



# MUSCLE

## Network of Excellence

Multimedia Understanding through Semantics, Computation and Learning

Project no. FP6-507752

Bi-monthly progress report # 22: Sept-Oct 2007

Due date of deliverable: 15.12.2007

Actual submission date: 31.01.2008

Start date of project: 1 March 2004

Duration: 48 months

Name of responsible editor(s): Patricia Ho-Hune (ERCIM)

version 1.0

<b>Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)</b>		
<b>Dissemination Level</b>		
<b>PU</b>	Public	✓
<b>PP</b>	Restricted to other programme participants (including the Commission Services)	
<b>RE</b>	Restricted to a group specified by the consortium (including the Commission Services)	
<b>CO</b>	Confidential, only for members of the consortium (including the Commission Services)	

## **Contents**

<b>1 OVERVIEW ACTIVITIES IN WP1</b>	<b>3</b>
<b>2 OVERVIEW ACTIVITIES IN WP2</b>	<b>3</b>
<b>3 OVERVIEW ACTIVITIES IN WP3</b>	<b>6</b>
<b>4 OVERVIEW ACTIVITIES IN WP4</b>	<b>17</b>
<b>5 OVERVIEW ACTIVITIES IN WP5</b>	<b>21</b>
<b>6 OVERVIEW ACTIVITIES IN WP6</b>	<b>29</b>
<b>7 OVERVIEW ACTIVITIES IN WP7</b>	<b>34</b>
<b>8 MUSCLE EFFORT TABLE</b>	<b>39</b>

# 1 Overview activities in WP1

## 1.1 General scientific and administrative coordination

### Activities & Achievements

- Administrative and financial coordination of the network
- Organisation of regular audio-conferences.
- E-Teams and showcase extension monitoring and grant validation
- Modifications of MUSCLE financial cost statements and Audit certificates for period 3 requested by the Commission's service.
- MUSCLE Annex I including JPA4 was modified according to the project officer's request; Version 4 sent on 7 September 2007
- Reimbursement of MUSCLE integration expenses (fellowship, sponsorship)
- Preparation of MUSCLE participation to the CeBIT Eurasia exhibition in Istanbul, Turkey in October ([www.cebitbilisim.com](http://www.cebitbilisim.com)).
- Preparation of MUSCLE participation in Intersec'08 exhibition in Dubai, UAE in January 2008.
- Preparation of the last MUSCLE Plenary meeting to be held in November 2007 in Paris.
- Preparation of the MUSCLE conference (final event) to be held in February 2008.
- Preliminary work for preparation of the MUSCLE brochure to be published before the end of the project.
- Maintenance of MUSCLE website by ERCIM
- Cooperation with related European projects:
  - Chorus meeting on 10 October 2007 in Geneva
  - Contacts with PASCAL Network of Excellence
  - Liaison with VITALAS Integrated Project

# 2 Overview activities in WP2

## 2.1 Contribution by Bilkent University

### Researchers involved

Ugur Gudukbay, Ozgur Ulusoy, Muhammet Bastan, Hayati Cam

### Activities & Achievements

We continued our collaboration with ISTI-CNR about "Integration of structural and semantic models for multimedia metadata management" which is an e-team under workpackage WP2. Within this context, we are currently working on the implementation of an MPEG-7 Feature Extractor and Query Processing Modules for BilVideo Video Database System. Currently, we work on the extraction of low-level features. This will be used for querying videos based on MPEG-7 features. We use an XML-based format for storing MPEG-7 features.

## **2.2 Contribution by ISTI-CNR**

### **Researchers involved**

Patrizia Asirelli, Sara Colantonio, Massimo Martinelli, Ovidio Salvetti, Marco Tampucci

### **Activities & Achievements**

The work focused on further developing the 4M Infrastructure. The Database Unit has been extended in order to develop the queries for retrieving the features extracted from an image region. The unit extension regarded also the development of XUpdate query to increase the metadata management efficiency. Now the administrator users can completely manage the semantic annotation thanks to the possibility to modify and delete them (when the last annotation for an image is deleted even the RDF document is deleted). The administrator users can also delete the multimedia documents stored into the 4M infrastructure. When a multimedia document is deleted even the XML document which contains the MPEG-7 features is deleted and, if the document is an image and if there are some semantic annotations for this, even the RDF document is deleted. In order to obtain a better control we have modified the RDF documents creation. Now, if no annotation is associated to the image the document is created by the controller, otherwise the existing document is simply updated. The Integration Unit has been extended and modified in order to guarantee the units cooperation and preserve the consistency between the interface and the system. Within an activity aimed at defining an ontological approach to image understanding, the conceptualization of the domain has been detailed and the definition of the necessary suite of ontologies started. Current results of the activity have been detailed in a paper presented at the 8th International Conference on Pattern Recognition and Image Analysis: New Information Technologies, PRIA-8, held in Yoshkar Ola on 8-13 October 2007. Furthermore, the E-Team web site has been updated by the partners inserting new publications and events.

### **Events**

Digital Signal Processing Application Day 2007, Dept. of Computer Science and Communication, University of Milan, Milan (Italy), 17 September 2007. Master in Computational Science and Supercomputing, NEC-CESIC (Supercomputing Center for Computational Engineering), Cosenza (Italy), 23-24 October 2007. 8th International Conference on Pattern Recognition and Image Analysis: New Information Technologies PRIA – 8, Yoshkar-Ola, Russian Federation, 8-13 October 2007

### **Publications**

P. Asirelli, S. Colantonio, I.B. Gurevich, M. Martinelli, O. Salvetti, Yu. Trusova - Ontology Driven Approach to Image Understanding. 8th Int. Conf. on Patt. Rec. & Im. An. New Inf. Tech. PRIA8, Yoshkar-Ola, Russ. Fed., 8-13 Oct. 2007, Vol. 1, 67-71

## **2.3 Contribution by TU VIENNA-PRIP**

### **Researchers involved**

Allan Hanbury, Julian Stöttinger

### **Activities & Achievements**

The Third MUSCLE / ImageCLEF workshop on the evaluation of image and video retrieval took place in Budapest on the 18th of September 2007. Nine papers were presented at the workshop. These include two invited papers by Marcel Worring, University of Amsterdam (Netherlands) (Chair of the IAPR TC12) and Thijs Westerveld, CWI (Netherlands) (Co-organiser of the XML Multimedia Track at INEX). The proceedings of the workshop can be found here: [http://muscle.prip.tuwien.ac.at/ws\\_proceedings\\_2007.php](http://muscle.prip.tuwien.ac.at/ws_proceedings_2007.php)

The MUSCLE CIS Coin competition workshop took place on the 30th of August 2007. It turned out that the presence of occlusion, the main challenge in this years dataset, increased the difficulty of the problem immensely. Of 5 registrations, only 2 groups submitted results. The best performing system was submitted by Laurens van der Maaten of Maastricht University, the Netherlands. However, as the performance of this system was not up to the standard of the winner in 2006 (i.e. the occlusion problem was not solved satisfactorily), it was decided to halve the prize money. It is planned to write a journal paper including the results of all competition participants in 2006 and 2007. Negotiations on continuing the running of this competition within an EU project on coin classification have also started.

We also approached the IAPR TC5 (International Association on Pattern Recognition Technical Committee 5) on Benchmarking and Software about taking over the hosting of the MUSCLE Benchmarking server in 2008. Interest was expressed, and possibilities will be examined (Allan Hanbury is a member of the Steering Committee of this TC).

### **Events**

The Third MUSCLE / ImageCLEF workshop on the evaluation of image and video retrieval.

### **Publications**

Proceedings of the Third MUSCLE / ImageCLEF workshop on the evaluation of image and video retrieval (also MUSCLE Deliverable DN2.1).

A. Hanbury, B. Mičušík and J. Stöttinger, The MUSCLE Live Image Retrieval Evaluation Event, Proceedings of the Third MUSCLE/ImageCLEF Workshop on Image and Video Retrieval Evaluation (2007), pages 2–8, Budapest, Hungary.

H. Müller, T. Deselaers, M. Grubinger, P. Clough, A. Hanbury and W. Hersh, Problems with Running a Successful Multimedia Retrieval Benchmark, Proceedings of the Third MUSCLE/ImageCLEF Workshop on Image and Video Retrieval Evaluation (2007), pages 9–18, Budapest, Hungary.

M. Grubinger, P. Clough, A. Hanbury and H. Müller, Overview of the ImageCLEF 2007 Photographic Retrieval Task, Working Notes of the 2007 CLEF Workshop, Budapest, Hungary.

T. Deselaers, A. Hanbury, V. Viitaniemi, A. Benczúr, M. Brendel, B. Daróczy, H. J. Escalante Balderas, T. Gevers, C. A. Hernández Gracidas, S. C. H. Hoi, J. Laaksonen, M. Li, H. M. Marin Castro, H. Ney, X. Rui, N. Sebe, J. Stöttinger and L. Wu, Overview of the ImageCLEF 2007 Object Retrieval Task, Working Notes of the 2007 CLEF Workshop, Budapest, Hungary.

## **3 Overview activities in WP3**

### **3.1 Contribution by UCL**

#### **Researchers involved**

Fred Stentiford, Shijie Zhang, Rob Shilston.

#### **Activities & Achievements**

##### **Visual Saliency**

Work on attention based motion detection and estimation was presented at ICIP in September. The algorithms have been developed further on a wider range of video data and are now producing more consistent results. Further experiments on comparative performance are underway and a region tracking system is being designed.

A series of experiments are planned to evaluate models of visual quality including versions based upon visual attention. The following subjective questions will be answered: How closely do the models match the best subjective choice of focussed/defocussed images? How do different digital cameras perform when capturing an identical scene? How does a single camera perform when metered at different positions in the image?

##### **Events**

ICIP 2007

### **3.2 Contribution by TUG**

#### **Researchers involved**

Martina Uray, Martin Winter, Helmut Grabner, Peter Roth, Horst Bischof

#### **Activities & Achievements**

1. We did exagenuous performance evaluations of the behaviour of LDA. We are especially interested in the direct comparison of the results achieved by our local vocabulary tree framework and the global subspace method. Concerning robustness there are ongoing experiments concerning the breakdown point of LDA in contrast to PCA. 2. We developed a simple approach for person detection in surveillance for static cameras. The basic idea is to train a separate classifier for each image location which has only to discriminate the object from the background at a specific location. This is a considerably simpler problem than the detection of persons on arbitrary backgrounds. Therefore, we use adaptive classifiers which are trained on-line. Due to the reduced complexity we can use a simple update strategy that requires only a few positive samples and is stable by design. This is an essential property for real world applications which require operation for 24 hours a day, 7 days a week. We evaluated the method on publicly available sequences and compared it to state-of-the-art methods which reveal that despite the simple strategy the obtained performance is competitive.

