ERCIM NEWS

European Research Consortium for Informatics and Mathematics
www.ercim.org

Number 55, October 2003

Special: Machine Perception

Contents

Cover: Kurt3D – An Autonomous Mobile Robot (see article on page)

Next issue:
January 2004

Special theme:
Analysis, Diagnosis, Planning and Simulation of Industrial Systems

SpaRCIM becomes 17th ERCIM Member Institute

SpaRCIM, the Spanish Research Consortium for Informatics and Mathematics, joined ERCIM on 1st July 2003.

The SpaRCIM consortium was founded under the auspices of the Spanish Ministry of Science and Technology (MCYT). Its main goal is to spread the activities of the ERCIM consortium to the informatics and mathematics research communities within Spain. The consortium is composed of five universities and a research institute dependent on the Spanish Research Council (CSIC):

- Artificial Intelligence Research Institute (IIIA)
- University of Malaga (UMA)
- Technical University of Catalonia (UPC)
- Technical University of Madrid (UPM)
- Technical University of Valencia (UPV)
- Rey Juan Carlos University (URJC).

These institutions include a number of groups that are considered to be representative of Spanish research in computer science. These groups are working on a great variety of research areas within information technology and applied mathematics.

IIIA – Instituto de Investigación en Inteligencia Artificial (Artificial Intelligence Research Institute)
The IIIA is a centre devoted to research in Artificial Intelligence (AI), and belongs to the Spanish Scientific Research Council (CSIC). The IIIA was created in 1985, and currently has around thirty members. Topics of research include learning systems, intelligent agents, logic reasoning and search, electronic marketplaces, autonomous robots and music with AI. Intensive collaborations, mostly within the framework of European community programs, have taken place with industries and academic institutions in many countries.

UMA – Universidad de Málaga (University of Malaga)
The UMA was created in 1972, and thirty years later, it plays an important role in the cultural life and the scientific, economical and technological development of the local region. The participation of the UMA in SpaRCIM is currently managed by researchers from the Departamento de Lenguajes y Ciencias de la Computación (LCC). Research focuses on various aspects of software engineering in the field of distributed systems, and in particular, specification and design methodologies and open and heterogeneous systems. The objectives are to study (and deal with) the basic problems of these techniques and systems, and to develop applications which can be supported by them, including telecommunication services, telematic applications, electronic commerce etc.

UPC – Universitat Politècnica de Catalunya (Technical University of Catalonia)
The UPC was created in 1971 and is one of eleven universities in Catalonia. Almost all technical fields are taught at the faculties and technical schools of the UPC and in addition, the UPC provides academic supervision to a number of associated schools and hosts several research institutes. The participation of the UPC in SpaRCIM is currently managed by researchers from the Departament de Llenguatges i Sistemes Informàtics (LSI), which is composed of more than 180 people. Topics of research include algebraic foundations of system specification and design, logic in computer science (automated deduction, term rewriting etc), natural language processing, machine learning, autonomous agents etc.

UPM – Universidad Politécnica de Madrid (Technical University of Madrid)
While the UPM was created in 1971, most of its integrated centres have a long history dating back to the eighteenth and nineteenth centuries. The schools of engineering and architecture are the oldest in

ERCIM President Gerard van Oortmerssen (left) and SpaRCIM's representative on ERCIM's Board of Directors, Juan José Moreno from the Technical University of Madrid and the Spanish Ministry of Science and Technology.

SpaRCIM is composed of one national research institute and five universities: Artificial Intelligence Research Institute (IIIA); Technical University of Madrid (UPM); Technical University of Catalonia (UPC); Rey Juan Carlos University (URJC); Technical University of Valencia (UPV); University of Malaga (UMA).

Spain, and became part of the UPM after its foundation. The participation of the UPM in SpaRCIM is managed by researchers from the Departamento de Lenguajes y Sistemas Informáticos e Ingeniería del Software (LSIIS). Topics of research include design, semantics, implementation, applications of declarative programming and specification languages, distributed execution, the Internet and World Wide Web, enviromental software and modeling, global program analysis, programming language design, program transformation/optimisation and parallelism.

UPV – Universidad Politécnica de Valencia (Technical University of Valencia)
The UPV, founded in 1971, is a dynamic, innovative public institution dedicated to research and teaching, which maintains strong bonds with the local community and simultaneously has an important presence abroad. The participation of the UPV in SpaRCIM is managed by researchers from the Departamento de Sistemas Informáticos y Computación (DSIC), which comprises more than 150 researchers, staff and graduate students. Topics of research include computational logic, databases, formal methods in software engineering, multi-paradigm programming, natural language processing, object-oriented conceptual modeling and databases, object-oriented methods for software development, WWW engineering etc.

URJC – Universidad Rey Juan Carlos (Rey Juan Carlos University)
The URJC in Madrid was created in July 1996 at the behest of the government of the Comunidad de Madrid. Its medium-term goal is to achieve a balance between supply and demand for university resources in the region, thus complementing existing universities and enriching the students' individual and social opportunities. This public university is located in the southern area of the Comunidad de Madrid, and its faculties include Experimental Sciences and Technology, Health Sciences, Communication Sciences, and Law and Social Sciences. Topics of research relevant to SpaRCIM include image processing, neural networks, object-oriented databases, security etc.

These institutions already collaborate in a number of European networks and projects such as NAME, CoLogNet, AgentLink and PROENBIS. They also collaborate in a number of national networks and projects funded by the Spanish Ministry of Science and Technology (MCYT), and have founded a number of spin-off companies. The consortium is open to other Spanish institutions.

The members of SpaRCIM are interested in sharing experiences and developing coordinated projects in areas related to software technology. They expect to actively participate in the current working groups, and to promote new research in other fields of competence. There is also interest in reactivating, where possible, former working groups such as 'Programming Languages Technologies' and 'Database Research'.

Regarding SpaRCIM as a member of ERCIM, Prof. Juan José Moreno from the Technical University of Madrid (UPM) and the Spanish Ministry of Science and Technology (MCYT) is heading the ERCIM activities at SpaRCIM. Prof. Ernesto Pimentel from the University of Malaga (UMA) is the corresponding member of ERCIM's Executive Committee.

Links:
SpaRCIM website:
http://kybele.escet.urjc.es/SpaRCIM

Please contact:
Juan José Moreno
Universidad Politécnica de Madrid
Tel: +34 91 336 7458
E-mail: jjmoreno@fi.upm.es
http://lml.ls.fi.upm.es/~jjmoreno/

IM2IM – A New ERCIM Working Group on Informatics and Mathematics applied to Interventional Medicine

by Marc Thiriet

On an initiative by scientists from INRIA, a new ERCIM Working Group has been established on Informatics and Mathematics applied to Interventional Medicine.

Nowadays, many medical operations are based upon minimal-invasive, safe and cheap procedures. High precision medicine with as little invasion as possible, are required to improve patient comfort and reduce hospitalisation costs. Interventional medicine consists in placing a small effective and reliable medical device inside any anatomical duct. As endoluminal operations tend to involve precise, complex and risky techniques, doctors yearn more and more for simulation tools with which to train themselves, plan operations and ensure detailed follow-up on the state of their patients.

The new operational techniques demand indeed the acquisition of new gestures. This is due in particular to the screen interface, which does not provide a direct, three-dimensional view. In a way somewhat similar to what flight simulators are to aircraft pilots, simulation tools are very much in demand by medical doctors; either to learn the gestures, plan them, train, operate or follow-up the post-operation evolution. However, one of the main difficulties in achieving a sufficiently realistic reproduction not only of

the visual aspects, but also of the tactile aspects of the situation, is to correctly model the mechanical behaviour of the various involved organs and anatomical tissues. Overcoming these difficulties is an important challenge.

Applied mathematics and computer science are necessary tools in these advances in medical techniques. Image processing, computer graphics and virtual reality, modelling and simulation of the behaviour of biological tissues and robotics are all involved. The development of the computer-aided medical tool indeed requires several tasks:

- three-dimensional reconstruction of the vascular region of interest from medical imaging and computational mesh generation
- input data collection and control parameter selection
- numerical model coupling for interaction between the vessel wall, the owing blood and the medical device
- multiscale modeling to ensure suitable wave propagation and appropriate boundary conditions
- parallel computing to speed up the computational time for medical purpose
- numerical result processing and physical entity field visualisation
- error estimation and physics-based dynamical mesh adaptation
- virtual endoscopy with a navigation tool for endoluminal catheter displacement and device installation control
- augmented reality training simulator.

Current participants come from the ERCIM members INRIA, Fraunhofer-Gesellschaft and SARIT (EPFL), as well as from other institutions in Belgium, France, Greece, Italy and the UK. A kick-off workshop is planned in December this year.

Link:
http://www-rocq1.inria.fr/Marc.Thiriet/Im2im/

Please contact:
Marc Thiriet, INRIA and Université Pierre et Marie Curie
Tel: +33 1 39 63 59 16
E-mail: Marc.Thiriet@inria.fr

Euro-Legal

News about legal information relating to Information Technology from European directives, and pan-European legal requirements and regulations.

UK Government's SMART CARD Initiative

Targets set in the UK by the Modernising Government agenda require the use of Information Technology tools to deliver the changes towards better, more efficient, convenient, quality government services. In July 2003 the Office of the e-Envoy issued a draft Policy Framework proposing the use of SMART CARDS as such a technology enabler. The Policy Framework is aimed mainly at strategists in the Public Sector responsible for service delivery, but also aimed at raising the awareness of the informed citizen about the role SMART CARDS can play in the delivery of e-government services.

In 2002 one billion SMART CARDS were issued worldwide. This is expected to rise to 1.7 billion by the end of 2004. SMART CARDS already impact on all our lives in such forms as prepaid phone cards, debit/credit cards, shop loyalty cards, SIM cards etc. A widespread adoption of SMART CARDS would increase significantly the delivery of e-government initiatives over the next few years.

With the ability to hold a digital credential in a portable form, SMART CARDS provide one solution to the problem of authentication and security on the Internet. Similarly, they may be used to overcome the problem of authentication between government and citizen. The portability and security of SMART CARDS, plus their ability to manage and manipulate data, make them able to fulfil a number of different functions. These functions can include acting as access control to buildings or computers, credit cards, a more durable substitute for paper records, such as medical records, and, more importantly, can act as a device for executing a digital signature. Furthermore, a single card can host many services from one or more providers.

One of the strongest drivers for the uptake of SMART CARDS is the demand for secure on-line commerce and access to government services. Around 20% of the 500 or so government services available to the public require strong authentication. The use of SMART CARDS enable users to authenticate themselves, firstly through a "two-factor authentication", ie the possession of the card and the knowledge of the PIN number, or even a "three-factor authentication", where the card stores additional biometric information that can be used to test the physical identity of the cardholder, and, secondly, to provide a digital signature. The UK Government is developing policies aimed at encouraging the adoption of digital credentials and the market for trust service providers.

However, there are a number of obstacles that need to be overcome. Primarily, the lack of an industry-wide set of standards, which are well developed in terms of agreeing the physical size and shape of the cards, but at the security and application level are lacking, make it difficult to build interoperable and integrated systems and access channels. Also, the absence of a proven business case for a multi-purpose SMART CARD scheme, that meets social inclusion objectives such as access for people with disabilities and those without readily available installed infrastructure, at an affordable cost. Perhaps the most contentious of all is the need to safeguard citizens' rights in respect of the data held about them, ensuring that existing regulatory provisions in the areas of data protection, financial services, telecommunications etc are not compromised and that citizens are confident that additional personal data required to demonstrate they are who they say they are online, is held securely.

The full text of the draft Policy Framework can be found at:
http://fastlink.headstar.com/smart2.

by Heather Weaver, CCLRC
Heather Weaver regrets that she is unable to reply personally to emails or telephone calls seeking legal advice.

W3C's Speech Interface Framework

by Dave Raggett

W3C is working to expand access to the Web to allow people to interact via key pads, spoken commands, listening to prerecorded speech, synthetic speech and music. This will allow any of the world's 2 billion telephones to be used to access appropriately designed Web-based services, and will be a benefit to people with visual impairments or needing Web access while keeping their hands and eyes free for other things. It will also allow effective interaction with display-based Web content in the cases where the mouse and keyboard may be missing or inconvenient.

To fulfil this goal, the W3C Voice Browser Working Group is defining a suite of markup languages covering dialog, speech synthesis, speech recognition, call control and other aspects of interactive voice response applications. Specifications such as the Speech Synthesis Markup Language (http://www.w3.org/TR/speech-synthesis/), Speech Recognition Grammar Specification (see http://www.w3.org/TR/speech-grammar/), and Call Control XML (see http://www.w3.org/TR/ccxml/) are core technologies for describing speech synthesis, recognition grammars, and call control constructs respectively. VoiceXML is a dialog markup language that leverages the other specifications for creating dialogs that feature synthesized speech, digitized audio, recognition of spoken and DTMF key (touch tone) input, recording of spoken input, telephony, and mixed initiative conversations.

These specifications bring the advantages of web-based development and content delivery to interactive voice response applications. Further work is anticipated on enabling their use with other W3C markup languages such as XHTML, XForms (a specification of Web forms that can be used with a wide variety of platforms, see http://www.w3.org/MarkUp/Forms/) and the Synchronized Multimedia Integration Language SMIL™ (see http://www.w3.org/AudioVideo/). This will be done in conjunction with other W3C Working Groups, including the Multimodal Interaction Activity.

Some possible applications include:

- accessing business information, including the corporate 'front desk' asking callers who or what they want, automated telephone ordering services, support desks, order tracking, airline arrival and departure information, cinema and theater booking services, and home banking services
- accessing public information, including community information such as weather, traffic conditions, school closures, directions and events; local, national and international news; national and international stock market information; and business and e-commerce transactions
- accessing personal information, including calendars, address and telephone lists, to-do lists, shopping lists, and calorie counters
- assisting the user to communicate with other people via sending and receiving voice-mail and email messages.

VoiceXML 2.0 is designed based upon extensive industry experience for creating audio dialogs. For an introduction, a tutorial is available at http://www.w3.org/Voice/Guide/. Further tutorials and other resources can be found on the VoiceXML Forum Web site (http://www.voicexml.org/). W3C and VoiceXML Forum have signed a memorandum of understanding setting out mutual goals.

Based upon a small set of widely implemented extensions to VoiceXML 2.0, we anticipate an interim version of the dialog markup language called VoiceXML 2.1. These features will help developers build even more powerful, maintainable and portable voice-activated services, with complete backwards compatibility with the VoiceXML 2.0 specification. We expect to publish VoiceXML 2.1 as a small specification that describes the extensions to 2.0. The first Working Draft is expected to be published in September 2003. Future work on dialog markup, component of W3C's Speech Interface Framework, is described below.

The Speech Recognition Grammar specification (SRGS) covers both speech and DTMF (touch tone) input. DTMF is valuable in noisy conditions or when the social context makes it awkward to speak. Grammars can be specified in either an XML or an equivalent augmented BNF (ABNF) syntax, which some authors may find easier to deal with. Speech recognition is an inherently uncertain process. Some speech engines may be able to ignore "um's" and "aah's", and to perform partial matches. Recognizers may report confidence values. If the utterance has several possible parses, the recognizer may be able to report the most likely alternatives (n-best results).

The Speech Synthesis specification (SSML) defines a markup language for prompting users via a combination of prerecorded speech, synthetic speech and music. You can select voice characteristics (name, gender and age) and the speed, volume, pitch, and emphasis. There is also provision for overriding the synthesis engine's default pronunciation.

The Voice Browser Working Group is collaborating with the Cascading Style Sheets (CSS) Working Group to develop a CSS3 module for speech synthesis based upon SSML for use in rendering XML documents to speech. This is intended to replace the aural cascading style sheet properties in CSS2. The first Working Draft was published in May 2003.

Semantic Interpretation for Speech Recognition specification (see http://www.w3.org/TR/semantic-interpretation/) describes annotations to grammar rules for extracting the semantic results from recognition. The annotations are expressed in a syntax based upon a subset of ECMAScript, and when evaluated, yield a result represented either as XML or as a value that can be held in an ECMAScript variable. The target for the XML output is the Extensible Multimodal Annotation Markup Language (EMMA) which is being developed in the Multimodal Interaction Activity (see http://www.w3.org/TR/emma/).

W3C is working on a markup language called CCXML to enable fine-grained control of speech (signal processing)

resources and telephony resources in a VoiceXML telephony platform (http://www.w3.org/TR/ccxml/). The scope of these language features is for controlling resources in a platform on the network edge, not for building network-based call processing applications in a telephone switching system, or for controlling an entire telecom network. These components are designed to integrate naturally with existing language elements for defining applications which run in a voice browser framework. This will enable application developers to use markup to perform call screening, whisper call waiting, call transfer, and more. Users can be offered the ability to place outbound calls, conditionally answer calls, and to initiate or receive outbound communications such as another call.

W3C's Speech Interface Framework work is ongoing and the W3C Voice Browser Working Group is seeking participants to develop and/or give feedback on public drafts and give suggestions for requirements and directions as well. A public mailing list is also available for public discussion at http://lists.w3.org/Archives/Public/www-multimodal/.

Links:
Home page for W3C's Voice Browser Activity:
http://www.w3.org/Voice/

Published Technical Reports on Voice:
http://www.w3.org/TR/tr-activity.html#VoiceBrowserActivity

Voice Browser FAQ: http://www.w3.org/Voice/#faq

VoiceXML tutorial: http://www.w3.org/Voice/Guide/

Emerging Ontology Standard OWL strengthens Semantic Web Foundations

W3C has announced the advancement of the OWL Web Ontology Language to Candidate Recommendation. OWL is a language for defining structured, Web-based ontologies which enable richer integration and interoperability of data across application boundaries. OWL is used to publish and share sets of terms called ontologies, providing advanced Web search, software agents and knowledge management. Early adopters of these standards include bioinformatics and medical communities, corporate enterprise and governments. OWL enables a range of descriptive applications including managing web portals, collections management, content-based searches, enabling intelligent agents, web services and ubiquitous computing.

OWL is a Web Ontology language. Where earlier languages have been used to develop tools and ontologies for specific user communities (particularly in the sciences and in company-specific e-commerce applications), they were not defined to be compatible with the architecture of the World Wide Web in general, and the Semantic Web in particular. OWL rectifies this by providing a language which uses the linking provided by RDF to add the following capabilities to ontologies:

- ability to be distributed across many systems
- scalable to Web needs
- compatible with Web standards for accessibility and internationalization
- open and extensible

The OWL Documents produced by W3C

OWL is part of the growing stack of W3C Recommendations related to the Semantic Web. It facilitates greater machine interpretability of Web content than that supported by XML, RDF, and RDF Schema (RDF-S) by providing additional vocabulary along with a formal semantics. OWL has three increasingly-expressive sublanguages: OWL Lite, OWL DL, and OWL Full.

OWL is specified in 6 documents: each aimed at different segments of those wishing to learn, use, implement or understand the OWL language. They include:

- a presentation of the use cases and requirements that motivated OWL
- an overview document which briefly explains the features of OWL and how they can be used
- a comprehensive Guide that provides a walk-through of the features of OWL with many examples of the use of OWL features (in the area of describing food and wine).
- a reference document that provides the details of every OWL feature
- a test case document, and test suite, providing over a hundred tests that can be used for making sure that OWL implementations are consistent with the language design
- a document presenting the semantics of OWL and details of the mapping from OWL to RDF

Web Ontology Working Group includes Industrial and Academic Leaders, seeks Implementations

The advancement of the OWL Web Ontology Language to Candidate Recommendation is an explicit call for implementations. A large number of organizations have been exploring the use of OWL, with many tools currently available. In addition, both the US government (via DARPA and NSF) and the European Union (via the 5th and 6th generation frameworks of the IST program) have invested in web ontology language development. Most of the systems currently using DAML, OIL and DAML+OIL (the predecessor languages that OWL was based on) are now migrating to OWL.

Links:
OWL Overview: http://www.w3.org/TR/2003/CR-owl-features-20030818/
OWL Guide: http://www.w3.org/TR/2003/CR-owl-guide-20030818/
OWL Reference: http://www.w3.org/TR/2003/CR-owl-ref-20030818/
OWL Semantics and Abstract Syntax:
OWL Test Cases: http://www.w3.org/TR/2003/CR-owl-test-20030818/
OWL Use Cases and Requirements:
http://www.w3.org/TR/2003/CR-webont-req-20030818/
OWL FAQ: http://www.w3.org/2003/08/owlfaq/
Home page for W3C's Semantic Web Activity:
http://www.w3.org/2001/sw/

ERCIM is the European host of W3C.

Machine Perception and Understanding: Introduction

by Eric Pauwels

The unrelenting pace of innovation in computer and communication technology continues to affirm the central role played by information in modern society. Indeed, the typical amount of digital storage capacity is doubling every year, and bandwidth in both wired and wireless networks increases at an even faster rate, supporting instant and almost ubiquitous access to an abundance of resources. As a consequence, there is a pressing need for tools that can assist us in exploring the huge quantities of data with which we are confronted when managing large multimedia databases or monitoring complex sensor streams. These data can have intricate spatial, spectral or dynamic structures (eg text that refers to images, audio punctuating video) and are potentially an extremely valuable source of information. However, unless we can extract the knowledge buried in the bits and bytes, all this data serves little purpose. In this respect, it is becoming increasingly clear that in order to be efficient, data processing needs to be content-based. As the enormous size of these collections precludes comprehensive human supervision, the only viable alternative is the development of reliable machine perception and understanding, and in particular, the automatic creation of semantically rich metadata that can be used as input for ensuing high-level processing or decision support.

Addressing these challenges is quite a daunting task. Fortunately, we are witnessing prodigious activity in key scientific and technological areas that promise to have a profound impact on the way we tackle this deluge of information. First, progress in signal processing for the different modalities (image, audio, speech, etc) has given rise to the creation of sophisticated tools capable of performing reliably for specialised sub-problems (eg face- or motion-detection and text-to-speech translation). In addition, researchers are increasingly turning their attention to cross-modal integration, combining different modalities to maximise information extraction and robustness. Concurrently, progress in statistical and machine learning has boosted the wider acceptance of automated learning methodologies, and it has transpired that these techniques can contribute significantly to the automatic exploration and structuring of large datasets.

