



MUSCLE

Network of Excellence

Multimedia Understanding through Semantics, Computation and Learning

Project no. FP6-507752

Bi-Monthly Progress Report #12: Jan – Feb 2006

Due date of deliverable: 15.04.2006

Actual submission date: 20.04.2006

Start date of Project: 1 March 2004

Duration: 48 Months

Name of responsible editor(s):

- Eric Pauwels (eric.pauwels@cwi.nl), Remi Ronchaud (remi.ronchaud@cwi.nl)

Revision: 1.0

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

Keyword List:

Contents

1 Overview activities in WP 1	3
2 Overview activities in WP 2	3
3 Overview activities in WP 3	3
4 Overview activities in WP 4	5
5 Overview activities in WP 5	6
6 Overview activities in WP 6	23
7 Overview activities in WP 7	27
8 Overview activities in WP 8	29
9 Overview activities in WP 9	35
10 Overview activities in WP 10	36
11 Overview activities in WP 11	38
12 MUSCLE Publications: Papers and Preprints	40
13 Resource Tables	46

1 Overview activities in WP 1

1.1 Contribution by ERCIM and CWI

Researchers involved: Remi Ronchaud, Eric Pauwels, Valrie Gouet-Brunet, Nozha Boujema

Activities

- Administrative and financial coordination of the network
- Launch and Stimulation of integration activities (mobility, e-teams support, conference sponsorship, joint papers)
- Organisation of the joint DELOS-MUSCLE summerschool on *Multimedia Digital Libraries: Machine Learning and Cross-modal Technologies for Access and Retrieval*, San Vincenzo, 12-17 June 2006 (in collaboration with Nozha Boujema, Valerie Gouet-Brunet)
- Organisation of MUSCLE's 4th Scientific Meeting (Istanbul, 15-17 Feb, 2006) together with Enis Cetin.
- Editing the Newsletter, follow-up and posting of the deliverables
- Preparation of first draft of JPA3-document
- Preparation of Second Annual Review
- Organisation of regular audio conferences
- Reimbursement of MUSCLE integration expenses (mobility support grants)

2 Overview activities in WP 2

2.1 Contribution by CWI

Researchers involved: Eric Pauwels, Paul de Zeeuw, Margriet Brouwer

Activities Continued development of PHP/MySQL tools for administrative and scientific reporting in WPs. We mention in particular:

- Showcases and Newsletters;
- Deliverables and Paper Archive;
- Events agenda and Istanbul talks online;
- Mobility grants and reporting.

3 Overview activities in WP 3

3.1 Contribution by TUVienna-PRIP

Researchers involved: Allan Hanbury,

Activities Organising MUSCLE evaluation activities in 2006. These include:

- CIS coin competition (in collaboration with ARCS)
- Some tasks in the ImageCLEF evaluation campaign

Two MUSCLE-sponsored workshops related to these activities are also being organised. The CIS competition result workshop will take place on the 11th of September in Berlin, Germany (before the DAGM conference). The MUSCLE/ImageCLEF 2006 workshop will take place on the 19th of September in Alicante, Spain (before the CLEF workshop).

Publications MP-codes:

3.2 Contribution by TUVienna-PRIP

Researchers involved: Allan Hanbury, Lech Szumilas, Branislav Micusik

Activities Ground truth generation: Members of the E-team *Selecting features for CBIR and Automated Image Annotation* have been providing ground truth segmentation of animal images from the Corel database. In addition, each of the 60000 images in the James Wang collection has been labelled as containing an animal or not.

Publications MP-codes:

3.3 Contribution by UTIA

Researchers involved: Michal Haindl, S.Mikes, J.Filip

Activities The Prague Texture Segmentation Datagenerator and Benchmark (PTSB) was modified to speed up its performance and to improve its interface. PTSB was generalized for supervised classifiers evaluation and to test the noise resistance of single classifiers or segmenters. For all 100 textures used in the benchmark texture mosaics the corresponding separate trainee textures are available. Additional 15 test criteria were added as well as LaTeX for export of resulting evaluation tables. Five additional segmentation algorithms were tested on the benchmark. Among them a method developed in cooperation with INRIA Ariana.

A common work on dynamic texture modelling with SZTAKI using the MUSCLE DynTex database resulted in the article submitted for ICPR 2006 conference.

Publications MP-codes: 315

3.4 Contribution by INRIA-Imedia

Researchers involved: Valrie Gouet-Brunet, Nozha Boujema, Marin Ferecatu, Alexis Joly

Activities Imedia is involved in the French project ImagEval (Technovision program), and more particularly in task 1 (recognition of transformed images), in task 4 (detection of objects) and in task 5 (extraction of semantics).

Publications MP-codes:

3.5 Contribution by INRIA-Ariana

Researchers involved: Ian Jermyn, Josiane Zerubia, Avik Bhattacharya

Activities The work of Avik Bhattacharya, PhD student in INRIA-Ariana, concerns the use of new road network features for retrieval from remote sensing image databases. Part of this work has involved the extraction, both manually and using algorithms developed by INRIA-Ariana, of road networks from a remote sensing image collection. The resulting image/road network pairs may be used for testing road network extraction methods and image retrieval using road networks.

Publications MP-codes:

4 Overview activities in WP 4

4.1 Contribution by UPC

Researchers involved: Montse Pardas, Jose Luis Landabaso, Camilo Dorea, Veronica Vilaplana, Cristian Canton

Activities Submission to ICIP 2006 of papers 211, 212, 213 and 271.

Publications MP-codes: 211, 212, 213, 271

4.2 Contribution by INRIA-Vista

Researchers involved: Ivan Laptev, Patrick Bouthemy

Activities Organization of special session

We have launched a special session *Content Analysis and Representation* within International Workshop on Multimedia Content Representation, Classification and Security (IWMRCS) that will be held in Istanbul, September 2006.

Publications MP-codes:

4.3 Contribution by INRIA-Imedia

Researchers involved: Valerie Gouet-Brunet, Nozha Boujemaa

Activities

- Organization of the first DELOS-MUSCLE Summer school on "Multimedia Digital Libraries: Machine Learning and Cross-modal Technologies for Access and Retrieval", San Vincenzo, 12-17 June 2006.
- Nozha Boujemaa was invited by the Unit D2 to a consultation workshop that serves for the guideline discussion of the present call6 on audio-visual search engines: http://www.cordis.lu/ist/audiovisual/neweve/e/ws150905/ws150905_2.htm (15 September 2005, Bruxelles)

Publications MP-codes:

4.4 Contribution by INRIA-Ariana

Researchers involved: Ian Jermyn, Josiane Zerubia

Activities Dissemination activities include: invited presentation on Muscle research at the *Workshop on Mathematics and Imaging* at the University of Nice; creation of Shape modelling e-team web site. Training activities include direction of the PhD students, interns and postdocs in INRIA-Ariana. In particular, Giuseppe Scarpa, joint ERCIM postdoc with UTIA, currently at INRIA-Ariana, made a successful application for a permanent position as Researcher at the Electronic and Telecommunications Department of the 'Federico II' University in Naples, Italy.

Publications MP-codes:

5 Overview activities in WP 5

5.1 Contribution by TUVienna-PRIP

Researchers involved: Allan Hanbury,

Activities Collaboration with Beatriz Marcotegui of ARMINES-CMM in the framework of the E-team on "Choosing Features for CBIR and Image Annotation" on the following topics:

- Segmentation of images using the waterfall algorithm on colour-texture gradients
- Matching of images using 2D colour histograms

Publications MP-codes: 181

5.2 Contribution by TUVienna-PRIP

Researchers involved: Allan Hanbury, Lech Szumilas, Branislav Micusik

Activities E-team on "Choosing Features for CBIR and Automated Image Annotation". Collaboration with Beatriz Marcotegui of ARMINES-CMM in the framework of the E-team on "Choosing Features for CBIR and Image Annotation" on the following topics:

- Segmentation of images using the waterfall algorithm on colour-texture gradients
- Matching of images using 2D colour histograms

Two MUSCLE mobility grants have been used in this collaboration. Lech Szumilas visited the CMM in October 2005 for a course on morphological segmentation. Allan Hanbury visited the CMM in April 2006 to continue work on the above topics.

Alireza Tavakoli Targhi (KTH) visited PRIP using a MUSCLE mobility grant in March 2006. Work was done on texture features and automatic scale detection for animal recognition.

Publications MP-codes: 181

5.3 Contribution by TUVienna-PRIP

Researchers involved: Allan Hanbury, Lech Szumilas, Branislav Micusik and Masters students

Activities Development of image segmentation algorithms which segment an image based on a sample of the texture to be found. This sample must be specified in advance by the user. The algorithm then attempts to mark all the regions in the image which correspond to the specified texture. It makes use of the minimum cut/maximum flow algorithm on a graph representation of the image. This problem is an instance of the one-class classification problem, as we have information on the texture to be located, but no information on the "background" (the rest of the image). This algorithm has been further developed to be fully automatic, yielding good results which will be presented at the ECCV 2006. Improving the efficiency of this method by creating a graph based on the results of a Maximally Stable Extremal Region (MSER) detector instead of directly on the pixels is currently being investigated.

A texture detector for automatically locating textures in an image has been developed. It tries to find alternating colour patterns within an image using a clustering technique. The detected texture patches are then passed to the segmentation algorithm described above to attempt to locate the whole texture. An algorithm for classifying images into categories such as city/nature and inside/outside has been developed. It has been entered into the ImagEVAL evaluation campaign.

We have also investigated the keywords which have been used to annotate images in currently available image datasets. This has resulted in the creation of a combined keyword list.

Publications MP-codes: 200,204,218,291,292

5.4 Contribution by AUTH

Researchers involved: Constantine Kotropoulos, Nikoletta Bassiou, Athanasios Papaioannou

Activities Task 4: Text and natural language processing: Word and document clustering

Two methods for interpolating the distanced bigram language model are examined which take into account pairs of words that appear at varying distances within a context. The language models under study yield a lower perplexity than the baseline bigram model. A word clustering algorithm based on mutual information with robust estimates of the mean vector and the covariance matrix is employed in the proposed interpolated language model. The word clusters obtained by using the aforementioned language model are proved more meaningful than the word clusters derived using the baseline bigram. A novel method for updating probabilistic latent semantic indexing (PLSI) when new documents arrive has been developed. The proposed method adds incrementally the words of any new document in the term-document and derives the updating equations for the probability of terms given the class (i.e. latent) variables and the probability of documents given the latent variables. The performance of the proposed method is compared to that of the folding-in algorithm, which is an inexpensive but potentially inaccurate updating method. It is demonstrated that the proposed updating algorithm outperforms the folding-in method with respect to the mean squared error between the aforementioned probabilities as they are estimated by the two updating methods and the original non-adaptive PLSI algorithm. A paper on this topic has been submitted to the 4th Hellenic Conference on Artificial Intelligence.

Publications MP-codes: 81,231

5.5 Contribution by AUTH

Researchers involved: Ioannis Pitas, Costas Cotsaces, Nikolaos Nikolaidis

Activities Face indexing in video (Task 2: Image and video processing; Sub-task 1. Low-level feature extraction for visual content description).

A semantic video signature can be defined as a video signature based on high-level content information rather than on low-level features of the video stream. Its major advantage is that it is invariant to nearly all types of distortion. A major semantic feature of a video is the appearance of specific people in specific frames. Because of the great amount of research that has been performed on the subject

of face detection and recognition, the extraction of such information is nowadays generally feasible. AUTH has developed an indexing method that uses the pre-extracted output of face detection and recognition modules to perform fast semantic indexing and retrieval of video segments. The biggest advantage of the proposed approach is that the evaluation of similarity is convolution-based, and is thus resistant to perturbations in the signature and independent of the exact boundaries of the query segment. The extraction of face information is not dealt with in this work, since ample work has been performed on the subject. The proposed method tries to solve the problems of consistency and continuity with regards to face-based indexing, to represent face information with minimal redundancy, and also to find a fast (logarithmic-time) search method. The developed method is based on reducing the information extracted regarding the existence of persons in a video to only four values:

- The time instant a person's face starts being visible/recognizable in the video.
- The time instant a person's face stops being visible/recognizable in the video.
- On average, how recognizable this person's face is in this interval (which closely related to the size of the face on the video)
- Finally, the identity of the person we are dealing with (we assume the existence of a face recognition module along with the face detection module).

Each such quartet is considered as a signature element, or sigel for short. In effect, the signature of the video consists of a set of pulse series (face indicator functions), with each pulse series of the set corresponding to one person and each pulse corresponding to a sigel. In order to detect the similarity of two video segments, an integration on the pulse series obtained by evaluating the minimum (at each time instant) between two corresponding pulse series is performed. Additionally, if we want to search for the best fit between two video segments, we need to find where this integral has its maximum value when sliding one pulse series over the other. This is very similar to a convolution, except that instead of product we use the minimum operator. A significant part of the work performed so far deals with the organization of the search algorithm so that it performs with near-logarithmic efficiency with respect to video size. This is achieved by utilizing an RDBMS, and using indexes on persons and temporal intervals. It is factually impossible to process (i.e. perform face detection and recognition) the very large volume of videos (several thousand hours) which is necessary in order to test the efficiency of the proposed method, either manually or automatically. In addition, automatic processing would not allow us to control/modify the noise introduced in the semantic signatures due to face detection & recognition errors. For these reasons artificial data for the appearance and identity of persons (i.e. faces) have been generated. To achieve this, we have formulated a probabilistic model based on the analysis of the motion picture production process. We have set the parameters of this model using statistics extracted from a corpus of real video data. Preliminary results have demonstrated real-time performance in a database of 10000 artificial videos containing approximately 108 signature elements, and general robustness with respect to noise introduced from simulated face detector/recogniser errors. The extraction of a digital signature from a video segment in order to uniquely identify it, is often a necessary prerequisite for video indexing, copyright protection, and other tasks. Semantic video signatures are those that are based on high-level content information rather than on low-level features of the video stream, their major advantage being that they are invariant to nearly all types of distortion. Since a major semantic feature of a video is the appearance of specific people in specific frames, we have developed a method that uses the pre-extracted output of face detection and recognition to perform fast semantic indexing and retrieval of video segments. We give the results of the experimental evaluation of our method on an artificial database created using a probabilistic model of the creation of video.

