

Final Project Report

Multimedia Understanding through Semantics,
Computation and Learning



MUSE



www.muscle-noe.org

MUSCLE is a Network of Excellence
co-funded by the European Commission

Multimedia Understanding through Semantics, Computation and LEarning



MUSCLE IS A NETWORK OF EXCELLENCE ON MULTIMEDIA UNDERSTANDING THROUGH SEMANTICS, COMPUTATION AND LEARNING, PARTIALLY FUNDED BY THE EUROPEAN COMMISSION IN THE SIXTH FRAMEWORK PROGRAMME FROM 2004 TO 2008.

THE MAIN OBJECTIVE OF THE MUSCLE NETWORK OF EXCELLENCE IS TO CONDUCT A JOINT PROGRAMME OF ACTIVITIES IN ORDER TO FOSTER CLOSE COLLABORATION AND INTEGRATION BETWEEN RESEARCH GROUPS IN THE FIELD OF MULTIMEDIA DATA MINING ON THE ONE HAND AND MACHINE LEARNING ON THE OTHER, THUS CREATING A NETWORK OF DIFFERENT RESEARCH COMMUNITIES AT A EUROPEAN LEVEL.

THIS BROCHURE PRESENTS THE ACTORS OF THE MUSCLE CONSORTIUM AS WELL AS THE ACTIVITIES AND ACHIEVEMENTS OF THE NETWORK OF EXCELLENCE OVER THE FOUR YEARS OF ITS EC-FUNDED LIFE-TIME.

Project N°: FP6-507752

Duration: 48 months.

Start Date: 1 March 2004

End Date: 29 February 2008

Overall network cost: € 14 094 579

European Commission Funding:

€ 6 900 000

MUSCLE home page:

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

Funding Agency:

European Commission FP6

Unit E2 : Knowledge Management

and Content Creation

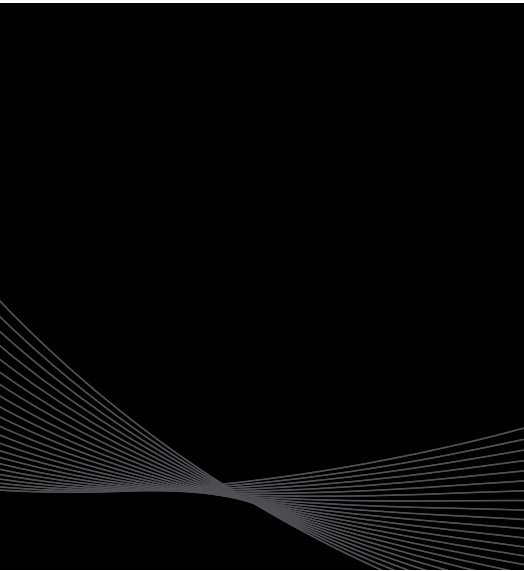
Executive Summary



MUSCLE Objectives

MUSCLE aims at creating and supporting a pan-European Network of Excellence to foster close collaboration between research groups in multimedia data-mining and machine learning in order to make breakthrough progress towards the following objectives:

- Harnessing the full potential of machine learning and cross-modal interaction for the (semi-) automatic generation of meta-data with high semantic content for multimedia documents.
- Applying machine learning for the creation of expressive, context-aware, self-learning, and human-centred interfaces that will be able to effectively assist users in the exploration of complex and rich multimedia content.
- Improving interoperability and exchangeability of heterogeneous and distributed (meta)data by enabling data descriptions of high semantic content (e.g. ontologies, MPEG7 and XML schemata) and inference schemes that can reason about these at the appropriate levels.



Major MUSCLE work areas

The MUSCLE Network of Excellence has actively brought together 45 leading research teams in the field of multimedia analysis. The actual integration of these 45 teams around a joint scientific agenda has been initiated successfully to address some of the research priorities described hereafter.

- Ensure durable integration and collaboration through the creation of a virtual lab that facilitates the easy and immediate access to people, data and ideas.
 - Through dissemination, training and industrial liaison, contribute to the distribution and uptake of the technology by relevant end-users such as industry, education, and the service sector. In particular, close interactions with other Integrated Projects and Networks of Excellence in this and related activity fields were implemented.
 - Through accomplishing the above, facilitate the broad and democratic access to information and knowledge for all European citizens (e.g. e-Education, enriched cultural heritage).
- **Moving from modelling to learning:** Exploiting the potential of machine learning and cross-modal interaction to semi-automatically generate robust meta-data with high semantic value for multimedia documents. In particular, MUSCLE researchers developed software tools and research strategies that enable users to move away from labour-intensive case-by-case modelling of individual applications, and allow them to take full advantage of generic adaptive and self-learning solutions that need minimal supervision.
 - **Improving interoperability through understanding:** Interoperability and exchangeability of heterogeneous and distributed meta-data can be improved by enabling data descriptions at high semantic levels and adding inference schemes. To this end MUSCLE researchers contributed to relevant international standards and protocols.
 - **Creation of expressive and adaptive interfaces:** In the same vein, improve the human-machine interface by exploring how machine learning can invigorate the creation of expressive, context-aware, and human-centred interfaces that will be able to effectively assist users in the exploration of complex and rich multimedia databases. With regard to these topics, MUSCLE research contributed to viability studies and proof-of-principle demonstrators.
- To stimulate cohesion, the Network sets itself two **grand challenges**. These are ambitious research projects that involve the whole spectrum of expertise represented within the consortium. As such they also require the collaboration of a large number of groups and therefore act as focal points for the consortium:
 - **Grand Challenge 1:** Natural high-level interaction with multimedia databases. In this vision it is possible to query a multimedia database at a high semantic level. This is an extremely challenging problem and will involve a wide range of techniques: natural language processing, interfacing technology, learning and inferencing, merging of different modalities, federation of complex meta-data, appropriate representation and interfaces, etc.
 - **Grand Challenge 2:** Detecting and interpreting humans and human behaviour in videos. Many important applications of multimedia data mining revolve around the detection and interpretation of human behaviour. Applications are legion: surveillance and intrusion detection, face recognition and registration of emotion or affect, automatic analysis of sports videos and films, etc. Success will depend heavily on the integration and interpretation of various modalities such as vision, audio and speech.

MUSCLE Achievements

Initiated in March 2004, the project has achieved the following goals over its four years:

- The MUSCLE portal (<http://www.muscle-noe.org>) federating institutes around the network's banner was established. In 2006, the MUSCLE portal has been redesigned to provide further collaborative tools to support the integration of research partners.
- Posters and flyers have been produced and distributed at international workshops and conferences where MUSCLE presented its latest scientific results. With the partners acting as relays for the promotion of MUSCLE, the network has gained a high visibility.
- The Internal Fellowship Programme providing advanced training to future researchers was implemented. This Fellowship Programme enables bright young scientists from all over the world to work on a challenging problem as fellows of leading European research centres.
- MUSCLE e-teams have been established to address core scientific challenges in the field of Multimedia understanding.
- In 2006, the MUSCLE Community launched several internal projects aimed at delivering highly visual results able to promote the technology to the research community, with a particular dissemination focus on Industry. The underlying objective of this showcasing activity is not only to take integration a step further, but also to stimulate the uptake of MUSCLE technology by the industrial stakeholders in the field. Details of these showcases can be found on the MUSCLE website at the URL: <http://www.muscle-noe.org/content/blogcategory/19/64/>
- MUSCLE participated in several professional exhibitions, namely:
 - CeBIT 2007 in Hanover Germany on 15-21 March 2007
 - Demo session and Practitioner day at ACM CIVR 07 in Amsterdam, The Netherlands, on 9-11 July 2007
 - International Broadcasting Convention (IBC) in Amsterdam, The Netherlands, on 7-11 September 2007
 - CeBIT Eurasia in Istanbul, Turkey, on 2-7 October 2007
 - Intersec Middle East 2008 in Dubai, UAE, on 13-15 January 2008

MUSCLE received a positive feedback from industrial contacts at these fairs.

- MUSCLE has most of its researchers participating in international workshops and conferences where they presented their latest scientific development and achievements in dedicated papers.

- The Joint MUSCLE-DELOS Summer School was organised in San Vincenzo, Italy on 12-17 June 2006. This event gathering experts and students was an excellent opportunity to promote advanced experiences and improve knowledge in the field of "Multimedia Digital Libraries, Machine learning and cross-modal technologies for access and retrieval".

- MUSCLE also participates in European Commission events, in order to liaise and collaborate with related projects and initiatives which have already received support.

Multimedia Understanding through Semantics, Computation and Learning

Contents

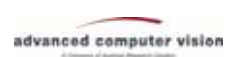
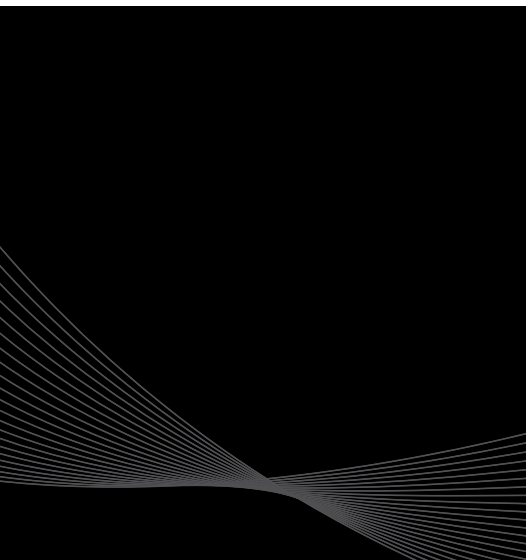
Executive Summary	4-6
Contents	7
MUSCLE Consortium	8-9
MUSCLE work packages	10-17
WP1: Coordination	11
WP2: Evaluation, Integration and Standards	12
WP3: Visual Content Indexing	13
WP4: Content Description for Audio, Speech and Text	14
WP5: Multimodal Processing and Interaction	15
WP6: Machine Learning and Computation Applied to Multimedia	16
WP7: Dissemination to Industry: Bringing the Grand Challenges to the End Users	17
Showcasing Activity	18-29
Showcase: BEYOND: The Virtual MUSCLE Experience	19
Showcase: Articulatory Talking Head Driven by Automatic Speech Recognition	20
Showcase: Augmented Assembly Using a Multimodal Interface	21
Showcase: Film Summarisation and Skimming Demonstrator	22
Showcase: Recognition Demonstrator	23
Showcase: Evaluation Showcase	24
Showcase: Object Categorisation	25
Showcase: Content-Based Copy Detection for Video and Still Images	26
Showcase: Dynamic Texture Detection in Video	27
Showcase: Real-Time Detector for Unusual Behaviour	28
Showcase: ACADI: Automatic Character (in Audiovisual Document) Indexing	29
E-Team Activity	30-33
Integration Activities	34-35
Fellowship Programme	34
Vocational Training	35
Spreading Excellence	36-37
Dissemination and Communication	36
Professional Exhibitions	37
Expertise of the MUSCLE Community	38-41
Finances	42

Multimedia Understanding through Semantics, Computation and Learning



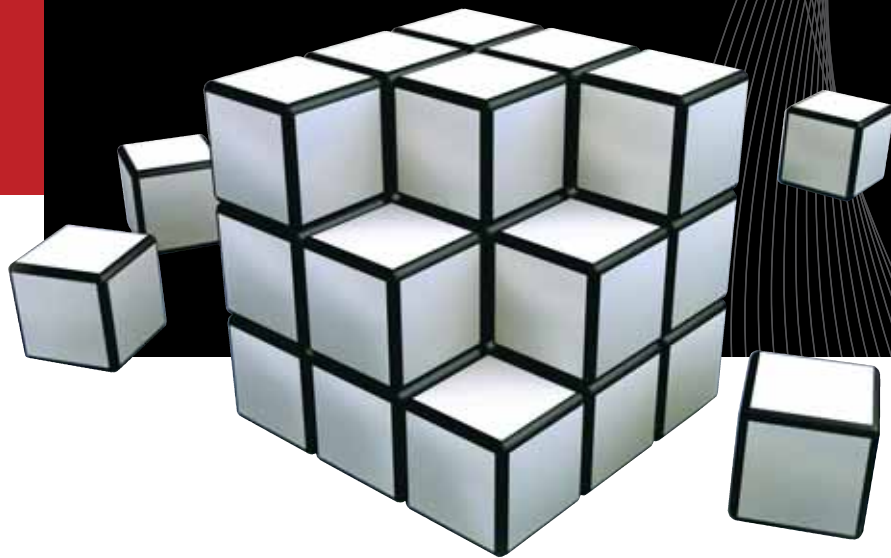
MUSCLE Consortium

GEIE ERCIM	ERCIM	France
Institut National de Recherche en Informatique et en Automatique	INRIA	France
Technische Universitaet Wien	TU Vienna-IFS/PRIP	Austria
Universiteit van Amsterdam	UVA	Netherlands
Institute of Communication and Computer Systems	ICCS	Greece
Telecommunication System Institute (Technical University Crete)	TSI-TUC	Greece
The Provost, Fellows and Scholars of the College of the Holy and Undivided	TCD	Ireland
Trinity of Queen Elizabeth near Dublin		
National University of Ireland, Dublin	NUID / UCD	Ireland
Bilkent Universitesi	BILKENT	Turkey
Stichting Centrum voor Wiskunde en Informatica	CWI	Netherlands
Kungliga Tekniska Högskolan	KTH	Sweden
Magyar Tudományos Akadémia Számítástechnikai és Automatizálási Kutató Intézeze	MTA SZTAKI	Hungary
Consiglio Nazionale delle Ricerche	CNR-ISTI	Italy
Commissariat a l'Energie Atomique	CEA	France
Advanced Computer Vision GMBH	ACV	Austria
ARC Seibersdorf Research GMBH	Seibersdorf Research	Austria
Technical Research Centre of Finland	VTT	Finland
Groupe des Ecoles des Telecommunications	GET	France
Université Paul Sabatier – Toulouse III	UPS – IRIT	France
University College London	UCL	United Kingdom
Technische Universitaet Graz	TUG	Austria
Universitat Politecnica de Catalunya	UPC	Spain
Academy of Sciences of the Czech Republic: Institute of Information Theory and Automation	UTIA	Czech Republic
Aristotle University of Thessaloniki	AUTH	Greece
Association pour la Recherche et le Développement des Méthodes et Processus Industriels	ARMINES	France
The University of Surrey	UNIS	United Kingdom
Ecole Nationale Supérieure de l'Electronique et de ses Applications	ENSEA	France
Forschungsverein E-Commerce Competence Center - EC3	EC3	Austria
Université Pierre et Marie Curie – Paris 6	UPMC	France
University of Ulster	UU	United Kingdom
Albert-Ludwigs-Universitaet Freiburg	UFR	Germany
Technion -Israel Institute of Technology	Technion-ML/MM	Israel
Institut für Bildverarbeitung und angewandte Informatik e.V	IBAI	Germany
Tel Aviv University	TAU-speech/visual	Israel
Foundation for Research and Technology - Hellas	FORTH	Greece
LTU Technologies	LTU	France
The Chancellor, Masters and Scholars of the University of Cambridge	UCAM-DENG	United Kingdom
Centre National de la Recherche Scientifique	CNRS	France