To underscore the urgency of the data-mining problem and its potential impact on society and industry, the European Commission has made semantic-based knowledge systems a priority theme in its call for Information Society Technologies for the upcoming Sixth Framework. ERCIM has positioned itself to play an important role by submitting a Network of Excellence (NoE) on Multimedia Understanding through Semantics, Computation and Learning (MUSCLE, currently in the negotiation phase). This NoE will run for four years

and should stimulate closer collaboration between European groups working on research projects that aim to integrate machine perception and learning for multimedia data-mining. The consortium members have agreed to focus on different aspects of single- and cross-modal processing (in video, audio and speech) as well as various flavours of statistical learning.

To encourage close coordination of effort and durable scientific integration, MUSCLE will set itself two 'Grand Challenges'. These are ambitious research projects that involve the whole spectrum of expertise that is represented within the consortium and as such, will act as focal points. The first challenge focuses on natural high-level interaction with multimedia databases. In this vision it should become possible to query a multimedia database at a high semantic level. Think *Ask Jeeves* for multimedia content: one can address a search engine using natural language and it will take appropriate action, or at least ask intelligent, clarifying questions. This is an extremely complicated problem and will involve a wide range of techniques, including natural language processing, interfacing technology, learning and inferencing, merging of different modalities, federation of complex meta-data, appropriate representation and interfaces, etc. The second Grand Challenge is related more closely to machine perception and addresses the problem of detecting and recognising humans and their behaviour in videos. At first glance, this might seem rather a narrow scope, but it has become clear that robust performance will heavily rely on the integration of various complementary modalities such as vision, audio and speech. Applications are legion: surveillance and intrusion detection, face recognition and registration of emotion or affect, and automatic analysis of sports videos and movies, to name just a few. For more information on this Network, we invite the reader to visit the MUSCLE Web page (see below).

This special issue highlights the breadth and depth of the research related to machine perception and understanding that is currently being conducted by various ERCIM groups. There continues to be a strong interest in biologically inspired approaches, which is hardly surprising, since nature still easily outperforms the most intricate technical solutions. Another trend that re-affirms itself is the reliance on computationally intensive methodology to extract statistical information and simulate computational models. Equally varied are the applications on display, ranging from mobile robots to support for virtual studios. Researchers are also enthusiastically accommodating the advances in sensor and (wireless) communication technology that support the creation of networks of interacting and context-aware components, thus moving closer to the vision of genuine ambient intelligence. It is our hope that this Special Issue will offer the reader a taste of the exciting developments on which the various ERCIM laboratories are working.

Links:
MUSCLE: http://www.cwi.nl/projects/muscle

Please contact:
Eric Pauwels, CWI
E-mail: eric.pauwels@cwi.nl

ARTICLES IN THIS SECTION

Towards Bio-Inspired Neural Networks for Visual Perception of Motion

by Claudio Castellanos Sánchez, Frédéric Alexandre and Bernard Girau

Data from neuroscience are used to implement cortex-like maps on an autonomous robot, allowing it to detect and analyse motion in the external world.

Visual perception of motion is a major challenge in machine perception research, since it constitutes an important parameter in a wide variety of tasks: path-finding, perception of shape from motion, depth segmentation, estimation of coincidence (time to collision, time to task achievement), determination of motion direction and speed, perception of gestures, orientation and movement control and determination of the three-dimensional structure of the environment.

Figure 1: Three images from a visual sequence and the corresponding motion extracted by the neuronal model.

Recent research in neuroscience has led to a more extensive knowledge of the computational properties of individual neurons and of the interactions of neural circuits in the visual pathways. It has provided us with a solid basis for understanding how the image processing of natural scenes derives from the structure and function of real brains. One of the goals is to study the intricate relations between motor and sensory patterns that emerge during active vision.

It is now known that the detection and analysis of motion are achieved by a cascade of neural operations. This starts with the registration of local motion signals within restricted regions of the visual field. These signals are then integrated into more global descriptions of the direction and speed of object motion.

In the domain of autonomous robotics, the notion of emergence may serve as a major guideline for the integration of different individual tasks. This integration implies the capacity of the robot to manage complex and changeable environments in an autonomous way. The sturdy, adaptive and embeddable properties of both learning and computation inherent in connectionist models are thus particularly well suited here.

Until now, the strong constraints of autonomy had not been satisfied by most neural models for perceptive tasks, in particular visual ones. Our team is working to define a set of connectionist models that perform various tasks of real-time dynamic visual perception bound to autonomous robotics. It is our aim that the methods and resultant algorithms be integrated into autonomous mobile robots working in unknown and dynamic environments. This work is situated at the intersection of two of the research areas of our team: on one hand the conception of connectionist models inspired by the cerebral cortex, which will perform various perceptive and cognitive tasks for autonomous systems, and on the other hand the definition of connectionist models of computation in embedded reconfigurable circuits (to ease the satisfaction of real-time constraints).

Concerning our approach to the visual perception of motion, we mainly focus on two fundamental problems in the treatment of a sequence of images. Firstly, the computation of their optical flow (a three-stage process: pre-processing based on filters, extraction of elementary characteristics and integration into a 2D optical flow), and secondly, the extraction of the egomotion. This work faces many concrete difficulties, such as specular effects, shadowing, texturing, occlusion and aperture problems. Moreover, the complexity of this task must be dealt with within the implementation constraint of real-time processing.

We are currently developing a bio-inspired neural architecture to detect,

extract and segment the direction and speed components of the optical flow within sequences of images. The structure of this model derives directly from the course of the optical flow in the human brain. It begins in the retina and receives various treatments at every stage of its magnocellular pathway through the thalamus and the cortex. More precisely, our model mostly handles the properties of three cortical areas called V1, MT (middle temporal), and MST (middle superior temporal): the MT area detects patterns of movement, while spatio-temporal integration is made at the local level by V1 and at the global level by both MT and MST, so that a multi-level detection and integration may discriminate egomotion from movements of objects in a scene and from the scene itself.

Our first attempts have dealt mainly with manipulation of the optical flow by testing various spatio-temporal filters followed by neural processing based on shunting inhibitions (an example is given in Figure 1). Current efforts are looking at the definition of a model of the magnocellular pathway for the detection of the movement of one or several objects in various dynamic scenes after egomotion extraction. Further research will focus on another challenging mechanism of perception-action in autonomous robotics: the interpretation of the movements of a target in dynamic environments.

Link:
http://www.loria.fr/equipes/cortex

Please contact:
Claudio Castellanos Sánchez, Frédéric Alexandre and Bernard Girau, Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA)/INRIA
Tel: +33 3 8359 2053
E-mail: {castella,falex,girau}@loria.fr

Coupling Perception with Actions via Mirror Neurons

by Jiří Wiedermann

A recent discovery in the brains of primates, mirror neurons are special neurons that show activity both when a subject performs an action and when it observes the same action performed by itself or another (possibly conspecific) subject. Mirror neurons present a versatile self-learning mechanism for the coordination of perceptional and motor actions. They also seem to present a potential bridge between minds. An understanding of the coupling of perception with actions via the mirror neurons, and the management of the respective internal representations, has opened the road to elucidating the computational principles underlying human cognition.

Operating within the framework of long-term research into the theory and applications of neural nets and computational intelligence, a project on machine perception and embodied cognition commenced in 2003 at the Institute of Computer Science of Academy of Sciences of the Czech Republic. The immediate aim has been to elucidate the role of mirror neurons in sensorimotor control and the consequences for imitation learning; a more distant goal is the study of perception-motor coordination in relation to abstract concept formation, language development, understanding, acquisition and generation, and to human-like cognition in general.

The research was initiated by the original hypothesis on the role of mirror neurons in cognition. It has been conjectured that the experimentally observed properties of mirror neurons are only part of a larger associative mechanism that serves to complete cross-modal (perception and motor) information. In this way, the mirror neurons allow 'virtual' (off-line) information to be processed. This hypothesis could therefore explain the mechanisms behind sensorimotor coordination, imitation learning and thinking, since these activities are viewed as specific modes of work of the mirror neural net.

In order to support this conjecture, a computational model of an embodied cognitive agent incorporating the mirror neural net has been designed. The model consists of two neural nets called the mirror neural net and the cogitoid, respectively. Attached to these nets are so-called perception-motor units (PMUs). Each PMU delivers corresponding perception information to the mirror net and the cogitoid sends instructions for motor actions to the PMUs. Perception information is of two kinds: sensory, from the sensors delivering data from the environment, and proprioceptory, delivered by the sensors within each PMU and reporting the current state of that PMU. In the mirror net, perception information meets the instructions that have been issued for PMUs. The resulting stream of data is checked for consistency and completeness by comparing it with previous similar cases, and recovered if necessary. The completed information then proceeds to the cogitoid. Its task is to create concepts corresponding to frequently performed actions and to build associations among them that are used for selecting the next action and instructing the PMUs. The cogitoid was described in ERCIM News No. 53, 2003. The resulting neural architecture of an embodied cognitive agent is depicted in Figure 1. Note that the mirror

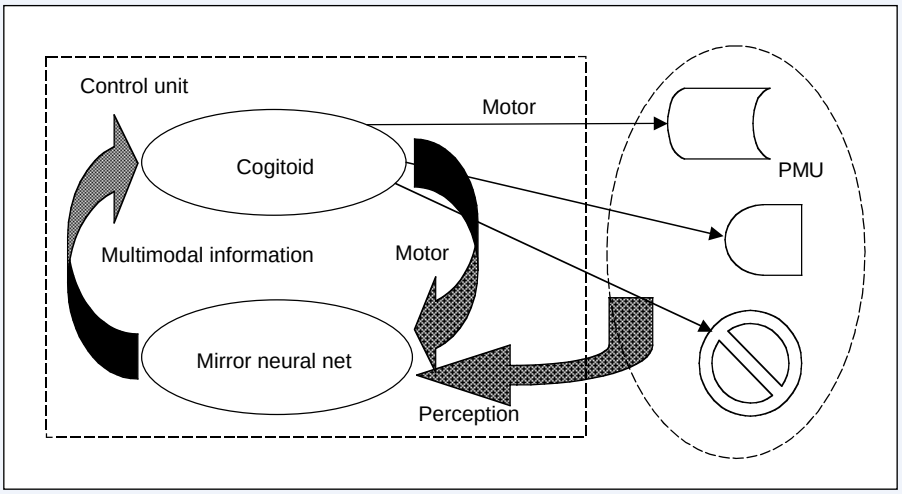


Figure 1: The neural architecture of an embodied cognitive agent.

net takes care of the grounding of concepts, while its associative properties provide 'situatedness' to the agent under circumstances when the perception or motor part of the multimodal information is missing.

The first theoretical results from this formal model of an embodied cognitive agent are well in line with some current theories that, however, are not systematically based on a computational model and usually do not span the whole range of human cognition. This model leads to a clear distinction between data processing and cognitive tasks. It also allows definitions of various cognitive tasks in terms of interactive 'games' between a situated agent and its environment, or between several agents. Its algorithmic properties enable a plausible insight into mechanisms of the mind that were hitherto only anticipated. In particular, the coupling of perception with actions via a mirror net, its associative properties and the introduction of a mechanism for concept and association formation (the cogitoid) have opened the door to understanding the mechanism of understanding, the evolution of inter-agent communication leading to language emergence, and indeed to the development of a rudimentary mind. So far, the project is of a more speculative, rather than experimental nature, but it is assumed that in the near future, the practical potential of such an embodied neural architecture for imitation learning will be realised. The use of Khepera robots and their engagement in various imitation games is envisaged for such a purpose. This phase of the project relies on the involvement of PhD students from the Faculty of Mathematics and Physics at Charles University in Prague, and it is also expected that the Laboratory of Machine Perception at the Czech Technical University will be part of the collaboration. This research has been partially supported by GA CR grant No. 201/02/1456

Links:
See the related technical report V894-03.ps by the author at http://www.cs.cas.cz/

Please contact:
Jiří Wiedermann, Institute of Computer Science, Academy of Sciences of the Czech Republic (CRCIM)
Tel: +4202 6605 3520
E-mail: jiri.wiedermann@cs.cas.cz

Mechanisms for Cognition Development Inspired by the Mammalian Paradigm

by Michalis Maniadakis and Panos Trahanias

In recent years, mechanisms supporting the emergence of intelligence have attracted a lot of attention from researchers. Since the hard-wired, human-designed approach offers limited capabilities in the development of intelligent artificial organisms, modern approaches investigate mechanisms that allow cognition to develop based on the organism's interaction with the environment. In the Computational Vision and Robotics Laboratory (CVRL) of ICS-FORTH, an ongoing research effort studies the mammalian paradigm to develop mechanisms for cognition development in artificial organisms.

During the last few decades, many research efforts have focused on the development of intelligent machines. In the field of robotics, this is usually done by employing detailed algorithms to determine the behaviour of robots in certain environments. These purely computational approaches have facilitated the solution of well-defined problems. However, achieving ubiquity requires that robots be able to dynamically adapt to new circumstances, since real environments can be highly unpredictable. In recognising this fact, it is clear that contemporary, detailed computational approaches offer limited capabilities in approaching real-world problems.

New ideas for approaching the problems related to the development of machines

that operate in the real world can be gained through observation of the everyday actions of biological organisms. Substantial research efforts have therefore been devoted to building 'artificial models' of biological organisms.

Since mammals exhibit the highest level of intelligence among biological organisms, they could be used as an excellent prototype for the development of machines with high cognitive abilities. Modern theories for the explanation of mammalian cognition argue that the observed behaviour of animals is a result of their free interaction with the environment. During environment exploration, subjectively important features are stored in the appropriate structures of the Central Nervous System (CNS). Evidently, this hypothesis could also form a basis for the development of cognition in artificial organisms.

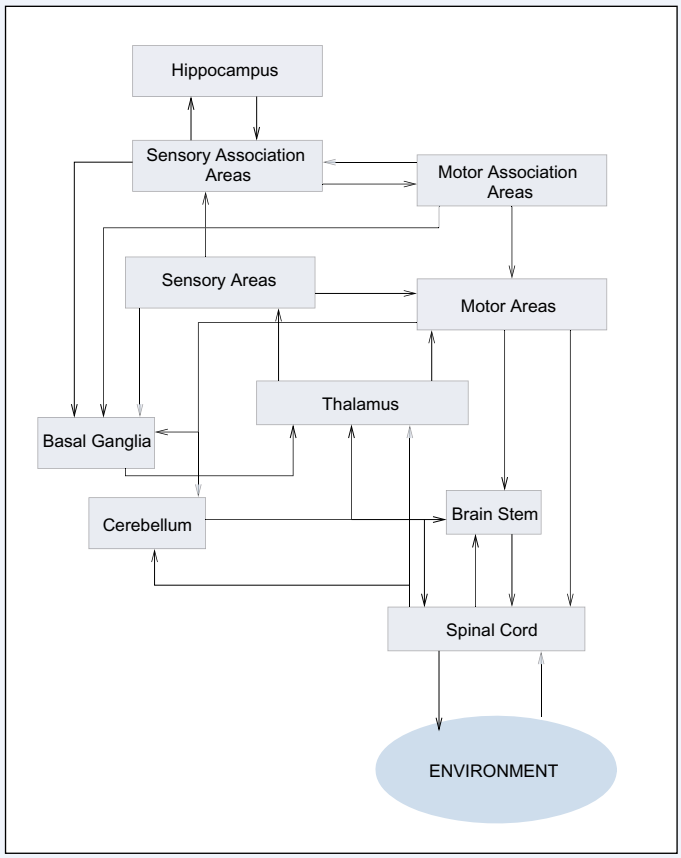


Figure 1: A model of the Mammalian Central Nervous System.

In the Computational Vision and Robotics Laboratory (CVRL) of ICS-FORTH, the functional and anatomical properties of mammalian CNS structures are being investigated in parallel with the development of corresponding computational models. Functional success of the models is gauged by employing a robotic platform to support interaction with the environment. Cooperation between primitive computational modules follows the mammalian paradigm, constituting a global computational model of the mammalian CNS. The model illustrated in Figure 1 shows the basic CNS structures necessary to process sensory and motor information for a mobile robot.

A general but flexible neural network structure has been designed to represent each structure of the CNS. All primitive computational structures have different internal dynamics (eg learning rules), appropriately defined to support the emergence of the desired functionality in the module. The synaptic weights in each neural network are obtained based on interaction with the environment (subjective experience) and internal dynamics. Consequently, the performance of the computational model is not predefined in detail by a human observer to mimic biological activation, but is an emergent result of the robot's experiences and the dynamics which guide the performance of its structural components.

The selection of internal dynamics in each module is performed following an evolutionary approach. A genetic algorithm is employed for each computational module to estimate the internal dynamic parameters. These allow the emergence of a similar functionality to that of the associated area of the mammalian CNS, after a certain number of robot-environment interactions. This approach has already been demonstrated through the implementation of various modules of the CNS model that facilitate robotic localisation and motion control within the workspace.

The global CNS model is obtained following a developmental approach, starting from the performance of simple tasks to gradually accomplishing more complex ones. The cooperation within computational CNS modules is achieved by employing a new evolutionary scheme to select the dynamics of the intermediate connective synapses within primitive modules. This is the current working direction, where primitive behaviours cooperatively form more complex ones.

The development of a robotic CNS (R-CNS) may shed light on the intricacies of real-world problems. Since the R-CNS is intended to support the emergence of adaptive robot behaviours which will be self-organised by subjective experiences, this approach might offer guidelines for the development of intelligence in artificial organisms in general. From a neurobiological perspective, the R-CNS computational model may offer the possibility of integrating human understanding of the mammalian CNS (M-CNS), while new ideas concerning the functionality of M-CNS could be further tested and evaluated.

Link:
http://www.ics.forth.gr/cvrl/

Please contact:
Panos Trahanias, ICS-FORTH
E-mail: trahania@ics.forth.gr

Iterative Orientation Tuning for Contrast Detection in Images

by Marina Kolesnik

A biologically-motivated model for contrast detection in images – the iterative orientation tuning model – has been developed at the Fraunhofer Institute for Media Communication. The model has been inspired by a regular orientation preference layout of simple cells in the primary visual cortex of mammals. The model allows the efficient detection of contours in complex images.

About forty years ago, Hubel and Wiesel discovered that simple cells in cat primary visual cortex (V1) are tuned for the orientation of light/dark borders. The inputs to V1 come from the lateral geniculate nucleus (LGN), whose cells are not significantly orientation selective. The orientation selectivity of simple cells in V1, as proposed by Hubel and Wiesel, is derived from an oriented arrangement of the input from the LGN. Namely, *ON*-centre LGN inputs have receptive fields (RFs) centres aligned over simple cell's *ON* subregions, and similarly for *OFF*-centre inputs. Because of this input arrangement, simple cells perform a linear spatial summation of light intensity in their fields, and their RFs have an elongated shape. The orientation preference of simple cells is quite narrow, and turning a bar-stimulus through more than 20° from the preferred orientation greatly reduces the cell's firing rate.

Although many aspects of simple cell responses are consistent with this linear model, there also are important violations of linearity. For example, scaling the contrast of a stimulus would identically scale the responses of a linear cell. At high contrasts, however, the responses of simple cells show clear saturation. Such behaviour of the simple cells is referred to as contrast invariance of orientation selectivity.

Several computational models have applied the principle of orientation selectivity of simple cells for the processing of natural images generating a pattern of image contours. In fact, any model emulating the functionality of simple cells, has the property of an efficient edge detector. But what is the role of the impressive regularity of simple cells in V1 for the processing of visual stimuli? Simple cells in V1 are aligned into ocular dominance stripes. When the orientation preference of cells in the ocular dominance stripes was related to their position, an astonishingly systematic organisation emerged: the orientation preference changed linearly with its position across V1. Because, furthermore, the orientation columns tended to be at right angles to ocular dominance stripes, their regular layout was nicknamed the Icecube layout.

We have suggested a model for iterative orientation tuning, which explicitly involves the morphology of simple cells for the processing of edges in images. In this model, an activated simple cell amplifies the activity of proximate cells, so that these are becoming locally tuned for distinguished orientation. After several cycles of iterative tuning, the response of simple cells to visual input stabilises. The model architecture is built upon a simple cell circuit composed of

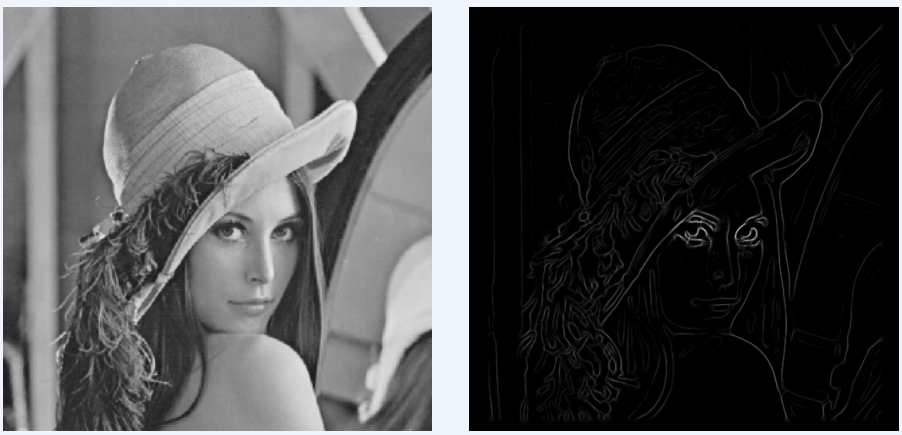


Figure 1: Natural input image and the edge image with eyes enhanced via attentional feedback projection to the LGN.