Publications MP-codes: 82,232

5.6 Contribution by AUTH

Researchers involved: Constantine Kotropoulos, Euthymius Ziogas, Maria Marinaki

Activities Detection of voice disorders (Task 3: Audio and speech processing; Sub-task 3: Speech analysis)

A combined scheme of linear prediction analysis is proposed for feature extraction along with linear projection methods for feature reduction followed by known pattern recognition methods on the purpose of discriminating between normal and pathological voice samples. Two different cases of speech under vocal fold pathology are examined: vocal fold paralysis and vocal fold edema. Three known classifiers are tested and compared in both cases, namely the Fisher linear discriminant, the K-nearest neighbor classifier, and the nearest mean classifier. The performance of each classifier is evaluated in terms of the probabilities of false alarm and detection or the receiver operating characteristic. The datasets used are part of a database of disordered speech developed by Massachusetts Eye and Ear Infirmary. The experimental results indicate that vocal fold paralysis and edema can easily be detected by any of the aforementioned classifiers.

A two-class pattern recognition problem has been studied, namely the automatic detection of speech disorders such as vocal fold paralysis and edema by processing the speech signal recorded from patients affected by the aforementioned pathologies as well as speakers unaffected by these pathologies. The data used were extracted from the Massachusetts Eye and Ear Infirmary database of disordered speech. The linear prediction coefficients are used as input to the pattern recognition problem. Two techniques are developed. The first technique is an optimal linear classifier design, while the second one is based on the dual-space linear discriminant analysis. Two experiments were conducted in order to assess the performance of the techniques developed namely the detection of vocal fold paralysis for male speakers and the detection of vocal fold edema for female speakers. Receiver operating characteristic curves are presented. Long-term mean feature vectors are proven very efficient in detecting the voice disorders yielding a probability of detection that may approach 100% for a probability of false alarm equal to 9.52%.

Publications MP-codes: 69,233

5.7 Contribution by AUTH

Researchers involved: Ioannis Pitas, Costas Cotsaces, Nikos Nikolaidis

Activities Video shot boundary detection (Task 2: Image and video processing; Sub-task 5: Image sequence features).

A review of basic information extraction operations that can be performed on video is presented in this paper. Specifically, the review focuses on shot boundary detection and condensed video representation (also called summarization and abstraction). Shot boundary detection is the complete segmentation of a video into continuously imaged temporal video segments. Condensed video representation is the extraction of video frames or short clips that are either semantically representative of the corresponding video. Both tasks are very significant for the organization of video data into more manageable forms. An overview of the fundamental issues in each task is provided, and recent work on the subject is described and is critically reviewed.

Publications MP-codes: 237

5.8 Contribution by AUTH

Researchers involved: Constantine Kotropoulos, Dimitrios Ververidis

Activities Emotional speech analysis and classification (Task 3: Audio and speech processing; Sub-task 5: Events detection, segmentation, and classification for audio streams).

Emotional speech recognition aims to automatically classify speech units (e.g., utterances) into emotional states, such as anger, happiness, neutral, sadness and surprise. We have rated the discriminating capability of a set of features for emotional speech recognition when gender information is taken into

consideration. A total of 87 features have been calculated over 500 utterances of the Danish Emotional Speech database. The Sequential Forward Selection method has been used in order to discover the 5-10 features which are able to classify the samples in the best way for each gender. The criterion used in SFS is the crossvalidated correct classification rate of a Bayes classifier where the class probability distribution functions (pdfs) are approximated via Parzen windows or modeled as Gaussians. When a Bayes classifier with Gaussian pdfs is employed, a correct classification rate of 61.1% is obtained for male subjects and a corresponding rate of 57.1% for female ones. In the same experiment, a random classification would result in a correct classification rate of 20%. When gender information is not considered a correct classification score of 50.6% is obtained.

A novel method to control the number of crossvalidation repetitions in sequential forward feature selection algorithms has been proposed. The criterion for selecting a feature is the probability of correct classification achieved by the Bayes classifier when the class feature probability density function is modeled by a single multivariate Gaussian density. Let the probability of correct classification achieved by the Bayes classifier be a random variable. We demonstrate by experiments that the probability density function of the latter random variable can be modeled by a Gaussian density. Based on this observation, a method for reducing the computational burden in Sequential forward selection algorithms is proposed. The method predicts the number of crossvalidation repetitions by employing a t-test to guarantee that a statistically significant improvement in the probability of correct classification is obtained by increasing the number of selected features. The proposed method is twice faster than the sequential forward selection algorithm that uses a fixed number of crossvalidation repetitions and it maintains the performance of the sequential floating forward selection algorithm.

Publications MP-codes: 68,81

5.9 Contribution by AUTH

Researchers involved: Ioannis Pitas, Irene Kotsia

Activities Dynamic facial expression recognition using support vector machines (Task 2: Image and video processing; Sub-task 5: Image sequence features).

The system consists of two subsystems, one for information extraction and one for information classification. The information extraction has been performed by a tracking system, developed based on deformable models. The system is semi-automatic, in the sense that the user has to manually place in the beginning some of the Candide wireframe model grid points to a face depicted at the first frame of the image sequence under examination. The wireframe node tracking is performed by a pyramidal variant of the well-known Kanade-Lucas-Tomasi (KLT) tracker. The loss of tracked features is handled through a model deformation procedure that increases the robustness of the tracking algorithm. The tracking system follows the facial expression evolving through time to reach its highest intensity, thus producing the grid that corresponds to it. A subset of the Candide grid points that is responsible for the formation of the movements described by the Facial Action Coding System (FACS) is chosen. The number of grid nodes finally used, is equal to 62. The facial expressions to be recognized are 6, as defined by the psychologists (anger, disgust, fear, happiness, sadness and surprise). The geometrical displacements of grid nodes, defined as the difference of each point's coordinates between the first and the last frame of the image sequence are collected.

For information classification, a six-class SVM system is used where each class corresponds to one of the 6 basic facial expressions to be recognized. The SVM system's kernel used is a 3rd degree polynomial function. Experiments were performed using the Cohn-Kanade database, using the leave-one-out method and the results indicate an accuracy of 97.75% with a speed of 20 frames per second for testing, therefore making it suitable for real-time applications.

Publications MP-codes: 77

5.10 Contribution by AUTH

Researchers involved: Ioannis Pitas, Stylianos Asteriadis, Nikos Nikolaidis

Activities Task 2: Image and video processing, Sub-task 1: Low-level feature extraction for visual content description.

Work was conducted for eye and lip area detection using geometrical information. A method for eye detection and eye center localization has been developed. The work on this method is a continuation and extension of work performed in this topic in the past. According to this method, edge detection is performed on facial regions and a vector pointing to the closest edge pixel is assigned to every pixel. Length and slope information for these vectors is used to detect the eyes. For eye center localization, edge and intensity information is used. The proposed method can work on low-resolution images and has been tested on two face databases with very good results. Work is continuing towards producing a variant of this method that can detect mouth regions.

Publications MP-codes:

5.11 Contribution by AUTH

Researchers involved: Constantine Kotropoulos, Emmanouil Benetos, Tomas Lidy, Andreas Rauber

Activities Task 3: Audio and speech processing; Sub-task 3: Speech analysis; Sub-task 5: Events detection, segmentation, and classification for audio streams.

Joint research work with (Vienna University of Technology, Dept. of Software Technology and Interactive Systems) was performed on musical instrument classification [2]. A class of algorithms for automatic classification of individual musical instrument sounds is presented. Two feature sets were employed, the first containing perceptual features and MPEG-7 descriptors and the second containing rhythm patterns developed for the SOMEJB project. The features were measured for 300 sound recordings consisting of 6 different musical instrument classes. Subsets of the feature set are selected using branch-and-bound search, obtaining the most suitable features for classification. A class of supervised classifiers is developed based on the non-negative matrix factorization (NMF). The standard NMF method is examined as well as its modifications: the local and the sparse NMF. The experiments compare the two feature sets alongside the various NMF algorithms. The results demonstrate an almost perfect classification for the first set using the standard NMF algorithm (classification error 1.0%), outperforming the state-of-the-art techniques tested for the aforementioned experiment.

Publications MP-codes: 284

5.12 Contribution by CWI

Researchers involved: Eric Pauwels, Elena Ranguelova, Mark Huiskes

Activities Detection and automatic categorization of pigmentation markings on whale flukes for photo-identification of individual animals. Subsequent matching is based on local configuration.

In addition, Sandor Fazekas (SZTAKI) has visited CWI from Apr 3 until Apr 13, 2006. In collaboration with Mark Huiskes (CWI) and Renaud Peteri (La Rochelle, also visiting CWI from Apr 3 to Apr 8), he has worked on extending the DynTex database of dynamic textures.

The DynTex database was initially set up by Peteri and Huiskes. It aims to provide the dynamic texture research community with a comprehensive set of sequences of dynamic textures for a variety of testing purposes. The DynTex team, of which Sandor Fazekas has become an active member, has used the collaboration period at CWI for the following activities:

- Writing a joint paper documenting the DynTex database in detail. The paper includes a discussion on the definition of dynamic textures, the various purposes of the database, a metadata annotation scheme (see below) and the collection protocol for the sequences.
- Setting up the DynTex annotation scheme: an XML-based metadata scheme facilitating, amongst others, the description of the perceptual qualities of dynamic textures. It also includes a classification hierarchy for dynamic textures based on their underlying physical processes.
- An extension of the DynTex database by collecting, processing and annotating a large set of new dynamic texture sequences.

Publications MP-codes: 260

5.13 Contribution by TCD

Researchers involved: Simon Wilson,

Activities Discussed with Julien Fauqueur (UCAM) and Nozha Boujemaa (INRIA) about a chapter on CBIR for the proposed book on Machine Learning for Multimedia Data, edited by Matthieu Cord. We have agreed a table of contents for the chapter and an action list.

Also we have discussed running experiments using my Bayesian CBIR system for mental image search that will take as a prior image from the Visual Thesaurus idea of Fauqueur and Boujemaa. UCAM will supply database features and queries, and TCD will implement on their CBIR system.

Publications MP-codes:

5.14 Contribution by TAU-visual

Researchers involved: Nahum Kiryati, Tomer Amiaz

Activities We develop new algorithms for dense optical flow computation. Dense optical flow schemes are challenged by the presence of motion discontinuities. In state of the art optical flow methods, over-smoothing of flow discontinuities accounts for most of the error. A breakthrough in the performance of optical flow computation has recently been achieved by Brox et al. Our algorithm embeds their functional within a contour-based segmentation framework. Piecewise-smooth flow fields are accommodated and flow boundaries are crisp. Experimental results show the superiority of our algorithm with respect to alternative techniques.

Publications MP-codes: 238

5.15 Contribution by TAU-visual

Researchers involved: Nahum Kiryati, Tammy Riklin-Raviv, Nir Sochen

Activities Object detection and segmentation can be facilitated by the availability of a reference object. However, accounting for possible transformations between the different object views, as part of the segmentation process, remains a challenge. Recent works address this problem by using comprehensive training data. Other approaches are applicable only to limited object classes or can only accommodate similarity transformations. We suggest a novel variational approach to prior-based segmentation, which accounts for planar projective transformation, using a single reference object. The prior shape is registered concurrently with the segmentation process, without point correspondence. The algorithm detects the object of interest and correctly extracts its boundaries. The homography between the two object views is accurately recovered as well. Extending the Chan-Vese level set framework, we propose

a region-based segmentation functional that includes explicit representation of the projective homography between the prior shape and the shape to segment. The formulation is derived from two-view geometry. Segmentation of a variety of objects is demonstrated and the recovered transformation is verified.

Publications MP-codes: 243

5.16 Contribution by TAU-visual

Researchers involved: Nahum Kiryati, Nir Sochen, Leah Bar

Activities Image restoration and segmentation are both classical problems, that are known to be difficult and have attracted major research efforts. This activity shows that the two problems are tightly coupled and can be successfully solved together. Mutual support of image restoration and segmentation processes within a joint variational framework has been theoretically motivated, and validated by successful experimental results. The proposed variational method integrates semi-blind image deconvolution (parametric blur-kernel), and Mumford-Shah segmentation. The functional is formulated using the Gamma-convergence approximation and is iteratively optimized via the alternate minimization method. While the major novelty of this work is in the unified treatment of the semi-blind restoration and segmentation problems, the important special case of known blur is also considered and promising results are obtained.

Publications MP-codes: 240

5.17 Contribution by TAU-visual

Researchers involved: Nahum Kiryati, Artemy Baxanski

Activities The volume, location of the centroid, and second order moments of a three-dimensional star-shaped object have been determined in terms of the spherical harmonic coefficients of its boundary function. Bounds on the surface area of the object are derived in terms of the spherical harmonic coefficients as well. Sufficient conditions under which the moments and area computed from the truncated spherical harmonic series converge to the actual moments and area are established. The proposed method is verified using a scanned head model and by recent measurements of the 433 Eros asteroid. An extension to non-star-shaped objects of genus 0 is provided. The computational complexity of our method is shown to be equal to that of the discrete spherical harmonic transform, which is $O(N^2 \log^2 N)$, where N is the maximum order of coefficients retained in the expansion.

Publications MP-codes: 241

5.18 Contribution by TAU-visual

Researchers involved: Nahum Kiryati, Tammy Riklin-Raviv, Nir Sochen

Activities Visit of Tammy Riklin-Raviv, a TAU-VISUAL PhD student, to INRIA-ARIANA. This activity is related to the Shape-Modelling e-team.

Publications MP-codes:

5.19 Contribution by TAU-speech

Researchers involved: Nahum Kiryati, Tal Darom

Activities A method for representation and lossy compression of textured surfaces was developed. The input surfaces are represented by surfels (surface elements), i.e., by a set of colored, oriented, and sized disks. The position and texture of each surfel are mapped onto a sphere. The mapping is optimized for preservation of geodesic distances. The components of the resulting spherical vector-valued function are decorrelated by the Karhunen-Loeve transform and represented by spherical wavelets. Successful representation and reconstruction has been demonstrated. Methods for geodesic distance computation on surfaces represented by surfels have also been presented.