## **Publications**

Helmut Grabner, Peter Roth, Horst Bischof Is Pedestrian Detection Really a hard Task? IEEE Workshop on Performance Evaluation of Tracking and Surveillance, 2007 (PETS'07), to appear

## **3.3 Contribution by MTA SZTAKI - GMCV lab**

### **Researchers involved**

S.Fazekas, D.Chetverikov

### **Activities & Achievements**

1. Development of a novel method for dynamic texture and background modelling in presence of periodic motion. Collecting test data with human motion on escalator. Testing the method on the collected data. Implementing a reference method (MMG, Stauffer-Grimson) for comparison with the developed one.
2. Started development of a novel method for fast and accurate optical flow calculation, based on lattice Boltzmann.

### **Problems**

More periodic backgrounds needed, not just escalators that have roughly constant colour making them less challenging as background.

### **Events**

MUSCLE-CLEF Workshop organised in SZTAKI on September 18, 2007. Local organiser: D. Chetverikov.

### **Publications**

A joint paper on dynamic texture detection and segmentation submitted to International Journal of Computer Vision together with T.Amiaz and N.Kiryati (TAU-Visual).

## **3.4 Contribution by Universitat Politecnica de Catalunya**

### **Researchers involved**

Montse Pardas, Veronica Vilaplana, Ferran Marques

### **Activities & Achievements**

- Presentation of the paper “Face detection and segmentation on a hierarchical image representation” (Veronica Vilaplana, Ferran Marques) at EUSIPCO 2007. This paper presents a face segmentation technique that works on a hierarchical region-based representation of the image. The algorithm bases its analysis strategy on a reduced set of regions that represent the image content at different scales of resolution. For each region, a set of simple one-class classifiers that rely on different shape, color and texture attributes is evaluated. The outputs of the classifiers are combined into a final face likelihood. The proposed system has been tested on a large set of images, providing very good results both in detection rate and accuracy of the segmented faces.
- Presentation of the paper “On building a hierarchical region-based representation for generic image analysis,” (Veronica Vilaplana, Ferran Marques) at ICIP 2007. This paper studies the procedure to create a hierarchical region-based image representation aiming at generic image analysis. This study is carried out in the context of bottom-up segmentation algorithms and, specifically, using the Binary Partition Tree implementation. The different steps necessary to create a hierarchical region-based representation are analyzed; namely, (i) the creation of the initial partition in the hierarchy, which is split into the definition of the initial merging criterion and the proposal of a stopping criterion, and (ii) the merging criteria used to produce the different regions in the final hierarchical representation. For both steps, the proposed approach is assessed and compared with previous existing ones over a large data set using well-established partition-based metrics.
- Collaboration with Surrey University. Bud Goswami visited UPC from the 22nd of October till the 9th of November. The work was within the e-team Person Detection, Recognition and Tracking and focussed on lip tracking. Different approaches to lip tracking have been considered including snakes and particle filters based tracking.

### **Publications**

“On building a hierarchical region-based representation for generic image analysis,” Veronica Vilaplana, Ferran Marques, ICIP 2007.

“Face detection and segmentation on a hierarchical image representation” Veronica Vilaplana, Ferran Marques, EUSIPCO 2007.

## **3.5 Contribution by Advanced Computer Vision GmbH**

### **Researchers involved**



## **Activities & Achievements**

We started work towards pedestrian detection and tracking in crowded scenarios where due to frequent partial occlusions most of the time only body parts are visible. A contour-based body-part representation appears to be appropriate, since body parts typically exhibit no texture or distinctive color information at common spatial image resolutions (such as in visual surveillance video data). We use the particle filtering framework to track head and shoulder shapes. We use an approximated representation for object shape (or a set of shapes for deformable or articulated objects) enabling fast computation of the image likelihood for a given hypothesis (particle). First tests of the approach show promising performance.

## **3.6 Contribution by ISTI-CNR**

### **Researchers involved**

Umberto Barcaro, Sara Colantonio, Davide Moroni, Ovidio Salvetti

### **Activities & Achievements**

The previously developed method for the segmentation of echocardiographic image sequences and the computation of Left Ventricle Ejection Fraction has been refined and adapted to the images acquired by different echocardiography devices from various manufacturers. In particular, an IT infrastructure has been configured, whose core is a DICOM server that allows the users for storing, querying and retrieving image sequences. In this way, the data are organized in a structured format and may be used for constructing statistically-meaningful shape and appearance models, according to the aims of the e-team 4 on "Shape Modelling". Such models may be used for segmentation purposes via a shape prior term in a variational formulation of level sets. After this preliminary but important stage, the image database could be annotated with automatically computed shape descriptors both of regional and global nature, thus permitting similarity searches and data mining procedures. An activity concerning the segmentation of breast dense tissue in mammography has started with the aim of interpreting mammograms and providing an automatic classification according to the BI-RADS mammography assessment categories. In addition a fire detection system based on integrated Infrared-Visible cameras has been developed. The system includes two coupled, aligned and mobile cameras, one acquiring at Far-IR and the other at visible spectrum wavelength. The system is able to monitor, detect and localize forest fires in the area under surveillance. The method exploits a first module for the rough localization of fires, based on the comparison of the radiance gradient with reference values acquired a priori. Then, the result is refined exploiting the visible camera information: fragments of the visible camera images are compared with the corresponding fragments in the reference database, in order to give evidence of the previously detected fire location. Finally, the performance of the system is improved by fusing further information such as digital terrain models and by storing meteorological data and fire risk factors in a database. The method has been reported at the 9th International Workshop on Advanced Infrared Technology and Applications - AITA07 held in Leon (Mexico). In collaboration with the Physics Department of the University of Pisa, an activity concerning the analysis of tribological properties of polymers has started. The main aim is to provide a 3D reconstruction of an object from several views obtained by

scanning electronic microscopy technique. After reconstruction of the object, automatically computed differential geometric features (such as mean and Gaussian curvatures) may be related to physical relevant properties of the material. Dissemination activities have been carried out delivering seminars in two Italian University centres (see the Events section). At the DSP Application Day, organized by the Computer Science and Communication Department of the University of Milan, special emphasis has been given to topic related to the WP3 and to the e-team 4 on "Shape Modelling". Actually, methods for feature extraction, shape description and analysis have been reported, including techniques for visual content indexing, also in the framework of MPEG7 Visual. In another seminar delivered to the Master in Computational Science and Supercomputing, NEC-CESIC (Supercomputing Center for Computational Engineering) in Cosenza (Italy), aspects of biomedical image processing have been discussed. In particular, recent Level Set segmentation techniques and their possible applications to biomedical images have been treated extensively. Further, in the framework of e-team 4 on "Shape modeling", the problem of constructing suitable shape priors for smart segmentation has been addressed.

## **Events**

1-S. Colantonio, D. Moroni, O. Salvetti. Methods for Digital Image Processing. Seminar delivered to the DSP Application Day 2007, Dipartimento di Informatica e Comunicazione, University of Milan, Milan (Italy), 17 September 2007. 2-S. Colantonio, D. Moroni, O. Salvetti. Image Processing in Biomedical Applications. 4-hour long seminar delivered to the Master in Computational Science and Supercomputing, NEC-CESIC (Supercomputing Center for Computational Engineering), Cosenza (Italy), 23-24 October 2007

## **Publications**

G. Pieri, O.Salvetti: "Fire Detection System based on integrated Infrared-visible Cameras". Proceedings of the 9th International Workshop on Advanced Infrared Technology and Applications - AITA07. Leon, Mexico, October 8 - 12, 2007.

## **3.7 Contribution by TAU SPEECH**

### **Researchers involved**

Arie Yeredor (PI)

### **Activities & Achievements**

We have revised our paper (Efrat Be'ery and Arie Yeredor, "Blind Separation of Superimposed Shifted Images Using Parameterized Joint Diagonalization", submitted to IEEE Transactions on Image Processing) in accordance with the remaining comments from one of the reviewers, mainly regarding the possibility to consider spatial-domain reconstruction instead of our frequency-domain approach. Following submission of the revised version, the paper has now been formally accepted for publication.

## 3.8 Contribution by UvA

### Researchers involved

Jasper Uijlings, Nicu Sebe, Cees Snoek

### Activities & Achievements

1. Cooperative Object Tracking We developed a system that can track objects with multiple calibrated Pan-Tilt-Zoom (PTZ) cameras in a cooperative fashion. Tracking and calibration results are combined with several image processing techniques in a statistical segmentation framework through which the cameras can hand over targets to each other. 2. Object Recognition using Correlograms When using the Generalized Correlogram (GC) as an image representation, the object is described as a constellation of GCs where each one encodes information about some local part and the spatial relations from this part to others (i.e. the part's context). We show how such a representation can be used with fast procedures that learn the object category with weak supervision and efficiently match the model of the object against large collections of images. In the learning stage, we show that by integrating our representation with Boosting the system is able to obtain a compact model that is represented by very few features, where each feature conveys key properties about the object's parts and their spatial arrangement. In the matching step, we propose direct procedures that exploit our representation for efficiently considering spatial coherence between the matching of local parts. Combined with an appropriate data organization such as Inverted Files, we show that thousands of images can be evaluated efficiently. The framework has been applied to different standard databases and we show that our results are favorably compared against state-of-the-art methods in both computational cost and accuracy. 3. Color Interest Points for Image Retrieval In image retrieval scenarios, many methods use interest point detection at an early stage to find regions in which descriptors are calculated. Finding salient locations in image data is crucial for these tasks. Observing that most current methods use only the luminance information of the images, we investigate the use of colour information in interest point detection. A way to use multi-channel information in the Harris corner detector is explored and different colour spaces are evaluated. To determine the characteristic scale of an interest point, a new colour scale selection method is presented. We show that using colour information and boosting salient colours results in improved performance in retrieval tasks. 4. Personalized Multimedia Information Retrieval In our overview we are considering several aspects of this problem. On one hand, the user will want to have a personalized access to his image/video collections and this can be achieved by providing intuitive and natural browsing capabilities and customized features. Furthermore, the system is required to perform user profiling and to adapt the existing parameters of the system to the user needs and interest. On the other hand, it is also important to consider the devices and applications in which this technology is going to be deployed. Mobile media is a high growth area but the state-of-the-art technologies are lagging behind the consumers expectations. In this overview we are addressing these three important aspects. 5. Detecting Attention through Telepresence Attention is the cognitive process of selectively concentrating on one aspect of the environment while ignoring other things. As such, attention is one of the most intensely studied topics within psychology and cognitive neuroscience. In this research, we present a way to detect attention during a remote communication session. Our approach considers the difference between face-to-face and remote communication. In the first case, participants directly share the physical space, information, and emotional cues, while a remote

communication scenario will be hampered by some limitations. To address these, we enhance awareness by adding a cognitive reaction to an event: the system is aware if the user is conscious about the discussion topic and individual/global performance. In our research, we consider attention as a particular feature of awareness. We measure the attention level, based on user activities and his facial expressions. Activities are performed by participants during a remote video conferencing, through mouse clicks and keystrokes. Each user's face is analyzed to find the values of seven basic emotional states: neutral, happy, sad, fear, disgust, surprise and angry, following the codification of "universal emotions" suggested in the psychology field.