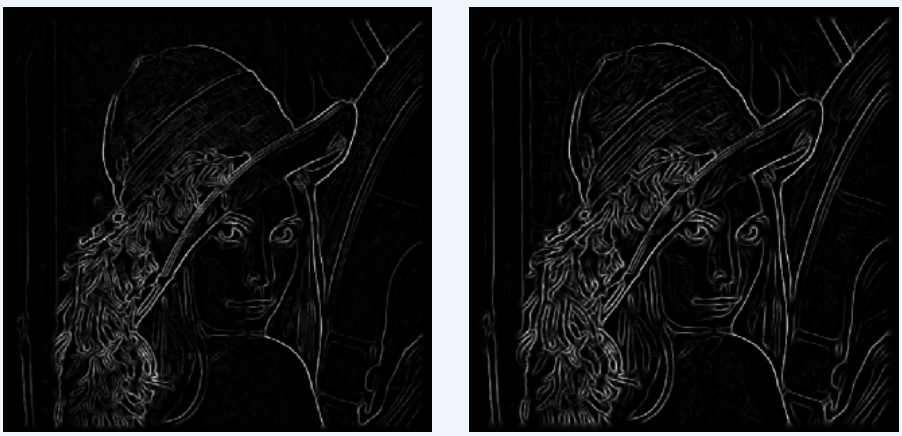


Figure 2: Edge images at 1st and 5th iteration. Closer visual inspection reveals how some initially weak edges become visible at 5th iteration while others, situated in the vicinity of strong edges and visible at 1st iteration, become suppressed and vanish.

segregated ON- and OFF-data streams interacting via a mechanism of opponent inhibition. There are two types of local intracortical interaction each is governed by the spatial layout of simple cells: 1) amplification of the activity level of simple cells that are spatially close within the Icecube layout, and 2) cross-orientation inhibition. The latter is based on considerable experimental evidence suggesting that stimuli at non-optimal orientations suppress the background activity of simple cells.

The iterative amplification of nearby cells enhances responses of both retinotopically proximate cells and cells tuned for similar orientations activating a process of selective orientation tuning. The cross-orientation inhibition affects the activity of simple cells in two ways. On the one hand, the activity is cancelled out by strongly activated retinotopically close cells of orthogonal orientation. On the other hand, the response of strongly activated cells will only be slightly inhibited by retinotopically close cells of orthogonal orientation if these are slightly activated. It follows that the cross-oriented inhibition has the potential to suppress weak responses in the vicinity of contrastive edges and increase model's robustness to noise.

Testing the iterative orientation tuning model with different images, revealed several remarkable properties:

1. The model's response to grey scale images is a fine pattern of image contours. Because of the favourable amplification of weak contrast variations over the strong ones at subsequent iterations, the iterative model is responsible for about 70% of contrast invariance of orientation selectivity.
2. The iterative amplification enhances the response to weak isolated contrast variations. The same level of contrast variations in the vicinity of strong edges does not trigger the amplification. In a sense, the model's response becomes locally attuned to distinguished orientations in the visual pattern. This capability of adjusting the response to either strong or weak contrast changes seems to be natural for biological systems (see Figure 2).
3. The model is highly robust to noise. Noise reduction is especially pronounced in the vicinity of sharp edges, where weak responses are 'cancelled out' by stronger responses due to local interaction.
4. All computational stages employ the local processing. This would make transition to parallel computations an easy task.
5. The computational scheme is flexible and can be extended by incorporating the response of other orientation selective cells, such as complex cells, or color-opponent cells. Further incorporation of memory association feedback mediated by projections to the LGN would turn the detection of contours into a selective process enhancing the response to edges of interest (see Figure 1).

The model's performance gives a powerful demonstration of how the mechanisms of neuronal visual processing, and the morphology of visual neurons can be exploited for the development of advanced image processing algorithms.

Link:
http://viswiz.imk.fraunhofer.de/~marina/

Please contact:
Marina Kolesnik, Fraunhofer IMK
Tel: +49 2241 14-3421
E-mail: marina.kolesnik@imk.fraunhofer.de

Learning and Understanding Bimanual Movements

by Atid Shamaie and Alistair Sutherland

Scientists at Dublin City University have researched a subset of human movements called bimanual movements. At different stages of this research they have approached the problems from the novel points of view. They believe that many machine learning problems can accommodate neuroscience and perceptual aspects of human movements for learning and recognising human behaviours.

Learning and recognising human movements have been given great attention of researchers around the world in the recent years. A broad range of applications from medicine to surveillance and security can benefit from this technology. Learning hand movements and recognising gestures are significant components of such technologies.

Bimanual movements in general form a large subset of hand movements in which both hands move simultaneously in order to do a task or imply a meaning. Clapping, opening a bottle, typing on a keyboard and drumming are some usual bimanual movements. Sign Languages also use bimanual movements to accommodate sets of gestures for communication.

Due to the involvement of both hands, understanding bimanual movements requires not only computer vision and pattern recognition techniques but also neuroscientific studies as a background to perceive the movements.

A cognitive system for bimanual movements learning and understanding entails three fundamental components (see Figure 1): low-level image processing to deal with sensory data, intelligent hand tracking to recognise the left hand from the right hand, and machine learning for understanding the movements.

Using a monochrome surveillance CCD camera the hands are extracted based on the hand grey-levels within the high contrast images. The second component is hand tracking, which is a significant problem due to the presence of hand-hand occlusion. When one hand covers the other partially or completely, they must be re-acquired correctly at the end of the occlusion period.

Studies in neuroscience show that the two hands are temporally and spatially coordinated in bimanual movements. In addition, the components of one hand are temporally coordinated too. These coordinations form the basis of our algorithm to track the hands in bimanual movements.

We have taken a general view of the tracking problem to cover many challenging problems in this area. For example, from a pure pattern recognition point of view a movement can be understood differently when it is seen from different camera view directions. By defining a general set of movement models independent of view angle we have developed the tracking algorithm so that it covers almost every camera view direction. It is trained in just one direction and can be used in other directions. This makes the algorithm independent of the position of the visual system.

Using the temporal coordinations both between limbs (the two hands) and within a limb (a hand and the fingers) the algorithm tracks the hands independent of the hand shapes even in movements where the shapes change. This is especially important from the processing speed point of view. Since processing and understanding the hands shapes is usually a time consuming process, as a component of an integrated real-time recognition system, the tracking algorithm must be fast enough to leave enough room for the other components.

The view-direction and hand-shape independence naturally lends itself to extending the concept of tracking towards mobile vision environments (eg active vision in robotics). We have developed a model to make the algorithm independent from the actual position and velocities. Consequently, it can be used in applications where the visual system (the camera) moves or turns. For example, assuming that the camera is installed on a humanoid robot, the algorithm tracks the hands of a subject while the robot walks.

The third component of the system is the recogniser. As a hierarchical cognitive system, it analyses the hand shapes at the bottom level, learns the individual partial movement of each hand at the intermediate level, and combines them at the top level to recognise the whole movement (see Figure 2). Statistical and spatio-temporal pattern recognition methods such as Principal Component Analysis and Hidden Markov Models form the bottom and intermediate levels of the system. A Bayesian inference network at the top level perceives the movements as a combination of a set of recognised partial hands movements.

The recogniser has been developed so that it learns single movements and recognises both single and concatenated periodic bimanual movements. The concatenated periodic bimanual movements are used particularly in Virtual Reality simulators for interacting with virtual environments. A virtual spacecraft controlled by bimanual gestures is an example.

In all parts of this research we have looked at the problems from the general point of view and developed general solutions. The tracking algorithm can be employed in a wide range of applications including recognition, Virtual Reality, and surveillance/security systems. The recogniser can be used in recognising both single and concatenated periodic bimanual movements.

Our plan for the future is to make the recognition component independent from the camera view direction. This will result in a system that can recognise the movements from the view directions that has not been trained for. Results of the ongoing research in this area will open significant doors towards the general learning and understanding of human movements.

Link:
http://www.computing.dcu.ie/~ashamaie/mvg/

Please contact:
Atid Shamaie,
Dublin City University
Tel: +353 1700 8449
E-mail: atid@computer.org

Alistair Sutherland
Dublin City University
Tel: +353 1700 5511
E-mail: alistair@computing.dcu.ie

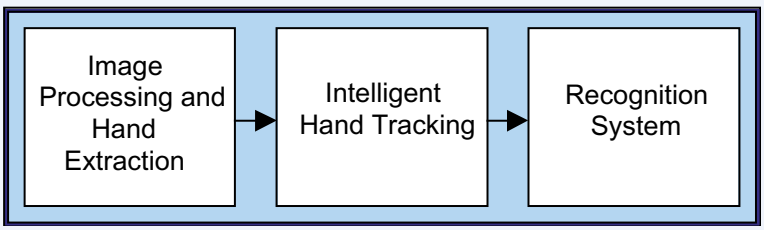


Figure 1: A bimanual movement recognition system.

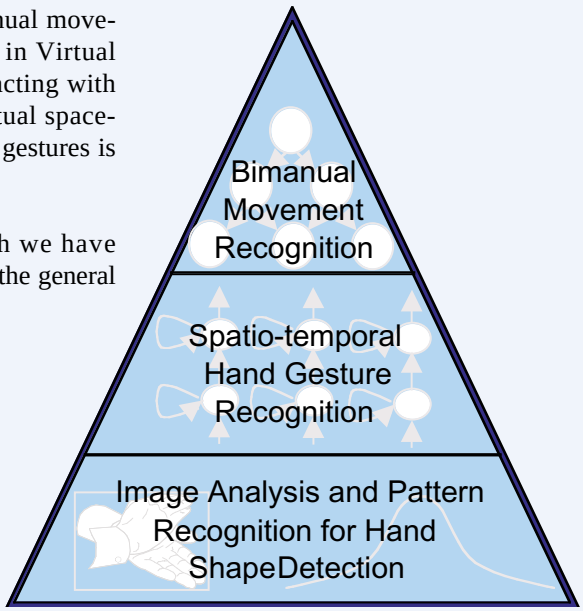


Figure 2. The recognition system.

What We See is What We Compute

by Jan-Mark Geusebroek and Arnold W.M. Smeulders

Understanding human vision is an intriguing challenge. Vision dominates our senses for personal observation, societal interaction, and cognitive skill acquisition. At the Intelligent Sensory Information Systems group of the University of Amsterdam, we have been researching computer vision and content-based multimedia retrieval for some time, and have recently started a new line of research into computational models for visual cognition.

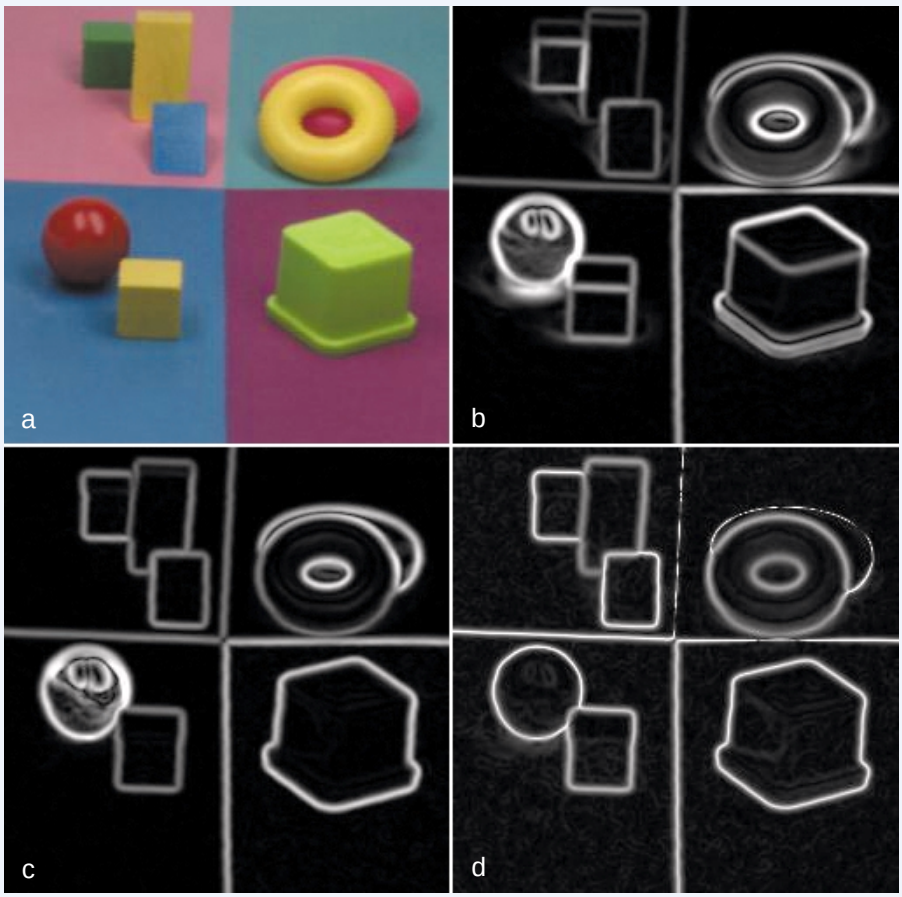
Cognitive vision as defined by the ECVision research network is the processing of visual sensory information in order to act and react in a dynamic environment. The human visual system is an example of a very well-adapted cognitive system, shaped by millions of years of evolution. Since vision requires 30% of our brain capacity, and what is known about it points to it being a highly distributed task interwoven with many other modules, it is clear that modelling human vision - let alone understanding it - is still a long way off. It is also clear that there is a close link between vision and our expressions of consciousness, but statements such as this add to the mystery rather than resolving it.

Understanding visual perception to such a level of detail that a machine could be designed to mimic it is a long-term goal, and one which is unlikely to be achieved within the next few decades. However, as computers are expected in the next ten years to reach the capacity of the human brain, now is the time to start thinking about methods of constructing modules for cognitive vision systems.

For both biological and technical systems, we are examining which architectural components are necessary in such systems, and how experience can be acquired and used to steer perceptual interpretation. Since human perception has evolved to interpret the structure of the world around us, a necessary boundary condition of the vision system must be the common statistics of natural images.

Neurobiological studies have found a dozen or so different types of receptive fields in the visual system of primates. As the receptive fields have evolved to capture the world around us, they are likely to be dual to our physical surrounding. These fields must be derived from the statistical structures that are probed in visual data by integration over spatial area, spectral bandwidth and time. In our cognitive vision research, we have initially derived several receptive field assemblies, each characterising a physical quantity from the visual stimulus.

As the visual stimulus involves a very reductive projection of the physical world onto a limited set of visual measurements, only correlates to relevant entities can be measured directly. Invariants transform visual measurements to true physical quantities, thereby removing those degrees of freedom not relevant for the observer. Hence, a first source of knowledge involved in visual interpretation is the incorporation of physical laws. In the recent past, we have used colour invariance as a well-founded principle to separate colour into its correlates of material reflection, being illuminant colour, highlights, shadows, shading components and the true object reflectance. Such invariants allow a system to be sensitive to obstacles, while at the same time being insensitive to shadows. The representation of the



Courtesy of Theo Gevers, University of Amsterdam.

Figure 1: Invariants for color separating shadow (b), highlights (c), and true object contours (d) from all contours (a).

Figure 2:
Focal attention at regions where spatial statistics break with the generally observed statistics in natural scenes, that is, at regular patterns.

visual input into a plurality of invariant representations is a necessary information-reduction stage in any cognitive vision system.

To limit the enormous computational burden arising from the complex task of interpretation and learning, any efficient general vision system will ignore the common statistics in its input signals. Hence, the apparent occurrence of invariant representations decides what is salient and therefore requires attention. Such focal attention is a necessary selection mechanism in any cognitive vision system, critically reducing both the processing requirements and the complexity of the visual learning space, and effectively limiting the interpretation task. Expectation about the scene is then inevitably used to steer attention selection. Hence, focal attention is not only triggered by visual stimuli, but is affected by knowledge about the scene, initiating conscious behaviour. In this principled way, knowledge and expectation may be included at an early stage in cognitive vision. In the near future, we intend to study the detailed mechanisms behind such focal attention mechanisms.

Links:
http://www.ecvision.info/home/Home.htm,
http://www.science.uva.nl/~mark/

Please contact:
Jan-Mark Geusebroek and Arnold W.M. Smeulders, University of Amsterdam
Tel: +31 20 525 7552
E-mail: {mark, smeulders}@science.uva.nl

Coarse-to-Fine Object Detection

by François Fleuret and Hichem Sahbi

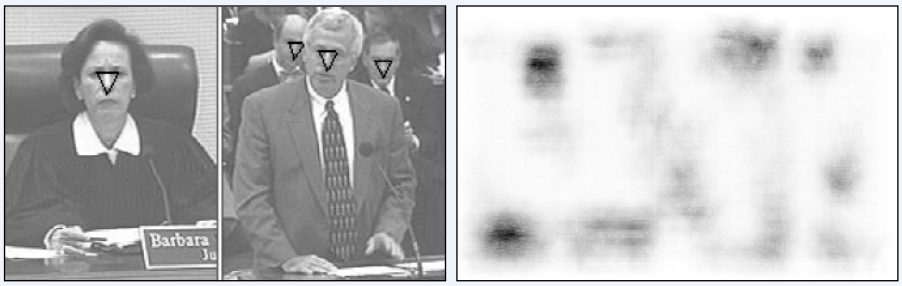
Of all the techniques currently available, object recognition remains one of the most promising approaches to content-based image retrieval. The advent of software able to detect and automatically recognise a wide range of objects in pictures would lead to a new generation of technologies based on high-level semantics. Customers could, for instance, be provided with interactive on-line catalogues, or with improved search engine for TV network archives, which contain usually hundreds of thousands of hours of video..

In the classical approach, object detection is related to the broader field of pattern classification with statistical methods. Given a large number of examples, those techniques build classifiers that are able to label small patches of scenes as object or non-object. Such a classifier is used on an entire scene, and at all possible scales, to achieve object detection. As it can be expected, putting aside issues related to learning, such a brute-force approach necessarily has a very high computational cost. This is an important drawback, since the major areas using detection, such as real-time detection in video or indexing of large databases of pictures, require very low computation times in order to process several scenes per second.

INRIA's IMEDIA research team is studying a family of algorithms which explicitly address the trade-off between error rate and computational cost. We have developed several detectors based on the same idea of a hierarchical composition of classifiers, which reflects a hierarchical decomposition of the views of the object to be detected. Each one of these complex detectors is an adaptive sequence of tests: at each step, a classifier is chosen according to the responses received so far, and its own response is then computed.

This global approach has several important advantages. Firstly, it concentrates the computation on ambiguous areas; trivial areas are rejected early in the process, after a very small computation investment. The second advantage is that it dispatches the representation of the object being detected to a large number of learning machines, which can be very simple. For instance, this approach does not expect the classifiers to learn invariance to geometrical transformations of pictures, as such transformations are taken into account at the global level of the detector. The approach is also generic in respect of the type of classifiers. We have used either very simple edge-counting schemes or more sophisticated wavelet-based support vector machines.

Given a large set of classifiers of varying cost and statistical power, we have



The figure on the left shows the result of face-detection. Each triangle indicates the location of a detected face, with an accurate pose estimation. The figure on the right gives a graphical representation of the computation intensity on the various parts of the pictures. Intensity is estimated by counting the number of times the algorithm takes into account the value of each pixel during the processing of the scene. White stands for a couple of such access, while the dark areas correspond to several hundreds of them.

studied how they can be combined in order to obtain the optimal average cost when processing a scene. This optimisation is based on a statistical model of the relation between the cost and the error rate of the individual classifiers. This leads to a constrained optimisation problem that can then be solved. As might be intuitively expected, both power and cost must grow, so the process begins with simple, almost trivial tests, and goes into details later. More precisely, the growth in complexity is exponential. The resulting strategy is able to reject trivial parts of the picture - like an empty sky or a wall - with very little computational cost (see Figure 1).

Future work will address the handling of large families of objects. Because it dispatches representations to a large number of classifiers, we expect the coarse-to-fine approach to enable control of the growth of both representation and computation when dealing with larger families of objects.

Links:
http://www-rocq.inria.fr/~sahbi/Web/these.html
http://www-rocq.inria.fr/~fleuret/ctf.html

Please contact:
Hichem Sahbi, INRIA
Tel: +33 1 3963 5870
E-mail: Hichem.Sahbi@inria.fr

François Fleuret, INRIA
Tel: +33 1 3963 5583
E-mail: Francois.Fleuret@inria.fr

Video Understanding and Indexing for Surveillance: Image Perception, Quality and Understanding

by Tamás Szirányi

Motion tracking and scene analysis, especially in surveillance systems, require a high-level interpretation of possible shapes and their events, even in the case of incomplete vision conditions and transient motion. To this end, a multi-camera surveillance system has been developed and new efficient algorithms constructed at the Analogical and Neural Computing Systems Laboratory of SZTAKI, with fragments of motion and bodies being grouped through methods of statistical inference.

What is the common thread in the indexing of archive films and the registration of objects in surveillance tasks? In both cases, some understanding of the scene is required, and in the case of films we can be sure that the director's intention lies behind the movement of the camera. In surveillance, however, there is no director, area of focus or even structured scene. Statistical inference is needed to extract meaningful information when it is required. Clever effects such as illumination or associations can help to elucidate incomplete visual data to support human visual understanding. These effects are exploited in composed films, and could be also helpful in interpreting poor surveillance data.

If surveillance problems are related back to better-defined scenes, then Bayesian approaches can lead to an understanding of unanticipated events. The main problem involves the duration of events. While film sequences have a beginning and an end, surveillance events are usually transient dissolving scenes. The arbitrary motions of low-resolution objects (think of a police camera surveying a whole street) are not easily interpreted. However, if we compose beforehand a set of possible events with motion samples and statistical inductions, then real-life surveillance events can be more easily analysed. While it is an achievement to successfully connect the two distinct areas of visual analysis, problems can occur on both sides: definitions of objects and motion in transient events, indexing of sequences and interpretation of scenes.

Motion and shape can hardly be described in real-life applications of noisy video surveillance. Using a greater number of

cameras can help to obtain super-resolution or registered tracks, but this entails calibration of the cameras. The definition of shape or motion in outdoor applications is ambiguous. Unfixed cameras and transient motion of objects together present the challenge of applying statistical inference and optimisation to simultaneously estimate the relative positions of cameras and moving objects. In developing a multi-camera surveillance system, our task is the detection of motion, and the segmentation and characterisation of moving objects. The information coming from the various camera units is interchanged to obtain more precise data concerning direction and vectors of motion, and to recognise the object and its behaviour. These tasks need fast data transmission, comparison of images, and continuous evaluation and explanation of image characteristics. This in turn requires high-speed processing to ensure real-time operation. We are working on relaxation-based object-tracking methods for indefinite object shapes, where moving objects following different paths may obscure each other, or partially fade into other objects in low contrast areas.