Publications MP-codes: 239

5.20 Contribution by UPC

Researchers involved: Montse Pardas, Cristian Canton, Xavier Giro, Veronica Vilaplana, Miriam Leon, Jose Luis Landabaso, Camilo Dorea

Activities

- Coordination of E-team session on Person detection, recognition, tracking and analysis for the Plenary meeting at Istanbul.
- Attendance and presentation of the work at the Plenary meeting at Istanbul.
- **Image sequence description using long-term motion features** In this work we proposed a technique for generating partition sequences of regions presenting long-term homogeneity in color and motion coherency in terms of affine models. The technique is based on region merging schemes compatible with hierarchical representation frameworks and can be divided into two stages: Partition Tracking and Partition Sequence Analysis. Partition Tracking is a recursive algorithm whereby regions are constructed according to short-term spatio-temporal features, namely color and motion. Partition Sequence Analysis proposes the Trajectory Adjacency Graph (TAG) to exploit the long-term connectivity relations of tracked regions. A novel trajectory merging strategy using color homogeneity criteria over multiple frames is introduced. Algorithm performance is assessed and comparisons to other proposals, are drawn by means of established evaluation metrics . This work has been done in cooperation with ARMINES-CMM.
- **Tree Structured-based Caption Text Detection Approach** Nowadays superimposed text in both images and video sequences provides useful information about their contents. The aim of this paper is to introduce a method, which allows us to extract this kind of information, focused on working as independently as possible from the content, quality or font. Some pre-processing tools can be applied in order to reduce the number of false positives as well as the computational cost. The input image is represented by means of a Max-tree. This structure allows us to perform text localisation as a tree pruning. The pruning is performed applying connected operators based on geometric features of the letters. As a result, a set of potential text regions are obtained. The output of this first stage shows promising results. A second stage will be necessary to extract text as a whole, a set of unconnected regions with a unique meaning, allowing us to discard those regions not accomplishing text features.
- **Face recognition** We have developed a technique for face recognition in smart environments. The technique takes advantage of the continuous monitoring of the scenario and combines the information of several images to perform the recognition. Appearance based face recognition techniques are used given that unobtrusive systems are required in this type of applications and that the scenario does not ensure high quality images. Models for the users are created on-the-fly and subsequent face images of the same individual are gathered into groups. Images within a group are jointly compared to the models for identification and verification purposes. Reliable face images are used to update the users models. The proposed technique is assessed with the BANCA and CHIL databases.

- **From Partition Trees to Semantic Trees** This work proposes a solution to bridge the gap between semantic and visual information formulated as a structural pattern recognition problem. Instances of semantic classes expressed by Description Graphs are detected on a region-based representation of visual data expressed with a Binary Partition Tree. The detection process builds instances of Semantic Trees on the top of the Binary Partition Tree using an encyclopedia of models organised as a hierarchy. At the leaves of the Semantic Tree, classes are defined by perceptual models containing a list of low-level descriptors. The proposed solution is assessed in two very different environments such as smart rooms and remote sensing.
- **Object representation using colour, shape and structure criteria in a Binary Partition Tree** Binary Partition Trees (BPTs) are a well known technique used for region-based image representation and analysis. BPTs are usually created as a result of a merging process based on homogeneity properties, such as colour or motion. In this work, we develop a BPT creation technique based on a general merging algorithm, where the homogeneity criteria are neither low-level (pixel oriented, intra-region), nor high-level features (object oriented, semantics), but a combination of several criteria including region-based and structural features such as shape and partial-inclusion. We are thus combining intra-region homogeneity (e.g colour-based) with inter-region homogeneity (structural), with the long term aim of bridging the gap in region-based image analysis from low-level features to a higher level interpretation of the image by the intermediate description of the image structure (which we call "syntactic visual features"). Syntactic visual features are geometric relationships among regions based on shapes and the spatial configuration of neighboring regions in the image and can be found by structure analysis (or syntactic analysis). The aim of this work is to present how the addition of combined pixel-region and structural features leads to better object BPT representation.

Publications MP-codes: 210, 211, 214

5.21 Contribution by Bilkent University

Researchers involved: A. Enis Cetin, Mehmet Turkan, Ibrahim Onaran, Berkan Dulek

Activities We continue to develop a human face detection method in images and video. After determining possible face candidate regions using color information, each region is filtered by a high-pass filter of a wavelet transform. In this way, edges of the region are highlighted, and a caricature-like representation of candidate regions is obtained. Horizontal, vertical and filter-like projections of the region are used as feature signals in dynamic programming (DP) and support vector machine (SVM) based classifiers. It turns out that the support vector machine based classifier provides better detection rates compared to dynamic programming in our simulation studies.

Publications MP-codes:

5.22 Contribution by UCL

Researchers involved: Fred Stentiford, Li Chen, Adetokunbo Bamidele, Shijie Zhang

Activities Work at UCL has concentrated on detecting new forms of saliency that are relevant to the understanding of visual content (Task 2: Image and Video Processing; Sub-task 4: Saliency detection and visual features configuration).

A framework for attention mechanisms has been developed that enables different structures to be identified in images by incorporating transforms that correspond to the features being sought. Reflective symmetries are detected using an attention mechanism in which pixel configurations are transformed through reflections before translation and checking for a match. Peaks in the distributions of reflection axis angles at which matches are found indicate the locations and strengths of the symmetries present

in the image. The mid points of lines joining corresponding fork pixels lie along the axis of symmetry of the shape. Forks include some pixels that mismatch each other; this ensures that emphasis is given to image regions that contain attentive material. Papers were presented at ICAPR 2005 and MMSP 2005.

In a similar fashion measures of perspective are computed using a scaling transform before testing for a match. Peaks in the distributions of matches across the image indicate the locations of vanishing points in the image. A paper has been submitted to ICIP 2006.

Earlier work on attention-base similarity has led to the development of a new algorithm that extracts colour correction parameters from pairs of images and enables the perceived illumination of one image to be imposed on the other. This time colour transforms are used as a means of obtaining the relative illumination between two images. The colour shift that stimulates the highest frequency of matches represents a measure of the relative illumination of the two images. We apply the reverse colour shift to pixels in the second image to obtain a transformation that approximates the illumination present in the first image. The work was presented at SPIE in January 2006 and an internet service has been established to evaluate usability and acceptability (<http://colourcorrection.bat.bt.co.uk/ColourCorrection/>). Attention based methods for motion estimation are being investigated that do not depend upon specific features which are thought to characterise foreground objects or background regions. The technique maps attention mechanisms in time and generates motion vectors for each frame in a moving video sequence. Results have been compared with motion vectors derived from MPEG video encoders. A paper has been submitted to ICPR 2006.

Whilst experimenting with real time DSP implementations of visual attention algorithms applied to video camera outputs, it has been observed that the highest measures of saliency in images are obtained at the point of best focus. This work promises to offer new techniques for optimising the informativeness of images during focusing operations or through the variation of other parameters e.g. spectrum of illumination, brightness, etc. A paper has been submitted to ICIP 2006.

Visual attention algorithms have also been used to guide the selection of seed points for region growing in image segmentation. Points of low attention are normally present in large areas of self-similar background regions that are easy to segment, whereas high attention regions are often in the vicinity of rapidly changing features where segmentation needs to be constrained. A paper has been submitted to ICPR 2006.

Several findings from research on Content Based Image Retrieval have been published that have used attention based similarity measures. Most significantly attributes derived from a combination of contextual metadata and image similarity have been used to indicate the location at which camera phone images were taken. In this work the problem of image classification by grouping images into visual clusters and combining this with contextual metadata was undertaken with some success using a very diverse set of images. This work was presented at SPIE in January 2006.

Personnel exchanges with INRIA are planned for the end of April during which ideas taken from saliency research will be applied to problems in copy detection.

Publications MP-codes: 206, 207, 220

5.23 Contribution by UCL

Researchers involved: Fred Stentiford, Li Chen, Adetokunbo Bamidele, Shijie Zhang

Activities Work has continued to investigate eye gaze behaviour during image search and retrieval. (Task 2: Image and Video Processing; Sub-task 4: Saliency detection and visual features configuration). Models of visual attention have been extended to encompass specific aspects of visual saliency such as reflective symmetry, perspective and colour constancy. New methods for determining image similarity have been combined with analysis of automatically acquired contextual metadata associated with images taken by cameraphones to substantially improve the performance of face and place recognition algorithms. Two fellowship proposals have been submitted in collaboration with CWI and INRIA. In

addition an exchange of personnel is planned with INRIA to explore the application of ideas arising from saliency research to copy detection.

Publications MP-codes: 190, 195-9, 201

5.24 Contribution by MTA-SZTAKI

Researchers involved: Tamas Sziranyi, Lszl Havasi

Activities Knowledge of the vanishing-point position is the key for the geometrical modeling of reflective surfaces or cast shadows. An automatic method is presented using motion statistics to determine correspondences, and an improved fitting function for final parameter estimation which takes into account the statistical properties of image-points. The experiments show that our approach gives robust results in the context of widely different environments especially in cases where the correspondences are corrupted with considerable amounts of noise.

Publications MP-codes: 187

5.25 Contribution by MTA-SZTAKI

Researchers involved: Dmitry Chetverikov, Sandor Fazekas

Activities Developed, implemented and tested new method for temporal periodicity estimation of dynamic textures. Paper submitted to BMVC 2006.

Publications MP-codes:

5.26 Contribution by ACV

Researchers involved: Herbert Ramoser, Csaba Beleznai, Julia Puckmayr, Markus Clabian

Activities

- All papers submitted for the 2nd Workshop on Applications of Computer Vision (held in conjunction with ECCV 2006) have been review. 17 out of 20 submitted papers were accepted for presentation.
- A software tool (Matlab) for annotation of video sequences has been developed. The tool allows input of object bounding boxes and stores object type (e.g. human, vehicle), object position and trajectories for multiple frames.
- Two video sequences have been annotated using above tool. The sequences contain moving persons and vehicles in a surveillance scenario and can be used to assess the performance of algorithms.

Publications MP-codes:

5.27 Contribution by INRIA-Parole

Researchers involved: Khalid Daoudi, Christophe Cerisara, Yves Laprie

Activities Task3: Speech and audio processing.

Sub-task: Robustness of speech recognition systems: We proposed a new strategy of noise compensation for robust speech recognition. This strategy is an hybrid approach between the well-established denoising and adaptation techniques. The basic principle is to use stereo training data to build a speech denoiser which parameters can be easily and efficiently adapted to new environments using test data. The preliminary experiments we carried out show that this strategy is very promising and leads to better performances than classical methods.

Sub-task: Speech Analysis: We developed the first version of the audiovisual-to-articulatory inversion. The main idea is to add constraints provided by the view of the visible articulators (lips and lower jaw). These constraints are used to select articulatory regions. Then formant frequencies are used to recover the other articulatory parameters. Preliminary tests were conducted and showed that the expected acoustic precision should be decreased in order not to discard correct articulatory trajectories. This is probably due to the model discrepancy between the analyzing articulatory model and the speakers vocal tract.

Publications MP-codes: 287,288

5.28 Contribution by INRIA-Vista

Researchers involved: Ivan Laptev, Frdric Cao, Thomas Veit, Patrick Bouthemy

Activities Coherent motion detection with an a contrario approach

We investigated the following question: given a region of an image in a sequence, should it be considered as moving or not? The answer was given by the Helmholtz principle in [Veit et.al.2006]: if a certain well-defined motion quantity were randomly and uniformly spread over the data, then it would be too improbable to observe high concentration of motion as we actually do. This is measured by the Number of False Alarms of a region (NFA) which is related to the expected number of regions with the same concentration of motion quantity in a noisy situation. For a region R , the lower $NFA(R)$, the more meaningful the region as a moving region. Now, a question arises when regions are nested: a large region may be detected because it contains one or several moving regions. In this case, what is the best representation of the data in terms of motion? The answer we propose is that the best solution is exactly the one with the smallest NFA.

The second part of the work is concerned with the detection of small pieces of trajectories. Given, at each instant, the set of regions detected as moving, how to detect the spatio-temporal coherence between the instantaneous detections? It would allow us to i) eliminate detections with no coherence, ii) reinforce the detectability of moving regions by integration information through time, iii) initialize trajectories. Again, decision is made by observing how far the data is from randomness. Assume that local features are computed in the vicinity of moving regions. This locality is necessary in order to be robust to local perturbations of the images (shadows, occlusion, etc...). Pairs of features from consecutive images are considered. In a short time interval, the trajectory of a moving object is assumed to be close to rectilinear with a constant velocity. Hence, from a pair of features at times $(t, t + 1)$, one can compute a candidate velocity and a candidate initial position at time 0. If we now assume that these observations are independent for all instants and any position, we can compute how unlikely it is to observe a concentration of points around a given initial position and velocity. Hence, the problem of finding spatio-temporal coherence, can be turned into a clustering problem where the space of parameters is the space of initial position and velocity. The methodology is inspired by our works on grouping for shape analysis and recognition as well as by our work on motion detection [Veit et.al.2005, Cao et.al.2005].

Publications MP-codes: 293, 294, 295

5.29 Contribution by INRIA-Vista

Researchers involved: Ivan Laptev, Patrick Pérez

Activities Periodic motion detection and video alignment

Independent motion is a strong cue for detecting, segmenting and recognizing objects and activities in image sequences. Motion-based segmentation, however, is known to be a hard problem in scenes with motion parallax and scenes with multiple moving objects. In our work [MP-256] we have addressed this problem by exploiting not only the presence but also the type of motion as an informative cue, and thus, by addressing motion detection and recognition jointly (the two problems being usually treated separately). As one specific type of motions we considered periodic ones, noting their detection can be roughly seen as sequence alignment (a sequence is matched to itself over one or more periods). Hence, we first considered alignment of two video sequences obtained by independently moving cameras. Under assumption of constant translation, fundamental matrices and the homographies were shown to be time-linear matrix functions. These dynamic quantities can be estimated by matching corresponding space-time points with similar local motion and shape. For periodic motion, we matched corresponding points across periods and developed a ransac procedure to simultaneously estimate the period and the dynamic geometric transformations between periodic views. Using this method, we demonstrated detection and segmentation of human periodic motion in complex scenes with non-rigid backgrounds, moving camera and motion parallax. In the on-going work we extend this framework to align video sequences with similar type of activities, using appropriate local motion descriptors and global geometric constraints.

