatest obstacles for participation in large consortia is the cost of entry. What appears to be a reasonable membership fee in Western Europe, Japan or North America is prohibitive in other parts of the world. While W3C does embrace participation from individuals as invited experts, the consortium realized that more was necessary to engage organizations around the globe.

W3C is all about building Web technologies that can be of service to the world. This new fee structure for organizations from the developing world affirms the value W3C places on their participation in, contribution to and use of the standards and guidelines developed to drive the future of the World Wide Web.

This initiative, focused on regions of the world beginning to discover Web technologies, is only the most recent in W3C's international commitments. Through its technical Internationalization Activity, volunteer-based translation program, its fourteen Offices around the globe, as well as its Patent Policy, W3C can better meet the needs and requirements of diverse populations, and can help those regions develop sound, standards-based Web infrastructure.

### Links:

New fee structure: <http://www.w3.org/Consortium/fees>

Join W3C documentation: <http://www.w3.org/Consortium/join>

## Latest W3C Recommendations

- XML Key Management Specification (XKMS 2.0)  
*28 June 05, Shivaram H. Mysore, Phillip Hallam-Baker*
- XML Key Management Specification (XKMS 2.0) Bindings  
*28 June 2005, Shivaram H. Mysore, Phillip Hallam-Baker*
- Character Model for the World Wide Web 1.0:  
Fundamentals  
*15 February 2005, Tex Texin, François Yergeau, Richard Ishida, Martin J. Dürst, Misha Wolf*

### A complete list of all W3C Technical Reports:

<http://www.w3.org/TR/>

## W3C Internationalization Activity looks towards Africa

Richard Ishida represented the W3C Internationalization Activity in Casablanca beginning of June as a keynote speaker at a 3-day Pan-African Localization Workshop organized by the International Development Research Center. The workshop, the first of its kind, brought together participants from twelve African countries as well as experts from other continents to discuss how to better localize ICT into indigenous languages and scripts so as to promote rapid and fair development in Africa. The workshop was also visited by M. Rachid Talbi el Alami, Moroccan Minister-Delegate to the Prime Minister in Charge of General and Economic Affairs, and Carmen Sylvain, Canadian Ambassador to Morocco. Both expressed support for its aims. The workshop sets a foundation for future networking and information sharing via the development of a collaborative, Web-based site which will provide useful information and support the initiatives of a pan-African community of localizers.

Contributing to the workshop supports the aims of the W3C to increase participation by developing countries in the process of developing Web technologies. The W3C has recently revised its member fees to encourage participation by such countries.

The W3C Internationalization Activity has the goal of proposing and coordinating any techniques, conventions, guidelines and activities within the W3C and together with other organizations that allow and make it easy to use W3C technology worldwide, with different languages, scripts, and cultures.

### Articles and Tutorials on International Usage of W3C Technologies

W3C's Internationalization Activity reviews W3C technologies in production for internationalization concerns. It also regularly publishes articles and tutorials relating to international usage of W3C technologies. For example, the latest article describes the use of the language tags to indicate the language of text in HTML and XML documents, as well as in HTTP headers, SMIL and SVG switch statements, CSS pseudo-elements, etc. The tutorials list covers the multilingual web addresses usage, the ruby markup and styling, the character sets and encodings in XHTML, HTML and CSS, and many more.

The Internationalization Activity welcomes the participation of individuals and organizations around the world to help improve the appropriateness of the Web for multiple cultures, scripts and languages.

### Links:

W3C Internationalization Activity: <http://www.w3.org/International/>

Articles: <http://www.w3.org/International/articles/>

W3C Web Internationalization Tutorials:

<http://www.w3.org/International/tutorials/>

Language tags in HTML and XML:

<http://www.w3.org/International/articles/language-tags/>

# Multimedia Informatics

by Joachim Köhler

**Multimedia informatics** is a multidisciplinary research area looking at the creation, processing, transmission and consumption of audio-visual data. It combines expertise in signal processing, pattern recognition, coding technology, networking and protocols, data modelling and user interaction. Although these are all challenging research areas in their own right, multimedia informatics stresses their applied usage, exploitation and adaptation. It covers the whole chain of processing, from content authoring to media indexing of the archived content.

Media informatics provides the basis for many practical applications, such as video conferencing, advanced digital TV and Internet services. It has gained from enormous advances in data-capture and presentation hardware, and in broadband network technologies for real-time transmission of multimedia data services.

Different standardization organizations and funding agencies have recognized the important role of multimedia technologies. The MPEG community (<http://www.chiariglione.org/mpeg/>) was very successful in standardizing advanced audio and video coding technologies. MPEG-2 coding and transmission technology forms the basis for all digital video broadcasting (DVB) services and has generated huge revenue in the media industry. Advanced MPEG-4 audio and video codecs allow the transmission of audio-visual data over mobile networks. The increased availability of multimedia content has also generated the need to manage and organize audio-visual data. This is covered by the MPEG-7 standard, which contains a metadata description scheme to describe multimedia content for

indexing and retrieval applications. The goal of many research groups is to invent automatic methods to extract relevant metadata information, using methods from the area of signal processing and pattern recognition. This is a typical example of how algorithms and methods are applied and exploited for multimedia content processing.

Although several applications of multimedia technology already exist in our daily lives, there are still many advances and improvements to be made. This special issue contains 29 articles on a variety of research projects in this field being undertaken by ERCIM members. The invited article by Lynn Wilcox from FXPAL research describes an advanced system for managing digital photo collections using a face recognition engine to detect and classify persons on digital photos. Other work is the result of national and European projects, including SAVANT, MUSCLE and SIMILAR. The importance of international projects is clear, especially when benchmarking tests must be carried out to evaluate the performance of a multimedia system.



The articles are clustered into three areas, which are organized in the following manner:

### Topic 1: Multimedia Indexing and Retrieval

This topic contains most of the submitted articles and includes research work on content-based image retrieval, multimedia search engines and the automatic generation and management of metadata. The goal of several research groups is to increase the semantic knowledge of multimedia resources and to close the semantic gap between high-level features coming from text processing, and low-level features from image and audio processing.

### Topic 2: Multimedia Networking

Seven articles cover investigations on multimedia networking issues, and present distributed multimedia systems that are connected and managed with intelligent network components (eg proxies). Work on scalable streaming techniques and transcoding mechanisms for adapting the bandwidth for heterogeneous networks is subject of the articles from Sherwin (page 35) and Bonuccelli (page 39).

### Topic 3: Interactive Multimedia Applications

These articles describe research work on content authoring and show several interactive multimedia applications. With the toolkit LimSee2, developed by Romain Deltour (page 41), it is possible to create multimedia applications and content using the W3C standard SMIL. Another toolkit called TERESA from the research organization ISTI-CNR allows the development of multimodal user interfaces in multi-device environments (page 40). The virtual actor Marilyn (see article by Sepideh Chakaveh from Fraunhofer IMK on page 43) is applied as a virtual newscaster for an interactive TV application.

#### Please contact:

Joachim Köhler,  
Institute for Media Communication -IMK,  
Fraunhofer ICT Group, Germany  
E-mail: Joachim.Koehler@imk.fraunhofer.de

## ARTICLES IN THIS SECTION

### Introduction

#### 8 Multimedia Informatics

by Joachim Köhler

### Multimedia Indexing and Retrieval

#### Invited article:

#### 10 Managing Digital Photo Collections

by Lynn Wilcox

#### 11 Multimedia Indexing: The Multimedia Challenge

by Patrick Gros, Manolis Delakis, and Guillaume Gravier

#### 12 MediAssist: Managing Personal Digital Photo Archives

by Noel Murphy, Cathal Gurrin and Gareth J. F. Jones

#### 14 The MUSCLE Benchmarking Initiative

by Allan Hanbury and Michael Nölle

#### 15 New Testbed of One Million Images

by Gregory Grefenstette, Pierre-Alain Moëllic, Patrick Hédre, Christophe Millet and Christian Fluhr

#### 16 Managing the Growth of Multimedia Digital Content

by David Bainbridge, Paul Browne, Paul Cairns, Stefan Rüger and Li-Qun Xu

#### 18 Maps of Music

by Andreas Rauber, Thomas Lidy and Robert Neumayer

#### 19 Structuring Multimedia Archives with Static Documents

by Denis Lalanne and Rolf Ingold

#### 20 MultimediaN: Personalized Information Delivery

by Marcel Worring and Nellie Schipper

#### 21 SPIEGLE: A Multimedia Search Engine Generator

by Arjen de Vries

#### 23 Towards a 'Smart Content Factory'

by Georg Güntner

#### 24 Fischlär-News: Multimedia Access to Broadcast TV News

by Alan F. Smeaton, Noel E. O'Connor and Hyowon Lee

#### 25 Interactive Multimedia-Enabled Learning and Training

by Claire Kenny, Declan McMullen, Mark Melia and Claus Pahl

### Multimedia Networking

#### 27 Representation and Communication of Multimedia Data and Metadata

by Sara Colantonio, Maria Grazia Di Bono, Massimo Martinelli, Gabriele Pieri and Ovidio Salvetti

#### 28 A Cognitive Architecture for Semantically Based Medical Image Retrieval

by John Moustakas, Socrates Dimitriadis and Kostas Marias

#### 29 Personalized and Adaptive Multimedia Retrieval

by Joemon M. Jose and Jana Urban

#### 31 ADMITS: Adaptation in Distributed Multimedia IT Systems

by Laszlo Böszörményi

#### 32 Solutions for an Interpreter-Enabled Multimedia Conferencing System

by Ferenc Sárközy and Géza Haidegger

#### 34 Host Recommendation in the Adaptive Distributed Multimedia Server

by Ottó Hutter, Tibor Szkaliczki and Balázs Goldschmidt

#### 35 Scalable Audio Streaming to Mobile Devices

by Jonathan Sherwin and Cormac J. Sreenan

#### 36 A Networked Approach to TV Content Distribution

by Adrian Cahill, John Roche and Cormac J. Sreenan

#### 38 Presentation, Control and Collaboration in the Networked Classroom

by Leandro Navarro-Moldes and Manuel Oneto

#### 39 Video Transcoding Architectures for Multimedia Real Time Services

by Maurizio A. Bonuccelli, Francesca Lonetti and Francesca Martelli

### Interactive Multimedia Applications

#### 40 Designing Multi-Modal Multi-Device Interfaces

by Silvia Berti and Fabio Paternò

#### 41 LimSee2: A Cross-Platform SMIL Authoring Tool

by Romain Deltour, Nabil Layaïda and Daniel Weck

#### 43 Merging Virtual Reality and Television through Interactive Virtual Actors Marilyn — Multimodal Avatar Responsive Live Newscaster

by Sepideh Chakaveh

#### 44 Interactive Multimedia for Supporting the Quality of the Production

by George L. Kovács, Géza Haidegger and János Nacsa

#### 45 The CINEMA Project: A Video-Based Human-Computer Interaction System for Audio-Visual Immersion

by Renaud Dardenne, Jean-Jacques Embrechts, Marc Van Droogenbroeck, and Nicolas Werner

#### 46 Personalised Enriched Broadcast Experience

by Mounia Lalmas, Nick Bryan-Kinns and Alan Pearmain

# Managing Digital Photo Collections

by Lynn Wilcox

**The FXPAL photo application automatically organizes digital photo collections based on date, event, person, or place.**

The increased use of digital cameras is creating a need to manage large collections of photos. People want to organize their photos, browse their collections, search for a particular photo, and create slide shows to share with friends. However, people are typically unwilling to manually classify their photos for organization, and most often simply upload the images from the camera into folders on their computer.

The FXPAL photo application is a tool that allows users to automatically organize their photo collections on the fly. The user first tells the application where the photos are stored on the computer. He then selects the desired view and the system displays the photos accordingly. For example, if the user wants to see the photos organized by date, he can select the dates view to see photos grouped hierarchically by year, month, and day. Similarly, he can select the people view to see photos grouped according to who is in the picture and the place view to see photos grouped according to location. There is also an event view that automatically groups the photos into meaningful events, such as birthdays or weddings.

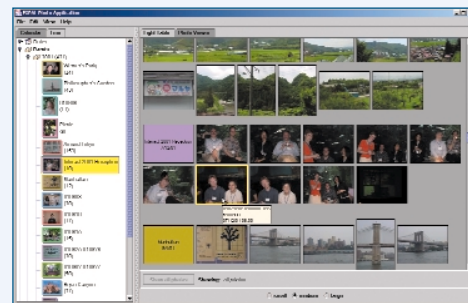
We achieve these groupings by automatically analyzing the photos' content and the metadata associated with the photos. By metadata, we mean information such as the date, time, and location the picture was taken. Grouping by time and date is straightforward — we simply create a hierarchy that allows users to select all photos by year, month, or day. The FXPAL photo application provides a calendar interface that visualizes the photos within the calendar for easier selection. The use of space is optimized in sparse calendar views by expanding the days of the calendar where there are photos and shrinking the days where there are none.

Grouping photos by events such as holiday gatherings or vacations is a

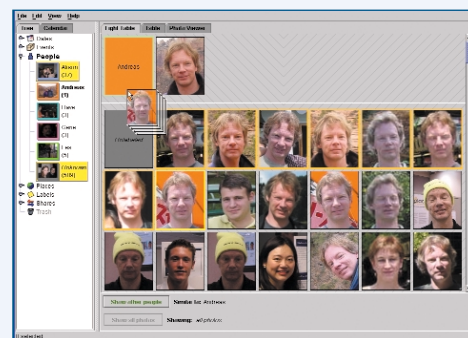
common practice among photographers. The FXPAL system automatically detects events by clustering photos according to the time they were taken and their content. The technique is based on similarity analysis, in which the self-similarity matrix of the photos, computed from temporal and content similarity, is partitioned into disjoint events. After photos have been automatically grouped into events, users can attach semantic labels, so that when the photos are displayed in the event view the contents of the collection are easily understood.

Another common method for grouping photos is according to people. Manually assigning names to people in photos is tedious and time consuming. We make use of the FSCA face detection and recognition system to make this task easier. Face detection is highly accurate and is run automatically. Faces are cropped from the photos by orienting an ellipse based on the line segment between the eye positions. Users can view only the cropped faces from the photos by selecting the face view.

To determine the identity of people in the photos, we use the FSCA face recognition system. Unlike face detection, face recognition is not reliable. Variations in lighting, pose, and eyewear cause the system to make errors. Thus we provide a user interface for semi-automatic classification of faces according to person. Using the face view, the user selects one or more faces of a particular person to create a model. The user then asks the system to find similar faces. This results in a display of faces ordered by similarity to the model. The user selects correct faces, usually found near the top of the list, and adds them to the set of faces corresponding to the person. These faces are used to update the model, thus increasing the accuracy when more faces



**Figure 1: Event View in the photo organizer. Events have been automatically determined. Semantic event labels are assigned by the user.**



**Figure 2: Face view in the photo organizer, showing cropped faces from photos and their similarity to the photo labeled 'Andreas'.**

of the person. This process is repeated until all faces for people in the collection have been labeled.

The FXPAL photo application also provides interesting ways to view photos. The Pan 'n Scan animated slide show displays full screen photo images with background music. The slide show is animated by panning and zooming through each image using rules based on location of faces. Another way to view photos is as a Stained Glass image. Selected photos are cropped using detected faces and assembled in a collage with irregular boundaries similar to stained glass windows.

The photo application created by FXPAL and FSCA is available for user testing. Please contact the author for use permission.

**Link:**

<http://www.fxpall.com/?p=PhotoApplication>

**Please contact:**

Lynn Wilcox, FX Palo Alto Laboratory, USA  
Tel: +1 650 813 7574  
E-mail: Wilcox@fxpal.com



# Multimedia Indexing: The Multimedia Challenge

by Patrick Gros, Manolis Delakis and Guillaume Gravier

**Multimedia indexing is a very active field of research, despite most works using only a single medium. This is mainly due to the fact that while they may be correlated, media are not strongly synchronized. Segment models appear to be a good candidate to manage such a desynchronization.**

Multimedia indexing has become a general label to designate a large domain of activities ranging from image description to description languages, from speech recognition to ontology definition. Of course, these fields existed before the expression ‘multimedia indexing’ became popular, and most continue to have an independent existence. However, the rise of multimedia has forced people to try to mix them together in order to manage properly big collections of multimedia documents. The global goal of multimedia indexing is to describe documents automatically, especially those containing images, sounds or videos, allowing users to retrieve them from large collections, or to navigate these collections easily. Such documents, which used to be rare due to the price of acquisition devices and because of the memory required, are now flooding our digital environment thanks to the camera-phones, webcams, digital cameras, as well as to the networks that allow the data to be widely shared. The question is no longer “How can I acquire a digital image?”, but rather “How can I retrieve the image I want?”

## What Does Multimedia Change?

While it is possible to study images or audio tracks alone for some documents, such approaches appear to be very limited when applied to multimedia documents like TV streams. This limitation is twofold. First, users (who are not specialists or documentalists) would like to access such documents semantically; second, users face huge sets of documents. As a consequence, many techniques that reduce semantics to syntactic cues in the context of small sets of documents are no longer useful, and no single medium can provide acceptable access to document semantics.

If one considers a TV stream, it is apparent that images are not able to provide a lot of semantic information. The information that can be extracted from this medium includes segmentation information (shot detection, clustering of neighbouring shots), face detection and recognition capabilities, and text and logo detection. It is possible to do a lot more but only in very limited contexts, like news reports or sports broadcasts. In such contexts, syntactic cues like outdoor/indoor classifications have a pertinent semantic translation (anchor person/outdoor reports), but these tricks cannot be used in open contexts. The situation is similar in audio analysis. Cries and applause are good indications of interesting events in sport reports, but not in drama and films. On the other hand, audio can provide useful segmentation information (music or speech detection), speaker detection and recognition, key sound detection, or speech transcription capabilities. There may be several sources of interesting text, eg internal sources like closed captions, text included in the images, speech transcription or external sources such as program guides.

## The Big Challenge: Mixing Media

The best way to describe a document is to make use of all the information it carries, and thus all the media it includes. If this statement seems obvious, it nevertheless implies many practical difficulties. The various media within a document are not synchronized temporally and spatially: the speaker is not always visible on the TV screen, the text related to an image may not be the closest thing to this image, audio and video temporal segmentations have different borders. To make things worse, audio and video do not work at the same rate (100Hz for audio, and 24, 25 or 30Hz for video).

From a more general point of view, audio, video and text are studied using different backgrounds, which are not always easy to mix. Text requires natural language-processing tools that use data analysis or symbolic techniques, while image and audio are branches of signal processing and use a lot of statistical tools but in the continuous domain. Other domains like geometry are also used. Mixing all these tools in one integrated model is one facet of the problem.

Two common solutions to this problem exist in the literature. The first is to use the media in a sequential manner. One medium is used to detect some event, and another medium is then used to classify it. For example, audio can be used to find the most important events in a soccer game, while video is necessary to understand what kind of event it is. Such an approach does not require a theoretical framework, remains ad-hoc and is not so difficult to implement, and is a good starting point for many problems. The second uses Hidden Markov Models (HMMs) to describe and recognize sequences of events. Markov models are of common use in sound and image processing and are very suited to identifying sequences of events. This is thanks to the Viterbi algorithm, which is based on a dynamic programming approach and provides a global optimal solution at a reasonable cost.

## Segment Models: A Promising Approach

In the context of multimedia documents like video streams, HMMs have strong limitations due to the fact that each state may correspond to one and only one observation. On the other hand, this observation can contain a visual and an audio part. In the context of video documents, this means that a single temporal

granularity must be chosen for the observations, and to align sound information on video units (images or shots) or vice versa. We used such models to retrieve the structure of videos of tennis, and despite the limitation, these models performed well in terms of precision of shot classification.

We propose using an enhanced version of these models called segment models (SMs). In these models, each state can accept a variable number of observations, this number (or its distribution) being a new parameter of the state. On the one hand, such a model allows a different number of visual and audio observations for a given audio-visual event. On the other hand, it adds some complexity to learning the conditional probabilities of the observations, and to identifying the duration of each state in the data streams. Our first results show that segment models can outperform Markov models. However, the main work is now to determine how much flexibility we can gain, and what can be done that was impossible before.

We present the performance of HMMs and SMs on a test set of three tennis games. The task is to segment the complete video into predefined scenes,

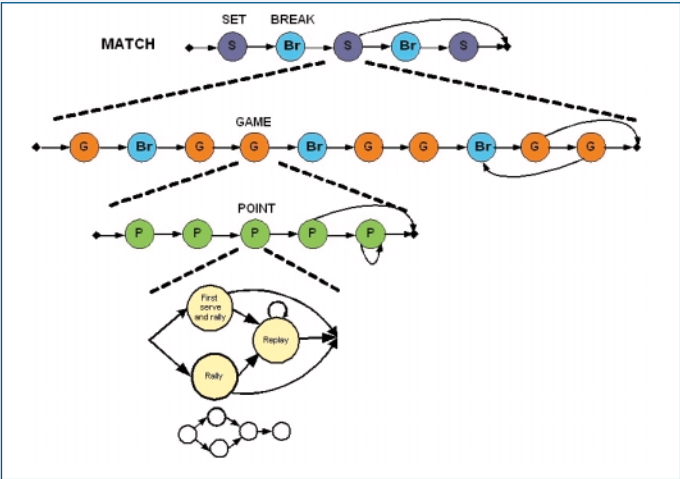


Figure 1: Structure of a HMM for tennis videos analysis.

namely ‘first missed serve and exchange’, ‘exchange’, ‘replay’, and ‘break’. In this context, a state in SMs represents a complete scene rather than a shot, as in HMMs. Performances are measured in terms of percentage of shots assigned with the correct scene label (C), and recall (R) and precision (P) rates on

	C	P	R
HMMs	74.57	73.69	82.51
SM1gram	76.95	72.28	72.47
SM2gram	79.17	75.11	80.13

Table 1: The performance of Hidden Markov Models (HMMs) and Segment Models (SMs) on a test set of three tennis games.

scene boundaries detection. We use shot-based audio descriptors in HMMs. The video part of the observations of a scene for SMs is modelled via HMMs, operating as observation scorers. The audio part is modelled by unigram models of audio events (SM1gram) that fuse audio descriptors as in HMMs but at the scene level, or by bigram models (SM2gram) that can capture a succession of audio events inside a scene.

**Please contact:**  
Patrick Gros, IRISA – CNRS, France  
Tel: +33 2 99 84 74 28  
E-mail: Patrick.Gros@irisa.fr

# MediAssist: Managing Personal Digital Photo Archives

by Noel Murphy, Cathal Gurrin and Gareth J. F. Jones

MediAssist is creating a range of applications to help people to manage their digital photo archives by employing context to automatically annotate photo collections.

Recent years have seen a revolution in photography with a move away from analog film towards digital technologies. Many users of digital cameras are now accumulating very large numbers of personal digital photos. While digital storage offers ample capacity to store these collections, technology for managing digital photos has not kept pace with advances in capture and storage technologies. The MediAssist project at the Centre for Digital Video

Processing (CDVP) at Dublin City University is developing applications to enable users to efficiently search their image archives. Users often remember where and when a photo was taken, and indeed may recall this more clearly than the actual contents of a photo, so clearly context information of this type should be effective when searching digital photo archives. MediAssist applications use the capture and exploitation of context data as the basis for organising

and searching personal photo collections. The time and date of photo capture are easily accessible from the camera, and this can be augmented with location information using coincident GPS data.

A typical operational scenario for a user using MediAssist tools would require a GPS enabled digital camera and although available, these are currently expensive. While awaiting the arrival to market of consumer-grade digital





**Figure 1:**  
MediAssist desktop  
search application.

In addition to the desktop interface, we are developing a mobile photo management application. This is based on the same underlying architecture as the desktop application, with an interface tailored to suit mobile devices.

It is known that the reduced screen space and limited options for interaction with mobile devices means that user interaction with mobile devices should differ greatly from those of the standard desktop applications. Figure 2 shows the interface to the MediAssist mobile application, this addresses interactivity issues by primarily presenting a user with a personalised list of recommended 'favourite' photos. As a secondary access methodology it also supports location and feature based searching of the user's archive. The search window in the mobile device is hidden from view and only appears when the user wishes to search the photo archive. The search options are more limited than in the desktop environment, and are primarily based on location and easily selected useful features, such as season and weather. The result of a search is a list of events, with a single most representative photo chosen for each event. If a user wishes, they can further browse the event by selecting the orange arrow.

The current MediAssist applications will be enhanced to incorporate advances in the annotation technologies as these become available within the CDVP. These improvements will include advances in recognition of content features within the photos, and also extensions to the annotation process to include further external knowledge.

The support of the Informatics Directorate of Enterprise Ireland is gratefully acknowledged.

**Link:**

Centre for Digital Video Processing:  
<http://www.cdvp.dcu.ie/>

**Please contact:**

Noel Murphy, Centre for Digital Video  
Processing, Dublin City University / Irish  
Universities Consortium  
E-mail: [murphyn@eeng.dcu.ie](mailto:murphyn@eeng.dcu.ie)

cameras with integrated GPS, we use a separate GPS device and we automatically match photos with their location from a GPS tracklog. Photos are uploaded from the camera to a PC which then automatically annotates each photo. This annotation stage extends the time labels to include weekday, weekend, month, season and year of capture, and the GPS location labels the town, city/state and country. Data from the camera is used to determine whether the photo was taken indoors or outdoors. The MediAssist tools use external information sources to further annotate images based on the time and location of capture to include whether the environment was light or dark, and even the prevailing weather conditions. In addition, various automatic content analysis tools are employed to add further content to a digital photo archive. Based on other research activities within the CDVP, we have integrated technologies which annotate each photo to identify (re-occurring) faces and buildings.

Once a user has their digital photos organized in the archive, MediAssist applications provide support for searching and browsing. We have two streams of research, producing a desktop and a mobile device interface.

Figure 1 shows the MediAssist desktop search application. Using this the user can search for photos based on location and/or time of capture, and to filter results based on light status, weather conditions, indoor or outdoor, and whether the photos contain people or buildings. Location based searching

allows the user to select combinations of desired country, city, and town, while time based searching enables selection of a specific time period, eg summer, or time ranges, eg between February and April.

When examining the results of a search, a user is presented either with an exhaustive list of images or a list of events, where each event is a combination of photos taken at the same time and place, eg a birthday party, family outing, or trip to the zoo. Each event is represented in the desktop interface with up to five key photos automatically extracted from the event, to best represent the varied themes within the event. Selecting an event automatically displays all photos from that event.



**Figure 2:** Mobile MediAssist photo  
management application.

# The MUSCLE Benchmarking Initiative

by Allan Hanbury and Michael Nölle

**Evaluating the performance of multimedia retrieval algorithms is part of the MUSCLE, a EU Network of Excellence administrated by ERCIM.**

The MUSCLE Network of Excellence aims at fostering close collaboration between research groups in multimedia data-mining and machine learning. Around 40 research groups are involved in the MUSCLE network, which has been in existence since March 2004.

The network aims at achieving two grand challenges in the area of multimedia learning: natural high-level interaction with multimedia databases and interpreting human behaviour in videos. In order to measure the progress towards achieving these challenges, a benchmarking initiative has been started within MUSCLE. The activities taking place within this benchmarking initiative are also open to research groups not directly involved in MUSCLE.

A very important component of a benchmarking initiative is test data. A repository for such data has been set up within MUSCLE (see link below). Most of the data in this repository is publicly available, while some is restricted to use by members of MUSCLE. The publicly available data includes: video sequences of artificially generated humans in natural scenes for evaluating motion detection and tracking algorithms; videos of different human gestures and videos of various types of basketball shot for evaluating human behaviour interpretation algorithms. A collection of 10 000 images of coins for evaluating classification algorithms and content based image retrieval approaches is also available. This Coin Image Seibersdorf database (CIS) is a result of the changeover from 12 European currencies to the Euro. After the changeover, large volumes of mixed coin collections had to be returned properly sorted to the national banks of the originating countries. The database consists of roughly 2000 patterns (classes) of coins from many different countries. Additionally, there are 100 000 coin images collected

during an automatic sorting process carried out at the ARC Seibersdorf research GmbH, which will later serve as test and benchmarking data. A carefully generated and manually verified ground truth accompanies the data. This makes the CIS database and benchmark definition ideally suited to a large scale evaluation of classification or object recognition algorithms.

All researchers in the multimedia information retrieval field are encouraged to contribute useful benchmarking data or software to this repository. This is particularly encouraged if results obtained by using this data have been published, as this allows other researchers to evaluate their algorithms on the same data. Furthermore, any

ground truth or annotation for the data collections can also be contributed.

Apart from making benchmarking data available, a number of benchmarking campaigns are being organised. These campaigns are open to all groups doing research in multimedia retrieval. The first campaign is the MUSCLE coin classification competition. This is an educational initiative aimed at encouraging senior students interested in pattern recognition and machine learning. Participants in the competition will submit code implementing algorithms for classifying the coin database currently available on the benchmarking webpage. These algorithms will be tested on a part of the database which has not yet been made public, and the author of the best performing algorithm will receive a prize sponsored by MUSCLE. The call for participation in this competition will appear on the MUSCLE benchmarking webpage in autumn 2005.

An evaluation campaign aimed at benchmarking image retrieval is planned for 2006 in collaboration with the CLEF image retrieval track (ImageCLEF). The first step in this collaboration is a workshop that will be held on the 20th of September 2005 in Vienna, in conjunction with the CLEF workshop and the ECDL 2005 (European Conference on Digital Libraries). The workshop is aimed at stimulating discussion on the current state of image retrieval evaluation (with well-known researchers in this field) as well as planning the 2006 evaluation campaign. More information is available on the MUSCLE benchmarking webpage.

An evaluation campaign linked to MUSCLE is ImagEVAL, funded by the French Ministry of Research. This campaign consists of the following five tasks related to image retrieval: recogni-

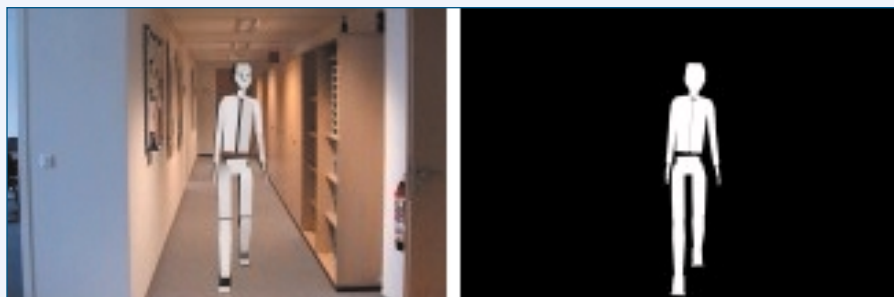


**Figure 1: Image from the CIS database (provided by ARC Seibersdorf Research GmbH).**



**Figure 2: Image from the basketball shots database (provided by INRIA-Vista).**





**Figure 3:** Image from the motion detection database (provided by Advanced Computer Vision GmbH – ACV). The image containing a computer-generated human is shown on the left, and the ground truth image (actual position of the human) is shown on the right.

tion of transformed images (rotation, translation, scaling, etc.); search for photographs illustrating a text using text and image analysis; extraction and recognition of text areas in an image; object detection; and automatic image classification (night/day, indoor/outdoor, city/nature, etc.).

Participation in these campaigns is encouraged as it will lead to an objective evaluation of the current state-of-the-art in the multimedia information retrieval research area. As the best results of these campaigns will be publicised, the groups having submitted the best-performing algorithms can benefit from the publicity.

#### Links:

MUSCLE NoE website:

<http://www.muscle-noe.org>

MUSCLE benchmarking page:

<http://muscle.prip.tuwien.ac.at>

Cross-Language Evaluation Forum (CLEF):

<http://www.clef-campaign.org/>

ImageCLEF: <http://ir.shef.ac.uk/imageclef>

ImageEVAL competition:

<http://www.imageval.org>

European Conference on Digital Libraries

2005 (ECDL): <http://www.ecdl2005.org>

#### Please contact:

Allan Hanbury, Technical University

Vienna, PRIP, Austria

Tel: +43 1 58801 18359

E-mail: [hanbury@prip.tuwien.ac.at](mailto:hanbury@prip.tuwien.ac.at)

Michael Nölle,

Quantumtechnology, ARC Seibersdorf

Research, Austria

E-mail: [Michael.noelle@arcs.ac.at](mailto:Michael.noelle@arcs.ac.at)

## New Testbed of One Million Images

by Gregory Grefenstette, Pierre-Alain Moëllic, Patrick Hède, Christophe Millet and Christian Fluhr

**The Commissariat à l'Energie Atomique (CEA) in France has produced an image database of one million images that will allow researchers in content-based image retrieval to test their system on a life-size collection.**

Content-based image retrieval involves searching a collection of images for those relevant to a given 'query image'. The user submits an image (or sometimes only the description of an image) and wants to find the images in the collection that most closely match. Currently, researchers have been limited to small collections in developing and testing their content-based image retrieval techniques. The largest test collections currently used (such as the University of Columbia databases, or the Corel databases) contain from hundreds up to 60,000 images. To serve as a test collection, the image database must contain labeled images, where the labels show which image is relevant to which other image. Any image can then serve as a test query, and researchers can measure the performance of their system using values such as precision (the proportion of retrieved images that actually are relevant to the query) and recall (the percentage of relevant images in the database that the system retrieves).