Tracking and center definition of street scenery of slowly and discursively moving objects of indefinite shapes.

This research has strong connections with theoretical work being done at the Laboratory of Image Processing in Veszprém University (led by Tamás Szirányi), where a new semi-automatic digital restoration system is being introduced for motion-picture restoration for film archives. The automation is controlled by occasional operator interactions. For this purpose, the film analysis is supported by cut detection and film indexing based on colour information and motion activity. Data representation in XML also aims to create well-defined and controllable processes. When restoring defective films, corresponding scene sequences must be registered with each other by scene-based indexing.

Our surveillance work is supported by appropriate hardware and network tools. These include intelligent camera units with processing and optimised networking capabilities, ultra-fast image-analysis engines of Cellular Neural Network processors, and wired and radio transmission of images and data. The camera system is robust in arbitrary connections and geometry. We are developing the system for automatic grouping and calibration.

Another important issue involves human sensation and the information content of the digital image, including the artistic interpretation of a scene. We have run several human tests for qualifying methods in which artefacts and objects must be sequestered.

Our project is partly supported by the Hungarian National Research and Development Program: TeleSense NKFP 2001/02/035. With this activity, we joined a new Network of Excellence project run by ERCIM: MUSCLE (Multimedia Understanding through Semantics, Computation and Learning).

Links:
http://www.sztaki.hu/~sziranyi
http://lab.analogic.sztaki.hu

Please contact:
Tamás Szirányi, SZTAKI
Tel: +36 1 279 6106
E-mail: sziranyi@sztaki.hu

Psychovisual Evaluation of Image Database Retrieval and Image Segmentation

by Ian Jermyn

Retrieval from large image and multimedia repositories is a pressing problem for everyone from home digital camera users to research hospitals. But how can one evaluate the performance of proposed retrieval systems and their components in an objective way, and thus measure progress and direct further research?

MOUMIR (Models of Unified Multimedia Information Retrieval), is a research training network (HP-99-108) funded by the European Community within the Fifth Framework. Its goal is to advance the state of the art in multimedia database retrieval. A defining part of this effort is the evaluation of retrieval systems and the image or multimedia processing engines that drive them. This problem is fraught with methodological difficulties; the rest of the article describes one attempt to surmount these difficulties in the case of image databases.

Work began with an analysis of possible evaluation methodologies for different classes of image repository. Two examples were taken as paradigmatic: a collection of aerial images of the Ile-de-France region around Paris, courtesy of the French Mapping Agency (IGN), and

Figure 1: on the left, an image from the IGN collection (©IGN); on the right, an image from the Bridgeman collection (©Bridgeman Art Library).

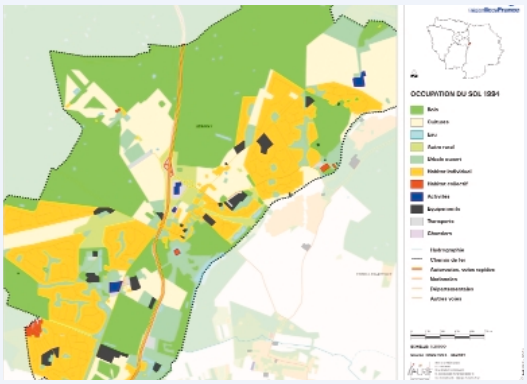


Figure 2: an example of ground truth for part of the IGN image in Figure 1 (©IAURIF).

a varied collection of scanned images of fine art, courtesy of the Bridgeman Art Library. Example images from each collection are shown in Figure 1.

Central to the analysis was the notion of repository semantics: the set of statements that one would like to make about the images in the repository, and which form the basis for retrieval queries. In the absence of such a semantics, the retrieval problem is ill-defined, since we literally do not know what we wish to retrieve. (In particular, this applies to the use of query-by-example.) In the presence of such a semantics, the notion of evaluation via retrieval performance, although well-defined, becomes redundant: if the image processing engine used in the retrieval system can reproduce the semantics, accurate retrieval can be achieved.

Unfortunately, specification of the semantics is very hard for many databases. In this respect, our two reference collections differ fundamentally. In the IGN collection, a semantics for the queries likely to be made of a retrieval system is available in the form of land use maps compiled by the Institute for Urban Planning and Development of the Paris Ile-de-France Region (IAURIF). An example is shown in Figure 2. Evaluation for this collection is thus straightforward: the more accurately an image processing engine can reproduce this ground truth, the better will be the retrieval performance. On the other hand, the BAL images possess a very complicated semantics, even if we consider only statements for which there exists consensus, such as "the image in Figure 1 contains two people", and ignore artistic and emotional judgements. There is little hope of making this semantics both explicit enough to be used for evaluation and simultaneously broad enough to cover a significant part of the semantics of the collection. In this case, can anything be said about how well a retrieval system or the image processing engine that powers it might perform?

It is clear that the semantics associated with the BAL images will at least involve the identification of areas in the images corresponding to different 'objects'. This implies that the output of image segmentation algorithms, which split the image area into a number of disjoint pieces, may capture a reduced and implicit portion of the full semantics and therefore could provide the first stage of the image processing engine for a future retrieval system for the BAL database. Following the reasoning above, an evaluation of the semantic content of these methods then gives an indication of the potential success of retrieval systems based on them. However, since the methods make no explicit semantic statements about the images, no direct comparison of output with semantics is possible. The alternative is to use a psychovisual test with human subjects to compare the semantic content of different segmentation methods, and consequently this was the methodology followed in MOUMIR. The test was designed to answer two questions. First, in the environment of this reduced and implicit semantics, does a consensus about a 'best' segmentation exist? Second, if a consensus exists, what do we learn about the value of different segmentation methods for retrieval?

Figure 3: what the subject saw.

The test worked as follows. First, a subset of the images in the BAL database were segmented using a number of segmentation algorithms of different types. Each subject was then given a series of trials, in each of which they were shown a screen like that in Figure 3. In the centre of the screen was the original image, while on either side were the results of two segmentation algorithms. Each result was displayed in two different ways: the boundaries of the regions above, and the regions themselves below. The subject was told to click on the side of the screen that contained the segmentation that "made most sense". If undecided, they could click in the centre. The choice and the time taken to make it were recorded, the latter being used as a measure of the 'certainty' of the judgement.

The results give a pairwise ordering of the segmentation methods involved, since they are tested two at a time. It is a first sign that a consensus exists that the pairwise ordering is consistent with a single total ordering on the methods. Quantitative statistical tests confirm that the results are not due to chance, thus implying that there is some subject-independent sense in which one segmentation is better than another, at least for this image collection. Of further interest is the fact that the two schemes that performed best were more or less tied, while at the same time being the least and the most complex of the methods tested. The first split the image into areas of approximately equal grey level, while the second used complex colour and texture models. This seems to confirm the folklore that intensity edges, as captured by the first model, are as important as, but no more important than the colour and texture information captured by the second model, but further investigation is needed to analyse the reasons why these schemes performed so well.

The work described above was performed in collaboration with Nick Kingsbury and Cián Shaffrey of the Cambridge University Engineering Department Signal Processing Laboratory. A research report (INRIA RR4761), and publications resulting from this work, are available at the author's home page and on the MOUMIR web site, reached via the links below.

Links:
http://www.inria.fr/ariana
http://www.moumir.org

Please contact:
Ian Jermyn, INRIA.
Tel: +33 4 92 38 76 83
E-mail: Ian.Jermyn@sophia.inria.fr

Searching and Browsing Databases of Digital Decoration Designs

by Mark Huiskes and Eric Pauwels

The FOUNDIT project, coordinated by CWI, aims to develop content-based image retrieval systems particularly geared towards requirements for browsing and searching digital decoration designs. Research is currently directed towards the construction of automatic design descriptions in MPEG-7. Results of this project are currently being tested in collaboration with an Italian fashion designer.

Content-based image retrieval (CBIR) through the searching or browsing of databases of decorative designs such as clothes, textiles and wallpaper remains a challenging problem, especially when the user's subjective appreciation is involved. In these applications, the only way to elucidate the user's preference is by continuously soliciting his/her feedback. This feedback is then harnessed to estimate for each image in the database the likelihood of its relevance with respect to the user's goals whereupon the most promising candidates are displayed for further inspection and feedback.

The goal of the FOUNDIT project is to build a CBIR search engine based on sound principles of inference that can handle the requirements typically encountered in decoration-related image and design databases. The FOUNDIT system comprises the following three modules. The graphically oriented interface (see Figure 1) allows the user to provide the system with relevance feedback by selecting examples and counter-examples which are collected in separate bins. By coupling it to mathematical features, this qualitative feedback is then transformed by the relevance inference engine into a probabilistic measure for each image in the database. The inference engine therefore relies on the availability of pre-computed features that characterise the visual appearance of the images. This feature database is generated off-line by the feature extraction engine.

The feature extraction engine consists of a large collection of algorithms for quantitative image characterisation. The routines are not restricted to computation of low-level features such as global colour and texture measures, but try to establish a link to the more semantically meaningful categories that are typically used by humans when making aesthetical judgments on designs.

In recognition of their vital role in capturing the essence of a design, much effort has been directed towards the detection of so-called salient design elements. Two main strategies are followed to this end: figure-ground segregation (see Figure 2), and grouping of primitives. The figure-ground segregation is based on colour-texture region extraction and

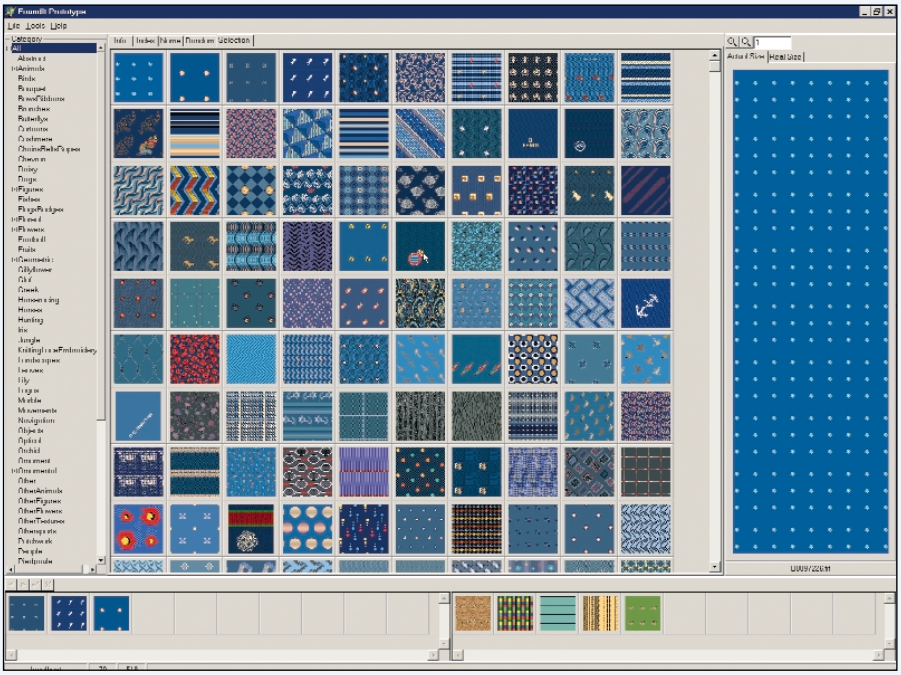


Figure 1: Screenshot of the FOUNDIT prototype interface. Feedback is supplied by selecting positive and negative examples which are collected on the display bar at the bottom (positive examples on the left, negative on the right).

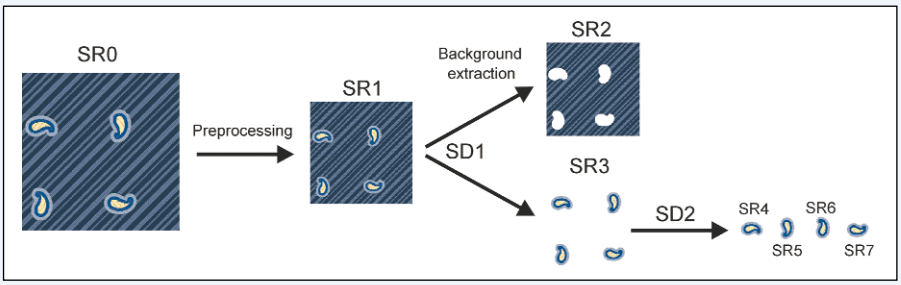


Figure 2: Example of a design segment decomposition using the MPEG-7 StillRegion (SR) and SegmentDecomposition (SD) Description Schemes. The boxes represent semantical entities, the arrows semantical relations.

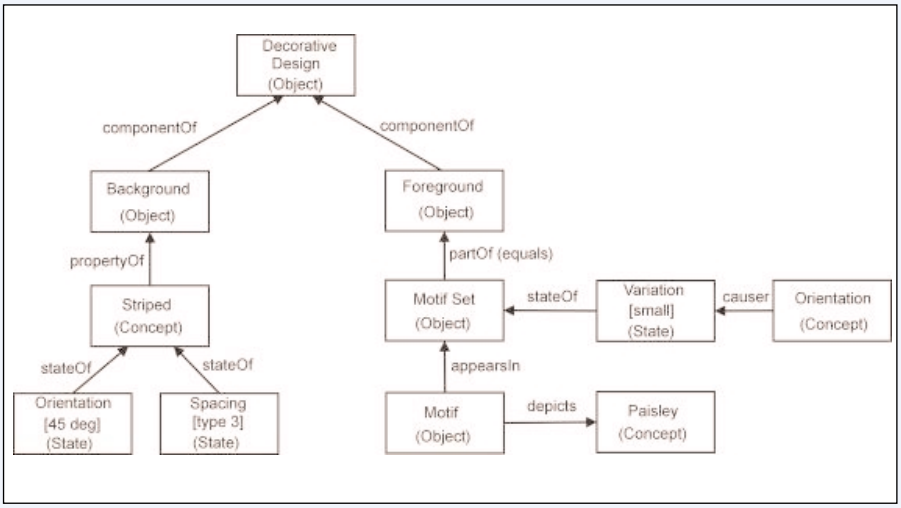


Figure 3: Example of a semantic content description of the design shown in Figure 2.

subsequent region classification based on regional property variables such as relative size, connectedness and compactness. Primitive grouping is directed at finding objects by analysing the configurations of primitive image elements such as edges. In this manner we may, for instance, detect the occurrence and arrangement of homogeneous strips. Based on the decomposition of a design into a ground and one or more salient regions or objects, the feature computation process can be further specialised. For salient objects we compute, among other properties, size, orientation, colour and shape (region and contour-based), and in case of several objects, spatial organisation, occurrence of periodic patterns and motif variation (colour, shape, orientation). The background regions may be similarly characterised, for instance, in terms of colour and textural properties.

Based on the various elements, properties and relationships thus obtained, it becomes feasible to construct an automatic interpretation of the design. The MPEG-7 semantic content-description schemes provide a convenient framework to this end.

An example of a semantic content description is shown in Figure 3 for the decorative design of Figure 2. The decorative design consists of a background and a foreground entity. The background is associated with the abstract concept of 'striped-ness'. Two state entities further specify the quality of the stripes (their orientation and their size). The foreground consists of a set of motifs. The type of motif set may be further clarified by a variation state (indicating to what extent the motifs vary within the set). The motifs that occur within the set, in this case so-called paisley motifs, may then be described independently.

The FOUNDIT Project is partially supported by the European Commission under the IST Programme of the Fifth Framework.

Link:
http://www.cwi.nl/~foundit

Please contact:
Eric Pauwels, CWI
Tel: +31 20 592 4225
E-mail: Eric.Pauwels@cwi.nl

Kurt3D – An Autonomous Mobile Robot for Modelling the World in 3D

by Hartmut Surmann, Andreas Nuechter, Kai Lingemann and Joachim Hertzberg

Kurt3D is an autonomous mobile robot equipped with a reliable and precise 3D laser scanner that digitalizes environments. High quality geometric 3D maps with semantic information are automatically generated after the exploration by the robot.

Precise digital 3D models of indoor environments are needed in several applications, eg, facility management, architecture, rescue and inspection robotics. Autonomous mobile robots equipped with a 3D laser range finder are well suited for gaging the 3D data. We have equipped the autonomous mobile robot KURT2 and a mobile 3D laser range finder for the automatic generation of compact and precise 3D models. The proposed method consists of four steps.

First we use an autonomous mobile robot for gaging in 3D the environment. We use the mobile robot KURT2 and the AIS 3D laser range finder that is built on the basis of a 2D range finder by extension with a mount and a servomotor. The 2D laser range finder is attached to the mount for being rotated. The rotation axis is horizontal. Due to odometry errors, the robot's self localization is an unprecise measurement and therefore can only be used as a starting point for registration of the 3D scans in a common coordinate system. We have developed a very fast registration algorithm which extends the well known ICP (Iterative Closest Point) algorithm and aligned scans in less than 2 seconds.

Second we extract features, ie, planes from registered unmeshed range data. After the 3D scans are acquired we detect planar surfaces in the registered point cloud. No prior meshing algorithms need to be applied. Our algorithm is a mixture of the RANSAC (Random Sample Consensus) and the ICP algorithm. To detect a surface the algorithm randomly selects a point and estimates a plane through two neighbored data points. An ICP based optimization is started, if more than a certain number of data points exist in an epsilon area spanned by the estimated plane. Data points form the model set and their projections to the plane to form the data set for each iteration of the ICP algorithm. The plane is aligned given the measured data within a few iterations. The time-consuming search within the ICP algorithm is replaced by direct calculation of the closest point and the transformation is efficiently calculated by the use of quaternions. The extracted 3D planes are unbounded in size. Surfaces are finally extracted from the points by mapping them onto the planes. A quadtree based method generates the surfaces.

Third the computed planes are labeled based on their relative orientation. These scene interpretation uses the found features, ie, the planes. The background for interpretation comprises generic architectural knowledge. A model of an indoor scene is implemented as a semantic net. Nodes of a semantic net represent entities of the world/model.

Figure 1: The autonomous mobile robot Kurt3D mapping a construction site.

The relationship between the entities are encoded using different connections. Possible labels of the nodes are Wall, Floor, Ceiling, Door, NoFeature. The relationships between the features are parallel, orthogonal, above, under, equal height. The labels above and under are relative to their plane and hence not commutative. A depth first search (backtracking) is implemented to assign the labels to the set of planes according to the constraints in the semantic net. The

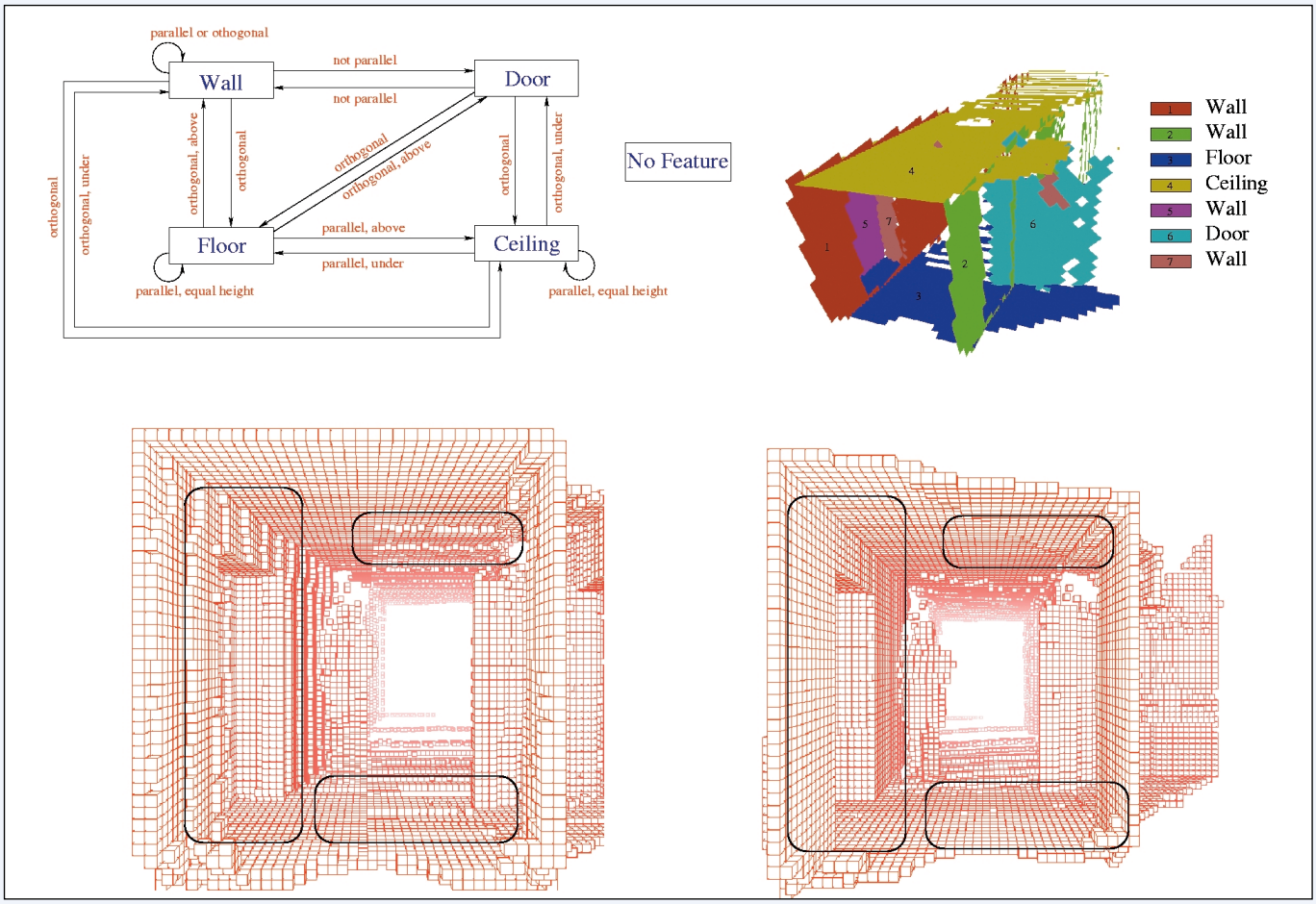


Figure 2: Top left: The semantic net used for the interpretation. Top right: A typical scene (corridor) with extracted labeled planes. Bottom: unconstrained and constrained mesh.

computational expense is reduced by backtracking pruning and reusing (caching) of constraint tests.