Publications MP-codes: 256

5.30 Contribution by INRIA-Vista

Researchers involved: Ivan Laptev, Patrick Pérez

Activities Multiple-object tracking and Bayesian tracking with auxiliary variables

Handling in a principled way a varying (and unknown) number of entities to be tracked is a difficult and important problem, which keeps receiving a great deal of attention from both trajectography and computer vision communities. Our recent contributions to these efforts included various sequential Monte Carlo techniques, focusing on the data association problem in case the number of objects is fixed. We addressed the problem of starting and terminating automatically individual tracks as objects enter and exit the volume of observation. To this end we have introduced a binary “existence” variable within a generic multi-object “detect-before-track” state model. We have proposed a general treatment of this augmented model that impacts as little as possible on existing tracking algorithms, so that software can be reused, and that allows implementation with Kalman filters, extended Kalman filters, particle filters, etc. We have applied the proposed framework to color-based tracking of multiple objects, where we adopted a mixture Kalman filter implementation.

Based on above-mentioned work on tracking with existence [MP-298], we considered the more general context of sequential estimation problems with an auxiliary discrete process as part of the state space. These discrete variables usually follow a Markovian process and interact with the hidden state either via its evolution model or via the observation process, or both. Examples of such auxiliary variables include depth ordering for occlusion handling, switches between different state dynamics, exemplar indices, etc. We have considered a general model that encompasses all these situations, and show how Bayesian filtering can be rigorously conducted in this general setup. The resulting approach facilitates easy re-use of existing tracking algorithms designed in the absence of the auxiliary process. In particular, we have shown how particle filters can be obtained based on sampling only in the original state space instead of sampling in the augmented space, as it is usually done. In [MP-299] we have demonstrated how this framework facilitates solutions to the critical problem of appearance and disappearance of targets, either upon scene entering and exiting, or due to temporary occlusions. We have illustrated this in the context of “track-before-detect” color-based tracking with particle filters.

Publications MP-codes: MP-298, MP-299

5.31 Contribution by INRIA-Vista

Researchers involved: Ivan Laptev, Patrick Pérez, Nicolas Gengembre, Venkatesh Babu Radhakrishnan, Patrick Bouthemy

Activities Robust visual tracking without prior

In this research we are interested in the problem of tracking arbitrary entities along videos of arbitrary type and quality. Such a tracking cannot rely, as classically done, on a priori information regarding both the appearance of the entities of interest (shape, texture, key views, etc.) and their visual motion (kinematical constraints, expected dynamics relative to the camera, etc.).

The first crucial step is then the definition and the estimation of the reference appearance model on which the tracking, no matter its precise form, will rely on. Roughly, two extreme types of representations are routinely used in the literature: detailed pixel-wise appearance models subject to rapid fluctuations (e.g. intensity template instantaneously refreshed) and rough color models very persistent over time (e.g., color histogram instantiated at initialization time and kept unchanged). They are both interesting and complementary. For these reasons, we have combined them in a simple probabilistic tracker, where intensity template matching, refined by gradient ascent on color similarity to a fixed global model, feeds a Kalman filter [MP-296, MP-297]. Such a simple combination has already permitted to improve robustness compared to trackers based on each of the standard representations alone. We have also shown that the discriminative power of the color representation could be noticeably improved by selecting carefully the selected color space and by performing background color subtraction (which we have done in various ways). To handle the problem of partial and global occlusions we have further introduced a random partitioning of the target object into a constellation of parts. These parts are individually tracked and then merged after occluded parts are detected and ignored. Complete occlusions, however, remain a critical problem which must be addressed. Although probabilistic generic trackers have proved better at handling them, as compared to their deterministic counterparts, we are still lacking ways to fully treat the problem. One interesting output of our methodological work on Bayesian tracking with discrete auxiliary variables is that it can be used to seamlessly incorporate a binary “visibility” variable in existing state space models for tracking. Not only this improves slightly the robustness of tracking under severe occlusions, but it also allows sequential evaluation of the posterior visibility probability within the tracking procedure.

Publications MP-codes: MP-296, MP-297

5.32 Contribution by INRIA-Imedia

Researchers involved: Valérie Gouet-Brunet, Nozha Boujemaa, Julien Law-To, Olivier Buisson, Alexis Joly

Activities Local descriptors have been proved to be very useful for image indexing with application to object or sub-image retrieval. A number of recent techniques have been proposed to identify points of interest or regions of interest in images. If directly applied to image sequences, one of the drawbacks of such descriptors is its spatio-temporal redundancy. When considering applications like real-time monitoring, it is necessary to build a very compact description. In this context, we have studied the behaviour of local descriptors. Studying the trajectories of local descriptors is an efficient solution to avoid temporal redundancy and, moreover an interesting way to strongly fingerprint the video sequence. We have proposed to modelize and label the different kinds of behaviours of such local descriptors according to their trajectories. Such a content-based video description is richer and more compact than the usual ones (see ref. MP-302). This work is done in collaboration with the INA (French National Institute of Audiovisual), within the scope of a CIFRE thesis. The main application

considered is real-time monitoring of huge databases of videos (about 100000 hours). At present, we focus more particularly in video content-based copy detection (CBCD).

In parallel, we have proposed a density-based method to select discriminant local features in images or videos, by introducing a new fast density estimation technique using a simple grid index structure and specific queries based on the energy of the gaussian function (see ref. MP-303).

We are also working on the enhancement of scenarios involving local descriptors, by exploiting the geometric distribution or the relative positions of the local descriptors. During this year, we have shown that CBCD scenarios, the robust estimation of a global geometric transformation model after the search is widely profitable to improve the discrimination of the detection. However, for other scenarios, using the geometry remains a challenging task. For a few months, we are investigating the use of non parametric geometric consistency measurements such as mutual information and robust correlation ratio and we plan to combine them with some robust local geometric properties that could be included in the descriptor itself in order to limit the number of matches during the second step.

Within the WP5, we have proposed the Visual Saliency e-team (<http://www-rocq.inria.fr/imedia/Muscle/WP5/eteams.html>). This e-team will investigate and develop saliency operators for images and video sequences, e.g. based on local descriptors with point-of-interest detection and description. Once salient regions have been detected they have to be represented by visual features and their organization (relations) has to be described. We will work on perceptually motivated high-level features in computer vision such as symmetry, anisotropy, regularity, complexity, symmetry of textures and shapes, on multi-resolution behavior and also spatio-temporal based saliency features. We will also focus on combined representation of salient image information through region and point of interest. This e-team is coordinated by Imedia and involves 13 Muscle partners.

Publications MP-codes: 302, 303

5.33 Contribution by INRIA-Ariana

Researchers involved: Ian Jermyn, Josiane Zerubia, Johan Aubray, Peter Horvath, Ting Peng, Giuseppe Scarpa, Dan Yu

Activities Ariana's work in WP5 involves two areas: modelling images and modelling regions in the image domain. Together, as likelihood and prior, they enable the extraction of the regions in images corresponding to particular entities.

Image modelling: the work on adaptive wavelet packet texture models has continued. The quartic models of the bimodal statistics of adaptive wavelet packet coefficients developed by Johan Aubray, stagiaire in INRIA-Ariana last year (a paper has been accepted to EUSIPCO 2006 on this topic), are now being extended, by Dan Yu, postdoc in INRIA-Ariana this year, to include inter-subband correlations. The result will be texture models with non-trivial rotation invariance: that is, typical texture images under the probability distribution can be oriented, unlike rotation invariant Gaussian models, but no particular orientation will be preferred. This will obviate the need for pose estimation in the application of these models to images. At the same time, the models will retain the ability to represent the bimodal statistics so critical to the description of texture structure.

Image modelling: work has continued on the construction of image models for the extraction of road networks from very high resolution (0.5m) satellite images. The existence of phenomena at multiple different scales in very high resolution images suggests the use of multiscale models. Ting Peng, joint PhD student of INRIA-Ariana and the LIAMA Institute in Beijing, studied the one- and two-point statistics of the scaling coefficients, and the one-point statistics of the wavelet coefficients, at different scales, for both roads and the background. These statistics are modelled well by one or at most two component Gaussian mixtures, the mixture components having clear physical interpretations (for example, for the background, one component corresponds to shadows and bodies of water). The resulting likelihoods will now be used with a phase field higher-order active contour prior to extract road networks.

Region modelling: the higher-order active contour model for a 'gas of circles', which describes regions composed of a number of circles all of approximately the same radius and with short-range mutual repulsion, has been further advanced by Peter Horvath, joint PhD student of INRIA-Ariana and the University of Szeged. The previous model made the circles involved into energy minima, which creates problems for the gradient descent algorithm used to minimize the energy: it can become stuck in these local minima, thus producing phantom circles even when there is no supporting data. An analysis of the energy enabled the calculation of parameters making the circles energy inflection points, there by ensuring that they can be made to disappear in the absence of supporting data. As a side-effect, this constraint halves the number of parameters in the energy, and severely restricts one of the two that remain. This is a major advantage in an energy-based approach where parameter setting is problematic. The model has been applied to tree crown extraction, and despite the limited parametric freedom, produces results as good as or better than the previous model. A paper will be submitted to BMVC 2006 on this work.

Region modelling: Tammy Riklin-Raviv, PhD student with TAU-Visual, visited INRIA-Ariana for several days to begin work on the 'Shape modelling' e-team. During the visit, energies incorporating both phase field higher-order active contour models and the reference region priors developed in TAU-Visual were constructed, and experiments were conducted to investigate their behaviour, with promising results. The next steps in the collaboration will be to study the robustness and feasibility of the combined model, and to incorporate projective distortions of the object sought la Riklin-Raviv *et al* . (MP-136). Giuseppe Scarpa, joint ERCIM postdoc of INRIA-Ariana and UTIA, has continued his work on the development and evaluation of a novel segmentation algorithm for textured images. In particular, his research activity focused on a refinement of the finite-states model-based algorithm. The reliability of the technique was analysed, and as a result the model was modified. Further modifications are in progress, after which research will switch to the testing of the modifications.

Publications MP-codes: 314

5.34 Contribution by UTIA

Researchers involved: Michal Haindl, Stanislav Mikes, Jiri Grim, Jiri Filip, Petr Somol

Activities UTIA work in WP5 has focused on developing new probabilistic multidimensional image models mainly for Bidirectional Texture Function (BTF) compression and synthesis, multichannel image restoration, image and range map segmentation and benchmarking. These results were published in more than 25 publications in journal papers, conference proceedings, research reports and one PhD thesis.

Bidirectional Texture Function Compression and Modelling Methods

The BTF textures represent majority of natural textures for which the Lambert law does not hold and their appearance changes due to varying illumination and viewing conditions. The BTF function is acquired by large number of measurements for all possible combinations of illumination and viewing positions and its representation typically requires to store tera bytes of texture data which is far out of limits for any current real time hardware. Several novel methods allowing to compress BTF measurements were published (MP-120,126,318,319,320). These methods are either based on the set of Markov random field models or on the non-linear reflectance models. The proposed BTF Markovian representation enables to reach the several order improvement of the compression ratio (1:million) above any other known compression method and it is the best currently available method in this area. These methods simultaneously allow efficient BTF texture modelling directly in the graphical processing unit. An alternative method of BTF texture modelling based on Gaussian distribution mixtures have been proposed MP-321. The synthesized texture is obtained by means of a step-wise prediction of the texture image and the method can be viewed as a statistically controlled sampling.

Unsupervised Multispectral Texture Segmentation

An unsupervised multispectral texture segmentation algorithm (MP-317) based on a causal simultaneous autoregressive random field representation and the Gaussian mixture model was developed.

Text Document Classification

Multinomial mixture models were applied as the models for class-conditional distribution for multiclass text document classification task. Two new feature selection criteria, modifications of the mutual information criterion for feature/word selection have been proposed for the purpose of text classification.

Multichannel Image Restoration

A novel multiimage / multichannel restoration method based on the underlying spatial probabilistic image model for astronomical image restoration if degradation obeys a linear degradation model with the unknown possibly non-homogeneous point-spread function was published in MP-316.

Publications MP-codes: 126,315,316,317,318,319,320,321

6 Overview activities in WP 6

6.1 Contribution by AUTH

Researchers involved: Constantine Kotropoulos, Margarita Kotti, Emmanouil Benetos, Ignacio Martinez de Lizarrondo, Gustavo Martins, Jaime Cardoso

Activities Speaker turn detection

Task 3: Cross-modal Integration for Multimedia Analysis and Recognition

Subtask 3.1: Video Analysis and Integration of Asynchronous Time-evolving Modalities

The activity is related to E-team 2 on Audio-Visual Understanding and in particular with the application area entitled "Dialogue detection in movies".

Unsupervised speaker change detection is a necessary step for several indexing tasks. We assume that there is no prior knowledge on the number of speakers. New features, included in the MPEG-7 Audio Prototype, are investigated such as the AudioWaveformEnvelope and the AudioSpectrumCentroid. The model selection criterion is the Bayesian Information Criterion (BIC), which can achieve a reliable segmentation performance. A multiple pass algorithm has been developed that uses a novel dynamic thresholding and a fusion scheme so as to refine the segmentation results. The experimental results on recordings extracted from the TIMIT database demonstrate that the performance of the proposed multiple pass algorithm is better than that of the existing approaches.

Further developments aim at speeding up the BIC-based speaker change detection by employing subset feature selection for dimensionality reduction, second-order statistical measures (such as the sphericity measures applied to the covariance matrices of the reduced features vectors, the Euclidean distance between MFCCs and the T2 Hotelling measure applied to MFCCs) before resorting to BIC.

Publications

Publications MP-codes: 235,283

6.2 Contribution by AUTH

Researchers involved: Constantine Kotropoulos, Ioannis Pitas, Margarita Kotti, Vassiliki Moschou, Panagiotis Antonopoulos, Spyridon Siatras

Activities Multimodal dialogue dataset

Task 3: Cross-modal Integration for Multimedia Analysis and Recognition

Subtask 3.1: Video Analysis and Integration of Asynchronous Time-evolving Modalities

The activity is related to E-team 2 on Audio-Visual Understanding and in particular with the application area entitled dialogue detection in movies.