## Events

ACM International Workshop on Human-centered Multimedia, September 2007, Augsburg, Germany  
ACM International Workshop on Multimedia Information Retrieval, special session on Personalized Multimedia Retrieval, September 2007, Augsburg, Germany  
IEEE International Workshop on Human-computer Interaction, October 2007, Rio de Janeiro, Brasil

## Publications

- Cooperative Object Tracking with Multiple PTZ Cameras I. Everts, N. Sebe, G. Jones, International Conference on Image Analysis and Processing (ICIAP), Modena, Italy, September 2007.
- Facial Expression Recognition: A Fully Integrated Approach R. Valenti, N. Sebe, T. Gevers, International Workshop on Visual and Multimedia Digital Libraries, Modena, Italy, September 2007.
- Class-Specific Binary Correlograms for Object Recognition J. Amores, N. Sebe, P. Radeva, British Machine Vision Conference, Warwick, UK, September 2007.
- Do Colour Interest Points Improve Image Retrieval? J. Stoettinger, A. Hanbury, N. Sebe, T. Gevers, International Conference on Image Processing, San Antonio, TX, USA, September 2007.
- Personalized Multimedia Retrieval: The New Trend? N. Sebe, Q. Tian, ACM Multimedia Information Retrieval Workshop, Augsburg, Germany, September, 2007.
- Jaimes and N. Sebe, Multimodal Human-computer Interaction: A Survey, Computer Vision and Image Understanding, volume 108, number 1-2, pages 116-134, October/November 2007.
- J. Amores, N. Sebe, and P. Radeva, Context-Based Object-Class Recognition and Retrieval by Generalized Correlograms, IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 29, number 10, pages 1818-1833, October 2007.
- M. Lew, E. Bakker, N. Sebe, and T.S. Huang, Human-computer Intelligent Interaction: A Survey, IEEE International Workshop on Human-computer Interaction (HCI'07), Rio de Janeiro, Brazil, October 2007.
- S. Levialdi, A. Malizia, T. Onorati, E. Sanginetto, and N. Sebe, Detecting Attention through Telepresence, 10th Annual International Workshop on Presence, Barcelona, Spain, October 2007.
- N. Sebe, M. Lew, T. S. Huang, E. Bakker (Eds.), Computer Vision in Human-Computer Interaction, Proceedings of the 4th International Workshop on Human Computer Interaction (HCI07) - in conjunction with the IEEE International Conference on Computer Vision (ICCV 2007), Lecture Notes in Computer Science, vol. 4796, Springer-Verlag, ISBN 3-540-75772-6, October 2007.

## **3.9 Contribution by ARMINES-CMM**

### **Researchers involved**

Youssouf Chherawala, Beatriz Marcotegui and Francis Bach

### **Activities & Achievements**

First classification experiments have been performed in the framework of the “Choosing Features for CBIR and Automated Image Annotation” e-team. The SVM classifier has been used with the classical Gaussian kernel. Many colour and texture features are available to describe the regions of the images. The results have shown that the BIC feature (Border Interior pixel Classification) performed better (55 % of accuracy) than other colour and texture features (Accuracy: rgb 43%, LEP 25%, HSV 15%). Also ACP allows to reduce the BIC vector size (from 412 to 25) without significant loss of accuracy (53%), the data size reduction allows to reduce also the computing time for the SVM. We are currently running classification tests based on local features computed on segmented regions. Multiple Kernel Learning (MKL) and segmentation neighbourhood will also be introduced to improve our results.

## **3.10 Contribution by AUTH**

### **Researchers involved**

I. Kotsia, I. Pitas

### **Activities & Achievements**

A novel method for the recognition of the six basic facial expressions plus the neutral state in videos that incorporate geometrical information and SVMs has been developed. The system first extracts the deformed Candide facial grid that corresponds to the facial expression depicted in the video sequence. The mean Euclidean or Hausdorff distance of the deformed grids is then calculated to create a new metric multidimensional scaling. This scaling is used to define an embedding in a new multidimensional Euclidean space. In this space, a multi-class SVM system is used to perform classification of the sample under examination to one of the 7 possible classes of facial expressions, i.e. anger, disgust, fear, happiness, sadness, surprise and neutral. Face recognition was also attempted, including images taken at different time instances as well as variations in facial expressions (open/closed eyes, smiling/non-smiling) and facial details (glasses/no glasses). This work resulted in a journal paper that is currently under review in the IEEE Transactions on Neural Networks.

### **Events**

AUTH hosted M. Haindl and P. Somol who gave lectures to the researchers of the lab. Joint research initiatives were explored.

## **3.11 Contribution by TU VIENNA-PRIP**

### **Researchers involved**

Allan Hanbury, Lech Szumilas

### **Activities & Achievements**

Work on object recognition using image keypoints based on a measure of symmetry combined with a new feature describing the shape of the area around keypoints has continued. Emphasis has been on the development of an algorithm for orientation and scale invariant matching of these features. The novelty lies in the fact that the invariance is not encoded in the features, but in the matching algorithm.

### **Publications**

L. Szumilas, H. Wildenauer, A. Hanbury and R. Donner. Radial Edge Configuration for Semi-Local Image Structure Description, Third International Symposium on Visual Computing (ISVC 2007), Lake Tahoe, Nevada, USA

## **3.12 Contribution by GET-ENST**

### **Researchers involved**

Beatrice Pesquet-Popescu, Wided Miled, Aurélie Fraysse

### **Activities & Achievements**

A new activity started with Aurélie Fraysse, concerning the stochastic modeling of textures through multifractional Brownian motion sheets. This model provides a very flexible description of complex textures. Our work involves parameter estimation for texture synthesis and interpolation. Another new activity started with Wided Miled, concerning the disparity estimation for stereo video sequences, by using variational methods. This disparity field is an important element for indexing stereo sequences. Finally, a contact has been taken with Sara Parrilli and Gianni Poggi, from Univ. Federico II of Naples, Italy, to continue the work on content adaptive spatial representations. Sara Parrilli will come to visit GET-ENST in January, for a period of 6 months.

## **3.13 Contribution by CWI**

### **Researchers involved**

Eric Pauwels, Paul de Zeeuw, Elena Ranguelova

### **Activities & Achievements**

We have worked on algorithms that perform image-based shape matching within the domain of tree taxonomy. As such, it serves as an example relevant to many other potential

applications within the field of biodiversity and photo-identification. Unsupervised matching results are produced through a chain of computer vision and image processing techniques, including segmentation and automatic shape matching. The matching itself is based on a nearest neighbours search in an appropriate feature space, constructed automatically by training the system on a annotated set of exemplars. To capture the shape we computed a large set of semantically overlapping shape features. We intentionally introduced redundancy in the feature set to improve robustness. The idea is that it is often prohibitively costly to check whether features accurately capture the visual characteristics they have been designed for (i.e. whether they have successfully bridged the so-called numerical gap). We therefore use the redundancy in the feature set to flush out problem cases. Indeed, if two features — designed to quantify the same visual characteristic albeit using complementary methods — yield divergent results, then that image is earmarked for inspection by a human supervisor. This way we can quickly search the database for outliers that might be indicative of aberrant shapes or, (perhaps even more importantly) flaws in the assumptions subsumed by the features. We are currently extending these techniques for applications to other biodiversity datasets.

## **Publications**

E.J. Pauwels, P.M. de Zeeuw, E. Ranguelova: Computer-assisted Tree Taxonomy by Automated Image Recognition. Submitted to Special Issue of "Engineering Application of Artificial Intelligence"

## **3.14 Contribution by IRIT-UPS**

### **Researchers involved**

Philippe Joly, Elie El Khoury, Christine Sénéac

### **Activities & Achievements**

The segmentation method initially designed for 'audio clips' boundaries detection is now tested in the domain of video data for shot boundaries detection and program boundaries detection. This method seems robust and is really generic as it doesn't need any knowledge in order to identify correctly true shot boundaries from the following: fast object or camera motion, fast illumination changes, reflections, sudden change due to explosion and flash photography. Results, obtained on the corpora of the Argos campaign and the Trecvid campaign, are among the best. This method is also being tested in order to detect program boundaries on the TV stream. First results obtained with a 100 hours corpus are promising (precision and recall of 87%).

## **3.15 Contribution by INRIA-Ariana**

### **Researchers involved**

Josiane Zerubia, Ian Jermyn, Avik Bhattacharya, Aymen El Ghoul, Peter Horvath, Ting Peng.