Fourth we have implemented a knowledge based approach for the automatic model refinement. The merging of the views as well as the scanning process itself produces noisy data. Due to unprecise measurements or registration errors, the 3D data might be erroneous. These errors lead to inaccurate 3D models. The semantic interpretation enables the software to automatically refine the model. The planes are adjusted such that the planes explain the 3D data and the semantic constraints like parallelism or orthogonality are enforced. The main optimization process uses an error function to enforce the parallelism or orthogonality constraints. The error function consists of two parts. The first part accumulates the point to plane distances and the second part accumulates the angle differences given by the constraints. A suitable optimization algorithm for the the generated error function is Powell's method, because the optimal solution is close to the starting point. Powell's method finds search directions with a small number of error function evaluations. Gradient descent algorithms have difficulties, since no derivatives are available.

The proposed methods have been tested in several experiments with our autonomous mobile robot Kurt3D in several environments eg Birlinghoven Castle and and various office corridors and hallways. The figure shows the reduction of the jitters at the floor and ceiling (circled). The orientation of the model in the bottom image is transformed along the axis of the coordinate system and the meshing algorithm produces flat walls. This transformation is done by using the semantic interpretation. It is not necessary to transform the model interactively into a global coordinate system or to stay in the coordinates given by the first 3D scan.

Future work will concentrate on the aspect of integrating camera images and enhancing the semantic interpretation by fusing color images with range data. The aperture angle of the camera will be enlarged using a pan and tilt unit to acquire color information for all measured range points. Furthermore we plan to concentrate on generating high-level descriptions and semantic maps including the 3D information, eg, in XML format. The semantic maps contain spatial 3D data with descriptions and labels.

Links:
http://www.ais.fraunhofer.de/ARC/3D/
http://www.ais.fraunhofer.de/ARC/kurt3D/

Please contact:
Hartmut Surmann, Fraunhofer AIS
Tel: +49 2241 14 25 18
E-mail: hartmut.surmann@ais.fraunhofer.de

Automatic Fusion and Interpretation of 2D and 3D Sensor Data

by Dmitry Chetverikov

The Geometric Modelling and Computer Vision (GMCV) Laboratory of SZTAKI is working on several projects related to the automatic fusion and high-level interpretation of 2D and 3D sensor data for building rich models of real-world objects and scenes.

The Geometric Modelling and Computer Vision Laboratory conducts research in various areas related to computer vision, image processing, pattern recognition, reverse engineering, computer representation and graphic visualisation of various objects, computational and algorithmic issues of complex curves, surfaces and volumetric objects. The computer vision research is carried out by the Image and Pattern Analysis (IPAN) group of the GMCV lab. The main areas of interest are shape, texture and motion analysis, as well as stereo vision.

A major research goal of the GMCV lab is to build photorealistic 3D models based on multimodal sensor data. We process and combine sensor data from various physical origins, such as camera, laser scanner and CT images, to obtain rich and geometrically correct models of real-world objects and scenes.

In our view, a photorealistic model has three major components - geometry, appearance and dynamics - which must satisfy the following requirements: precision, continuity, high-level description (geometry), texture, realistic surface models, presentation at varying levels of detail (appearance), motion and deformable shapes (dynamics).

To achieve this, we develop, implement and test algorithmic tools for the fusion of 2D and 3D data, feature detection in images and shapes, segmentation of images, curves, point sets and surfaces, classification and matching, surface and curve fitting, and morphing. In our research, we pay particular attention to the following three critical issues: bridging the gap between geometric modelling and computer vision, achieving robustness against noise and outliers and creating efficient and flexible data structures. (A characteristic example of the gap between GM and CV is reverse engineering, where vision still cannot provide 3D data precise enough for accurate geometric modelling.)

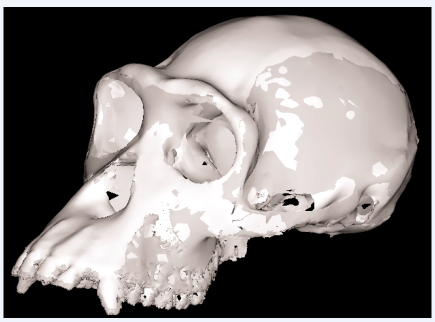


Figure 1: Robust automatic registration (fusion) of partial measurements.

Figure 3: Finding and matching high-level structural features in a wide-baseline stereo pair.

Our current projects are illustrated by a few typical results. Accurate and robust 3D data registration is an important step in reverse engineering and 3D medical image processing. In reverse engineering, the data sets consist of partially overlapping measurements of an object, usually produced by a 3D laser scanner. Figure 1 shows the registration result for a surface represented by more than 100,000 points, while Figure 2 illustrates high-level interpretation of measured data as the final outcome of a long chain of reverse engineering algorithms. In a large-scale national medical project, we have obtained, registered and segmented CT data of various modalities to build a new model of the human knee for knee surgery and prosthesis development.

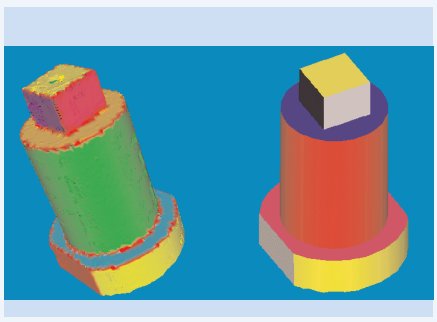


Figure 2: Reverse engineering is high-level interpretation of measured 3D data.

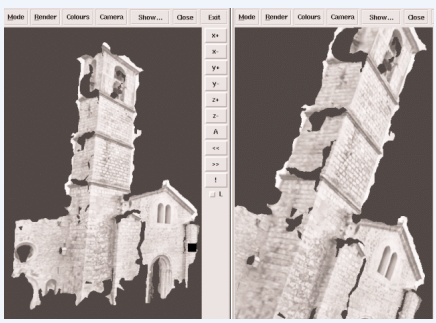


Figure 4: Surface reconstruction from the wide-baseline pair.

A related project run by IPAN is devoted to scene reconstruction from multiple points of view. Special attention is paid to the so-called wide-baseline stereo, where two representations of a scene differ significantly due to divergent viewing angles. Corresponding features of the images therefore have significantly different positions and are subject to affine distortion. This basic problem is often complicated further by occlusions. For these reasons, a critical step of the reconstruction process entails establishing the

initial (sparse) correspondences and building the epipolar geometry.

In our approach, we automatically detect and identify (match) the corresponding high-level structural features of a wide-baseline stereo pair, as illustrated in Figure 3. The structural features considered are the dominant, compact periodic structures, known as periodic distinguished regions, or PDRs.

The initial correspondence between the PDRs is not sufficiently precise to build an accurate epipolar geometry. However, it can be used to obtain a rough affine alignment of the images, in which corresponding local features, such as Harris corners, are much closer to each other than in the initial pair. Once this has been done, a conventional, close-range feature matching procedure is used to find a sufficient number of precise correspondences.

Based on the epipolar geometry, we rectify the two images and apply a recently developed dense matching procedure based on affine region growing. The procedure accounts for affine distortion of the local features that is typical for wide-baseline stereo images. Figure 4 shows an example of surface reconstruction from a wide-baseline pair. Due to the occlusion, some regions are missing, which is typical for this task. Comparison with conventional dense matching techniques shows visible improvement in the quality of the reconstruction.

Building photorealistic 3D models based on multimodal sensor data is a challenging research area. Despite the significant progress made in the last few years, many critical issues remain open. Solving the basic problems described here will open the way to seamless integration of computer vision, geometric modelling and computer graphics, and to the creation of next-generation, high-level photorealistic models. We hope our results will contribute to achieving this ambitious goal.

Links:
http://www.sztaki.hu/gml/
http://visual.ipan.sztaki.hu/.

Please contact:
Dmitry Chetverikov, SZTAKI
Tel: +36 1 209 6510
E-mail: csetverikov@sztaki.hu

AmbientDB: P2P Database Technology for Ambient Intelligent Multimedia Applications

by Peter Boncz

Applications providing intuitive and proactive access to heterogeneous multimedia collections are an important part of the vision of ambient intelligence. CWI has tackled the problem of bridging the gap between such applications and multiple diverse sources of multimedia content using the paradigm of peer-to-peer (P2P) database technology.

Ambient Intelligence refers to digital environments in which multimedia services are sensitive to people's needs, personalised to their requirements, anticipatory of their behaviour and responsive to their presence. Some obvious examples are playing music appropriate to the situation, the location (at a party, in the car) and the mood (stressed, happy), or showing pictures relevant, for instance, to a holiday being discussed. Such audio and video content will be pervasively available in highly distributed multimedia databases found on the Internet (itunes.com), in our homes (PC, TiVo) and cars, and in a mobile form (phones, portable media players).

We envision an environment in which media-rich and intelligent end-user applications run on mobile devices such as PDAs and smartphones. These devices connect in P2P fashion with other devices in the neighbourhood, sharing multimedia content and associated metadata, as well as context data (temperature, light, mood, persons present) stemming from both sensors within the devices and external environmental sensors.

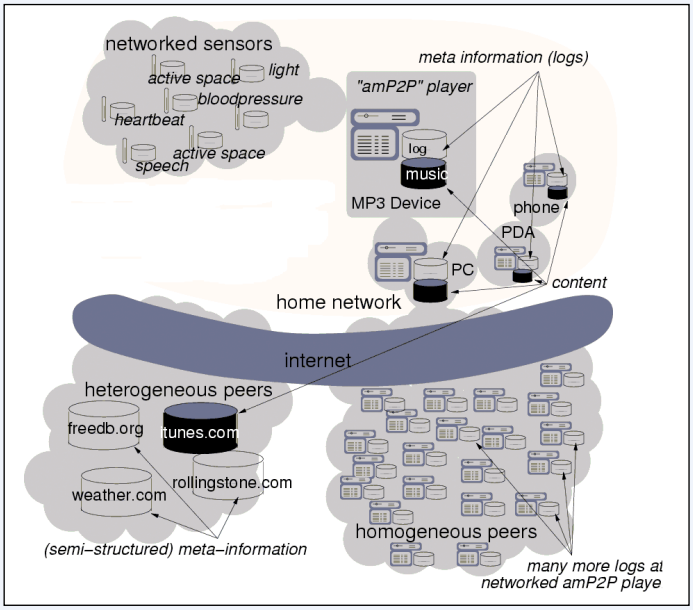
In such a scenario, heavier (non-mobile) semantic multimedia servers that hold and index large multimedia collections may be found on the Internet or in local base stations. Such servers extract features and context data from multimedia objects, and perform off-line indexing and data-mining operations on these, while offering on-line multimedia retrieval services. Such services might allow, for example, nearest neighbour top-N queries on video data using probabilistic frameworks such as Gaussian mixture models.

An important problem in realising ambient intelligent applications is how to exploit all these possible data sources and services. Clearly it is undesirable to hardcode all context and content information management facilities in each application. Such an approach is inherently static, generates a tremendous duplication of effort among applications with overlapping functionalities, and ignores the crucial need for applications to dynamically share both content and context information in order to interact intelligently with each other. Consequently, there is a need for an adaptive and resource-aware middleware layer that shields such applications from these complexities.

The premise of the AmbientDB project is that this middleware role of integrating context-awareness and access to distributed multimedia retrieval servers could be handled by P2P database technology. Such technology allows devices

to self-organise in ad-hoc overlay networks, without the need for central coordination. The principal functionalities it aims to provide are:

- ad-hoc integration of possible diverse information sources into a global schema. This concerns both integration of data access as well as services (queries) supported on this data by the participating peers
- facilities for ad-hoc query execution over this integrated global schema. By using high-level query languages, applications can concentrate on the issue of 'what' to ask, while the database technology concentrates on 'how' to execute such requests in an optimal manner
- a framework for expressing diverse replication, and update/ synchronisation strategies.



Scenario: "generate a playlist of the available content that fits the likes and moods of all current listeners".

One important research question in AmbientDB is which data model and query formulation framework is sufficient to support non-trivial semantic multimedia retrieval queries. The fact that AmbientDB target devices may be mobile and are frequently disconnected, also opens up questions as to alternative update propagation frameworks that can, for example, aim at convergence rather than pure transactional serialisation. Additionally, we look at formalisms that allow expression of application-specific (loose) consistency constraints, replication mechanisms and conflict resolution strategies, using tools such as gossip protocols. Finally, query execution in an ad-hoc network of highly heterogeneous devices poses a number of resource optimisation challenges, which form the current focus of our research.

The AmbientDB project started in the summer of 2002 at the INS1 group of the Information Systems cluster at CWI. This group researches novel database architectures and their integration into advanced application areas, in particular semantic multimedia retrieval. We focus on the area of multimedia in advanced consumer environments, and cooperate in this area with Philips Natlab, which is looking at P2P database techniques to provide integrated data managements for its future generation of networked consumer electronics. This cooperation is part of the MultimediaN national Dutch knowledge infrastructure project. In particular, Twente University will use AmbientDB as its vehicle for a semantic context-based query tool in this project. The MultimediaN project is expected to start in the winter of 2003 and run through to 2007.

Links:
http://www.cwi.nl/~boncz
http://www.cwi.nl/ins1
http://db.cwi.nl/rapporten/abstract.php?abstractnr=1355

Please contact:
Peter Boncz, CWI
Tel: +31 20 592 4309
E-mail: boncz@cwi.nl

Real-time Tracking of Moving Objects for Virtual Studios in TV and Cinema Production Sets

by Florica Mindru, Wolfgang Vonolfen and Ulrich Nütten

Virtual studio technology can be used in several types of applications, such as TV productions, cinema productions, virtual advertisement, and so on. The purpose of this technology is to create a realistic look of images obtained by integrating real and computer-generated audio/video entities.

The continuous improvements in computing power and in the price and quality of the sensors and hardware have led to significant improvements in this technology over the recent years. A virtual studio used as a production set can now provide increasingly higher quality, and it is becoming an affordable production solution. Current developments are focusing on the real-time recovery of the camera model and the scene models, which allows the automatic and correct integration of all elements with respect to the specific models, as well as providing better interaction between the real and the virtual objects.

The virtual studio technology available at the Fraunhofer Institute for Media Communication (FhG-IMK) in Germany is a competitive production tool. This is due to its state-of-the-art camera tracking system, the high quality, real-time rendering tools of the 3D virtual scenes and to the efficient, in-

house developed software system (3DK) at the core of the virtual studio, which were already used successfully in several productions.

One of the objectives of IMK with respect to improving the virtual studio applications is the design of a real-time tracking system that can monitor moving non-rigid objects within the observed scene, and which is able to meet the requirements imposed by two particular applications of the virtual studio technology.

One of these applications is related to virtual studios for TV productions, which are a further extension of the traditional blue-box, or blue-screen technology. The purpose is to create the impression that the moderator is moving and speaking in a virtual world. Virtual sets can be graphically generated with the help of 3D modeling tools. One of the current limitations of the state-of-the-art techniques for virtual studios is related to the seamless integration of the synthetic and real worlds, where the production and integration processes should run in real time and the resulting images should meet the high requirements of picture quality in professional broadcasting. The integration process is still mainly based on a simple mixing technique that does not include the spatial and physical relationships between the elements. Virtual studios offer a large number of new options. For example, animation can be integrated in real time to create more dynamic situations, and connections can be created with interactive interfaces, which can have a direct influence on the set. For integrating such features, keeping track of the position of the moderator(s) in the (real) studio would allow a more realistic integration of the moderator into the generated 3D virtual scene. The goal here is to set up a PC-based, low-cost, virtual studio system, which will broaden the field of application by adding the feature of real-time tracking of the moderator's trajectory in the 3D scene.

The second application is related to the sound system based on wave-fieldsynthesis recently developed at the Fraunhofer Institute for Integrated Circuits IIS in Ilmenau, Germany. The wave-fieldsynthesis theory was developed at the Delft Technical University (Netherlands). The new process developed by Fraunhofer IIS researchers not only records the sound, but also the sonic characteristics of the surrounding space, and information regarding the spatial arrangement of the acoustic source. This makes it possible to reproduce a more natural spatial sound, covering a wide area of the theater. The cinema "Lindenlichtspiele" in Ilmenau is already equipped with the new system.

The blue box studio for TV productions at Fraunhofer IMK.

In order to produce sound that takes into account the spatial arrangement of the acoustic sources, these sources must be identified, and their position tracked during movie recording. For the time being, most of the methods used to achieve the sound sources positioning rely heavily on manual work. Automating as much as possible of this process would reduce the time and costs related to the post-production phase. Again, of course, the goal is to obtain a PC-based, low-cost, real-time tracking system.

Given the applications at which we are aiming, several constraints are imposed on the tracker configuration. In the case of television stations, an important problem to consider is the real-time capability of the hardware and software. Synchronization between the different recording devices is also vital. Movie productions typically involve recording of scenes with various types of backgrounds. In the case of virtual studios, most of the existing recording studios are based on the traditional blue-box technology, and the moderators are segmented out from the studio images with the help of chromakeyers. There is a need, however, for porting this technology also to more general backgrounds.

The tracking system should therefore be able to cope with various types of background, as well as with clutter and partial occlusions. It should also be possible to detect and track various kinds of objects. Within the virtual studios the moderator is of prime importance, but other kinds of objects can also be included in the scenarios. The tracker we want to develop should nevertheless be able to follow general types of objects, because in tracing acoustic sources within movie settings, not only people (the characters), but also other object types should be taken into account. Particular design issues need therefore to be addressed concerning the main steps of the tracking system (moving object detection, tracking, and 3D localization). The system is for the time being in the stage of development and testing.

Links:
http://www.3dk.de/
http://www.emt.iis.fhg.de/virtual.html

Please contact:
Please contact:
Wolfgang Vonolfenn, Fraunhofer IMK
E-mail: wolfgang.vonolfen@imk.fraunhofer.de

Articles in this Section

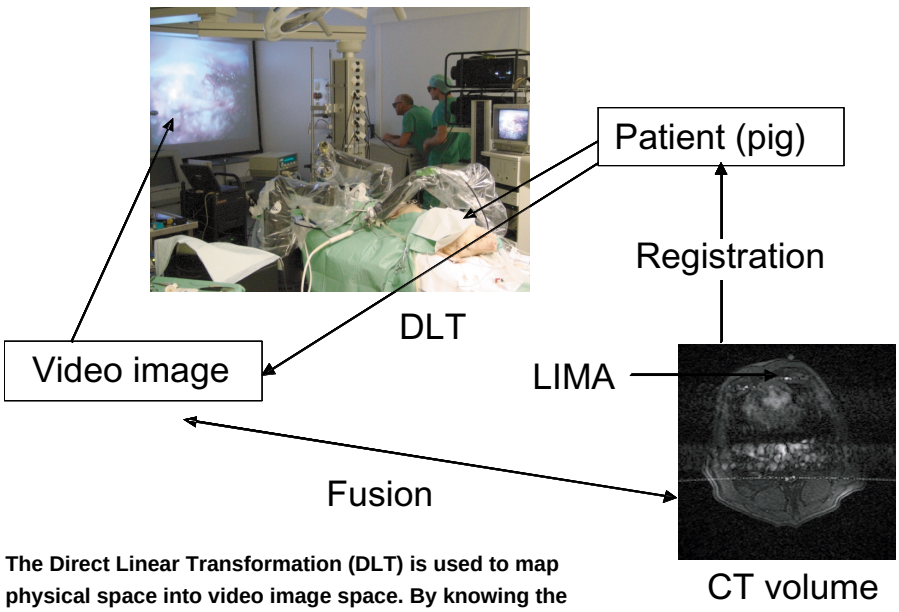
Extrinsic Registration of Heart Volume Images

by Mohamed El Ansari, Ilangko Balasingham, Tor A. Ramstad and Erik Fosse

Minimally invasive therapy has recently become one of the main areas of research in clinical medicine. The procedure is cost effective for both patients and hospitals. Advancements made in medical imaging have initiated the rapid development of minimally invasive surgery or key-hole surgery. The procedure is assisted by video-scopes. Its main drawback is manipulating the surgical fields through small incisions, which reduces direct vision and dexterity, and decreases the tactile feedback. It is hoped that enhanced vision through the addition of multimodal imageries will lead to a reduction in the difficulties faced by surgeons performing these procedures.

Minimally invasive heart surgery is a very useful technique for by-passing obstructed coronary arteries. The left internal mammary artery (LIMA) is a blood vessel located in the chest cavity, and in about 90 percent of all coronary bypass operations it constitutes the best available conduit for a surgical bypass to the major arteries of the heart. However, the soft tissue prevents the LIMA from being seen through the videoscope, and consequently it is difficult for the surgeon to localise it. By fusing vessel anatomy segmented from computed tomography (CT) angiography with the videoscopic images, this procedure can be carried out faster and more safely. The fusion can be done by means of registration, which consists of deriving a mapping transformation between the image space (CT) and the physical space, namely, the patient (see Figure 1).

Volume registration is a new area of research in medical processing and visualisation. Registration of multimodal images makes it possible to combine different types of structural and functional information for diagnosis and surgical planning. Registration of images acquired with the same modality at different times or under different conditions allows the control of disease progression or regression. Registration of preoperative images with the physical space occupied by the patient during surgery is a fundamental step in image-guided heart surgery. It allows the fusion of the segmented LIMA from CT volume with endoscopic video images, thus enhancing the surgeon's field of view and the visualisation of the surgical anatomy.



The Direct Linear Transformation (DLT) is used to map physical space into video image space. By knowing the DLT and the registration between the CT image and physical space, we can fuse the CT and video images.

Our approach is an extrinsic registration method, which is based on the implementation of spherical markers on the skin of the patient. A rigid-body transformation modelled as a rotation matrix and translation vector is considered. The registration approach is achieved in four steps:

- localisation of markers in CT images. A global thresholding method is applied to the CT images which produces binary images. The two morphological operations, erosion and dilatation, are applied to the binary images to remove small details and to fill whole appeared on the markers.
- localisation of markers in patient space. A stereoscopic method is used to compute the marker centroids in patient space
- correspondence between markers on image and physical spaces. The correspondence is achieved by using a heuristic measure of point pair affinity
- computation of the mapping transformation. The rotation matrix and the translation vector are computed by using a least squares method.