Multimedia Understanding through Semantics, Computation and Learning

MUSCLE Work Packages



The activities of the Network are organised around 7 work packages

WP1 : Coordination.

The coordinators assume the responsibility of the project management and optimally support the efficient and timely execution of the Joint Programme of Activities. The team also ensures the efficient communication flow among partners and work packages.

WP2: Evaluation, Integration and Standards. This WP centralises all the roles the Network plays as technical enabler: The benchmarking databases and protocols enable researchers to test and compare their algorithms while data and metadata standards enable groups to exchange and integrate their data. The involvement of WP7 is motivated by the fact that part of the WP7 task is to collect (and advise on the use of) computational software, enabling other groups in the consortium to gain more proficiency in the use of sophisticated statistical procedures.

WP3: Visual Content Indexing (collecting on image and video) aims at prepa-

ring numerical visual data for knowledge extraction and multimodality semantics association. Tasks will focus on visual content indexing and mining (still images and video).

WP4: Content Description for Audio, Speech and Text has the objective of preparing numerical multimedia data to knowledge extraction and multimodality semantics association, with tasks focused on each single modality of text, audio (music, sound, speech), as well as combinations thereof to extract semantic concepts.

WP5: Multimodal Processing and Interaction propose to research, design and build natural and efficient human-computer interfaces for performing multimedia information retrieval tasks that allow for negotiation (dialogue) between the user and the system.

WP6: Machine Learning and Computation applied to Multimedia includes development, extension and integration

of existing and new Machine Learning techniques specific to multimedia understanding.

WP7: Showcases and Dissemination towards Industry has been assigned the task of assembling showcases and generating interest from potential end users (in particular commercial ones). The principal objective of WP7: **Dissemination**. The showcases offer opportunities to intensify the Network's knowledge transfer and forge new research partnerships in preparation for the 7th Framework Programme. These new partnerships need to be pursued. In particular, the showcases should be used to attract more substantial involvement from commercial and industrial partners.

This WP aims to facilitate and coordinate:

1. Integration into system-level showcases of components supplied by individual teams.
2. Dissemination of the results to both internal and external "end-users".

Coordination



INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE



The coordination team ensures the scientific as well as the financial and administrative coordination of the Network of Excellence:

- Create the necessary conditions for successful work and results according to the contract
- Provide administrative support for the technical and scientific work performed by the work packages
- Manage all administrative and financial issues and organise the project reviews
- Take care of the quality assurance for the actions undertaken by the consortium

Scientific Coordination



Nozha Boujema
nozha.boujema@inria.fr
Head of the IMEDIA research group -
INRIA Rocquencourt

Administrative and Financial Coordination



Patricia Ho-Hune
Project Coordinator - ERCIM
patricia.ho-hune@ercim.org



Rémi Ronchaud
European Relations Manager - ERCIM
remi.ronchaud@ercim.org

Evaluation, Integration and Standards



Source: Fraunhofer IPK Berlin

The winner of the MUSCLE CIS Coin Classification Competition 2006, Marco Reisert (centre), together with the organisers: Michael Nölle (left) and Allan Hanbury.



WP Leader:

Allan Hanbury

WP Leader Institution Name:

Vienna University of Technology – PRIP.

hanbury@prip.tuwien.ac.at

Objectives

This work package had three main components: Evaluation, Standards and Computation.

The Evaluation objective involved the organisation of evaluation campaigns, the collection of test datasets and the manual annotation of some of the data in the test datasets.

The Standards objective was concentrated on the development of an Infrastructure for MultiMedia Metadata Management (4M) based on Semantic Web technology. The Computation objective promoted making available software for free use by network members and published resources on computational methods, such as software repositories and tutorial material.

Main Activities & achievements

- A repository of evaluation datasets and software was collected, and will continue to be available.
- Three workshops on Image and Video Retrieval Evaluation were organised in cooperation with the ImageCLEF campaign in 2005, 2006 and 2007. Highlights of the workshops were invited talks by well-known researchers involved in evaluation.

- The MUSCLE CIS Coin Classification Competition was organised in 2006 and 2007, with the task being to recognise coin types in a dataset of 10000 coins with as few false classifications as possible.
- The Infrastructure for MultiMedia Metadata Management (4M) was developed.

The repository, workshop proceedings and other information are at: <http://muscle.prip.tuwien.ac.at>

Future orientations and foreseen research challenges

The evaluation of image and video retrieval system continues to be a challenge. In particular, the adoption of these techniques by industry requires the possibility of objectively evaluating them, allowing the best to be chosen. Remaining research challenges include the creation of realistic use cases, large datasets with ground truth and suitable evaluation metrics.

Visual Content Indexing



Objectives

- Prepare numerical visual data to knowledge extraction and multimodality semantics association. Focus on visual content indexing and mining (still images and video).
- Develop novel, innovative methods and algorithms for processing visual information from still images and video sequences, that can enhance symbolic and semantic object/event detection and description to allow efficient indexing of visual repositories;
- Evaluate the new algorithms on standardised datasets;
- Foster integration through close e-team cooperation
- Disseminate the work through showcases

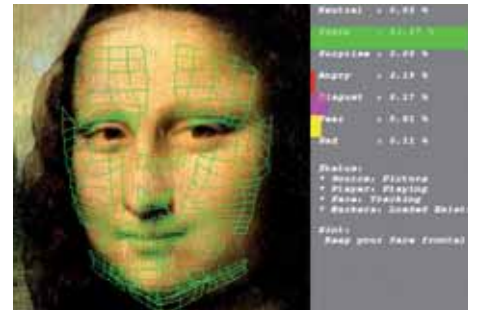
Main Activities & achievements

Low-level feature extraction for visual content description

We addressed the issue of visual feature extraction for modelling the visual appearance of images and their use for indexing the database. A variety of feature extraction algorithms were evaluated including colour, shape (2D and 3D) and texture analysis and modelling (2D and 3D) with multiscale consideration.

Visual Saliency

Within this task we investigated and developed saliency operators for images and video sequences, e.g. based on local descriptors with point-of-interest detection and description. We also worked on perceptually motivated high-level features in computer vision such as symmetry, anisotropy, regularity, complexity; symmetry of textures and shapes, on multi-resolution behaviour and also spatio-temporal based saliency features. One focus was on combined representation of salient image information through region and point of interest.



Symbolic Object Detection and Recognition

We developed methods for recognising arbitrarily complex individual objects under realistic real world viewing conditions (i.e., background clutter, illumination variations, occlusions, missing data, etc.). This task dealt with extraction of semantic information by using structural properties of images and patterns in terms of mid-level and high-level features. Particular prominent cases of study were:

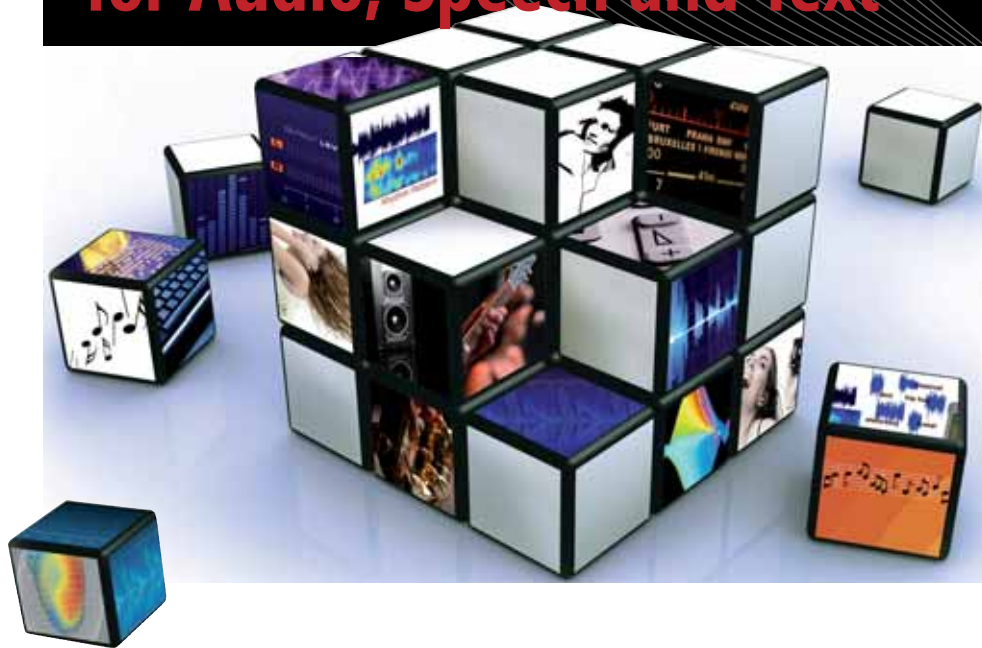
- Face detection and extraction, emotion detection, pedestrian motion in 3D
- Line structure (eg. roads in satellite images) extraction, representation and matching
- Events detection, scene classification.



WP Leader:
Nicu Sebe

WP Leader Institution Name:
University of Amsterdam
nicu@science.uva.nl

Content Description for Audio, Speech and Text



Objectives

The focus of this work package was on content extraction of audio, speech and text data for provide numerical descriptions of multimedia data to knowledge extraction and semantic association. The workpackage focused on each single modality of text and audio (music, sound, speech) and collaborated with WP2 on Evaluation, WP3 on Content Description for Image and Video and WP5 on Multi-Modal Integration for Multimedia Content, to drive the cross-modal integration of the developed multimedia knowledge extraction methods. A tight cooperation with WP6 ensured the direct usability of the methods with machine learning.



WP Leader:

Dr. Andreas Rauber.

WP Leader Institution Name:

Department of Software Technology
and Interactive Systems, Vienna
University of Technology, Austria
rauber@ifs.tuwien.ac.at

The e-Team on Semantics from Audio fostered the integration of the various approaches.

Main Activities & achievements

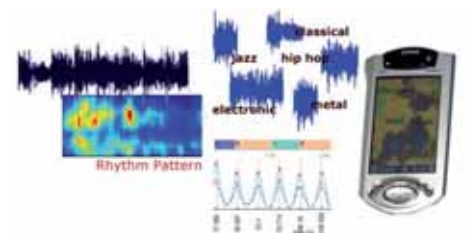
Audio processing

Reliable techniques for audio indexing and recognition (speech, music, instruments, sound) have been developed, and multiple sets of feature extractors and new classification techniques have been successfully evaluated on benchmark databases and in international evaluation campaigns, such as MIREX and NIST. Applications of these methods are in the recognition of sound classes (e.g. speech activity detector, speaker verification), music genre classification, audio search and retrieval, and enhanced visualisation of audio databases through clustering. Moreover, methods for audio segmentation, audio (source) separation and for the enhancement of the dynamics of recorded music have been developed. Furthermore, the problem of robustness of automatic speech recognition systems has been addressed, with a specific focus on European small or medium population languages. Software is provided via the MUSCLE web page as download or partially also as an automatic online audio feature extraction web service. Results of audio analysis and description methods have been shown in the Evaluation Showcase, with interactive demos of novel map-

based music browsing applications at public exhibitions and online in the MUSCLE 3D virtual world showcase.

Text and natural language processing

Although content-based information retrieval can be performed using image, sound or video samples, the preferred query mechanism of users is still that which uses text. We investigated the specific problems of natural language processing for describing multimedia content, including an analysis of the specific vocabulary used for describing multimedia data. Methods for analyzing and integrating textual information complementary to multimedia data, such as captions, scripts, lyrics, artist biographies, etc. have been developed, with the goal of contributing to and improving multimedia indexing and semantic knowledge extraction methods.