A protocol to collect multimodal dataset capturing dialogues in movies is proposed. A sample multimodal dataset of 33 movie scenes extracted from 6 movies has been collected. The dataset is comprised

from the audio tracks, the visual channel, and text transcriptions. Proper annotation rules employed to describe the ground truth are investigated.

Publications MP-codes:

6.3 Contribution by AUTH

Researchers involved: Constantine Kotropoulos, Ioannis Pitas, Margarita Kotti, Bartosz Ziolkó, Vassiliki Moschou

Activities Dialogue detection

Task 3: Cross-modal Integration for Multimedia Analysis and Recognition

Subtask 3.1: Video Analysis and Integration of Asynchronous Time-evolving Modalities

The activity is related to E-team 2 on Audio-Visual Understanding and in particular with the application area entitled dialogue detection in movies.

We investigate two dialogue detection rules that are based on indicator functions. The first rule relies on the value of cross-correlation function at zero time lag that is compared to a threshold. The second rule is based on the cross-power in a particular frequency band that is also compared to a threshold. Experiments are carried out in order to validate the feasibility of the aforementioned dialogue detection rules by using ground-truth indicator functions determined by human observers from six different movies. Almost perfect dialogue detection is reported for every distinct threshold.

Publications MP-codes:

6.4 Contribution by AUTH

Researchers involved: Ioannis Pitas, Panagiotis Antonopoulos, Spyridon Siatras

Activities Face clustering and mouth activity detection for dialogue detection

Task 3: Cross-modal Integration for Multimedia Analysis and Recognition

Subtask 3.1: Video Analysis and Integration of Asynchronous Time-evolving Modalities

The activity is related to E-team 2 on Audio-Visual Understanding and in particular with the application area entitled dialogue detection in movies.

Scale invariant feature transform (SIFT) features have been used to perform face clustering on video scenes. Dialogue detection based on the successive occurrence of faces that belong in the produced clusters are investigated.

A method used to visually determine whether a person displayed in a video document is speaking using mouth intensity and edge detection information is studied.

Publications MP-codes:

6.5 Contribution by AUTH

Researchers involved: Constantine Kotropoulos, Marios Kyperountas, Ioannis Pitas

Activities Audiovisual scene change detection

Task 3: Cross-modal Integration for Multimedia Analysis and Recognition

Subtask 3.1: Video Analysis and Integration of Asynchronous Time-evolving Modalities

The activity is related to E-team 2 on Audio-Visual Understanding.

A novel audiovisual scene change detection algorithm has been developed and evaluated experimentally. An enhanced set of eigen-audioframes is extracted to create an audio signal subspace that aims at discovering the audio background changes. Visual information is used to align audio scene change indications with neighboring video shot changes and, accordingly, to reduce the false alarm rate. Moreover, video fade effects are identified and used independently in order to track scene changes. The detection methodology that processes the audio and video signals in complementary manner was tested on newscast videos provided by the TRECVID2003 video test set yielding a recall and precision rates of 80.6% and 82.1%, respectively.

Publications MP-codes:

6.6 Contribution by UVA

Researchers involved: Thang Pham, Arnold Smeulders, Michiel van Liempt, Theo Gevers, Nicu Sebe,

Activities We participated in discussion for potential collaboration within WP3. In particular, Dr. Nicu Sebe has received a mobility grant to go to Vienna for the period of 6-14 May 2006 to setup a joint research proposal.

We participated in the scientific meeting of WP5. We participate in the e-team "body detection and tracking" within the workpackage.

We contributed to WP6 two software modules for object tracking in video. These software modules are implementation of state-of-the-art tracking algorithms for video analysis.

We participated in the Dublin meeting of WP8. The aim is to form an e-team on semi-supervised learning. The discussion is on-going. We also participate in the scientific workshop organized by the workpackage.

Publications MP-codes: 280

6.7 Contribution by Bilkent University

Researchers involved: Yigithan Dedeoglu, B. Ugur Toreyin, Yigithan Dedeoglu, A. Enis Cetin

Activities We continued our work on multi-modal falling person detection. Human motion in video is modeled using Hidden Markov Models (HMM) in this study. In addition, the audio track of the video is also used to distinguish a person simply sitting on a floor from a person stumbling and falling. Audio channel data based decision is also reached using HMMs and fused with results of HMMs modeling the video data to reach a final decision. We worked on extending this method to recognize different human actions using the same multi-modal approach.

Publications MP-codes:

6.8 Contribution by INRIA-Textmex

Researchers involved: Patrick Gros, Manolis Delakis

Activities In the last two months, we investigated the use of information carried out by the player positions for tennis video structure analysis with Hidden Markov Models and Segment Models. We want to distinguish between the court views of missed serves and that of exchanges, where the actual game takes place. To this end, we used so far audio features regarding the ball hits and applause sound classes. We experienced a certain degree of difficulty during their automatic detection, especially for ball hits. Using information from the player positions at the beginning of the exchange can further support the audiovisual features. According to tennis rules, after a missed serve the player has to

repeat while both players being at the same location. On the contrary, after an exchange, both players have to change position. In order to detect changed positions, we firstly filter out non court views shots. We then examine by pairs successive court views shots using their first frames. Instead of trying to extract error-prone high-level features and attempt court reconstruction, we applied some simple, fast and yet effective basic image operations on the two frames. The player blobs are detected using color filtering, then image dilation is performed on the binarized frames. Finally we examine if the difference of the two processed frames gives a symmetrical image or not. In the first case, the players have changed position, while in the second one, they have not. While this technique is quite reliable, wrong estimations are unfortunately introduced by idle court views (no game action at all) or missed ones (false court view detection). Nevertheless, using these estimations as a binary extra feature in our models, we achieve a visible performance improvement.

Publications MP-codes: 285 286

6.9 Contribution by INRIA-Imedia

Researchers involved: Valrie Gouet-Brunet, Marin Ferecatu, Nozha Boujemaa, Michel Crucianu

Activities We addressed the challenge of semantic gap reduction using a hybrid visual and conceptual representation of the content within an active relevance feedback context. We introduced a new feature vector, based on the keyword annotations available for the images, which makes use of conceptual information extracted from an external lexical database, information represented by a set of "core concepts". Evaluation results show that the use of the proposed hybrid conceptual and visual feature vector dramatically improves the quality of the relevance feedback results (see ref. MP-304).

Publications MP-codes: 304

6.10 Contribution by ICCS

Researchers involved: George Papandreou, Athanassios Katsamanis, Vassilis Pitsikalis, Petros Maragos

Activities

Audio-Visual Speech Recognition

Research into this field aims at improving the performance of automatic speech recognition systems in noisy environments by exploiting speech-related information extracted from video depicting the face of the speaker. Audio-visual speech recognition, besides being an important research field in itself, serves as a major test-bed for methods and algorithms for cross-modal interaction potentially applicable to other multimedia integration scenarios. Part of our research on Audio-Visual speech recognition is done in collaboration with the TSI-TUC team.

We have been developing an integrated audio-visual speech recognition system. The visual front-end is based on statistical shape and appearance generative models, which track the speaker's shape and capture speech-related information into a compact set of visual speech features. As part of the visual analysis work in the above on-going research we have also been developing a statistically motivated scheme for enabling the synergy between object recognition and image segmentation with application to the problem of speakers face detection.

These visual features are combined with auditory features and enhance the performance of speech recognition systems; the improvement is most profound in low audio SNR environments. Training of the models and audiovisual ASR recognition experiments have been conducted on the CUAVE audio-visual speech database (obtained from Clemson University), while the Vid-Timit audio-visual database (obtained from the University of Adelaide) might be used in the future for additional experiments.

Our latest research in the field has focused on adaptive methods for fusing the audio and visual modalities. We show that if the speech degradation under noise is explicitly modelled and the uncertainty of the features is properly taken into account, fully adaptive weighting of the two modalities can be achieved and the performance of the system significantly improves. Our approach can be easily integrated with architectures such as Product Hidden Markov Models, which explicitly model the asynchrony between audio and video in speech.

Publications MP-codes: 313

6.11 Contribution by ICCS

Researchers involved: Kostas Rapantzikos, N. Tsapatsoulis, Y. Avrithis, S. Kollias

Activities

Spatio-Temporal Visual Attention and Ontologies

We study the interactions between low- and high-level information in multimedia ontologies. Progress has been done in refining a spatiotemporal model for visual attention with applications to video classification/retrieval and segmentation. Recently, we extended our research on multimedia ontologies by examining interactions between low- and high- level information. Much has also been done in refining a spatiotemporal model for visual attention with applications to video classification/retrieval and segmentation. We have also studied the use of the developed spatiotemporal visual attention scheme in achieving better video classification results. We tested the proposed approach in the sports domain and obtained promising results. Classification based on features extracted from the selected salient regions is more successful than the one based on feature extraction from the whole sequence. We have recently elaborated more on tuning the developed spatiotemporal visual attention scheme in order to obtain spatiotemporal regions that most probably correspond to meaningful areas (foreground/background, objects etc). For the time-being we tested the proposed approach in classifying sport video clips and obtained promising results. We soon expect to combine the proposed scheme with audio/speech detection/recognition towards robust audio-visual understanding.

Publications MP-codes:

7 Overview activities in WP 7

7.1 Contribution by INRIA-Ariana

Researchers involved: Ian Jermyn, Josiane Zerubia, Marie Rochery

Activities Higher-order active contours have recently been reformulated as non-local phase field models. This reformulation offers many advantages: at the algorithmic level via an immediate increase in computational efficiency, and by opening the way to further efficiency gains using multigrid, multiscale, and adaptive grid/wavelet algorithms, as well as the possibility of parallelization. A paper was published in ICCV 2005 on the phase field formulation, and Ariana hopes to have a PhD student working on this topic starting in 2006.

Publications MP-codes: 245,246,250

7.2 Contribution by INRIA-Ariana

Researchers involved: Ian Jermyn, Josiane Zerubia, Ting Peng

Activities The extraction of road networks from very high resolution remote sensing images, the thesis subject of Ting Peng, joint PhD student of INRIA-Ariana and the LIAMA Institute in Beijing (<http://liama.ia.ac.cn>) poses many problems, not the least of which is the large size of the images concerned and the consequent need for fast algorithms. Multiscale likelihood models for these images are already being developed, and several possibilities for the incorporation of these models into multiscale algorithms are being examined.

Publications MP-codes:

7.3 Contribution by TCD

Researchers involved: Simon Wilson,

Activities Simon Wilson visited ISTI-CNR in Pisa to discuss the use of Bayesian methods for source separation in an astronomy application - reconstruction of the Cosmic Microwave Background. A full Bayesian solution is highly computationally demanding. We will build on existing work in the area. Simon Wilson also presented his work on Bayesian CBIR to the group.

Publications MP-codes:

7.4 Contribution by TAU-visual

Researchers involved: Nahum Kiryati, Leah Bar, Nir Sochen

Activities We developed a variational approach for deblurring and impulsive noise removal in multi-channel images. A robust data fidelity measure and edge preserving regularization are employed. We considered several regularization approaches, such as Beltrami flow, Mumford-Shah and Total-Variation Mumford-Shah. The latter two methods are extended to multi-channel images and reformulated using the Gamma-convergence approximation. Our main contribution is in the unification of image deblurring and impulse noise removal in a multi-channel variational framework. Theoretical and experimental results show that the Mumford-Shah and Total Variation Mumford Shah regularization methods are superior to other color image restoration regularizers. In addition, these two methods yield a denoised edge map of the image.

Publications MP-codes: 242

7.5 Contribution by INRIA-Imedia

Researchers involved: Valrie Gouet-Brunet, Nozha Boujemaa, Nouha Bouteldja, Michel Scholl

Activities When considering objects or parts of image described with a set of local descriptors, searching in the feature space is usually done independently and sequentially for each local descriptor. We have studied multiple queries approaches, existing in the Database community, before applying and adapting them to the retrieval of groups of local descriptors. Two directions have been investigated:

- Reduction of the I/O costs, by studying a new approach for searching in the multidimensional structure ; the structure considered is the SR-Tree one, which achieves good performances for feature spaces based on local descriptors;
- Reduction of the CPU costs, by considering relations existing on distances computed between several query points and points of the feature space, for instance the triangular inequality. The structure considered is also the SR-tree one, but the proposed improvements could be applied to any tree structure. We have revisited two existing lemmas and proposed three novel ones.

The evaluation has been done over 15 datasets of 1 million of real local descriptors. The results show that the joint use of such lemmas allows to accelerate the search, whatever the dimensions (see ref. MP-305).

All these studies are done in collaboration with the CEDRIC/Vertigo research group.

Publications MP-codes: 305

7.6 Contribution by UTIA

Researchers involved: Michal Haindl, Jiri Filip, Giuseppe Scarpa

Activities UTIA work in WP7 has focused on texture segmentation and dynamic texture modelling. Fast Synthesis of Dynamic Colour Textures

We proposed a hybrid method for colour DTs modelling. The method is based on eigen-analysis of DT frames and subsequent modelling of temporal interpolation eigen-coefficients using a multidimensional probabilistic model. This method (submitted to ICPR 2006 conference) compresses significantly the original data and enables extremely fast synthesis of artificial sequence, which can be easily performed by the means of contemporary graphics hardware.

Unsupervised Texture Segmentation by Spectral-Spatial-Independent Clustering

A segmentation algorithm combining the color-based clustering and spatial-based clustering (SBC) was developed by a joint postdoc with INRIA Ariana, Giuseppe Scarpa. SBC, as well as the subsequent growing algorithm, make use of a characterization of the regions based on shape and context.

Publications MP-codes:

8 Overview activities in WP 8

8.1 Contribution by AUTH

Researchers involved: Constantine Kotropoulos, Vassiliki Moschou

Activities Clustering by using Self-Organizing Maps

Work on clustering N -dimensional patterns that are represented as points on the $(N - 1)$ -dimensional simplex has been performed. The elements of such patterns could be the posterior class probabilities for N classes, given a feature vector derived by the Bayes classifier for example. We are interested in reducing the number of clusters to $N - 1$, in order to redistribute the features classified into a particular class in the $N - 1$ simplex, according to the maximum a posteriori probability principle, over the remaining $N - 1$ classes in an optimal manner by using a self-organizing map. An application of the proposed solution to the re-assignment of emotional speech features classified as neutral into the emotional states of anger, happiness, surprise, and sadness on the Danish Emotional Speech database is presented.