### **Activities & Achievements**

Ting Peng, joint PhD student of INRIA Ariana and the LIAMA Institute in Beijing, has continued to work on models for road network segmentation from very high resolution (0.5m) satellite images. As described in the previous report, Ms Peng has fixed on two competing models for the extraction of narrower roads. One involves asymmetrizing the interaction in the higher-order energy term to allow it to vary with the relative orientation of the contour normal vectors, while the other uses a hitherto unrecognized linear term that takes into account the relative orientation of the vector between the two interacting points and the normal vectors. The results obtained with both models are very good, but the second is preferable computationally. Ms Peng is continuing to test both models to see which performs best. She has also prepared and submitted a publication to the IEEE Transactions on Geoscience and Remote Sensing, and a conference publication to ISPRS 2008. Aymen El Ghouli began his PhD at INRIA-Ariana in September. He is working on extending the higher-order active contour model of road networks, and its phase field formulation, to the cases of grid-like road networks and river networks, with the aim of extracting these entities from high resolution satellite and aerial images. As described in the previous report, as part of his internship work, he performed a stability analysis for a long straight bar analogous to that performed by Mr. Horvath for a circle, and tested it against the results of gradient descent with success. Subsequently, he has performed the same type of stability analysis for the new nonlinear model developed by Ting Peng. Mr El Ghouli is currently performing extra experiments on the latter model to add to the INRIA Research Report and subsequent journal paper. He has also submitted a paper to RFIA 2008, and he attended a Summer School on "Level set methods in image processing" at Montpellier. Peter Horvath, joint PhD student of INRIA Ariana and the University of Szeged, submitted his thesis to the University of Szeged in August, and to the University of Nice Sophia Antipolis in September. He will defend on December 3. He has already found employment at the Institute for Biochemistry of ETH Zurich, a position he took up in September. Avik Bhattacharya, joint PhD student of INRIA Ariana and ENST, submitted his thesis to ENST in October. He will defend on December 14 at ENST in Paris.

## **Publications**

A paper has been published in EUSIPCO 2007, and another in BMVC 2007, on the research of Peter Horvath and Ting Peng respectively.

## **3.16 Contribution by UniS**

### **Researchers involved**

A. Khan, B. Christmas

### **Activities & Achievements**

We improved the speed and accuracy of our SVM / level set based segmentation method, with application to detection of lip pixels. In the original method the set of training images was chosen subjectively (i.e. those that "visually" seemed representative of a large range of skin types). Subsequently we added more faces to the training set corresponding to cases where the method had partially or completely failed, with the intention that they could represent face types for which the SVM had not been given sufficient information (via training). We then tried some methods (such as the CONDENSE method, that removes "interior" samples from



the training set) to reduce the number of training samples. This improved the number and quality of the successful results somewhat, and also speeds up the SVM training process, but it also means that the training set was derived from a larger number of face images.

## **Publications**

F. Yan, W.J. Christmas and J Kittler, "All Pairs Shortest Path Formulation for Multiple Object Tracking with Application to Tennis Video Analysis", Proceedings of British Machine Vision Conference , September, pages 650-659, vol. 2, [www.ee.surrey.ac.uk/CVSSP/Publications/papers/yan-bmvc-2007.pdf](http://www.ee.surrey.ac.uk/CVSSP/Publications/papers/yan-bmvc-2007.pdf)

# **4 Overview activities in WP4**

## **4.1 Contribution by CNR-ISTI**

### **Researchers involved**

Graziano Bertini, Vincenzo Di Salvo, Massimo Magrini, Leonello Tarabella.

### **Activities & Achievements**

- Preparation of non compressed and compressed musical tracks for objective and subjective evaluation of ARIA algorithm (release 2).
- Testing of Pandora Multi Media System performance with the projection of special video effects controlled by audio parameters extracted from live instrument's sounds.
- Going on porting of pCM++ audio synthesis&processing framework from MacOS to Windows platform. - Preparation of a set of musical examples jointly with partners for final demo and evaluation results of MODEM EU project.

### **Events**

- MIDIA (1° Mercato Internazionale dell'Immagine Animata - Digital Entertainment & New Media, 4-8 October 2007 Nuova Fiera di Roma, Italy), show/demo of Pandora System and conferences participation of ISTI audio team.
- MODEM EU project final meeting (Deffenu Secondary School, 12 October 2007, Olbia, Italy)

### **Publications**

"Real-time sound-parameters tracking in a multimedia system". In: "DSP Application Day" conference(Milano, 17 settembre 2007). Proc. on CD-ROM, Mario Malcangi Editor, DICO Dept. Informatica e Comunicazione, Univ. di Milano.

## **4.2 Contribution by TAU SPEECH**

### **Researchers involved**

PIs: Arie Yeredor, David Burshtein

### **Activities & Achievements**

1) Support vector machine training for improved hidden Markov modeling: We have continued our work by conducting experiments on the N-best version of the algorithm for connected speech recognition. Results on the tidigits database have shown very significant improvements on the test set compared to plain EM training of HMMs. We have also started to summarize this work in a paper that we plan to submit for publication. This work is also related to WP6. 2) Single channel audio separation: In our continued work on single-channel audio separation, we applied some further improvements and corrections to the algorithm proposed by Pearlmutter and Olsson. New code for finding a discriminating dictionary for the audio sources has been written and tested in Matlab. The code uses the L<sub>1</sub> "sparse" dictionaries constructed using Pearlmutter and Olsson's algorithm as initial guesses, and then minimizes an L<sub>2</sub> distance measure using gradient optimization. A multi-band HMM separation algorithm based on Reyes-Gomez et al. ("Multiband audio modeling for single-channel acoustic source separation", ICASSP 2004) has also been considered.

## **4.3 Contribution by KTH**

### **Researchers involved**

Bjorn Granstrom, Preben Wik

### **Activities & Achievements**

A demonstration of the articulatory talking head was given at IBC in Amsterdam, September 7-11. This formed a final presentation of the showcase MuscleHead that KTH has been a partner in. A special feature of the head is that prosody information is displayed. Moreover, auditory and visual cues are combined to give speaker friendliness. Previous work on texture analysis has resulted in a submitted paper on retrieval of pictures containing animals and on recognition of animals. This work has been done in the visual salience E-team.

## **4.4 Contribution by AUTH**

### **Researchers involved**

E. Benetos, C. Kotropoulos

### **Activities & Achievements**

Joint research with TU Vienna-IFS was conducted within the CAS showcase. The goal of this showcase is to bring together and demonstrate the wide range of semantic analysis and annotation capabilities that are present within MUSCLE. The feature sets that were extracted

from the audio parts of the videos in the database are described. Four different sets of features were used. Rhythm Patterns, Statistical Spectrum Descriptors and Rhythm Histograms were developed by TU Vienna - IFS and perform an analysis of a psycho-acoustically transformed spectrogram on the Bark-scale. AUTH extracted MPEG-7 based descriptors and general audio data (GAD) features. In order to reduce the feature set cardinality, a suitable subset for classification has to be selected. The optimal feature subset should maximize the ratio of the inter-class dispersion over the intra-class dispersion. For the CAS classification experiments, 80 features were selected out of the 187 for the AUTH dataset. In addition, 100 features were selected out of the 1440 rhythm pattern descriptors. Another selection of 100 features was performed out of the concatenation of rhythm histograms, rhythm patterns, and the statistical spectrum descriptor, which have a feature cardinality of 1668. A final selection of 100 features was performed using all feature sets from TU Vienna-IFS and AUTH. Classifiers, such as Support Vector Machines with a 2nd order polynomial kernel, Non-negative Tensor Factorization and cosine classifiers were tested. On average 65.96% classification accuracy has been obtained.

## **Events**

C. Kotropoulos attended 2007 European Signal Processing Conference at Poznan, Poland. E. Benetos has visited TU-Vienna IFS in order to conduct joint research within the framework of CAS showcase.

## **4.5 Contribution by CNRS LLACAN**

### **Researchers involved**

Fathi Debili, Zied Ben Tahar

### **Activities & Achievements**

On one hand, we continued our work on the syntactic analysis of Arabic and the interactive annotation. On the other hand, we worked on the recognition of the named entities too. A corpus of approximately 250000 words was annotated manually (vowelized, lemmatized and labelled) and was checked. Programs of syntagmatic and dependency annotation were done too. Now, we have to use them in a massive way, improve the automatic component, and measure the performance of the automatic annotation vs the interactive cost.

### **Publications**

S. Guillemin-Lanne, F. Debili, Z. Ben Tahar, C. Gaci. Reconnaissance des entités nommées en Arabe. VSST'2007, Marrakech, Octobre 2007, 13 pages.

## **4.6 Contribution by IRIT-UPS**

### **Researchers involved**

Helene Lachambre, Julien Pinquier, Régine André-Obrecht

## **Activities & Achievements**

In the context of audio structuration, we have worked on improving the jingle detector developed in our team [1]. The idea is to use well known jingles to not only locate these types of jingles, but also to discover other ones. From several well known jingle prototypes, we first detect candidate segments and then we identify each of them as known jingles or unknown ones. So the jingle reference database increases. For that purpose, we use only spectral representations and Euclidean distances coupled with some heuristic rules. The data base ESTER is currently used to assess the method. [1] Julien Pinquier, Régine André-Obrecht. Jingle detection and identification in audio documents. In : Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'2004), Montréal, Canada, 17/05/2004-21/05/2004, Volume IV, IEEE, p. 329-332.

## **4.7 Contribution by TU Vienna - IFS**

### **Researchers involved**

Jakob Frank, Thomas Lidy, Robert Neumayer, Andreas Rauber

### **Activities & Achievements**

At the Workshop on Video and Multimedia Digital Libraries (VMDL'07) which was held in conjunction with the International Conference on Image Analysis and Processing (ICIAP'07) in Modena, Italy, September 10-13 2007, we presented a paper subsuming our work on content-based audio analysis and its applications, specifically on mobile devices. At the International Conference on Music Information Retrieval (ISMIR 2007) in Vienna, Austria, we presented our work done on "Improving Genre Classification By Combination Of Audio And Symbolic Descriptors Using A Transcription System". We participated at the 2007 Music Information Retrieval Evaluation eXchange (MIREX), where we made the following achievements: rank 2 (of 7) in Audio Genre classification, rank 3 (of 9) in Audio Mood classification, rank 5 (of 7) on Composer Identification, rank 4 (of 7) on Artist Identification and rank 5 (of 12) for Audio Music Similarity and Retrieval. In October a researcher from the University of Thessaloniki (AUTH) visited our group at TU Vienna IFS. Experiments on a feature combination approach on audio analysis of the video database of the Evaluation Showcase were done.