Within the EU-sponsored Network of Excellence MUSCLE (Multimedia Understanding through Semantics, Computation and Learning), a new testbed image collection of one million test images has been created. The database is called CLIC (for CEA List Image Collection) and has been produced by the LIC2M team at the CEA, which is one of the MUSCLE partners. LIC2M stands for Laboratoire d'Ingénierie de la Connaissance Mutlimédia Mutlilingue, and is a laboratory outside of Paris that specializes in image and text processing in many languages.

The CLIC image collection contains labelled images, each of which can be used as a query image. The image collection was created by hand-labelling photographs that were donated to the project by colleagues. Any photographs containing identifiable persons were removed. The remaining 15,200 photos were classified into a shallow hierarchy:



**Some images from classes of the kernel of the CLIC content-based image retrieval testbed.**

- Food: images of food, and meals
- Architecture: images of architecture, architectural details, castles, churches, Asian temples
- Arts: paintings, sculptures, stained glass, engravings
- Botanic: various plants, trees, flowers
- Linguistic: images containing text areas
- Mathematics: fractals
- Music: images of musical instruments
- Objects: images representing everyday objects such as coins, scissors etc
- Nature&Landscapes: landscapes, valley, hills, deserts etc
- Society: images with people
- Sports&Games: stadiums, items from games and sports
- Symbols: iconic symbols, road signs, national flags (real and synthetic images)
- Technical: images involving transportation, robotics, computer science
- Textures: rock, sky, grass, wall, sand etc
- City: buildings, roads, streets etc.

These labelled images form the kernel of the collection (see the figure for exam-

ples). Each image in the kernel was then altered in 69 different ways. The transformations applied to each original kernel image included: geometric transformations (such as rotation, translation, projection, and splitting), chromatic transformations (such as negative, saturation, black-and-white, and quantification) and various other transformations (such as low-pass filtering, noise addition, border addition, text incrustation, mosaic, resizing, and edge outlining). The altered images were added to the collection with the same labels, thereby generating the one million labelled images in CLIC. Any image can be used as a query, in the knowledge that at least 69 other images are relevant to the query. The transformations were designed to cover a wide variety that occur in natural image manipulation. For example, it is common to find slightly altered and cropped pictures to avoid detection for copyright infringement.

The database will be distributed to the research community through the

MUSCLE Network of Excellence, and will prove useful for testing text-based content-based image retrieval (using the labels), evaluating algorithm behaviour over large databases, testing the invariance of algorithms towards transformations, for automatic classification (using the hierarchy), for object and person recognition, and for the detection of text in images. The database will be distributed for research free of charge: please check the MUSCLE Web site for the latest information on its availability.

**Link:**

<http://www.muscle-noe.org>

For further information about the CLIC testbed, please consult the following article: Pierre-Alain Moëllic, Patrick Hède, Gregory Grefenstette, Christophe Millet, "Evaluating Content Based Image Retrieval Techniques with the One Million Images CLIC TestBed", Proceedings of the Second World Enformatika Congress, WEC'05, February 25-27, 2005, Istanbul, Turkey, pp 171-174.

**Please contact:**

Gregory Grefenstette,  
Commissariat à l'Energie Atomique, France  
Tel: +33 1 46 54 96 56  
E-mail: [Gregory.Grefenstette@cea.fr](mailto:Gregory.Grefenstette@cea.fr)

## Managing the Growth of Multimedia Digital Content

by David Bainbridge, Paul Browne, Paul Cairns, Stefan Rüger and Li-Qun Xu

**The growth in multimedia documents including collections of photos, music and videos etc has been phenomenal in recent years. Effective management of these multimedia collections has become a necessity for both businesses and the general public. Indeed with the advent of large volume and cheap storage devices and increasingly adopted broadband connections, home storage systems are becoming a reality for ordinary users to organise multimedia files. Current approaches - used mainly by fully trained professionals - are expensive and complicated. Can the media growth be effectively managed?**

A multimedia digital libraries project has been set forward to address the complex issues related to media management with a view to developing more effective approaches for searching and browsing multimedia content: The Multimedia Information Retrieval group at Imperial College London collaborates with Human-Computer-Interaction experts from University College London, Digital Library experts from the Greenstone group at the University of Waikato, New

Zealand, and Media Management experts at BT Research. The group is pulling together digital content from archives of the BBC, the British Library, the New Zealand Digital Library and the Victoria & Albert Museum.

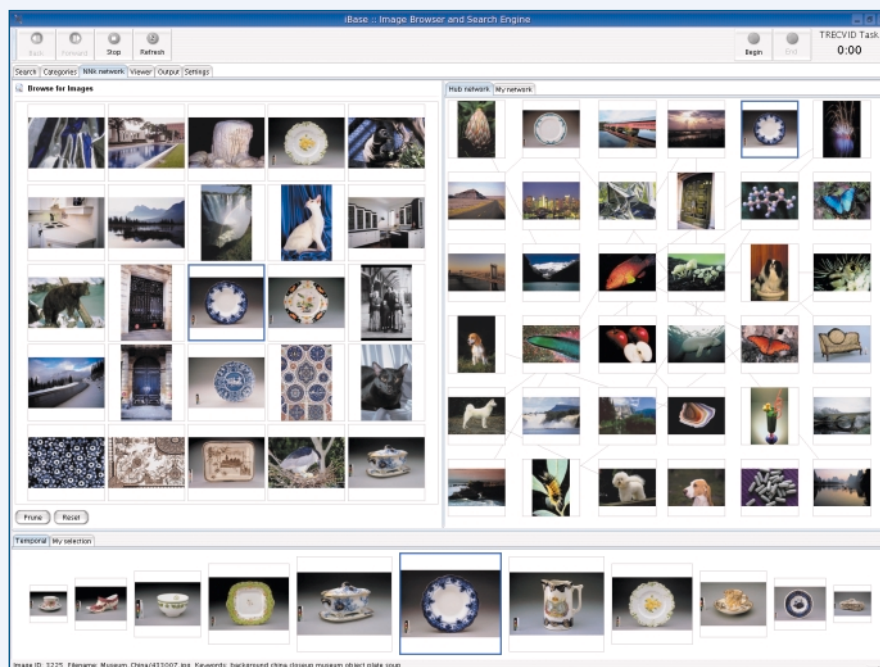
This project has the following four main objectives:

- develop a query-by-example retrieval approach using automated content-based analysis, in conjunction with

meta-data and text-based search when and if necessary

- devise new tailored search and browsing approaches appropriate for different media collections and users' needs
- reduce information overload by presenting and summarising search results in a semantically meaningful manner
- define new interfaces which can integrate and adapt to different media and user knowledge.

For example, we will automatically create new relations between multimedia objects (ie photos and videos segments) by automatically inserting links from each object to all those objects which are most similar under different criteria, thus creating a list of nearest neighbours with what we call 'lateral associations'. This generates a network which is eminently suitable for browsing and exploration. The right panel in the figure summarises a large image database by displaying all highly connected images in the network; they can serve as entry points for browsing. The lateral neighbours of a particular image are shown in the left panel. The bottom panel shows objects in their context, which may be temporal in the case of videos or semantic (eg same category or genre) in the case of image collections.



Interface design for lateral browsing.

### Who really needs Multimedia Digital Library Management?

The beneficiaries of such a system include students and scholars, the medical profession with their specific databases and the general public. Organisations that create and manage digital libraries could also benefit; they include museums, art galleries, and photo & movie libraries. Indeed, even casual users are starting to see a need for an effective way to manage their growing collections of digital photos, music and personal movies.

The project is meant to have a defining impact on the way documents / information are presented and accessed in digital libraries. It aims to provide easy access to digitised material that is not yet fully annotated. Such methods are likely to ultimately change the way products, goods and services are presented in digital catalogues and how they are marketed on the Internet. It will revolutionise the design of large scale video databases as well as spawn a range of new revenue opportunities for museums, art galleries and image libraries.

### Managing Future Multimedia Collections

Traditional libraries, digital or not, use meta-data, catalogues and classification systems to facilitate access to documents. Search engines such as Google have enhanced this process by indexing

and searching the entire documents to give access to the meta-data, ie, the link to the document. This project intends to go a step further by providing visual access modes, eg, through visual search boxes into which images can be dropped. Those visual access modes will be in addition to traditional meta-data search in libraries and full-text search boxes. As such, we amalgamate digital library functionalities given, eg by the successful Greenstone Digital Library (<http://www.greenstone.org>) with content-based multimedia access.

As applications evolve over time their functionality and user interface complexity tend to increase substantially. User interfaces do not generally offer different approaches for people having different individual characteristics (Novice, Intermediate, or Professional). Adaptable and responsive interfaces are important for the search and browse of media libraries.

### Our Multimedia Collections

The multimedia digital libraries project development requires a diverse collection of media in order to design and evaluate appropriate search and browsing approaches according to users' requirements. The varying levels of meta-data available within the collections will

provide a realistic testbed for validating the underlying technologies. The test collections include the following:

- 50 hours of television content from the BBC
- 1 million page images from a selection of 200 newspaper titles supplied by the British Library
- the University of Waikato's Maori newspaper collection
- the V&A collection of 29,000 images with annotation
- Imperial College London's digital Shoebox collection of 6,000 personal photos.

The multimedia digital libraries project aims to solve the complex media management issues by offering a tailored browsing and search access that is adaptable to users' needs as well as different digital library collections.

Software that arises from this project will become open source and will be disseminated under the GNU General Public License.

**Link:**  
<http://mmir.doc.ic.ac.uk/pr-mmdl-2005/>

**Please contact:**  
 Stefan Rüger, Imperial College London, UK  
 E-mail: [s.rueger@imperial.ac.uk](mailto:s.rueger@imperial.ac.uk)



# Maps of Music

by Andreas Rauber, Thomas Lidy and Robert Neumayer

**Manually sorting and searching through directory structures is no way to enjoy music, no matter how much we can pack onto a single, tiny device. Players should be 'intelligent' enough to organize the music for us, allowing us to pick only what we want to listen to.**

Music is becoming one of the dominant goods in internet traffic and private storage, with according business models slowly starting to catch up. Yet, to select from this wealth of music, to choose which songs to listen to in a specific situation or particular mood, we are still forced to turn to cumbersome clicking, to manually sort titles into playlists, again listen only to pre-compiled albums, or to revert the dulllest automatic selection method, namely random play.

In order to enjoy the plethora of music available, new techniques and interfaces are required that will free us from burdensome selection of tracks. While private music collections and audio players are desperately in need of more intuitive methods of organizing and easily selecting music, commercial vendors most probably are as well. This need is evident from the popularity of currently available recommendation techniques, such as the dominant "customers who bought this album also bought this" style recommendations. Allowing customers to casually select and browse sections of vast stocks of audio tracks and discover titles they didn't know before requires a different approach.

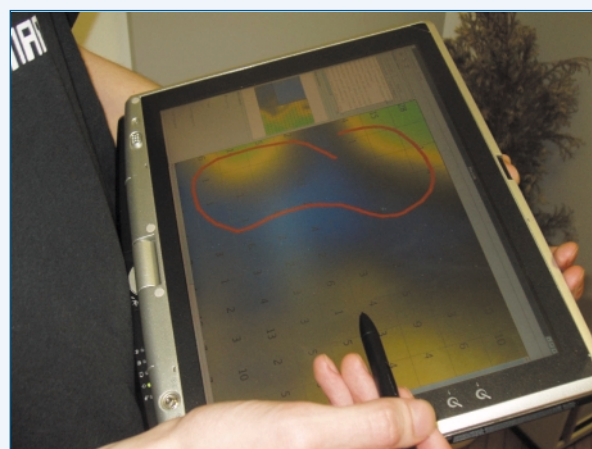
To this end we are developing methods that allow us to organize audio repositories by the way we perceive music, grouping audio tracks by their perceived acoustic similarity. The SOM-enhanced JukeBox system (SOMeJB) provides automatic indexing and organization of music repositories based on perceived sound similarity of single tracks. A map metaphor is used for visualization, with similar songs being placed into similar regions on the map.

Using a variety of feature extraction techniques, the audio signal in the form of WAV or MP3 files is analysed to extract representations that allow us to compute the perceived similarity of two pieces of music. Specifically, we use - amongst other statistical features - Rhythm Patterns, modelling the amplitude modulation frequency in different frequency bands while incorporating a range of psycho-acoustic transformations. In a two-stage feature extraction process, the specific loudness sensation in different frequency bands is first computed, and is then transformed into a time-invariant representation based on the modulation frequency. These features describe the complex rhythmic

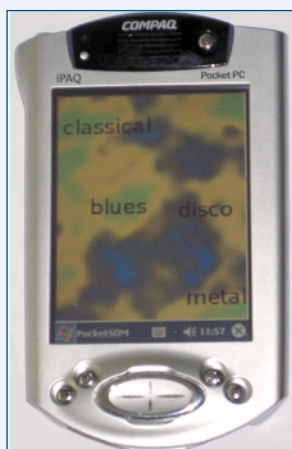
interactions that are characteristic for different musical styles.

On top of this, we can apply standard information retrieval techniques to search for specific pieces of music, or for tracks from a certain musical genre. Classifiers can be trained to sort audio into pre-defined genre categories. With the SOMeJB system, the SOM-enhanced JukeBox, we go one step further by clustering individual audio tracks using a self-organizing map, allowing us to overcome traditional genre boundaries. This system provides a mapping of the high-dimensional feature spaces into a two-dimensional map space that can be conveniently explored. Different visualization techniques, such as Smoothed Data Histograms, reveal the cluster boundaries on the map and result in 'Islands of Music' being depicted, with each island containing music of a specific type or style. Vector field visualizations on top of these produce weather charts that help users to interpret the map, telling them where more aggressive or quieter music is located on.

With the PlaySOM and the PocketSOMPlayer, we added two novel interface modules allowing us to browse a music collection by navigating a map of clustered music tracks and to select regions of interest containing similar tracks for playing. The PlaySOM system is primarily designed to allow interaction via a large, preferably touch-screen device, whereas the PocketSOMPlayer is implemented for mobile devices, supporting both local and streamed audio replay, or acting as a remote control for the audio server.



**Figure 1:**  
PlaySOM playlist selection by trajectory on a tablet PC.



**Figure 2:** PocketSOMPlayer running on an IPAQ.

Music is selected simply by marking an area on the map, or – in a slightly more sophisticated fashion – by drawing a trajectory on the map, along which music is played. This allows us to start

from, say, some soft classical piano music, move on to somewhat more dynamic orchestral pieces, before returning again to softer violin pieces. Or, in a different area of the map, starting with some electronic music, moving on to more dynamic pop and rock pieces, and gradually increasing in dynamics to reach the metal area on the map.

This approach allows users to quickly explore the range of music available and to effortlessly select styles of music

according to their preferences, simply by selecting the appropriate areas on the map. Current work is focusing on further improvements in the feature extraction process to even better capture the perceived characteristics of audio, as well as on methods for integrating different views of and interaction possibilities with large music collections.

Part of this work was supported by the European Union in the 6th Framework Programme, IST, through the Networks

of Excellence 'DELOS' on Digital Libraries, and 'MUSCLE' on Multimedia Understanding through Semantics, Computation and Learning.

**Link:**

<http://www.ifs.tuwien.ac.at/~andi/somejb>

**Please contact:**

Andreas Rauber

Vienna University of Technology, Austria

E-mail: [rauber@ifs.tuwien.ac.at](mailto:rauber@ifs.tuwien.ac.at)

## Structuring Multimedia Archives with Static Documents

by Denis Lalanne and Rolf Ingold

**If we consider static documents as structured and thematic vectors towards multimedia archives, they can be used as a tool for structuring events such as meetings. Here we present a method for bridging the gap between static documents and temporal multimedia data, such as audio and video. This is achieved by first extracting electronic document structures, then aligning them with multimedia meeting data, and finally using them as interfaces to access multimedia archives.**

Interfaces to textual-document libraries are improving, but search and browsing interfaces in multimedia-document libraries are still in the early stages of development. Most existing systems are mono-modal and allow searching either for images, videos or sound. For this reason, much current research in image and video analysis is focusing on automatically creating indexes and pictorial video summaries to help users browse through multimedia corpora. However, such methods are often based on low-level visual features and lack semantic information. Other research projects use language-understanding techniques or text captions derived from OCR, in order to create more powerful indexes and search mechanisms. Our assumption is that in a large proportion of multimedia applications (eg lectures, meetings, news etc), classical printed documents or their electronic counterparts (referred to by the term 'printable') play a central role in the thematic structure of discussions.

Unlike other multimedia data, static documents are highly thematic and structured, and thus relatively easy to

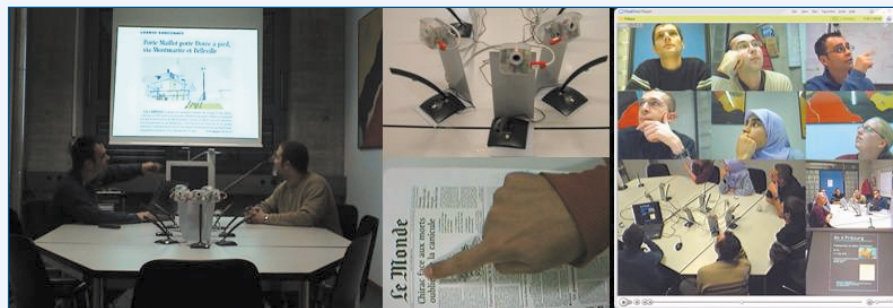
index and retrieve. Documents carry a variety of structures that can be useful for indexing and structuring multimedia archives, but such structures are often hard to extract from audio or video. It is therefore essential to find links between documents and multimodal annotations of meeting data, such as audio and video.

Recently there has emerged a significant research trend toward recording and analysing meetings. This is done mostly in order to advance research on multimodal content analysis and multimedia information retrieval, which are key features for designing future communication systems. Many research projects aim at archiving recordings of meeting in forms suitable for later browsing and retrieval. However, most of these projects do not take into account the printed documents that often form part of the information available during a meeting. We believe printable documents could provide a natural and thematic means for browsing and searching through large multimedia repositories.

For this reason, we have designed and implemented a tool that automatically extracts the hidden structures contained in PDF documents. The semantics of the information behind layout and logical structures is largely underestimated and we believe their extraction can drastically improve both document indexing and retrieval, and linking with other media.

In order to browse multimedia corpora using documents as interfaces, it is necessary to build links between printable documents, which are inherently non-temporal, and other temporal media. We use the term 'temporal document alignment' to refer to the operation of extracting the relationships between a document excerpt at variable granularity levels, and the meeting presentation time. Temporal document alignment creates links between document extracts and the time intervals in which they were in either the speech focus or the visual focus. It is thus possible to align document parts with audio and video extracts, and by extension with any annotation of audio, video and/or gesture.

In the FRIDOC multimedia browser that we have developed, users can first search at a cross-meeting level by typing in a set of keywords: this will retrieve all relevant documents. Clicking on a document or an article then allows users to view the related multimedia data attached to this element and to directly jump to the portions of meetings in which it was in focus. At the intra-meeting level, all the components (documents, audio/video, transcription and annotations) are synchronized through the meeting time, thanks to the document alignments; clicking on one of them causes all the components to visualize their content at the same time. For instance, clicking on a journal article cues audio/video clips from the time at which it was discussed, cues the speech transcription from the same time period, and displays the document that was projected.



**Multimedia meeting.**

This work demonstrates the role of static documents as structured and thematic vectors towards multimedia archives and proposes a method for bridging the gap between static documents and multimedia meeting archives. The results obtained so far through user evaluations tend to prove that documents are an efficient means of accessing multimedia corpuses, such as multimedia meeting

repositories or multimedia conference archives.

**Link:**

<http://diuf.unifr.ch/im2/>

**Please contact:**

Rolf Ingold,  
University of Fribourg, Switzerland  
E-mail: [rolf.ingold@unifr.ch](mailto:rolf.ingold@unifr.ch)

## MultimediaN: Personalized Information Delivery

by Marcel Worring and Nellie Schipper

**How can we make the best use of the abundance of multimedia information? The multidisciplinary PID project in the Netherlands is developing methods for capturing information, and then automatically indexing and presenting it to users in an optimal way.**

The amount of multimedia information being captured and produced is constantly increasing. While providing the right information to a user is already difficult for structured information, it is much harder in the case of multimedia information. When multimedia collections become large, complete manual annotation is no longer an option. As a consequence, automatic indexing of multimedia is becoming an essential ingredient in any modern information system.

The state-of-the-art in automatic video indexing is evaluated in the yearly TRECVID, an international video retrieval benchmark that focuses on news data. Our Mediamill team (UvA/TNO) has participated in all editions. Much progress has already been made in this field, and the perfor-

mance of automatic indexing techniques has proven to be useful for interactive retrieval. In the TRECVID we have shown that for successful indexing, all information about the data should be employed. For news video this means combining information from both the speech and the visual channel. Furthermore, analysis should not be restricted to the content of the two channels, but should also consider how the data is captured and what recurring use of style can be observed. For TRECVID2004 we indexed 32 concepts, and we are now scaling up to 50-100. From there we use ontologies to scale up by orders of magnitude the number of concepts for which an index is available.

Ultimately this will lead to automatic annotation both for produced video, like news and film, and non-produced video

captured with a security camera or by someone walking around with a camera. In the latter case, the user can employ the speech channel for spoken annotation. Further, it is clear that the use of video restricts us to a two-dimensional representation of the world. Ideally, the three-dimensional world and the objects within it could also be stored in the database. We are working on 3D reconstruction methods from video for this purpose.

In the end, our information systems will be filled with large collections of images, videos, and 3D worlds, together with annotations of these data items. Deciding what to present to the user depends on a number of different factors. What is the device the person is using? Is the user sitting behind her PC in her office, or is she walking around in the

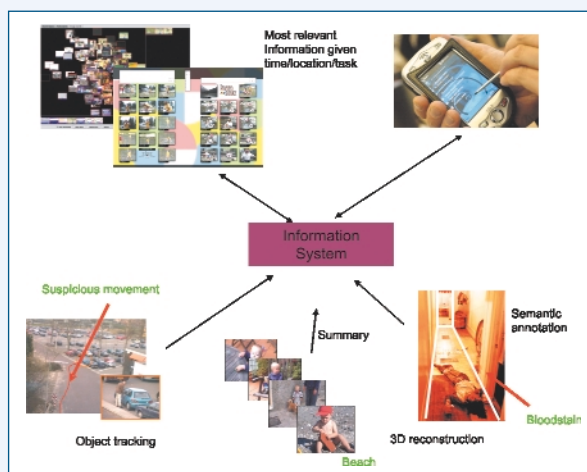


field with her PDA? Furthermore, it depends on the task being performed and the context in which it is performed. A cognitive engineering approach is vital, in order to provide the user with the right information by taking into account the task, context and user capabilities.

Clearly only a multidisciplinary approach can bring together all of the above. Experts are needed in computer vision, machine learning, information systems, information visualization and human computer interaction. These different disciplines have been brought together in the PID project.

### Context

The PID project is part of the large-scale MultimediaN project funded by the Dutch government. It started on 1<sup>st</sup> April, 2004, and will have a total duration of four years. The PID consortium consists of research institutes (University of Amsterdam, TNO), system integrators (LogicaCMG, Compans/Ziuz) and application holders (Dutch Olympic Committee, Dutch



Various applications of personalized information delivery.

Forensic Institute, the police). Thus, the project covers the whole chain from research to applications.

### Applications

To study the above methodologies a number of concrete applications are being pursued. For each, the whole chain from data capturing to presentation is considered, but each application has its emphasis on one of the elements. The application being developed with the Dutch Forensic Institute is the 3D reconstruction of crime scenes using video

cameras. Indexing of the crime scene will be performed using a combination of speech and visual analysis. A project in collaboration with the Dutch Olympic Committee is developing a personal coach, which captures the 3D movement of athletes, combining this information with data from other sensors, such as heart rate. This information is then used for tracking and improving the athletes' performance. For the police, the emphasis is on the cognitive side, aiming at equipping police on the job with attentive mobile devices that provide

them with relevant information. Finally, home videos are considered where the emphasis lies on creating summaries of the data and finding relations within the data.

#### Link:

<http://www.multimedien.nl>

#### Please contact:

Marcel Worring  
University of Amsterdam, The Netherlands  
Tel: +31 20 5257521  
E-mail: [worring@science.uva.nl](mailto:worring@science.uva.nl)  
Nellie Schipper, TNO, The Netherlands  
E-mail: [nellie.schipper@tno.nl](mailto:nellie.schipper@tno.nl)

## SPIEGLE: A Multimedia Search Engine Generator

by Arjen de Vries

When the efficiency of multimedia search system engineering is improved, multimedia search effectiveness will follow. The Semantic Multimedia Access project in the Netherlands has developed 'Spiegel', a parameterized search-engine generator. It generates a search system specialized for a particular set of circumstances, including user, form of data collection and background. The Netherlands Institute for Sound and Vision - the national audio-visual archive - and Van Dale publishers will test the system.

Multimedia searching is never a goal on its own, but in reality is embedded in user tasks. These vary in complexity, in the collections accessed, and in their contextual parameters. For example, a film maker may seek a suitable shot to include in a documentary, a teacher might look for an animation to illustrate a lecture, a group of friends may search for photos to accompany their stories

told at a wedding party, while a DJ spinning his records in a popular night club might try to find that perfect blues lick to sample over his beats.

It is to be expected that a search strategy that works well in one scenario will not necessarily be the best choice for another. Ideally, each user task would be matched by a specialized retrieval

strategy. In other words, a retrieval engine should be context-aware, or at least be adaptable to the context in which it will operate. For example, while the journalist could find relevant shots in a national archive, those looking for wedding photos would be better served on the Web, or in their friends' collective folders of digital photos. In addition, while the teacher may be satisfied with a

familiar example, the best answers to the DJ's query would be unusual musical samples.

Researchers from CWI, the University of Amsterdam and the University of Twente are collaborating on MultimediaN's Semantic Multimedia Access project (also known as MN-N5) to create new technology to ease this adaptation of search system to user task. The project's main goal is to develop Spiegle, a 'parameterized search engine

models with the probabilistic models developed for image and video retrieval, and will use the resulting system for multimedia search in both TRECVID and INEX 2005.

The second building block is the RAM database front-end for processing queries over arrays. RAM (Relational Array Mapping) was originally developed to express the retrieval models involving multimedia content. A recent article by Roelleke and others has

Search systems generated with Spiegle will be put to the test with a variety of search tasks in scientific evaluations such as TREC, TRECVID, INEX, and CLEF. Perhaps more interestingly though, the project also involves various end-user organisations, including the Netherlands Institute for Sound and Vision, and Van Dale publishers, who offer great case studies for further validation of our research results.

Sound and Vision is not only the business archive of the national broadcasting corporations, but also a cultural history institute and a unique media experience for its visitors. The institute intends to open its archive to program makers and researchers, as well as for educational purposes. Using our technology, it should be easier to support each of these user groups with search systems specialized to their search tasks. We also hope to reduce the burden of annotation by integrating search functionality into the annotation process.

Our work with Van Dale, a prominent publisher of dictionaries, demonstrates that project results are not limited to the 'multimedia search engine'. Both detection and tracking are important forms of searching, and we have applied our technology to track the development of language in terms of word usage and the shifting of meaning over time. We think the insights resulting from this project are also applicable in searching collections that span decades of text data.

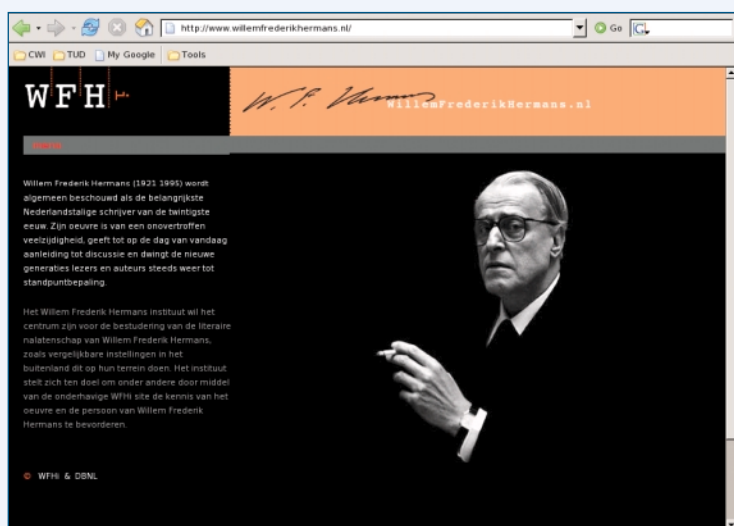
In summary, the main innovation of this project is its goal of working toward a system architecture that accommodates different types of searching, using different sets of a priori knowledge and exhibiting a varying degree of heterogeneity (or homogeneity). This will simplify considerably the comparison of different types of retrieval model instantiations on a series of search problems.

#### Links:

<http://monetdb.cwi.nl/projects/trecvid/MN5/>  
<http://www.spiegle.nl/>

#### Please contact:

Arjen de Vries, CWI, The Netherlands  
 Tel: +31 20 592 4306  
 E-mail: [Arjen.de.Vries@cwi.nl](mailto:Arjen.de.Vries@cwi.nl)



**The Willem Frederik Hermans (WFH) portal contains the complete primary and secondary bibliography, the author's huge private archive, as well as interviews broadcast on radio and television. Courtesy Willem Frederik Hermans instituut (WFHi) and the Stichting DBNL (Digitale Bibliotheek voor de Nederlandse Letteren).**

generator'. Spiegle takes two inputs: firstly, a collection schema and secondly, a declarative specification of a retrieval strategy suitable for the user search task. It then generates a search system specialized for the particular context at hand, including the combination of this user, this collection, and the specific background knowledge available.

Spiegle combines the results of two existing research lines, both based upon probabilistic methods for information retrieval. The first building block is the TIJAH structured document retrieval system, developed in the Cirquid project. TIJAH retrieves document components from XML documents, matching on both content and structure. We are currently extending its text retrieval

demonstrated how to express many well-known retrieval models in a general matrix framework. The corresponding matrix expressions are easily expressed in RAM, so it seems the ideal starting point for specifying search strategies declaratively.

Consequently, the Spiegle parameterized search engine can be realized by integrating TIJAH and RAM. TIJAH provides the techniques to adapt the search system to different collections and background knowledge, and RAM provides the declarative language for specifying the retrieval model. Since both are implemented as front-ends on the same database back-end (the open source database system MonetDB), this should be feasible without too many complications.

# Towards a 'Smart Content Factory'

by Georg Güntner

**A project at Salzburg NewMediaLab (SNML) introduces a pragmatic approach to the implementation of a knowledge-based infrastructure for search and retrieval in audiovisual repositories.**

While the consumer market provides us with increasing possibilities to create rich media content (cameras, smart phones etc), the 'professional' content industries (broadcasters, news agencies etc) continue to digitize their complete content life-cycle management processes. This results in a semantic gap between the ease of content creation and the need for the utilization of content in a context-aware, individually tailored way.

## The Smart Content Factory

In October 2003 a group of Austrian technology and science partners (X-Art ProDivision, ORF, Joanneum Research, coordinator: Salzburg Research) started a project entitled 'Smart Content Factory' (the Factory). The project aims to develop a knowledge-based infrastructure for search and retrieval in an audiovisual archive: The approach of the project is not to reinvent existing media management technology, but to create a framework superimposing a 'semantic layer' on top of state-of-the-art technology. The general objective of the project is to define a system architecture supporting a wide range of 'knowledge-intensive' user scenarios for the utilization of rich media content in the business to business (B2B) and business to consumer (B2C) areas.

The project is one of the lead projects at Salzburg NewMediaLab (SNML), the Austrian research centre in the area of digital content engineering.

