MUSCLE

Network of Excellence

Multimedia Understanding through Semantics, Computation and Learning

Project no. FP6-507752

Bi-Monthly Progress Report #13: Mar − Apr 2006

Due date of deliverable: 15.06.2006
Actual submission date: 25.06.2006

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

1.1 Contribution by CWI and ERCIM

Researchers involved: Eric Pauwels, Remi Ronchaud

Activities

- Administrative and financial coordination of the network;

- Most coordinating activity during this period focussed on the preparation and presentation of the 2nd Annual Review (Nice, 27-28 April 2006);

- Preparation of successive drafts of JPA3-document;

- Organisation of regular audio conferences

- Reimbursement of MUSCLE integration expenses (mobility support grants);

- Continued organization of the DELOS-MUSCLE summerschool (San Vincenzo, 12-16 June 2006) in collaboration with Nozha Boujemaa (MUSCLE), Rita Cucchiara and Alberto Del Bimbo (DE-LOS). More information can be found at the school's webpage:

  http://www-rocq.inria.fr/imedia/DelosMuscleSummerSchool2006/index.php

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;

- Preparation of webportal for 2nd Annual Review;

- Mobility grants and reporting.

3 Overview activities in WP 3

3.1 Contribution by Seibersdorf Research

Researchers involved: Michael Noelle, Gustasvo Fernandez, Michael Rubik

Activities Preparation of MUSCLE CIS Benchmarking Competition

Publications MP-codes:

3.2   Contribution by CEA

Researchers involved: Moëllic Pierre-Alain, Grefenstette Gregory, Patrick Hede, Christophe Millet

Activities   ImagEVAL campaign :

- evaluation of the the test runs (5 tasks)
- organization of the official evaluation
- start discussion for the final meeting

Publications MP-codes:

3.3   Contribution by CEA

Researchers involved: Mollic Pierre-Alain, Millet Christophe

Activities   Works for the E-Team : Choosing Features for CBIR and Automated Image Annotation

- Selection of the features
- Test for automatic recognition of Animals using SVM
- Construction of the WIKI for the E-Team web page

Publications MP-codes:

3.4   Contribution by CWI

Researchers involved: Mark Huiskes, Mark Huiskes, Renaud Peteri

Activities   Extension of the DynTex database of dynamic textures; in collaboration with Sandor Fazekas, SZTAKI

Publications MP-codes:

3.5   Contribution by INRIA-Ariana

Researchers involved: Ian Jermyn, Josiane Zerubia, Giuseppe Scarpa

Activities   The performance of the textured image segmentation algorithms developed by Giuseppe Scarpa, joint ERCIM postdoc of INRIA-Ariana and UTIA, has been assessed by means of the Prague Texture Segmentation Datagenerator Benchmark created by UTIA. The benchmark system provides users with texture mosaics to be segmented, and the uploaded segmentations are then evaluated and scored using different accuracy indicators.

Publications MP-codes: 454,456

4   Overview activities in WP 4

4.1   Contribution by UPC

Researchers involved: Montse Pardas, Jose Luis Landabaso, Cristian Canton, Josep Ramon Casas

Activities

- Creation of web page for e-team Person Detection, Recognition and Tracking.
- Submission of papers MP425 and MP426 to EUSIPCO 2006 MUSCLE special session.

Publications MP-codes: 425, 426

4.2   Contribution by TUVienna-IFS

Researchers involved: Andreas Rauber, Thomas Lidy, Rudolf Mayer, Rawia Awadallah, Georg Poelzlbauer

Activities   Built testdatabase for CAS e-Team consisting of Video recordings from several partners. Cleansing and re-structuring

Publications MP-codes:

4.3   Contribution by INRIA-Ariana

Researchers involved: Ian Jermyn, Josiane Zerubia

Activities   Dissemination activities include: Josiane Zerubia presented INRIA Ariana's activities within Muscle to the company Magellium in Toulouse, France.

Publications MP-codes:

5   Overview activities in WP 5

5.1   Contribution by AUTH

Researchers involved: Constantine Kotropoulos, Nikoletta Bassiou, Athanasios Papaioannou

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

- N. Bassiou and C. Kotropoulos, "Interpolated Distanced Bigram Language Models for Robust Word Clustering," in CDROM Proc. *2005 IEEE-EURASIP Workshop Nonlinear Signal and Image Processing* , Sapporo, Japan, May 2005.