Audio indexing and segmentation methods enable classification and new audio browsing applications.

Multimodal Processing and Interaction



Objectives

This work package addresses research on the theory and applications of multimedia analysis approaches that improve robustness and performance through cross-modal integration. It also focuses on interaction with multimedia content, with special emphasis on multimodal interfaces for accessing multimedia information. Its general research objectives include several scientific and technological goals and can be grouped into the following categories:

1. State-of-the-Art Reviews and Book on Multimodal Processing and Interaction
2. Audio-Visual Speech Analysis, Synthesis, and Recognition
3. Multimodal Fusion and Integration for Multimedia Analysis and Recognition
4. Interfaces to Multimedia Content
5. Coordination of research, E-teams, Showcase demos, and Dissemination of results



Main Activities & Achievements:

1. Unimodal processing advances: (a) Advanced Computer Vision Approaches for Face detection, modelling, and visual feature extraction. (b) Nonlinear Speech Modelling for better audio & articulatory feature extraction.
2. Multimodal Feature Fusion and stochastic modelling with several data streams / several temporal rates / weakly synchronised multimodal data.
3. Multimodal dialogue detection and film structure analysis.
4. Real-time Audio-Visual Speech Recognition demonstrator.
5. Integrated multimodal Film Summarisation demonstrator.
6. Feature Fusion, Salient Feature & Event Detection in Films or Sports Videos.
7. Audio-Visual to Articulatory Speech Inversion.
8. Text Integration/Association with other Modalities, including Speech, Music, Image.
9. Multimodal interfaces to multimedia data.
10. Speech interfaces, Eye-tracking interfaces, Mobile interfaces.
11. Edited Book on: Multimodal Processing and Interaction: Audio, Video, Text, P. Maragos, A. Potamianos, P. Gros (Eds), Springer-Verlag, 2008.



WP Leaders:

Petros Maragos,
ICCS - Natl. Tech. Univ. Athens
maragos@cs.ntua.gr

Alexandros Potamianos,
TSI - Tech. Univ. of Crete
potam@telecom.tuc.gr

Machine Learning and Computation Applied to Multimedia



Objectives

The objective of this work package was the development and application of Machine Learning techniques for processing multimedia data. The techniques researched in this work package are as follows:

1. Supervised Learning
2. Kernel Methods
3. Dimension Reduction and Feature Transformation
4. Unsupervised Learning and Clustering
4. Semi-Supervised Learning & Active Learning

Main Activities & Achievements

Multimedia data processing presents particular challenges for machine learning algorithms due to the nature and context of the data. Activity in this workpackage has focused particularly on dimension reduction and feature selection techniques because multimedia data is typically of very high dimension. Because the potential applications for machine learning techniques in multimedia do not fit well with the standard supervised/unsupervised dichotomy in machine learning we have also developed novel strategies for semi-

supervised learning and active learning that work well for multimedia data.

Some of the research output from this workpackage will be published in March 2008 in a book by Springer:

Machine Learning Techniques for Multimedia: Case Studies on Organisation and Retrieval.

Springer Series: Cognitive Technologies, Cord, Matthieu; Cunningham, Pádraig (Eds.) 2008, 320 pages

Future orientations and foreseen research challenges

It is clear that the different contexts in which multimedia data exist creates specific challenges for Machine Learning techniques. The traditional distinctions between labelled and unlabelled data is less clear-cut with multimedia data. Often only some of the available data is labelled, data of one class only is labelled or large quantities of unlabelled data are available. These special characteristics require the development of new Learning techniques. Some progress in this direction has been made in the MUSCLE Network but many challenges still exist.

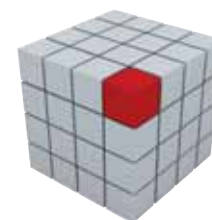


WP Leaders:

Simon Wilson,
School of Computer Science and
Statistics - Trinity College Dublin
swilson@tcd.ie

Pádraig Cunningham,
School of Computer Science and
Informatics - University College Dublin
padraig.cunningham@ucd.ie

Dissemination to Industry: Bringing the Grand Challenges to the End Users



Objectives

The principal objectives of WP7 were Research Integration and Dissemination. To achieve this aim WP7 provided a framework for the development of small internal research projects, the results of which will be destined essentially to showcasing activities. These small research groups served for both research integration through their collaborative work and dissemination through the showcase demos presented at several fairs, exhibitions, conferences and contests. To come up with different showcase demos, these groups shared their expertise with other institutes while mutually benefitting from each other's expertise. In this way a common research environment and research integration was achieved.

Main Activities & Achievements

To disseminate the MUSCLE results to the scientific community, the consortium participated in several events. MUSCLE presented highly visible interactive demonstrations at CeBIT Hannover, the largest IT event worldwide, which was held in March 2007.

In July 2007, two events were organised at CIVR'07 conference held in the Netherlands. The first event was a dedicated MUSCLE demo session in CIVR. Scientists who attended the conference had a chance to see the research topics and showcases of the MUSCLE community. The second event was the "Video Olympics" contest at which video-image retrieval algorithms of several institutions and companies (including MUSCLE's video retrieval showcase algorithm) competed with each other.

In August 2007, MUSCLE attended one of the world's most famous and important conventions on Broadcasting, IBC, which was held in Amsterdam. IBC has been very important for industrial dissemination of the project. MUSCLE had a stand at the "New Technology Campus" hall of the exhibition, where all of the demos were presented. Several representatives

from companies came to our stand and were interested in our new technologies. For many people, we were presenting the new technologies that other big companies will buy and use in the future. In October 2007, CeBIT-EURASIA was held in Istanbul. MUSCLE had a stand in the "FUTURE PARC" hall of the exhibition. We showcased all projects, particularly the "Real-Time Fire and Smoke Detection" demo which attracted much attention. On January 2008 we attended the INTER-SEC fair in Dubai. This fair is mostly about Security issues. Hence, Real-time Fire and Smoke Detection and Unusual Behaviour Detection were presented.



WP Leader:
A. Enis Çetin

WP Leader Institution Name:
Bilkent University
cetin@bilkent.edu.tr

Multimedia Understanding through Semantics, Computation and Learning

Showcasing Activity



Actively pooling research expertise and software resources, the MUSCLE network developed targeted showcasing research projects designed to deliver highly demonstrable results.

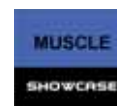
Below is a short list of MUSCLE on-going research destined to produce prototypes for highly visual public showcasing and demonstrations.

- Beyond : The Virtual MUSCLE Experience
- Articulatory talking head, driven by automatic speech recognition
- Augmented assembly using a multimodal interface
- Film Summarisation
- Audio-Visual Speech Recognition Showcase
- Evaluation Showcase
- Object categorisation showcase
- Content-Based Copy Detection for Videos and Still Images
- Dynamic Texture, driven by Detection in Video
- Unusual behaviour detection
- Automatic Character in Audiovisual Document Indexing (ACADI).

The complete list of Showcases and demonstration are freely accessible from the MUSCLE web site

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

BEYOND: The Virtual MUSCLE Experience



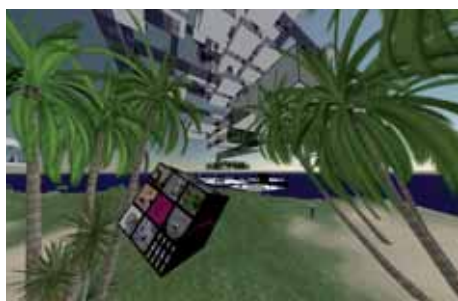
Objective

Generating, publishing and, most importantly, experiencing content in 3D virtual spaces is an emerging facility on the Internet with Second Life being the most prominent representative.

The "Virtual MUSCLE Experience" puts this forward by creating a platform that provides for enhanced sustainability and visibility of both the MUSCLE community (its members and work packages) and its research outcomes.

Moreover, it represents an environment for science communication by putting on display selected projects teaching the underlying scientific and technological building blocks.

"BEYOND: The Virtual MUSCLE Experience" will become the immersive one-stop information centre beyond the duration of the MUSCLE network ensuring continued sustainability of the MUSCLE achievements.



Its concept is based on the MediaSquare, a 3D Multimedia Environment where users are impersonated as avatars enabling them to browse and experience multimedia content by literally walking through it. Users may engage in conversations with other members of the community, exchange experiences or simply enjoy the featured content. On the one hand, such 3D virtual worlds address the aspect of social interaction by providing instruments to interact and to exchange experiences with other users that go beyond the possibilities of conventional text-based chat rooms. Especially, one's inherent presence in space and the awareness of others facilitate the initiation of social contacts. On the other hand, using 3D virtual worlds has the advantage of communicating via commonly accepted spatial metaphors. Beyond these aspects, "Virtual MUSCLE" allows the user to browse and experience multimedia content by literally walking through it. Users may engage in conversations with other users, exchange experiences as well as collectively explore and enjoy the featured content.

Research Challenges

Most virtual worlds are designed to be experienced in 3D, thus generating new challenges for users as well as designers.

Finding the ultimately usable 3D user interface is still a matter of research. For the specific problem of providing multimedia content in 3D environments, particular solutions for the case of Virtual MUSCLE have been found.

The combination of algorithms from the area of artificial intelligence with state-of-the-art 3D virtual environments creates an intuitive interface that provides access to manually as well as automatically structured multimedia data. Multimedia data are thereby presented by taking advantage of spatial metaphors, which are provided by the use of clustering and other machine learning methods.

Potential applications

These immersive 3D environments are suitable for scientific result dissemination, specifically virtual conferences. They are also applicable as learning environments for complex topics, enabling real-time communication, building of communities, and efficient consumption of multimedia content.

"BEYOND: The Virtual MUSCLE Experience" is accessible through:

<http://slurl.com/secondlife/MUSCLE/125/128/22/>

Showcase Leading Institution

EC3

<http://ispaces.ec3.at/muscle.php>

Contact person

Johannes Sperlhofer

johannes.sperlhofer@ec3.at

Office Phone: 0043 1 5227171

Office Fax: 0043 1 522717171

Additional Partner Involved

**TU Vienna, IFS
Andreas Rauber**

Articulatory Talking Head Driven by Automatic Speech Recognition



Objective

The objective is to re-create in real-time the articulatory movements of a speaker with an articulatory talking head, using the speech signal only. This articulatory talking head aims at giving a realistic display of the movements of speech articu-

lators that are normally hidden, such as the tongue. The talking head will also display some cues about the physics of speech production and provide some visual prosody information.

Research challenges

Research challenges include speech recognition, voice activity detection, building of a static and dynamic model of the articulators, articulators display, speech prosody extraction, prosody display.



Showcase Leading Institution

INRIA, LORIA-PAROLE

Contact person

Jacques Feldmar

Feldmar@free.fr

Phone: +33 6 80 03 60 22

Office Fax: +33 3 83 27 83 19

Additional Partner Involved

KTH, Department of Speech, Music and Hearing

Name of Researchers

Jonas Beskow, Olov Engwall, Jacques Feldmar, Bjorn Granström, Benjamin Husson, Yves Laprie, Samer Al Moubayed Jonathan Ponroy, Preben Wik.

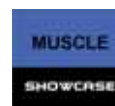
Potential applications

Articulation, prosody and physics of speech production are complementary pieces of information which provide highly valuable visual information for:

- Second language learning/teaching,
- Pronunciation training for hearing-impaired children,
- Augmented speech understanding for hard-of-hearing persons.



Augmented Assembly Using a Multimodal Interface



Objective

We developed a demonstration system for augmented assembly work with a multimodal user interface. This application demonstrates how to support the assembly of a complex product by augmenting the assembly instructions for the user. We used an assembly task of a wooden 3D puzzle box as an example application. The portable demonstration system consists of a PC, a camera and a display. The display can be a normal monitor, or a lightweight HMD (Head Mounted Display) with a small monitor in front of the user's eye. With a HMD we attach the camera to the user's headset, while with the other display types the camera can be mounted in any place with a good view of the assembly table.

With the augmented instructions, the user can assemble complex entities without browsing the instruction manual. We evaluated different input and output modalities during the development. We use audio and visual modalities for both input and output. The output from the system consists of voice feedback and visual augmentations on the display. In addition, the user can give both speech and gesture commands as input for the system.

The possible commands include control commands (e.g., moving from one task to the next) and information requests (e.g., look up information in the manual). The visual and speech modalities were integrated seamlessly on this application to clearly demonstrate synergies between the input and output modalities. System uses hybrid (marker-based and markerless) visual tracking method for calculating camera pose for correct 3D augmentations.

Research challenges

Real-time visual tracking, Automatic content creation, Automatic visual verification of the correct assembly.

Potential applications

A support system for a complex assembly task.



Showcase Leading Institution

VTT Technical Research Centre of Finland

Contact person

Sanni Siltanen

sanni.siltanen@vtt.fi

Office Phone: +358 20 722 6048

Office Fax: +358 20 722 7066

<http://www.vtt.fi/multimedia/>

Film Summarisation and Skimming Demonstrator

MUSCLE
SHOWCASE



Objective

As the amount of video data available (film, TV programmes, clips) in a personal recorder or computer are becoming increasingly large, intelligent algorithms for efficiently representing video data and presenting them to the user are becoming important. Video summarisation, film summarisation and film skimming are

increasingly popular research areas with immediate applications. The proposed demonstrator will have the ability to detect and decompose a video sequence into its audiovisual salient parts with the additional ability of breaking the synchrony of the audio/video streams and selectively present audio or video information in a summarised (time-compressed) form.