## Objectives and Results

This section outlines the project objectives and shows the approach the project team took during the prototype phase to meet the requirements.

### Re-use of existing content:

By using content 'as is' there is no need to reformat or transform existing content into new data structures. Hence there is no need for redundancy of high-volume data storage. All operations

related to the media clips use unique references to the media objects.

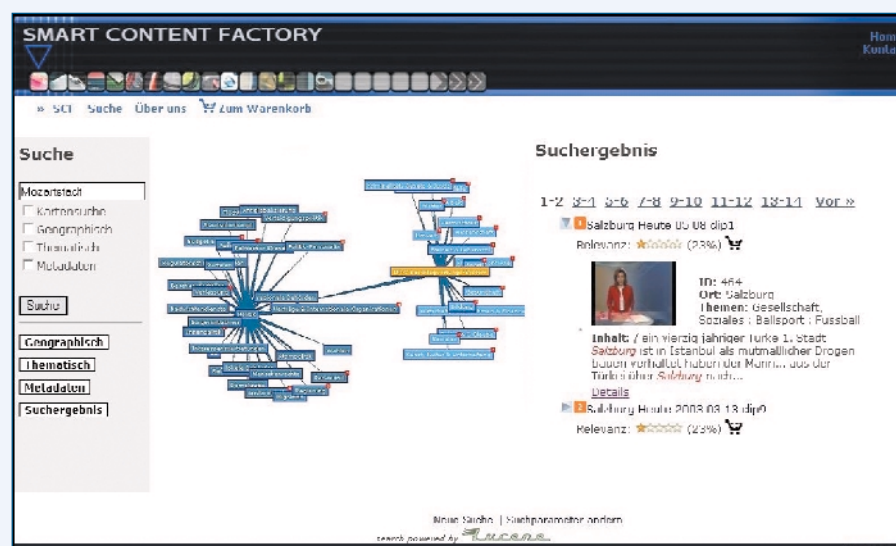
### Aspects of integration:

The Factory's indexing components offer interfaces for integration into the content production workflow. For example, when a new digital video editing system is introduced, the Factory will be informed of newly created content and the indexing process will be triggered as soon as new content is published. Currently the Factory supports an indexing pipeline consisting of two steps: a primary content-based indexing and a secondary semantic

programming interface. In the course of the project an MPEG-7 interface was created to avoid dependencies from the proprietary system.

### Adaptability and extensibility:

The Factory will be adaptable and extensible to various domain knowledge models. Currently three 'pluggable' knowledge models form the knowledge base of the Smart Content Factory: (i) the 'Geoname Thesaurus' developed by Salzburg Research; (ii) the IPTC thesaurus provided by the International Press and Telecommunications Council; and (iii) a service of (German) synonyms



The prototype of the Smart Content Factory at work (April 2005).

indexing using the concepts of the knowledge base.

### Interoperability:

The Factory makes use of state-of-the-art digital asset management technology. Wherever possible, standard-based interface layers are used to keep the core of the Factory independent of the underlying data and knowledge layer. The Factory uses Virage's VS Archive™, a media management system offering powerful content-based indexing features and a well documented

provided by the University of Leipzig. The first two of these are represented in RDF according to the SKOS Core Meta Model (W3C).

### Usability:

The Factory makes use of advanced user interface paradigms for easy exploration and navigation in the repository, including reasoning and recommendations. In the prototype, hyperbolic tree navigation (based on TouchGraph) is offered for browsing the knowledge models. The detailed result view ('video



summary') offers a synchronized presentation of the video-, audio-, speech-, keyframe- and stripe image-track.

### Outlook

The project has now delivered a first vertical prototype for a knowledge-based search and retrieval scenario in an audio-visual archive. Future work will be dedicated to the extension of the core infrastructure to the needs of other 'knowledge-intensive' scenarios, such as the automatic creation of a summary of historical video clips related to a certain event or person. We therefore plan to integrate additional knowledge models

(eg temporal categories and categories related to Austrian history and policy), and to adapt the content retrieval methods to the user's intuitive understanding of 'generalization' and 'specialization' (using inferencing and reasoning based on the knowledge models).

The prototype has shown that our approach to integrating and extending existing technology with semantic features has proven feasible. However complex it seems to implement a fully featured, knowledge-based infrastructure for audiovisual repositories (and some earlier projects got lost in this

complexity), one must start with manageable and pragmatic steps to approach the vision of a 'Smart Content Factory'.

#### Links:

<http://www.newmedialab.at/>  
<http://www.iptc.org/>  
<http://www.touchgraph.com/>  
<http://www.virage.com/>  
<http://www.w3.org/2004/02/skos/>

#### Please contact:

Georg Güntner  
 Salzburg NewMediaLab, Austria  
 Tel: +43 662 2288 400  
 E-mail: [georg.guentner@newmedialab.at](mailto:georg.guentner@newmedialab.at)

## Fischlár-News: Multimedia Access to Broadcast TV News

by Alan F. Smeaton, Noel E. O'Connor and Hyowon Lee

**Fischlár-News is an operational system which provides content-based access to a growing archive of broadcast TV news.**

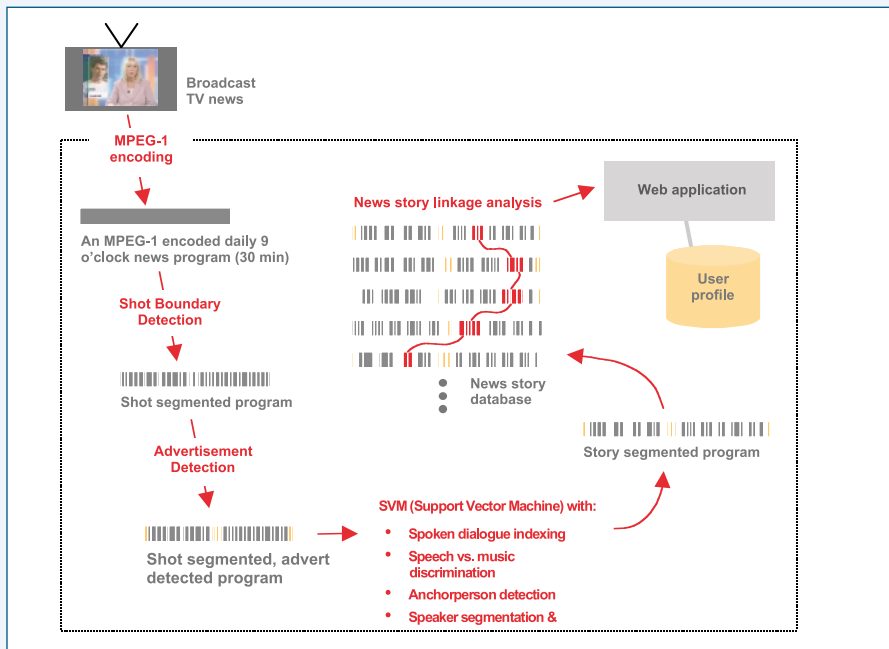
The Centre for Digital Video Processing at Dublin City University has been carrying out leading research into video analysis and structuring since the mid-1990s and this has been done with a view towards supporting content-based operations such as summarisation (movies), highlight detection (sports) and searching/browsing (TV News). Part of this work has culminated in the Físchlár-News system as part of a project which commenced in 2000. The resulting system captures broadcast TV news, nightly, and automatically analyses and structures the entire broadcast into an MPEG-7 annotation. These analysis processes include shot boundary detection, detection of keyframes, identification of the exact start/end of the program, TV advertisement detection, speech/music discrimination and automatic detection of anchorperson shots. The outputs of these analyses are fed into a trained Support Vector Machine (SVM) which segments the broadcast into discrete news stories which we use as the units for retrieval. The analysis of a broadcast news program takes about real time and the program is available on

the system about 30 minutes after its transmission.

A user uses Físchlár-News to access TV news video through a conventional web browser and when a user logs on he/she is presented with several ways in which to locate TV news stories. The most popular modality is access by date and a calendar option provides fast access to all the news stories for a given date in this way. Users can also enter keywords into a text search box which are then matched against a text representation for each news story using standard information retrieval approaches, captured via the closed captions associated with the broadcast. Once a story is located it is presented as a set of keyframes, including an anchorperson shot plus keyframes for any outside footage for that story, inter-twined with the text dialogue of that story as taken from the closed captions. For each story presented we also present a list of "related" stories which are those from the archive (currently about 10,000 stories, and growing daily) which are most similar to the story in question. This allows a user to follow a thread of stories of, for

example, a criminal event, the tracking down and capture of the criminals, their trial, sentencing, etc. It also allows a user to follow links to related stories. For example, when viewing a story of a murder in Dublin city a user would be shown links to stories of other murders in Dublin from other dates, etc. Finally, we also incorporate a personalisation and recommender system into Físchlár-News which tracks users' viewing of stories, as well as their explicit story ratings, and we use this to provide a "recommended stories" feature which is used when users have been away from the system for some time. This is also useful to bring older stories which users may have missed, or forgotten, to their attention. The diagram in the figure shows the process of news video analysis in the system and apart from some periodic sanity-checking of the performance of the SVM, the process is entirely automatic.

The Físchlár-News system has been shown to be very useful for users but it is primarily a showcase for our underlying research in multimedia content analysis and automatic annotation using



Process of News Video Analyses in Físchlár-News.

MPEG-7. It demonstrates a variety of ways in which multimedia content (digital video in our case) should be indexed and subsequently accessed. The system has been operational on the University campus for nearly 3 years and over 1,500 unique users have used the system to keep themselves informed of TV news. It is especially useful for

people when traveling as it allows them to access their own TV local news, from abroad using only a web browser and their password.

Although the research project which funded the development of Físchlár-News was finished in late 2003 we have kept the system operational, mostly

because of the demands of our users who find it too useful to be without ! The system also acts as an operational showcase of how video analysis and multimedia analysis and annotation can lead to useful systems to allow searching and browsing of that same multimedia content. Several extensions to the system have also been suggested including recording more than 1 TV news program per day, recording TV news from more than 1 broadcaster and allowing more explicit temporal browsing through topic threads. These suggestions came from an extensive user study we performed recently based on analysis of usage data, user diaries and pre- and post-study questionnaires. It is our intention to incorporate as many of these suggestions as we can.

The support of the Informatics Directorate of Enterprise Ireland is gratefully acknowledged.

**Link:**

Centre for Digital Video Processing:  
<http://www.cdvp.dcu.ie/>

**Please contact:**

Alan F. Smeaton, Centre for Digital Video Processing, Dublin City University / Irish Universities Consortium, Ireland  
E-mail: [Alan.Smeaton@dcu.ie](mailto:Alan.Smeaton@dcu.ie)

## Interactive Multimedia-Enabled Learning and Training

by Claire Kenny, Declan McMullen, Mark Melia and Claus Pahl

Multimedia technology is an ideal platform to support advanced forms of learning and training. Active learning, for instance, requires a high degree of interactivity in different forms — for which multimedia technology provides an infrastructure solution. The IDLE system — an interactive educational multimedia system supporting database learning and training — shall illustrate the benefits and supporting architectures for multimodal, interactive learning and training.

Database learning, like many other topics in various disciplines, requires an understanding of foundational concepts combined with skills that can only be obtained in a realistic environment. As part of the INVITE project — INfrastructures for VIRTUAL Teaching and learning Environments — at the School of Computing at Dublin City University, we

have developed a range of interactive multimedia features in an integrated environment — called IDLE — to support active database learning and training. IDLE, the Interactive Database Learning Environment, is a Web-based, multi-modal educational media environment used in undergraduate teaching for more than five years.

### An Educational Multimedia Architecture

Interactive multimedia features have to structure and guide the learner's access to educational content. Learning technology systems need to communicate content in the most appropriate form to a learner. Multimedia ideally suits these requirements. Learning content is a

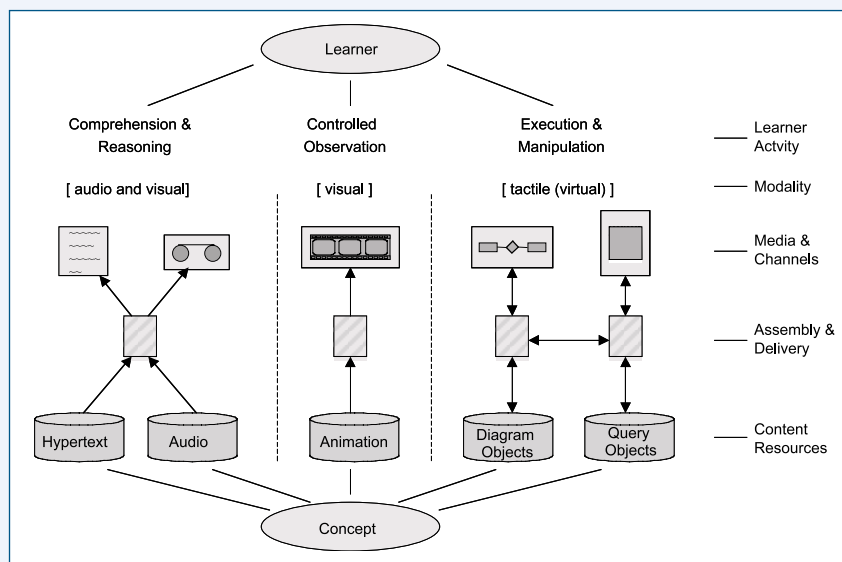
collection of stored media resources that presents the learner with different views and activities relating to the central concepts of the subject domain. At the core of the learning environment is a multimedia delivery systems that allows the learner to access content resources and to interact with content in the most appropriate, educationally sound way. The figure shows the multimedia architecture of IDLE. The main architectural elements are:

- content resources, eg spoken and written words enhanced by images, moving pictures, or active objects
- delivery infrastructure and media players, eg basic Web browser functionality (hypertext), audio player (audio stream), animation player (animations), advanced Web browser and server functionality (active, dynamic pages using applets, HTML forms, or servlets).

### Educational Multimedia Design

Concepts are at the centre of organising educational content. Usually, various perspectives on the presentation of concepts in content exist in terms of the learning and training context. Aspects such as declarative or factual knowledge, procedural knowledge, and skills form these perspectives. All three perspectives can be related to the same concept. For example, learning about database queries requires an understanding of the conceptual relational data model background and the operational aspects of query execution as well as trained skills in query formulation and execution. These different perspectives arise from the different learner objectives in relation to a given concept:

- for declarative knowledge such as data model definitions presented through synchronised audio and hypertext, learning objectives include abstraction, comprehension, and reasoning
- for procedural knowledge such as query execution presented through sequenced individual animations, controlled observation is a means to understand the operational aspects
- or skills such as query definition and formulation presented through applet- and servlet-supported active Web pages, execution and manipulation with feedback are paramount.



The IDLE Multimedia Architecture.

Consequently, the learning and training requires different forms of interaction of learners with content.

Multimodality is one of the central characteristics of comprehensive learning technology systems that enable a successful learning experience and support learning objectives. A multimodal media architecture is needed to support these learner objectives, ie, to facilitate the corresponding learning activities. We can associate learning activities and suitable multimedia modalities and channels. Abstraction, comprehension, and reasoning is usually enabled through spoken and written language, ie using audio and text media. Controlled observation is based on a visual learning experience using moving pictures or animations in a computer-supported learning environment. Execution and manipulation can be supported through an (almost) tactile form, where virtual objects are manipulated. Different media types in learning technology systems such as IDLE enable interactivity in order to support successful active learning.

### Authoring of Interactive Educational Multimedia

The importance of multimedia in learning and training requires a systematic Interactive Educational Multimedia (IEMM) engineering approach. Learning and training activities need to be mapped to the human-computer interface and

implemented through multimedia features. Due to the complexity of the domain, the design of channels and interaction languages is ideally supported in a domain-specific framework. We have used education-specific channels – such as declarative knowledge, procedural knowledge, skills, learning sequencing, feedback, and coaching – as domain-specific channels in the design of IDLE and other learning technology systems. Each of the channels is different in the way learners use and work with the content that is communicated over the channels.

### Conclusions

The benefits of a multimodal learning experience that enables a variety of content interactions are undisputed. The complexity of designing these experiences and authoring IEMM content requires a domain-specific approach using educational channel notions as abstractions and particular media types to enable certain learner-content interactions. Only such a development and authoring approach will help to reduce the high development costs associated with educational content development and to make multimedia learning objects more reusable.

#### Please contact:

Claus Pahl,  
Dublin City University, School of Computing /  
Irish Universities Consortium  
Tel: +353 1 700 5620  
E-mail: Claus.Pahl@dcu.ie



# Representation and Communication of Multimedia Data and Metadata

by Sara Colantonio, Maria Grazia Di Bono, Massimo Martinelli, Gabriele Pieri and Ovidio Salvetti

In recent years the increasing role of Multimedia (MM) data, in the form of still pictures, graphics, 3D models, audio, speech, video or their combination (eg MM presentations), in the real world, has lead to a demand for better procedures for the automatic generation and extraction of both low level and semantic features from multi-source data in order to enhance their potential for computational interpretation and processing.

MUSCLE (Multimedia Understanding through Semantics, Computation and Learning) is a European Network of Excellence (NoE) that aims fostering close collaboration between research groups in multimedia data mining and machine learning. Within MUSCLE, our research is focused on investigating standards and tools that allow interoperability of heterogeneous and distributed (meta)data also by enabling data descriptions of high semantic content (eg ontologies, MPEG-7 and XML schemata) and inference schemes that can reason about these at the appropriate levels.

Metadata are used to represent the value-added information that describes the technical and semantic characteristics associated with MM data. Metadata make data more processable, allowing more efficient retrieval or classification, quality estimation and prediction based on Machine Learning techniques in both single and multiple-modality.

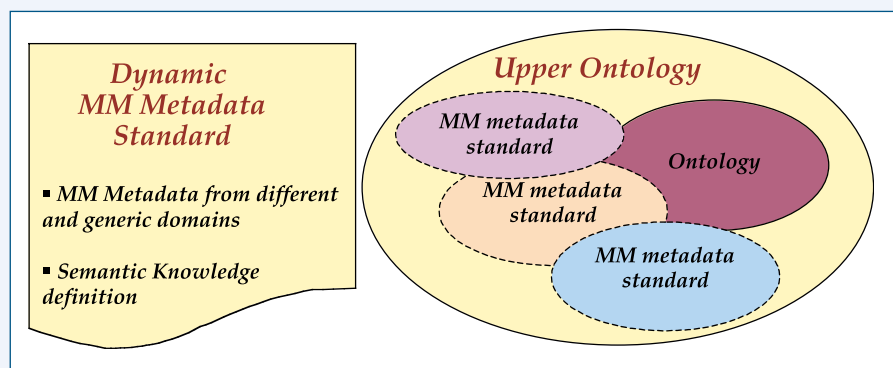
Many initiatives for metadata standardisation have been proposed in order to describe multimedia content in various domains. Scientific and industrial communities tend to create their own standards tailored on their particular needs. This could cause an unrestricted growth in the number of available standards making the integration and sharing of MM data between different communities (vision, speech, text, ...) very difficult. A recent approach is to combine a specific MM metadata standard with other standards that can be used to describe similar application domains, in order to provide a more comprehensive

characterisation of heterogeneous MM data without creating a new standard.

From a recent survey of the state-of-the-art, we have identified two main approaches to MM data processing. On the one hand, people who employ MM data for scientific purposes use consolidated MM data processing algorithms, on the other hand, applications following a Content Based Query (CBQ) paradigm require content representation.

to “represent” and “translate” different metadata sets whose elements are correlated on the basis of the same or similar meanings so that MM applications can use ontology knowledge in addition to the metadata (see the figure).

In the first case, one strategy could be to use MPEG-7, which is currently the most mature MM metadata standard, due to its generality and extensibility. MPEG-7 permits an extensive descrip-



Two possible solutions for MM data integration and interoperability: dynamic single metadata standard (left) or definition of an upper ontology (right).

This scenario, which involves MPEG-7 or metadata models tailored on the specific requirements of a given community, highlights a possible limit for interoperability among different communities. We feel that two important issues must be considered in order to achieve an efficient and integrated use of MM metadata: (i) a common MM standard format able to describe and represent the intrinsic heterogeneous nature of MM data and their semantics must be defined; (ii) more abstract models (eg ontologies) and the related mapping tools are needed

tion of multimedia content not only at the low-level feature level (visual, audio, multimedia) but also at the higher semantic level. However, its freedom in terms of structures and parameters is such that, in general, it is not easy to interpret MPEG-7 produced by others. To overcome this problem, a possible solution could be to build an MPEG-7 ontology.

In the second case, the introduction of a high-level ontology covering multiple domains could be convenient; this solu-

tion would have the advantage of being more independent of the lower metadata standards. However, the definition of a completely new ontology is a very complex task.

When constructing ontologies, standard tools like XML, RDF and OWL should be considered. XML supports the definition of constraints for structure, cardinality and data-types but does not support the definition of semantic knowledge. RDF provides mechanisms to describe MM resources, group them and

represent their semantic relationships, but does not offer mechanisms to define general axioms, which can provide a stronger semantic representation. OWL can be used to derive logical semantic consequences but has expressivity limitations, which could be overcome by using specific rule languages (eg RuleML, SWRL, ORL, ...) extending OWL axioms with additive and more expressive rules.

The use of a higher level ontology would seem the best way to approach the

problem of offering a simpler high level access to MM data for processing and interpretation purposes. We are now investigating this solution.

**Link:**

<http://www.muscle-noe.org> (Workpackage 9)

**Please contact:**

Ovidio Salvetti, ISTI-CNR, Italy

Tel: +39 050 3153124

E-mail: [Ovidio.Salvetti@isti.cnr.it](mailto:Ovidio.Salvetti@isti.cnr.it)

or Eric Pauwels,

MUSCLE Scientific Coordinator

CWI, The Netherlands,

Tel: +31 20 592 4225

E-mail: [Eric.Pauwels@cwi.nl](mailto:Eric.Pauwels@cwi.nl)

## A Cognitive Architecture for Semantically Based Medical Image Retrieval

by John Moustakas, Socrates Dimitriadis and Kostas Marias

**The automatic extraction of meaningful image semantics is an important step towards the development of intelligent systems for Content-Based Image Retrieval (CBIR). Such systems have the potential to become useful clinical decision support tools, by retrieving medical images with established diagnoses that are 'similar' to the images the clinician must read. In the Institute of Computer Science - FORTH, we are developing and implementing experimental platforms for the investigation of CBIR, based on biologically inspired multi-agent architectures. In this news article, we present a novel platform based on a two-level architecture inspired by human cognitive mechanisms. These two levels share the computation of generic similarity and medical image semantics.**

Humans undoubtedly possess the ability to process visual information efficiently and to identify images as being similar based on their visual content. However, computational approaches currently fall short of matching this ability. At FORTH-ICS we aim to develop and implement CBIR mechanisms that are perceptually motivated and based on biologically inspired architectures. Our recent work is concerned with medical image retrieval and aims to provide a reliable framework that can be customized for several imaging applications, and potentially be combined with DICOM functionality as well. Crucial to this goal is the automatic extraction of semantic information from medical images (eg asymmetry or pathology detection). Nevertheless, the semantic content of images is subjective, and depends on the specific image class. For this reason, generic similarity CBIR approaches often fall into the 'semantic gap' problem, meaning that the computed

features can't always properly describe the real characteristics of the image. At FORTH-ICS we developed a novel two-tier CBIR platform inspired by the human cognitive architecture. The key idea underlying our work emanates from psychological and neuroscientific studies which indicate that the human visual system processes information in several stages.

The visual system retains independent retinotopic maps for different primitive visual features (colour, form etc). In a pre-attentive or early stage of vision, the processing on these feature maps is undertaken independently and in parallel, whereas in the subsequent attentive stage the visual modalities engage in cooperative work. In other words, the pre-attentive level decomposes the optical scene in its primitive characteristics, which to a large extent are processed independently, in parallel and autonomously.

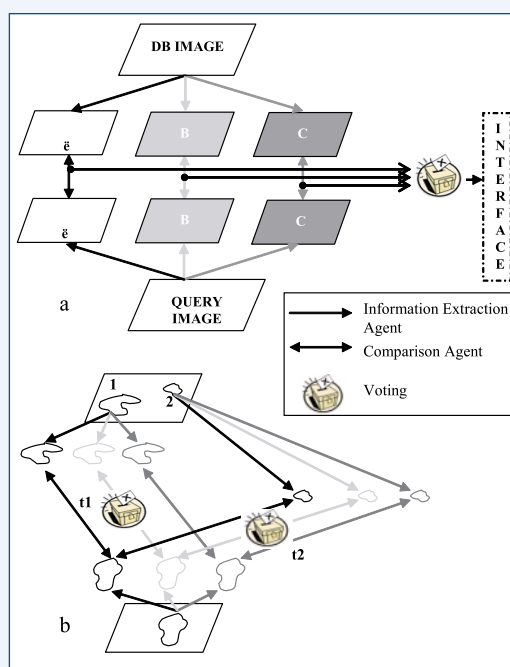
After the first stage of fixed-time pre-attentive processing, the human visual system performs a serial and selective examination of semantic objects that draw the subject's attention – that is, the attentive level of perception.

Based on this biological paradigm we developed a two-tier CBIR platform featuring both a pre-attentive and an attentive level of retrieval by extending our previous work on agent-based single-tier CBIR (ERCIM News No. 53, April 2003). In order to be able to assess the value of the proposed architecture, our platform was customized for a specific domain, ie brain MRI image retrieval. The pre-attentive layer of the proposed architecture produces independent, parallel feature maps (A, B, C, in Figure), each coding an independent visual feature. During the retrieval stage, each autonomous agent compares the computed values between the query and

each database image. The comparison scores from all relevant agents are driven to the voting system, resulting in the final score for the candidate retrieval image. The voting scheme is selected by the user.

An additional, 'attentive' layer is designed for the agents to receive, one by one, semantic regions of interest (ROIs). A specialized group of 'attentive' agents then carefully examines and compares ROIs in a serial fashion (first '1', then '2', and so on). In our implementation, the attentive similarity of a given pair of MRI images is defined as the similarity of their 'closest' pair of ROIs. It is obvious that in order to implement the attentive retrieval level, the semantics must be defined for the specific application (brain MRI retrieval), since it is still difficult to automatically define important semantics in any image class without incorporating any prior knowledge.

The definition of semantic regions for brain MRI retrieval was based on novel algorithms for brain symmetry detection. It is well known that a normal human brain exhibits a remarkable degree of symmetry with respect to the mid-



**Two-tier architecture for medical image retrieval consisting of a pre-attentive (a), and an attentive level (b).**

sagittal plane. In addition, the identification of regions of asymmetry is often indicative of diseases such as schizophrenia, epilepsy, and Alzheimer's. We developed algorithms for segmenting and analysing asymmet-

rical regions for the 'attentive' level of CBIR. The authors will report initial retrieval results on publicly available data ('The Whole Brain Atlas' from <http://www.harvard.edu>) at the forthcoming IEEE International Conference on Multimedia & Expo (ICME2005).

The CBIR system was developed in JavaTM, and makes use of the Java Advanced Imaging package. It is fully scalable and can be easily extended with additional agents or voting schemes in order to take into account the specific requirements of different experiments and classes of images. While the attentive level of retrieval can be customized for any application, provided that semantic features can be automatically defined, this is an extremely hard task. Researchers across disciplines have tried for many years to shed light on the mechanisms of decomposing any image to its primitive visual features, and extracting the true underlying semantics.

**Please contact:**

Kostas Marias, ICS-FORTH, Greece  
Tel: +30 2810 391696  
E-mail: [kmarias@ics.forth.gr](mailto:kmarias@ics.forth.gr)

## Personalized and Adaptive Multimedia Retrieval

by Joemon M. Jose and Jana Urban

Data accumulation has become an integral part of our life, be it text (eg email) or multimedia data (eg photographs). This is mainly due to the proliferation of computers, networking devices and consumer devices like digital cameras and camcorders. However, our creation prowess is not matched by any comparable search facilities. As a result, finding relevant information from such archives becomes a cumbersome process. Even after a decade of research in content-based image retrieval (CBIR), users must plough manually through their media archives in order to find what they need.

There are a number of reasons for this scenario. The major one is that the state-of-the-art in image processing does not allow us to identify and extract any meaningful segments from images. Images are still represented using low-level features. However, the images' low-level feature representation does not reflect the high-level concepts the user

has in mind (semantic gap) and – partly due to this – users can experience serious difficulties in effectively formulating and communicating their information need (query formulation problem).

Any reasonable solution to CBIR should address the issue of semantic gap, and providing help with query formulation is

one way to do this. In our approach to image retrieval, we employ an adaptive scheme for image retrieval.

### Adaptive Models for Image Retrieval

Difficulties with query formulation are addressed using an adaptive query learning scheme and an innovative



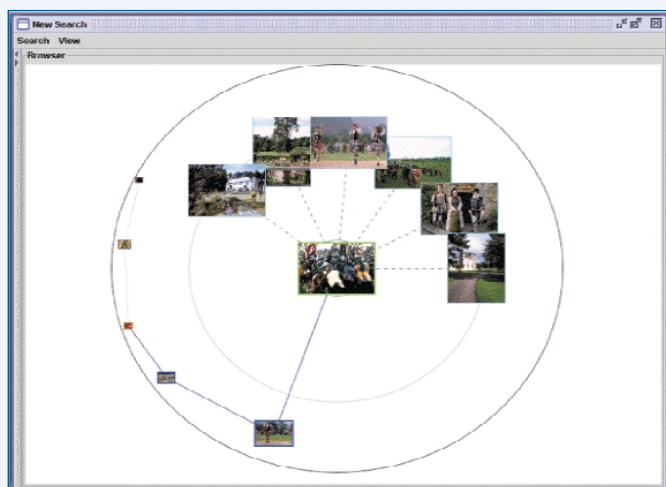


Figure 1: An interface of the system.

search interface. We have developed an interface (see Figure 1) in which a user starts browsing with one example image. Subsequently, a new set of similar images is presented to the user.

As a next step, the user – through selecting one of the returned images – updates the query, which now consists of the original image and the selected image from the returned set of images. After a couple of iterations the query is based on a set of images. In this scheme, the retrieval process is iterative: updating the system's knowledge of the user's information need based on the user's implicit feedback. In the underlying interaction model, the user builds up a browsing tree of interesting images by choosing one image from a recommended set to be appended to the browsing path in each iteration. The system's recommendations are then based on a query constructed from the current path of images.

In this approach, the emphasis is placed on the user's activity and the context, rather than any predefined internal representation of the data. A path represents a user's motion through information, and taken as a whole is used to build up a representation of the instantaneous information need.

In a nutshell, our method supports both browse-based and query-based approaches. It supports a query-less

interface, in which the user's selection of an image is interpreted as evidence of its relevance to her/his current information need. It therefore allows direct searching without the need for formally describing the information need. For the query, each image in the path is considered relevant, but the degree of relevance is dependent on the age: it decreases over time when new images are appended. In this way, the retrieval model is a special kind of relevance feedback model, in which a query is implicitly refined by the user's selection of images for feedback. It recognizes and addresses the dynamic nature of information needs, and has the advantage of allowing for an intuitive and user-centred search process.

A user study involving design students demonstrated the effectiveness of this approach. The evaluation showed that people preferred the search process in the ostensive browsing scheme, felt more comfortable during the interaction, and found the system more satisfactory to use compared to a traditional CBIR interface.

### Image Retrieval by Group Organization

Another approach to providing effective image retrieval is to develop mechanisms to support the handling of the retrieved set of images. This will allow users to handle retrieved sets of images more easily by organizing them into relevant groups. The system can then

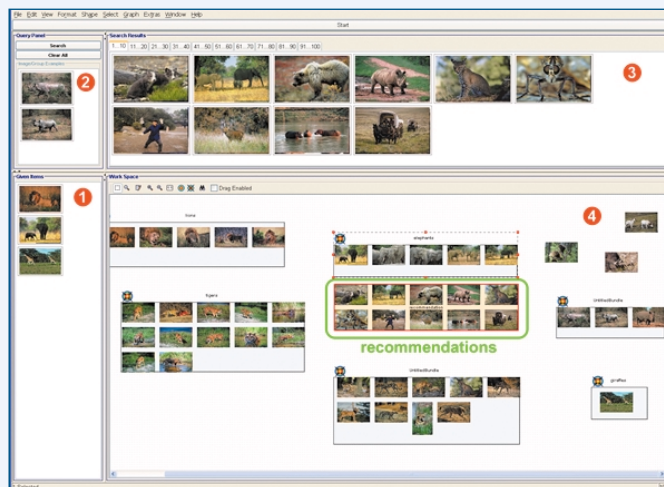


Figure 2: The EGO interface.

capture the context of the user, and will thus be able to recommend images more tuned to their needs.