- C. Kotropoulos and A. Papaioannou, "A novel-updating scheme for probabilistic latent semantic indexing," in Proc. *4th Panhellenic Artificial Intelligence Conf. (SETN-06),* , May 19-20, Heraklion, Greece.

Publications MP-codes: 81,231

5.2 Contribution by GET

Researchers involved: Beatrice Pesquet-Popescu, Olivier Crave

Activities Following our work on adaptive wavelet decompositions, we have proposed a new method for image interpolation using a flexible lifting 3-band scheme. This method allows to perform an invertible conversion from a low resolution image to a 3/2 interpolated image.

Publications MP-codes:

5.3 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

- C. Cotsaces, N. Nikolaidis, and I. Pitas, "Video shot boundary detection and condensed representation: A review," *IEEE Signal Processing Magazine* , Special Issue on Semantic Retrieval of Multimedia, Vol. 23, No 2, pp 28-37, March 2006.

Publications MP-codes: 237

5.4 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

- S. Asteriadis, N. Nikolaidis, A. Hajdu, I. Pitas, A novel eye detection algorithm utilizing edge-related geometrical information, in Proc. *2006 European Signal Processing Conference (EU-SIPCO 06)* , Florence, Italy, September 2006

Publications MP-codes: 406

5.5   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

- E. Benetos, C. Kotropoulos, T. Lidy, and A. Rauber, Testing supervised classifiers based on non-negative matrix factorization to musical instrument classification, in Proc. *XIV European Signal Processing Conf.* , Florence, September 2006 accepted.

Publications MP-codes: 284

5.6   Contribution by MTA-SZTAKI

Researchers involved: Tams Szirnyi, Zoltn Szlvik, Lszl Havasi

Activities Submit a concise description of your WP activities. (You can use simple HTML tags to structure your text.) We describe a general approach that can be used in several different situations for the estimation of scene geometry. Our methods are based on the extraction of temporally coherent features from video sequences. Camera registration and vanishing point estimation show that the proposed co-motion based methods can successfully be used for dynamic scene modeling.

Publications MP-codes: 419

5.7   Contribution by TUG

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

Activities We formulated the tracking problem as a binary classification task in order to achieve robustness by continuously updating the current classifier of the target object with respect to the current surrounding background. For this we used an on-line AdaBoost feature selection algorithm for tracking. The advantage is the capability of updating a model (classifier) during tracking. This allows on the one hand that a classifier can adapt to any object and on the other hand to handle appearance changes (e.g. out of plane rotations, illumination changes) quite naturally. Moreover, depending on the background the algorithm selects the most discriminating features for tracking resulting in stable tracking results. By using fast computable features (e.g. Haar wavelets, integral orientation histograms, local binary patterns) the algorithm runs in real-time. Using this approach we developed a framework for observing static scenes that can be used to detect unknown objects (i.e., left luggage or lost cargo) as well as removing or changing objects in the scene (i.e., theft or vandalism).

Publications MP-codes: 411, 416

5.8   Contribution by UCL

Researchers involved: Fred Stentiford, Li Chen, Rob Shilston, Shijie Zhang

Activities Task 2: Image and Video Processing Sub-task 4: Saliency detection and visual features configuration Work at UCL has continued to explore attention-based mechanisms. The detection of perspective in colour images has yielded good results and a paper is to be presented at ICIP 2006. Investigations of similar architectures for motion detection and estimation show that information on the geometry of the moving object may be extracted. Work is continuing with surveillance videos. An on-line internet demonstration of attention-based focusing has been set up using remote PTZ cameras across the world. Focusing commands are issued depending on the results of image analysis. A paper is to be presented at ICIP 2006. The work is currently directed at the optimization of other image parameters for image informativeness. Attention-based similarity has been applied to the problem of copy detection. The precision and recall results are encouraging. A paper on similarity is to be published in Pattern Recognition and the work on copy detection is to be reported at ICANN 2006 in the Special Session on Visual Attention Algorithms and Architectures for Perceptual Understanding and Video Coding. Further papers have been submitted to CVMP 2006 and IEEE Trans on Circuits and Systems for Video Technology. Li Chen visited INRIA in April to explore areas of mutual interest in video copy detection and make plans for exchanging data. A return visit to UCL is expected in September or October.

Publications MP-codes: 421, 422, 423, 424.