The assessment of the clustering produced by two variants of the self-organizing map (SOM) that are based on order statistics, such as the marginal median SOM and the vector median SOM, has been performed. We have employed the well-known IRIS data set and we assess their performance with respect to the accuracy and the average over all neurons mean squared error between the patterns that were assigned to a neuron and the neuron's weight-vector. Both figures of merit favor the marginal median/vector median SOM against the standard SOM. Based on the aforementioned findings, the marginal median SOM and the vector median SOM have been used to re-distribute emotional speech patterns from the Danish Emotional Speech database that were originally classified as being neutral to four emotional states such as hot anger, happiness, sadness, and surprise.

Publications MP-codes: 236

8.2 Contribution by ENSEA

Researchers involved: Matthieu Cord, G. Chavez, Ph. Gosselin

Activities We work in the Machine learning framework for multimedia data processing. We have activities on image and video data:

- Image retrieval: we introduce active learning schemes based on new utility criteria dedicated for information retrieval different from the ones optimized for classification process. Some work about supervised learning to optimize Gram matrices for image retrieval has been presented to the ICIP 2005 conference.
- Video processing: we start to develop new techniques to segment video data.

Dissemination:

- Mini-course at UFGM, Brazil on machine learning for image retrieval,
- Muscle wp8 report about video analysis

Publications MP-codes: 312

8.3 Contribution by ENSEA

Researchers involved: Matthieu Cord, PH. Gosselin, G. Chavez

Activities We continue our activities on both video analysis and content-based data indexing and retrieval. As far as the video field is concerns, we focus on video cut detection, and experiment methods using machine learning techniques for doing the cuts (PhD of G. Chavez). On the indexing part, we start in the postdoctoral work of PH. Gosselin to explore kernel-based solutions for object class recognition. M. Cord gave a talk on the whole activities of the ENSEA in the CBIR field at the MUSCLE meeting in Istanbul.

Publications MP-codes:

8.4 Contribution by ENSEA

Researchers involved: Matthieu Cord, G. Cachez, Ph. Gosselin

Activities We develop some work on Content-based image retrieval and Video cut detection. We investigate Kernel methods for concept learning. The PhD defence of Ph. Gosselin was in december 05 on this subject. Some part of this work has been published in PR journal (MP-252). Experiments have been done on video data for sharp cut detection analysis.

Publications MP-codes: MP-252

8.5 Contribution by TSI-TUC

Researchers involved: Alex Potamianos, Ilias Iosif, Kelly Zervanou, Apostolis Pangos

Activities Unsupervised Semantic Class Induction:

Many applications dealing with textual information require classification of words into semantic classes including spoken dialogue systems, language modeling, speech understanding and machine translation applications. Manual constructing semantic classes is a time consuming task and often requires expert knowledge. Also, semantic features are sensitive to domain changes. Unsupervised induction of semantic classes is also the first step towards unsupervised learning of semantics from text, the holy grail of natural language processing.

The main idea underlying our approach is the principle that similarity of context implies similarity of meaning. Two semantic similarity metrics that are variations of the Vector Product distance are used in order to measure the semantic distance between words and to automatically generate semantic classes. The first metric computes wide-context similarity between words using a bag-of-words model, while the second metric computes narrow-context similarity using a bigram language model as discussed in Pargellis et al, 2004. A hybrid metric that is defined as the linear combination of the wide and narrow-context metrics is also proposed and evaluated. The semantic metrics are evaluated on two corpora: a semantically heterogeneous web news domain (HR-Net) and an application-specific travel reservation corpus (ATIS). For the hybrid metric, precision exceeds 85% for both corpora at 50% recall, a very significant improvement on the state-of-the-art.

Publications MP-codes: 274, 374

8.6 Contribution by TSI-TUC

Researchers involved: Alex Potamianos, Eduardo Sanchez-Soto

Activities Feature Fusion for Statistical Classifiers

A common practice for combining information sources in a statistical classification framework is the use of feature streams. A fundamental assumption behind streams is that the information sources/features are independent of each other and that the probability distribution functions (pdfs) of the two streams can be multiplied to obtain the global observation pdf. However, often this independence assumption does not hold or the reliability (estimation/modeling error) of each stream is different. In these cases, it has been empirically shown that stream weights (exponents weighting the contribution of each stream pdf) can reduce the total classification error.

In this work, we compute analytical results for the selection of stream weights as a function of single-stream estimation and misclassification errors. Optimality is investigated in terms of multi-stream classification error minimization for the two class problem. It is shown that under certain conditions, the optimal stream weights are inversely proportional to the single-stream classification error. We apply these results to the problem of audio-visual speech recognition and experimentally verify our claims. The applicability of the results to the problem of unsupervised stream weight estimation is also being investigated. The analytical results are verified for an AV-ASR multi-stream application. Using more complex statistical classifiers (e.g., Bayesian network) for feature fusion is also being investigated.

Publications MP-codes:

8.7 Contribution by TUVienna-PRIP

Researchers involved: Allan Hanbury, Branislav Micusik

Activities We have continued developing the segmentation of an image viewed as a one-class classification problem (described in Deliverable 8.1). At present, we are attempting to reduce the dependence of the algorithm on the choice of the gradient.

Publications MP-codes:

8.8 Contribution by UTIA

Researchers involved: Michal Haindl, Mikes, Scarpa, Novovicova, Grim

Activities UTIA work in WP8 focused on the unsupervised texture segmentation, feature selection and supervised learning.

Two new segmentation algorithms have been developed and their robustness against the additive contamination noise was tested. Two article were submitted to ICPR 2006 conference.

Feature Selection

Concurrent advance has been taking place regarding Branch & Bound in form of a) the prediction mechanisms that replace slow criterion evaluations by fast estimates, b) an idea to trade space for speed. Floating feature selection methods were investigated and several new improvements were tested. These results were accepted for S+SSPR 2006 conference publication. The chapter "Filter methods for feature selection" was submitted for WP8 deliverables.

Supervised Learning

A new statistical background of machine learning based on sequential estimation of distribution mixtures was studied considering a special type of distribution mixtures with product components.

UTIA participated on the WP8 Dublin MUSCLE workshop and started to work in several WP8 eteams.

Publications MP-codes: 170

8.9 Contribution by TCD

Researchers involved: Simon Wilson, Pdraig Cunningham, Rozenn Dahyot

Activities We have begun writing a chapter for the proposed book of Machine Learning for Multimedia Data, to be edited by Matthieu Cord. The chapter will be on Bayesian Methods and Decision Theory.

Publications MP-codes:

8.10 Contribution by TUG

Researchers involved: Horst Bischof, Martina Uray, Helmut Grabner, Michael Grabner, Peter Roth

Activities

1. Realtime training from unlabeled data using a combination of PCA and a MSER-tracker is almost running. Additionally the automatic initialisation problem was solved. Next steps are evaluation of robustness and application to several datasets.
2. Work on incremental LDA is continuing. The retraining concept and the claim of keeping full image information has been proved. Currently we are running tests on large scale databases.
3. Work on incremental Adaboost is continuing. As a new application the conservative learning for automatically learning objects was added to the existing framework.

Publications MP-codes: 217, 262, 263

8.11 Contribution by IBAI

Researchers involved: Petra Perner, Horst Perner

Activities E-Team is working on a paper on Novelty Detection on multi-media data.

Publications MP-codes:

8.12 Contribution by UU

Researchers involved: Niall Rooney, Mykola Galushka, David Patterson

Activities We collaborated with Petra Perner at IBAI and Eric Pauwels of CWI on a paper titled *Novelty Detection - an Overview*. The paper was intended for submission to ICDM 2006. This work was based through participation in the e-team *Unknown pattern recognition*.

Publications MP-codes:

8.13 Contribution by INRIA-Vista

Researchers involved: Ivan Laptev, Patrick Bouthemy, Jean-Pierre Le Cadre, Alexandre Hervieu

Activities Video content recognition using motion trajectories

In November 2005 we started analysis of dynamic content in video using local motion trajectories. We use feature tracking methods previously developed in our team to supply a set of trajectories in a given video segment. The first step will be to structure these trajectories in coherent space-time clusters and to design a graph-based representation of them. After a supervised learning stage, a statistical classification method should allow us to detect events of interest. Trajectories are interesting motion features since they convey elaborated information on the temporal behaviour of the moving entities, which should enable a fine categorization of the dynamic content.

Publications MP-codes:

8.14 Contribution by INRIA-Parole

Researchers involved: Khalid Daoudi, J. Louradour

Activities SVM classification of dynamic sequences.

We proposed a new approach for sequence classification which is based on the framework of Reproducing Kernel Hilbert Spaces. We developed the theoretical material which leads to the formulation of an original sequence kernel, that we implement in a SVM scheme. We used a speaker verification task using NIST SRE data as vehicle to demonstrate the potential of the new kernel. The results show that our new sequence kernel significantly outperforms the generative approach with Gaussian Mixture Models. They also show that it generally outperforms the powerful Generalized Linear Discriminant Sequence kernel, while offering more efficiency and flexibility.

This work led us create an E-team on *Kernel methods* and to establish a collaboration with another MUSCLE partner (CMM) on sequence kernels. As a result of this collaboration we have conceived a new family of kernels for the classification of dynamic observations. The application to SVM speaker verification led to very good performances as compared to existing methods.

Publications MP-codes: 289,290,222

8.15 Contribution by INRIA-Vista

Researchers involved: Ivan Laptev, Patrick Bouthemy

Activities Action recognition with kernel methods

Kernel-based methods in combination with SVM has shown great potential in solving supervised recognition problems. When dealing with time-dependent data, an improvement to standard SVM methods can be achieved by taking the causality of the data into account. One class of such methods involving Sequence Kernel developed by Khalid Daoudi has been successively applied to speaker verification from audio data. As a part of collaboration in E-team *Kernel methods in dynamic problems* we investigate applications of Sequence Kernels to the problem of action recognition in video.

Publications MP-codes:

8.16 Contribution by INRIA-Vista

Researchers involved: Ivan Laptev, Jian-Feng Yao, Patrick Bouthemy, Gwénaëlle Piriou

Activities Auto-models with mixed states for motion texture analysis

Markov random fields models are now a standard tool in image analysis. However, to our knowledge, the existing models deal either with continuous observations, or with discrete observations, but never with observations that can take values of both types. Indeed, any discrete information is usually examined by the introduction of a label process. However, the label process is a latent process and the resulting statistical inference methods need in general to restore it. We have begun to investigate a different approach. The aim is to design a model which automatically deals with the two types of observations, without the introduction of any latent process. The basic idea is then to introduce mixed-state variables (or distributions) in a random field set-up. We have introduced new models for random fields called auto-models with mixed states. We have developed an estimation procedure of the model parameters based on the pseudo-likelihood function. We are exploiting these mixed-state auto-models for the modeling of dynamic textures in videos of natural scenes (such as views of rivers, sea-waves, moving foliage, fire, steam, smokes,...). The considered observations are (locally averaged) normal flow magnitudes which take their values in 0 and $(0, \infty)$ where 0 is here a discrete (symbolic) value since it accounts for no motion. We call them *motion textures*. We have implemented positive Gaussian mixed-state auto-models defined on a spatial 4-nearest-neighbour system. They involve four parameters only. Simulations have demonstrated that this (simple) model is able to reproduce important texture characteristics such as density, contrast, granularity and spatial orientation, and the influence of each parameter was identified [MP-300]. Experiments on real image sequences have shown that basic properties like isotropy/anisotropy, spatial stationarity/non stationarity are well captured by the model. However, this first implemented model needs to be improved. Current extensions are dealing with larger neighborhoods (8-nn, 2 more parameters), multi-scale models, other continuous distributions (the non zero-mean Gaussian distribution has also been introduced, adding four more parameters, and simulations have proven its interest), in order to model long-distance spatial correlations and correlations between continuous values. The addition of the time dimension in the auto-models is also envisaged. Besides, we are investigating the use of these mixed-state auto-models for the segmentation of motion textures in image sequences.

Publications MP-codes: MP-300

8.17 Contribution by INRIA-Imedia

Researchers involved: Valrie Gouet-Brunet, Nizar Grira, Michel Crucianu, Nozha Boujemaa

Activities By including information provided by the user, semi-supervised categorization can produce results that come closer to user's expectations. To make such an approach acceptable for the user, the supervision information must be of a very simple nature and the amount of information the user is required to provide must be minimized. We proposed an effective semi-supervised clustering algorithm, Active Fuzzy Constrained Clustering (AFCC), that minimizes a competitive agglomeration cost function

with fuzzy terms corresponding to pairwise constraints provided by the user. To minimize the amount of constraints required, we define an active mechanism for the selection of candidate constraints. The comparisons performed on a simple benchmark and on a ground truth image database show that with AFCC the results of clustering can be significantly improved with few constraints, making this semi-supervised approach an attractive alternative in the categorization of image databases. See ref. MP-306, MP-307 and MP-308.

Within the WP8, we have proposed the *Active Semi-Supervised Learning* e-team: Multimedia databases are characterized by large amounts of rather cheap data and small amounts of quite expensive knowledge. Among the many existing frameworks for learning, these characteristics support those frameworks that can exploit large amounts of data while using a minimum amount of expensive knowledge. This expensive knowledge can be either available a priori or actively acquired by interaction with users. The aim of this e-team is to:

- Better identify the working assumptions of the existing frameworks (supervised learning, transduction, semi-supervised clustering, unsupervised clustering...) in the context of applications to multimedia data,
- Identify gaps in this spectrum of assumptions,
- Put forward new methods for those gaps that deserve being filled and
- Align existing semi-supervised learning methods to the requirements of multimedia content.

This e-team is coordinated by Imedia.

Publications MP-codes: 306, 307, 308

8.18 Contribution by INRIA-Ariana

Researchers involved: Ian Jermyn, Josiane Zerubia, Johan Aubray, Dan Yu

Activities Dan Yu, a postdoc in INRIA-Ariana this year, began work on parameter estimation for the rotation-symmetric quartic adaptive wavelet packet texture models she is developing. The work of Johan Aubray, stagiaire in INRIA-Ariana last year, on parameter estimation for quartic models of the bimodal statistics of adaptive wavelet packet coefficients of texture was accepted for publication in EUSIPCO 2006. Initial steps in the estimation of the parameters of higher-order active contour models were also taken, with promising results. Saloua Bouatia, an intern in INRIA-Ariana starting work in April 2006, will continue work on this topic.