### **Events**

14th International Conference on Image Analysis and Processing (ICIAP'07), 1st Workshop on Video and Multimedia Digital Libraries (VMDL'07), Modena, Italy, September 10-13 2007. 8th International Conference on Music Information Retrieval (ISMIR 2007), Vienna, Austria, September 23 - 27 2007. Music Information Retrieval Evaluation eXchange (MIREX) 2007. Exchange visit from University of Thessaloniki (AUTH) at TU Vienna IFS.

### **Publications**

Robert Neumayer, Jakob Frank, Peter Hlavac, Thomas Lidy, and Andreas Rauber. Bringing mobile based map access to digital audio to the end user. In Proceedings of the 14th International Conference on Image Analysis and Processing (ICIAP'07), 1st Workshop on

Video and Multimedia Digital Libraries (VMDL'07), pages 9-14, Modena, Italy, September 10-13 2007. IEEE.

## **4.8 Contribution by UniS**

### **Researchers involved**

C. Hory, W.J. Christmas

### **Publications**

C Hory and W Christmas: "Cepstral features for classification of an impulse response with varying sample size dataset", Proceedings of European Signal Processing Conference, September, pp. 1546 - 1550, <http://www.ee.surrey.ac.uk/CVSSP/Publications/papers/hory-eusipco-2007.pdf>

# **5 Overview activities in WP5**

## **5.1 Contribution by VTT**

### **Researchers involved**

Sanni Siltanen, Petri Honkaamaa, Seppo valli, Charles Woodward

### **Activities & Achievements**

Contribution to e-Team book "Multimodal Processing and Interaction"

## **5.2 Contribution by INRIA - Texmex**

### **Researchers involved**

Patrick Gros Gwenole Lecorve Guillaume Gravier Pascale Sebillot Stephane Huet

### **Activities & Achievements**

From a scientific point of view, our work focuses on speech processing for TV streams analysis. TV presents several specific difficulties: a relatively short video can concern many topics (advertisements, news reports), these topics can be quite short in term of number of words... while speech recognition systems are very dependent of their language model and often require long speech records. We have started to develop a system adapted to such data which aims to adapt automatically the language model based on a first blind recognition and a web based adaptation. During the last two months, we have mainly sought to refine the Web-based topic language model adaptation method developed beforehand. First, the previous topic characterization technique was sometimes unreliable because of the rather small size of

the document reference corpus used. To circumvent this problem, this corpus has been replaced a larger one : the French newspaper Le Monde corpus composed of 800,000 articles dated from 1987 to 2003. Then, new experiments on the whole adaptation technique have been carried out on a set of broadcast news reports much larger than the one used during the research master (currently 160 vs. 22 previously). On the one hand, these experiments led to confirm that the proposed Web-based topic adaptation is relevant to improve the transcription of spoken documents. On the other hand, these results have highlighted that the current method still needs to be made more robust according to, for instance, the nature of encountered topics (e.g. sports should probably be processed in a different way than politics) or the length of considered documents (e.g., in small documents, it is rather difficult to characterize well the topic). Second, in the light of these conclusions, new reflections have been engaged to determine other possible improvements of the current method like gathering thematically similar documents to factorize the adaptation process or using an iterative procedure to make the keyword spotting task more robust. These ideas will probably be jointly studied with Stéphane Huet. Finally, one article which summarized these results have been submitted to the ICASSP conference and a long abstract have been proposed to the LREC conference to present the effectiveness of using the Web as a linguistic resource.

## **Events**

October 4th: MUSCLE WP5 meeting in Chania. Participation of Patrick Gros. Discussion on the WP5 book, scientific presentation.

## **5.3 Contribution by AUTH**

### **Researchers involved**

V. Moschou, C. Kotropoulos, D. Ververidis

### **Activities & Achievements**

A new release (Release 1.3) of AUTH MUSCLE Movie Database is in preparation that includes three clips of duration between 7 to 10 sec in order to enable audiovisual saliency detection. The annotation is jointly performed with NTUA. Common annotation protocols have been defined and used. On-going research on speaker diarization within the showcase “Movie Summarization and Skimming Demonstration” has been continued aiming at dialogue detection in movie scenes. The overall classification error has been reduced to 8.89%, when outliers are not included in the clustering. However, the algorithm tends to over-estimate the number of clusters. Joint research with NTUA using speech activity detection using the multiband modulation energy has been launched.

## **Events**

C. Kotropoulos and V. Moshcou attended IEEE Workshop on Multimedia Signal Processing at Chania, Greece. They have demonstrated speaker clustering and attended the WP5 showcase meeting.

## 5.4 Contribution by ICCS-NTUA

### Researchers involved

G. Papandreou, A. Katsamanis, V. Pitsikalis, P. Maragos

### Activities & Achievements

## Audio-Visual Interaction for Speech Processing

In our on-going effort in the field of multimodal fusion for Automatic Audio-Visual Speech Recognition and Processing, we have extended our approach to automatic multimodal fusion by uncertainty compensation by incorporating the effect of measurement uncertainty into model training procedures as well. This allows model training from noisy data and improves speech recognition performance, especially at slow SNR values.

### Publications

G. Papandreou, A. Katsamanis, V. Pitsikalis, and P. Maragos, “Multimodal Fusion and Learning with Uncertain Features Applied to Audiovisual Speech Recognition”, in Proc. IEEE Workshop on Multimedia Signal Processing (MMSP-2007), October 2007.

## 5.5 Contribution by ICCS-NTUA

### Researchers involved

ICCS-NTUA (G. Papandreou, P. Maragos) TSI-TUC (M. Perakakis, A. Potamianos, E. Sanchez-Soto) INRIA-Textmex (G. Gravier, P. Gros)

### Activities & Achievements

#### Real-time Audio-visual Automatic Speech Recognition Demonstrator Showcase

We have continued during the reporting period the effort to build a real-time audio-visual automatic speech recognition demonstrator, as part of the Muscle Showcasing initiative. We have spent significant effort in streamlining the visual front-end of the prototype. Trying to overcome the limitations of our previous research-level visual front-end module, which was designed for off-line feature extraction in well-defined scenarios, such as processing videos shot under carefully controlled conditions, we have made the following improvements: We have integrated a fast adaboost-based face detector in our visual front-end. The role of the face detector is to initialize the face tracker at the first frame, and also whenever tracking fails. We have made algorithmic advances in the techniques for AAM fitting, yielding efficient and accurate computer vision algorithms for parametric model fitting and tracking. Repeated image resampling at updated shape warps (texture mapping in computer graphics terminology) constitutes a significant part of the computational load during model fitting. We build on advanced features of modern graphics cards (GPUs) and perform texture-mapping very efficiently on the GPU. We are currently working on integrating all these components into our prototype's visual front-end.

## Events

Regarding dissemination activities, a first version of the demonstrator was showcased as a Technical Demo at the IEEE Workshop on Multimedia Signal Processing (MMSP-2007), held in Chania, Greece, on October 1-3 2007 and exhibited at the IBC exhibition held in Amsterdam, Holland, on September 12-16 2007.

## 5.6 Contribution by ICCS-NTUA

### Researchers involved

G. Evangelopoulos, K. Rapantzikos, and P. Maragos

### Activities & Achievements

#### Audiovisual Attention Modeling and Salient Event Detection

Although human perception appears to be automatic and unconscious there exist complex sensory mechanisms that form the preattentive component of human understanding and lead to awareness. Considerable research has been carried out into these preattentive mechanisms and computational models have been developed and employed to common computer vision or speech analysis problems. The separate audio and visual modules may convey explicit, complementary or mutually exclusive information around structures of audiovisual events. In any video sequence the two streams are processed in parallel. Based on recent studies on perceptual and computer attention modeling, we extract attention curves using features around the spatiotemporal structure of video and sounds. The potential of intra-module fusion and audiovisual event detection is demonstrated in applications such as key-frame selection, video skimming and summarization and audio/visual segmentation. During the reported period, there has been considerable progress in building an image saliency Matlab toolbox which will soon be put on the public domain. More specifically, this code is an implementation of T. Lindeberg's scale-invariant primal sketch used for object detection and shape analysis, and detects through an image scale-space salient edge and ridge contours. These contours are then post-processed to obtain a sparse set of straight line segments, which can then be utilized for higher level recognition tasks.

### Publications

K. Rapantzikos, G. Evangelopoulos, P. Maragos, and Y. Avrithis, "An audiovisual saliency model for movie summarization", in Proc. IEEE Workshop on Multimedia Signal Processing (MMSP-2007), 2007.

## 5.7 Contribution by ICCS-NTUA

### Researchers involved

ICCS-NTUA (P. Maragos, G. Evangelopoulos, K. Rapantzikos, I. Avrithis) AUTH (C. Kotropoulos, P. Antonopoulos, V. Moschou, N. Nikolaidis, I. Pitas) INRIA-Textmex (P. Gros, X. Naturel) TSI-TUC (A. Potamianos, E. Petrakis, M. Perakakis, M. Toutoudakis)



## **Activities & Achievements**

### Movie Summarization and Skimming Demonstrator Showcase

During the reporting period, we have launched an effort to build a Movie Summarization and Skimming Demonstrator, as part of the Muscle Showcasing initiative. Participating partners are ICCS-NTUA (leader), TSI-TUC, AUTH, and INRIA-Textmex. As the amount of video data available (movie, TV programs, clips) in a personal recorder or computer are becoming increasingly large (100h in VCRs or hundreds of hours on a PC) intelligent algorithms for efficiently representing video data and presenting them to the user are becoming important. Video summarization, movie summarization and movie skimming are increasingly popular research areas with immediate applications. In this showcasing project we will: (i) use combined audio and video saliency detectors to identify the importance of movie content to the user and (ii) design an interface that presents the audio and video information to the user in a compressed form, thus saving time with little or no loss of information. The demonstrator will have the ability to render a movie from its typical 2h duration down to 30' by skimming over (fast forwarding or omitting) non-salient movie scenes while playback at regular speed parts of the movie with salient audio and video information. The interface will also have the ability to break the synchrony of the audio/video streams and selectively present audio or video information.

## **Events**

Regarding dissemination activities, a first version of the demonstrator was showcased as a Technical Demo at the IEEE Workshop on Multimedia Signal Processing (MMSP-2007), held in Chania, Greece, on October 1-3 2007 and exhibited at the IBC exhibition held in Amsterdam, Holland, on September 12-16 2007.