Research challenges

- Advance audio and visual saliency detectors
- Fuse the audio and video saliency detectors in order to identify perceptually-important events
- Design an interface that presents the salient audio and video information to the user in a compressed form, thus saving time with little or no loss of semantic information



Showcase Leading Institution

ICCS (Inst. Comm. Comp. Syst.)
NTUA (Natl. Tech. Univ. Athens)

Contact person

Prof. Petros Maragos

maragos@cs.ntua.gr

Office Phone: +30 210 772360

Office Fax: +30 210 7723397

Additional Partners Involved

ICCS-NTUA, Greece

K. Rapantzikos, G. Evangelopoulos,

Y. Avrithis, A. Zlatintsi

AUTH, Greece

C. Kotropoulos, P. Antonopoulos,

V. Moschou, N. Nikolaidis, I. Pitas

TSI-TUC, Greece

A. Potamianos, M. Perakakis

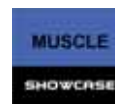
INRIA-Textmex, France

P. Gros

Potential applications

Key-frame detection, video summarisation, film summarisation, film skimming.

Recognition Demonstrator



Objective

One of the most promising approaches to improve the performance and extend the applicability of Automatic Speech Recognition (ASR) systems is to integrate visual information into the recognition process. Towards practically deployable Audio-Visual Automatic Speech Recognition (AV-ASR), we have built a proof-of-concept laptop-based AV-ASR prototype which: (i) uses consumer microphone and camera to capture the speaker; (ii) performs visual/audio feature extraction, as well as speech recognition on the laptop in real-time; (iii) is robust to failures of a single modality, such as visual occlusion of the speaker's face; and (iv) automatically adapts to changing acoustic noise levels.



Research challenges

- Adaptive multimodal fusion in unconstrained environments
- Robust real-time face tracking and visual feature extraction on commodity hardware

Potential applications

Robust ASR in adverse environments, such as car cabin and airplane cockpit.

Showcase Leading Institution

TSI-TUC

Contact person

Alexandros Potamianos

potam@telecom.tuc.gr

Office Phone: +30-28210-37221

Office Fax: +30-28210-37542

Additional Partners Involved

TSI-TUC

Manolis Perakakis

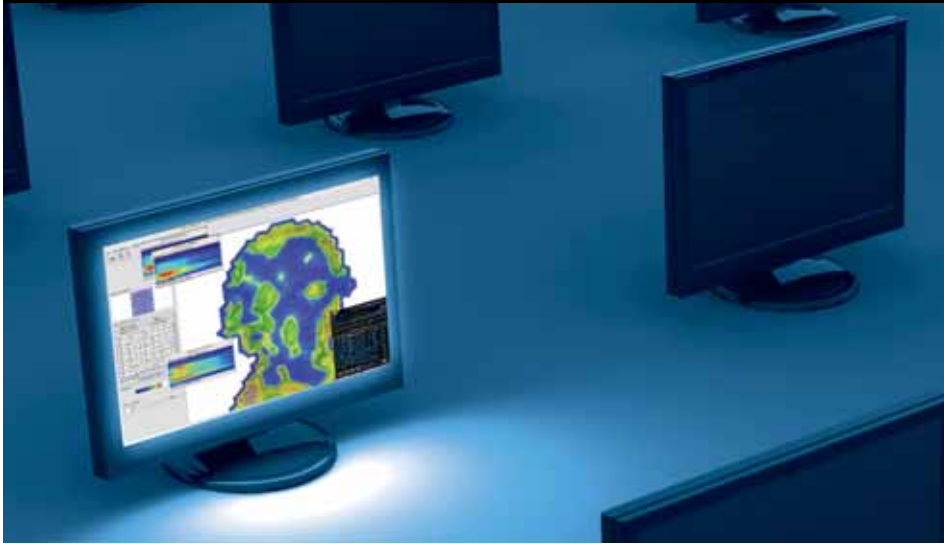
ICCS-NTUA

**George Papandreou
and Petros Maragos**

INRIA-TEXMEX

Guillaume Gravier and Patrick Gros

Evaluation Showcase



Showcase description and objective

This showcase is a demonstration of the wide range of semantic analysis and annotation capabilities developed within MUSCLE. The main goal of the showcase is to perform an objective evaluation of these capabilities. Several important activities of the showcase were:

- Organisation of VideOlympics which consisted of live evaluation events at the CIVR 2007 Conference on video retrieval and image retrieval
- Organisation of the 3rd MUSCLE / ImageCLEF Evaluation Workshop
- Organisation of the 2nd MUSCLE coin competition
- Assembly and annotation of an evaluation database
- An interactive Website allowing the user to experience a clustering of the video database by visualizing different modalities based on MUSCLE's content analysis methods

- Dissemination of the results at scientific and industrial events (e.g. CeBIT and IBC 2007)
- An evaluation web service allowing future evaluations of content analysis methods for the community
- Organisation of CIVR video copy detection showcase, which was the first benchmark initiative for exploring and comparing video copy detection technologies. A dedicated corpus (MUSCLE-VCD-2007) was created for this live event and shared with the community for further research on this topic.

Showcase Leading Institution

University of Amsterdam

Contact person

Nicu Sebe

nicu@science.uva.nl

Office Phone: +31-205257580

Office Fax: +31-205257490

Additional Partners Involved

University of Amsterdam

Cees Snoek, Marcel Worring

TU Vienna-PRIP

Allan Hanbury, Julian Stöttinger

TU Vienna-IFS

Andreas Rauber, Thomas Lidy,

Rudolf Mayer

INRIA-IMEDIA

Alexis Joly, Julien Law-To,

Nozha Boujema

IRIT

Philippe Joly

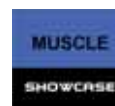
AUTH - AIIA

Constantine Kotropoulos,

Emmanouil Benetos



Object Categorisation



Objective

The objective of the showcase was to create an efficient web-based real-time object categorisation system.

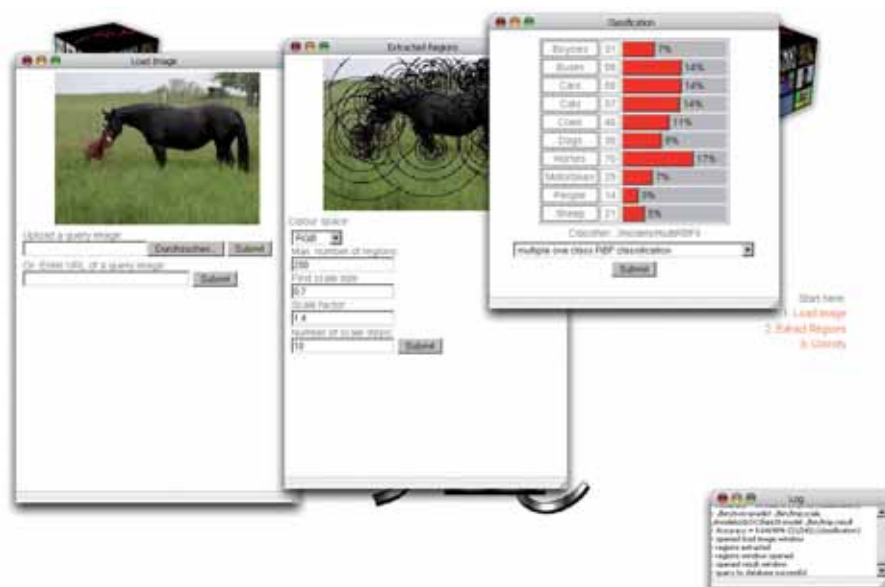
Research challenges

The main challenge was to get an acceptable system response time of a few seconds to categorise the objects in a provided image. One of the approaches investigated to attain this was the use of

colour interest points. As these take more information into account, they should be more salient. It is therefore possible to categorise an object based on fewer interest points and hence using fewer feature vectors, resulting in better performance.

Potential applications

Image search for particular objects, and automatic categorisation of images based on the objects they contain.



Showcase Leading Institution

TU VIENNA-PRIP

Contact person

Allan Hanbury

hanbury@prip.tuwien.ac.at

Office Phone: +43 1 58801 18359

Office Fax: +43 1 58801 18392

Additional Partners Involved

TU Vienna-PRIP

Julian Stöttinger

University of Amsterdam

Nicu Sebe, Theo Gevers

INRIA-IMEDIA

Nozha Boujemaa, Jaume Amores

Content-Based Copy Detection for Video and Still Images

MUSCLE
SHOWCASE



Objectives

Due to the increasing broadcasting of digital content (TV Channels, Web, Video Blogs, etc), finding copies in a large video and photo database has become a critical new issue and content-based Copy Detection presents an alternative to the watermarking approach to identify video sequences and still images. It provides a useful tool to control the copyrights of digital content on the web and on TV, and also a powerful tool to analyse and explore these data. The goal of this showcase is to introduce the motivations and

the issues of copy detection on the web and in TV streams and to present efficient methods for video and still image.

Research challenges

A crucial difficulty is the fundamental difference between a copy and the notion of similar image encountered in Content-Based Retrieval: a copy is not an identical or a near replicated video sequence but

rather a transformed video sequence. These photometric or geometric transformations (gamma and contrast transformations, overlay, shift, etc) can greatly modify the signal, and therefore a copy can in fact be visually less similar than other kinds of videos that might be considered similar. The other main challenge is the huge amount of data that requires the use of efficient indexing and search structures.

Showcase Leading Institution

INRIA - IMEDIA

Contact persons

Nozha Boujema

nozha.boujema@inria.fr

Alexis Joly

alexis.joly@inria.fr

Julien Law-to

julien.law-to@inria.fr

Additional Partners Involved

INA (Institut National de l'Audiovisuel)

D.Terougi, O. Buisson

(jointly with INRIA)

INRIA - VISTA team

I. Laptev



Example of two copies made from the same original video material but having a very low degree of similarity:



Dynamic Texture Detection in Video



Objective

Examples of dynamic textures include fire, smoke, clouds, trees, sky, sea and ocean waves etc. In this showcase, we present solutions to video-based fire, flame, smoke detection and swaying tree leaves and bush detection problems.

Description and Research Challenges

Researchers extensively studied 2-D textures and related problems in the field of image processing. However, there is very little research on dynamic texture detection in video. It is well-known that tree leaves in the wind, moving clouds etc., cause major problems in outdoor video motion detection systems. If one can initially identify bushes, trees, and clouds in a video, then such regions can be excluded from the search space or proper care can be taken in such regions, and this leads to robust moving object detection and identification systems in outdoor video. One can take advantage of the research in 2-D textures to model the spatial behaviour of a given dynamic texture. To be able to detect and segment dynamic textures in challenging real world applications, differences in dynamics must be also analysed. Two different approaches are studied in this showcase. In the first approach, dynamic textures are

classified as weak dynamic textures and they are analysed with standard optical flow algorithms relying on the brightness constancy assumption. However, self-occlusion, material diffusion, and other physical processes not obeying the brightness constancy assumption make such algorithms inappropriate for strong dynamic textures. An alternative to the brightness constancy assumption, the brightness conservation assumption enables the brightness of an image point to propagate to its neighbourhood and thus to model complex brightness changes. A non-regular optical flow calculation based on the brightness conservation assumption provides a better model for strong dynamic textures.

Potential applications

Video-based fire and smoke detection algorithms developed within the scope of this showcase project can be integrated to a CCTV system monitoring large and open spaces like atriums, shopping centres, aeroplane hangars, depots, ship compartments, forests, etc. The algorithm detects the presence of fire/smoke within seconds once they become visible. Smoke due to a forest fire is also detected with a similar algorithm developed in the course of this project. It is well-known that tree leaves and branches in the wind, moving

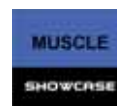


Fig 1: Smoke and flame detection in regular video.



Fig 2: Flame detection in infrared video.



Fig 3: Detection of swaying leaves in video.

clouds, etc., are a main source of false alarms in outdoor video analysis. One other application of the methods developed here is to be able to exclude regions with bushes, trees and clouds from the search space in a video. In this way, proper care can be taken in such regions.

Showcase Leading Institution

Contact person
Bilkent University
Enis Cetin
 cetin@bilkent.edu.tr

MTA SZTAKI
Dmitry Chetverikov
 csetverikov@sztaki.hu

Real-Time Detector for Unusual Behaviour

MUSCLE

SHOWCASE



Objective

In this showcase we present a distributed, network-based infrastructure for automatic detection of unusual behaviour in video flows. The framework is well-suited to integrate the different modules of the partners. Such modules are: disorderly motion (e.g. fight), fall detection, gate passing detection.