Publications MP-codes: 314

9 Overview activities in WP 9

9.1 Contribution by CNR-ISTI

Researchers involved: Ovidio Salvetti, P. Asirelli, M. Martinelli,

Activities

- Study of a common representation model (ontology model) able to integrate and describe MM data and their relations and properties in the specific NoE scientific domain. In this framework we started the design of an OWL ontology for image processing and understanding, exploiting also an active collaboration with RAS Moscow.

- Study and development of aid tools for the implementation of a NoE metadata-service available to all the partners to facilitate interoperability and MM data integration. An architecture has been studied and parts of the system have already been implemented;
- The study of the requirements for the development of an XML database for an adequate management of media descriptions has been completed.
- Cooperations have been launched with: aceMedia (Multimedia Ontology Group), Distributed Systems Technology Centre - DSTC Australia, E-team on "Integration of structural and semantic models for multimedia metadata management".
- Continuation of the interaction activity with W3C Italian office
- Cooperation activity with the Networked Multimedia Information Systems Laboratory - Delos partner
- Development of the 4M (MultiMedia Metadata Management) prototype architecture
- Application for an Internal Fellowship in contacts with Jane Hunter
- Integration activity with CEA LIST, Bilkent, IBAI and UPC in the fields of database architectures for multimedia content access, video database system, data preparation of web log files, text analysis and XML metadata extraction and collection.

Events: Workshops, conferences, etc.

- Istanbul Scientific Meeting
- 2005 IEEE Conference "Standards for Global Business - The European Conference on Collaborative Trends in European and Global Standardization" , Munich
- Participation to MultiMedia Ontology Group workshop, London 2005
- Contribution to the Call for a Common MultiMedia Ontology (MultiMedia Ontology Group)
- WWW 2006, Workshop on Semantic Web Annotations for Multimedia (SWAMM), Edinburgh (paper accepted)

Publications MP-codes:

- P. Asirelli, M. Martinelli, O. Salvetti, et al., EWIMT London 2005
- P. Asirelli, M. Martinelli, O. Salvetti, Contribution for a Multimedia Ontology Framework
- Gregory Grefenstette, Fathi Debili, Christian Fluhr, Svitlana Zinger. "Exploiting text for extracting image processing resources" Genoa, Italy, LREC'2006, May 2006
- Meike Reichle, Petra Perner, Klaus-Dieter Althoff, "Data Preparation of Web Log Files for Marketing Aspects Analyses", ICDM 2006, Leipzig 2006.

10 Overview activities in WP 10

10.1 Contribution by TSI-TUC

Researchers involved: Alex Potamianos, Manolis Perakakis, Michalis Toutoudakis

Activities Multi-Modal Dialogue Systems: Modality Selection

The emergence of powerful mobile devices such as personal digital assistants (PDAs) and smart-phones raises new design challenges and constraints that could be better addressed by a combination of more than one modalities such as speech and visual modalities. Combining multiple modalities efficiently is a complex task and requires both good interface design and experimentation to determine the appropriate modality mix. Issues such as device size, interface latency, speech recognition accuracy can seriously affect the efficiency and naturalness of the multimodal interface and bias the optimum modality mix. Few guidelines exist for selecting the appropriate mix of modalities. It is established that the visual modality is more efficient than speech while speech is the more natural interaction mode. However, it is often the case when designing multimodal user interfaces, that the developer is biased either towards the voice or the visual modalities. This is especially true if the developer is voice-enabling an existing graphical user interface (GUI)-based application or building a GUI for an existing voice-only service. Our goal is to follow an approach that respects both modalities, creating an interface that is both natural and efficient. Towards that goal we implement a modality-selection multimodal system that is a mixture of the click-to-talk and open-mike multimodal modes. Modality-selection implements a simple version of the adaptive modality tracking algorithm introduced in Potamianos et al 2003. The unimodal and three multimodal systems are built using the Bell Labs Communicator platform and are evaluated on a travel reservation task. Our focus in this work, is on creating and evaluating various multimodal interaction modes in order to gain valuable experience in the area of modality integration and selection.

Bibliography:

- M. Perakakis, M. Toutoudakis, and A. Potamianos, Modality selection for multimodal dialogue systems, in *Internat. Conf. on Multimodal Interfaces*, (Trento, Italy), Oct. 2005.
- A. Potamianos, E. Ammicht, and E. Fosler-Lussier, Modality tracking in the multimodal Bell Labs Communicator, in *Proc. Automatic Speech Recogn. and Underst. Workshop*, (St. Thomas, U.S. Virgin Islands), Dec. 2003.

Publications MP-codes:**10.2 Contribution by UCL**

Researchers involved: Fred Stentiford, Wole Oyekoya, Li Chen

Activities Work at UCL has concentrated on detecting new forms of saliency that are relevant to the understanding of visual content.

Task 5: Multimedia Information Retrieval Using Novel Interfaces: Eye-tracking.

Work in this area has concentrated on exploring how eyetracking data may be used to expedite the retrieval of visual information. It is to be expected that eye gaze behaviour will change according to the visual search task and it is the objective of this research to use this information to retrieve images from a database.

Experiments have yielded significant evidence that target images may be located faster by eye than by using a mouse interface. Results were reported in a paper presented at EWIMT 2005. A retrieval interface has now been built and tested and a video of a sample saccade sequence can be seen on the Muscle demo web page.

Further experiments have been designed to test the effectiveness of the eye tracking retrieval interface. A set of 1000 images was selected from the Coral database and processed to obtain a matrix of inter-image similarities using an attention based measure. Subjects were asked to search for target images in screens of thumbnails which were successively replaced with new screens of thumbnails based on gaze behaviour and the associated inter-image similarity values. It was found that target images were identified significantly faster than if thumbnails were selected randomly. This work was published in a paper presented at ETRA 2006. More recently it has been found that extremely short screen exposure times have not degraded performance despite complaints from subjects that they felt out of control. It is conjectured that pre-attentive attention is playing a significant part in the search.

Publications MP-codes: 219, 221

10.3 Contribution by UCL

Researchers involved: Fred Stentiford, Wole Oyekoya, Li Chen

Activities Task 5: Multimedia Information Retrieval Using Novel Interfaces: Eye-tracking

Recent experiments have indicated that visual target selection is faster with an eyetracker than with a mouse. This mechanism has now been developed into an interface for image retrieval in which the user is able to search a database of images for a target image just using eye movement. A network of pre-computed inter-image similarities are used to provide successive displays of thumbnail images that are similar to the images that attracted attention in previous displays.

Publications MP-codes: 58, 61, 205, 219, 221, 278

10.4 Contribution by INRIA-Imedia

Researchers involved: Valrie Gouet-Brunet, Nozha Boujemaa, Anne Verroust-Blondet

Activities We have worked on a graphical interface based on Euler-like diagrams for traditional library environments. In [MP-310], interactive Euler diagrams have been introduced to help the user: to build graphically and interactively a complex boolean query, to obtain a partial cartography of the database content from a set of indexes, and to visualize and control the properties and the relevance of the documentary structuration and indexation.

We have also worked on a tool drawing automatically "Euler like" diagrams. In [MP-309], we have presented our first results and proposed graphical methods to enhance the readability of such diagrams.

Publications MP-codes: 308, 309, 310

11 Overview activities in WP 11

11.1 Contribution by UPC

Researchers involved: Montse Pardas, Cristian Canton

Activities

- **3D human action recognition in multiple view scenarios** In this work we developed a novel view-independent approach to the recognition of human gestures of several people in low resolution sequences from multiple calibrated cameras. In contraposition with other multi-ocular gesture recognition systems based on generating a classification on a fusion of features coming from different views, our system performs a data fusion (3D representation of the scene) and then a feature extraction and classification. Motion descriptors introduced by Bobick et al. for 2D data are extended to 3D and a set of features based on 3D invariant statistical moments are computed. Finally, a Bayesian classifier is employed to perform recognition over a small set of actions. Results are provided showing the effectiveness of the proposed algorithm in a Smart-Room scenario.

Publications MP-codes: 212

11.2 Contribution by Bilkent University

Researchers involved: Ugur Gudukbay, Ozgur Ulusoy, Yigithan Dedeoglu

Activities We continued the development of a vision-based Human-Computer Interface that takes input from the user by hand gestures. Movements and the postures of the hand control the position of the mouse cursor and simulate standard mouse actions, such as click or wheel mouse events. The interface will be used to input queries to a Video Database System, called BilVideo.

Publications MP-codes:

11.3 Contribution by INRIA-Ariana

Researchers involved: Ian Jermyn, Josiane Zerubia, Giuseppe Scarpa

Activities Integration activities include: a visit to INRIA-Ariana by Tammy Riklin-Raviv, PhD student with TAU-Visual, to begin work on Shape modelling e-team; a successful application for a Muscle Postdoctoral Fellowship with TAU-Visual to work on Shape modelling e-team; a joint ERCIM postdoc with UTIA, Giuseppe Scarpa, currently at INRIA-Ariana for the second half of his fellowship.

Publications MP-codes:

12 MUSCLE Publications: Papers and Preprints

For more details, consult the MUSCLE online paper archive (www.muscle-noe.org > Research > Papers).

- **MP-134:** Wilson, Simon; Stefanou, George; *Bayesian Approaches to Content-based Image Retrieval*
- **MP-135:** Bar, Leah; Sochen, Nir; Kiryati, Nahum; *Variational Pairing of Image Segmentation and Blind Restoration*
- **MP-137:** Bar, Leah; Sochen, Nir; Kiryati, Nahum; *Image Deblurring in the Presence of Salt and Pepper Noise*
- **MP-146:** Aronowitz, Hagai; Burshtein, David; *Efficient Speaker Identification and Retrieval*
- **MP-147:** Aronowitz, Hagai; Irony, Dror; Burshtein, David; *Modeling Intra-Speaker Variability for Speaker Recognition*
- **MP-187:** Havasi, Laszlo ; Szirnyi, Tams; *Estimation of Vanishing Point in Camera-Mirror Scenes Using Video*
- **MP-217:** Grabner, Michael; Grabner, Helmut; Bischof, Horst; *Fast approximated sift*
- **MP-258:** Huiskes, Mark J. ; Pauwels, Eric J.; *Indexing, Learning and Content-based Retrieval for Special Purpose Image Databases*
- **MP-259:** Argyros, Antonis A. ; Lourakis, Manolis I.A.; *Real-Time Tracking of Multiple Skin-Colored Objects with a possibly Moving Camera*
- **MP-260:** Rangelova, Elena; Pauwels, Eric J.; *Saliency Detection and Matching for Photo-Identification for Humpback Whales*
- **MP-262:** Grabner, Helmut; Bischof, Horst; *On-line Boosting and Vision*
- **MP-263:** Skocaj, Danijel; Uray, Martina; Leonardis, Ales; Bischof, Horst; *Why to Combine Reconstructive and Discriminative Information for Incremental Subspace Learning*
- **MP-264:** Roth, Peter; Bischof, Horst; Skocaj, Danijel; Leonardis, Ales; *Object detection with bootstrapped learning*
- **MP-265:** Grabner, Michael; Bischof, Horst; *Extracting Object Representations from Local Feature Trajectories*
- **MP-266:** Fraundorfer, Friedrich; Winter, Martin; Bischof, Horst; *MSCC: Maximally Stable Corner Clusters*
- **MP-267:** Winter, Martin; Bischof, Horst; Fraundorfer, Friedrich ; *Maximally Stable Corner Clusters: A novel distinguished region detector*
- **MP-268:** Fraundorfer, Friedrich; Bischof, Horst; *A novel performance evaluation method of local detectors on non-planar scenes*
- **MP-271:** Landabaso, Jose Luis; Pardas, Montse; Casas, Josep Ramon; *Reconstruction of 3D shapes considering inconsistent 2D silhouettes*
- **MP-272:** Potamianos, Alexandros; Sanchez-Soto, Eduardo; Daoudi, Khalid; *Stream weight computation for multi-stream classifiers*

- **MP-273:** Karagiorgakis, Panagiotis; Potamianos, Alexandros; Klasinas, Ioannis; *Towards incorporating language morphology into statistical machine translation systems*
- **MP-274:** Pangos, A; Iosif, E; Potamianos, A; *Combining statistical similarity measures for automatic induction of semantic classes*
- **MP-275:** Perakakis, M; Toutoudakis, M; Potamianos, A; *Modality selection for multimodal dialogue systems*
- **MP-276:** Perakakis, M; Toutoudakis, M; Potamianos, A; *Modality selection for multimodal dialogue systems*
- **MP-277:** Oyekoya, Oyewole; Stentiford, Fred; *Exploring the significance of visual attention by eye tracking*
- **MP-280:** Pham, Thang; Smeulders, Arnold; *Efficient projection pursuit density estimation for background subtraction*
- **MP-281:** Schreiber, T. *Dobrushin-Kotecky-Schlosman theorem for polygonal Markov fields in the plane*
- **MP-282:** van Lieshout, M. N. M.; Schreiber, T. *Perfect simulation for length-interacting polygonal Markov fields in the plane*
- **MP-283:** Kotti, Margarita; Martins, Gustavo; Benetos, Emmanouil; Cardoso, Jaime; Kotropoulos, Constantine; *Automatic speaker segmentation using multiple features and distance measures: A comparison of three approaches*
- **MP-284:** Benetos Emmanouil; Kotropoulos, Constantine ; Lidy, Thomas; Rauber, Andreas; *Testing supervised classifiers based on non-negative matrix factorization to musical instrument classification*
- **MP-285:** Delakis, Manolis; Gravier, Guillaume; Gros, Patrick; *Multimodal Segmental-Based Modeling of Tennis Video Broadcasts*
- **MP-286:** Delakis, Manolis; Gravier, Guillaume; Gros, Patrick; *Audiovisual Fusion with Segment Models for Video Structure Analysis*
- **MP-287:** Ouni, Slim; Laprie, Yves; *Modeling the articulatory space using a hypercube codebook for acoustic-to-articulatory inversion*
- **MP-288:** Daoudi, Khalid; Cerisara, Christophe; *The MAP-SPACE denoising algorithm for noise robust speech recognition*
- **MP-289:** Louradour, Jrme; Daoudi, Khalid; *Conceiving a new Sequence Kernel and Applying it to SVM Speaker Recognition*
- **MP-290:** Louradour, Jrme; Daoudi, Khalid; *SVM Speaker Verification using a new Sequence Kernel*
- **MP-291:** Micusik, Branislav; Hanbury, Allan; *Automatic Image Segmentation by Positioning a Seed*
- **MP-292:** Hanbury, Allan; *Analysis of Keywords used in Image Understanding Tasks*
- **MP-293:** Veit, Thomas; Cao, Frdric; Bouthemy, Patrick; *An a contrario decision framework for region-based motion detection*
- **MP-294:** Veit, Thomas; Cao, Frdric; Bouthemy, Patrick; *A maximality principle applied to a contrario motion detection*