## **5.8 Contribution by ICCS-NTUA**

### **Researchers involved**

Petros Maragos (ICCS-NTUA), Alexandros Potamianos (TSI-TUC) and Patrick Gros (INRIA-Textmex)

## **Activities & Achievements**

### Book on "Multimodal Processing and Interaction: Audio, Video, Text"

The book under preparation is covering the thematic areas of WP5 (former JPA2 WP6 & WP10). It comprises two main parts: Part A, which is a comprehensive State-of-the-Art review of the area and Part B which consists of selected research contributions / chapters by Muscle WP5 members. The book proposal has been approved by Springer-Verlag. During the reporting period, all book contributors have submitted a first version of their corresponding chapter, and have received reviews by at least two referees. In the current phase the chapter authors are incorporating the revisions suggested by the reviewers in their documents.

## 5.9 Contribution by ICCS-NTUA

### Researchers involved

A. Katsamanis, A. Roussos, G. Papandreou, P. Maragos (ICCS-NTUA) Y. Laprie (INRIA-LORIA)

### Activities & Achievements

#### Audio-Visual Speech Inversion

In this research track we address the problem of audiovisual speech inversion, namely recovering the vocal tract's geometry from auditory and visual speech cues. In our recent work, which has culminated into a submitted paper during the reported period, we approach the problem in a statistical framework, combining ideas from multistream Hidden Markov Models and canonical correlation analysis, and demonstrate effective estimation of the trajectories followed by certain points of interest in the speech production system. Our experiments show that exploiting both audio and visual modalities clearly improves performance relative to either audio-only or visual-only estimation. We report experiments on the QSMT database which contains audio, video, and electromagnetic articulography data recorded in parallel.

### Publications

A. Katsamanis, G. Papandreou, and P. Maragos, "Audiovisual-to-Articulatory Inversion Using Hidden Markov Models", in Proc. IEEE Workshop on Multimedia Signal Processing (MMSP-2007), 2007.

## 5.10 Contribution by ICCS-NTUA

### Researchers involved

A. Potamianos (TSI-TUC) P. Maragos (ICCS-NTUA)

### Activities & Achievements

#### MUSCLE WP5 Workshop

A WP5 scientific meeting was organized in Chania, Crete, on October 4, 2007, as a satellite workshop to the IEEE Conference on Multimedia Signal Processing (MMSP-2007). The purpose of the meeting was to consolidate research activities of the Muscle partners involved in the WP5 showcase projects and e-teams, in view of the Muscle Plenary Meeting to be held in Paris, 29-30 Nov 2007 and the Final Review in the beginning of Feb. 2008. The participants had the opportunity to closely collaborate on the e-teams and showcase projects and discuss in detail both scientific progress and practical issues such as code development, dataset exchange and planning of joint publications. Importantly, the participants also had the opportunity to resolve issues related to editing the Muscle WP5 book.

## 5.11 Contribution by ICCS-NTUA

### Researchers involved

A. Zlatintsi, P. Maragos, K. Rapantzikos, G. Evangelopoulos (ICCS-NTUA) V. Moschou, C. Kotropoulos (AUTH)

### Activities & Achievements

#### Video-Clip Multimodal Annotation Protocol Development

For the purpose of creating a rich multimodal annotation framework in the context of integrated multimedia understanding, the two participating teams have developed a common XML-based annotation protocol for video-clips, which features (1) detailed audio classification and saliency description, (2) visual saliency description, and (3) dialogue analysis.

## 5.12 Contribution by IRIT-UPS

### Researchers involved

Julien Pinquier, Philippe Joly, Zein Ibrahim

### Activities & Achievements

During this period, we have achieved the project “CHAPITRE” (Automatic and Hierarchical Video Classification for an Interactive Platform of Digital TV Enriched), supported by ANR-RIAM (French Network for Research and Innovation in Audiovisual and Multimedia). In the audiovisual context, this work was aimed at improving structuring techniques. In fact, work has focused on our macrosegmentation method based on similarity matrix [1], and especially:

- “Low-level” characteristics (audio and video) weighting, based on their ability to prove whether or not the structure of a TV stream portion.
- Results organization in a hierarchical structure based on an average length of the a priori segments known or on the segments number used in the browser's interface.
- Treatment adaptation to take account of the constraints imposed by “real-time” processing of a stream.
- Method customization for the treatment of specific collections tested in the project on the content of diverse nature (TV news, TV games, movies and soccer match). These different proposed changes will be published soon.

[1] Siba Haidar, Philippe Joly and Bilal Chebaro. Mining for Video Production Invariants to Measure Style Similarity. In International Journal of Intelligent Systems, Wiley, V. 21 N. 7, p. 747-763, july 2006.

## 5.13 Contribution by TSI-TUC

### Researchers involved

Manolis Perakakis, Alexandros Potamianos

## **Activities & Achievements**

The efficiency and usage patterns of input modes in multimodal dialogue systems was investigated for both desktop and personal digital assistant (PDA) working environments. For this purpose a form-filling travel reservation application was evaluated that combines the speech and visual modalities; three multimodal modes of interaction were implemented, namely: "Click-To-Talk", "Open-Mike" and "Modality-Selection". The three multimodal systems were evaluated and compared with the "GUI-Only" and "Speech-Only" unimodal systems. Mode and duration statistics were computed for each system, for each turn and for each attribute in the form. Turn time was decomposed in interaction and inactivity time and the statistics for each input mode were computed. Results show that multimodal and adaptive interfaces were superior in terms of interaction time, but not always in terms of inactivity time. Also users tend to use the most efficient input mode, although our experiments show a bias towards the speech modality.

## **Publications**

1. M. Perakakis and A. Potamianos, "The effect of input mode on inactivity and interaction times of multimodal systems," in Internat. Conf. on Multimodal Interfaces, (Nagoya, Japan), Nov. 2007.
2. M. Perakakis and A. Potamianos, "A study in efficiency and modality usage in multimodal form filling systems," IEEE Transactions on Audio, Speech and Language Processing, submitted, 2007.

## **5.14 Contribution by UNIS**

### **Researchers involved**

B. Goswami, W.J. Christmas

### **Activities & Achievements**

We are implementing a system to study the main modes of lip shape variation. Vanilla PCA algorithms assume a robust calculation of the statistical properties of the input data. However, the Point Distribution Model paradigm accommodates erroneous outliers in the data being analysed. Consequently, a method needs to be implemented that robustly estimates the statistics of a PDM in order to eliminate outliers. We are developing two lip trackers. The first is based on a simple dynamic Bayesian network paradigm. The state evolution model is purely based on current state and optical flow calculations. The observation model is based on ellipse instantiation using available control points and subsequent lip boundary finding using a log-likelihood measure of the likely class characteristics of a given pixel. The second tracker is based on a Particle Filter. The state space it uses is the spline control points of a given lip shape and it has no spline regularisation or prior models incorporated into it.

## **Publications**

I Kolonias, J Kittler, WJ Christmas and F Yan, "Improving the accuracy of automatic tennis video", Proceedings of the International Conference on Image Analysis and Processing, September [www.ee.surrey.ac.uk/CVSSP/Publications/papers/kolonias-iciap-2007.pdf](http://www.ee.surrey.ac.uk/CVSSP/Publications/papers/kolonias-iciap-2007.pdf)

# **6 Overview activities in WP6**

## **6.1 Contribution by TUG**

### **Researchers involved**

Michael Grabener, Amir R. Saffari A. A., Horst Bischof

### **Activities & Achievements**

1. Initial results from visual object recognition problems seems to be promising. In most of the experiments, we could achieve the same performance of well-known clustering algorithms but with faster training and testing time. We will continue doing more experiments.
2. We improved our detection, recognition and tracking system. In particular feature computations have been further investigated and improved in performance. Apart work has been done according to efficiently solve the multiclass classification problem.

## **6.2 Contribution by ISTI-CNR**

### **Researchers involved**

Luigi Bedini, Sara Colantonio, Ercan Kuruoglu, Massimo Martinelli, Davide Moroni, Gabriele Pieri, Emanuele Salerno, Ovidio Salvetti, Anna Tonazzini

### **Activities & Achievements**

Within the activities of e-team 13, the assessment of the MCMC separation technique has continued by analysing real data from the WMAP satellite mission. The partial results obtained are being evaluated with the help of the astrophysical community. A joint paper is in preparation.

A study on dependent component analysis on remote-sensed images has been continued by evaluating the performances of the MaxNG algorithm under different nongaussianity measures. Our Correlated Component Analysis technique in the Fourier domain is being experimented on a reference sky model set up by the ESA's Planck mission collaboration. A research on source separation based on an MRF model with variational approximation is being carried out in collaboration with the Department of Electronic Engineering at Bogazici University, Istanbul, Turkey. Some results were presented at the EUSIPCO 2007 conference, Poznan, Poland.

An integrated information system has been developed for automatically detecting, localizing and monitoring forest fires. The system employs robotized stations equipped with combined infrared (IR) and visible cameras, suitably integrated. The final decision about the presence of a fire, within an area initially identified by visual processing the multisource imagery, is made by applying a set of fuzzy rules which are defined a priori and process a group of heterogeneous features. The system has been described in a paper presented at 9th International Workshop on Advanced Infrared Technology and Applications - AITA07. An ontology-driven approach to image understanding has been defined by introducing the suite of ontologies, including an algorithm ontology. To illustrate the validity of the approach, a case study on cell image analysis has been considered, introducing the corresponding domain ontologies. Preliminary results have been presented at 8th International Conference on Pattern Recognition and Image Analysis: New Information Technologies, PRIA-8 in October.

The work on decision support has been summarized in a paper accepted for the International Conference on Health Informatics (HEALTHINF 2008). An activity concerning the segmentation and interpretation of mammograms has started aimed at categorizing the breast dense tissue according to BI-RADS classification. The editorial activity for the Special issue on Human-Activity Analysis in Multimedia Data of the EURASIP Journal on Advances in Signal Processing, Guest Editors: E.A. Çetin, E. Pauwels, O. Salvetti, has been finalized.