5.9   Contribution by ACV

Researchers involved: Herbert Ramoser, Csaba Beleznai, Julia Puckmayr

Activities Research on combination of motion detection based person localization and visual recognition by a boosted cascade has been started. The performance of single detection modules and that of their combination was analyzed in terms of ROC curves. Additional dependencies considered are for example image resolution and training data sets. Results obtained for the combined system indicate a significant drop in the false alarm rate at an unchanged detection rate.

Publications MP-codes:

5.10 Contribution by UPC

Researchers involved: Montse Pardas, Jose Luis Landabaso, Josep Ramon Casas

Activities

- **Cooperative Background Modelling using Multiple Cameras Towards Human Detection in Smart-Rooms** In multi-camera systems for human detection and analysis, Shape-from-Silhouette (SfS) is a common approach taken to reconstruct the Visual Hull, i.e. the 3D-Shape, of the bodies. The reconstructed 3D-Shape is later used in 3D-trackers and body fitting techniques. The Visual Hull is formally defined as the intersection of the visual cones formed by the backprojection of several 2D binary silhouettes into the 3D space. Silhouettes are usually extracted using a foreground classi- fication process, which is performed independently in each camera view. In this work we develop a novel approach in which 2D-foreground classification is achieved in 3D accordance in a Bayesian framework. In our approach, instead of classifying images and reconstructing the volume later, we simultaneously reconstruct and classify in the 3D space. Furthermore, the classification of the 3D space is used to obtain a more accurate model of the 2D-background.

Publications MP-codes: 425

5.11 Contribution by UPC

Researchers involved: Montse Pardas, Veronica Vilaplana, Ferran Marques

Activities During this period there has been a stage of Stylianos Asteriadis at UPC, within the context of the e-team "Person detection, recognition and tracking". Research was conducted around facial feature detection issues. The AUTH face detection method was combined with a face detection method provided by UPC. In this way, more accurate and reliable results were achieved and problems of scaling were minimized.

Publications MP-codes:

5.12 Contribution by CEA

Researchers involved: Mollic Pierre-Alain, Millet Christophe

Activities Works for the E-Team : Choosing Features for CBIR and Automated Image Annotation
- Selection of the features
- Test for automatic recognition of Animals using SVM
- Construction of "a WIKI" for the E-Team web page

Publications MP-codes:

5.13   Contribution by AUTH

Researchers involved: Stylianos Asteriadis, Nikos Nikolaidis, Ioannis Pitas

Activities   Facial feature extraction (Integration and Research Visit)
Task 2: Image and Video procesing
AUTH PhD student Stylianos Asteriadis visited UPC in order to conduct integration and joint research on issues related to facial feature detection. A face detection method that was used in AUTH (a variant of a face detection method based on Haar features) was combined with a face detection method provided by UPC. In this way, more accurate and reliable results were achieved and scaling problems were minimized. The improvement of the accuracy of the face detection results consisted in refining the face bounding boxes extracted by the Haar face detection used by AUTH, using an ellipse fitting method that gave precise boundaries of the actual face. This improvement in face localization resulted in better success rates for strict criteria on eye centre detection. As an example, considering as successful eye centre detection cases where the maximum distance of the found eye centre from the real one is less than 5% of the actual distance between the eyes, the results obtained are twice as good as the ones before face detection refinement. The results will be submitted for publication as a conference paper. Furthermore, collaboration on the topic of color-based techniques for facial feature detection in difficult cases (e.g. mouth) was initiated. A combination of the two methods used (geometry-based in the case of AUTH and colour-based in the case of UPC) are expected to give further improvements.

Publications MP-codes:

5.14   Contribution by AUTH

Researchers involved: Constantine Kotropoulos, Nikoletta Bassiou

Activities   Color Histogram Equalization
Task 2: Image and Video Processing
Color histogram equalization is revisited by employing techniques from statistical language modeling such as probability smoothing and maximum entropy.
In particular a novel histogram equalization approach is proposed that not only takes into consideration the correlation between color components in the color space, but it is also enhanced by a multi-level smoothing technique adopted from the field of language modeling. In this way, the correlation between color components is taken into account and the problem of unseen values for a color component, either considered independently or in combination with others, is efficiently dealt with. The proposed method is conducted in the HSI color space for intensity (I) component and saturation (S) component given the I component. The quality of the visually appealing equalized images was confirmed by means of the entropy and the Kullback-Leibler divergence estimates between the resulted color histogram and the multivariate uniform probability density function.