The method has been tested on synthesised data, and we hope to apply it to real data and exploit the results obtained to fuse the segmented LIMA from the CT volume into video images.

Mohammed El Ansari currently holds an ERCIM fellowship at NTNU.

Please contact
Mohammed El Ansari, NTNU
E-mail: melansari@yahoo.com

A Machine Vision System Controlling the Cutting of Animal Hide

by Enrico Fantini, Fabio Ganovelli, Paolo Pingi

Based on the integration of image acquisition techniques and real-time systems, an innovative system for cutting raw hides has been developed at ISTI-CNR. The aim is to partly automate the cutting process, so that minimal human intervention is needed.

The current procedure for cutting animal hide is completely manual. The hide is spread out on a bench and expert operators decide the best cutting lines on the basis of the location of specific features. The hide is then manually cut using ad hoc knives and the parts are removed from the bench. The work of cutting the hide is the most time consuming step, and requires three or four workers.

Problem Specifications

Any system to support the cutting of animal skins needs to address the following operational parameters and conditions:

- the dimensions of raw hides can vary up to a maximum width and length of 3.5m
- the whole process should take less than one minute per hide, which is the time currently required by one skilled operator and three assistants
- the cutting system should work within a tolerance of approximately 1cm, measured as the maximal distance between the cuts performed and the ideal cuts. The error in terms of area should not exceed 2.0% of the total area - this is the tolerance allowed when the cutting is accomplished manually.

There are two main sources of error:

- the animal's coat visibly goes against the grain and either the operator does not spend enough time and attention on planning the cuts correctly or mistakes are made while cutting
- the ideal cutting lines of the flanks are not arranged symmetrically about the axis of folding. This error is essentially due to the working methodology since, to save time, the skin is folded in order to make two cuts at once.

To be commercially competitive, any automatic system must be at least as accurate as a human operator and possibly faster. The parameters above clearly define these acceptable levels of performance.

System Description

The system work cycle proceeds as follows:

1. The skin is loaded by hand onto a mobile carriage and left unfolded. Reference marks are positioned on the skin and the carriage is started.
2. A picture of the skin is taken by the camera.
3. A computer vision module recognises the skin contour and marker positions and computes the cutting lines.
4. The computed lines are transferred to a numerical control unit which controls the cutting.
5. The carriage moves back to its initial position and the skin is unloaded.

Image Acquisition System Specifications

The image acquisition system is composed of a camera with suitable optics positioned 4.5m above the hide, a light source positioned near the camera, and the frame grabber.

These specifications are defined by the size of the area to be acquired (up to 3.5 by 3.5m), the limit on the height of the camera (at most 4.5m), and the required precision (0.5cm per pixel). The use of an 8.5mm lens allows these constraints to be met with the introduction of an acceptable distortion to the image.

In the calibration phase, the constants of radial distortion are found, along with the correspondence between the points in the image and the world coordinate. Calibration is only necessary after a hardware modification has occurred (eg if the relative position of camera, water-jet machine or carriage have changed).

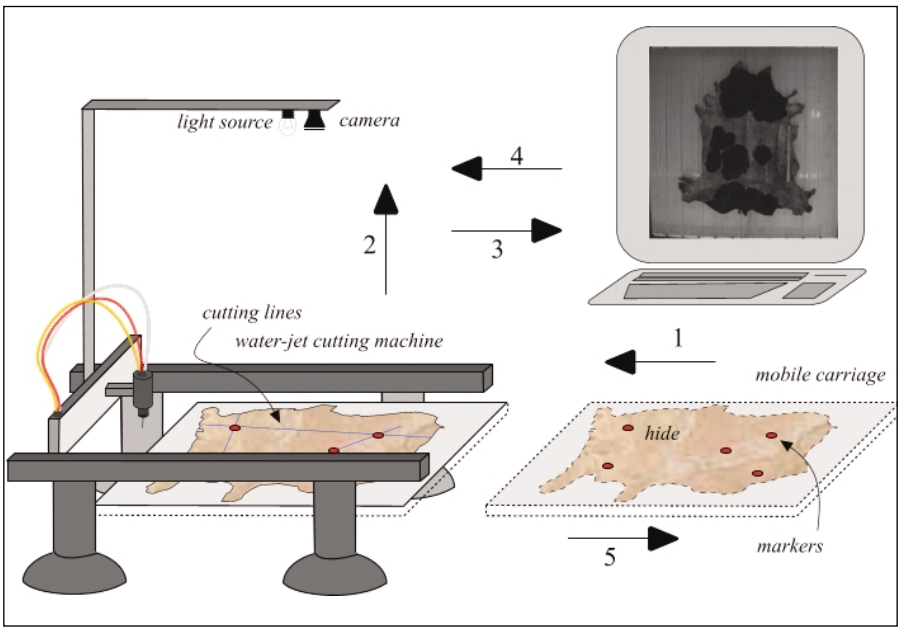


Figure 1: The Cutting System.

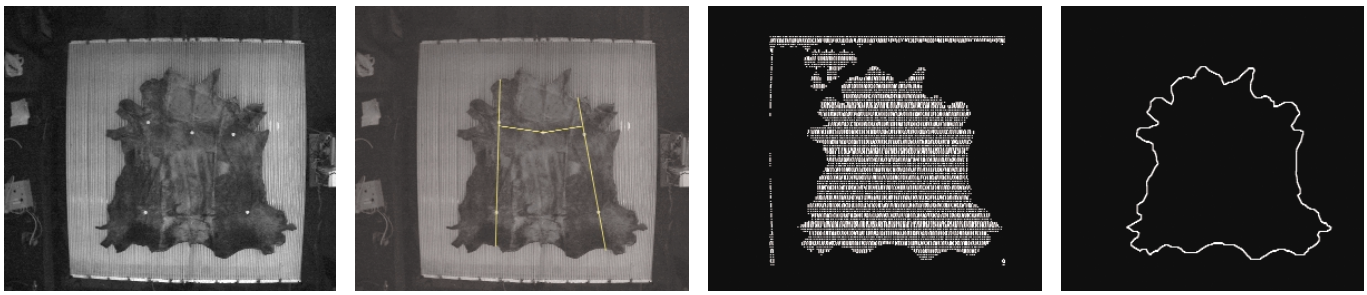


Figure 2: a) The input image; b) Cutting lines; c) detection of inside points; d) contour.

Computer Vision Module

This software module takes as its input the image acquired by the camera, and computes the cutting lines. These are completely specified by the skin contour and the markers' position.

The markers are round-shaped pieces of common white paper, which show up brightly in the image and are distinguishable even on white skin. To detect the skin contour is harder because skins show wide variations in size, shape and colour. The software works on the difference between the carriage plane when it is empty and when the skin is present. More precisely, the carriage plane is made up of many white parallel bars. The empty space between the bars allows the water from the water-jet machine to flow away. Since the bars move under the weight of the loaded skin and also become soiled by the skins, pixel-by-pixel difference between images is not usable. Instead, the system uses an approach loosely based on Snakes (proposed by Kass et al) to detect the white bars, which are known to be present, and to classify the points where bar detection has failed as skin contour. Figure 2 shows the steps of the algorithm.

The process of acquisition and feature detection takes around 5 seconds to be performed. The time for the whole cycle is around 30 seconds.

Water-Jet Cutting Machine

The water-jet has a 0.35mm nozzle operating at a cutting pressure of 3 500 bar and a maximum cutting speed of 30m/min. The cutting bench incorporates specific mechanical modules and the conveyor belt.

The system described has been developed in cooperation with C.G.S. Sas - Ricerca Scientifica Avanzata, Pisa, within the framework of an EC project in the IST programme (IST1999-20188).

Please contact:
Paolo Pingi, ISTI-CNR
Tel.:+39 050 315 29 25
E-mail: paolo.pingi@isti.cnr.it

The European CORBA Components Open Source Initiative

by Marc Born, Andreas Hoffmann, Axel Rennoch, Julia Reznik, Tom Ritter and Alain Vouffo

The building of distributed applications is going to become easier, faster, more efficient and even cheaper. The CORBA Component Model (CCM) standard [formal/02-06-65] of the Object Management Group (OMG) has been a first milestone on this road. The European Commission adopted this programming approach and investigates it in the Open Source project COACH that realises a flexible and complete component framework based on the CCM standard.

CCM is the first non-proprietary, and language and platform independent component architecture. It allows application programmers to design and implement a distributed application independent of specific programming languages, operating systems or vendor-specific communication infrastructures. With CCM it's even easier to build such applications than with plain CORBA.

In the past software developers of CORBA servers had to know about using POA (Portable Object Adapter) or have been restricted to one interface per CORBA object. CCM brings innovations like IDL 3.0, component containers and the new CCM deployment features. Now you can define several configurable attributes as well as provided and requested interfaces (facets and receptacles). Furthermore asynchronous event communication can be introduced due to the definition of event sources and sinks. Developers can concentrate on the business logic instead of reinventing technical infrastructure. They can think about where to place executables instead of how to couple and combine the software. Reusing existing components is

simple and reduces development costs and time to market.

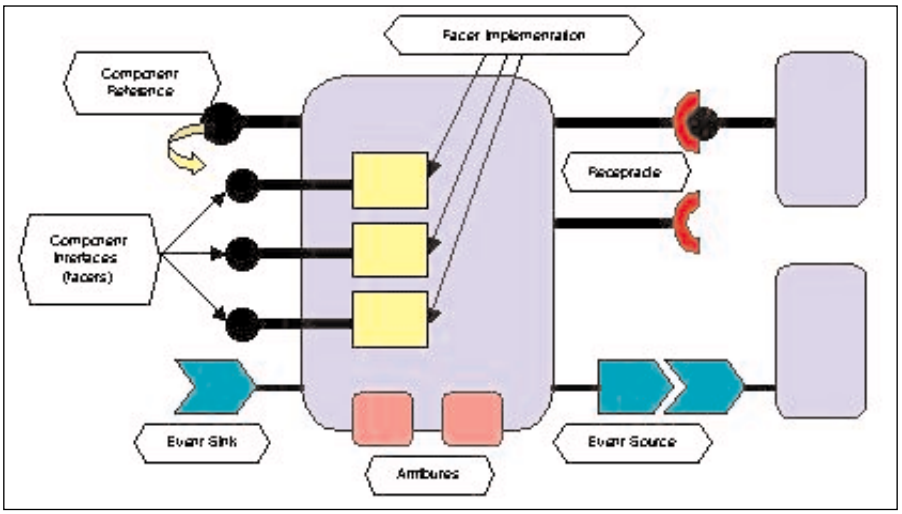
CORBA Components

The European Commission had already identified these challenges and set up the R&D project Component Based Open Source Architecture for Distributed Telecom Applications (COACH). The COACH project provides the CCM infrastructure to the Open Source community and brings the benefits of component programming to everybody who is interested in independency from vendor or programming constraints.

COACH allows the rapid transformation of architecture and design level component to execution level ones. In order to automate and improve CORBA component software developments, COACH provides a complete CCM tool chain supporting the latest version of OMG Interface Definition Language (IDL), the Component Implementation Definition Language (CIDL), components packaging and assembling and XML descriptors.

The complete Open Source CCM platform will consists of IDL/CIDL compilers, code generators for Java and C++, a new flexible runtime container technology, packaging and assembling tools, and a distributed deployment infrastructure. Furthermore, telecom specific CCM adaptations and a support for multiple ORB products will be developed.

As part of COACH, a component test environment and a new security architecture is addressed. The component test environment includes a test framework for test specification, test implementation and test execution, runtime visualization of component interactions and interactive or scripted component testing. The Security architecture covers platform independent definition of security policies, policy definition language, and platform specific enforcement for CCM.



CORBA components.

Two distributed telecommunication applications, a Network Management Framework and a Parlay Platform are going to be implemented to evaluate the suitability of the COACH framework. The use of CORBA components for network management should bring a common solution for integrating the different network management legacy in terms of switching technology (different suppliers) of network nature (terrestrial, submarine) and of network management standards. Taking into account the Parlay Platform will allow the consideration of secure application requirements, access control and usage policy schemes.

The project has a duration of two years and will end in March 2004. You can even benefit today from the COACH working results that are offered with the (Java-based) OpenCCM or the (C++-based) QEDO CCM infrastructure implementations. Furthermore COACH has a great impact on OMG standardization due to several proposals and drafts for new specification standards, eg on Deployment and Configuration of Components, the Unified Modelling Language (UML) profile for CCM, and the UML profile for Modelling QoS and Fault Tolerance Characteristics and Mechanisms.

The availability of a full CCM platform in open source will increase the general know-how in Europe on this technology. Furthermore, it allow SMEs, which typically do not have the means to invest in large developments, to start from the open source for developing business components and finding new markets.

Industry, research labs and universities from five EU countries do work together in COACH: T-Systems Nova (D), Humboldt-University (D), Intracom (GR), Lucent Technologies (NL), Thales Communications (F), LIFL (F), ObjectSecurity (UK), University Paris 6 (F), CNRS (F), and Fraunhofer (D). ERCIM is represented by its German member Fraunhofer Institute for Open Communication Systems (FOKUS) which participates with its two competence centers for Distributed Object Technology, Platforms and Services (PLATIN) and for Testing, Interoperability, and Performance (TIP).

Links:

Project websites:
http://www.ist-coach.org
http://www.fokus.fraunhofer.de/tip/projects/coach

CCM implementation websites:
http://qedo.berlios.de
http://openccm.objectweb.org

Please contact:
Marc Born, Fraunhofer FOKUS
Tel: +49 30 3463 7235
E-mail: born@fokus.fraunhofer.de

INTERMON – Advanced Architecture for Inter-Domain Quality of Service Monitoring, Modelling and Visualisation

by Ulrich Hofmann and Christof Brandauer

Salzburg Research Advanced Networking Centre (ANC) is working jointly with European partners from industry and research institutions on Internet inter-domain performance modelling and visualisation for network planning and inter-domain routing.

Over a number of years, the integration of QoS (Quality of Service) provisioning into Internet technologies was the motivation for several European 5th Framework Program projects (eg AQUILA and TEQUILA). Internet-Telephony, video conferencing and content distribution via the Internet were the driving forces. The requirements are well known (eg telephony requires a transmission delay less than 100ms), but the current QoS provisioning standard 'Integrated Services' architecture has failed because of the large overhead generated by the per-flow resource reservations. The recently developed 'Differentiated Services' provisioning schemes avoid this overhead. These standards and algorithms must be integrated into routers, and must be managed by the network providers per network domain ('intra-domain'). With these mechanisms, the QoS provisioning task can be technically solved per network domain.

However, a large portion of Internet traffic carried by a provider comes from and goes to other domains of the provider or other providers' networks. In such a scenario the provider acts as a transit carrier. Many network providers want to increase their competitiveness in this business area by offering global QoS provisioning. They therefore need a global view of the transmission qualities of other possible transit providers on the path to specified destinations, to find the optimal (eg by cost/performance) inter-domain path.

In Figure 1 for example, network provider A receives a transmission request from the video conferencing application server: {destination = application client, QoS_delay < 100 ms, QoS_rate = 1 Mbit/s }. Provider A knows from the inter-domain routing protocol BGP (Border Gateway Protocol) the two possible paths to the destination provider D: path_1=ABCD, path_2=ACD. Because each of the providers B and C wants to get the contract for the transit transmission from A to D, both offer information on their transmission quality and costs to provider A. The information from provider B to provider A is aggregated for the path BC and path CD (which provider B got from provider C). Now provider A is able to select between the two paths and informs the application server of the delay and the cost of a transmission via domain A to the application client. In order to support such business scenarios, the INTERMON project is focusing on the inter-domain traffic measurement and modelling aspects.

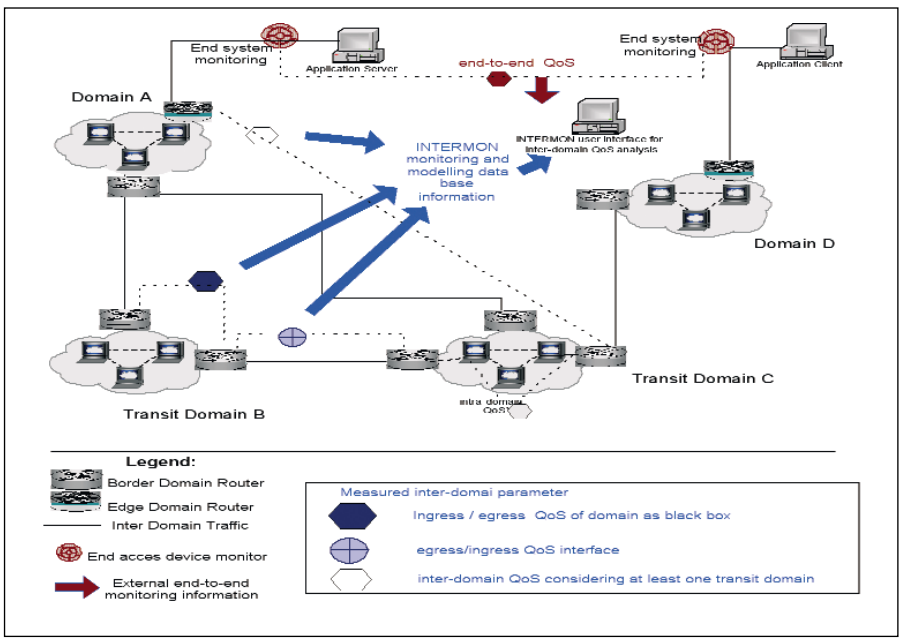


Figure 1: Inter-domain QoS analysis with policy controlled data collection.

Links between network providers generally have high capacities (ranging from 100 Mbit/s to 10 Gbit/s). However for transit traffic, this bandwidth may be reduced by the peering contracts between the providers. To monitor the link load, the data packets must firstly be monitored with the parameters {arrival time [µs or ns], length [bytes], flow_id}. The flow_id allows per-flow traffic monitoring, eg for incoming transit flows with different destinations. So the domain B provider in Figure 1 is able to monitor the traffic between the application servers at the entry and exit routers. In the INTERMON project this is called 'policy-driven monitoring and measurement', because different monitoring application requirements need different monitoring methods (eg for delay measurements, two monitoring points are necessary).

The next step is for this information to become the input for a 'what if' analysis,

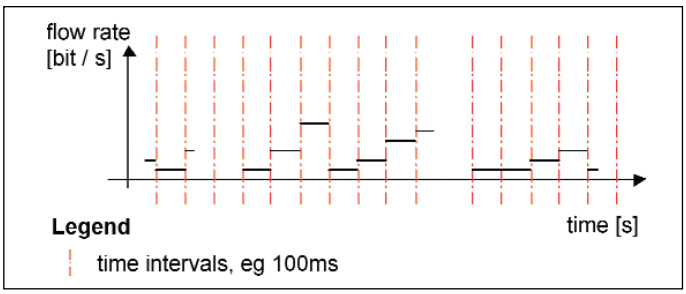


Figure 2 : The simulation process is driven by traffic streams that are modelled as chunks of fluid flows.

eg: "How much increase of delay or loss is caused by an increase of the transit traffic by x%?" To find the answer, the INTERMON project derives inter-domain QoS simulation models. One of these is a so-called fluid approach. Obviously the simulation of complex inter-domain scenarios cannot be done by a per-packet simulation, due to the explosion in the number of simulation events that occur in such large-scale scenarios. For each event the simulator has to update the system state.

The first step towards a scalable simulation model is an approach in which traffic streams are modelled as chunks of fluid flows. The monitored individual packet arrivals are aggregated to traffic load events, eg per 100 ms, and the simulation process is now triggered by these traffic-load events (see Figure 2).

A rigorous next step in this traffic modelling is the transformation of this discrete load process into a continuous 'fluid' process. To retain the most important process characteristics - mean, variance, and autocorrelation - the continuous fluid process is derived from the discrete fluid process by a newly developed iterative algorithm. With this abstraction, inter-domain links become continuous queuing systems and the dynamic relations between incoming fluid-traffic, service link rate, buffer occupation, loss rate etc, are described by differential equations.

On the basis of these models, the 'what-if' scenarios can be executed with the help of powerful existing continuous simulation modelling tools like SIMULINK. To optimally support the network operator/planner, a high degree of automation is employed to create the simulation. Topology information is imported into a graphical user interface (GUI) where the scenarios of interest ('what-if' changes) are configured, and the current load situation is fed from the monitoring database (IPFIX standard) into the simulation.

In general, the fluid simulation algorithms run on a digital computer and the time progress of the differential equations will be approximated by discrete 'sufficiently small' time steps. The dependencies between simulation performance (accuracy, simulation time) and the size of the simulation time step will be investigated in the next project period. Running the fluid simulation with sufficiently small time intervals may be interpreted as a step backward to the granularity of the criticised per-packet simulation, but today's powerful numerical processors are able to solve the approximations for differential equations very efficiently. Finally, a very high degree of accuracy may not always be the main interest in large inter-domain scenarios and the ability to arbitrarily choose the trade-off between simulation accuracy and performance is a powerful feature.

Link:
http://www.ist-intermon.org/

Please contact:
Christof Brandauer
Salzburg Research Forschungsgesellschaft
E-mail:
christof.brandauer@salzburgresearch.at

Ulrich Hofmann,
Salzburg Research Forschungsgesellschaft
and Salzburg University of Applied Sciences
and Technology
E-mail: ulrich.hofmann@salzburgresearch

Inference for Random Sets

by Marie-Colette van Lieshout

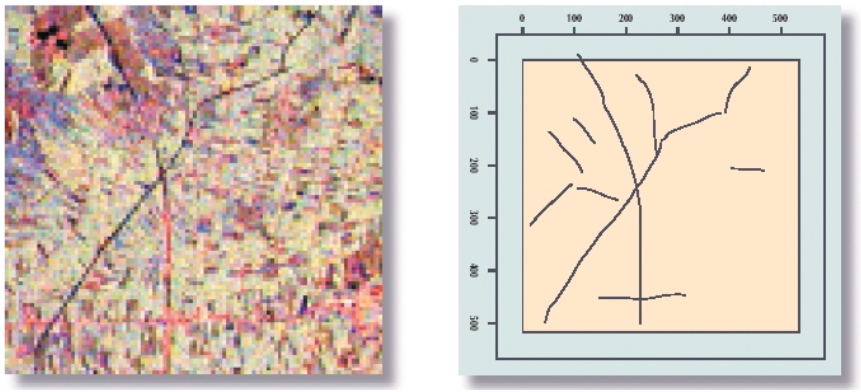
Image analysis and spatial statistics are widely used in medical diagnostics (eg body scans), satellite technology (eg cartography) and analysis of spatial correlation (eg in forestry and epidemiology). At CWI, scientists in the group Signals and Images have studied the problem of extracting linear features such as road networks from remotely sensed images. A new set of methods has been constructed based on Monte Carlo Markov Chain simulation. The corresponding new simulation procedures are faster and more precise than earlier methods.