Showcase Leading Institution

MTA - SZTAKI

Contact person

Tamas Sziranyi

sziranyi@sztaki.hu

Tel: +36 1 2796106

Fax: +36 1 2796292

Additional Partners Involved

MTA-SZTAKI

**Laszlo Havasi, Levente Kovács,
Istvan Petras, Zoltan Szlavik**

**Universitat Politecnica de Catalunya
Montse Pardàs, Cristian Canton-
Ferrer**

**Advanced Computer Vision GmbH
Csaba Beleznai**

BILKENT University

**Yigithan Dedeoglu, B. Ugur Toreyin,
U Gudukbay, A. Enis Cetin**

Show case description

Visual surveillance and activity analysis has attained great interest in the field of computer vision research. Several algorithm libraries are available on-line. However their integration into a complex system is hindered by the heterogeneity of the implementation language, format, processing speed, etc. The aim of this work is to produce a flexible, transparent system for activity analysis. The actual implementation integrates diverse algorithms forming a test-bed for unusual activity detection. Various surveillance-related algorithms, such as human and body action, tracking and unusual motion activity algorithms are integrated into one system. The architecture according to the current trend and software tools is as flexible as possible. The modules can be distributed over the network; they are organised into a hierarchical structure.

Research challenges

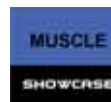
- Mathematical and algorithmic definition of "Behaviour" or "Event"
- All problems are mathematically badly posed.
- All algorithms are expected to run in real-time.

Potential applications

Security industry, Video surveillance, Public area surveillance, Sport events.



ACADI: Automatic Character (in Audiovisual Document) Indexing



Objective

We propose a system which can be used to describe and structure audiovisual documents with neither training nor corpus knowledge, and to visualise with an interface the principal interventions. It displays the most significant person list of the processed documents (news, TV games, variety programmes, film, etc.). A person is considered as significant if she/he speaks or appears on the screen during a minimum time lapse. Thanks to this person list, it is possible to listen and/or to view all interventions of each

character by clicking on the representation of the selected one. The system is based on the INRIA/Textmex face detection tool and the IRT speaker and costume segmentation tools.

With the interface, we can:

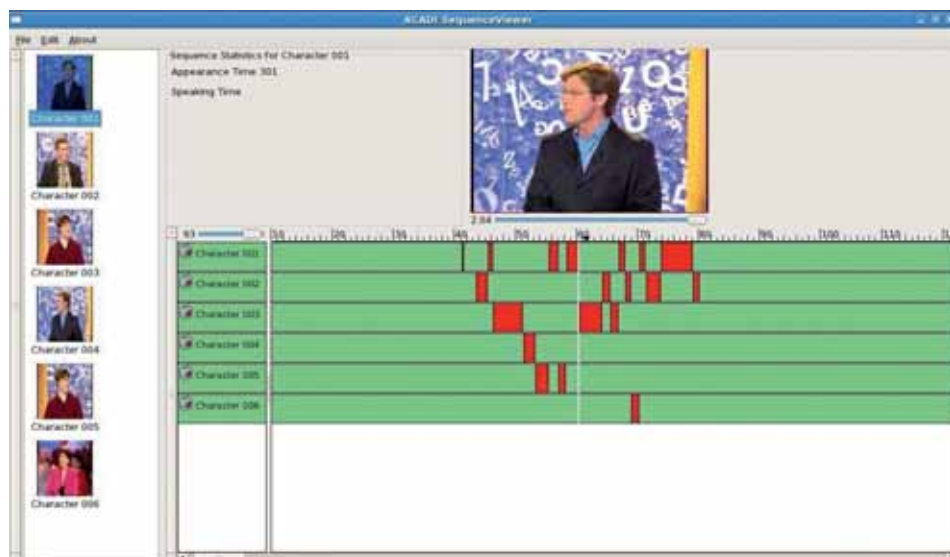
- watch (and/or listen) only those segments where the character of interest appears,
 - modify automatic segmentations,
 - visualise simultaneously several segmentations (speaker segmentation, costume segmentation, shot segmentation, etc.).
- This should lead to easy comparison and analysis, and thus enhance the fusion.

Research challenges

- Association of audio and video features for multimedia indexing, without *a priori* knowledge,
- Visualisation and annotation interface.

Potential applications

- Direct access to relevant information (via characters) in TV-enriched context,
- Measure of the audience of each character on a TV programme.



Showcase Leading Institution

UPS-IRIT

Contact person

Julien PINQUIER

pinquier@irit.fr

Office Phone: +33 5 61 55 74 34

Office Fax: +33 5 61 55 62 58

Additional Partners Involved

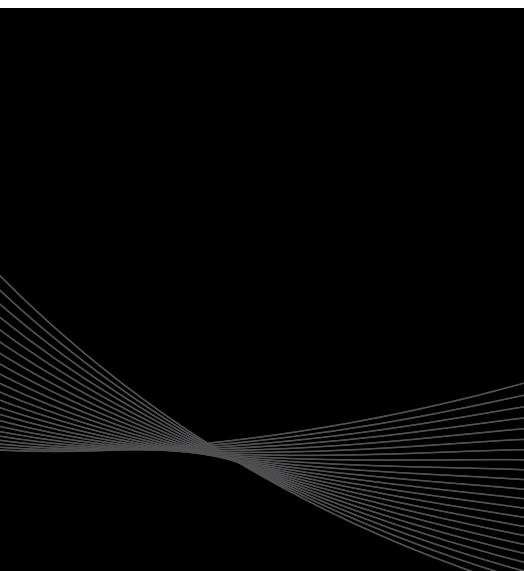
INRIA-Textmex

Patrick GROS, Ewa KIJAK

E-Team Activity



The purpose of E-Team activities is to stimulate scientific collaboration between different MUSCLE partners on research themes of common interest. This first innovation was introduced at the end of the first year and has proved very successful. As the E in E-teams suggests, collaboration is primarily conducted through electronic means. The Network continues to support active E-teams through grants for mobility and collaboration. Such teams were encouraged to organise focus meetings in the wake of international conferences or workshops (as most members of an E-Team share similar interests, they are very likely to attend the same international conferences).



E-Team list

■ ET1: Integration of structural and semantic models for multimedia metadata management

Participants: **CNR-ISTI**, Bilkent University, CEA LIST, IbaI,

This e-team investigates new techniques for extracting and integrating multimedia data and metadata from different modes (text, images, video). The e-team also undertakes research for a better understanding of the semantic models and their requirements for integration and dissemination of media.

■ ET2: Visual saliency

Participants: **INRIA Imedia**, UCL, UFR, INRIA/Vista, MTA SZTAKI, KTH, ENST, TU Graz, CWI, UvA

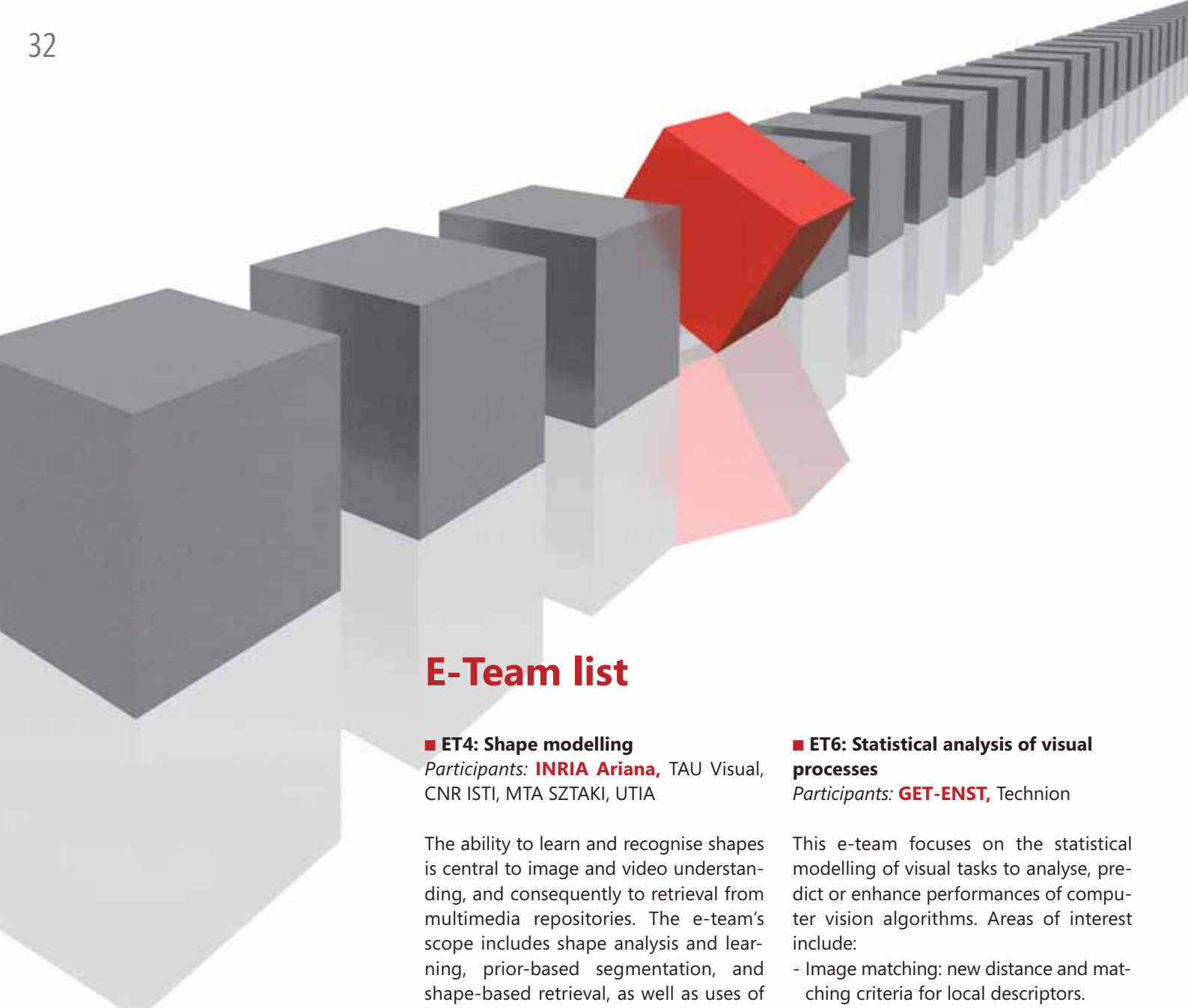
The e-team investigates and develops saliency operators for images and video sequences, e.g. based on local descriptors with point-of interest detection and description. The team works on perceptually motivated high-level features in computer vision such as symmetry, anisotropy, regularity, complexity, symmetry of textures and shapes, on multi-resolution behaviour and also spatio-temporal based saliency features.

■ ET3: Person detection and recognition, tracking and analysis

Participants: **UPC**, ACV, Bilkent University, MTA-SZTAKI, Technion-ML, UniS, UvA

This group focuses on:

- Face and facial features detection and tracking
- Body detection: Background learning techniques in both single and multi-camera environments.
- Body tracking: By means of models (e.g., templates, 3D models) and appropriate motion prediction.
- Body analysis



E-Team list

■ ET4: Shape modelling

Participants: **INRIA Ariana**, TAU Visual, CNR ISTI, MTA SZTAKI, UTIA

The ability to learn and recognise shapes is central to image and video understanding, and consequently to retrieval from multimedia repositories. The e-team's scope includes shape analysis and learning, prior-based segmentation, and shape-based retrieval, as well as uses of these techniques, their incorporation into broader models, etc., in order to aid in the solution of other problems.

■ ET5: Choosing Features for CBIR and Automated image annotation

Participants: **TU Vienna-PRIP**, ARMINES-CMM, CEA-LIST, KTH, TCD, UFR, UPC

The e-team works on:

1. Feature extraction and selection for CBIR and image annotation.
2. CBIR using Bayesian methods operating on well-chosen features.
3. Automated image annotation.