We introduced a system called EGO (Effective Group Organization), which facilitates retrieval in context (see Figure 2). EGO is a system for the management of image collections, supporting the user through the process of personalization and adaptation. In this system, we stress the need for a search system that provides flexible and extensible interfaces. The idea is to help the user in query formulation: by facilitating qualitatively high interaction, the system learns the needs of its users and can provide searches accordingly.

Our approach encourages the user to group and organize their search results and thus provide more finely grained feedback for the system. It combines the search and management process, which helps the user to conceptualize their search tasks and to overcome the query formulation problem. The system assists the user by recommending relevant images for selected groups. The user can therefore concentrate on solving specific tasks rather than having to think about how to create a good query in accordance with the retrieval mechanism.

#### Please contact:

Joemon M Jose, University of Glasgow, UK  
E-mail: jj@dcsg.gla.ac.uk

# ADMITS: Adaptation in Distributed Multimedia IT Systems

by Laszlo Böszörményi

The ADMITS project, at the Institute of Information Technology of the University of Klagenfurt (ITEC), has developed an experimental distributed multimedia system for investigations in adaptation, an increasingly important tool for resource and media management in distributed multimedia systems.

In recent years, a large number of papers have been published on the subject of adaptation in multimedia systems. These papers reflect a diversity of interpretations, however, generally have a rather limited scope. Best-effort scheduling and worst-case reservation of resources are two extreme cases, neither being well suited to cope with large-scale, dynamic multimedia systems. A middle course can be found in a system that dynamically adapts its data, resource requirements, and processing components to achieve user satisfaction. Nevertheless, there is no agreement about the questions of where, when, what and who should adapt.

A distributed multimedia system comprises several types of components, such as media servers, meta-databases, proxies, routers and clients. In addition, a large number of adaptation possibilities exist, from simple frame dropping up to virtual server systems that dynamically allocate new resources on demand. The main problem is determining which kind of component can best be used for each kind of adaptation.

In the frame of the ADMITS project, we are seeking answers to exactly this basic question, and to a number of related questions. In building the experimental system, we explore a number of possible adaptation entities (server, proxy, clients, routers), and implement and evaluate different algorithms for media, component and application-level adaptations. Experimental data is also collected in order to gain insight into when, where and how to adapt, as well as how individual, distributed adaptation steps interoperate and interact with each other. The overall architecture of the experimental system is depicted in the Figure.

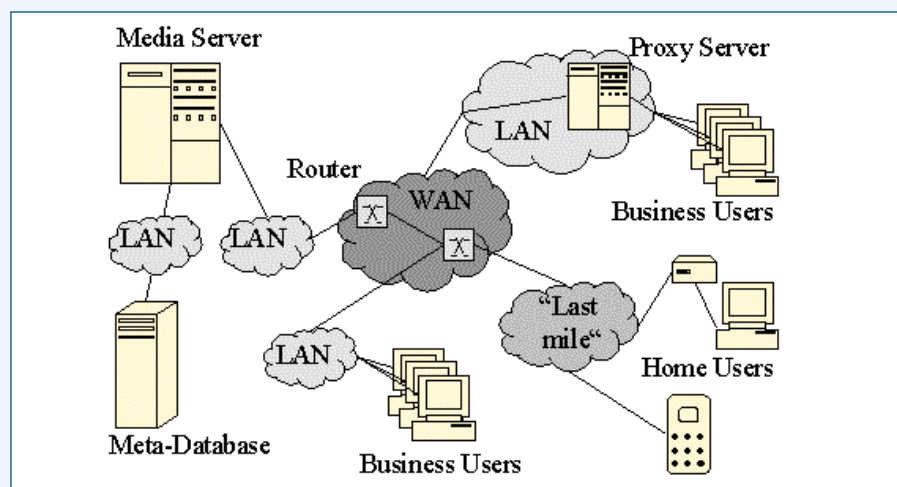
The individual components play different roles in the adaptation process. They are connected physically by the network, and 'semantically' by MPEG-7 Multimedia Descriptions and MPEG-21 DIA Descriptions that flow over the network. The reliance on these international standards makes the components interoperable with any standard-compliant components developed elsewhere. To our knowledge there exists no other system that handles adaptation in

such a comprehensive and interoperable way.

Since ADMITS is a research project, its main emphasis is on the publication of papers. A list of these can be found on the home page of ITEC (<http://www.ifi.uni-klu.ac.at/ITEC/>), under 'Publications'. However, in order to validate our ideas, but also with the aim of future industrial applicability, a large repository of software tools (called ViTooKi, for Video Toolkit) has been developed. The software tools are highly interoperable and available as open-source software (<http://vitooki.sourceforge.net/>). The repository is unique in providing a very rich set of interoperable video tools. The following non-exhaustive list gives an overview of the tools and their major features:

- **Media server:**

- standard compliant media streaming by using RTSP and RTP/UDP
- streams all media formats supported by the ffmpeg library (eg MPEG-1, MPEG-2, MPEG-4)
- communicates terminal capabilities of the client device and user preferences using standardized MPEG-21 descriptors
- supports real-time adaptation of media content according to the clients' terminal capabilities, the user preferences, and the available network resources; for example, mobile devices get a lower stream quality than high-performance workstations with good network access
- implements standardized RTP extensions to allow intelligent retransmission of lost video frames where necessary
- can be run in a distributed environment that supports proactive service and content replication and migration



The Experimental Distributed Multimedia System.

operations; this is especially helpful when content adaptation steps are not allowed due to legal constraints or the user insists on the original stream in its full quality

- supports proactive adaptations by actively measuring and forecasting available server and network resources on and between server nodes.
- **Proxy server:**
  - incorporates both a server and a client implementation (since a proxy must act as a server to the client and a client to the server)
  - caches elementary streams in different quality versions
  - implements quality-aware replacement strategies
  - can be dynamically relocated in the vicinity of requesting clients.

- **Meta-database:**

- multimedia database schema based on the MPEG-7 standard
- multimedia indexing framework
- cost-based query optimization for range and k-nearest neighbour searches
- application-level libraries for content-based image retrieval systems, audio recognition tools, video browsing tools, and quality aware MPEG-4 proxies.

- **Media player:**

- standard compliant control of RTP-based media streams by using RTSP
- supports parallel presentation of many videos in different viewers, in different qualities
- implements a general framework for SMIL- and BIFS-based multi-scene presentations.

One of the most important and unique characteristics of the ADMITS project is the consequent combination of fundamental research with international standards. To reach this goal, ITEC has decided to participate in the standardization processes of the Moving Picture Experts Group (MPEG), especially in the area of MPEG-7 and MPEG-21. A number of contributions in the context of Digital Item Adaptation (DIA) have been submitted by ITEC (in partial collaboration with Siemens AG) and accepted by MPEG in the recent years.

**Link:**

<https://www.ifi.uni-klu.ac.at/ITEC/>

**Please contact:**

Laszlo Böszörményi  
Institute of Information Technology (ITEC),  
Klagenfurt University, Austria  
Tel: +43 463 2700 3611  
E-mail: laszlo@uni-klu.ac.at

## Solutions for an Interpreter-Enabled Multimedia Conferencing System

by Ferenc Sárközy and Géza Haidegger

As video conferencing is likely to become a commonplace tool in the near future, a demand may also arise for an interpreter-enabled video conferencing system. Despite the fact that current technology is able to accomplish this, no such product is actually available. The Global Conference Network project aimed at establishing a distributed, multi-lingual, multimedia conferencing system. Research and development work at the CIM Research Laboratory of SZTAKI contributed to this project.

Video conferencing has become a widely used tool, since it is available to anyone with a mainstream computer and a broadband (xDSL, cable) Internet connection. Despite telecommunication service parameters having improved remarkably, they still represent a bottleneck in some professional applications of video conferencing systems. Progress in this area has slowed recently, mostly for business-related reasons. Consequently, it is hoped that the market push will make further development probable, as has been the case in Japan. The enhancement of the local loop bandwidth and other QoS (quality of service) parameters can open up the way for such demanding applications as interpreted video conferences (IVC).

The Global Conference Network (GCN) system attempts to establish audio and video connections independently from language and actual place of residence. In order to achieve this general goal, GCN has to meet the following requirements:

- to integrate interpreters in the conference: this is the fundamental difference between GCN and currently available video conferencing solutions
- to take over the role of existing local conference systems as the software of standard hardware elements
- to integrate all kinds of participants through an appropriate Internet connection
- to integrate local network-based conferences and individual participants through the Internet.

The most important advantages of an IVC system are the following:

- to widen the applicability of video conferencing itself
- to reduce the cost of conferences by integrating remote participants, particularly interpreters
- to increase potential audiences for e-learning through interpretation.

The GCN system has three main functional parts. Firstly, the GCN database (GCNDB) contains all the information about any conferences organized with the help of that GCN system. This is a MySQL database and it is maintained by the GCN Web Server (GCNWS). Secondly, GCNWS is a PHP-based Web application. It provides Web pages that



handle the entire life cycle of the conference data, but the emphasis is on pre- and post-conference functionalities. The conference structure is created during the set-up process and after the conference, the GCNWS can be used as an archive viewer.

The third part of the GCN system - the GCN Applications (GA) - is an extendable set of applications which interact with each other and with the floor control messages of GCN. The base set of these and their roles are the following:

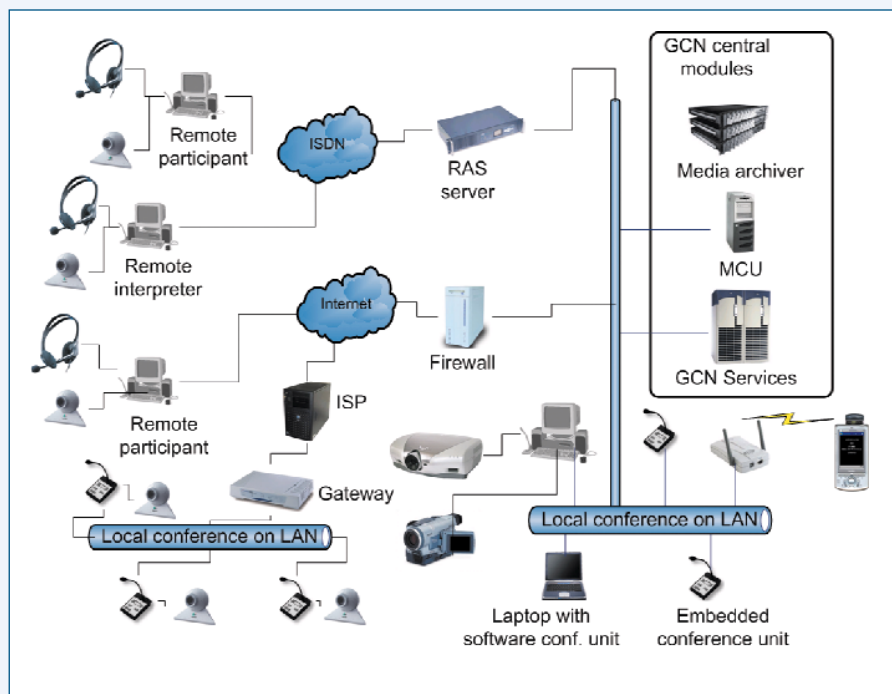
- a Conference Organizer (CO), which stores the conference's actual status and can be queried by client applications to initialize themselves after login or reconnect, authenticates and authorizes the users, manages the voting sessions and archives GCN messages
- client applications for the different kinds of participant roles, such as organizer, chairman, interpreter, speaker and observer
- interfaces to the media handling applications.

Our aim was to create a portable solution, so all these components are realized on a Java platform.

Media-handling tools are loosely integrated into the GCN system. At the beginning of the development none of the available tools had a distinct advantage over the others. Since we did not want to make such an important commitment at the beginning, so we designed GCN clients to be able to include different kind of media-handling plug-ins.

As most standards and standard protocols in tele- and video-conferencing have some deficiencies, we must also deal with protocol development.

The Open H.323 project aims to create an open-source implementation of the H.323 teleconferencing protocol. H.323 and therefore the Open H.323 project offer a solution based on a Multipoint Control Unit (MCU). One of our media-handling solutions is based on this open-source software.



General configuration of the GCN system.

Using an MCU leads necessarily to a centralized system. JMF (Java Media Framework) and the VideoLAN project (see below) allow a decentralized, media server-less solution. The requirement is that the underlying network should be multicast-enabled. JMF allowed us to implement a distributed media-handling solution. However, during the implementation, an unforeseen bug in JMF caused some problems.

The Open Source VideoLAN project aims to provide a video and audio streaming tool. One of the result applications of VideoLAN is an open-source program, named 'VLC', which is a media-processing unit. It can handle codecs that scale very well, and so works also on low bit-rates. VLC can stream with RTP, either unicast or multicast. VLC itself is a stand-alone application, but offers several possibilities for programmed access to its services.

The solutions based on VLC and H.323 worked equally well on a local area network. VLC provides much better quality, but also requires more processing power and is based on a lower level standard. On an entry level ADSL connection (384/128kbps), the

upload bandwidth was a serious bottleneck for VLC. In the download direction the bandwidth was sufficient, and both image and sound quality were satisfactory. The H.323-based solution provided the quality that two NetMeeting clients can achieve on this connection. During the modification of the MCU we developed the know-how for the integration of better codecs into the system, and as a consequence the final solution will be based on the H.323 standard.

The project leader, DIGITON Ltd, is an innovative Hungarian SME, working on solutions to match future, novel needs. This project inspired new ideas and plans, and the consortium is open to further scientific challenges.

#### Links:

<http://www.videolan.org>  
<http://www.digiton.hu>  
<http://www.sztaki.hu/sztaki/divisions.jhtml>

#### Please contact:

Ferenc Sárközy, Géza Haidegger,  
 SZTAKI, Hungary  
 Tel: +36-1-2796207  
 E-mail: [sarkozy@sztaki.hu](mailto:sarkozy@sztaki.hu),  
[haidegger@sztaki.hu](mailto:haidegger@sztaki.hu)

# Host Recommendation in the Adaptive Distributed Multimedia Server

by Ottó Hutter, Tibor Szkaliczki and Balázs Goldschmidt

The Adaptive Distributed Multimedia Server (ADMS) developed at the Klagenfurt University is able to add and remove its components to and from different nodes of the network. This novel feature of the multimedia server enables the dynamic placement of the server components according to the current requests and Quality of Service (QoS) parameters of the network. Researchers at SZTAKI and the Budapest University of Technology and Economics have developed and implemented algorithms for the optimal placement of nodes for hosting the ADMS components in the network.

The Adaptive Distributed Multimedia Server (ADMS) was developed at the Klagenfurt University. It is an adaptive architecture for distributing video streams through networks covering large territories, and manages both network and node overloads by migrating and replicating server components in the network. The Host Recommendation module developed at the Budapest University of Technology and Economics (BUTE) is responsible for finding proper placement of the components according to the dynamically changing network parameters and client demands. SZTAKI is also participating in the development of the Host Recommendation module, in cooperation with the Klagenfurt University and BUTE. The module was tested as a stand-alone in a simulated network environment using randomly generated client requests, and is integrated into the

ADMS at the Klagenfurt University for experimental tests.

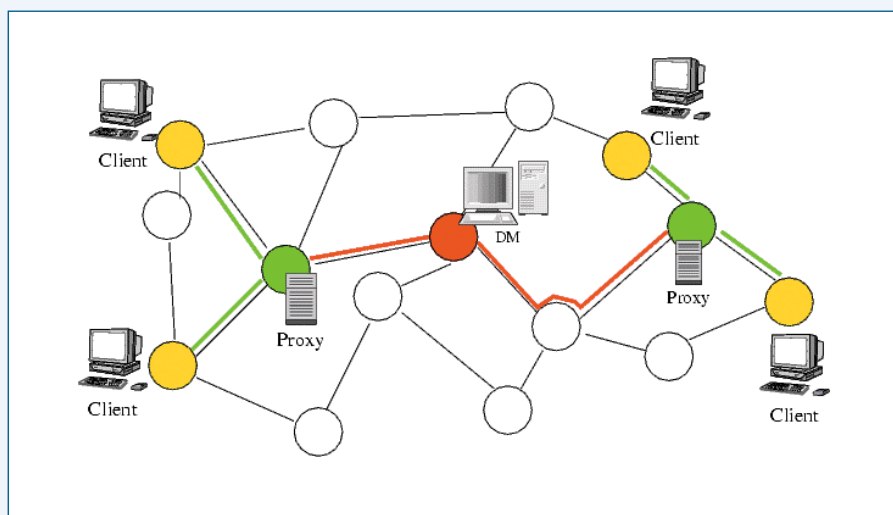
First, we concentrated on where to put proxies in the network. The task was to find suitable locations for proxies of the distributed multimedia server, while maximizing the clients' satisfaction and minimizing the network load. The clients receive the same video in parallel.

Proxy and cache placement is a heavily studied area, but we found significant differences between the placement of Web proxies and ADMS proxies. In our case, the clients do not need accurate delivery but rather strict Quality of Service. Moreover, ADMS delivers huge media streams instead of small files. Nodes hosting server applications in a network are usually assigned in a static way. However, this approach is not applicable to the ADMS. The delivery of

the data-streams with proper quality can start only after the placement of the server nodes. As a consequence, the dynamic reconfiguration of the server requires an automatic host recommendation whose running time should comply with strict time constraints. The Technical University of Budapest began work on this problem by analysing two methods, namely the greedy algorithm and the particle swarm algorithm. SZTAKI joined the server development by proposing and implementing two other methods (linear programming rounding and incremental algorithm) in 2003 and 2004.

The greedy approach was the first algorithm to be implemented, and has proven to be successful in many similar problems. The particle swarm algorithm is a kind of evolutionary algorithm. The incremental algorithm finds an initial solution as fast as possible and then the algorithm incrementally improves it, complying with the time constraints in order to approximate the optimal placement. Another approach is the linear programming rounding technique, where the integer programming formulation of the problem is derived, its linear programming relaxation is solved and the results are rounded.

We compared the results gained by running the implementations of the four proposed algorithms on different test networks. The greedy algorithm was clearly the worst of the four, while the particle swarm algorithm produced the best result. In terms of running time, the incremental algorithm was the best. There was no single winner from the



Our goal is to find a suitable location for multimedia proxies and data managers in a network.

particle swarm, linear programming and incremental algorithms; they can be applied in a smart combination.

After the stand-alone tests, our next task is to integrate this module into the ADMS and to test it in an experimental environment before the multimedia service becomes available for a wider audience. This year, the three partners participating in ADMS development started a new project to complete the server and to introduce pilot applications. This is running in cooperation with two SMEs within the framework of the Hungarian Economical Competitiveness Operative Programme.

The Klagenfurt University implemented and evaluated a mechanism called Client Behaviour Prediction, which determines the most likely locations of client

requests in the near future. This mechanism enables the server to execute the time-consuming process of the placement of the data managers and videos in advance, without significant delay in serving the client requests. For this reason, the placement of the videos recently came into the focus along with that of the proxies.

Numerous online training materials contain videos that can be streamed to the user. SZTAKI intends to apply the research results to e-learning applications containing video-based courseware. Our objective is to create an Adaptive Learning Management System through the integration of standard Learning Management Systems (LMS), our own standardized courseware repository, and the ADMS platform. This integrated system delivers the videos

appearing in the training materials through the ADMS. Student behaviour in requesting the training videos can be modeled relatively well. For this reason, the student and course data gained from the LMS can significantly facilitate the prediction of the client requests.

**Link:**

<http://143.205.180.128/ITEC/Publications/pubfiles/pdf/2004-0005-BGAT.pdf>

**Please contact:**

Ottó Hutter, SZTAKI, Hungary  
Tel: +36 1 279 6191  
E-mail: [hutter@sztaki.hu](mailto:hutter@sztaki.hu)

Tibor Szkaliczki, SZTAKI, Hungary  
Tel: +36 1 279 6172  
E-mail: [sztibor@sztaki.hu](mailto:sztibor@sztaki.hu)

Balázs Goldschmidt, Budapest University of Technology and Economics, Hungary  
Tel: +36 1 463 2649  
E-mail: [balage@inf.bme.hu](mailto:balage@inf.bme.hu)

## Scalable Audio Streaming to Mobile Devices

by Jonathan Sherwin and Cormac J. Sreenan

**Telecoms operators are introducing new audiences to video and audio streaming as they roll out their 3G mobile phone technology throughout Europe and elsewhere. However, current implementations are really only the beginning, limited as much by commercial as by technical considerations. In comparison to the Internet, mobile environments will pose significant challenges for the quality delivery of streaming media. The use of scalable audio streaming offers promise for adapting quality in the face of mobility, but to date there has been a dearth of practical proposals and implementation experiences.**

A project at University College Cork in Ireland has resulted in an experimental platform for scalable audio streaming to mobile endpoints. This research was supervised by Professor Sreenan and carried out for a research Master's degree by Jonathan Sherwin, a Lecturer at Cork Institute of Technology. In the near future, a mobile device will be capable of connecting to many different wireless networks, eg 3G, GPRS, 802.11, Bluetooth. The choice of which interface to activate depends on factors such as availability, cost, and bandwidth, as well as coverage areas and mobility. The research addresses the issue of how to manage audio quality in such an environment.

Two key technical challenges arise due to variations in delay and variations in band-

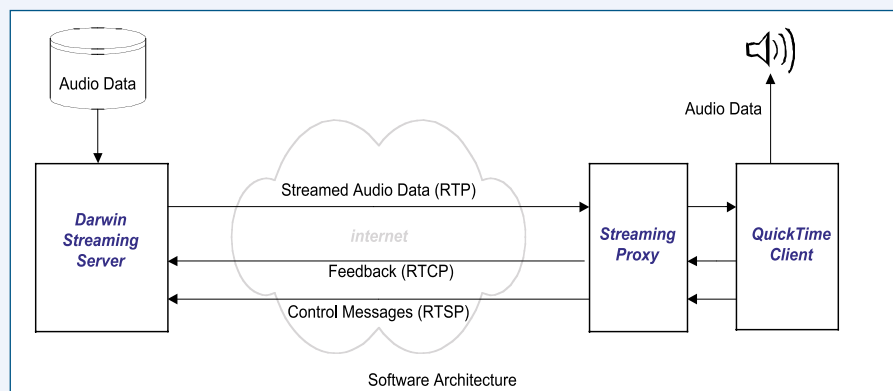
width. Streaming applications require uninterrupted playback that is isolated from the vagaries of network jitter. In the Internet this is commonly achieved using a playback buffer, but moving between networks can result in much larger delays due to handover protocols and the need to re-establish the flow of streamed data. In regard to bandwidth variability, in the Internet various techniques are used to determine the bandwidth available to a user and to adjust to mild and infrequent changes. But in a mobile environment extreme changes of bandwidth are likely as a mobile device roaming between heterogeneous networks.

It was decided to design and implement a platform based on open standard protocols, and to implement an application-layer solution to mobility issues,

meaning that the software should be able to run on a device connected to any network through which a TCP/IP connection could be established with the streaming server. The design made use of Apple Computer's Darwin Streaming Server (DSS) and their QuickTime Player. These use the standard protocols for streaming - RTP/RTCP and RTSP. The approach taken was to add functionality directly to DSS since source code is made available by Apple, and indirectly at the client end of the network by implementing a client-side proxy since the QuickTime Player is closed-source (see the figure).

The purpose of the proxy is to shield the client from the effects of changes of bandwidth or moves between networks. Achieving this initially required addition





Software Architecture.

of support for mid-stream scalability to DSS and implemented in the streaming proxy. Then, components were added to the proxy to enable bandwidth monitoring and probing, and to re-establish the connection to the server in case of a move between networks. A buffer manager was designed for the proxy to maintain a continuous supply of data to the client even during a move between networks, and to replenish the buffer to a minimum level by varying the rate of transmission from the server as necessary.

The implementation was tested using a network emulator and some interesting results and insights were obtained.

Firstly, maintaining continuous playback to the user during a move between networks was demonstrated to be feasible. However the duration of data in the buffer depended on a number of factors, most critically the length of time for which the mobile device is not connected to any network, and the length of time it takes to obtain an IP address on the new network. For example, obtaining an IP address via DHCP can take several seconds. These effects were quantified as part of the experiments.

Secondly, managing the buffer in the face of changing levels of bandwidth is a complex problem. For example, if an extreme reduction in available bandwidth

is detected – eg by falling levels of buffered data – the proxy requests a lowering of the quality of data sent from the server. However, before the server has received the request and acted on it, a certain amount of data is already buffered within the network, awaiting delivery. While this data trickles through, the proxy's buffer level continues to fall, possibly to a critical level. The implementation has allowed clear observation and analysis of this problem.

The key contributions of the project are firstly, an open-source standards-based scalable streaming platform, and secondly a set of valuable results and practical insights. Areas of future activity include the incorporation of a passive method of bandwidth measurement and an adaptive predictive control method to better cope with varying levels of network delay.

**Link:**

<http://www.cs.ucc.ie/misl>

**Please contact:**

Jonathan Sherwin, Cork Institute of Technology, Ireland.  
E-mail: [jsherwin@cit.ie](mailto:jsherwin@cit.ie)

Cormac J. Sreenan, University College Cork / Irish Universities Consortium, Ireland.  
E-mail: [cjs@ucc.ie](mailto:cjs@ucc.ie)

## A Networked Approach to TV Content Distribution

by Adrian Cahill, John Roche and Cormac J. Sreenan

**The combination of broadband access and packet internetworking removes the existing spatial constraints for TV viewing by allowing residential users to obtain content that originates in any country. Allied with high-capacity networked storage, the temporal constraints can also be removed by ensuring ubiquitous access to a comprehensive library of stored TV content. This compelling vision presents several important research challenges in the area of efficient and scalable content distribution networks for high-quality streaming media.**

Research taking place at University College Cork in Ireland under the direction of Prof. Sreenan is creating the elements that are needed to make this vision a reality. For his PhD degree, Adrian Cahill is investigating novel approaches for cost-effective management of TV content on a content distribution

network (CDN) – work that is part-funded by AT&T Labs USA. John Roche has designed and implemented an experimental network-based digital video recorder as part of his MSc research. The research commenced in 2001 and is nearing completion.

This research is motivated by several factors, including the penetration of broadband access and digital set-top boxes, availability of sophisticated techniques for bandwidth management and quality of service, and the popularity of personal digital video recorders. The latter are a class of home appliance that

allow a user to record TV content with features for shifted time viewing, but strictly limited in their range of available content and capacity. The proposed system model is shown in the figure and envisages set-top box clients accessing network-based proxy storage servers which are organised into a large-scale CDN.

The various elements of the architecture are:

- proxies that store and deliver video objects on-demand
- a database is used as an indexing server to track what content is available and where it is located within the network
- a management server is used to keep track of accounting and managerial aspects of the network, such as what proxies are active
- a search server provides a user interface to a cached replica of the index-database.

Clients connect to the Search Server and select the object to view, the search server redirects the client to the most appropriate Proxy Server that can deliver the object, and a streaming session is initiated. The system is technically a hybrid CDN/Peer-to-Peer architecture, consisting of proxies that are leased on-demand from Internet Service Providers and Network Operators. The proposed architecture enlists the use of idle ISP proxies during times of high load, and later, when the requests abate, the proxies are released from the network. The use of leased servers in this manner is an especially interesting and challenging feature of the work, and one which appears attractive from a commercial perspective in relation to deployment costs and operational flexibility.

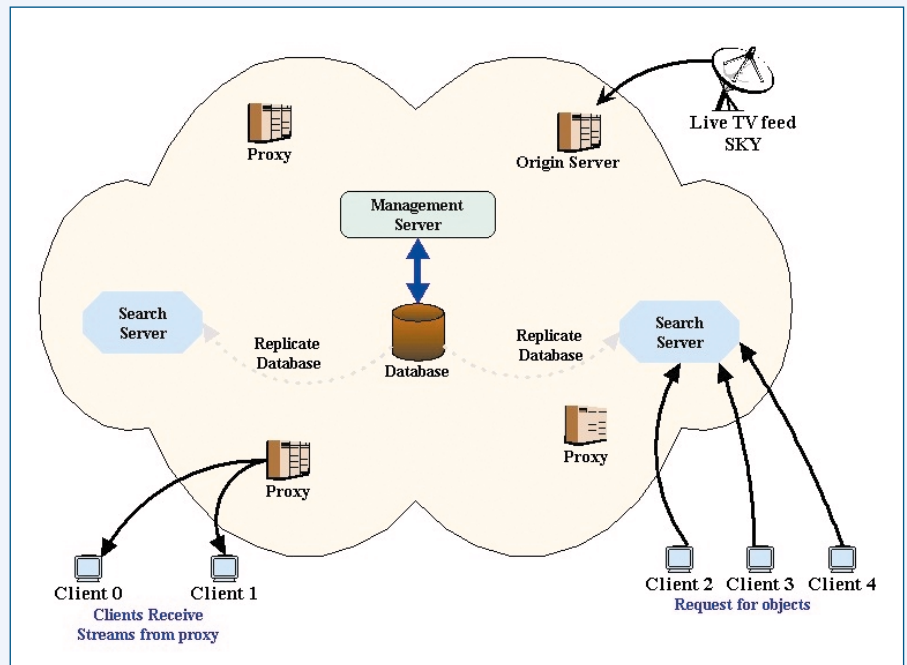
Currently, the main focus of the work is on the computational placement problem, which involves deciding upon the number and location of the replicas within the network. The effectiveness of our architecture can be greatly affected by the placement strategy in use. If too many replicas were created, then storage resources would be wasted, whereas too few replicas would increase the distance (network hops) between the clients and the servers, possibly resulting in

degraded performance. These factors and others need to be considered when deciding upon an optimal placement strategy. Finally, this placement strategy needs to be constantly evaluated, as object freshness and popularity diminishes over time.

One solution to finding the optimal placement for video objects is to take all aspects of the problem and formulate a

- broadcast channel
- broadcast time and date
- language
- episode number.

Yet-to-be-broadcast programmes that are identified in search results may be recorded by simply selecting an option on the GUI. Recording requests are sent to the network-based Management Server where a suitable Proxy is identi-



**High-Level System Architecture.**

cost function. This cost function evaluates all placement instances and identifies the placement layout that yields the lowest resource usage. This is computationally expensive so a heuristic approach is used based on a hierarchical architecture that first decides the general region of the network for the replica placement and then performs an in-depth evaluation of all possibilities within this region.

An experimental networked digital video recorder has been implemented to gain experience with the approach. Programme listings for all available channels are provided through an Electronic Programme Guide (EPG) based on ETSI naming standards. A GUI enables searching on a variety of information including:

- keywords

fied to complete the recording task. Whilst a programme is airing it is encoded by an Origin Server and transmitted by multicast to the designated Proxy and any clients wishing to view the programme in real-time. Proxy servers provide VCR-like functionality and content delivery is enabled by the utilisation of standard streaming protocols - Real-Time Streaming Protocol (RTSP) and Real-Time Transport Protocol (RTP).

**Link:**  
<http://www.cs.ucc.ie/misl>

**Please contact:**  
 Adrian Cahill, University College Cork / Irish Universities Consortium, Ireland  
 E-mail: [a.cahill@cs.ucc.ie](mailto:a.cahill@cs.ucc.ie)

# Presentation, Control and Collaboration in the Networked Classroom

by Leandro Navarro-Moldes and Manuel Oneto

**Classrooms with networked PCs can be augmented by software to facilitate learning. Presentation:** an IP multicast-based application allows screen content, streaming video, files and Web pages to be viewed in all PCs efficiently. **Control:** the instructor can observe a mosaic of thumbnail views of all PCs, and can select one for remote support. **Collaboration:** the instructor can define groups of students who can share screens, files or messages to facilitate collaborative learning.