Publications MP-codes: 429

5.15   Contribution by ARMINES

Researchers involved: Beatriz Marcotegui,

Activities   Content-based image retrieval usually works in two steps:

1. Feature extraction.

2. Machine learning techniques that through the definition of a distance in the feature domain, are able to classify images in predefined classes or, given an image, are able to find the most similar images in the data base.

Features can be divided in two classes, global and local features. While global features consider the whole image as a unit, local ones take into account spatial information. Local features may be calculated on square blocks, but their performance may be improved if their support corresponds to objects in the image. For that purpose, segmentation approaches are introduced in the indexing schemes.

Many segmentation approaches exist in the literature, and their evaluation for a given application is a crucial point. In the context of the e-team, we have chosen a data base of animal pictures as demonstration field. Manual segmentations have been generated by the e-team members and are considered as ground truth. Several segmentation approaches have been considered, with different parameters and an evaluation method has been proposed to choose the best adapted approach for our application.

The segmentation evaluation method considers the distance of each ground-truth contour point to the closest automatic generated contour. The average of these distances for all ground truth pixels tells us how close are the automatic generated contours to the interesting ones. The same operation is performed inverting the role of ground truth and automatic contours in order to know if the precision of contours is obtained at the price of an oversegmentation, in which case, the distance of automatic generated contours to the closest ground truth pixel would be large.

Publications MP-codes:

5.16   Contribution by MTA-SZTAKI

Researchers involved: Dmitry Chetverikov, Sandor Fazekas

Activities   A novel method for quantitative evaluation of temporal periodicity of dynamic texture has been developed, tested and used for DT classification. The paper has been accepted for publication at British Machine Vision Conference 2006, where the acceptance rate this year is only 25%.

A related paper on dynamic texture recognition with optic flow has been submitted to IEEE Tr. Image Processing.

Publications MP-codes:

5.17   Contribution by UCAM-DENG

Researchers involved: Julien Fauqueur, Nick Kingsbury, Ryan Anderson

Activities   We progressed on the design of our DTCWT wavelet based multiscale keypoint detector: we improved the keypoint energy measure and proposed a new robust scale selection technique. A preliminary version of this work was presented at the BMVA Visual Recognition workshop in March. The most recent progress on this work was submitted and accepted for publication for ICIP 2006.

We have developed a method called Cluster Context Descriptors to match planar objects regardless of viewpoint. It is a local collection of edges whose properties are relatively unique, fast to calculate, and stable to most affine transformations.

We also had two papers accepted for the EUSIPCO conference in Italy this September:

- Rotation-invariant local feature matching with complex wavelets

- Rotation-invariant object recognition using edge profile clusters

Julien Fauqueur is collaborating with Simon Wilson (TCD) and Nozha Boujemaa (INRIA-IMEDIA) on the chapter for the proposed book on Machine Learning for Multimedia Data.

Publications MP-codes: 186,431,432

5.18   Contribution by FORTH

Researchers involved: Panos Trahanias, Antonis Argyros, Manolis Lourakis, Iason Economides

Activities   Regarding symbolic object detection and recognition, FORTH has continued work on visual detection and 2d/3d tracking of objects according to colour information. FORTH has also worked on methods to utilize motion information as a means to improve results in cases that the camera is not moving. Work on the application of the developed tracking mechanisms on problems related to human/computer and human/robot interaction has also been continued.

Publications MP-codes:

5.19   Contribution by GET

Researchers involved: Beatrice Pesquet-Popescu, Grgoire Pau

Activities   A flexible M-band filter bank structure was proposed for fractional scalability. This involves the modification of the synthesis filter bank, without altering the analysis filter bank (and therefore allowing to remain compliant with existing coding standards). The proposed modification allows to change the scale ratio of the reconstructed image without the need of full reconstruction and then downsampling. A paper was published in IEEE Signal Processing Letters and another one (more focused on coding) was accepted for presentation at EUSIPCO 2006.