■ ET6: Statistical analysis of visual processes

Participants: **GET-ENST**, Technion

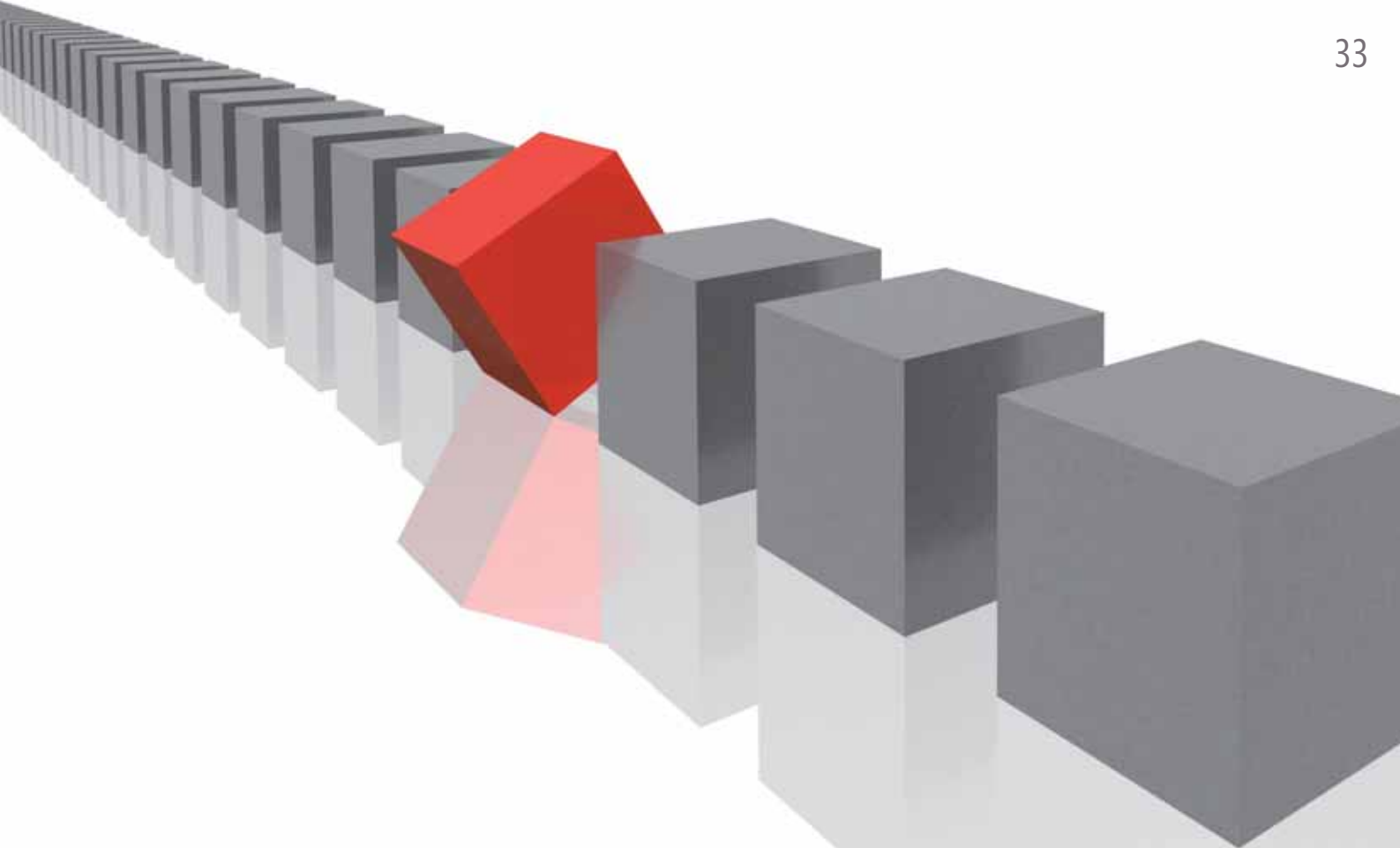
This e-team focuses on the statistical modelling of visual tasks to analyse, predict or enhance performances of computer vision algorithms. Areas of interest include:

- Image matching: new distance and matching criteria for local descriptors.
- Stochastic modelling of visual attention: global approach to the visual attention process, "E-saliency" algorithm using Bayesian networks
- Object recognition: statistical analysis of part-based recognition algorithms

■ ET7: Semantic from Audio

Participants: **TU VIENNA-IFS**, CNR-ISTI, AUTH, UCAM-DENG, EC3

This e-team collaborates on different methods for audio feature extraction and their application in both supervised classification as well as unsupervised organisation as a mean to access and explore audio holdings such as sound archives or, particularly, music. The e-team also investigates and develops special methods for extracting features sets in audio and musical signals.



■ ET8: Audio-Visual Speech Analysis & Recognition

Participants: **ICCS-NTUA**, TSI-TUC, IRIT, INRIA-Parole, INRIA-Textmex, AUTH, KTH

Research areas of this e-team include:

- Active-Appearance (and other Deformable) Models and Statistical Approaches for Face (or only mouth area) detection, modelling and feature extraction.
- Nonlinear Speech Modelling for better audio & articulatory feature extraction.
- Audio-Visual Feature Fusion
- Audio-visual to Articulatory Speech Inversion.

■ ET9: Multimodal Processing and Multimedia Understanding

Participants: **ICCS-NTUA**, INRIA-Textmex, AUTH, TU VIENNA-IFS, TUC, VTT, UCL, Bilkent University

Focus Themes:

- Theme 1: Stochastic modelling with several data streams / several temporal rates / weakly synchronized data and Multimodal Feature Fusion.
- Theme 2: Salient Feature & Event Detection in Films or Sports Videos.
- Theme 3: Integrate/Associate Text (and possibly NLP) with other Modalities, e.g. Speech, Music, Image.

■ ET10: Multimodal Interfaces

Participants: **TUC**, ICCS-NTUA, VTT, UCL, Bilkent University, AUTH, TU VIENNA-IFS

Some research areas include:

- multimodality,
- annotation of multimedia databases,
- interface efficiency,
- eye-tracking interfaces,
- speech interfaces and mobile interfaces.

■ ET 11: Dynamic Kernels

Participants: **IRIT-UPS**, ARMINES-CMM, CWI

Kernel methods have shown their power in static classification problems. The best example is definitely the support vectors machine (SVM) which is the most popular algorithm for supervised classification of static data. In many applications (such as speech, video, bioinformatics), however, one needs to classify dynamic sequences of observations. Classical kernel methods fail to provide satisfactory solutions in such applications. The goal of this e-team is to develop new kernel-based methods for the classification of dynamic sequences of observations

■ ET 12: Active and Semi-Supervised Learning

Participants: **INRIA-Imedia**, ENSEA, SZTAKI, TU VIENNA-IFS, UPMC

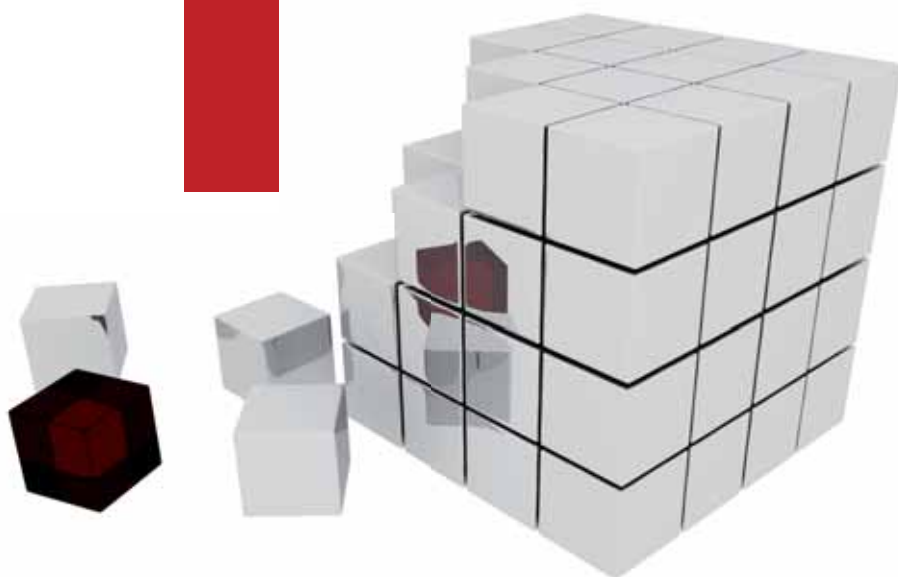
Multimedia databases are characterised by large amounts of rather cheap unlabelled data and small amounts of quite expensive knowledge (such as class labels). This expensive knowledge may sometimes be available a priori, but can usually be actively acquired by interaction with users. Our main focus is on the study of the active elicitation of such knowledge and on the joint use of these two sources of information – knowledge and unlabelled data.

■ ET 13: Unsupervised image segmentation

Participants: **ISTI-CNR**, TCD

Many measurement procedures are intended to extract information on physical systems by exploiting spectral diversity. This implies that the system components must be identifiable from their emission spectra. The aim of this e-team is to study this problem theoretically and develop methods to solve it in particular applications, when no prior assumption on the “spectral” features of the individual components is made.

Integration Activities



MUSCLE Fellowship Programme

The MUSCLE Fellowship Programme enables bright young scientists from all over the world to work on a challenging problem as fellows of leading European research centres. In addition, a MUSCLE fellowship helps widen and intensify the network of personal relations and understanding among scientists.

The programme offers MUSCLE fellows the opportunity to:

- Improve their knowledge of European research structures and networks
- Become familiarised with working conditions in leading European research centres
- Promote cross-fertilisation and cooperation, through the fellowships, between research groups working in similar areas in different laboratories.
- Collaborate closely with two hosting institutes

Each application was reviewed by senior scientists. Selection of the candidates was made taking into account the qualification of the applicant and the overlap of interest between the applicant and the hosting institution.

List of MUSCLE fellowships

Understanding and benchmarking of content-based image retrieval methods
Gareth LOY hosted by KTH

Ontology object construction
Svitlana ZINGER hosted by CEA

Video coding, image and video indexing and machine learning for representing multimedia content
Huicheng ZHENG hosted by TCD and FT

Semantic understanding of audio-visual data streams for Sports
Cyril HORY hosted by TCD and University of Surrey

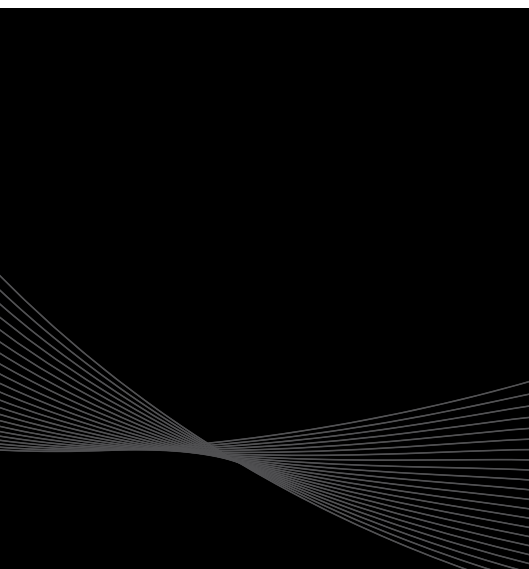
New features for automatic speech recognition process
Eduardo Sanchez SOTO hosted by TSI-TUC and INRIA Parole

Computer vision and pattern recognition
Eugen ZAHARESCU hosted by Bilkent

Multimedia metadata: bridging the gap from low-level media specific features to high-level domain-specific semantic terms
Suzanne LITTLE hosted by CNR and IBal

Monte Carlo Learning (MCL) methods for event recognition in multimodal sources
Onay URFALIOGLU hosted by CNR and Bilkent

Feature subset selection and multi-sensor fusion techniques for affect recognition
Mohammad Hossein SEDAAGHI hosted by AUTH



Vocational Training

In cooperation with the DELOS Network of Excellence, the MUSCLE community organised a summer school entirely dedicated to "Multimedia Digital Libraries, Machine learning and cross-modal technologies for access and retrieval". From 12 to 17 June 2006, this event gathered experts and students from around the world in San Vincenzo, Italy.

The aim of the school was to promote advanced experiences and to improve knowledge in the fields of information systems, knowledge representation, computer vision, audio and image processing, compression and storage, machine learning and information retrieval, multimedia data mining, cross-media analysis, user interfaces and interoperability.

On each of these topics, the School programme has planned highly qualified lectures given by international researchers, addressing theoretical approaches and practical solutions.

The school was opened to researchers and practitioners interested in audio-visual and multimedia digital libraries and in advanced solutions and technologies for their implementation. It also contributed to the definition of a common vision of the future ICT technologies in multimedia digital libraries.

Ultimately, this event was a significant opportunity to exchange experiences, to share common problems and to start new joint activities.

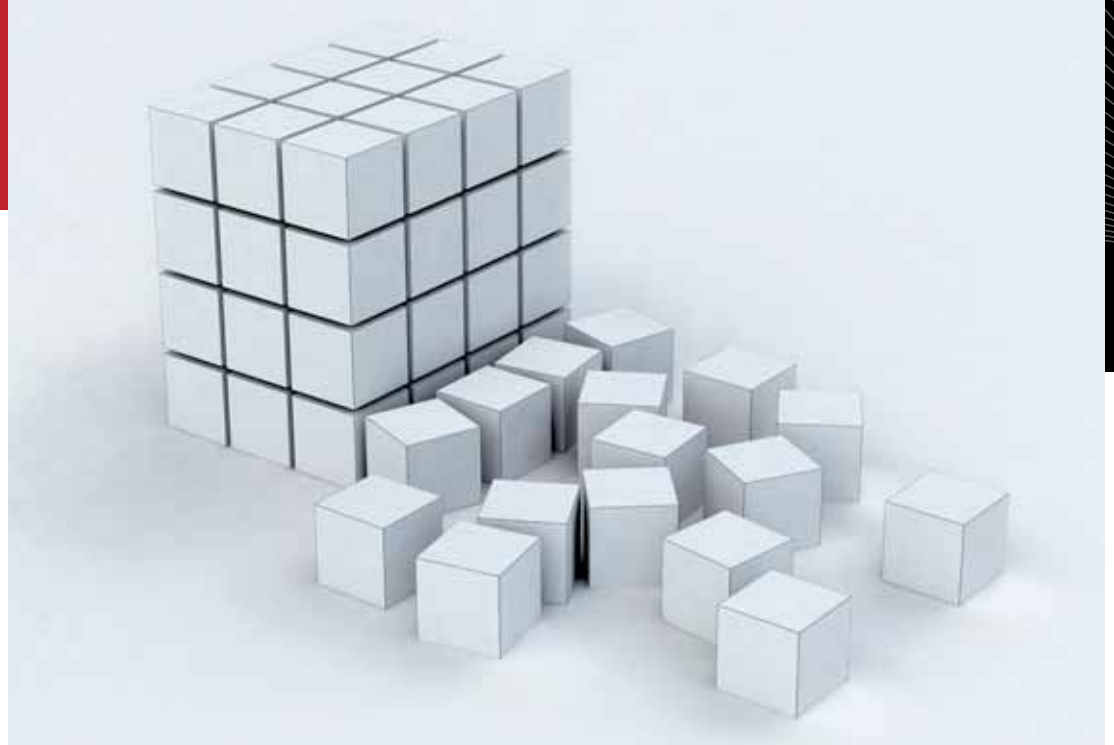
All the summer school lectures have been recorded and the corresponding videos are accessible at: http://videolectures.net/dmss06_san_vincenzo/

Again, we would like to thank the DELOS Network of Excellence for its contribution to this event.