This project was undertaken by the Distributed Systems group at Universitat Politècnica de Catalunya (UPC) in Barcelona (Spain), with the support of a Spanish publishing company (edebé). It was conceived in 2000 as an evaluation of the learning and collaboration possibilities in the growing number of classrooms in schools and universities that are equipped with networked PCs. The idea was to define, pilot and develop networked applications using simple PCs connected via Ethernet with no additional hardware. We started considering reliable multicast transports for screen-sharing applications based on the Remote Frame Buffer (RFB) protocol, as a software replacement for expensive and inconvenient screen projectors. As part of the research, pilot tests with hundreds of schools from many educational levels all over Spain were carried out, which provided feedback and suggestions from the field. As a result of this process, an initial application was developed and a spin-off company (Rededia.com) was initiated in the summer of 2002. This company specializes in multicast-based synchronous

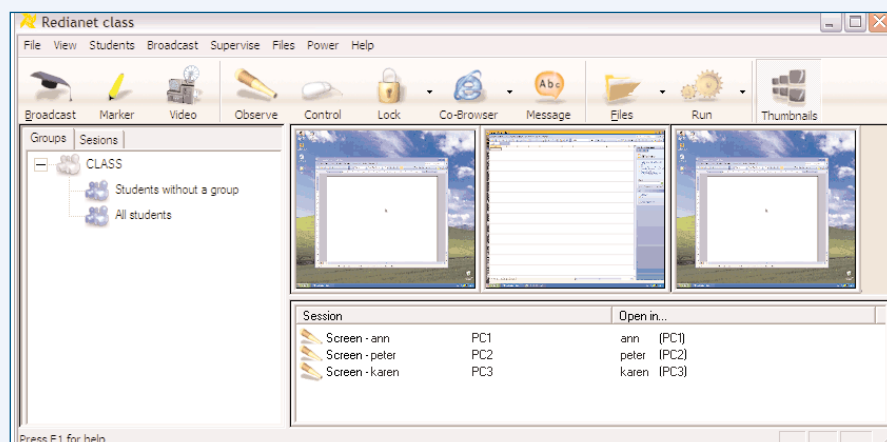
applications and is also providing technical support for the product.

From that experience, the software evolved towards a complete reimplementation of a Windows COM component-based synchronous and multicast-based middleware for 1-N, N-1 and N-N communication over a LAN supporting several synchronous applications. The current application goes well beyond simply replacing a screen projector: it supports interaction between the instructor and all or several PCs in the class in both directions, and also within groups of students defined by the instructor. Screen content, files, streaming multimedia, Web URLs and objects can be efficiently shared among PCs in the class.

The research group and the company are working together in complementary topics. While Rededia.com is supporting, maintaining and commercializing the Redianet class product, on the research side the group at UPC is working on extensions of the transport for additional media, for larger-range distribution

beyond LAN towards supporting multi-site collaboration, and on integrating sensors for tracking the location of people and PCs and automatically detecting the formation of groups. These new mechanisms would provide the applications with the context awareness to automatically and immediately offer collaboration support for emerging groups of people who are simply sitting at the same table, or move close to one other.

This long-term collaboration between a university research group and a technology company has proven mutually productive. Rededia.com and UPC recently applied for research and development project grants in national and European funding bodies with other ERCIM research institutions. Given the strong research component of our work, there is strong interest in collaborating with other ERCIM members, as well as offering the product of our work to ERCIM members (particularly universities: an evaluation version of the product can now be downloaded in Catalan, Spanish and English (<http://www.redianet.com>), and feedback is both appreciated and useful for the evolution of the product).



Screenshot of the application from the viewpoint of the instructor.

## Links:

Universitat Politècnica de Catalunya:  
<http://www.ac.upc.edu>  
<http://www.redianet.com>  
<http://www.rededia.com>

## Please contact:

Leandro Navarro-Moldes,  
 Politécnica de Catalunya, Spain  
 Tel: +34 9 3401 6807  
 E-mail: [leandro@ac.upc.edu](mailto:leandro@ac.upc.edu)

Manuel Oneto,  
 REDEDIA, Spain  
 Tel: +34 9 3413 7952  
 E-mail: [manuel@rededia.com](mailto:manuel@rededia.com)



# Video Transcoding Architectures for Multimedia Real Time Services

by Maurizio A. Bonuccelli, Francesca Lonetti and Francesca Martelli

The video transcoding project at PisaTel, a laboratory located at ISTI-CNR in a joint collaboration between ISTI, Ericsson Lab Italy, the 'Scuola S. Anna' and Pisa University, aims at developing efficient solutions for real-time video coding and transcoding.

A key technology for many applications, such as Digital TV Broadcasting, Distance Learning, Video on Demand, Video Telephony and Video Conferencing, is Digital Video Coding. In third generation telecommunication systems, communication technologies are extremely heterogeneous. Adapting the media content to the characteristics of different networks (communication links and access terminals) in order to obtain video delivery with acceptable service quality is thus an important issue. Video transcoding converts one compressed video bitstream into another with a different format, size (spatial transcoding), bit rate (quality transcoding), or frame rate (temporal transcoding). The goal of transcoding is to enable the interoperability of heterogeneous multimedia networks reducing complexity and run time by avoiding the total decoding and re-encoding of a video stream.

We are interested in temporal video transcoding. This process skips some frames in order to change the frame rate of a video sequence without decreasing the video quality of non-skipped frames. In third generation mobile telecommunication systems (UMTS), the bandwidth of a coded video stream must be drastically reduced in order to cope with the constrained transmission channel. Frame skipping is a promising approach for transcoding one video sequence into another with a lower bit rate, while maintaining good video quality. Many multimedia services (such as videoconferencing, video telephony) have real-time features, so transcoding must guarantee a fixed communication delay. We have

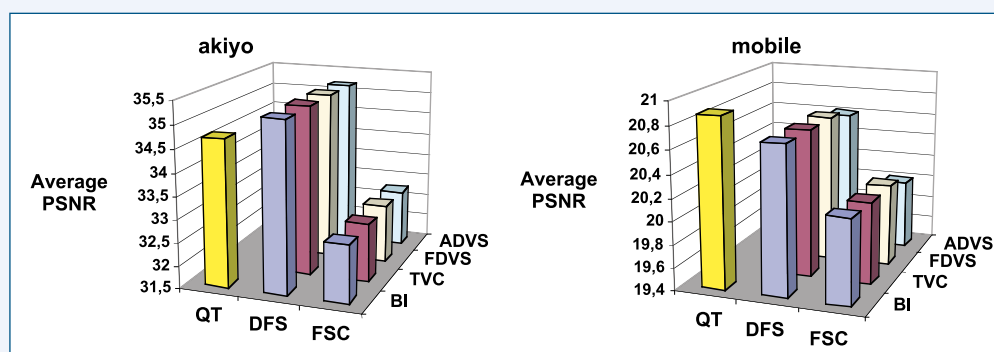
concentrated on this aspect, and have developed and evaluated two temporal transcoding architectures.

## Temporal Transcoding Architectures

In a video sequence, many frames are coded with reference to previous frames, using motion vectors and prediction errors. In temporal transcoding, when a frame is skipped, the references of the next frame are no longer valid. Motion Vector Composition (MVC) is a procedure that computes the new motion

produces re-encoding errors, and frames are skipped on the basis of the effect of re-encoding errors and motion activity. The goal of this strategy is to minimize the re-encoding errors and to preserve the motion smoothness of the transcoded frames.

The real time features of many advanced multimedia applications are not taken into account by the above architectures. In order to meet the needs of such applications, we have modified both architectures so that the output bit rate is



Average PSNR (dB) of temporal (DFS and FSC) and quality transcoder (QT) for 'akiyo' and 'mobile' video sequences (input bit rate = 64 Kbps, output bit rate = 32 Kbps).

vectors of the non-skipped frames. Once new motion vectors have been computed, new prediction errors are also needed for the transcoded frames. Another important issue in temporal transcoding is the choice of frames to be skipped. A first frame rate control architecture, Dynamic Frame Skipping (DFS), dynamically adjusts the number of skipped frames according to motion activity. This gives a measure of the motion in a frame and frames with much motion are not skipped. Another temporal transcoding architecture, Frame Skipping Control (FSC), computes the prediction errors; this

constant, and the maximum communication delay is fixed. We achieved this by introducing a transcoder output buffer, and by skipping frames according to the buffer occupancy. The maximum communication delay depends on the buffer size.

## Simulation Results

We implemented an MPEG4-based temporal transcoder and evaluated the performance of both our architectures over several benchmark videos. The results, in terms of PSNR (a measure indicating the quality of the transcoded sequence) are compared with those of a

quality transcoder (QT). The comparison shows that better performance is achieved by quality transcoding for videos with a lot of motion and by temporal transcoding (DFS and FSC) for videos with little motion (see the figure). Moreover, we observed that the DFS architecture has the better performance

since in FSC many frames are skipped because of re-encoding errors. We obtained similar results using different MVC algorithms (Bilinear Interpolation (BI), Telescopic Vector Composition (TVC), Forward Dominant Vector Selection (FDVS), Activity Dominant Vector Selection (ADVS)).

**Link:**  
Pisatel:  
<http://pacinotti.isti.cnr.it/ERI/project4.htm>

**Please contact:**  
Francesca Martelli, ISTI-CNR, Italy  
Tel: +39 050 315 3468  
E-mail: [f.martelli@isti.cnr.it](mailto:f.martelli@isti.cnr.it)

## Designing Multi-Modal Multi-Device Interfaces

by Silvia Berti and Fabio Paternò

The increasing availability of new types of interaction platforms raises a number of issues for designers and developers of interactive applications. There is a need for new methods and tools to support the development of Multi-Modal Multi-Device applications. The TERESA tool supports multi-modal user interfaces in multi-device environments.

Life today is becoming a multiplatform experience in which people are surrounded by different types of interactive devices, including mobile phones, personal digital assistants (PDAs), pagers, car navigation systems, mobile game machines, digital book readers, and smart watches through which they can connect to networks in different ways.

This situation poses a number of challenges for the designers and developers of multi-device interfaces. A further complication is that these devices can use different modalities (graphics, voice, gestures, and so on, in different combinations). Thus, although there are now a number of multimodal systems in circulation, their development still remains a difficult task.

A promising solution to handle the complexity involved is to use logical device-independent, XML-based languages to represent concepts, such as user tasks and communication goals, along with intelligent transformers that can generate user interfaces in different implementation languages for different platforms depending on their interaction resources and modalities.

We have developed an authoring environment, TERESA, within the EU IST project CAMELEON to address issues related to multi-device interfaces. An



**A User Interface generated by the authoring environment.**

extension was implemented in the SIMILAR Network of Excellence on Multi-Modal User Interfaces in order to support multi-modal user interfaces in multi-device environments. TERESA incorporates intelligent rendering to decrease the cost of developing multiple interface versions for the different target platforms, allowing designers to concentrate on logical decisions without having to deal with a variety of low-level details at the level of implementation languages.

In the design and development process, it is important to consider that there are tasks that may be meaningful only when using some specific platform or modality. For example, watching a long movie makes sense in a multimedia

desktop system, whereas accessing information from a car in order to avoid a traffic jam can be done only through a mobile device, and if this task is performed while driving, it can be supported only through a vocal interface. The modality involved may also impact on how to accomplish a task, for example vocal or graphical mobile phone interfaces require the user to perform tasks sequentially which could have been done concurrently in a desktop graphical interface.

In our approach, a user interface is structured into a number of presentations. A presentation identifies a set of interaction techniques that are enabled at a given time. The presentation is structured into interactors (logical descriptions of inter-

action techniques) and composition operators, which indicate how to put the interactors together. While at the abstract level, the interactors and their compositions are identified in terms of their semantics in a modality independent manner, at the concrete level their description and the attribute values depend on the modality involved.

In the case of both vocal and graphical support, multimodality can be exploited in different manners. Modalities can be used alternatively to perform the same interaction. They can be used synergistically within one basic interaction (for example, providing input vocally and showing the result of the interaction graphically) or within a complex interaction (for example, filling in a form partially vocally and partially graphically). A certain level of redundancy can also be supported, for example when feedback for a vocal interaction is provided both graphically and vocally. We have decomposed each interactor into three parts and the above properties can be applied to each part: prompt, represents the interface output indicating

that the system is ready to receive an input; input, represents how the user can actually provide the input; feedback, represents the output of the system after the user input.

The composition operators indicate how to put such interactors together and are associated with communication goals. A communication goal is an effect that designers aim to achieve when they structure presentations. Grouping is an example of a composition operator that aims to show that a group of interface elements are logically related to each other. It can be implemented in the graphical channel through one or multiple attributes (fieldset, colour, location...), whereas in the vocal channel the grouping of elements is achieved by inserting a sound or a pause at the beginning and the end of the grouped elements. In the case of multimodal interfaces we have to consider the actual resources available. Thus, grouping on a multimodal desktop interface should be mainly graphical and the use of the vocal channel can be limited to providing additional information on the elements

involved. Instead, the grouping for a multimodal PDA interface uses the vocal channel more extensively, while the graphical one is dedicated to important or explicative information (see Figure).

The current environment supports design and development for various types of platforms (form-based desktop, interactive graphical desktop, form-based mobile, interactive graphical mobile, vocal device, multimodal device) and generates the corresponding user interfaces in various implementation languages (XHTML, XHTML Mobile Profile, VoiceXML, SVG, X+V). Future work will be dedicated to supporting additional modalities, such as gestural interaction.

#### Links:

TERESA: <http://giove.isti.cnr.it/teresa.html>

SIMILAR Network of Excellence:  
<http://www.similar.cc>

HIIS Laboratory at ISTI-CNR:  
<http://www.isti.cnr.it/ResearchUnits/Labs/hiis-lab/>

#### Please contact:

Fabio Paternò, ISTI-CNR, Italy  
Tel: +39 050 315 3066  
E-mail: [fabio.paterno@isti.cnr.it](mailto:fabio.paterno@isti.cnr.it)

## LimSee2: A Cross-Platform SMIL Authoring Tool

by Romain Deltour, Nabil Layaïda and Daniel Weck

**LimSee2 is an open-source and cross-platform authoring tool dedicated to the manipulation of time-based multimedia documents for the Web. It relies on the SMIL standard of the World Wide Web Consortium (W3C).**

With the rapid diversification of the Web (access devices, communication networks etc), the utilization of multimedia documents is becoming common practice. For this reason, in 1998 the W3C designed a standard mark-up language dedicated to the description and synchronization of multimedia content: the Synchronized Multimedia Integration Language (SMIL, pronounced 'smile'). However, authoring multimedia information remains a real challenge, as the traditional WYSIWYG paradigm is difficult to apply in non-deterministic multimedia presentations involving rich user interaction. Taking such parameters into account, the Web Adaptation and

Multimedia (WAM) team of INRIA has developed LimSee2, an authoring tool that provides a powerful graphical user interface designed to assist in the manipulation of time-based multimedia and thereby increase productivity.

LimSee2 has a multi-view solution that renders the structure of the SMIL document at different levels during the authoring process: timing and synchronization, spatial layout, XML tree etc. The different views are synchronized (a modification in one view is immediately rendered in all the other views) and provide functionality that allows a user to manipulate and fine-tune a SMIL

document without requiring a full knowledge of the language.

### Spatial View

The spatial layout of a SMIL document can be edited in LimSee2 in a 2D canvas, which constitutes a WYSIWYG environment for a fixed time in the temporal scenario. SMIL regions can be easily moved, resized or created in a few clicks, media content can be directly previewed (for images, texts and videos), and region z-indexes can be adjusted with an intuitive drag-and-drop mechanism. The 2D canvas also provides traditional features such as a zooming tool or a customizable snap grid.



## Timing View

Visualizing the temporal scenario of a document during the authoring process is one of the key challenges of multimedia authoring. To fulfil this requirement, LimSee2 features a timeline view: a temporal element is represented by a box, with its length standing for the duration of the element, and its position standing for the start-time of the element. An arrow linking two boxes then represents a synchronization relation. Hence, the user can easily and intuitively adjust media synchronization by moving and resizing the boxes in the timeline. Moreover, a cursor may be moved through the timeline in order to see in the spatial view the state of the multimedia presentation at a specific time. This is very useful for previewing the results of the synchronization. In addition to SMIL-specific features (for instance SMIL temporal containers represented as collapsible box containers) the timing view also provides a zooming tool and media preview ability.

## XML Features

As SMIL is an XML language, many well-established XML technologies are directly involved in LimSee2. First of all, the DTD-awareness of the application ensures the validity of the underlying SMIL document model (provided that it is modified within the Limsee2 GUI). Additionally, a built-in validator can check on demand the conformance

of the document with SMIL 1.0 or 2.0 syntaxes, and allows the user to correct possible errors interactively. LimSee2 also features two XML-dedicated views: the Structure View, which renders the hierarchical structure of the SMIL document as a collapsible tree, and the Attributes View, which is a high-level DTD-aware attributes editor allowing fine-tuning of the presentation directly from the attributes values. As for the users who want to directly access the source of their SMIL documents, LimSee2 integrates an XML source editor. This features traditional syntax highlighting, pretty formatting, and incremental search and replace, and seamlessly synchronizes with the other views.

## Other Features

In addition to the major features described above, LimSee2 also provides the following:

- the character encoding of documents is detected at opening and can be changed on-the-fly
- the LimSee2 interface has just been internationalized, and is available in English, French, and Japanese
- a native player (eg RealPlayer, Ambulant) can be launched from the application to externally visualize the edited document
- a 'storing' mechanism means entire SMIL presentations can be saved into a single directory (source, media

content, linked documents), even for distant data accessed by HTTP

- slides can be automatically imported as JPG images from Microsoft PowerPoint or OpenOffice.org presentations
- a slideshow builder tool allows the creation in a few clicks of multimedia slideshow presentations, which synchronize a video/audio track, an interactive table of content, a navigation panel and imported slide images.

## Collaboration and Perspectives

A new version of the application is under development. It will be more user-oriented, and will feature powerful authoring functionality such as template-based editions or reusable models of documents. In parallel, the WAM Team is cooperating closely with NRCD (National Rehabilitation Center for Persons with Disabilities, Japan) on accessibility features in LimSee2, in the context of a project aiming at establishing a natural disasters preparedness system.

## Project Details

The development of LimSee2 was initiated by the Web Adaptation and Multimedia research team at the French National Institute for Research in Computer Science and Control (INRIA), Grenoble, France.

The project started in October 2002. The first public release was made on the 13th June, 2003. The latest available version is 1.5.2 and was released on the 28th March, 2005.

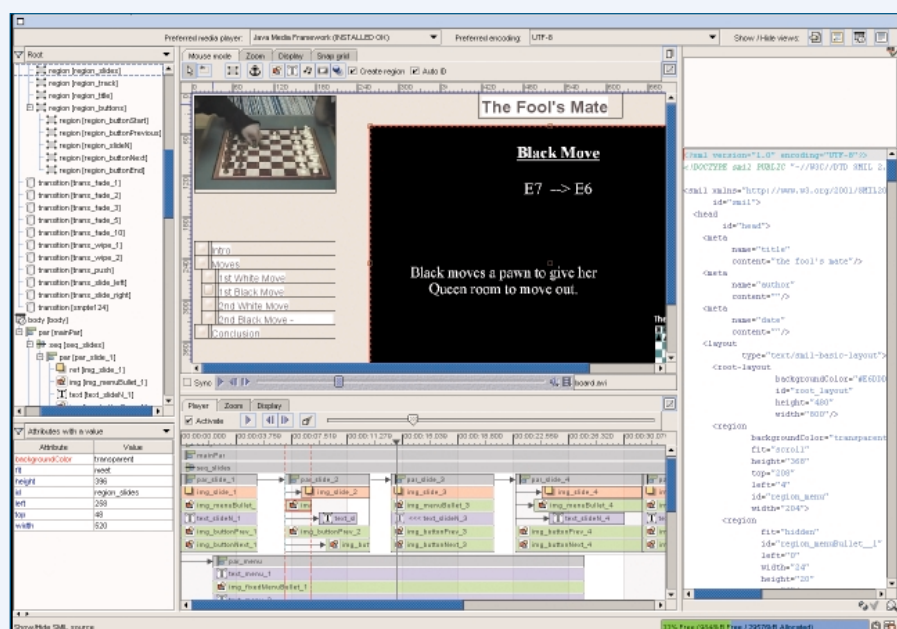
LimSee2 is developed with Java J2SE 1.4.2. It is currently available on Windows, Linux, and Mac OS X platforms.

### Links:

LimSee2 Home Page:  
<http://wam.inrialpes.fr/software/limsee2/>  
 WAM Team Home Page:  
<http://wam.inrialpes.fr/>  
 W3C Synchronized Multimedia Page:  
<http://www.w3.org/AudioVideo/>

### Please contact:

Nabil Layaïda, Romain Deltour,  
 WAM Team, INRIA, France.  
 Tel: +33 4 76 61 52 84, +33 4 76 61 52 18  
 E-mail: Nabil.Layaïda@inrialpes.fr,  
 Romain.Deltour@inrialpes.fr



The LimSee2 SMIL authoring tool.

# Merging Virtual Reality and Television through Interactive Virtual Actors Marilyn — Multimodal Avatar Responsive Live Newscaster

by Sepideh Chakaveh

A recent and innovative research activity of the ERCIM E-Learning Working Group and the Fraunhofer Institute for Media Communication (IMK) merges virtual reality, artificial intelligence and television.

In recent years we have seen the Internet and television come closer in many ways. This began when streaming technology gradually opened the way to watching television programs on the Internet. IMK-ITV is now considering reversing this process, that is, implementing Internet-type applications on television (TV). Moreover, interactive TV (iTV) aims to combine traditional TV with additional services previously only available on the Internet. This is leading to the disappearance of the boundaries between two forms of content, namely infotainment and edutainment.

The development of iTV depends on several factors. The first of these is the development of the iTV production chain. This involves producing media content (ie a TV program), broadcasting it on a particular channel and receiving it using a TV view system. The second is the development of both the medium and paradigm of interactivity with iTV. So far, interaction with TV content is still limited, with 'clickable' interactivity as the sole paradigm of interaction, and the use of remote controls as the sole medium of interaction. The third factor is the development of artificial intelligence (AI) techniques that serve an intelligent iTV (IiTV) platform. To this end, speech recognition, natural language understanding and decision support systems should be deployed in the development of (IiTV). This has enormous potential, as the number of Internet users, while rapidly

increasing, is still small compared to television users.

Marilyn (Multimodal Avatar Responsive Live Newscaster) is a new system for interactive television, where a virtual reality three-dimensional facial avatar responds to the remote control in real time, speaking to the viewer and providing the requested information.

Marilyn informs the viewer with a click of a button on daily financial news. Here the focus is on the provision of choice as

well as personalization of information in an entertaining manner. As well as offering live financial data from leading stock exchanges such as New York, London, Frankfurt and Tokyo, multilingual aspects of the information are also catered for.

Traditionally, financial news has been regarded as content-based and rather rigid in format. In contrast, the edutainment aspects of Marilyn can make such a program entertaining as well as informative.

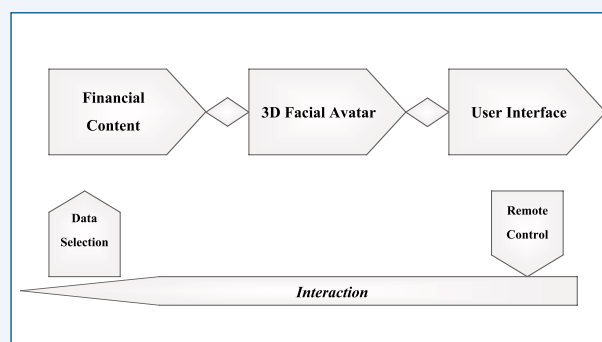


Figure 1: A schematic of the structure of Marilyn.



Figure 2: Marilyn screenshot presentation of the Tokyo stock exchange.

Figure 1 shows a schematic of the structure of Marilyn. The application consists of three stand-alone but interrelated sections.

## Financial Content

This section, which is primarily developed in Java, M3D (a Java3D language developed at Fraunhofer) and XML, defines the core information centre of Marilyn. Financial content from various locations (eg Frankfurt, London and New York) is fed in live to the system, and updated continuously.

## 3D Facial Avatar

This, which forms the heart of Marilyn, is a 3D humanoid face of the late actress Marilyn Monroe, which reads the financial news tickers live in respective languages. The facial avatar is implemented using XML, MPEG4 and M3D1. The program implements natural mouth and eye gestures during speech, as in the case of a real human being.

The avatar may of course have any other face mapped to it, but, owing to the different bone structures of male and female faces, the sex of the facial avatar is pre-defined. This means the mesh structure for the avatar must be created in advance for a male or female presenter.

Equally, the voice of a male or a female speaker should be defined, as well as the desired language or accent. For example, if Marilyn reads the news in English, then information provided from the

London stock exchange could be read with a British accent, or that from the New York stock exchange with an American accent. In addition, the financial data from the Frankfurt stock exchange can be read in German.

#### User Interface

Through the user interface the viewer may communicate with the facial avatar, which in turn selects the desired content. By clicking on the remote control it is possible to choose the stock exchange and in which language the information

should be read. The user interface is first realized using FLASH and it is then implemented using MPEG4.

This project has been supported by Dr. Soha Maad, an ERCIM fellow, hosted by the Institute for Media Communications for 18 months.

#### Please contact:

Sepideh Chakaveh  
Institute for Media Communication,  
Fraunhofer ICT Group  
ERCIM E-Learning Working Group  
E-mail: Chakaveh@imk.fraunhofer.de

## Interactive Multimedia for Supporting the Quality of the Production

by George L. Kovács, Géza Haidegger and János Nacsa

In the frame of DIGITAL FACTORY, a Hungarian national R&D project on factory automation, researchers are investigating, developing and applying interactive multimedia technologies to assist design, production and test processes. The project is being undertaken at the CIM (Computer Integrated Manufacturing) Research Laboratory of SZTAKI and in the factory partner's sites. Over the past three years, a generalized software environment (IMUTA) and some use cases have been developed.

Within a factory environment, there are several areas in which it is essential that human experts observe, test, control, and understand the details of the production processes. This is the case in both the design and the planning stages, as well as during manufacturing, testing, and verification. The new features of interactive multimedia services raise the level of quality, increase the efficiency, observability and controllability of production, and also allow the customer and the end-user to actively take part in the manufacturing processes.

Since tele-presence and interactive multimedia involve the integration of humans, our research work looks at man-machine interaction, and the definition of VMDs (Virtual Manufacturing Devices) devoted to human entities within the production area. A survey on wearable computers and head-mounted monitors was used to analyse current technology with a view to applying computer platforms on the human body.

Our research focused on a generalized software environment that enables tele-

presence operations with interactive multimedia features. The tangible result of this work is called IMUTA: Interactive Multimedia for Tele-presence Applications. In this framework we extended Enterprise Resources with multimedia information and functionality. The Extended Resource model provides a richer and more human-oriented view of resources and processes. For example:

- machining can be observed remotely and in real time; relevant process data, product drawings, technology plans and quality reports can be chosen and shown according to actual requests
- documentation of value products can be extended with multimedia attachments to ensure the customers' satisfaction
- customers can observe the production flow of a product.

Significant effort has been devoted to the planning of some pilot demonstrations. We looked for demonstration sites and service-intensive applications that would inspire the hosts and potential customers

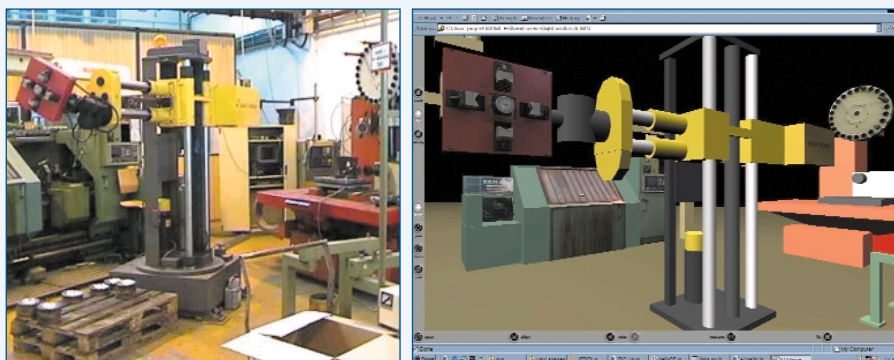


Figure: The real and the simulated manufacturing cell.



to implement similar functions. Models were developed for the specific application scenarios in order to derive a generalized framework.

Some of these scenarios are the following:

- reuse of high-value waste workpieces are supported with complex databases, including production and design information, and quality multimedia data; users can reach these entities via an intranet environment
- a wearable computer-based test environment can support customer witness tests
- predefined deviations in the production parameters activate multimedia data acquisition at the necessary machines and operations; the effects of the time delay of the statistical data are

solved with a circular multimedia buffer mechanism.

By integrating real-life vision and computer-animated entities, virtual and augmented reality services are offered for specific user applications (see Figure). For that purpose, we developed a distributed simulation and visualization environment for flexible manufacturing systems. It enables users to visualize the results of a simulation of an FMS (flexible manufacturing system) with three-dimensional graphics, or to visualize the actual data of a real system. On the client side, our system uses VRML and Java, so it needs only a Web browser with the appropriate plug-ins to use it. The users can view the scene provided by the simulation system via the Internet, as well as choosing the

perspective from which they view it and interacting with it in the simulation phase. The IMUTA concept supports the integration of augmented reality.

These developments will make interactive multimedia a more widely accepted tool not only in simulation and experimentation, but in real factory applications. This will, however, require further organization and management.

#### Links:

<http://www.sztaki.hu/sztaki/ake/cim/>  
<http://www.digital-factory.sztaki.hu>  
[http://www.ercim.org/publication/Ercim\\_News/enw51/monostori.html](http://www.ercim.org/publication/Ercim_News/enw51/monostori.html)

#### Please contact:

George L. Kovács, Géza Haidegger,  
 János Nacsá, SZTAKI, Hungary  
 Tel: +36 1 279 6140  
 E-mail: [gkovacs@haidegger@nacsá@sztaki.hu](mailto:gkovacs@haidegger@nacsá@sztaki.hu)

## The CINEMA Project: A Video-Based Human-Computer Interaction System for Audio-Visual Immersion

by Renaud Dardenne, Jean-Jacques Embrechts,  
 Marc Van Droogenbroeck and Nicolas Werner

Numerous studies are currently focusing on the modelling of humans and their behavior in 3D environments. The CINEMA project has similar goals but differentiates itself by aiming at enhanced interactions, the creation of mixed reality, and the creation of interactive and reactive acoustical environments. Part of the project consists in gesture recognition of a user, who is given the real-time control of auralization and audio spatialization processes.

The CINEMA project is a collaborative effort between teams at four Belgian universities. The project foresees the development of a real-time system that includes 3D user and environment modelling, extraction of motion parameters, gesture recognition, computation of depth maps, creation of virtual spaces, and auralization and spatialization. The merging of these techniques is expected to produce a tractable 3D model allowing users to interact with a virtual world.

Although much progress has been made in the areas of video and audio processing, there is still a need to find a way of properly combining these elements. The production of a tool for

augmented or virtual reality that mixes video and audio could be used for a number of applications: as an educational tool for example, or at cultural places like museums.

This project started on January 2003 and will end in September 2006. It is funded by the Walloon Region, Belgium.

Following are details regarding the video-handling techniques and the interaction with the sound system:

### The Video Analysis Engine

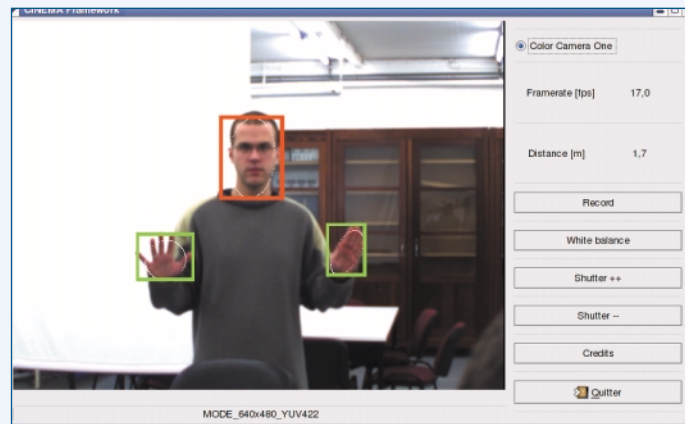
The major goal of the video part of the system is to identify simple human gestures. As real time operation is a compulsory requirement for any interac-

tive system, we committed to the use of simple tools to achieve gesture recognition. Therefore our algorithm detects motion based on a known algorithm for background extraction, combined with a skin detection algorithm to isolate the head and the hands. To increase the robustness of the system, we tested several motion detection algorithms and combined them. For indoor scenes, a background extraction using a static background and a thresholding algorithm in the HSV colour space for skin detection appeared to suffice, except for the presence of shadows. Fortunately shadows have a specific range of colours and particular shapes that enable us to discard them.