Publications MP-codes:

5.20   Contribution by INRIA-Imedia

Researchers involved: Julien Law-To, Nozha Boujemaa, Valrie Gouet-Brunet, Alexis Joly

Activities   Within the e-team 'Visual saliency', we worked jointly with UCL (Li Chen, Fred Stentiford) and Vista (Ivan Laptev, Patrick Bouthemy) on a comparative evaluation framework for different visual content description applied to the video copy detection problem.
Ivan Laptev is providing space time interest points; Li Chen is providing global features with videos from BBC. We are providing a framework for the test and our local features dedicated to video copy detection. An objective comparison with the same set of data is important to understand the difference and the complementarities of the different types of features developed in different teams.
The discussion during the two weeks aim at setting the context with different parameters: the data (set of videos), the evaluation method and the different transformations that can occur. These scientific exchanges between our teams benefit of a financial support from Muscle for2 weeks mobility to foster direct discussion and meetings.

Publications MP-codes:

5.21   Contribution by CNR-ISTI

Researchers involved: Ovidio Salvetti, Sara Colantonio, Davide Moroni

Activities   - Study and development of methods for high-level description of complex shape models: analysis of the relevant types of features involved. - Deformation analysis through shape statistics and, more generally, 3D shape analysis and computational anatomy. The results will be presented at the Mass-Data Analysis Workshop, Leipzig, Germany, July 14-15, 2006.

Publications MP-codes: 445

5.22 Contribution by INRIA-Ariana

Researchers involved: Ian Jermyn, Josiane Zerubia, Giuseppe Scarpe, Dan Yu, Peter Horvath, Ting Peng, Saluoa Bouatia

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 of Dan Yu, postdoc in INRIA-Ariana this year, on adaptive wavelet packet texture models has continued. In order to incorporate non-trivial rotation invariance (that is, the probability distributions are rotation invariant, but high probability textures are not necessarily) into the models, she has been examining various bases with improved angular discrimination when compared to standard wavelet bases. A brushlet basis was eventually selected, and Ms Yu has now begun examining the one- and two-point statistics of the brushlet coefficients for a number of textures taken from the Brodatz album and remote sensing images.

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. Ting Peng, joint PhD student of INRIA-Ariana and the LIAMA Institute in Beijing, has been using the models of the one-point statistics of the scaling coefficients of the on-road and off-road regions previously developed to perform maximum likelihood classification of VHR satellite images at different scales. The results are promising: at low resolutions, the new image models performed well even without a region prior, suggesting that these simpler models can be used instead of the gradient-based model developed for medium resolution images. At higher resolutions, however, the results were not that good using maximum likelihood alone. Ms Peng then studied the statistics of the wavelet coefficients of the different classes at different scales. The results suggest that wavelet coefficients can be used to complement the low-resolution information at medium resolution. It appears, however, that at the highest resolution, there is little difference in the one-point statistics of the wavelet coefficients, suggesting that two point statistics may be necessary to extract precise road boundaries.

Region modelling: the higher-order active contour (HOAC) 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. This model is being applied to the extraction of tree crowns from remote sensing images. One of the remaining sources of error in the results obtained arises when the area between two nearby trees strongly resembles a tree crown. The current model tends to form a dumbbell shape, with two circles linked by a 'bridge'. In order to cure this problem, the prior region model must be developed, because local image terms are incapable of distinguishing the background from a real tree. Mr. Horvath has begun work on a new interaction function for the HOAC model that will encourage the 'bridge' to collapse once its width is less than a certain threshold, thus separating the trees.

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. The texture segmentation algorithm based on spectral-spatial-independent clustering has been further improved by replacing the one-step spatial-based clustering with an iterative optimization schedule which eventually provided better performances as highlighted by the experimental results.

Publications MP-codes: 454,455

5.23 Contribution by CNR-ISTI

Researchers involved: Graziano Bertini, Massimo Magrini, Leonello Tarabella, Tommaso Giunti

Activities

Sub-task 3.5. Events detection, segmentation and classification for audiostreams Presentation of an e-team proposal regarding topics related to real-time methods for audio/music processing, restoration and synthesis at WP5 First Focus Meeting (Rocquencourt (FR) 1-2 dec. 2005. With the contribution of the other partners (i.e. IFS-TU Wien) we have re-focused and planned the future activity and the name of the e-team is defined as Semantic from audio: features, perception and synthesis. (see details on wp5 e-teams web-site)

Publications MP-codes:

5.24 Contribution by CNR-ISTI

Researchers involved: Graziano Bertini, Massimo Magrini, Leonello Tarabella, Tommaso Giunti

Activities

Sub-task 3.5 & 3.6 Events detection, segmentation and classification for audiostreams & High-level features extraction for audio Organization of a session and a talk with the progress reports of the e-team Semantic from Audio partners, at the 4th Scientific Meeting (Feb 15-17 , Istanbul, Turkey). An ISTI-CNR TU Wien jointly proposal about Transient enhancemnt of recorded music by semantic driven control has been presented.
Improvements and tests on several genres of recorded music by ARIA methods have been carried out: a preliminary version of demo-list has been arranged and demonstrate in a not-scheduled way at the meeting.