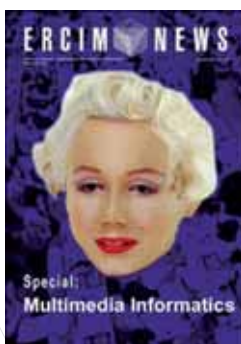


Multimedia Understanding through Semantics, Computation and Learning

Spreading Excellence



Dissemination and Communication

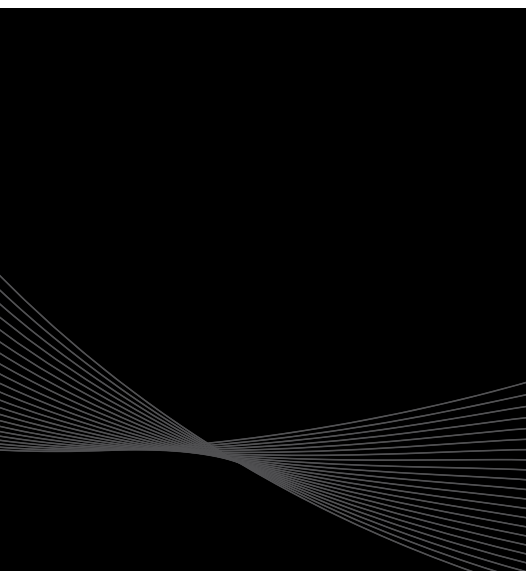


MUSCLE web site

The MUSCLE web site has public and private sections, allowing the visitors to get information about the network while hosting the scientific research community in the field.



Multiple dissemination and promotion material has been produced : logos for download, brochures, posters, articles published on AlphaGalileo and CORDIS, as well as ERCIM News N°62, July 2005.



Professional Exhibitions

MUSCLE-NoE participated to several events and fairs to disseminate MUSCLE showcase demonstrations and results to industry and a large public.

MUSCLE @CeBIT 2007

The Consortium presented its activities and showcases at CeBIT'07, the largest IT event worldwide, from 15-21 March 2007 in Hanover, Germany. Several partners of the Network presented their multimedia expertise live through highly visible interactive demonstrations. Live demos allowed visitors to try out and experiment each showcase.

Many visitors were attracted by the MUSCLE booth. People could try out the demos and were deeply impressed by the prototypes shown. Several companies as well as visitors involved in applied research showed their interest in the systems presented.

MUSCLE@IBC 2007

MUSCLE participated in the International Broadcasting Convention (IBC'07), which was held in Amsterdam, The Netherlands between 7-11 September 2007. The Exhibits showcased the latest technology and foremost business ideas in broadcasting and new media. The exhibition is internationally renowned for being run by the industry for the industry and offers unrivalled expertise and opportunities in the fields of content creation, management and delivery.

MUSCLE was located in the New Technology Campus, part of IBC2007, together with the NEM platform (Networked and Electronic Media), one of the European Industrial Initiatives. Live, working demonstrations of the technology develop-

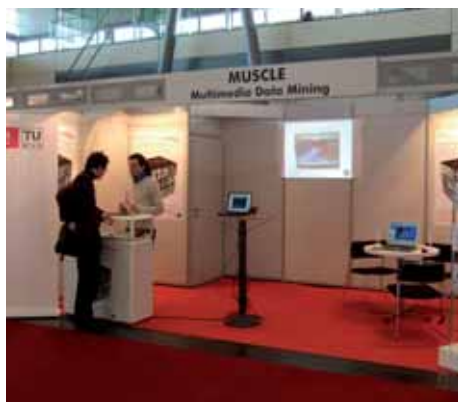
ped by MUSCLE members were staged and received much interests from visitors. This event offers great business to business and networking opportunities.

MUSCLE@CeBIT Eurasia 2007

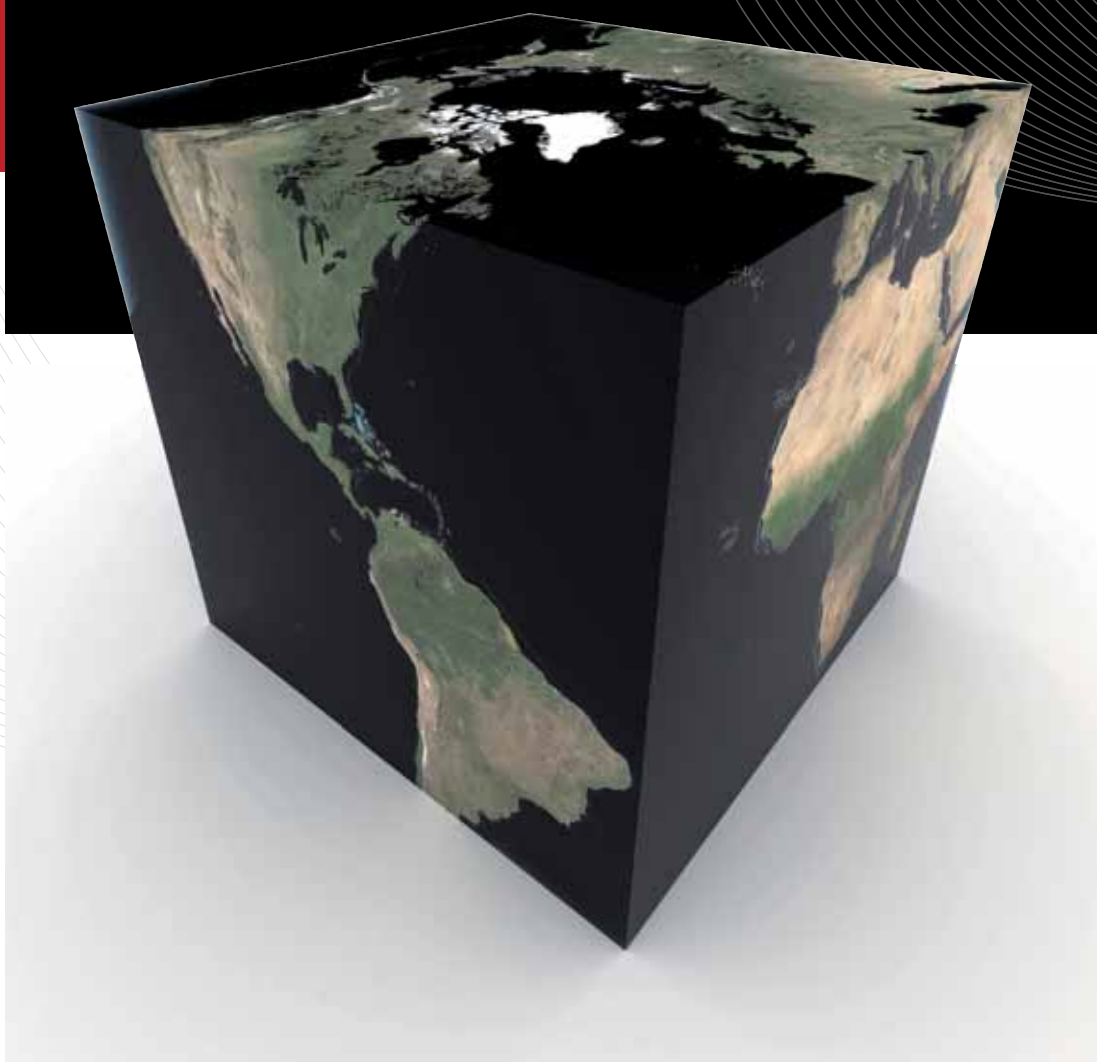
MUSCLE attended CeBIT Bilisim Eurasia which was held in Istanbul, Turkey, between 2-7 October 2007. All the MUSCLE showcase projects were presented and demos of Real-Time Fire Detection were given. The stand was visited by several industrial company representatives from Turkey. Moreover, reporters from local newspapers and internet sites were interested in the MUSCLE showcases and they published articles and interviews about the "Real-Time Fire and Smoke Detection" showcase. Besides industry, a large audience, especially university students, showed their interest in MUSCLE activities.

MUSCLE @Intersec 2008

MUSCLE participated in Intersec'08 which was held in Dubai, UAE between 13-15 January 2008. Researchers from "Real-Time Fire and Smoke detection" and "Unusual Behaviour Detection" showcases demonstrated their achievements. The event hosts four different areas of security in one fair; Commercial Security, Fire and Rescue, Safety and Health, Homeland Security and Policing. MUSCLE had a stand in the Commercial Security hall. This year more than 700 exhibitors from 50 different countries attended the fair.



Expertise of the MUSCLE Com



ERCIM

ERCIM, the European Research Consortium for Informatics and Mathematics, founded in 1989, is an EEIG (European Economic Interest Grouping) which provides effective management for joint projects between its members in association with external partners. ERCIM has its central office located in France, in Sophia Antipolis, the largest Science Park in Europe, and acts as a front-end to access the scientific expertise of its 19 European members. ERCIM members are all research organisations, or national consortia of research organisations and universities, with strong activity in I.T. research.

Institut National de Recherche en Informatique et en Automatique INRIA IMEDIA

The expertise of IMEDIA is in the field of content-based image indexing techniques and interactive search and retrieval methods for browsing large multimedia databases by content. We study "generic" image databases (e.g., the web) as well as "domain-specific" databases (e.g., facial imagery, medical imagery). IMEDIA focuses on modelling and structuring the feature space, automatic extraction of image signatures and 3D shape descriptors.

Community

INRIA ARIANA

Expertise in remote sensing image processing, in particular, shape and texture modelling:

1. Segmentation of entities from images, incorporating sophisticated prior information about the shape of the region in the image occupied by the entity.
2. Image modelling using probabilistic adaptive wavelet packet models of texture.
3. Image segmentation using hierarchical Markov random field approaches.

INRIA PAROLE

INRIA-Parole has good expertise in speech indexing and recognition, as well as natural language processing for multimedia content description. It is also involved in audio-visual speech analysis/recognition research for multimodal interaction.

INRIA TEXMEX

TEXMEX works on developing new techniques to exploit, manage, navigate and search the large collections of digital multimedia documents (TV, archives etc.). Our research is organised in three directions: - developing advanced algorithms of data analysis, description and indexing, - searching new techniques for linguistic information acquisition and use, - building new multimodal processing tools for audiovisual.

INRIA VISTA

Vista research work is concerned with various types of spatio-temporal images, (video images, meteorological satellite images, video-microscopy, X-ray images).

We are investigating methods to analyse dynamic scenes and dynamic phenomena, within image sequences. We usually rely on statistical approaches, resorting to: Markov models, Bayesian inference, robust estimation, particle filtering and machine learning.

CWI

CWI investigates mathematical methodologies to generate content-specific descriptions of images for efficient retrieval from large image databases. We are focusing on image databases for which additional context information is unavailable. Computer-assisted photo-ID of animals in image collections generated by biodiversity studies is one important application area. CWI also has expertise in stochastic geometry and statistical image analysis with particular interest in object recognition, segmentation and tracking.

KTH

Department of Speech, Music and Hearing KTH-Speech covers basic research in speech analysis, synthesis and recognition and in several language technology areas, including multimodal spoken dialogue systems and communication aids for persons with disabilities. The work is characterised by a search for understanding spoken human-human interaction and how this knowledge can be used in human-machine interaction.

Bilkent University

Research expertise:

1. We carry out research in multimedia, especially video databases. Our main focus is on data modelling and query processing in video databases. We developed a prototype video database system, called BilVideo, which provides an integrated support for queries on spatio-temporal, semantic and low-level on video data.
2. Dynamic Textures and Video Surveillance
3. Audio search capability

TU VIENNA-PRIP

The PRIP Group has expertise in image and video analysis, including image segmentation, object recognition, scene classification, colour and texture analysis, hierarchical representations, 3D reconstruction and object tracking. Retrieval systems for archaeological applications have been created, in particular for databases of images of historical coins and of 3D scans of pottery sherds.

TU VIENNA-IFS

The Information & Software Engineering Group (IFS) of the Department of Software Technology and Interactive Systems concentrates research on Information & Knowledge Engineering, Process Engineering, Software Engineering, Web Engineering, E-Learning, Security, Media Processing, Digital Preservation, Data Mining, and Information Retrieval. A core focus within the field of Information Retrieval is on the retrieval of audio documents based on cross-modal analysis of textual and acoustic content.

MTA SZTAKI

The DEVA group deals with the interpretation and organisation of information coming from signals of freely distributed or embedded sensors, mainly of video feeds coming from multimedia and sound-sensors as well as the evaluation, recognition and classification of events occurring in different locations and at different times. The research activities include texture and motion segmentation, surveillance systems for panoramic and multiple camera systems, digital film restoration etc.