Once regions with skin have been detected, we must distinguish between the head and the hands. It is hard to propose a fast and general method that would consider all possible pathological cases for the relative positions of skinned regions. Therefore we rely on the assumption that the head lies in the middle of the region with motion. The final system is capable of recognizing simple hand positions (both hands raised above the head, both hands on the same side of the body etc), and sending appropriate instructions to the audio subsystem (see Figure 1).

### The Audio Subsystem

To build a realistic and immersive world, we have implemented a sound spatialization system. This audio system is able to synthesize localized sound sources, which means that we can decide on the orientation and distance of all the sources. In addition, the system uses real-time software to produce a realistic sound. This last technique, called auralization, is fully configurable and takes into account the room configuration or the virtual audio environment via its impulse response, and the positions of the sound sources and the listener. For the sound rendering, a multichannel setup is used, and the audio signals are



Head and hands detection.

distributed to the loudspeakers according to an amplitude panning technique.

### The Communication Channel between the Video and Audio Subsystems

The video and the audio subsystems communicate through a socket interface. This approach offers the choice of running all the algorithms on a single machine (using the loopback) or on separate machines. As all the machines are presumably located on a LAN, a communication channel based on the UDP protocol was chosen. The subsystems interoperate, and in order to allow for further extensions, both understand a language that uses tags and is based on the XML language.

#### Link:

<http://www.ulg.ac.be/telecom/research/cinema.html>

#### Please contact:

Nicolas Werner, Université de Liège, Belgium  
Tel: +32 4 366 2652  
E-mail: [nwerner@ulg.ac.be](mailto:nwerner@ulg.ac.be)

Renaud Dardenne,  
Université de Liège, Belgium  
Tel: +32 4 366 2686  
E-mail: [R.Dardenne@ulg.ac.be](mailto:R.Dardenne@ulg.ac.be)

Jean-Jacques Embrechts,  
Université de Liège,  
Belgium  
Tel: +32 4 366 2650  
E-mail: [jjembrechts@ulg.ac.be](mailto:jjembrechts@ulg.ac.be)

Marc Van Droogenbroeck,  
Université de Liège, Belgium  
Tel: +32 4 366 2693  
E-mail: [M.VanDroogenbroeck@ulg.ac.be](mailto:M.VanDroogenbroeck@ulg.ac.be)

## Personalised Enriched Broadcast Experience

by Mounia Lalmas, Nick Bryan-Kinns and Alan Pearmain

**The SAVANT European research project developed integrated broadcast and Internet technologies, allowing users to access enriched broadcast content in an intelligent and transparent manner on a range of devices (eg TV, PC and PDA) and under varying network conditions. One aim of SAVANT was to personalise the enriched broadcast experience.**

Interactive digital television (iTV) is becoming widely available and is being promoted by broadcasters as a means of attracting viewers to digital TV and as an additional revenue stream. iTV provides additional services to the viewer, which enrich the broadcast experience, by including, for example, time-shifted viewing; bidirectional feedback; and additional content, from subtitles to interactive games; which all take the user's experience to a higher level than

the linear, passive format of traditional TV. However, this comes at a cost.

Increasing the amount of content available whilst retaining current interaction devices increases the complexity of the viewer interaction resulting in more buttons being pressed to locate additional content, and more information being crammed onto screens to show available options. Viewers could be provided with new forms of interaction devices such as personal computers, or

increasingly complex remote controls. However, this would change the nature of the viewing experience from a shared, family, or social situation, to a personal tele-visual experience where one person has control, and possibly prime viewpoint.

A more promising approach is to personalise the experience so that to reduce the content presented to what is believed of interest to viewers. This is most commonly realised as personalised elec-

tronic programme guides (EPG). However, as the amount of content increases personalisation needs to be introduced within programmes and sub-programmes, especially in information rich domains such as news programmes. In this article we describe our personalisation system developed as part of the SAVANT project (Synchronised and Scalable Audio Visual content Across NeTworks).

### Personalisation

A service consists of a number of service components with a component consisting of either a segment of the main broadcast or an item of additional related content (eg MPEG-4 clips, HTML pages, 3D graphics). A service is personalised by modified it according to user preferences so that only those service components that are of interest to the user are shown or recommended.

A SAVANT screen, for the news domain, provides an alphabetical listing of content items from the daily archive, grouped under their respective topic headings. Figure 1 shows how content is personalised according to user interests, which are expressed in user profiles. Personalisation is applied to the topic listings as well as the individual main and additional content items of a news story within each topic.

The personalisation system also built recommendations according to user profiles (see Figure 2). A recommendation screen only include the ranked list of content items that the user is likely to be interested in, whereas a personalisation screen include all content items, but place the user's 'favourites' at the top of the list.

### Implementation

Personalisation was applied to the main broadcast content and any additional content included in the digital TV service itself. Any type of additional content, together with segments of the main content, was considered as service components within the complete service of a TV programme. A novel metadata model based on the existing standards



Figure 1: Personalisation.



Figure 2: Recommendation.

TV-Anytime, MPEG-7 and MPEG-21 was developed to transparently segment the traditional linear programme into sub-components.

Users profiles were represented as a list of likes and dislikes with associated probability values reflecting the degree of interest. By examining over time the type of content that users choose to view, user profiles were created and continually maintained to reflect the changing interests of the user, by employing a relevance feedback algorithm.

Information retrieval technologies were used to prioritise news items and to build recommendations. The metadata associated with news items was indexed, and this index was matched against user profiles. Matching resulted in the probability that the news item is relevant to the user profile, where the higher the probability value, the higher the news item is estimated to be of interest to the user.

### Evaluation

A study was carried out to evaluate the appropriateness of the personalisation of iTV services. We considered the appropriateness of a recommended news clip to be measurable in terms of its perceived relevance to some defined other semi-randomly generated.

Our results indicate that the recommendations were on the whole better than the semi-random generated ones, which means that appropriate recommendations were made to users. Post-feedback showed some skepticism from the participants in terms of the utility of personalised services. Participants expressed their concerns about using recommendations to structure their viewing, but were still interested in receiving such recommendations. They also indicated that personalisation may be of use when there is a lot of content available and some pre-filtering is required, but would not be appreciated if it actually selected additional possible items to view and so contributed to information overload. So, it is the filtering aspect that was seen by the participants as its key utility in their interactive TV experience.

### SAVANT

This project was funded by the European Commission IST programme. Our partners were Brunel University, UK; Expway, France; Fraunhofer IPSI, Germany; Institut für Rundfunktechnik, Germany; Nederlands Omroepproductie Bedrijf, The Netherlands; Rundfunk Berlin-Brandenburg, Germany; Siemens, Germany; STT, Spain; Telenor, Norway; TNO Telecom, The Netherlands. The interface was designed by IPSI and the content was provided by RBB (Rundfunk Berlin-Brandenburg).

Link:  
<http://www.savant.tv/>

Please contact:  
Mounia Lalmas  
Queen Mary University of London, UK  
E-mail: [mounia@dcs.qmul.ac.uk](mailto:mounia@dcs.qmul.ac.uk)  
<http://qmri.dcs.qmul.ac.uk>



## Articles in this Section

48 **LIGHT: XML-Innovative Generation for Home Networking Technologies**by Luca Tarrini  
and Vittorio Miori49 **Modelling of Authentic Reflectance Behaviour in Virtual Environments**

by Michal Haindl and Jiří Filip

51 **Haptic Training Systems in Virtual Surgery**

by José San Martín, David Miraut, Carolina Gómez and Sofía Bayona

52 **Computer Recognizes Whale Tails**by Annette Kik, Eric Pauwels  
and Elena Rangelova53 **Text Document Classification**

by Jana Novovičová

54 **Advancing Black-Box Reuse in a Multimedia Application Framework**

by Bernhard Wagner

55 **CASSEM: Vibration Control in the Smart Way**

by Salim Belouettar

57 **Point6: The IPv6 Skill Centre — Moving to the Next-Generation Internet Protocol**

by César Viho and Annie Floch

58 **Working Slower with More Powerful Computers**by Lorenz M. Hilty, Andreas Köhler, Fabian van Schéele  
and Rainer Zah59 **Coordinating IST Research across Europe**

by Simon Lambert

60 **Software Automation meets Interactive Media Development**by Dirk Deridder, Thomas Cleenewerck, Johan Brichau  
and Theo D'Hondt

# LIGHT: XML-Innovative Generation for Home Networking Technologies

by Luca Tarrini and Vittorio Miori

The LIGHT project aims at designing and developing a standard, service-oriented architecture for the next generation of embedded devices in order to render home networking easier and more effective for the end-user. LIGHT will exploit the Service Oriented Architecture (SOA) paradigm in a new domain, the home environment, proposing advanced and interoperable profiles for household appliances and devising solutions for the problems of overhead associated with hosting Web services on embedded systems.

LIGHT (xml-Innovative Generation for Home networking Technologies) is a research project of the Domotics Lab at ISTI-CNR, Pisa, in collaboration with two local SMEs: SISTER (Sistemi-Territoriali), expert provider of professional GIS services and consultancy, and MicroComm (Microwave Communications), specialized in microwave communications circuits and systems. The project began in November 2004 and is currently developing a lightweight middleware, WSED (Web Services for Embedded Devices), for the seamless networking of embedded systems through the Web services protocols.

Increasingly, many appliances, such as video recorders, MP3 players, mobile phones and personal digital assistants have small embedded computing units that allow their users to interoperate in a ubiquitous environment. The actual trend of including a growing number of transistors into ever smaller devices, while decreasing their power consumption and costs, is pointing the Web services domain in a new and interesting direction, the home environment.

We believe that the advent of service-oriented protocols opens the way to universal, platform and language neutral connectivity among appliances. This technology is offering much of real interest to the home networking community, in particular with respect to system architecture. In recent years, a number of architectures for home networking have been proposed, all aiming at facilitating dynamic cooperation among devices in order to be able to announce the presence of new services in a network, or to

discover services and use them. However, so far, no single architecture provides all the capabilities of the others, and none is yet sufficiently mature to dominate the field.

Our research is thus focused on the integration of a rich suite of Web service specifications with resource-constrained devices to enable the typical networking functions. Figure 1 shows our proposed architecture.

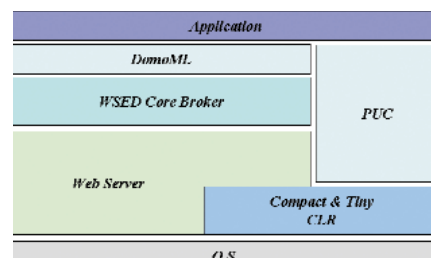
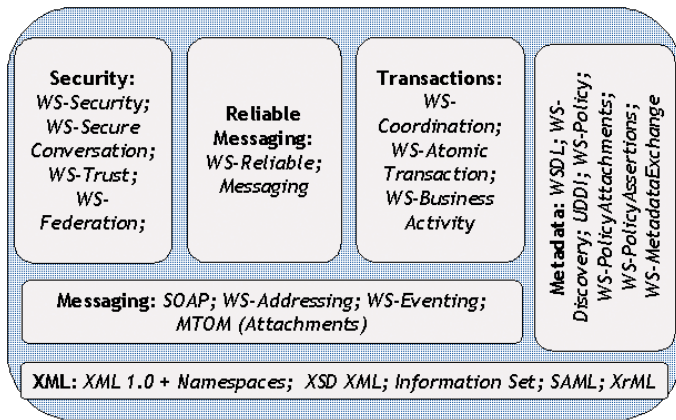


Figure 1: LIGHT Architecture.

The WSED Core Broker integrates current and emerging Web service specifications (see Figure 2), so that these can be used efficiently to offer new services with better discovery, description, and eventing models, and advanced security. In fact, the next generation of services in the home environment will be autonomous and platform-independent computational elements, described, published, discovered, orchestrated, and programmed using standard protocols. This will significantly reduce the cost and time that users have to spend in configuring and managing their pervasive device.

To satisfy the specific needs for interoperability between the WSED core and



**Figure 2:**  
**Web Services Specification**

other home computing platforms, LIGHT employs DomoML (Domotics Markup Language) an XML language for communication between household appliances, standardizing the exchange of messages independently of transport media. DomoML provides a unified vocabulary for the description of devices, services and applications, thus guaranteeing interoperability among heterogeneous domotics architectures such as UPnP, Konnex, X10, and so on. The heart of DomoML is the common type system which can be used to describe services and applications.

The Personal Universal Controller (PUC) improves the interface of household

appliances. It downloads a specification of the appliance's features and then automatically generates an interface to control that appliance. The XML-based specification language abstractly describes the various functions of the appliances. This specification has been provided by the Pebbles project and we thank Jeffrey Nichols and Brad A. Mayers of Carnegie Mellon University for the software.

Another objective of the project is to integrate these XML-based languages into a single architecture to satisfy the need for interoperability among intelligent appliances and the generation of remote control interfaces for complex appliances. However, in order to inte-

grate all these components within an embedded device, LIGHT must provide solutions to minimize the resource requirements. In fact, the current obstacle to hosting Web services on embedded devices is the high requirement of resources in terms of processing power, available memory, and network bandwidth. This is due to the overhead associated with the execution of applications, data processing and marshalling techniques. The processing and memory limitations of embedded devices may severely compromise the usability of XML, unless suitable optimizations are employed. LIGHT will thus investigate and offer efficient solutions in the middleware layer to overcome the constraints of limited resources typically imposed by embedded devices.

This research has been supported partly by the 'Curiosity Driven' programme of ISTI-CNR.

**Link:**

<http://light.isti.cnr.it/>

**Please contact:**

Vittorio Miori, ISTI-CNR, Italy  
Tel: + 39 050 315 3007  
E-mail: vittorio.miori@isti.cnr.it

Luca Tarrini, ISTI-CNR, Italy  
Tel: + 39 050 315 2607  
E-mail: luca.tarrini@isti.cnr.it

## Modelling of Authentic Reflectance Behaviour in Virtual Environments

by Michal Haindl and Jiří Filip

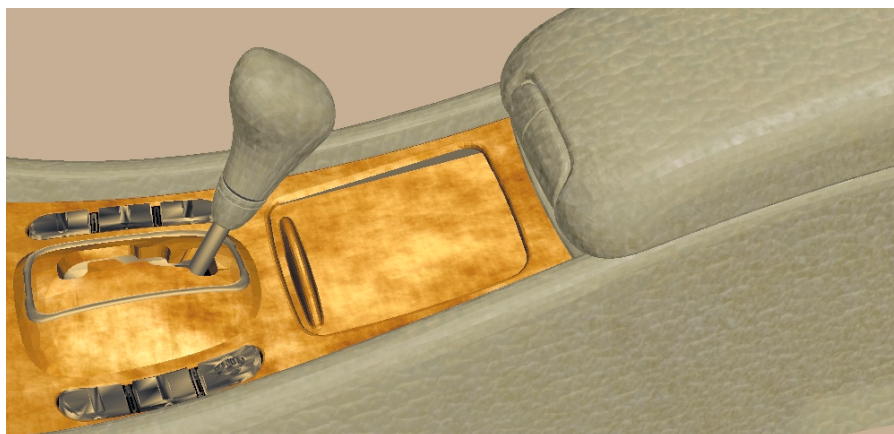
Recent advances in computer hardware and virtual modelling allow the view and illumination dependencies of natural surface materials to be respected. The corresponding texture representation in the form of Bidirectional Texture Function (BTF) enables significant improvements in the realism of virtual models, at the expense of an immense increase in material sample data space. Consequently, the introduction of a method for the fast compression, modelling and rendering of BTF data is inevitable. The development of such mathematical models is among the major objectives of the European RealReflect project.

The goal of this project is to develop a novel, image-based, physically correct visualization technology for VR-systems. The limited capabilities of current graphics systems restrict the appearance of materials covering virtual objects to low-quality unnatural approximations. The project aims to overcome

this restriction by developing a physically correct simulation of light distribution and reflection, as well as an image-based real-time visualization technology for synthetic objects with complex reflectance behaviour. RealReflect – (Real-time visualization of complex reflectance behaviour in virtual proto-

typing) is a joint research project between the University of Bonn; Vienna Technical University; Max Planck Institute for Computer Science, Saarbrücken; Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic (UTIA), Prague; French National Institute for





**Figure 1: The Mercedes C100 gearbox (3D model courtesy of DaimlerChrysler) covered with synthetic BTFs generated by probabilistic models.**



**Figure 2: BTFs generated by the synthetic reflectance model applied to part of the C100 car interior.**

Research in Computer Science and Control (INRIA), Grenoble; DaimlerChrysler AG, Germany; ICIDO GmbH, Germany; Faurecia, France; and Virtual Reality Architects, Austria. The project is funded by the EU IST-Programme. Several major applications have already come from the project, including realistic virtual models in architecture and the automotive industry; for example, the simulation of safety issues such as blinding of the driver by interior lights in a night-driving situation.

UTIA is responsible for development of mathematical models capable of simulating and compressing BTF textures. Modelling of a natural texture is a very challenging task, due to the unlimited variety of possible surfaces, illumination and viewing conditions along with the strong discriminative functionality of the human visual system. The BTF measure-

ments span the whole hemisphere of possible light and camera positions in observed material sample coordinates, according to selected quantization steps. We used the best recent BTF measurements from our University of Bonn project partners, which contain 6561 images per material sample. In their lossless compressed form, these data require about 5 GB per material sample. Such materials present in a car interior VR scene can easily reach the range of terabytes.

UTIA has developed several BTF modelling methods, which can be assorted into three distinct groups: probabilistic models, reflectance models and sampling-based models. As with other texture applications, we have confirmed the experience that there is no ideal BTF model. Models in each group have complementary properties and consequently also optimal application areas.

Our probabilistic models are based either on a set of underlying Markov random fields or probabilistic mixtures, and allow unlimited texture enlargement, BTF texture restoration, huge BTF space compression (up to 1:1000 000) and even modelling of previously unseen BTF data. These methods require neither the storing of original measurements nor any pixel-wise parametric representation. However, such models are non-trivial, and several unsolved theoretical problems exist that must be circumvented. Nonlinear reflectance models offer BTF modelling with excellent visual quality and mild compression ratio (1:200) as well as a fast graphics hardware implementation. Sampling approaches rely on sophisticated sampling from original BTF measurements. They offer high visual quality for most textures, negligible computation complexity but only a moderate compression ratio (1:4). Another drawback of these methods is that unlike our probabilistic models, they do not allow a BTF data space restoration or modelling of unseen (unmeasured) BTF space data. Finally a hybrid method based on Gaussian distribution mixtures was developed with the aim of combining the advantages of both basic texture-modelling approaches. This hybrid model can be used either to directly synthesize textures or to statistically control sampling from the original data.

Regardless of single models traits, they all meet comprehensive requirements such as unlimited seamless BTF texture enlargement, high visual quality and compression. In addition, they include some less obvious features like the strict separation of analytical and synthetic parts, possible parallelization or implementation inside a graphics-processing unit.

#### Links:

<http://www.realreflect.org>  
<http://www.utia.cas.cz/RO>

#### Please contact:

Michal Haindl,  
 CRCIM (UTIA), Czech Republic  
 Tel: +420 2 6605 2350  
 E-mail: [haindl@utia.cas.cz](mailto:haindl@utia.cas.cz)



# Haptic Training Systems in Virtual Surgery

by José San Martín, David Miraut, Carolina Gómez and Sofía Bayona

**In the near future, the learning process of different surgery techniques is going to be based on surgical simulators. On the one hand, emerging Minimal Invasive Surgery (MIS) means minimum patient trauma. On the other hand, the only surgeon assistance is an indirect visualization of the operation, via a video monitor. In this situation, the sense of touch has become a key feature in the training of surgeons. Consequently, there is a need for the development of virtual reality haptic training systems.**

Virtual Reality is acquiring increasing importance due to its potential in many areas of science and technology. Its aim is the manipulation of human senses through computer-generated virtual scenes such that users can interact with the virtual environment intuitively and in real time, without realising it is not real. In order to achieve these goals, virtual reality hardware (eg head-mounted displays, data gloves, tracking systems and haptic devices) is being used.

Thanks to this potential, virtual reality has been applied to education, military training, architectural design, psychology (for phobia therapy) and medicine, in particular for anatomical teaching and surgical training systems such as Minimal Invasive Surgery (MIS).

Typically, most efforts have been focused on composing and recreating synthetic images and sounds. However, this kind of virtual environment, which involves only visual and auditory sensations, is very limited in its ability to interact with users. A growing research community has realized the necessity of the sense of touch in performing precise tasks like surgical simulations and the remote operation of robotics in hazardous environments. During these procedures, the handling of assorted objects is critical, and haptic assistance clearly improves performance.

Recent advances in computing and haptic devices now allow the pressure and temperature sensory receptors in human skin to be 'cheated'. In order to create the sensation of touching and manipulating virtual objects in real time, we need to generate the reaction force over them. These receptors are spread over a large area and are extremely fast,

so the reflecting force must be recalculated over 1000 times per second. This is a complex challenge when we deal with deformable bodies as in MIS.

MIS techniques offer important advantages, such as decreasing patient trauma and reducing costs, but they also have an important drawback: they are complex procedures and difficult to master. Traditional learning methods, assuming the importance of force feedback, use real surgical instruments with phantoms



**A training session with Laparoscopy Impulse Engine.**

(plastic models) for training. However this set of instruments is expensive and delicate; moreover, the plastic models degrade due to the incisions until they become unusable.

Here virtual reality and haptics have come up with the solution. Our team, in cooperation with the company GMV, is developing a successful advanced arthroscopy training simulator, based on virtual reality. Using this simulator, the practitioner learns to handle the surgical instruments and to recognize pathologies. The incorporation of a haptic

device facilitates triangulation and navigation learning, making it possible to feel and distinguish between different tissues.

Current haptic MIS training systems can be classified as either general purpose devices (eg from SensAble PHANTOM) or brand new devices dedicated to surgical simulation (eg Immersion's Laparoscopy Impulse Engine, Virtual Laparoscopic Interface-VLI without force-feedback, and Laparoscopic Surgical Workstation with force-feedback). We have designed a dedicated system, but there remain some limitations that we want to solve.

Shoulder arthroscopy requires a very flexible working area, especially regarding inter-trocar distance (this is fixed in current devices at 135mm). Human body dimensions differ from one patient to another, and this technique necessitates the simultaneous positioning of each

trocar on every side of the patient's shoulder.

We need both surgical instruments to face on, one trocar against the other: this is not normally possible because of the main stand structure. It is therefore necessary to include this additional degree of freedom.

It is crucial to ensure that even if the instruments face each other, and the minimum lateral distance between them has been established, they do not touch one another.

The suggested design, the Laparoscopy Training System, provides the usual five degrees of freedom (including a scissors handle option in order to simulate grasping forces), plus two additional degrees. These are a lateral displacement of every actuator set (adaptive inter-trocar distance) and a rotation of these sets around the device symmetry axis. Moreover, the design allows an ergonomic positioning of the surgical

instruments depending on the preference of the trainee.

Our work is focused on designing an MIS training system to simulate specific laparoscopy techniques that are not generally considered. Having finished the design stage, a prototype is under construction, and a working device should be achieved by summer 2005.

Nowadays, the features of physical devices restrict programmers due to limitations in the number of degrees of freedom. We think that this design will help to solve these constraints.

**Link:**

<http://dac.escet.urjc.es/investigacion/GMRV/>

**Please Contact:**

José San Martín, Universidad Rey Juan Carlos/SpaRCIM, Spain  
E-mail: jose.sanmartin@urjc.es

## Computer Recognizes Whale Tails

by Annette Kik, Eric Pauwels and Elena Rangelova

**How many whales are there in the ocean and how do they migrate? To answer these questions it is important to identify individual animals. Until now, biologists have tried to search by hand through vast numbers of photographs. In the European EUROPHLUKES project that ended in 2004, CWI researchers developed a method for semi-automatic pattern recognition of whale tails and dorsal fins. This is a first step towards the automatic photo-identification of individual animals.**

Governments and organizations such as the International Whaling Commission want to know more about whale populations in order to protect biodiversity and to make informed choices about possible hunting within certain limits. Using parameters such as the number and age of female animals, scientists can estimate mathematically how the population will evolve. Identification is an important tool in collecting these data for stock management. One of the most convenient ways is photo-identification. It is less intrusive than harpooning whales for a sample of DNA, and is more extensive because of the potentially large collections of photographs from biologists, sailors and tourists.

The EUROPHLUKES project commenced in 2001. Its brief was to develop a photo-ID system and database for cetaceans - whales, dolphins and porpoises. The objective was to be able to identify if a particular cetacean had already been photographed, and if so, where and when. The network comprises more than forty partners and participants, mostly marine biologists. It is coordinated

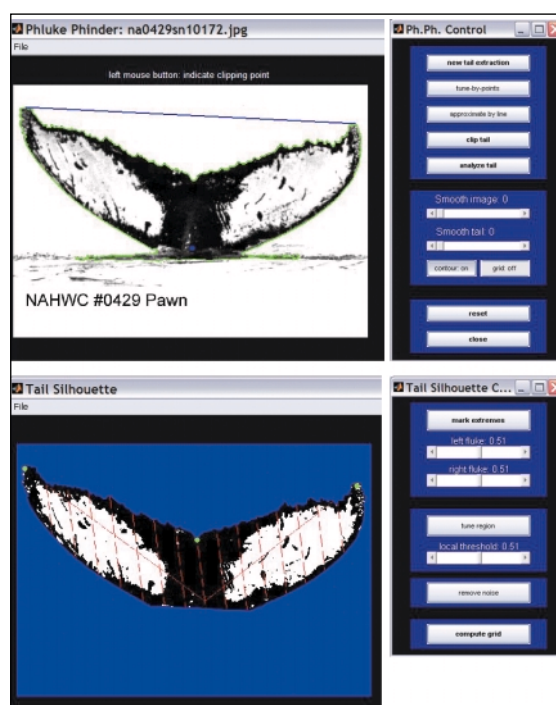
by the Universiteit Leiden and is funded by the Fifth Framework of the European Union. To deal with specific computer vision problems, researchers from CWI's Signals and Images group, a member of ERCIM's WG on Image and

Video Understanding, were invited to the team.

### Watershed Method

Human beings can easily identify individual whales, due to the unique spots and scars on the animals' skin and the shape and indentations of their tails or dorsal fins. However, they can only compare a few pictures at a time, whereas photographic collections are growing rapidly. A computer on the other hand, can quickly compare thousands of pictures in databases but it has great difficulties in spotting similarities. For instance, a tail can look different when it is turned, waves can occlude specific marks and the picture quality can vary enormously. In a black and white picture it is not always easy to distinguish between tail and water. In addition, a computer is not intelligent so it cannot immediately recognize the most important marks: it has to compare all features, big and small.

To recognize individual characteristics semi-automatically, CWI researchers combined and applied several mathematical techniques:



**Individual whales can be identified by their markings. CWI's software system Phluke\_Phinder detects the shape of the tail and the location of its various spots.**

image segmentation, contour and feature extraction and finally, comparison of these data with an image database. First, the grey-level of the image is represented as a three-dimensional picture: white is high, darker is lower. At the edge between sea and tail the difference in grey-scale will be large, or in other words, the gradient will be high. The picture of this gradient can be viewed as a 'topological surface': It has mountains, plains and valleys. In the so-called watershed method, virtual water floods this surface. When the water is so high that two lakes in valleys are about to merge, a virtual watershed is placed. This procedure has been programmed in MATLAB, a technical computing language, allowing the computer to robustly identify regions of similar grey scales and thus extract the contours of the tail or dorsal fin.

### Spots and Scars

To compare spots and scars on a tail photographed from nearby with one that is photographed from a larger distance or a different viewpoint, the researchers attach a virtual grid to the tail that is always the same, or in other words 'invariant under affine transformations'. This grid is used as a coordinate system. To define it, the user specifies three anatomical points in the tail: the middle notch of the fluke and both its tips. Assuming that the tail is not too flexible, parallelism and relative distances can be used to divide the tail into a large number of small regions (eg thirty). The proportion of spots to background in each of these regions can then be represented in a 30-dimensional vector, which can be compared with other vectors in the database of identified animals.

The performance can be improved by combining the above feature vector with a more detailed mathematical description of specific, salient spots and scars on the animal. With morphological processing these marks can be found - both their centre of gravity with respect to the grid and a computed approximating ellipse. Using these data, the computer gives a top list of potential matches with pictures from the database. The user can then pick the actual match from this shortlist or confirm that there is no matching animal in the database. This method makes it possible to compare pictures with larger cetacean databases.

#### Links:

<http://www.europhlukes.net>  
<http://homepages.cwi.nl/~ely/projects.htm>  
<http://homepages.cwi.nl/~pauwels/PNA4.1.html>

#### Please contact:

Eric Pauwels, CWI, The Netherlands  
 Tel: +31 20 592 4225  
 E-mail: [Eric.Pauwels@cwi.nl](mailto:Eric.Pauwels@cwi.nl)

## Text Document Classification

by Jana Novovičová

**During the last twenty years the number of text documents in digital form has grown enormously in size. As a consequence, it is of great practical importance to be able to automatically organize and classify documents. Research into text classification aims to partition unstructured sets of documents into groups that describe the contents of the documents. There are two main variants of text classification: text clustering and text categorization. The former is concerned with finding a latent group structure in the set of documents, while the latter (also known as text classification) can be seen as the task of structuring the repository of documents according to a group structure that is known in advance.**

Document classification appears in many applications, including e-mail filtering, mail routing, spam filtering, news monitoring, selective dissemination of information to information consumers, automated indexing of scientific articles, automated population of hierarchical catalogues of Web resources, identification of document genre, authorship attribution, survey coding and so on. Automated text categorization is attractive because manually organizing text document bases can be too expensive, or unfeasible given the time constraints of the application or the number of documents involved.

The task of text categorization (TC) is the focus of a research project at the

Institute of Information Theory and Automation, at the Academy of Sciences of the Czech Republic (UTIA). The dominant approach to TC is based on machine learning techniques. We can roughly distinguish three different phases in the design of TC systems: document representation, classifier construction and classifier evaluation. Document representation denotes the mapping of a document into a compact form of its content. A text document is typically represented as a vector of term weights (word features) from a set of terms (dictionary), where each term occurs at least once in a certain minimum number (k) of documents. A major characteristic of the TC problem is the extremely high dimensionality of

text data. The number of potential features often exceeds the number of training documents. Dimensionality reduction (DR) is a very important step in TC, because irrelevant and redundant features often degrade the performance of classification algorithms both in speed and classification accuracy.

DR in TC often takes the form of feature selection. Methods for feature subset selection for TC tasks use some evaluation function that is applied to a single feature. The best individual features (BIF) method evaluates all words individually according to a given criterion, sorts them and selects the best subset of words. Since the vocabulary usually contains several thousand or tens of



thousands of words, BIF methods are popular in TC. However, such methods evaluate each word separately, and completely ignore the existence of other words and the manner in which the words work together.

UTIA proposed the use of the sequential forward selection (SFS) method based on novel improved mutual information measures as criteria for reducing the dimensionality of text data. These criteria take into consideration how features work together. The performance of the proposed criteria using SFS compared to mutual information, information gain, chi-square statistics and odds ratio using the BIF method has been investigated. Experiments using a naive Bayes classifier based on multinomial model, linear support vector

machine (SVM) and k-nearest neighbour classifiers on the Reuters data sets were performed and the results were analysed from various perspectives, including precision, recall and F1-measure. Preliminary experimental results on the Reuters data indicate that SFS methods significantly outperform BIF based on the above-mentioned evaluation functions. Furthermore, SVM on average outperforms both Naive Bayes and k-nearest neighbour classifiers on the test data.

Currently, text classification research at UTIA is heading in two directions. First, investigation of sequential floating search methods and oscillating algorithms (developed in UTIA) for reducing dimensionality of text data; and second, design of a new probabilistic model for

document modelling based on mixtures for simultaneously solving the problems of feature selection and classification. These phases of the project rely on the involvement of PhD students from the Faculty of Mathematics and Physics at Charles University in Prague.

This research is partially supported by GAAV No A2075302, the EC MUSCLE project FP6-507752 and the project 1M6798555601 DAR.