Publications MP-codes:

5.25 Contribution by CNR-ISTI

Researchers involved: Graziano Bertini, Massimo Magrini, Leonello Tarabella, Tommaso Giunti

Activities

Sub-task 3.5 & 3.6 Events detection, segmentation and classification for audiostreams & High-level features extraction for audio Progress in the development of a software plug-in of ARIA algorithm that implements a special dynamic transients enhancement of digitally stored music on PCs machines.
A strategy to estimate the degree of compression of a song based on a suitable energetic features investigation is under study jointly with TUWien (partner and co-leader in Semantic from Audio e-team). These parameters and others derived from a-priori knowledges will be used to optimally set ARIA method in order to match the input music genre.
Partnership of ISTI audio team in a new EC project (M.O.D.E.M. -Music Open Distance Exange Model, Contract n I/05/B/F/PP-154059) regarding a "web-based environment for exchanging multi-track audio/musical signals" development. Some MUSCLE e-team partners could be officially involved later in the system testing.

Publications MP-codes:

5.26 Contribution by UTIA

Researchers involved: Michal Haindl, Stanislav Mikes, Jiri Grim, Jiri Filip, Pavel Vacha

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 segmentation and document classification. This work is described in 8 accepted papers in ICPR 2006 in Hong Kong.

Publications MP-codes:

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 AudioSpecrtumCentroid. 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 MFFCs) before resorting to BIC.
Publications

- M. Kotti, E. Benetos, and C. Kotropoulos, "Automatic speaker change detection with the Bayesian Information Criterion using MPEG-7 features and a fusion scheme," in Proc. *2006 IEEE Symp. Circuits and Systems* , Kos, Greece, May 2006.

- M. Kotti, L. G. P. M. Martins, E. Benetos, J. Cardoso, and C. Kotropoulos, "Automatic speaker segmentation using multiple features and distance measures: A comparison of three approaches," in Proc. *2006 IEEE Int. Conf. Multimedia and Expo* , Toronto, Canada, July 2006.

Publications MP-codes: 235,283

6.2 Contribution by AUTH

Researchers involved: Constantine Kotropoulos, Ioannis Pitas, Margarita Kotti, Bartosz Ziolko, 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

- M. Kotti, C. Kotropoulos, B. Ziolko, I. Pitas, and V. Moschou, "Dialogue detection for movie content description," in Proc. *Int. Workshop Multimedia Content Representation, Classification, and Security* , Istanbul, September 11-13, 2006, accepted.

Publications MP-codes:

6.3   Contribution by FORTH

Researchers involved: Panos Trahanias, Antonis Argyros, Haris Baltzakis

Activities   Work on the development of cross-modal human-friendly interfaces for application in mobile robotic platforms performing in public places such museums, exhibitions and trade fairs, has been continued. More specifically the work focuses on the utilization of a large variety of input (speech input, input via a touch screen, visual input from cameras, and information acquired by a large number of environment sensors) and output modalities (speech, emotion expression via a mechanical face, and robot motion) to provide a multimodal user-robot interface.

Publications MP-codes:

6.4   Contribution by ICCS

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

Activities

Audio-Visual Interaction for 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 speaker's face. 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 developed a statistically motivated scheme for enabling the synergy between object recognition and image segmentation with application to the problem of speakers face detection.

The visual speech 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 audiovisual speech database (obtained from Clemson University).

Our research in the field during the reporting period has focused on adaptive methods for fusing the audio and visual modalities. We have shown 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. We have also extended this framework to generalized sequence models which can account for the asynchrony between audio and visual modalities, such as Product HMMs (P-HMM), obtaining further improvements.