University of Ulster

Our expertise is related to information retrieval systems and the focus within MUSCLE has been the extension of the SOPHIA search engine platform. SOPHIA is a novel search engine technology which organises a collection of documents into conceptual clusters which are indexed. These clusters are discovered using a proprietary algorithm which tries to identify concepts of high specificity and coverage of the given themes running throughout the corpus. Documents are then assigned to the cluster related to the most relevant theme.

CNR, ISTI "Alessandro Faedo"

ISTI's "Signals & Images" Lab is working on many aspects of multimedia content management, such as signal acquisition and processing, image understanding and artificial vision, high performance and distributed computing, real-time data collection and transmission. The general goal of the Lab is to increase both theoretical and applied knowledge in these fields.

Technische University Graz

The expertise of the Institute for Computer Graphics and Vision is in the field of object recognition and visual learning. We have developed methods for autonomous and

unsupervised learning of object detectors (persons, cars, faces, etc) in videos. Novel tracking methods based on a recently developed on-line boosting method have shown robust performance even under rough conditions. The developed real-time object recognition methods can deal with thousands of objects.

Universitat Politècnica de Catalunya

The research of the Image and Video Processing Group focuses on the areas of compression, analysis, indexing and representation of visual content. We specialise in basic tools for nonlinear filtering, mathematical morphology, segmentation, object tracking, face detection and recognition, emotion analysis and modelling of human activity, which have been the basis of applications related to content-based video coding, indexing and the creation of tables of content and contributions to standardisation processes.

ALBERT-LUDWIGS-UNIVERSITAET FREIBURG

Computer Science Department, Chair for Pattern Recognition and Image Processing We have extensive expertise in the following fields: Methods for position invariant pattern recognition in images and 3D volumes, content-based image retrieval, optimal design of classifiers, motion and depth estimation based on stereo and multiple camera views as well as image sequences, digital image filtering and restoration.

Academy of Sciences of the Czech Republic

The Institute of Information Theory and Automation has gained expertise mainly in unsupervised image segmentation, semi-supervised learning, multiple classifier systems, feature selection methods, probabilistic neural networks, bi-directional texture function compression, modelling and recognition, dynamic textures synthesis and detection, texture editing, image and video restoration, illumination invariants, range measurement segmentation, content-based image retrieval, and automatic acquisition of virtual and augmented reality models.

University of Amsterdam

The ISLA research group at UvA conducts research on the theory, practice, and implementation of multimedia information analysis, computer vision, and human-centred computing. The main research activities are in multimedia information processing, annotation of and retrieval from images and video data repositories, data space visualisation, theory of computer vision, colour

and colour invariants in computer vision, and multimedia data mining and multimodal human-computer interaction.

Aristotle University of Thessaloniki

Artificial Intelligence and Information Analysis Laboratory

Our research expertise covers Shot cut detection, speech segmentation, audiovisual scene boundary detection; audiovisual dialogue detection, audiovisual fingerprinting; audiovisual film genre recognition; audiovisual affect recognition; musical instrument classification, music genre classification; cross-modal probabilistic latent semantic indexing.

CEA

Laboratoire d'Intégration des Systèmes et des Technologies

The competencies used by CEA LIST for the MUSCLE project include image processing tools for content-based image retrieval; image ontology tools; image evaluation benchmarks; and image-to-text research, natural language processing tools and research in English, French, German, Italian, Arabic, Japanese, Chinese, Spanish, Russian, etc.; and multilingual document and video retrieval.

Advanced Computer Vision GmbH

Industry-oriented contract research and development in the area of digital image processing in all stages of the innovation process, from feasibility studies and concept creation to design, prototyping and ready-made solutions. Additionally, Advanced Computer Vision, in conjunction with 7 national and international industrial partners, is conducting a 7-year industrial research programme on computer vision.

ISRAEL INSTITUTE of TECHNOLOGY TECHNION-ML

The Machine Learning Group

Anytime Learning, Cost-sensitive Learning, Selective Learning, Speedup Learning, Feature Generation, Active Learning, Anytime Algorithms, Text categorization, Semantic Analysis. Computer Vision, Image Processing and Learning, Statistical Analysis of Recognition and Grouping Algorithms, Computational Attention Models and Algorithms

TECHNION-MM

The Multimedia group

Our areas of expertise are:

Object Tracking, Behaviour Analysis of Moving Objects, Graphical Models for Behaviour, Analysis, 3D Tracking of Articulated Human Body, Classification of Human Actions, Computer Vision, Image Understanding, Image Processing.

IBal

The institute conducts basic and applied research in computer vision, data mining, machine learning and for image databases. The institute has an excellent staff of highly-qualified researchers from various fields such as computer science, mathematics, electrical engineering and physics. It has modern equipment and excellent facilities for conducting scientific research such as Image Interpretation Data Mining and Image Data Base.

ICCS-NTUA

The Institute of Communication and Computer Systems, ICCS-NTUA, is a university research institute carrying out R&D activities on all aspects of computer and telecommunications systems and their applications. The Computer Vision, Speech Communication and Signal Processing Group's (CVSP) current activities span the following areas:

- 1) Computer Vision and Image Processing:
- 2) Computer Sound Processing and Speech Communication
- 3) Multimodal Signal Processing and Machine Learning: Audio-Visual speech recognition.

Telecommunications Systems Inst., Technical Univ. of Crete

TSI-TUC 's current research and development interests include the following research areas: Digital Communication Systems, Computer Networks, Wireless Communication Networks, Signal Processing for Physical Layer Communications, Speech and Language Processing, Multimedia Signal Processing and Automated Services over Communication Networks.

ARMINES

The Centre for Mathematical Morphology (CMM) is an image processing laboratory of the Paris School of Mines. Always motivated by industrial applications, the CMM has developed performing image processing tools, such as connected filters, that reduce the noise without blurring the interesting contours, and hierarchical segmentation that, combined with graph theory, lead to very efficient algorithms.

TEL-AVIV UNIVERSITY

Tel-Aviv University - Speech

Robust discriminative speech recognition and support vector machine (SVM)-type training; Fast and robust speaker recognition in large audio archives using indexing methods; Separation and segmentation of audio signals using single and multi-chan-

nel observations; Blind separation of superimposed images and video with relative spatial shifts.

Tel-Aviv University - Computer Vision

TAU-VISUAL specialises in the extraction and processing of visual information from video and still images. Typical application domains are medical image databases, image libraries, video surveillance, and environmental monitoring. Recent accomplishments include mutual image segmentation in multiple-sensor environments and smoke extraction in video sequences with moving camera.

Austrian Research Centers GmbH

The teams at Austrian Research Centers GmbH have dedicated their work to the innovation of smart real-time systems: all-in-one solutions for systems with IT-based intelligence, which support human users in many areas of life and various fields of applications. Our core competence is dependable signal, image- and data processing on chip, in embedded systems and on computer systems; simulation techniques and measurement & test engineering.

FORTH

FORTH is a non-profit research organisation. The Institute of Computer Science (ICS) is one of the seven Institutes constituting FORTH. FORTH-ICS participates in MUSCLE with its Computational Vision and Robotics Lab (CVRL) which has been and is currently particularly active in the topics of image description and retrieval, video analysis, motion extraction and scene structure estimation.

VTT

VTT Technical Research Centre of Finland is the biggest contract research organisation in Northern Europe. The Virtual Models and Interfaces knowledge centre of VTT represents special expertise in areas of Virtual and Augmented Reality, and associated technologies such as vision-based tracking and physical browsing, as well as mobile platforms and applications.

GET-ENST

The Groupe des Ecoles des Télécommunications (GET) is made up of three major French Graduate Schools in the field of Information Technology, the first among which is ENST - Ecole Nationale Supérieure des Télécommunications. The Multimedia Group at ENST is a federation of teams working on multimedia. The group's activity focuses on research and development in the field of multimedia technologies with the goal of facilitating the use of multimedia content in every day life.

LTU Technologies

LTU Technologies is a worldwide editor of image search, classification and filtering software. LTU Technologies' expertise in multimedia search engines is renowned world-wide in particular since it developed an innovative image search system with French search engine Exalead. The company plans to further the development of technologies that are scalable and fit for large-scale deployment.

University of Surrey

The MUSCLE-related research at the University of Surrey was carried out in the Centre for Vision, Speech and Signal Processing (CVSSP). Within CVSSP, the Multimedia Signal Processing & Interpretation Group clusters research interests in signal processing and sensory data analysis. The primary aim of our endeavour is to develop techniques that enable machines to understand and act on the information content of the data (e.g. speech recognition).

The Chancellor, Masters and Scholars of the University of Cambridge

Research expertise:

1. Automated Audio analysis tools: Bayesian computational methods, automated polyphonic music transcription, blind source separation, background noise reduction.
2. Automated video and image analysis complex wavelet methods, feature detection, object recognition, object detection.

ENSEA

Equipe Traitements des Images et du Signal

ETIS is a joint research unit of the ENSEA (Ecole Nationale Supérieure de l'Electronique et de ses Applications), the University of Cergy Pontoise, and the CNRS

Theoretical expertise includes: kernel approach applied to vectors, bags of features and more recently on graphs, fuzzy segmentation and graph matching, - active learning, long-term learning.

CNRS

The Llacon (Laboratoire Langage, Langues et Cultures d'Afrique Noire) has developed several multimedia search engines and resources in the domain of linguistics focusing on African and Arabic languages. Various tools have been made to enable the automatic morphosyntactic analysis and the interactive annotation of Arabic. The analysis and functionality of the system are related to: Vowelization, Lemmatization, Tagging, Segmentation in nominal and verbal chains, etc.

Université Paul Sabatier - Toulouse III

Institut de Recherche en Informatique
The main studies of the team SAMoVA (Structuring, Analysis, Modelling of Video-Audio documents) deal with the correlated extraction of the audio and video information and its exploitation to propose an intelligent access. This research implies studies on signal processing, on modelling and learning and also on content structuring, Signal Processing, Learning and Modelling, and Audiovisual Content Structuring.

E-Commerce Competence Center

EC3, in particular the iSpaces Research Group, has long standing experience in information visualisation and data mining. In the context of information visualisation, the area of immersive virtual worlds and game engines has been used extensively for simulating complex relationships in an immersive three dimensional space. As an example, consider the "MediaSquare" a synthetic 3D multimedia environment, which enables users, impersonated as avatars, to browse and experience multimedia content by literally walking through it.

Université Pierre et Marie Curie-Paris 6

The LIP6 Machine learning team at UPMC is specialised in the application of Machine Learning to multimedia retrieval: text, image and video. The team expertise includes content-based image and video retrieval, interactive systems, online relevance feedback strategies for web-based applications and artwork contexts. The team's expertise in machine learning definitively concerns Support Vector Machines and Kernel-based algorithms for classification and information retrieval.

University College Dublin

The Machine Learning Group in UCD has extensive experience in machine learning techniques for processing high dimension data. It has experience in dimension reduction techniques and in supervised machine learning techniques such as support vector machines and lazy learning techniques. Within Muscle the UCD group has worked on text and audio data.

University College London

The vision group at the UCL Adastral Park Campus carries out research on Computer Vision mainly with application to content-based Image Retrieval. The work has concentrated upon understanding saliency in human vision. Understanding what is important and most informative in visual material is central to the problem of categorisation and retrieval.

Finances



The MUSCLE Network of Excellence has been allocated a budget of € 6 900 000 by the European Commission to fulfil the objectives of its work programme.

This grant was distributed over the four-year period of the project among all the partners according to their activities and commitment in the network.

The expenditures of the project budget can be divided into the following activities:

- Integration activities
- Dissemination activities
- JPA grant to Scientists
- Scientific coordination
- Administrative coordination

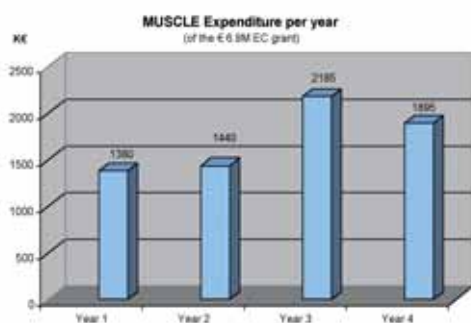
beginning of the project. However, the budget breakdown per activity was established each year in order to adjust and match the network's needs and evolution.

The budget for Integration activities consisted of several items, including:

- Fellowship programme for young scientists
- Mobility programme for researcher's visits
- E-team activities
- Showcasing activity
- Participation in professional exhibitions

The total estimated costs for the full duration of the project were evaluated at more than € 14 million, while the EC contribution did not exceed € 6.9 million.

The financial plans for the full duration of the project were established at the



Authors:

Nozha Boujemaa (INRIA)

Patricia Ho-Hune (ERCIM)

Rémi Ronchaud (ERCIM)

With the contribution of the MUSCLE Steering Committee and all MUSCLE Partners

Designed and printed by:

dzuf communication

2323 Chemin de St Bernard - Space 6 - BP 255 - Parc de Sophia Antipolis - 06227 Vallauris cedex (France)

Photos: Silke Heyer, Aleksey Bakaleev, Gregor Buir, Viktor Gmyria, Brock Meeler, Wellford Tiller, Kirsty Pargeter, Christopher Dodge, Julien Bastide, Franck Boston, Alex Slobodkin, Konstantinos Kokkinis, Rafal Zdeb, Denis Vorobyev, Mustafa Deliormanli, Eva Serrabassa, Sergey Ostroukh



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

February 2008