**Link:**

<http://www.utia.cas.cz/RO/FD>

**Please contact:**

Jana Novovičová,  
CRCIM (UTIA), Czech Republic  
Tel: +420 2 6605 2224  
E-mail: novovic@utia.cas.cz

## Advancing Black-Box Reuse in a Multimedia Application Framework

by Bernhard Wagner

**MET++ is a tool for producers of multimedia packages. A visual programming environment avoids the need to set up a development environment to author multimedia applications.**

Multimedia production environments like Macromedia Director face the challenge of presenting inherently complex technologies in a simplified, accessible way to the multimedia producer. When it comes to interactivity that exceeds the most basic operations (eg following links, 'next/previous slide' or 'start/stop animation'), these production environments opt for scripting. While a proprietary scripting language allows for a vocabulary exactly fitted to the domain at hand, such an approach can become complex and cumbersome, due to limitations of the metaphors used or because the difficulty of scripting is often underestimated.

Like most frameworks, the object-oriented multimedia application framework MET++ has evolved into a state where most tasks can be achieved by black-box rather than white-box reuse, or by composition rather than inheritance. To avoid having to set up a whole

development environment to author multimedia applications, a visual programming environment has been conceived and implemented as an orthogonal control environment to the abstractions already available within MET++.

### MET++

MET++ is a portable object-oriented C++ multimedia application framework developed at the University of Zürich. It is based on the object-oriented application framework ET++. ET++ consists of several frameworks, which support the development of desktop applications with graphical user interfaces. ET++ has a layered architecture addressing the following goals: portability among operating systems and windowing systems, generic data structures, support for graphic user interfaces, and desktop applications. The abstractions in ET++ are highly integrated and anticipate all generic interaction between application

components. Thus a developer using the framework need only fill predefined slots with the application-specific content.

MET++ is built on top of ET++, adhering entirely to the architecture and style defined by ET++. The multimedia extensions provided by MET++ are:

- 3D graphics
- audio and music
- video
- time synchronization
- visual programming.

MET++ has proven its worth in numerous multimedia projects and is used in commercial applications as well.

Experience in student assignments has shown that for a developer versed in the MET++ framework, application development is very efficient. The newcomer, however, must expend significant time and effort to get used to the framework and its abstractions. The difficulty comes

from the fact that, like typical frameworks, MET++ and ET++ have evolved into a state where most tasks can be accomplished by black-box reuse and composition of already existing classes. A software system functioning in a black-box manner is hard to understand if one has only the source code to examine.

This observation led to the development of an authoring environment that not only visually displays the current application as a composition of black-box components, but also allows applications to be built visually: the visual programming environment of MET++.

### Visual Programming

The metaphor of the visual programming environment in MET++ is that of a data-flow engine. Its building blocks are so-called DataUnits and DataPorts. There are several categories of DataUnits: mathematical functions, GUI components, wrappers, data containers and data mappers. The DataPorts provide the input/output to the DataUnits.

The visual programming environment in MET++ uses the Adapter design Pattern

to wrap existing media abstractions, thus enabling their control from within the visual programming environment. Using this environment, a user can explore the behaviour and protocol of a media abstraction available in MET++ before programming against its API using C++.

A special DataUnit type has been developed explicitly for the visualization of data, namely, the Mapper. This unit provides the necessary infrastructure for ascertaining the number and cardinality of rectangular data and for subsequently iterating over all dimensions. The data responds to the index sets generated by the mapper with the corresponding data content, which is then visualized by the mapper. Several kinds of mappers have been developed that allow visualization as two- or three-dimensional graphics and sonification of the data.

Entire applications can be built in the visual programming environment. A special HTML-Browser has been developed that allows visual programs to be embedded in normal HTML pages in the same way as applets or flash components. Likewise, plug-ins have been

developed for the Mozilla Firebird and Microsoft Internet Explorer browsers.

### Applications

The visual programming environment has been successfully employed in the areas:

- interactive data visualization
- animation
- sonification of animation
- visualization of sound.

### Outlook

The author is currently preparing didactic material for introductory lectures in 3D graphics, using the visual programming environment of MET++. This material will interactively demonstrate concepts of computer graphics and will also contain student exercises.

#### Links:

MET++ material:

<ftp://ftp.ifi.unizh.ch/pub/projects/met++/>

Commercial use of MET++:

<http://www.perspectix.com>

Publications about Visual Programming in MET++: <http://xmlizer.biz/priorArt.html>

#### Please contact:

Bernhard Wagner

University of Zurich, Switzerland

E-mail: [micre0505@xmlizer.biz](mailto:micre0505@xmlizer.biz)

## CASSEM: Vibration Control in the Smart Way

by Salim Belouettar

**The ambition of the CASSEM project is to define the 'best' models and techniques that will permit us to model, simulate and validate the development of a more efficient vibration control.**

In many industrial and defence applications, noise and vibration are important problems. In recent years, the control of sound and vibration has been the subject of much research, and there are now numerous examples of such applications. The most common general classification of vibration control differentiates between passive, active and hybrid passive-active control. Passive control involves the use of reactive or resistive devices that either load the transmission path of the disturbing vibration or absorb vibratory energy. Active control also loads the transmission path but achieves this loading through the use of actuators that generally require external energy. In

passive control, the material properties of structure such as damping and stiffness are modified so as to change the response of structure. In active control, the structural response is controlled by adding external effort to the structure.

Combining these two approaches, hybrid control integrates the passive approach with an active control structure, and is intended to reduce the amount of external power necessary to achieve control.

Normally an adaptive or smart structure contains one or more active or smart materials. It is the use of these materials that causes the whole structure to be

classified as 'smart'. These materials have the ability to change their shape, rheological properties (eg stiffness and damping), or internal electrical properties (eg dielectric constant or resistivity).

Depending on the relative positions of the viscoelastic layer and the piezoelectric actuator, the viscoelastic passive and piezoelectric active actions can operate either separate or simultaneous actions. Typically, these materials have a sandwich structure, in which a soft, thin viscoelastic layer is confined between identical stiff, elastic layers. These structures yield a superior energy absorption. In particular, they offer the advantage of

high damping with low weight addition. The interlayer-damping concept is highly compatible with the laminated configuration of composite structures and with their fabrication techniques, and provides an effective way to reduce vibrations and noise in structures. The damping is introduced by an important transverse shear in the viscoelastic layer. This is due to the difference between in-plane displacements of the elastic layers and also to the low stiffness of the central layer.

The performance of passive and hybrid control systems depends strongly on the viscoelastic material layer and piezoelectric material properties. In this project, numerical identification based on direct/inverse approaches will be developed. An advanced non-contact laser technique (ISI-SYS vibrograph system and Polytec scanning vibrometer) will be

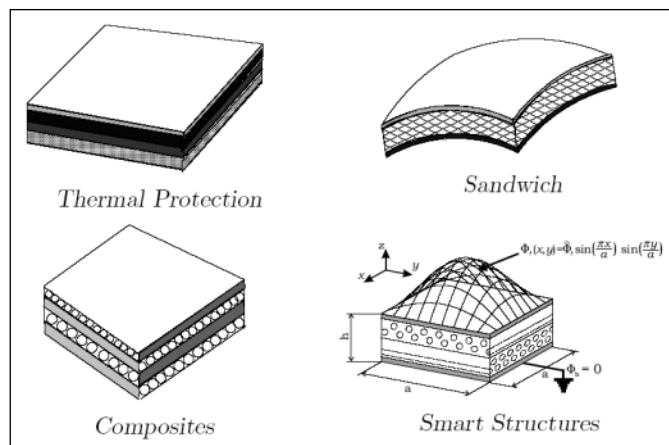
and experimental testing through neural networking.

The second approach considers optimization techniques. An error function, eg the least squares sum of the difference between modelling and experimental results, is minimized by changing the material parameters using different methods. These are either deterministic methods involving semi-analytical or numerical gradient procedures or stochastic methods like genetic algorithms or shooting methods.

The basic idea of the third proposed approach is that simple mathematical models (response surfaces) are determined only by the finite element solutions in the reference points of the experiment design. The function to be minimized describes the difference between

the measured and numerically calculated parameters of the response of structure. By minimizing the function, the identification parameters are obtained.

Another issue of this FP6 project is to develop a general analytical and numerical (Finite Element) framework to model:



#### Multi-layered structures.

applied for the vibration measurements. These experimental data will be used to determine the natural frequencies and corresponding loss factors by the developed modal analysis program.

Three approaches are put forward for retrieving the material parameters. The first approach is the use of neural networks as a regression analysis tool. With the development of neural networks, it has become possible to perform a meaningful parameter extraction without the knowledge of an analytical relationship between the material properties and test values. This allows mechanical parameters to be identified by combining finite element modelling

composite structures with piezoelectric sensors and actuators, (ii) thermal and pyroelectric effects in piezoelectric composites, and (iii) piezoelectric shunted damping.

In the context of the FP6 STREP Project 'CASSEM' (Composites and Adaptive Structure: Simulation, Experimentation and Modelling), we will also design a robust controller, which is stable in the presence of uncertainties of modelling and parameters, and ensures optimal disturbance rejection capability. In the implementation of the controller, actuators and sensors are needed. The locations of actuators and sensors over a structure determine the effectiveness of the controller in damping vibrations.

A variety of problems must be clarified before active systems can be implemented within structures. One of these - an important and not fully recognized problem - is the proper positioning of sensors and actuators on structures in the case of active systems, and the location of dampers in the case of passive systems. In active vibration control, actuator and sensor placement is a very significant issue, since it has a direct effect on the control efficiency and cost. For example, large flexible structures require many actuators for active vibration control, and the problem of optimizing their location becomes extremely significant in maximizing system controllability. An arbitrary choice of actuator positions can seriously degrade the system performance. The controllability index, the genetic algorithms, the gradient-based optimization procedure and the heuristic procedures were used to determine the proper sensor or actuator locations. The problem of positioning and size of passive dampers is also important, for similar reasons.

CASSEM is a highly interdisciplinary project combining engineering and physical sciences, for example, experimental material science, numerical modelling methods, mathematics, automation systems and mechatronics. The consortium consists of nine scientific and industrial partners from seven EU countries. The application of the project results will lead to long-term innovations in composites and adaptive structures. The development of vibration control systems will allow the areas of application of multi-functional composite materials to be extended based on the advanced knowledge and understanding of vibration response.

#### Links:

CASSEM: <http://www.cassem.lu>  
Centre de Recherche Henri Tudor:  
<http://www.tudor.lu>

#### Please contact:

Salim Belouettar, Centre de Recherche  
Henri Tudor, Luxembourg  
Tel: +352 54 55 80 500  
E-mail: [salim.belouettar@tudor.lu](mailto:salim.belouettar@tudor.lu)



## Point6: The IPv6 Skill Centre — Moving to the Next-Generation Internet Protocol

by César Viho and Annie Floch

IPv6 was initially conceived to anticipate the shortage of IP addresses, and to respond to the deficiencies in IPv4 that appeared in the 1990s. These requirements concern both emerging communication technologies (cell phones, PDAs, Internet-connected vehicles, xDSL etc) and billions of new users (in Japan, China, India etc), and they become more significant daily. It is therefore vital that IPv6 be introduced into infrastructures and new projects.

Since January 1995, IPv6 has been specified and standardized with the aim of meeting global needs generated by the huge growth of the Internet. In addition to the increased (virtually unlimited) address space, IPv6 offers numerous major design improvements over IPv4. These can remedy the 'old' protocol limits, as well as opening the door to innovative Internet-based products and services. Among the most significant of these enhancements are auto-configuration ('plug and play') and reconfiguration mechanisms, mobility services, end-to-end security - with IPSec encryption and authentication features - and enhanced support for multicast and QoS. Today, with these technical advantages, IPv6 is widely available from industry and supported by recent network equipments (routers, switches, desktops, servers and operating systems). It has also been successfully applied worldwide by educational establishments, government institutions, telecom companies and research organizations. However, European companies are still behind schedule in the acceptance and deployment of this technology. The common attitude appears to be: "IPv4 is well known and functional. Why should I change? What are the benefits and the associated costs and risks?"

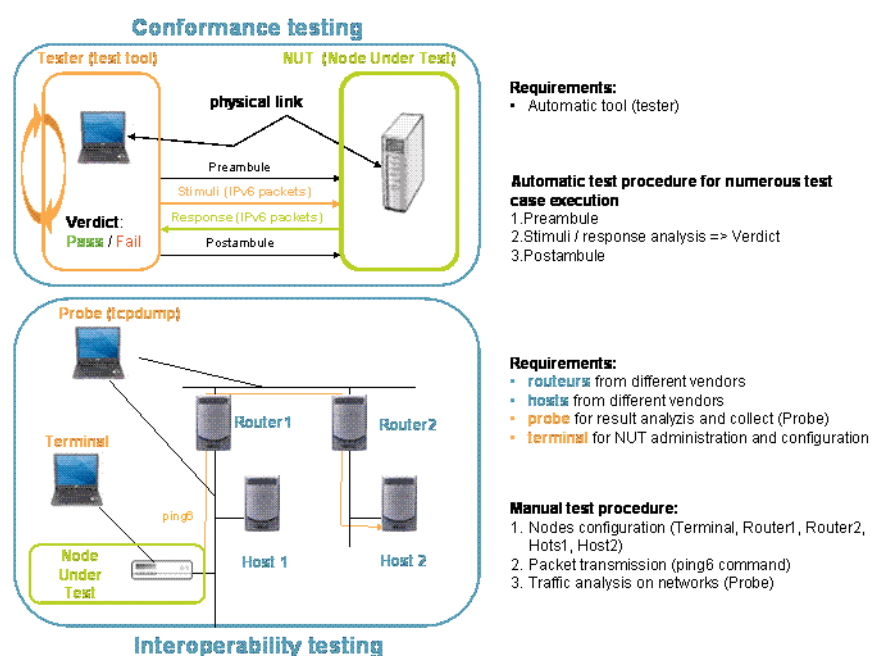
Researchers from the IRISA/INRIA and ENST Bretagne/GET laboratories have been involved in the IPv6 project from the very beginning. They co-developed the very first IPv6 stack ten years ago, and have been involved in the IPv6 interoperability testing and the international certification program (IPv6 Ready Logo) since 1999. This expertise develops daily through research activities, IETF standardization efforts, educational duties,

technical publications and conferences, prototype implementation and international interoperability test events. The team has been further reinforced by the creation of the Point6 skill centre in Rennes (France). This unique know-how is now available for IT and business managers and engineers who need to understand and control all the IPv6 related issues, prior to organizing any associated work.

IPv6 protocol is stable and mature but not well known, possibly because of a perception that it is difficult to learn. Indeed, IPv6 is specified by a collection of more than 60 RFCs (Requests for Comments - Internet standards). Companies that need to implement IPv6 in a product naturally ask the following questions: Which IPv6 skills should I

develop regarding my product specification? Where can I find an IPv6 stack for my real-time OS? Which part of my code do I have to modify to remain IPv6-compliant? How long will it take to complete the work?

The ambition of the Point6 experts is to understand and analyse the industry needs, provide efficient solutions, and assist the R&D teams in their adoption of the IPv6 protocol, from the initial requirements to the implementation details. From our experience, the main difficulty is the initial step of IPv6 knowledge acquisition. Once this is achieved, the technical issues are generally well known and decision are quite easy to take. Usually, products have to be both IPv4 and IPv6 compliant, and this can be managed with a minimum of effort.



Conformance and interoperability testing.

Another representative case is the network infrastructure migration from IPv4 to IPv6. Most organizations have difficulty making accurate decisions at the right moment, even though they know that IPv6 is now reliable and commonly used in products and from Internet Service Providers. Point6 experts can analyse existing infrastructure and are capable of developing customized migration plans. The main steps typically include IPv6 connectivity, address plan definition, equipment upgrade, routing configuration, network services reconfiguration (firewalls etc), service updates (mail, Web etc), end user equipment and metrics definition. This can be done gradually, retaining the same level of service as with IPv4.

Companies that have developed their IPv6 products or want to validate any IPv6 network configuration can use the Point6 platform for both conformance and interoperability testing facilities. Conformance tests verify that an implementation has been developed in strict accordance with its specifications (RFCs). Multi-vendor environments make it necessary to validate equipment compatibility by applying interoperability scenarios.

Testing is time-consuming, and requires skill, specialized resources (hosts, routers, probes etc) and rational impartiality. The Point6 test service is based on a platform supplied with up-to-date tests and managed by IPv6 experts. The

team has the relevant skills for helping companies to obtain the 'IPv6 Ready Logo', which requires 100% correct results for both conformance and interoperability tests. More tests for IPv6 are available in the laboratory for additional IPv6 routing protocols (RIPng, OSPFv3, BGP4+), mobility (Nemo) and transition mechanisms (NAT/PT, 6to4). Other test suites are also under development.

**Links:**  
Point 6: <http://www.point6.net>

IPv6 ready Logo Program:  
<http://www.ipv6ready.org/>

**Please contact:**  
César Viho, IRISA, France  
E-mail: [cesar.viho@irisa.fr](mailto:cesar.viho@irisa.fr)

# Working Slower with More Powerful Computers

by Lorenz M. Hilty, Andreas Köhler, Fabian van Schéele and Rainer Zah

**An empirical study conducted by the Swiss Federal Laboratories for Materials Testing and Research showed that using a more powerful PC can significantly slow down office workers in performing every-day office tasks.**

The average service life of a personal computer used for business is 2-3 years. After this short service life, it is usually replaced by a more powerful PC running new versions of system and application software. Therefore, it is worth considering whether the high replacement costs of PCs bring a comparable benefit. Whereas the benefits from having additional functionality (eg, in multimedia applications) are self-evident, it is not a priori clear whether pre-existing functionality can be accessed and used as efficiently as before when the user goes to a new machine. The study reported here addresses only the latter issue.

Our laboratory experiment was based on PC configurations typical of the years 1997, 2000 and 2003. As Table 1 shows, the hardware used differed greatly in clock speed, working memory and hard disk capacity. Since the test only involved file handling and text editing, it was sufficient to install the operating system and a word processor. The configurations A, B and C are reconstructions of PC configurations used

widely in business in the respective year. All software components were installed taking the default parameter settings. The appearance of the desktops was kept similar on all computers to optimally satisfy blind test conditions. A switch box made it possible to connect the same mouse, keyboard, and monitor to each of the three computers.

We defined two tasks for testing user performance on the three systems: a file handling and a text editing task. The file handling task included locating, copying, moving and deleting files and

groups of files varying in size from 24 kByte to 42 MByte. The text editing task focused on locating, inserting and deleting text, changing fonts and formats in a large document of 150 pages as well as copying, pasting and positioning photographs between two documents.

The laboratory experiment was performed on 42 subjects recruited from a service organization with 180 employees. Subjects had to execute each of the tasks twice on each computer system in randomized order. The whole test took 45-90 minutes per subject. We

measured the time the subjects needed to complete each task, logged user interactions and processor workload on the computer and video-taped the screen signal.

The test revealed a statistically significant decrease in user

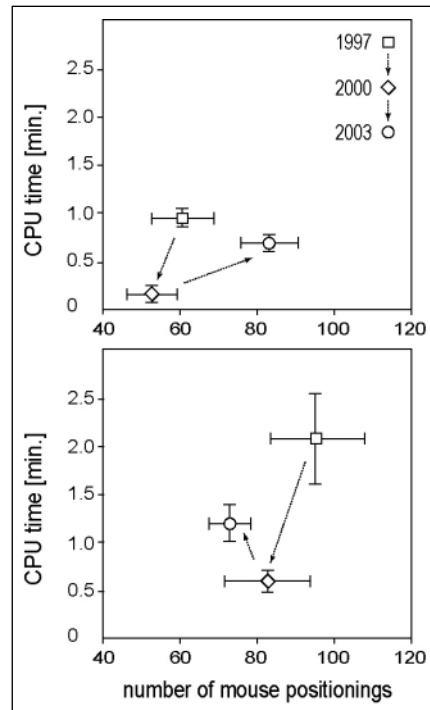
	A	B	C
Year	1997	2000	2003
CPU frequency	233 MHz	801 MHz	1,992 MHz
RAM capacity	64 MB	128 MB	256 MB
Hard disk capacity	2 GB	20 GB	56 GB
Operating system	Microsoft Windows NT	Microsoft Windows 2000	Microsoft Windows XP
Word processor	Microsoft Word 97	Microsoft Word 2000	Microsoft Word 2002

**Table 1:**  
**Description of the computer systems used for the experiment.**

performance on Windows XP as compared to Windows 2000 for the file handling task, despite the fact that Windows XP was run on considerably faster hardware. This result was totally unexpected because Windows 2000 had never been used in the organization in which the subjects were employed. Windows NT systems had been in use for five years. Approximately five months before the test, a migration from Windows NT to Windows XP had been initiated. By the time of the test, virtually every employee was working with Windows XP. Only four subjects reported that they usually worked with Windows 2000, and another four were still working with Windows NT.

In order to find explanations for our result, we examined the effort required from the user and the machine to complete the tasks in detail. Since most user actions during task execution were mouse positionings (usually followed by a single mouse click, less often by no click or a double click), we used the number of mouse positionings as an indicator of the user's effort. The computer's work was measured in CPU time, which we calculated by integrating processor utilization over time.

As Figure 1 (top) shows, the file handling task required much less CPU time on system B (year 2000) than on system A (1997), but again much more on system C (2003), and the same is true for the number of mouse positionings the users performed (significant difference,



**Figure 1: Human effort, measured by number of mouse positionings, versus machine work, measured by CPU time, for executing the tasks for the second time on a given computer. Top: mean values for the file handling task. Bottom: mean values for the text editing task. Error bars denote s.e.m. Arrows indicate the temporal sequence of systems.**

$P=0.016$ ). Whereas they managed with roughly 50 mouse positionings on Windows 2000 on the average, they needed more than 80 positionings for the same task on Windows XP.

The text editing task (Figure 1, bottom) shows a decrease in mouse positionings

from A to B to C, but an increase in CPU time from B to C. Given the fact that the C hardware executes more than twice as much many instructions per second than the B hardware, one would instead expect a significant decrease.

The main conclusion from this experiment is that changing over to faster a computer running newer software does not necessarily lead to better performance, at least for the types of tasks we used in our experiment. On the contrary, it is even possible that both – the machine and the user – need to work significantly longer in order to replicate a given task with the new system.

### Acknowledgements

The work reported here was conducted in cooperation with the Swedish Royal Technical University (KTH), Stockholm, and co-funded by the Board of the Swiss Federal Institutes of Technology (ETH board) as a part of the 'Sustainability in the Information Society' research program at Empa.

#### Links:

Empa: <http://www.empa.ch>

Technology and Society Lab at Empa: <http://www.empa.ch/TSL>

Sustainability in the Information Society Program: <http://www.empa.ch/SIS>

#### Please contact:

Lorenz M. Hilty, Technology and Society Lab, Empa, Swiss Federal Laboratories for Materials Testing and Research, Switzerland  
Tel: +41 71 2747 500  
E-mail: [lorenz.hilty@empa.ch](mailto:lorenz.hilty@empa.ch)

## Coordinating IST Research Across Europe

by Simon Lambert

**The CISTRANA project will help to enhance European competitiveness through international coordination of IST research activities**

The future prosperity of Europe depends on its competing successfully in a number of key knowledge-based industries, of which the so-called information society technologies (IST) form a large economic block. To achieve this economic success in the longer term, it is necessary to circumvent the barriers

posed by national boundaries and differing research policies and priorities, and to bring together a critical mass of resources from across Europe to research key technologies.

The CISTRANA project is a Coordination Action under the European

Commission's IST programme, with the aim of achieving coordination of national IST programmes with each other and with European programmes in order to improve the impact of all research and development efforts in Europe and to reinforce European competitiveness in this area.



CISTRANA is of three years duration, and involves five partners from across Europe. DLR in Germany is the coordinator, and the others are TEKES (Finland), NKTH (Hungary), ANRT (France) and CCLRC (UK, an ERCIM institute). The objectives of the project are to develop a map of the national research landscape in the area of IST; to pinpoint research topic areas and strategic themes where cooperation is essential; and to establish sustainable mechanisms including common methodologies and procedures to set up trans-national coordination initiatives between different countries.

It is evident that active participation and commitment of stakeholders is essential for the success of CISTRANA. A Steering Committee has been set up, with members from around thirty countries, that gives advice on information collection and guidance on priorities and implementation. The Steering Committee acts as the interface between the CISTRANA project and national ministries. Each of the CISTRANA partners, with the exception of CCLRC whose role is explained below, is a Country Group Representative, with the responsibility of gathering and assessing information collected at the national level. The information itself is provided by National Support Organisations, one

in each country, selected by the Steering Committee member.

The project started in September 2004 and so far a comprehensive survey has been conducted about national research policies, programmes and other activities. The response has been very good, and the results are currently being assessed to feed into other project activities. A series of workshops are being planned during the life of the project, aimed at various target audiences, with themes such as impact assessment, best practice in multi-national programme collaboration, and portals for information dissemination.

An additional important output of the CISTRANA project is a portal to IST research information. A study conducted for the European Commission addressed the perceived difficulty in accessing information on IST research at national levels. The main findings were that there are important barriers to accessing relevant information, including a lack of consistent thematic search possibilities; a lack of relevant information; and a lack of comparability between information sources. It concluded that the overall presence of information was insufficient in comparison with users, requirements, and that the barriers can be largely overcome by establishing an IST research

portal as a gateway providing access to national IST RTD information.

CCLRC is the partner responsible for developing the portal. The aim is to give access to diverse kinds of information about national programmes, policies and projects, either gathered specially for CISTRANA or in existing national databases. A system architecture has been developed that is sufficiently flexible to accommodate these possibilities, and in particular allows different options to national database owners for how they can make their data available via the portal. The CERIF standard is of key importance here: it is the Common Research Information Format that has been developed over many years as a recommendation to member states of the EU for Current Research Information Systems. It is expected that the portal will be of great interest to various user communities including policy makers, researchers and those in the industrial world, offering a way of conveniently finding out what is going on in other countries.

**Links:**

Project website: <http://www.cistrana.org>

CERIF standard:

<http://www.eurocris.org/en/taskgroups/cerif/>

**Please contact:**

Simon Lambert, CCLRC

Tel: +44 1235 445716

E-mail: [S.C.Lambert@rl.ac.uk](mailto:S.C.Lambert@rl.ac.uk)

## Software Automation meets Interactive Media Development

by Dirk Deridder, Thomas Cleenewerck, Johan Brichau and Theo D'Hondt

**Developing the software component of interactive media requires an advanced set of development tools and enforces a different view on software development in general. This is mainly due to the specific characteristics of the environment in which this development takes place. We propose a combination of software automation techniques to counter this effectively. Since the iMedia domain is in continuous flux, and these technologies are mostly designed for stable domains, the evolvability of the approach is guaranteed by rooting it into the heart of the system.**

Media production companies, in particular media broadcasters, are being challenged by their consumers to produce media in which the consumers can

actively participate and interact with TV shows and between peers. This new form of media, called interactive media, is a combination of traditional media and a

behavioural component (software). Examples of such components are online games, quiz software, virtual community worlds etc.

The challenge lies in the production of the behavioural component of the new form of media within the broadcasting environment. The production cycle is constrained by the imposition of extremely strict broadcasting times, the extremely short time-to-market situation caused by last-minute changes, and the extreme deployment prerequisites.

### IMedia Software Generation System

IMedia development requires more advanced development tools. The approach we propose combines existing research from the areas of generative programming, transformation systems and domain engineering. This results in a system that is best described as an

nology. The system and the techniques used to build the system were designed with evolution in mind: the impact of changes to the system is limited to individual easily identifiable modules.

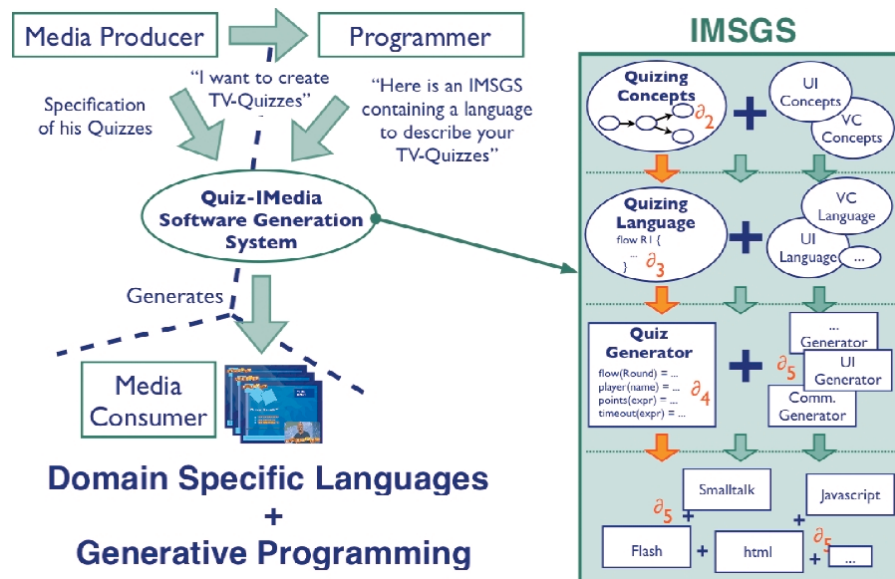
An IMSGS for a specific product range is divided into a set of program generators, each targeted at a specific concern in the product (eg the application logic, the graphical user interface etc). For each concern a domain model with CoBro is constructed that defines the concepts in the domain and the relationships between them. These models are then used to construct concise domain-specific languages (DSLs) compliant with the definitions in the domain model. The DSL compilers are program genera-

mentations in the quiz language. In this way, one can start at the level of the domain concepts to estimate which parts of the implementation ( $\delta 3$  in the figure) will be affected by the evolution ( $\delta 2$  in the figure). Moreover, connecting the domain knowledge to the implementation provides a valuable source of documentation of the assumptions made by the original developers.

The DSLs are constructed using the Linglet Transformation System via a composition of language components, which is expressed in a language specification. LTS modularizes the language components by specifying the necessary communication patterns among them in a separated language specification, through the customization of the language components. Hence the dependencies among the language components become explicit and are removed from their implementations. Consequently the impact of changes in the language ( $\delta 3$  and  $\delta 4$  in the figure) is isolated to the language specification and to individual identifiable components.

The composition of the program generators (DSL compilers) is realized with Generative Logic Meta-Programming. GLMP features a grey-box composition model of program generators that allows the specification of integration relationships among the subparts of different program generators. This mechanism is vital for adapting the generators so that they produce program parts; these can then be combined into a single application with no undesired interferences that could break their functionality ( $\delta 5$  in the figure).

This research is performed in the context of the Advanced Media Project, a collaboration between Vlaamse Radio- en Televisieomroep, Vrije Universiteit Brussel, Universiteit Gent and IMEC.



Overview of the IMedia Software Generation System with the major evolution d's in an IMSGS.

iMedia Software Generation System (IMSGS; see the figure), specialized for each product range. In an IMSGS, more autonomy and flexibility is given to the media producer to adapt the iMedia software product. This is achieved by generating different tailor-made 'instances' of the product range, given a high-level specification. The tailoring of particular instances is managed by the media producer (the domain expert).

### Evolution of the IMSGS

Our research focuses on the evolvability problems of a system that is based on DSL and generative programming tech-

tors implemented in Linglet Transformation System (LTS), which translate the DSL specifications into executable code in some generic language using generic libraries and frameworks. Finally these program generators are composed using Generative Logic Meta-Programming (GLMP) in order to integrate each of their generated program parts into one application.

The domain knowledge is described in CoBro. CoBro follows a concept-centric approach in which we couple the domain concepts to their corresponding imple-

#### Links:

<http://prog.vub.ac.be/>

<http://prog.vub.ac.be/Publications/2004/vub-prog-tr-04-19.pdf>

<http://www.xmt.be/sake.html>

#### Please contact:

Dirk Deridder

Vrije Universiteit Brussel, Belgium

E-mail: [Dirk.Deridder@vub.ac.be](mailto:Dirk.Deridder@vub.ac.be)

## Workshop on Challenges in Software Evolution

by Tom Mens



The annual meeting of the ERCIM Working Group on Software Evolution was held in Bern in Switzerland, 12-13 April 2005. Together with the scientific research network RELEASE, financed by the European Science Foundation (ESF), a two-day workshop was organized focusing on the most important challenges and emerging trends in software evolution research and practice.

The workshop was co-organized by Stéphane Ducasse, Radu Marinescu, Tom Mens and Oscar Nierstrasz. The workshop brought together people from academia and industry, identifying substantial obstacles to software evolution research and practice, and proposing ways to overcome these obstacles. In total, we received seventeen contributions, and there were 37 workshop participants coming from ten different countries.

The proposed challenges were diverse in nature, ranging from the fundamental (eg what are the laws governing software evolution?), to the pragmatic (eg how can we provide software evolution techniques and tools that scale up to industrial-size software applications?). The expected time horizon of the challenges also varied, ranging from short term (typically a couple of months) to long term (several years or even decades). A classification of all proposed challenges, together with a brief summary of each, has been compiled as a result of the workshop, and is available via the workshop Web site.