Publications MP-codes: 313,458

6.5   Contribution by ICCS

Researchers involved: George Evangelopoulos, Kostas Rapantzikos, Petros Maragos

Activities

Audiovisual Attention Modeling and Salient Event Detection
Although human perception appears to be automatic and unconscious there exist complex sensory mechanisms that form the preattentive component of human understanding and lead to awareness. Considerable research has been carried out into these preattentive mechanisms and computational models have been developed and employed to common computer vision or speech analysis problems. The separate audio and visual modules may convey explicit, complementary or mutually exclusive information around structures of audiovisual events. We focus on exploring the aural and visual sources of information for modeling attention and subsequent detection of salient (important) events. In any video sequence the two streams are processed in parallel. Based on recent studies on perceptual and computer attention modeling, we extract attention curves using features around the spatiotemporal structure of video and sounds. Audio saliency is captured by modulation-domain signal modeling and multi-frequency band features extracted through non-linear operators and energy tracking. Important audio events, e.g. speech, music, sound effects can then be identified by adaptive threshold-based detection mechanisms. Visual saliency is measured by means of spatiotemporal attention models that combine various feature cues (intensity, color, motion,...) and generate a single saliency map. Statistics are thus extracted in regions of interest obtained through segmentation of this map. Integration of audio and video attention curves is achieved by means of linear and non-linear fusion schemes resulting in a single attention curve, where events supported both from audio and video are enhanced while others may be suppressed or vanish. Event detection at this final audiovisual curve is processed in multiple scales and geometrical features such as local extrema and sharp transition points are extracted that signify the presence of important audiovisual events. The potential of intra-module fusion and audiovisual event detection is demonstrated in applications such as key-frame selection, video skimming and summarization and audio/visual segmentation.

Publications MP-codes:

6.6   Contribution by ICCS

Researchers involved: Petros Maragos, Alexandros Potamianos (TSI-TUC), Patrick Gros (INRIA-TEXMEX)

Activities

Book on *Multimodal Processing and Interaction: Audio, Video, Text* ,

Petros Maragos (ICCS-NTUA), Alexandros Potamianos (TSI-TUC) and Patrick Gros (INRIA-TEXMEX), Editors
The book planned will cover the thematic areas of WPs 6 and 10. It will comprise two main parts: Part A will be a comprehensive State-of-the-Art review of the area and Part B will consist of selected research contributions / chapters by Muscle WP 6/10 members. A rough tentative table of contents follows.

Part I: State-of-the-art report(s)   Merge WP6/WP10 state of the art reports and update

Part II: New research directions   Possible thematic areas

1. Multimodal Processing, Interaction and Understanding multimedia content

   - Audio-Visual ASR
   - Feature fusion
   - Video Analysis and Integration of Asynchronous Modalities

2. Audio-Visual Saliency

   - Audio-Visual Scene Change and Dialogue Detection
   - Audio-Visual Attention and Salient Event Detection

3. Searching multimedia content

   - Annotation of multimedia databases
   - Information retrieval for video or other multimedia databases
   - Integration of Vision + Text or Audio + Text

4. Interfaces to multimedia content

   - Multimodal dialogue interfaces
   - Eye-tracking interfaces for information retrieval
   - Mobile interfaces

Publications MP-codes:

6.7   Contribution by UTIA

Researchers involved: Michal Haindl, Pavel Zid

Activities   UTIA work in WP6 involved segmentation of range data.
A new fast multimodal (range and spectral data) range segmentation algorithm for scenes comprising general faced objects was developed. The range segmentation is based on a recursive adaptive probabilistic detection of step discontinuities which are present at object face borders in mutually registered range and intensity data. Detected face outlines guides the subsequent region growing step where the neighbouring face curves are grouped together. Region growing based on curve segments instead of pixels like in the classical approaches considerably speed up the algorithm.

Publications MP-codes:

7 Overview activities in WP 7

No WP7-specific activities reported for this 2-month period.

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 a 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 assessement 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

- C. Kotropoulos and V. Moschou, Self-organizing maps for reducing the number of clusters by one on simplex subspaces, in *Proc. 2006 IEEE Int. Conf. Acoustics, Speech, and Signal Processing* , May 14-18, Toulouse, France, accepted.

Publications MP-codes: 236

8.2 Contribution by MTA-SZTAKI

Researchers involved: Tams Szirnyi, Zoltn Szlvik

Activities A new Bayesian method is presented for the automatic extraction of common areas of images in multicamera systems through the detection of concurrently changing pixels. Unlike existing still-image and motionbased methods our approach does not need any a priori information about the scene, the appearance of objects in the scene, or their motion. The method is validated by demonstrating its successful use on several real-life outdoor stereo video image-sequence pairs.