Many images found in microscopy, material science and biology can be described as a set of independent, randomly placed particles. Think for example of the distribution of pine trees in a forest or, more aggressively, the distribution of fallen bombs. A formal description of such a random set is called a Boolean model.

Notwithstanding the strong independence assumptions, inference is not trivial because of the occlusion effect: Only an image of all the particles – rather than individual particles – is observable.

However, due to interaction, not all images can be described as completely random spatial patterns. Examples are the distribution of cells in the cat retina (eyes), metal particles in an alloy, and road networks. Cats' eye cells arrange themselves in two lattices, hard particles cannot penetrate each other and roads tend to be long and straight and have perpendicular crossings. Therefore, a

class of more realistic models was studied, which permitted correlation between the particles: Gibbs random set models, which are defined by a parametric density, quantifying the interaction structure. The goals of CWI's research were to investigate the maximum likelihood estimation of these models, to devise efficient and exact techniques for them, and to develop useful models for image analysis practice.



Pictures: NASA (left) and CWI (right).

Countryside region in Malaysia and extracted road map.

For the latter goal, CWI researched the practical problem of extracting linear features – such as road networks – from remotely sensed images, in collaboration with the Ariana research team at INRIA. They had previously introduced the so-called Candy model, a Gibbs model for random sets consisting of line segments. In contrast to the Boolean model, there is interaction between the segments in the sense that configurations in which the segments tend to be aligned or cross at angles of about 90 degrees are favoured over patterns with many short, isolated segments or with an abundance of sharp crossings.

Analysis of the interaction structure was used to define an efficient simulation algorithm. In this method the maximum likelihood estimation of the model parameters was carried out, and its uniqueness, consistency and asymptotic normality was established. The ideas were tested on publicly released satellite images obtained from the NASA/JPL website. An example is presented in the figures which show a rural region in Malaysia (left) and the extracted road network (right).

Various perfect simulation algorithms were also generalized to the random sets context and their relative efficiency was investigated for a range of random set models including the Candy model. Most perfect algorithms are based on well-known Monte Carlo samplers. For example, coupling from the past (CFTP) exhibits the dynamics of a spatial birth-and-death process, like the 'Life' simulator. It is particularly effective when the sampled distribution possesses some partial order structure. However, it generates a lot of objects that have no effect on the final outcome. In contrast, the clan of ancestors (ANCS) method aims to avoid births of objects that do not matter in the end, but does not take any model structure into account other than the range of interaction. It turns out that both methods take longer if the interaction strength increases, and that the ANCS method is more sensitive to the interaction range than CFTP.

Finally, a new method was proposed that for the first time allows the use of change moves that alter a single feature of the random set, a crucial step in using exact simulation in image analysis practice. A C++ library that incorporates all the above mentioned models and algorithms has been constructed.

Link:
http://www.cwi.nl/pna4
db.cwi.nl/projecten/project.php4?prjnr=125

Please contact:
Marie-Colette van Lieshout, CWI
Tel: +31 20 592 4008
E-mail: Marie-Colette.van.Lieshout@cwi.nl

Structured Peer-To-Peer Systems: The Impact of Lawyers and Music Piracy on Distributed Systems Research

by Sameh El-Ansary

What began as an initiative for music piracy has evolved into an interesting research area in distributed systems, due to legal rather than technical reasons.

After the decay of Napster as a mainstream peer-to-peer (P2P) system due to legal reasons, the high demand for file-sharing applications led to the appearance of traffic-demanding systems such as Gnutella or Freenet. The large amount of traffic induced by such systems due to their dependence on flooding algorithms constitutes a major concern to ISPs. This is the case even for optimised forms like the Kazaa system, which deploys the notion of super peers.

Initially, P2P systems development was driven by widespread demand for sharing of music files. The expensive solution of flooding-based systems started to attract the attention of research communities, which resulted in the tran-

sition to a new generation of P2P systems that provide a more elegant solution to the problem, namely, satisfying the constraints of decentralisation and scalability. The leading systems of that generation include Chord, CAN, Pastry, and Tapestry. Such systems were collectively said to implement a Distributed Hash Table (DHT) data structure as they provide the simple interface of a hash table: put(key, value), get(key). Unlike flooding-based systems which construct random overlays, DHT systems build structured overlays with desirable scalability properties.

A Unified View to Structured P2P

During our study of the leading DHT systems in the European project PEPITO, we have observed that while all DHTs have a similar interface, each system introduces a new overlay structure, eg, a ring, d-dimensional Cartesian space, torus, etc. Moreover, each system defines an algorithm for lookup, joining of new nodes and handling of leaving and failing nodes. However, having different structures and algorithms for each system makes understanding and reasoning about such systems complicated, and means research results are applicable only to a particular system rather than to the general concept of structured overlays.

Through close inspection of these systems we have reached two main conclusions:

- there exists a common framework based on the simple principle of a *k*-ary search, which is general enough to explain all DHT structures
- the concept of a structured overlay network could be used as a general technique for building scalable distributed systems even in non-P2P contexts.

For the first observation, the distributed *k*-ary search principle could be summarised as follows: given the basic assumptions of DHTs, ie, the mapping of a set of nodes and items into one identifier space of size *N*, each node needs to retain knowledge (routing information) about other nodes in order to lookup and insert items. Each node starts a lookup or insertion process by having a 'view' of the identifier space. A view is a division of the space into k equal intervals. For each interval, there is one 'responsible' node. That is, if a node n wants to insert item *x* in the DHT, it determines the interval to which *x* belongs, say interval *i*, and forwards the item or query about the item to node n_0, where n_0 is the node responsible for the interval *i*. Node n_0, in its turn, will have a 'view' of the interval *i*, ie, it will divide it into *k* parts as well and behave similarly to node *n*. The process ends after $\log_K(N)$ hops, when an interval of length 1 is determined and the node responsible for it is either the node which saves the item or that in which the item is to be inserted. Therefore, one can see that the number of routing entries kept by each node is $k \log_K(N)$ entries. This results from the division of intervals by a factor of *k* after each routing hop and the fact that we have *k* such intervals at each node. Since at each view of a node *n*, one out of the *k* intervals has *n* as a 'responsible', we therefore keep *k*-1 other responsibles per view, making the actual number of routing entries $(k\text{-}1)\log_K(N)$. Given this distributed *k*-ary search approach as a meta-model, we can instantiate other systems by deciding the function for the choice of the responsible of an interval. In Chord for instance, this function is the Successor function, ie, the first node encountered moving clockwise in a circular identifier space. In Pastry and Tapestry, it is a node whose identifier has a prefix with one extra digit match with the identifier of the item being looked up or inserted.

By using this framework, we were able to optimise a generalised form of the Chord system, and designed an optimal broadcast algorithm that takes advantage of the distributed search tree. We also developed a new meta-system based on the distributed *k*-ary search principle.

For the second observation, the consideration of structured overlays as a general architecture for scalable systems could be summarised as follows: given a service *S*, served by a set of resources *R*, we can connect the set *R* in any structured overlay following the distributed *k*-ary search principle. The benefit of such connectivity is twofold:

- no one resource in the set *R* is crucial to the delivery of *S*. This is in contrast to a system like DNS which depends on its root servers for operation
- if *S* needs to scale, the number of resources in *R* grows logarithmically and thus scalability is achieved.

Finally, we can only speculate about what would have happened if Napster had not been shut down due to legal problems. Napster would most probably have continued to operate successfully and solved scalability problems by adopting a Google-like model, ie, having a huge farm of clusters. However, the fall of Napster combined with the high popularity of the application of music sharing forced the research community to look at scalable distributed systems in an unprecedented way and to construct decentralised architectures that constitute the means for building scalable distributed systems.

Links:
http://www.sics.se/~sameh
http://www.sics.se/pepito

Please contact:
Sameh El-Ansary, SICS
E-mail: sameh@sics.se

The Discourse Meter: An Instrument for the E-discourse

by Viviane Wolff

The discourse meter for the discourse support system 'dito', developed at the Fraunhofer Institute for Autonomous Intelligent Systems (AIS), will help participants to become aware of the current state and progress of an on-going e-discourse. It remains to be proven, however, whether the effectiveness and efficiency of the e-discourse will be increased by using this instrument.

Through the continuing implementation of e-government solutions, it is likely that the participative elements of e-democracy will grow in importance. In particular, citizen participation will increasingly take place via e-discourse (electronic discourse). It is anticipated, however, that with the acceptance and growth of e-discourse, a level of complexity that is problematic for participants will soon be reached. This complexity will result from the large number of participants and contributions and the associated diversity of opinion, and it will require a measure of control to be exerted over all participants. The discourse meter is an instrument designed to support discourse awareness, making the process more transparent for participants, through the classification and evaluation of contributions. Complex discourse events will become more visible by the use of query mechanisms, monitoring and the use of graphs and metaphors.

Discourse Awareness

It is necessary for moderators to have access to information about participants' activities in order to efficiently and effectively moderate e-discourses. Information is also required by participants, who need to know about discourse events in order to take part in a concerted discussion. Since the focus on discourse awareness varies depending on the e-discourse phase and the participants' view, our discourse awareness tool is divided into five areas: argumentation, communication, moderation, location and navigation awareness.

Argumentation awareness asks, "What will the outcome be?" Answering this question provides information about an argument's structure and its progress. To this end, a discussion is analysed in order to extract indicators which can serve as measures of opinion and progress with respect to the state of the argument.

Location awareness is concerned with the question: "Where specifically do I go?" It points out, for instance, active and passive discussion sites to the moderator and participants respectively. Location awareness also provides information on how many participants reside in a given discussion section, as well as whether there are any currently published contributions and during which timeframe those contributions were made.

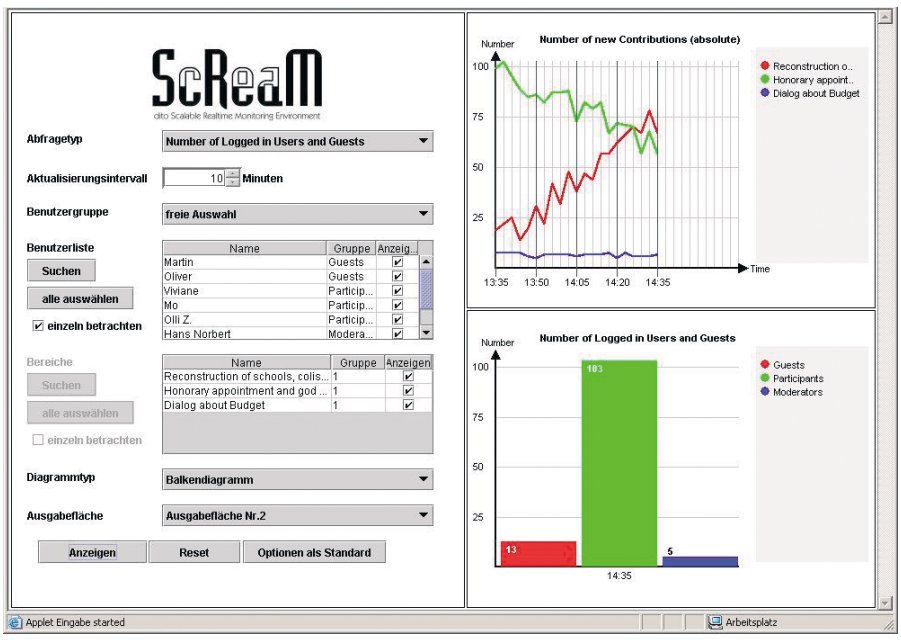
"Who is taking part and how?" This issue is analysed through *communication awareness*, which deals with participants' communication behaviour and detects whether participants are showing interest in each other and who is talkative or reserved.

Moderation awareness evaluates the question: "What does the moderator do and what effect does he/she have?". Moderators do not generally receive any direct feedback on their actions. Moderation awareness counters the lack of physical presence of the participants by providing an opportunity to check the effectiveness of the moderation. For example, has repositioning a 'dead' contribution or a call for a more lively discussion revived the discussion?

Navigation awareness answers the question: "What is the navigation behaviour of the participants?". This provides information on how an individual participant moves within an e-discourse.

Discourse Meter

The discourse meter consists of four components: the event recorder, queries, monitoring and the visualiser. The first component records actions in the form of



Monitoring discourse events.

discourse events. Such discourse events include making a contribution, entering into a discussion section or responding to a certain contribution. The query and monitoring components aggregate these discourse events, which the visualiser then shows as a graph or metaphor. While queries can provide information on past discourse events, the monitoring component is responsible for current discourse events within the hour and minute range.

Queries provide indicators and key numbers from various discussion sections in any given time period. Aside from determining the absolute number of readers and writers of contributions, a query can also calculate the ratio of these to emphasise the activity within an e-discourse. Furthermore, a query reports on the progress of an argument. If at the beginning of an e-discourse there are more questions than answers, then as the discourse progresses, stands will be taken on both sides of the argument. In this way its status can be made clear - whether it is in full swing, has not even started yet or has petered out so that the moderator can close it down.

The monitoring module is equipped with a configuration capability that allows it to accommodate various moderation preferences and discourse phases. Some moderators may be more interested in the actual communication among participants, while other moderators take a greater interest in the argumentation process. A moderator may also choose different configurations in order to control the phases of the e-discourse. Most moderators will most likely attend to the participants whom they have registered for the e-discourse at the beginning of the session. As the e-discourse continues, the moderator will be able to watch as the participants make their own contributions.

Each participant has a restricted view of the discourse meter, meaning participants' privacy can be protected. The user interface of the monitoring module is shown in Figure 1. The left side of this figure shows the input for the configuration parameter, and on the right, the individual output appears in the form of graphs. This example shows which participants are currently logged in, and the number of new contributions processed by the visualiser within the last hour.

Link:
http://zeno8.ais.fraunhofer.de/zeno/web

Please contact:
Viviane Wolff, Fraunhofer AIS
Tel: +49 22 4114 2665
E-mail: viviane.wolff@ais.fraunhofer.de

A Public Access Web Information System for the Italian Ministry of the Environment

by Nicola Aloia and Cesare Concordia

A Web Information Service has been implemented for the Italian Ministry of the Environment providing effective access for a large and diverse user community.

The Web-based Information Systems (WIS) group of ISTI-CNR has just completed the implementation of a public access Web information system for the 'Sustainable Development Service', a division of the Italian Ministry of Environment. The goal is to promote and coordinate nation-wide programmes and initiatives for the public on environmental awareness, education and training. The system provides tools for administration, research, and information dissemination. The 'human component' of the information system is distributed over the Italian territory and connected via Internet. The system users play different roles based on their skill and responsibility. Users with the role of 'Editorial Staff' control and coordinate the information publication activity. The authentication mechanism employs digital certification.

The system architecture has been designed with the following goals:

- to provide better scalability than the single hub-and-spoke architecture
- to build an information system capable of evolving as the needs of the organization grow and change
- to provide asynchronous interaction, eg for the submission and publication of new information
- to enable both 'many to many' and 'one to many' communication, as required, eg for event notifying
- to provide persistence for data and status via transitory storage. This feature is mainly needed for transaction management but is also required to handle problems caused by wireless connectivity, such as loss of connection
- to permit transaction management in order to support applications that involve several actors, each executing a part of the process.

The computer system has been implemented using a three tier architecture (see Figure 1).

The presentation logic is built through a 'multi-client' user interface, where the client can be based on a Web browser (ie HTML), a Java applet, a PDA microbrowser or a Wap mobile phone. The server is in a position to recognize the device from which the access request arrives and to activate the appropriate user interface. The Java applet GUI implements the desktop metaphor,

providing the typical functionality of a windows based system, enabling users to create windows, to magnify them, to move them, iconize them, etc. (see Figure 2). The desktop can contain active objects, such as, for example, the Messenger (active object for events notification), and can easily be extended with new services and functionalities (eg cooperative work features like directories shared via Web, etc).

The whole application layer is implemented using Java technologies to guarantee the platform independence requirement. The SVS server manages two kind of interactions: synchronous and asynchronous. The first ones are implemented by means of Java servlets, provided by an integrated HTTP server, and a set of RMI objects, while the latter are based on a Message Oriented Middleware (MOM). Clients can connect to the right server depending on the kind of operations to be performed.

As the information managed by the system is collected mostly in data-centric documents characterised by a regular structure and fine-grained data, we chose a relational database magagement system (RDBMS) to implement the data layer. The native features of the RDBMS allow high-volume, transactional, and secure database operations.

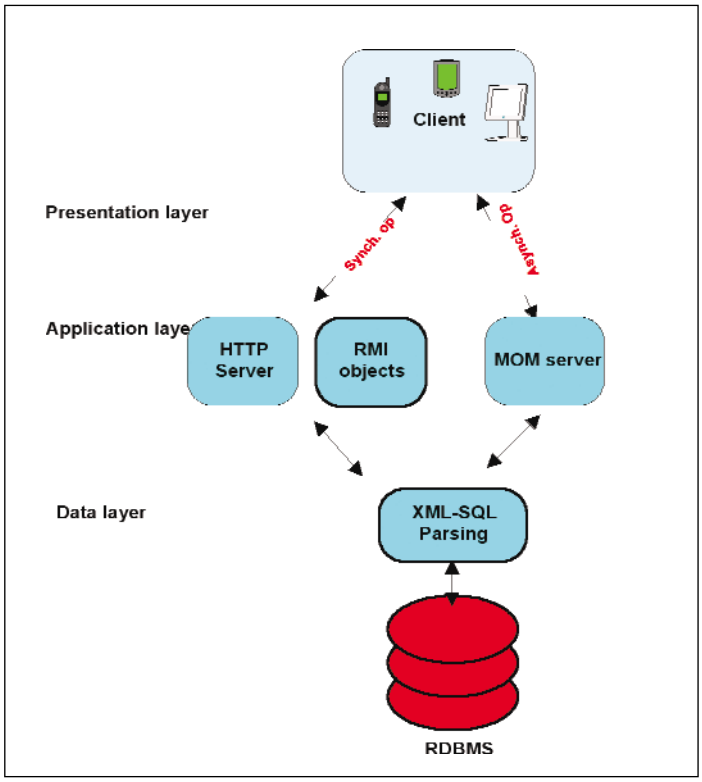


Figure 1: System architecture.

The technology used includes languages and tools mostly deriving from the open-source communities: the HTTP server and servlet container is Jakarta Apache Tomcat, the Message Oriented Middleware (MOM) is JORAM, Java Open Reliable Asynchronous Messaging, developed and maintained by INRIA. Since the data layer is implemented using Java Database Connectivity (JDBC) API, several RDBMS were tested; the production release runs on Microsoft SQL Server. XML has been used to obtain a specific data exchange language among system components. Extensible Stylesheet Language Transformations (XSLT) has been used to obtain the different renderings needed by the presentation layer and Extensible Stylesheet Language Formatting Objects (XSL-FO) for generating reports in various formats.

A few years ago, the opportunity for the public administration to adopt e-government strategies, was largely theoretical; today, thanks to Internet facilities there is a great opportunity to significantly enhance the organization of work in public institutions. Web based Information Systems are, in our opinion, a strategic choice.

Please contact:
Nicola Aloia, ISTI-CNR
Tel: +39 050 315 2997
E-mail: Nicola.Aloia@isti.cnr.it

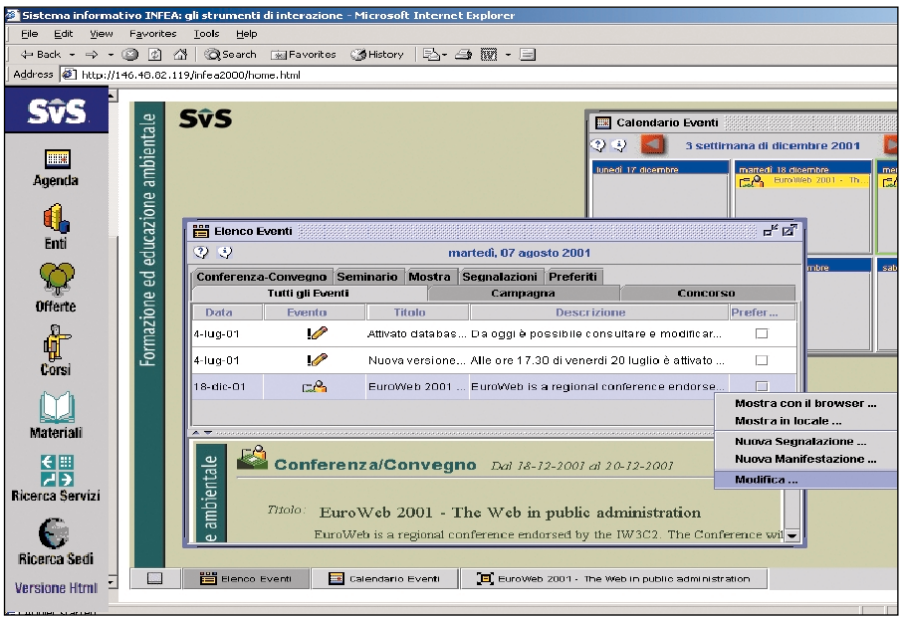


Figure 2: Applet based user interface.

CAPS Entreprise : New Technologies for Embedded Code Tuning

by Ronan Amicel and François Bodin

Designing an embedded system is searching for a trade-off between hardware and software. The fast evolution pace in hardware technology, combined with the increasing size of software, drives the need for new software tools. These tools must be able to speed up the development of new systems, while both decreasing the costs and increasing software portability and safety.

While microcontrollers and small DSP chips are still heavily used, more and more devices are now integrating a large processing power. This increased computational power comes at a price, which is more hardware complexity, as processors instruction-level parallelism and cache memories.