- **MP-295:** Cao, Frdric; Veit, Thomas; Bouthemy, Patrick; *Image comparison and motion detection by a contrario methods*
- **MP-296:** Venkatesh, Babu; Prez, Patrick; Bouthemy, Patrick; *Kernel-based robust tracking for objects undergoing occlusion*
- **MP-297:** Venkatesh, Babu; Prez, Patrick; Bouthemy, Patrick; *Robust tracking with motion estimation and kernel-based color modelling*
- **MP-298:** Prez, Patrick; Vermaak, Jaco; *Bayesian tracking with auxiliary discrete processes. Application to detection and tracking of objects with occlusions*
- **MP-299:** Vermaak, Jaco; Maskell, Simon; Briers, Mark; Prez, Patrick; *Bayesian visual tracking with existence process*
- **MP-300:** Bouthemy, Patrick; Hardouin, Ccile; Piriou, Gwnalle; Yao, Jian-Feng; *Mixed-state auto-models and motion texture modeling*
- **MP-301:** Turkan, Mehmet; Dulek, Berkan; Onaran, Ibrahim; Cetin, A. Enis; *Human Face Detection in Video using Edge Projections*
- **MP-302:** Law-To, Julien; Gouet-Brunet, Valrie; Buisson, Olivier; Boujemaa, Nozha; *Labelling the Behaviour of Local Descriptors for Selective Video Content Retrieval*
- **MP-303:** Joly, Alexis; Buisson, Olivier; *Discriminant local features selection using efficient density estimation in a large database*
- **MP-304:** Ferecatu, Marin; Boujemaa, Nozha; Crucianu, Michel; *Hybrid Visual and Conceptual Image Representation in an Active Relevance Feedback Context*
- **MP-305:** Bouteldja, Nouha; Gouet-Brunet, Valrie; Scholl, Michel; *Evaluation of strategies for multiple sphere queries with local image descriptors*
- **MP-306:** Grira, Nizar; Crucianu, Michel; Boujemaa, Nozha; *Semi-supervised fuzzy clustering with pairwise-constrained competitive agglomeration*
- **MP-307:** Grira, Nizar; Crucianu, Michel; Boujemaa, Nozha; *Active semi-supervised clustering for image database categorization*
- **MP-308:** Grira, Nizar; Crucianu, Michel; Boujemaa, nozha; *Semi-supervised image database categorization using pairwise constraints*
- **MP-309:** Thivre, J.; Verroust-Blondet, Anne; Viaud, Marie-Luce; *Drawing diagrams from labelled graphs*
- **MP-310:** Thivre, J.; Viaud, Marie-Luce; Verroust-Blondet, Anne; *Using Euler Diagrams in Traditional Library Environments*
- **MP-311:** Thivre, J.; Viaud, Marie-Luce; Verroust-Blondet, Anne; *Using Euler Diagrams in Traditional Library Environments*
- **MP-312:** Gosselin, PH; Cord, M; *Semantic kernel learning for interactive image retrieval*
- **MP-313:** Katsamanis, Athanassios; Papandreou, George; Pitsikalis, Vassilis; Maragos, Petros; *Multimodal Fusion by Adaptive Compensation for Feature Uncertainty with Application to Audiovisual Speech Recognition*
- **MP-314:** Aubray, Johan; Jermyn, Ian; Zerubia, Josiane; *Nonlinear models for the statistics of adaptive wavelet packet coefficients of texture*

- **MP-315:** Mikes, Stanislav; Haindl, Michal; *Prague Texture Segmentation Data Generator and Benchmark*
- **MP-316:** Haindl, Michal; Simberova, Stanislava; *Restoration of Multitemporal Short-Exposure Astronomical Images*
- **MP-317:** Haindl, Michal; Mikes, Stanislav; *Colour Texture Segmentation Using Modelling Approach*
- **MP-318:** Haindl, Michal; Filip, Jiri; *Modelling of Authentic Reflectance Behaviour in Virtual Environments*
- **MP-319:** Filip, Jiri; Haindl, Michal; *Efficient Image-Based Bidirectional Texture Function Model*
- **MP-320:** Haindl, Michal; Hatka, Martin; *BTF Roller*
- **MP-321:** Haindl, Michal; Grim, Jiri; Pudil, Pavel; Kudo, Mineichi; *A Hybrid BTF Model Based on Gaussian Mixtures*
- **MP-322:** Dobrynin, Vladimir; Patterson, David; Galushka, Mykola; Rooney, Niall; *SOPHIA: An Interactive Cluster Based Retrieval System for the OHSUMED collection*
- **MP-324:** Novovicova, Jana; *Text Document Classification*
- **MP-325:** Yan, Fei; Christmas, William; Kittler, Josef; *A Tennis Ball Tracking Algorithm for Automatic Annotation of Tennis Match*
- **MP-326:** Novovicova, Jana; Svitek, Miroslav; *Performance Parameters Definition and Processing*
- **MP-327:** Novovicova, Jana; Malik, Antonin; *Information-Theoretic Feature Selection Algorithms for Text Classification*
- **MP-328:** Novovicova, Jana; Malik, Antonin; *Text document classification based on mixture models*
- **MP-329:** Christmas, William; Kostin, Alexey; Yan, Fei; Kolonias, Ilias; Kittler, Josef; *A system for the automatic annotation of tennis matches*
- **MP-330:** Novovicova, Jana; Malik, Antonin; Pudil, Pavel; *Feature Selection using Improved Mutual Information for Text Classification*
- **MP-331:** Nlle, Michael; mer, Bernhard; *Representation of Cyclic Colour Spaces within Quantum Space*
- **MP-332:** Pflugfelder, Roman; Bischof, Horst; *Online auto-calibration in man-made worlds*
- **MP-333:** Marinetti, S.; Grinzato, E.; Bison, P.G.; Bozzi, E.; Chimenti, M.; Pieri, G.; Salvetti, O.; *Statistical analysis of IR thermographic sequences by PCA*
- **MP-334:** Chimenti, M.; Di Natali, C.; Mariotti, G.; Paganini, E.; Pieri, G.; Salvetti, O.; *An IR image processing approach for characterising combustion instability*
- **MP-335:** Di Bono, M.G.; Pieri, G. ; Salvetti, O.; *A tool for system monitoring based on artificial neural networks*
- **MP-336:** Colantonio, S.; Di Bono, M.G.; Pieri, G.; Salvetti, O.; *Processing multimedia biomedical information for disease evolution monitoring*
- **MP-337:** Colantonio, S.; Sartucci, F.; Salvetti, O.; *Automatic recognition and classification of cerebral microemboli in ultrasound images*

- **MP-338:** Pieri, G.; Benvenuti, M.; Carnier, E.; Salvetti, O.; *Object detection and tracking in an open and free environment with a moving camera*
- **MP-339:** Tonazzini, A.; Bedini, L.; Salerno, E.; *Independent component analysis for document restoration*
- **MP-340:** Tonazzini, A.; Salerno, E.; Mochi, M.; Bedini, L.; *Blind source separation techniques for detecting hidden texts and textures in document images*
- **MP-341:** Tonazzini, A.; Bedini, L.; Salerno, E.; *Image Analysis on the Archimedes Palimpsest*
- **MP-342:** Kuruoglu, E.E.; Zerubia, J.; *Modeling SAR Images With a Generalization of the Rayleigh Distribution*
- **MP-343:** Esmeir, Saher; Markovitch, Shaul; *Interruptible Anytime Algorithms for Iterative Improvement of Decision Trees*
- **MP-344:** Gabrilovich, Evgeniy; Markovitch, Shaul; *Feature Generation for Text Categorization Using World Knowledge*
- **MP-345:** Hamo, Yaniv; Markovitch, Shaul; *The Compset Algorithm for Subset Selection*
- **MP-346:** Gabrilovich, Evgeniy; Markovitch, Shaul; *Text Categorization with Many Redundant Features: Using Aggressive Feature Selection to Make SVMs Competitive with C4.5*
- **MP-347:** Esmeir, Saher; Markovitch, Shaul; *Lookahead-based Algorithms for Anytime Induction of Decision Trees*
- **MP-348:** Davidov, Dmitry; Gabrilovich, Evgeniy; Markovitch, Shaul; *Parameterized Generation of Labeled Datasets for Text Categorization Based on a Hierarchical Directory*
- **MP-349:** Lindenbaum, Michael; Markovitch, Shaul; Rusakov, Dmitry; *Selective Sampling for Nearest Neighbor Classifiers*
- **MP-350:** Benvenuti, M.; Colantonio, S.; Di Bono, M.G.; Pieri, G.; Salvetti, O.; *Tracking of Moving Targets in Video Sequences*
- **MP-351:** Di Bono, M.G.; Pieri, G.; Salvetti, O.; *Multimedia Target Tracking through Feature Detection and Database Retrieval*
- **MP-352:** Colantonio, S.; Di Bono, M.G.; Pieri, G.; Salvetti, O.; Cavaccini, G.; *System health monitoring using multilevel artificial neural networks*
- **MP-353:** Colantonio, S.; Di Bono, M.G.; Salvetti, O.; *Disease evolution prognosis based on multi-source signals and image analysis*
- **MP-354:** Colantonio, S.; Moroni, D.; Salvetti, O.; *A methodological approach to the study of periodically deforming anatomical structures*
- **MP-355:** Colantonio, S.; Moroni, D.; Salvetti, O.; *MRI Left Ventricle Segmentation and Reconstruction for the Study of the Heart Dynamics*
- **MP-356:** Colantonio, S.; Di Bono, M.G.; Martinelli, M.; Pieri, G.; Salvetti, O.; *Representation and Communication of Multimedia Data and Metadata*
- **MP-357:** Colantonio, S.; Moroni, D.; Salvetti, O.; *Shape comparison and deformation analysis in biomedical applications*
- **MP-358:** Costagli, M.; Kuruoglu, E.E.; Ahmed, A.; *Astrophysical separation images using particle filters. International Conference on Independent Component Analysis and Blind Signal Separation (ICA)*

- **MP-359:** Kuruoglu, E.E.; Tonazzini, A.; Bianchi, L.; *Source separation of astrophysical images modelled with Markov Random Field models*
- **MP-360:** Baccigalupi, C.; Perrotta, F.; De Zotti, G.; Smoot, G.F.; Burigana, C.; Maino, D.; Bedini, L.; Salerno, E.; *Extracting cosmic microwave background polarization from satellite astrophysical maps*
- **MP-361:** Tonazzini, Anna; Salerno, Emanuele; Mochi, Matteo; Bedini, Luigi; *Bleed-through removal from degraded documents using a color decorrelation method*
- **MP-362:** Salerno, E.; Tonazzini, A.; Kuruoglu, E.E.; Bedini, L.; Herranz, D.; Baccigalupi, C.; *Source separation techniques applied to astrophysical maps*
- **MP-363:** Gerace, Ivan; Cricco, Francesco; Tonazzini, Anna; *An Extended Maximum Likelihood Approach for the Robust Blind Separation of Autocorrelated Images from Noisy Mixtures*
- **MP-364:** Tonazzini, Anna; Gerace, Ivan; Cricco, Francesco; *JOINT BLIND SEPARATION AND RESTORATION OF MIXED DEGRADED IMAGES FOR DOCUMENT ANALYSIS*
- **MP-365:** Bedini, Luigi; Bottini, Sergio; Herranz, Diego; Kuruoglu, Ercan Engin; Salerno, Emanuele ; Tonazzini, Anna; *Separation of Dependent Sources in Astrophysical Radiation Maps using Second Order Statistics*
- **MP-366:** Costagli, Mauro; Kuruoglu, Ercan Engin; Ahmed, Alijah; *Astrophysical Source Separation Using Particle Filters*
- **MP-367:** Bedini, Luigi; Herranz, Diego; Salerno, Emanuele ; Baccigalupi, Carlo; Kuruoglu, Ercan Engin; Tonazzini, Anna; *Separation of Correlated Astrophysical Sources Using Multiple-Lag Data Covariance Matrices*
- **MP-368:** Tonazzini, Anna; Gerace, Ivan; *Bayesian MRF-based blind source separation of convolutive mixtures of images*
- **MP-369:** Tonazzini, Anna; Bedini, Luigi; Salerno, Emanuele ; *A Markov Model for Blind Image Separation by a Mean-Field EM Algorithm*
- **MP-370:** Tonazzini, Anna; Bonchi, Francesco; Gnesi, Stefania; Kuruoglu, Ercan Engin; Bottini, Sergio; *Analysis and Modelling of Genomic Data*
- **MP-371:** Tonazzini, Anna; Bedini, Luigi; Salerno, Emanuele ; *Fast correction of bleed-through distortion in grayscale documents by a blind source separation technique*
- **MP-372:** Caiafa, Cesar Federico; Salerno, Emanuele; Proto, Araceli Noemi; Fiumi, Lorenza; *Dependent component analysis as a tool for blind spectral unmixing of remote sensed images*
- **MP-373:** Cord, M.; Gosselin, PH; Philipp-Foliguet, S.; *Stochastic exploration and active learning for image retrieval*
- **MP-374:** Pargellis, A. ; Fosler-Lussier, E; Lee, C.-H. ; Potamianos, A. ; Tsai, A. ; *Auto-induced Semantic Classes*

13 Resource Tables

See updated figures in 6-monthly report.