## Events

- Anna Tonazzini and Emanuele Salerno participated in the 11th International Conference on Knowledge-based Intelligent Information and Engineering Systems, KES 2007, Vietri sul Mare, Italy, 12-14 September 2007, <http://kes2007.kesinternational.org/>.
- Emanuele Salerno chaired the Computational Learning Methods for Unsupervised Segmentation invited session at KES 2007 Digital Signal Processing Application Day 2007, Dipartimento di Informatica e Comunicazione, University of Milan, Milan (Italy), 17 September 2007. Master in Computational Science and Supercomputing, NEC-CESIC (Supercomputing Center for Computational Engineering), Cosenza (Italy), 23-24 October 2007.
- Dissemination of research carried out at ISTI by two invited seminars:
  - S. Colantonio, D. Moroni, O. Salvetti. Methods for Digital Image Processing. Seminar delivered to the DSP Application Day 2007, Dipartimento di Informatica e Comunicazione, University of Milan, Milan (Italy), 17 September 2007.
  - S. Colantonio, D. Moroni, O. Salvetti. Image Processing in Biomedical Applications. 4-hour long seminar delivered to the Master in Computational Science and Supercomputing, NEC-CESIC (Supercomputing Center for Computational Engineering), Cosenza (Italy), 23-24 October 2007

## Publications

- C. F. Caiafa, E. Salerno, A. N. Proto, “Blind source separation applied to spectral unmixing: comparing different measures of nongaussianity”, Lecture Notes in Artificial Intelligence 4694, Springer, 2007, pp. 1-8, doi: 10.1007/978-3-540-74829-8\_1.
- L. Bedini, E. Salerno, “Extracting Astrophysical Sources from Channel-Dependent Convolutional Mixtures by Correlated Component Analysis in the Frequency Domain”, Lecture Notes in Artificial Intelligence 4694, Springer, 2007, pp. 9-16, doi: 10.1007/978-3-540-74829-8\_2.

- E. Salerno, S. Wilson, "CLeMUS – Computational Learning Methods for Unsupervised Segmentation", Ercim News, No. 71, October 2007, p. 6.
- K. Kayabol, B. Sankur, E. Kuruoglu, "Source separation in images via MRFs with variational approximation", Proc. EUSIPCO 2007, Poznan, Poland, 4-8 September 2007.
- G. Pieri, O. Salvetti, "Fire Detection System based on integrated Infrared-visible Cameras", Proceedings of the 9th International Workshop on Advanced Infrared Technology and Applications - AITA07. Leon, Mexico, October 8 - 12, 2007.
- P. Asirelli, S. Colantonio, I.B. Gurevich, M. Martinelli, O. Salvetti, Yu. Trusova, "Ontology Driven Approach to Image Understanding", 8th Int. Conf. on Patt. Rec. & Im. An. New Inf. Tech. PRIA8, Yoshkar-Ola, Russ. Fed., 8-13 Oct. 2007, Vol. 1, 67-71

## **6.3 Contribution by ENSEA**

### **Researchers involved**

Prof. S. Philipp-Foliguet, J. Lebrun, Dr. P.-H. Gosselin, E. do Valle

### **Activities & Achievements**

Our work on kernels on graphs has been submitted to a French congress, Extraction et Gestion des Connaissances. It unifies and compares kernels initiated by Kashima, improves by Suard and our graph matching method FReBIR. This preliminary work has been pursued and will be submitted to main congresses. The PhD student E. Valle presented his work on knn for copy detection both in ICDAR, the main congress about document analysis in Curitiba in October and in ICHIM in Toronto in October. The work about segmentation evaluation is almost finished with a publication in revision.

### **Publications**

E. Valle, M. Cord, S. Philipp-Foliguet, "Matching High-Dimensional Local Descriptors for Image Identification on Cultural Databases", Brasil, ICDAR 2007. E. Valle, M. Cord, S. Philipp-Foliguet, "Content Based Image Identification in Cultural Databases", Toronto (Canada), ICHIM 2007.

## **6.4 Contribution by UPMC**

### **Researchers involved**

Matthieu Cord, David Picard

### **Activities & Achievements**

The papers " Kernel on Bags of Fuzzy Regions for fast object retrieval" and "Descriptor matching for image identification on cultural databases" have been presented at the IEEE ICIP conf., in San Antonio, Texas, September 2007 by Matthieu Cord. Matthieu Cord edited with Pdraig Cunningham the final version of the collective MUSCLE book on ML for Multimedia to be published by Springer. It has been sent to Springer for publication process. We made new developments for our distributed database retrieval system. This activity is connected

with the e-Team "Active and Semi-Supervised Learning". We also continued the collaboration with F. Precioso (ENSEA) and S. Philipp-Foliguet (ENSEA) on video content analysis.

## **6.5 Contribution by TU Vienna - IFS**

### **Researchers involved**

Rudolf Mayer, Thomas Lidy, Robert Neumayer, Jakob Frank, Andreas Rauber

### **Activities & Achievements**

In September, we presented both our new classification method called "Decision Manifolds" and our work done on "Component selection for the metro visualisation of the SOM" at the Workshop on Self-Organizing Maps (WSOM'07) in Bielefeld, Germany. Moreover, the Sky-Metaphor Visualisation for Self-Organising Maps was presented at the International Conference on Knowledge Management (I-KNOW'07) in Graz, Austria. The paper "Visualising Class Distribution on Self-Organising Maps" was presented at the International Conference on Artificial Neural Networks (ICANN'07), Porto, Portugal. In October, we released a first version of the web service for the training of Self-Organising Music Maps which allows the online creation of Music Maps.

### **Events**

6th International Workshop on Self-Organizing Maps (WSOM'07), Bielefeld, Germany, September 3 - 6 2007.

7th International Conference on Knowledge Management (I-KNOW'07), Graz, Austria, September 5 - 7 2007.

International Conference on Artificial Neural Networks (ICANN'07), Porto, Portugal, September 9 - 13 2007.

### **Publications**

- Robert Neumayer, Rudolf Mayer, and Andreas Rauber. Component selection for the metro visualisation of the SOM. In Proceedings of the 6th International Workshop on Self-Organizing Maps (WSOM'07), Bielefeld, Germany, September 3-6 2007.

- Georg Pözlbauer, Thomas Lidy and Andreas Rauber. Decision Manifolds: Classification Inspired by Self-Organization. In Proceedings of the 6th International Workshop on Self-Organizing Maps, Bielefeld, Germany, September 3-6, 2007

- Khalid Latif and Rudolf Mayer. Sky-Metaphor Visualisation for Self-Organising Maps. In Proceedings of the 7th International Conference on Knowledge Management (I-KNOW'07), Graz, Austria, September 5 - 7 2007.

- Rudolf Mayer, Taha Abdel Aziz, and Andreas Rauber. Visualising Class Distribution on Self-Organising Maps. In Proceedings of the International Conference on Artificial Neural Networks (ICANN'07), Porto, Portugal, September 9 - 13 2007. Springer Verlag.



## 6.6 Contribution by INRIA-Ariana

### Researchers involved

Ian Jermyn.

### Activities & Achievements

INRIA Ariana has a INRIA Associated Team project 'Shapes' with Professors Anuj Srivastava and Victor Patrangenu of the Statistics Department of Florida State University. As part of this collaboration, Ian Jermyn has been working with Professor Srivastava on new Riemannian metrics on shape spaces derived from the Fisher-Rao metric on the space of (probability) measures. Recent work has shown that the one-parameter family of 'elastic' metrics on the space of curves leads to flat spaces in two dimensions for all values of the parameter, but that in higher dimensions, the space is curved except for a unique value of the parameter, the same for all dimensions  $> 2$ . Euclidean coordinates have been found for all flat cases. For all values of the parameter, the spaces are conformally flat. In some cases, this leads to explicit analytical expressions for geodesics on shape spaces, that is, quotients of the space of parameterized curves by some symmetry group. This improves greatly the computation of such geodesics, which had previously only been achieved numerically.

## 6.7 Contribution by Technion - ML

### Researchers involved

Michael Lindenbaum, Shaul Markovitch, Saher Esmeir, Ido Leichter

### Activities & Achievements

Kernel-based trackers aggregate image features within the support of a kernel (a mask) regardless of their spatial structure. These trackers spatially fit the kernel (usually in location and in scale) such that a function of the aggregate is optimized. We propose a kernel-based visual tracker that exploits the constancy of color and the presence of color edges along the target boundary. The tracker estimates the best affinity of a spatially aligned pair of kernels, one of which is color-related and the other of which is object boundary-related. In a sense, this work extends previous kernel-based trackers by incorporating the object boundary cue into the tracking process and by allowing the kernels to be affinely transformed instead of only translated and isotropically scaled. These two extensions make for more precise target localization. Moreover, a more accurately localized target facilitates safer updating of its reference color model, further enhancing the tracker's robustness. The improved tracking is demonstrated for several challenging image sequences.

### Events

Saher Esmeir is attending NIPS 2007 to present our paper on cost-sensitive learning. Ido Leichter has attended ICCV 2007 to present our paper on Kernel-based tracking. Saher Esmeir attended the international workshop on constraint-based mining and learning at ECML/PKDD 2007 at Warsaw.

## **Publications**

- Visual Tracking by Affine Kernel Fitting Using Color and Object Boundary, Leichter, I. and Lindenbaum, M. and Rivlin, E., In Proc. IEEE International Conference on Computer Vision (ICCV'07), 2007.
- Evgeniy Gabrilovich and Shaul Markovitch. Harnessing the Expertise of 70,000 Human Editors: Knowledge-Based Feature Generation for Text Categorization. Journal of Machine Learning Research, 8:2297-2345, 2007
- Saher Esmeir and Shaul Markovitch. Anytime Learning of Decision Trees. Journal of Machine Learning Research, 8:891-933, 2007

# **7 Overview activities in WP7**

## **7.1 Contribution by TAU-VISUAL**

### **Researchers involved**

Mr. Tomer AMIAZ (TAU-VISUAL), Prof. Nahum KIRYATI (TAU-VISUAL), Dr. Sandor FAZEKAS (MTA-SZTAKI), Prof. Dmitry CHETVERIKOV (MTA-SZTAKI)

### **Activities & Achievements**

We completed the writing of a full-length journal article on the recognition and segmentation of dynamic textures in video sequences. The camera is not assumed to be static, so the proposed algorithm has to discriminate between motion involving dynamic textures and other types of motion. The experimental section includes many successful segmentations of smoke, fire and flowing water. The proposed method can be simplified to obtain real-time operation. It is the basis of our contribution to the dynamic texture showcase.