Instruction-level parallelism is used at many levels. First, the processor's functional units are pipelined and instruction-sets are extended with media instructions (eg Motorola's AltiVec for PowerPC) that rely on SIMD (Single Instruction Multiple Data) computations. To achieve still greater performance, VLIW (Very Long Instruction Word) processors are used (such as Philips TriMedia, Equator MAP-CA or Texas C6X), that can issue many instructions at each clock cycle. Contrary to the usual behavior of DSP code, faster code usually means increased code size. Furthermore, to reduce the gap between CPU and memory speeds cache memories are now used in embedded systems, introducing unpredictability in a program's performance, which may be very difficult to handle.

As a direct consequence of this hardware complexity, application developers are facing increasingly complex issues in the code optimization process. Hardware-dependent code optimizations are crucial to fully exploit the target processors. Small changes in the source code can result in dramatic performance changes (either positive or negative). In this context, compilers have a major role to play. First, compilers for embedded systems must now integrate code optimization techniques previously used only for supercomputers (eg vectorization techniques to exploit media instructions). However, unlike in the high performance computing context, embedded systems designers are looking for a trade-off between code size and performance. This constraint alone forbids a simple technology transfer from the high performance domain to the embedded systems world. Tools and compilers must be designed to give the programmer a better control on code quality and on memory usage. Furthermore, they need to bridge the gap between compiler-generated code and assembly code specially crafted by an expert.

In the past ten years, research in the CAPS project at IRISA (a common research lab between INRIA, CNRS, INSA and University of Rennes) has been focused on how to help programmers achieve the best performance on new processor architectures (see http://www.irisa.fr/caps/). To this end, the group developed innovative frameworks allowing to build automatic or interactive optimization tools or fast instruction-set simulators. Furthermore, the research group studied advanced compilation techniques able to deal at the application level with the code size versus execution speed trade-off.

A first environment worth mentioning is an interactive tool for code optimization that relies on artificial intelligence techniques (such as case-based reasoning) to suggest appropriate transformations to the programmer. The suggestions are based on a fine-grained analysis of the structure of program fragments and on their similarity with the situations in a knowledge base.

Another technology helps developers leverage custom instruction-set extensions, such as MMX or other media-oriented extensions, available in their target platform. By combining loop vectorization with a configurable pattern recognition engine, this tool can automatically transform C programs to take advantage of media-processing instructions.

The CAPS group also developed a fast and flexible compiled instruction-set simulation system that significantly reduces the time needed for simulator generation, therefore solving the major limitation of this technology.

CAPS Entreprise was founded by members of the CAPS research group to bring innovative software tools, solutions and services to the market of high performance embedded systems. The company aims at becoming a reliable partner for system builders, platform designers and developers seeking the best system performance, by helping them match their software to the specifics of the underlying hardware platform.

CAPS Entreprise offers standalone tools that are specialized for a given task (code transformation, simulation, worst-case execution time analysis, etc). These tools can act as building blocks in a software tool chain and are designed for seamless integration into common development environments. The company proposes global compilation solutions, tailored to the customer's needs. After a detailed study of the requirements and of the existing process, specific additions and enhancements to the previous code generation infrastructure are proposed and implemented. CAPS Entreprise finally offers custom consulting services, such as performance analyses or instruction-set evaluations. Through these services, customers benefit from the company's in-house expertise and tools, helping them make strategic decisions on complex technology issues.

Links:
http://www.caps-entreprise.com/

Please contact:
contact@caps-entreprise.com

HCI International 2003

by Constantine Stephanidis

HCI International 2003 was held in Crete, Greece, June 22-27, jointly with the Symposium on Human Interface (Japan) 2003, the 6th International Conference on Engineering Psychology and Cognitive Ergonomics, and the 2nd Conference on Universal Access in Human Computer Interaction (UAHCI). The Conference General Chair was Prof. Constantine Stephanidis, Deputy Director of ICS-FORTH and Chair of the ERCIM Working Group "User Interfaces for All" (http://www.ui4all.gr/).

This event, which was sponsored by ERCIM, attracted more than 1,500 participants from 65 countries (the biggest attendance so far in this Conference series) representing the research and academic communities as well as industry.

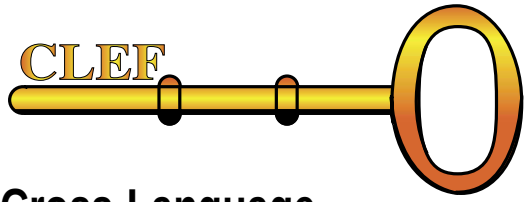
The Conference Programme was organised into 5 thematic areas, namely Human-Computer Interaction, Ergonomics and Health Aspects of Work with Computers, Human Interface and the Management of Information, Universal Access in Human-Computer Interaction and Engineering Psychology and Cognitive Ergonomics. The programme featured 2 plenary sessions, 204 parallel paper sessions, 27 tutorials, 4 workshops, 7 special interest groups, as well as 122 poster presentations and and 12 demonstrations. The keynote speakers were Prof. Ben Shneiderman, from the University of Maryland, USA, who presented "Leonardo's Laptop: Human Needs and the New Computing Technologies" and Prof. Jenny Preece, from the University of Maryland Baltimore County, USA, who talked about "Designing Sociable, Universally Usable Online Communities".

The Proceedings have been published by Lawrence Erlbaum Associates in four volumes containing 1,134 papers for a total of 5,909 pages. The first two volumes are dedicated to Human Computer Interaction (ISBN 0-8058-4931-0, ISBN 0-8058-4931-9, the third volume collects papers on cognitive, social and ergonomics aspects of computing (ISBN 0-8058-4931-7), while the fourth volume is dedicated to Universal Access in HCI (ISBN 0-8058-4931-5). The adjunct proceedings include the abstracts of 122 posters and 12 demonstrations.

HCI International 2005 will be held in Las Vegas, Nevada, USA, 22-27 July 2005, jointly with the Symposium on Human Interface (Japan) 2005, the 5th International Conference on Engineering Psychology and Cognitive Ergonomics, the 3rd International Conference on Universal Access in Human Computer Interaction, the 1st International Conference on Virtual Reality, and the 1st International Conference on Usability and Internationalization (http://hcii2005.engr.wisc.edu).

Links:
HCI 2003: http://www.hcii2003.gr/
ERCIM Working Group UI4All:
http://www.ui4all.gr/

Please contact:
Constantine Stephanidis, ICS-FORTH
E-mail: cs@ics.forth.gr



Cross-Language Evaluation Forum - CLEF 2003

by Carol Peters

The results of the fourth campaign of the Cross-Language Evaluation Forum were presented at a two-day workshop held in Trondheim, Norway, 21-22 August, immediately following the seventh European Conference on Digital Libraries.

The main objectives of the Cross-Language Evaluation Forum (CLEF) are to stimulate the development of mono- and multilingual information retrieval systems for European languages and to contribute to the building of a research community in the multidisciplinary area of multilingual information access. These objectives are realised through the organisation of annual evaluation campaigns and workshops.

Each year, CLEF offers a series of evaluation tracks designed to test different aspects of mono- and cross-language system performance. The intention is to encourage systems to move from monolingual text retrieval to the implementation of a full multilingual multimedia search service. CLEF 2003 offered eight tracks evaluating the performance of systems for monolingual, multilingual and domain-specific information retrieval, multilingual question answering, cross-language image and spoken document retrieval.

The main CLEF multilingual corpus now contains well over 1,600,000 news documents in nine languages, including Russian. A secondary collection used to test domain-specific system performance consists of the GIRT-4 collection of English and German social science documents.

CLEF 2003 posed a number of challenges, in particular with respect to the multilingual and bilingual tracks where the aim was to encourage work on many European languages rather than just those most widely used. There were two distinct multilingual tasks; the most challenging involved retrieving relevant documents from a collection in eight languages: Dutch, English, Finnish, French, German, Italian, Spanish and Swedish, listing the results in a single, ranked list.

Tasks offered in the bilingual track involved "unusual" language pairs: Italian -Spanish, German - Italian, French - Dutch, Finnish - German. We were very pleased by the number of groups that attempted these difficult tasks.

Another positive aspect of CLEF 2003 was the number of new tracks offered as pilot experiments. These included mono- and cross-language question answering, and cross-language image and spoken document retrieval. The aim has been to try out new ideas and develop new evaluation methodologies, suited to the emerging requirements of both system developers and users with respect to today's digital collections. This year's interactive track included full cross-language search experiments where the user attempts to find relevant documents using a complete interactive cross-language system which provides assistance in both query formulation and document selection.

Participation in the CLEF 2003 campaign was slightly up with respect to the previous year with 43 groups submitting results for one or more of the different tracks: 10 from N.America; 30 from Europe, and 3 from Asia. As in previous years, participants consisted of a nice mix of new-comers and veteran groups. Another important continuing trend is the progression of many returning groups to more complex tasks, from monolingual to bilingual, from bilingual to multilingual.

The campaign culminated in a Workshop held in Trondheim, Norway, 21-22 August. More than sixty researchers and system developers from academia and industry attended the Workshop in order to discuss the results of their experiments. In addition to presentations by participants in the CLEF campaign, talks included reports on the activities of the NTCIR evaluation initiative for Asian languages, and on cross-language information retrieval work at Moscow State University. The final session discussed proposals for future evaluation activities within the CLEF framework.

The presentations given at the CLEF Workshops and detailed reports on the experiments of CLEF 2003 and previous years can be found on the CLEF website at http://www.clef-campaign.org/

Please contact:
Carol Peters, ISTI-CNR
Tel: +39 050 3152897
E-mail: carol.peters@isti.cnr.it

RDP 2003 – Federated Conference on Rewriting, Deduction and Programming

by Salvador Lucas

The Federated Conference on Rewriting, Deduction and Programming was held in Valencia, Spain, from 8-14 June 2003. The meeting was organized by the Extensions of Logic Programming group at the Departamento de Sistemas Informáticos y Computación (DSIC) of the Technical University of Valencia. RDP 2003 consisted of the 14th Int'l Conference on Rewriting Techniques and Applications (RTA'03), the 6th Int'l Conference on Typed Lambda Calculi and Applications (TLCA'03), a number of workshops.

During the 12th International Conference on Rewriting Techniques and Applications (RTA 2001) in Utrecht, The Netherlands, in May 2001, the steering committes of RTA and the International Conference on Typed Lambda Calculi and Applications (TLCA) decided to jointly celebrate both conferences (for the first time) in Valencia, Spain, during 2003. A number of other meetings also decided to join the RTA and TLCA conferences to constitute the Federated Conference on Rewriting, Deduction and Programming (RDP 2003). Satellite workshops and meetings included:

- 5th Int'l Workshop on First Order Theorem Proving, FTP'03
- 4th Int'l Workshop on Rule Based Programming, RULE'03
- 16th Int'l Workshop on Unification, UNIF'03
- 12th Int'l Workshop on Functional and (Constraint) Logic Programming, WFLP'03
- 3rd Int'l Workshop on Reduction Strategies in Rewriting and Programming, WRS'03
- 6th Int'l Workshop on Termination, WST'03
- the annual meeting of the International Federation for Information Processing (IFIP) Working Group 1.6 on Term Rewriting.

The scientific program of RDP 2003 included 13 invited talks given by prominent researchers. From the many paper submissions, 141 scientific communications were accepted for presentation at RDP 2003. Among them, 19 correspond to system descriptions. The proceedings of the RTA and TLCA conferences were published by Springer-Verlag in the Lecture Notes in Computer Science (LNCS) series. The proceedings of the workshops were published as Technical Reports of the Departamento de Sistemas Informáticos y Computación. The final proceedings of FTP, RULE, WFLP and WRS workshops are published by Elsevier as part of its Electronic Notes in Theoretical Computer Science (ENTCS) series.

RDP 2003 was hosted by the Departamento de Sistemas Informaticos y Computacion (DSIC) at the the Technical University of Valencia. The conference took place at the ADEIT buildings of the 'Fundació Universitat Empresa de Valencia', a foundation created to promote collaboration between the university and industry. RDP 2003 was organised by the members of the Extensions of Logic Programming (ELP) group at the Universidad Politécnica de Valencia. See the URL below for further details about the members and activities of the group.

The conference received financial support from the European Network of Excellence in Computational Logic (CoLogNet), the European Thematic Network on Applied Semantics (APPSEM II), the Spanish Ministry of Science and Technology (MCYT), the Valencian Agency of Science and Technology (CiT), the Technical University of Valencia (UPV), and the Departamento de Sistemas Informáticos y Computación (DSIC). RDP 2003 was scientifically sponsored by the European Association for Programming Languages and Systems (EAPLS), the European Association for Theoretical Computer Science (EATCS), and ERCIM.

After this first edition of RDP, a second and third have already been proposed. RDP 2004 will be held in Aachen, Germany, from 1-5 June 2004. RDP 2005 will be held in Nara, Japan.

Links:
RDP 2003: http://www.dsic.upv.es/~rdp03
RDP 2004: http://www-i2.informatik.rwth-aachen.de/RDP04
ELP group:
http://www.dsic.upv.es/users/elp/elp.html

Please contact:
Salvador Lucas, Universidad Politécnica de Valencia / SpaRCIM
E-mail: slucas@dsic.upv.es
http://www.dsic.upv.es/~slucas

CALL FOR PAPERS

SMI&SM 2004 — International Convention on Shapes and Solids 2004

Genova, Italy, 5-11 September 2004

The International Convention on Shapes and Solids will be held in Genova, the European Capital of Culture in 2004, and will bring together in a common forum the two major international conferences in 3D modeling: Shape Modeling International (SMI) and the ACM Symposium on Solid Modeling and Applications (SM).

The Convention is organized by the Institute of Applied Mathematics and Information Technology, Dept. of Genova, of the Italian National Research Council (CNR).

The SMI, SM and Convention Chairs will coordinate the common part of the two events (tutorials, keynote speakers, reception). However, the distinctive features of the two conferences will be maintained, and each will be organised separately with their own programme committee. The SMI&SM04 Convention is chaired by Bianca Falcidieno, and Prof. Tosiyasu L. Kunii has been invited to be the Honorary Chair.

Important dates:

Shape Modeling International (SMI)

- October 20, 2003 Abstracts due
- November 15, 2003 Full papers due
- December 15, 2004 Short presentations (posters) due
- January 15, 2004 Notice of acceptance and reviews
- February 15, 2004 Final camera-ready papers due

ACM Symposium on Solid Modeling and Applications (SM)

- October 1, 2003: Abstracts due
- November 1, 2003: Full papers due
- February 15, 2004: Notice of acceptance and reviews
- March 31, 2004: Final camera-ready papers and extended abstracts due

More information:
http://smism04.ge.imati.cnr.it/

CALL FOR PARTICIPATION

DSN-2004 — International Conference on Dependable Systems and Networks

Florence, Italy, 28 June-1 July 2004

DSN 2004 represents the continuation of over three decades of tradition in the field of dependable computing. The conference will have multiple tracks of refereed papers, including the Dependable Computing and Communications Symposium (DCCS) and the Performance and Dependability Symposium (PDS). There will also be Workshops, Tutorials, a Student Forum, and Fast Abstracts to highlight late-breaking research. DSN is the premier forum for researchers, practitioners, and users to learn and exchange information on the latest research results in dependable systems and networks.

Topics include:

- Analytical and Simulation Techniques for Performance and Dependability Assessment
- Architectures for Dependable Computer Systems
- Dependability Benchmarking
- Dependability of High-Speed Networks and Protocols
- Dependability Modeling and Prediction
- Dependability in VLSI
- E-commerce Dependability
- Fault Tolerance in Transaction Processing
- Fault Tolerance in Distributed & Real-Time Systems
- Fault Tolerance in Multimedia Systems
- Fault Tolerance in Mobile Systems
- Internet Dependability and Quality of Service
- Survivable and Intrusion-Tolerant Systems
- Measurement Techniques for Performance and Dependability Assessment
- Safety-Critical Systems
- Software Testing, Validation, and Verification
- Software Reliability
- Tools for Performance and Dependability Assessment

More information:
http://2003.dsn.org/

CALL FOR PAPERS

ETAPS 2004 – European Joint Conferences on Theory and Practice of Software

Barcelona, Spain
27 March - 4 April 2004

The European Joint Conferences on Theory and Practice of Software (ETAPS) is the primary European forum for academic and industrial researchers working on topics related to Software Science. It is a confederation of five main conferences, a number of satellite workshops and other events. ETAPS 2004 is the seventh joint conference in this series. The conference is organized by the Technical University of Catalonia (UPC).

Conferences:

- CC 2004: International Conference on Compiler Construction
- ESOP 2004: European Symposium on Programming
- FASE 2004: Fundamental Approaches to Software Engineering
- FOSSACS 2004: Foundations of Sofware Science and Computation Structures
- TACAS 2004: Tools and Algorithms for the Construction and Analysis of Systems

ETAPS main conferences accept two types of contributions: research papers and tool demonstration papers.

Important dates:

- October 17, 2003: Submission deadline for conferences and tutorials
- December 12, 2003: Notification of acceptance/rejection
- January 9, 2004: Camera-ready version due
- March 29-April 2, 2004: ETAPS 2004 main conferences
- March 27-April 4, 2004: ETAPS 2004 satellite events

More information:
http://www.lsi.upc.es/etaps04

ERCIM NEWS

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 over 9000 copies.

ERCIM News online edition is available at 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
ISSN 0926-4981

Director: Bernard Larrouturou
Central Editor:
Peter Kunz
peter.kunz@ercim.org
Local Editors:
AARIT: Karin Hraby
hraby@ocg.at
CCLRC: Martin Prime
M.J.Prime@rl.ac.uk
CRCIM: Michal Haindl
haindl@utia.cas.cz
CWI: Annette Kik
Annette.Kik@cwi.nl
CNR: Carol Peters
carol.peters@isti.cnr.it
FORTH: Yannis Pachoulakis
yannisp@ics.forth.gr
FhG: Michael Krapp
michael.krapp@scai.fraunhofer.de
FNR: Patrik Hitzelberger
hitzelbe@crpgl.lu
INRIA: Bernard Hidoine
bernard.hidoine@inria.fr
NTNU: Truls Gjestland
truls.gjestland@ime.ntnu.no
SARIT: Harry Rudin
hrudin@smile.ch
SICS: Kersti Hedman
kersti@sics.se
SRCIM: Gabriela Andrejkova
andrejk@kosice.upjs.sk
SZTAKI: Erzsébet Csuhaj-Varjú
csuhaj@sztaki.hu
Irish Universities Consortium:
Ray Walshe
ray@computing.dcu.ie
VTT: Pia-Maria Linden-Linna
pia-maria.linden-linna@vtt.fi

Free subscription

You can 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/

CWI – **Lex Schrijver** received the Dantzig prize for his complete work in optimization on Monday 18 August 2003 during the International Symposium on Mathematical Programming (ISMP) 2003 in Copenhagen. Furthermore, he and **Bert Gerards** both received the Fulkerson prize at the same conference. The Dantzig prize is the highest award in mathematical optimization, awarded by the Mathematical Programming Society and SIAM. Schrijver received the prize for 'deep and fundamental research contributions to discrete optimization (...)'. 'Characterized by insights that are both broad and deep, and by a continual pursuit of simplification and unity, Schrijver's work is scholarship at its best.' The Fulkerson prize is awarded for outstanding papers in the area of discrete mathematics by the MPS and AMS. Bert Gerards won his Fulkerson prize for an article with Geelen and Kapoor in the Journal of Combinatorial Theory (B); Schrijver for his article in the same magazine. Both the Dantzig and Fulkerson prize are awarded every three years.

SARIT – ETH Zurich (The Swiss Federal Institute of Technology in Zurich) announced on 18 September the creation of a new research center, the **Zurich Information Security Center, ZISC**. The center is a consortium with four members: the ETH and the IBM Zurich Research Laboratory (both affiliated with SARIT), Sun Microsystems Laboratories, and Credit Suisse, one of the largest Swiss banks. The creation of the center is a result of our growing dependence on information technology and the growing threat to this technology from computer viruses, worms, Trojan horses and the need for increased data security. Consortium members will pool their considerable knowledge and experience, doing both basic and applied research. The center will also provide educational opportunities in the security area for ETH students and practicing professionals.

Fraunhofer ICT-Group – "**Research Roadmap for a European e-Learning and Knowledge Management Infrastructure**" is the title of a study recently published by the Fraunhofer ICT group. The roadmap shows the way to the integration of knowledge management with e-learning applications in order to enable easy content production and to interconnect working and learning processes through a learning service infrastructure across Europe. The chosen approach offers interesting perspectives and new possibilities for learning organisations. The study shows how life-long learning and continuous training on the job will look like in Europe in the year 2010. It points out the relevant topics for future applications and research, incentives and motivation models as well as realistic business models for knowledge sharing and content production. Because of the intensive cooperation between e-learning and knowledge management experts of the Fraunhofer ICT group, this study provides a rich state-of-the-art impression manner how the fusion of e-learning and knowledge management will have a considerable impact on the optimisation and effectiveness for learning scenarios and content production embedded into future business processes. The study was funded by the European Commission. For more information, see:
http://www.iuk.fhg.de/studien_eng.html

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/

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/

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/

Czech Research Consortium
for Informatics and Mathematics
FI MU, Botanicka 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

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/

Fonds National de la Recherche
20, Montee de la Petrusse
L-2912 Luxembourg
Tel. +352 26 1925 31, Fax +352 26 1925 35
http:// www.fnr.lu

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/

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

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/

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/

Spanish Research Consortium for Informatics
and Mathematics
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

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/

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/

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

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/

Trinity College
Department of Computer Science,
Dublin 2, Ireland
Tel: +353 1 608 1765, Fax: 353 1 677 2204
http://www.cs.tcd.ie/ERCIM

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

If you wish to subscribe to ERCIM News ***free of charge*** *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.*

Send, fax or email this form to:

ERCIM NEWS
Domaine de Voluceau
Rocquencourt
BP 105
F-78153 Le Chesnay Cedex
Fax: +33 1 3963 5052
E-mail: office@ercim.org

Name:

Organisation/Company:

Address:

Post Code:

City:

Country

E-mail:

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/