In addition to its scientific purpose, the workshop also hosted the annual steering committee meeting of the RELEASE network, financed by the European Science Foundation until autumn 2005, and the annual steering committee meeting of the ERCIM Working Group on Software Evolution. During the latter meeting, we discussed the current status of the network (including over thirty members originating from research institutes all over Europe, seventeen of which belong to ten different ERCIM partner institutes). We also reported on the successful meeting organised in Rome last year (October 2004), and planned a number of new activities for 2005 and 2006. Last, but certainly not least, we discussed concrete opportunities and plans for proposing new initiatives within the IST domain of the 5th call of the EU 6th Framework Programme (particularly with the strategic objectives 'embedded systems' and 'software and services'). In addition, we plan to set up a Marie Curie Research Training Network.

### Links:

WG Web site: <http://w3.umh.ac.be/evol/>  
Workshop website: <http://w3.umh.ac.be/evol/meetings/evol2005.html>  
RELEASE network website: <http://www.esf.org/release>

### Please contact:

Tom Mens, Institut d'Informatique, Université de Mons-Hainaut  
Tel: +32 65 37 3453; E-mail: [tom.mens@umh.ac.be](mailto:tom.mens@umh.ac.be)

## WWV 2005 — First Workshop on Automated Specification and Verification of Web Sites

by María Alpuente, Santiago Escobar and Moreno Falaschi

The First International Workshop on Automated Specification and Verification of Web Sites (WWV 05) was held in Valencia, Spain, during March 14-15, 2005.

This was the first workshop in a series aimed at promoting a common forum for researchers from the communities of rule-based programming, automated software engineering, and web-oriented research.

The workshop was attended by 42 participants from universities, research institutes, and companies from eleven countries. The two invited speakers gave highly interesting and stimulating presentations. Anthony Finkelstein, from the University College of London, UK, spoke about checking complex, distributed data against business rules, reference data and industry standards as a way to avoid operational errors and large, daily losses in data intensive businesses due to inconsistent data. Shriram Krishnamurthi from Brown University, USA, spoke on the pros and cons of two different approaches to web site verification: the static view of a web site as program source and the dynamic view of a web site as an entity with different contextual or temporal behaviors.

The Programme Committee had selected ten regular papers, two position papers, and six system descriptions or works in progress. All of them spanning formal methodologies and techniques as diverse and complementary as (i) formal models for describing and reasoning about web sites, (ii) testing, validation and categorization of web sites, (iii) accessibility evaluation, (iv) XML transformation and optimization, (v) Rule-based approaches to web site analysis and verification, and (vi) model-checking and static analysis applied to the web. The proceedings of the workshop are published as a technical report of the Departamento de Sistemas Informáticos y Computación of the Universidad Politécnica de Valencia. A selection of the papers will appear in the Elsevier series Electronic Notes in Theoretical Computer Science (ENTCS).

The programme and workshop chairs wish to thank the organisations that have supported the event, especially the EU-India project ALA/95/23/2003/077-054, Università degli Studi di Siena, CologNET, Technical University of Valencia, and the Spanish Ministry of Education and Science.

### Link:

WWV'05: <http://www.dsic.upv.es/workshops/wwv05/>

### Please contact:

María Alpuente, DSIC-UPV, Valencia, Spain  
Tel: +34 963 879 354; E-mail: [alpuente@dsic.upv.es](mailto:alpuente@dsic.upv.es)



## CALL FOR PARTICIPATION

**3rd International Workshop on Adaptive Multimedia Retrieval****Glasgow, UK, 28-29 July 2005**

This workshop is part of the IR Festival, which is a week full of IR activities in Glasgow, and it is co-located with the 19th International Joint Conference on Artificial Intelligence (IJCAI 2005) in Edinburgh. The goals of the workshop are to intensify the exchange of ideas between different research communities such as multimedia, information retrieval and AI, to provide an overview of current activities in this area and to point out connections between the diverse involved research communities and research in AI.

The workshop focuses especially on researchers that are working on feature extraction techniques for multimedia, computer linguistic approaches, (dynamic) data analysis methods, and visualisation methods as well as user interface design. Topics for the workshop are:

- multimedia retrieval systems (for text, image, audio, video and mixed-media)
- theoretical foundations of multimedia retrieval and mining
- intelligent multimedia data modelling, indexing and structure extraction
- adaptive Hypermedia and web based systems
- metadata for multimedia retrieval
- multimedia and multi-modal mining
- semantic content analysis for multimedia
- semantic web and ontologies
- adaptive query languages
- similarity measures (especially user adaptive measures)
- user and preference modelling (including feedback models)
- methods for adaptive data visualisation and user interfaces.

Registration is free and participation is open to everyone. The workshop is sponsored by the Multimedia Knowledge Management Network

(<http://www.mmkm.org>) which consists of research teams from seven UK universities who work in this new interdisciplinary field. The aim of this network is to enhance communication between the experts in both academia and industry, and to maintain shared resources for the direct benefit of the research community. The network is hosted at and maintained by the Multimedia Information Retrieval group at Imperial College London (<http://mmir.doc.ic.ac.uk>).

**More information:**

<http://www.dcs.gla.ac.uk/amr2005/>

If you need help in finding a partner for the EU IST programs:

**IDEAL-IST**

**IDEAL-IST is there to help you ... again !**

Expecting the deadline for the 5th Call in September 2005 for the submission of proposals in the IST-programme by September 2005 with a budget of: 638 Mio €, IDEAL-IST is ready for action !

Helping IST-proposers since 1996 in finding suitable partners for their IST-project ideas, IDEAL-IST has expanded by now to a consortium of 34 countries, linking over 55,000 organizations, with incomparable experience in the IST-programme as a whole and in the issue of Partner Search in particular.

**Three main services are offered free of charge:**

**1. Finding partners for your IST project idea**

Your project idea will be distributed to over 55,000 international contacts, interested in the IST Programme & subscribed to their national IDEAL-IST Mailing List. This service boasts a 92% success rate: proposers find appropriate partners for partnership building! The IDEAL-IST network provides suitable partners within weeks, even days, with an average of 44 responses per partner search! Partner searches are call-specific and rely on a proactive mechanism. That means that you will be contacted by the national IDEAL-IST Representative who becomes your personal advisor and supports you in the implementation of the final Partner Search application before publishing it on the public Website and distributing it to the 34 IDEAL-IST Representatives and Mailing List subscribers.

**2. Joining other IST proposals****3. Joining running IST projects**

Your national IDEAL-IST Representative keeps you informed about partner searches. Just register your interests in the national Mailing List and your national IDEAL-IST Representative will automatically send you the new searches.

Register your details today and take advantage of this service, free of charge.

**<http://www.ideal-ist.net>**

## CALL FOR PAPERS

## SOFSEM 2006 – 32nd Conference on Current Trends in Theory and Practice of Computer Science

**Merin, Czech Republic,  
21-27 January 2006**

SOFSEM (SOFTware SEMinar) is an annual international conference devoted to the theory and practice of Computer Science aiming at fostering co-operation among professionals from academia and industry working in various areas of Computer Science. SOFSEM offers a unique opportunity to quickly obtain a representative overview of the areas that are selected as the topics of the year:

- Foundations
- Wireless, Mobile, Ad Hoc and Sensor Networks
- Database Technologies
- Semantic Web Technologies.

SOFSEM is especially suited for young computer scientists. The program consists of series of invited talks, contributed talks, working sessions and the student research forum.

### SOFSEM 2006 Invited Speakers

- *Foundations:*  
S. Barry Cooper, UK: 'How Can nature help us Compute?'
- *Wireless, Mobile, Ad Hoc and Sensor Networks:*  
Sotiris Nikolettseas, Greece: 'Models and Algorithms for Wireless Sensor Networks (Smart Dust)';  
Christian Schindelhauer, Germany: n.c.
- *Database Technologies:*  
Georg Gottlob, Germany: 'Monadic Queries over Tree Structured Data'
- *Semantic Web Technologies:*  
Marie-Christine Rousset, France: 'Somewhere in Semantic Web'

### Submission Deadlines

- Abstracts: 15 August 2005
- Full papers: 22 August 2005

#### More information:

<http://www.cs.cas.cz/sofsem/06/>

## CALL FOR PARTICIPATION

## GRIDs @ Work : Middlewares, Components, Users, Contest, and Plugtest

**ETSI, Sophia Antipolis  
10-14 October 2005**

The purpose of this event is to test and experiment tools and protocols for Grid computing. This implies to be able to gain access to a Grid computing infrastructure and then to be able to have an application running on it. For that reason, it is necessary to gather workstations, clusters, supercomputers, computing Grids, etc., that may already be deployed all over the world, and to make them accessible during the Plugtest event. One technical challenge is to succeed to virtually merge all the gathered computing elements in order to form a single world-scale computing Grid. This yields to a huge Grid at work, through which it becomes easy to test the interoperability of the various cluster and Grid computing technologies. The effective test is done through the deployment of a single application on all CPUs available at once.

In order to set up such a world-scale grid, ProActive, a high-level Grid middleware will be used. ProActive is a library (Source code under LGPL license, developed by the OASIS team at INRIA Sophia-Antipolis, and a key part of the ObjectWeb consortium <http://www.objectweb.org>) for parallel, distributed, and concurrent computing, also featuring mobility and security in a uniform framework. Furthermore ProActive interoperates with the major protocols and tools for cluster and Grid computing (HTTP, ssh, ssh tunnelling, RMI, Globus GT2, GT3 and GT4, LSF, PBS, SunGrid Engine, etc). ProActive is open, ie, it is technically possible to enlarge the panel of Grid access tools with whom it is interfaced, in order to launch jobs for instance. As such, ProActive allows to orchestrate different heterogeneous Grid resources at once. ProActive has been used for the first Grid Plugtest organised by ETSI and INRIA in October 2004. It gathered 800 CPUs all over the world. This time, Grid infrastructures deployed by some of the major Grid actors such as EGEE and NorduGrid will join the testbed.

#### More information:

<http://www.etsi.org/plugtests/GRID.htm>

## This could be your advertisement!

**By advertising in ERCIM News, your company or institution  
will be able to speak to a highly qualified audience:**

**You can reach over 10,000 researchers, scientists and decision  
makers in the field of information and communication technologies.**

**For rates and formats, see:**

[http://www.ercim.org/publication/Ercim\\_News/ads.html](http://www.ercim.org/publication/Ercim_News/ads.html)

## ANNOUNCEMENT OF A COMPETITIVE CALL FOR ADDITIONAL PROJECT PARTNERS

The following project currently active in the *Sixth Framework programme of the European Community for research, technological development and demonstration activities contributing to the creation of the European research area and to innovation (2002-2006)* requires the participation of new project partners (mainly SMEs operating in collaborative networks) to carry out certain tasks within the project.

Project acronym and contract number	ECOLEAD, FP6 IP 506958 (Integrated project)
Project full name	European Collaborative networked Organizations LEADership initiative
Date of opening of call	August 2005; exact date to be published in <a href="http://www.ecolead.org">www.ecolead.org</a>
Date of close of call	October 2005; exact date to be published in <a href="http://www.ecolead.org">www.ecolead.org</a>
Language in which proposal should be submitted	English
Address for further information (call webpage & Project coordinator):	<a href="http://www.ecolead.org">www.ecolead.org</a> , Martin.Ollus@vtt.fi

### 1. ECOLEAD project and the role of new partners

ECOLEAD develops foundations, mechanisms and tools for establishing, operating and managing collaborative networked organizations (CNO). Three focus areas are addressed: Breeding Environments (BE), Dynamic Virtual Organizations (VO), and Professional Virtual Communities (PVC). The holistic approach is reinforced and sustained on two horizontal activities: the Theoretical Foundation for collaborative networks and the ICT Infrastructure, that both support and affect the three vertical focus areas. More info [www.ecolead.org](http://www.ecolead.org).

The searched and selected organizations will join the ECOLEAD consortium as new partners. **ECOLEAD will offer the new partners:**

- solutions and methods supporting operation and management of CNOs (BEs, VOs and PVCs),
- possibility to gain information and to have an influence on the latest development in collaborative networks,
- possibility to compare their experiences with other networked organizations,
- training into the main concepts and solutions & support in the demonstrations,
- partial funding from the European Commission (see below).

#### ECOLEAD expects the new partners:

- to represent a collaborative networked organization (CNO), that is a network, a virtual organization or a professional virtual community,
- to operate in the role of end-users of ECOLEAD solutions and to perform the demonstrations collaboratively each with its own CNO,
- to participate actively in the ECOLEAD collaboration and perform all the partner duties.

### 2. Summary of the tasks requested from the new partners:

#### The tasks of the new partners include:

1. **Demonstration tasks:** This is the main task of the new partners. The objective is to demonstrate, test and evaluate ECOLEAD solutions which have been developed in the three ECOLEAD focus areas (mentioned above). The demonstrated solutions may be tools (e.g. supporting software), but also processes, methods or operating rules, which support the operation of Collaborative Networked Organizations (CNOs).
2. **Research & development & innovation tasks:** The R&D tasks create the basis for the demonstrations and support the after-demonstration analysis. In the first phase ECOLEAD solutions are analysed and applied to create and document a partner-CNO-specific demonstration scenario. Further the task of the new SME network partners is to participate as end-users in the generation of ideas and future visions and further specification and development of solutions, based on experience derived from the demonstration. Participation in impact creation and dissemination of the results outside ECOLEAD is also encouraged.
3. **Management tasks:** A small amount of funding is reserved to take care about the administrative tasks of the project.

### 3. Demonstration targets and task description

The demonstrations support the ECOLEAD goal by generating information about the applicability and impact of the ECOLEAD developments.

The main candidate prototype demonstrations include:

- Breeding Environment (BE) partner profiles e-catalog & skill management system, VO Creation environment
- Virtual Organization management support & VO performance measurement process and support tool, VO e-services.
- Professional Virtual Community collaboration platform, operation model & management, PVC metrics, PVC expertise profiling

Demonstrations are the main task of the new partners. They will be organised in two different stages;

- First (trial) stages, whose main aim is to test and validate the ECOLEAD approach from an operative point of view, with the objective of provide feedback and improvement suggestions to the developed ECOLEAD products. As collaborative solutions are tested, the participation of collaborating partners in the demonstrations is envisaged. The new partner (proposer) coordinates its own demonstration in its Collaborative networked organization. Trial metrics will be prepared in advance by ECOLEAD.
- Second (take-up) stage, in which the impact of the adopted ECOLEAD methodologies and technologies will be assessed in terms of impact of the implementation of the methodology/ tool in the company practices and resulting business benefits. Take-up metrics will be prepared in advance.

Participation in collaborative workshops of the ECOLEAD consortium is expected. More detailed descriptions of the demonstrations can be found in [www.ecolead.org](http://www.ecolead.org).

### 4. Partner selection

The call is focused to end-users of collaborative network solutions, that is, mainly organizations managing or operating in SME enterprise networks or Professional Virtual Communities. The ECOLEAD project makes the potential contract only with one legal entity. The proposer is advised to name the main intended co-operating companies. The expected number of new partners is 6-8, from which some would participate only in the 1st stage (trial), and the rest to the both stages. Information about the selection criteria is available at [www.ecolead.org](http://www.ecolead.org).

### 5. Schedule and resources

Activity type	Estimated total (all the new partners together) costs	Commission funding percentage
Research	€ 520000	up to 50%
Demonstration	€1 070 000	up to 35%
Consortium management	€ 35000	up to 100%

Total Commission funding available for the new partners is € 669500. Expected duration of participation in project: between Project Month 21 and Month 48 (1.1.06-31.3.08). The project budget is revised yearly.

**Updated information about the call:** [www.ecolead.org](http://www.ecolead.org).



## News about legal information relating to Information Technology from European directives, and pan-European legal requirements and regulations.

### Digital Rights Management

Digital Rights Management (DRM) presents a real challenge for content communities in this digital age. Security and encryption issues as a means of avoiding unauthorized copying were the main issues for DRM in the past, basically locking away content unless it was paid for. Today DRM deals with much broader issues such as the description, identification, trading, protection, monitoring and tracking of all forms of rights usages over both tangible and intangible assets, including management of rights holders' relationships.

Typically a DRM framework is modelled around three areas of intellectual property:

- *Asset Creation and Capture:* To manage the creation of content so it can be easily traded. This includes asserting rights when content is first created (or reused and extended with appropriate rights to do so) by various content creators/providers.
- *Asset Management:* To manage the collection and trade of content. This includes accepting content from creators into an asset management system. Systems need to manage the descriptive metadata and rights metadata (eg, parties, uses, payments, etc.).
- *Asset Use:* To manage the use of content once it has been traded. This includes supporting constraints over traded content in specific desktop systems/software.

A DRM functional architecture addresses the following issues in each of these three main areas:

For asset creation and capture:

- *Rights Validation* – to ensure the content being created from existing content includes the rights to do so
- *Rights Creation* – to allow rights to be assigned to the new content, such as specifying the rights owners and permissible use
- *Rights Workflow* – to allow for content to be processed through a series of

workflow steps for review and/or approval of rights (and content).

For asset management:

- *Repository functions* – to enable the access/retrieval of content in potentially distributed databases and the access/retrieval of metadata. The metadata includes Parties, Rights and descriptions of the Works
- *Trading functions* – to enable the assignment of licenses to parties who have traded agreements for rights over content, including payments from licensees to rights holders (eg royalty payments). In some cases the content may be encrypted/protected or packaged for a particular type of desktop environment.

For asset use:

- *Permissions Management* – to enable the use environment to adhere to the rights associated with the content, such as taking account of any restrictions imposed under the conditions of the licence
- *Tracking Management* – to monitor use of which may impose restrictions under the license conditions, for example where the number of accesses is restricted or perhaps to track payment.

For more information on Digital Rights Management Architectures see <http://www.dlib.org/dlib/june01/iannella/06iannella.html>

DRM is heavily criticised by some supporters of the Open Access movement who believe it serves to 'lock up' content, rather being seen as a tool to ensure correct author attribution, to certify integrity and provenance, prevent plagiarism and to encourage authors to assert their creative rights rather than restrict them. Perhaps the most important area though is the use of rights metadata. The JISC-funded project RoMEO <http://www.lboro.ac.uk/departments/lis/disresearch/romeo/> found that 55% of researchers wanted to limit usage of their

works to certain purposes, eg educational or non-commercial use. However, an 'all rights reserved' model was more than most researchers wanted. The project developed an XML-based system designed to express rights and permissions in an OA environment. Around this time a number of intellectual property lawyers got together and founded Creative Commons <http://creativecommons.org/> a not for profit corporation founded on the principle that some creators may not want to exercise all the intellectual property rights the law affords them. Creative Commons offers a number of different licence solutions to creators such as Attribution - free use to others on the basis that credit is given to the creator; Non-commercial – restricted to non-commercial use only; No derivative works – work must be used as it is and not modified in any way; Share alike – allows others to distribute derivative works but only under an identical licence.

The creation of more and more compound digital objects, inheriting all of the existing rights of the individual digital components, along with the new rights of the newly-created object, makes Digital Rights Management in an e-Research environment more challenging than ever.

### Plagiarism

Anti-plagiarism software, used by universities to check student essays, is being adapted to review academic papers prior to publication. This is one of the latest initiatives to be embraced by the big publishers, like Elsevier (the world's largest scientific publisher which publishes around a quarter of a million papers each year) and Blackwell. The software is being tested out on the arXiv physics preprint server at Cornell University. It runs an algorithm that looks for any two documents that share six of the same words in a row. Other anti-plagiarism tools are being developed currently but little is known about them at present, but some of them could be free to use and targeted at editors and peer reviewers. Student anti-plagiarism software services detect not only duplicate publications but can check documents against large databases.

by Heather Weaver, CCLRC, UK

ERCIM News is the magazine of ERCIM. Published quarterly, the newsletter reports on joint actions of the ERCIM partners, and aims to reflect the contribution made by ERCIM to the European Community in Information Technology. Through short articles and news items, it provides a forum for the exchange of information between the institutes and also with the wider scientific community. This issue has a circulation of 10,500 copies.

## Advertising

For current advertising rates and conditions, see [http://www.ercim.org/publication/ERCIM\\_News/](http://www.ercim.org/publication/ERCIM_News/) or contact [office@ercim.org](mailto:office@ercim.org)

## Copyright Notice

All authors, as identified in each article, retain copyright of their work.

## ERCIM News online edition

[http://www.ercim.org/publication/ERCIM\\_News/](http://www.ercim.org/publication/ERCIM_News/)  
ERCIM News is published by ERCIM EEIG, BP 93, F-06902 Sophia-Antipolis Cedex  
Tel: +33 4 9238 5010, E-mail: [office@ercim.org](mailto:office@ercim.org)  
ISSN 0926-4981

**Director:** Jérôme Chailloux, ERCIM Manager

## Central Editor:

Peter Kunz, ERCIM office  
[peter.kunz@ercim.org](mailto:peter.kunz@ercim.org)

## Local Editors:

AARIT: n.a.  
CCLRC: Martin Prime  
[M.J.Prime@rl.ac.uk](mailto:M.J.Prime@rl.ac.uk)  
CRCIM: Michal Haindl  
[haindl@utia.cas.cz](mailto:haindl@utia.cas.cz)  
CWI: Annette Kik  
[Annette.Kik@cw.nl](mailto:Annette.Kik@cw.nl)  
CNR: Carol Peters  
[carol.peters@isti.cnr.it](mailto:carol.peters@isti.cnr.it)  
FORTH: Margherita Antona  
[antona@ics.forth.gr](mailto:antona@ics.forth.gr)  
Fraunhofer ICT Group:  
Michael Krapp  
[michael.krapp@scai.fraunhofer.de](mailto:michael.krapp@scai.fraunhofer.de)  
FNR: Patrik Hitzelberger  
[hitzelbe@crpg.lu](mailto:hitzelbe@crpg.lu)  
FWO/FNRS: Benoît Michel  
[michel@tele.ucl.ac.be](mailto:michel@tele.ucl.ac.be)  
INRIA: Bernard Hidoine  
[bernard.hidoine@inria.fr](mailto:bernard.hidoine@inria.fr)  
Irish Universities Consortium:  
Ray Walshe  
[ray@computing.dcu.ie](mailto:ray@computing.dcu.ie)  
NTNU: Truls Gjestland  
[truls.gjestland@ime.ntnu.no](mailto:truls.gjestland@ime.ntnu.no)  
SARIT: Harry Rudin  
[hrudin@smile.ch](mailto:hrudin@smile.ch)  
SICS: Kersti Hedman  
[kersti@sics.se](mailto:kersti@sics.se)  
SpaRCIM: Salvador Lucas  
[slucas@dsic.upv.es](mailto:slucas@dsic.upv.es)  
SRCIM: Gabriela Andrejkova  
[andrejk@kosice.upjs.sk](mailto:andrejk@kosice.upjs.sk)  
SZTAKI: Erzsébet Csuhaj-Varjú  
[csuhaj@sztaki.hu](mailto:csuhaj@sztaki.hu)  
VTT: Pia-Maria Linden-Linna  
[pia-maria.linden-linna@vtt.fi](mailto:pia-maria.linden-linna@vtt.fi)  
W3C: Marie-Claire Forgue  
[mcf@w3.org](mailto:mcf@w3.org)

## Subscription

Subscribe to ERCIM News free of charge by:

- sending e-mail to your local editor
- contacting the ERCIM office (see address above)
- filling out the form at the ERCIM website at <http://www.ercim.org/>



**INRIA** — It is with a profound sadness that INRIA officially announces the demise of **Isabelle Attali** and her two children Tom and Ugo, following the December 26, 2004 tsunami.

Isabelle Attali was the scientific head of the Oasis team since January 2000. She was very much involved in numerous national, European and international initiatives. She was Vice-President of the Institute's Evaluation Committee since 2003, head of a work group of the GridCoord European project launched in July 2004, and an active member of the ERCIM administrated CoreGRID network of excellence. Up to 2001, Isabelle Attali took care of the Young Researcher in Programming School she had contributed to creating in 1995. She also devoted much of her time and energy to the Telecom Valley association at Sophia Antipolis.

In December 2004, Isabelle and her companion, Denis Caromel, also a member of project Oasis, were on assignment in Sri Lanka with their children. They were teaching at a winter school organized by Cimpa and Unesco and were setting up a collaboration with the University of Ruhuna. As a tribute to her major commitment to all these endeavors, the Sophia Antipolis Foundation and INRIA have decided to create a fund to facilitate exchanges between Sri Lanka and Sophia Antipolis in the field of information and communication science and technology. This fund may be provisioned by individual and corporate donations. For further information about the fund, contact [info@sophia-antipolis.org](mailto:info@sophia-antipolis.org)

**VTT** — After many years in the ERCIM Executive Committee **Seppo Linnainmaa**, head of VTT's Information Systems department now represents VTT on ERCIM's Board of Directors succeeding Pekka Silvennoinen, Executive Director of VTT Information Technology. Seppo Linnainmaa was replaced in the Executive Committee by **Seppo Valli**, head of the Multimedia research group of the 'Networks' department.



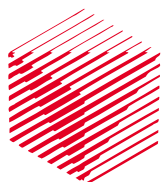
Photo: A. Wipenaar

**CWI** — **Alexander Schrijver** is awarded the Spinoza Prize 2005. This was announced by the Netherlands Organization for Scientific Research in The

Hague on 6 June. Lex Schrijver is researcher at CWI in Amsterdam, leader of the Probability, Networks and Algorithms research cluster (until 1<sup>st</sup> of July) and member of CWI's management team. He is also part-time professor at the University of Amsterdam. Schrijver received this prestigious prize, called the 'Dutch Nobel Prize', for his outstanding, pioneering and inspiring research in the field of combinatorics and algorithms. The Spinoza Prize is the most distinguished award in science in the Netherlands and consists of 1.5 million Euro – to be spent on research of choice – and a statue of Spinoza. The official ceremony will take place on Wednesday 23 November 2005.

**CWI** — **Scientists of CWI and the Universities of Twente and Konstanz** launched **MonetDB/XQuery** during the Holland Open Software Conference in Amsterdam, on May 31. Storing and querying high volumes of XML data require new software systems. MonetDB/XQuery is an open source system that provides a complete implementation of XQuery. This makes it possible to search both the content and structure of large XML documents. MonetDB/XQuery is built on the MonetDB Relational Database Management System. More information can be found on:

<http://pathfinder-xquery.org/>  
<http://www.cwi.nl/ins1>  
<http://www.hollandopen.nl/index.jsp?nr=2610>



ERCIM – The European Research Consortium for Informatics and Mathematics is an organisation dedicated to the advancement of European research and development, in information technology and applied mathematics. Its national member institutions aim to foster collaborative work within the European research community and to increase co-operation with European industry.



ERCIM is the European Host of the World Wide Web Consortium.



Austrian Association for Research in IT  
c/o Österreichische Computer Gesellschaft  
Wollzeile 1-3, A-1010 Wien, Austria  
Tel: +43 1 512 02 35 0, Fax: +43 1 512 02 35 9  
<http://www.aarit.at/>



INRIA

Institut National de Recherche en Informatique  
et en Automatique  
B.P. 105, F-78153 Le Chesnay, France  
Tel: +33 1 3963 5511, Fax: +33 1 3963 5330  
<http://www.inria.fr/>



Council for the Central Laboratory of the Research  
Councils, Rutherford Appleton Laboratory  
Chilton, Didcot, Oxfordshire OX11 0QX, United Kingdom  
Tel: +44 1235 82 1900, Fax: +44 1235 44 5385  
<http://www.cclrc.ac.uk/>



Norwegian University of Science and Technology  
Faculty of Information Technology, Mathematics and  
Electrical Engineering, N 7491 Trondheim, Norway  
Tel: +47 73 59 80 35, Fax: +47 73 59 36 28  
<http://www.ntnu.no/>



Consiglio Nazionale delle Ricerche, ISTI-CNR  
Area della Ricerca CNR di Pisa,  
Via G. Moruzzi 1, 56124 Pisa, Italy  
Tel: +39 050 315 2878, Fax: +39 050 315 2810  
<http://www.isti.cnr.it/>



Spanish Research Consortium for Informatics  
and Mathematics c/o Esperanza Marcos, Rey Juan Carlos  
University, C/ Tulipan s/n, 28933-Móstoles, Madrid, Spain,  
Tel: +34 91 664 74 91, Fax: 34 91 664 74 90  
<http://www.sparcim.org>



Czech Research Consortium  
for Informatics and Mathematics  
FI MU, Botanická 68a, CZ-602 00 Brno, Czech Republic  
Tel: +420 2 688 4669, Fax: +420 2 688 4903  
<http://www.utia.cas.cz/CRCIM/home.html>



Swedish Institute of Computer Science  
Box 1263,  
SE-164 29 Kista, Sweden  
Tel: +46 8 633 1500, Fax: +46 8 751 72 30  
<http://www.sics.se/>



Centrum voor Wiskunde en Informatica  
Kruislaan 413, NL-1098 SJ Amsterdam,  
The Netherlands  
Tel: +31 20 592 9333, Fax: +31 20 592 4199  
<http://www.cwi.nl/>



Swiss Association for Research in Information Technology  
c/o Prof. Dr Alfred Strohmeier, EPFL-IC-LGL,  
CH-1015 Lausanne, Switzerland  
Tel: +41 21 693 4231, Fax: +41 21 693 5079  
<http://www.sarit.ch/>



Fonds National de la Recherche  
6, rue Antoine de Saint-Exupéry, B.P. 1777  
L-1017 Luxembourg-Kirchberg  
Tel: +352 26 19 25-1, Fax +352 26 1925 35  
<http://www.fnr.lu>



Slovak Research Consortium for Informatics and  
Mathematics, Comenius University, Dept. of Computer  
Science, Mlynska Dolina M, SK-84248 Bratislava, Slovakia  
Tel: +421 2 654 266 35, Fax: 421 2 654 270 41  
<http://www.srcim.sk>



FWO  
Egmontstraat 5  
B-1000 Brussels, Belgium  
Tel: +32 2 512.9110  
<http://www.fwo.be/>

FNRS  
rue d'Egmont 5  
B-1000 Brussels, Belgium  
Tel: +32 2 504 92 11  
<http://www.fnrs.be/>



Magyar Tudományos Akadémia  
Számítástechnikai és Automatizálási Kutató Intézet  
P.O. Box 63, H-1518 Budapest, Hungary  
Tel: +36 1 279 6000, Fax: + 36 1 466 7503  
<http://www.sztaki.hu/>



Foundation for Research and Technology – Hellas  
Institute of Computer Science  
P.O. Box 1385, GR-71110 Heraklion, Crete, Greece  
Tel: +30 2810 39 16 00, Fax: +30 2810 39 16 01  
<http://www.ics.forth.gr/>



Irish Universities Consortium  
c/o School of Computing, Dublin City University  
Glasnevin, Dublin 9, Ireland  
Tel: +3531 7005636, Fax: +3531 7005442  
<http://ercim.computing.dcu.ie/>



Fraunhofer ICT Group  
Friedrichstr. 60  
10117 Berlin, Germany  
Tel: +49 30 726 15 66 0, Fax: +49 30 726 15 66 19  
<http://www.iuk.fraunhofer.de>



Technical Research Centre of Finland  
P.O. Box 1200  
FIN-02044 VTT, Finland  
Tel: +358 9 456 6041, Fax: +358 9 456 6027  
<http://www.vtt.fi/tte>

## Order Form

I wish to subscribe to the

If you wish to subscribe to ERCIM News  
**free of charge**

☐ **printed edition**

☐ **online edition (email required)**

or if you know of a colleague who would like to  
receive regular copies of  
ERCIM News, please fill in this form and we  
will add you/them to the mailing list.

Name: .....

Organisation/Company: .....

Address: .....

Post Code: .....

City: .....

Country: .....

E-mail: .....

Send, fax or email this form to:

**ERCIM NEWS**

**2004 route des Lucioles**

**BP 93**

**F-06902 Sophia Antipolis Cedex**

**Fax: +33 4 9238 5011**

**E-mail: [office@ercim.org](mailto:office@ercim.org)**

Data from this form will be held on a computer database.  
By giving your email address, you allow ERCIM to send you email

You can also subscribe to ERCIM News and order back copies by filling out the form at the ERCIM website at  
[http://www.ercim.org/publication/Ercim\\_News/](http://www.ercim.org/publication/Ercim_News/)