## **7.2 Contribution by Bilkent University**

### **Researchers involved**

A. Enis Cetin, Kivanc Kose, Erdem Dengel, Yigithan Dedeoglu, A. Enis Cetin

### **Activities & Achievements**

We participated CeBIT Bilisim Eurasia (Enformatics Fair) which was held in Istanbul Turkey at 2-7 October 2007. It was an important fair for creating a gateway to business success between the East and the West. 908 Companies from 20 Countries participated in this fair. It was organized by Hannover Fairs Interpro Inc. We demonstrated the posters and videos of MUSCLE showcases and distributed the flyers of the showcases in the fair. We made a general presentation of the MUSCLE project. We also made a live demonstration of dynamic texture detection in video. As a dissemination activity, the fair increased the visibility of the project in industry.

## Events

CeBIT Bilisim Eurasia (Enformatics Fair), Istanbul, Turkey, 2-7 October 2007.

## 7.3 Contribution by Bilkent University, INRIA, KTH, SZTAKI, Technical University of Crete, Technical University of Wien

### Researchers involved

Kivanc Kose, Erdem Dengel, Elias Iosif, Thomas Lidy, Helmut Berger, Istvan Petras, Preben Wik, Julian Stoettinger, Laurent Joyeux.

### Activities & Achievements

We participated in IBC 2007 (Visual Media Technologies Fair) held in Amsterdam, the Netherlands between 6-11 September 2007. We had a booth at "New Technology Campus" hall of the exhibition. In this hall innovative research projects of both EU and other individual universities were demonstrated. For many people this hall was the place at which the future technologies are shown. The attention to our stand was satisfactory. Each day 3 MUSCLE showcases were demonstrated. The following demonstrations were made at the exhibition.

- Day-1 (07.09.2007)
  1. PlaySOM
  2. Movie Summarization and Real Time Audio-Visual Automatic Speech Recognition Demonstration
  3. Dynamic Texture Detection
- Day-2 (08.09.2007)
  1. PlaySOM
  2. Movie Summarization and Real Time Audio-Visual Automatic Speech Recognition Demonstration
  3. Shaping 3D Environments and Unusual Behavior Detection
- Day-3 (09.09.2007)
  1. Articulatory Talking Head
  2. ACADI
  3. Shaping 3D Environments and Unusual Behavior Detection
- Day-4 (10.09.2007)
  1. Articulatory Talking Head
  2. ACADI
  3. Object Recognition and Video and Image Copy Detection
- Day-5 (11.09.2007)
  1. Dynamic Texture Detection
  2. Video and Image Copy Detection
  3. Object Recognition

## Events

IBC'07, 6-11 September 2007, Amsterdam, The Netherlands.

## **7.4 Contribution by VTT**

### **Researchers involved**

Sanni Siltanen, Petri Honkamaa

### **Activities & Achievements**

Demonstration and Poster at MMSP 2007 conference of the Showcase "Augmented Assembly Using a Multimodal Interface". Further development of the Showcase demo.

### **Events**

MMSP 2007 conference

### **Publications**

"Multimodal User Interface for Augmented Assembly" Siltanen, Sanni; Hakkarainen, Mika; Korkalo, Otto; Salonen, Tapio; Säski, Juha; Woodward, Charles; Kannetis, Theofanis; Perakakis, Manolis; Potamianos, Alexandros IEEE International Workshop on Multimedia Signal Processing ,MMSP2007. Chania, Crete, Greece, 1-3 October, 2007 (2007)

## **7.5 Contribution by TCD**

### **Researchers involved**

Rozenn Dahyot

### **Activities & Achievements**

I gave a seminar at Intel Shannon (Ireland) on the 4th of October on works done by the group GV2 (<http://gv2.cs.tcd.ie/>) in TCD, in computer vision. In particular, I presented the research in Multimedia understanding sponsored by the project MUSCLE including sport indexing, illicit content detection, video restoration for film database and colonoscopy videos, and also some applications based on audio-visual process and computer graphics for entertainment.

## **7.6 Contribution by TU VIENNA-PRIP**

### **Researchers involved**

Julian Stöttinger, Allan Hanbury

## **Activities & Achievements**

The object recognition showcase was demonstrated at the IBC in Amsterdam on the 10th and 11th of September.

Julian Stöttinger spent October at the University of Amsterdam to continue work on this showcase and integrate the results of the groups.

The work to improve the classification performance and the repeatability of the colour interest points has continued. In addition, the number of objects that can be recognised has been increased to 20.

The demonstrator was also used to participate in the object retrieval task of the ImageVAL 2007 campaign.

## **Events**

Demonstration at the IBC in Amsterdam on the 10th and 11th of September 2007.

## **7.7 Contribution by GET-ENST**

### **Researchers involved**

Beatrice Pesquet-Popescu, Maria Trocan, Enis Cetin (Bilkent University)

### **Activities & Achievements**

The common work on dynamic textures, between GET-ENST and Bilkent University, continued. Prof Enis Cetin had a visit at GET-ENST, including the participation to the PhD defence of Maria Trocan, who had a joint work with Bilkent University during the last years. It was also the opportunity to discuss on-going work on issues like indexing through compression, fire detection from temporal representation of motion vector fields, and copy detection.

## **7.8 Contribution by TSI-TUC**

### **Researchers involved**

Alexandros Potmianos (TSI-TUC), Petros Maragos (ICCS-NTUA)

### **Activities & Achievements**

Organized the IEEE Workshop on Multimedia Signal Processing (MMSP 2007, Oct 1st-Oct 3rd, Chania, Crete, Greece) as well as a demo session including relevant MUSCLE showcases (Oct 2nd, Chania, Crete, Greece). The demo session included MUSCLE showcases on movie summarization, augmented reality interfaces (see also MMSP publication below) and the audio-visual speech recognition showcase. This was work related to cross-modal processing and multimodal interfaces that fit well under the charter of MMSP.

## **Publications**

S. Siltanen et al, "Multimodal user interface for augmented assembly," in Proc. Intern. Workshop on Multimedia Signal Processing, (Chania, Greece), Oct. 2007.

## **7.9 Contribution by TU Vienna - IFS**

### **Researchers involved**

Thomas Lidy, Jakob Frank, Rudolf Mayer, Andreas Rauber

### **Activities & Achievements**

We presented our work on audio analysis, content processing and music maps at the International Broadcasting Convention (IBC 2007, a Visual Media Technologies Fair) in Amsterdam, the Netherlands, between September 6-11 2007. Live-demos were presented to people interested, at our MUSCLE stand in the "New Technology Campus" exhibition hall. We organized and hosted the 8th International Conference on Music Information Retrieval (ISMIR 2007) in Vienna, Austria, from September 23 to 27, for which much organization and preparation was necessary. In October a researcher from the University of Thessaloniki (AUTH) visited our group at TU Vienna IFS. Work was done within the MUSCLE Evaluation Showcase and the e-Team Semantics from Audio.

### **Events**

IBC 2007, Amsterdam, The Netherlands, September 6-11 2007. 8th International Conference on Music Information Retrieval (ISMIR 2007), Vienna, Austria, September 23 - 27 2007. Exchange visit from University of Thessaloniki (AUTH) at TU Vienna IFS.

## **7.10 Contribution by INRIA-Ariana**

### **Researchers involved**

Josiane Zerubia, Ian Jermyn.

### **Activities & Achievements**

In September, Professors Anuj Srivastava and Victor Patrangenaru of the Statistics Department of Florida State University visited INRIA-Ariana for a week each as part of the INRIA Associated Team project 'Shapes', and research continued on the Associated Team topics, which fall within Muscle's research ambit. Peter Horvath and Josiane Zerubia both attended EUSIPCO 2007, while Ting Peng attended BMVC 2007; all presented Muscle-related work. Josiane Zerubia and Ian Jermyn presented Muscle-related work to the French Space Agency CNES as part of a contract. The PhD work of Aymen El Ghoul is funded by this contract. INRIA-Ariana has submitted a joint proposal, with MBDA France, a division of EADS, to the EADS Foundation to pay for a Masters student to work on a joint project.

## 8 MUSCLE Effort Table

Person-Months report for period :

September - October 2007

Institute	WP1a	WP1b	WP2	WP3	WP4	WP5	WP6	WP7	Total
01 - ERCIM	2.1	0	0	0	0	0	0	0	2.1
03 - UCL	0	0	0	2	0	0	0	0	2
04 - KTH	0	0	0	0	0.2	0.7	0	0	0.9
05 - BILKENT	0	0.2	1.2	0	0.2	0.2	0.2	4.25	6.25
06 - VIENNA PRIP	0	0	0.3	2	0	0	0	2	4.3
07 - MTA SZTAKI	0	0	0	0.5	0	0	1.3	1.5	3.3
09 - CNR-ISTI	0	0.32	0.64	0.16	0.32	0	0.32	0	1.76
11 - TUG	0	0.1	0	2.2	0	0	1.5	0.2	4
12 - UPC	0	0	0	1.5	0	0	0	0	1.5
16 - AUTH	0	0	0	1.23	0.93	1.94	0	0	4.1
18 - TU VIENNA IFS	0	0	0	0	1.7	0	2	1.8	5.5
19 - ACV	0	0	0	0.8	0	0	0	0	0.8
25 - ARMINES	0	0.1	0	1.5	0	0	0	0	1.6
26 - TAU-SPEECH	0	0	0	0.5	2.5	0	0.3	0	3.3
27 - TAU-VISUAL	0	0	0	1	0	0	0	2	3
31 - VTT	0	0	0	0	0	0.5	0	0.7	1.2
32 - INRIA Ariana	0	0.26	0	0.53	0	0	0.44	0.44	1.67
32 - INRIA Tex Mex	0	0	0	0	0	1.2	0	0	1.2
33 - GET	0.1	0	0	1.4	0	0	0	1.7	3.2
37 - ENSEA	0	0	0	0.35	0	0	0.35	0	0.7
38 - CNRS	0	0	0	0	1.5	0	0	0	1.5
39 - UPS – IRT	0	0	0	1	1	1	0	1	4
41 - UPMC	0	0.3	0	0	0	0	0.6	0	0.9
42 - NUID / UCD	0	0.1	0	0	0	0	2.1	0	2.2
<b>Total</b>	<b>2.2</b>	<b>1.38</b>	<b>2.14</b>	<b>16.67</b>	<b>8.35</b>	<b>5.54</b>	<b>9.11</b>	<b>15.59</b>	<b>60.98</b>