Publications MP-codes: 418

8.3 Contribution by TUG

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

Activities

1. Real-time training from unlabeled data with an automatic initialisation using a combination of PCA and a MSER-tracker is running. The main idea is to combine a robust background model, a tracker and an on-line learning method feeding the learning algorithm with patches obtained by the tracker. The tracker is initialized automatically by change detection, thus no user interaction is needed.

2. A robust Active Shape Model was developed that can handle corrupted shape data. Furthermore the model can be expanded on-line using a robust incremental PCA algorithm. Thus, an already partially learned Active Shape Model can be used for segmentation of a new image and the result of this segmentation process contributes to an on-line update of the robust model. This approach is computationally much more efficient than previous ASMs using batch or iterated batch PCA.

Publications MP-codes: 412, 413, 414, 415

8.4   Contribution by UU

Researchers involved: Niall Rooney, David Patterson, Mykola Galushka

Activities   We carried out further development in search technoologies for document collections indexed using SOPHIA cluster -based indexing. We looked at both specific and broad based query mechanisms. We carried out studies into appropriate subclustering mechanisms consistent with the thematic nature of SOPHIA clusters.We sucessfully applied for funding to visit Simon Wilson at TCD, to discuss extending this work for image and textual documents.

Publications MP-codes:

8.5   Contribution by ENSEA

Researchers involved: Matthieu Cord, David Picard, Guillermo Chavez, Philippe Gosselin

Activities   We have done work on CBIR and video analysis. Our content-based retrieval system is extended to the distributed database context. Some experiments have been proposed on a local network. This work is managed by D. Picard. A publication related to this developpement is available MP-428.
M. Cord also manages the Machine Learning book project. M. Cord also met F. Bach (from armines-cmm partner) for discussions on kernel functions for vector sets.

Publications MP-codes: MP-428

8.6   Contribution by ENSEA

Researchers involved: Matthieu Cord, David Picard, Guillermo Chavez, Philippe Gosselin

Activities   We have done work on CBIR and video analysis. Our content-based retrieval system is extended to the distributed database context. Some experiments have been proposed on a local network. This work is managed by D. Picard. A publication related to this developpement is available MP-428.
M. Cord also manages the Machine Learning book project. M. Cord also met F. Bach (from armines-cmm partner) for discussions on kernel functions for vector sets.

Publications MP-codes: MP-428

8.7 Contribution by AUTH

Researchers involved: Constantine Kotropoulos, Dimitrios Ververidis

Activities Feature selection based on mutual correlation
Feature selection is a critical procedure in many pattern recognition applications. There are two distinct mechanisms for feature selection namely the wrapper methods and the filter methods. The filter methods are generally considered inferior to wrapper methods, however wrapper methods are computationally more demanding than filter methods. A novel filter feature selection method based on mutual correlation is proposed. We assess the classification performance of the proposed filter method by using the selected features to the Bayes classifier. Alternative filter feature selection methods that optimize either the Bhattacharrrya distance or the divergence are also tested. Furthermore, wrapper feature selection techniques employing several search strategies such as the sequential forward search, the oscillating search, and the sequential floating forward search are also included in the comparative study. A trade off between the classification accuracy and the feature set dimensionality is demonstrated on both two benchmark datasets from UCI repository and two emotional speech data collections.

Publications MP-codes:

8.8 Contribution by IBAI

Researchers involved: Petra Perner, Horst Perner

Activities Preparation of the Muscle sponsored event Mass Data Analysis of Signals and Images, MDA 2006 www.mda-signals.de
Further work on novelty detection paper.

Publications MP-codes:

8.9 Contribution by FORTH

Researchers involved: Panos Trahanias, Stathis Kalsderidis, Haris Baltzakis

Activities During the reporting period, FORTH continued to be active in tasks 3 and 5 and more specifically in tasks related to extraction and selection of features from multimedia streams acquired by mobile robots. The extracted features are used for the creation of a robot concept system and the corresponding abstraction mechanism that will not depend on predefined feature sets but select/discover the appropriate features on the fly. Development of a unified representation of concepts will be followed so as to transform precepts to concepts to be further used by the reasoning process for action generation and learning percepts.

Publications MP-codes:

8.10 Contribution by FORTH

Researchers involved: Panos Trahanias, Antonis Argyros, Haris Baltzakis, Manolis Lourakis

Activities During the reporting period, FORTH continued to be active in fields related to the second of the two grand challenges related to WP9 (Detecting and interpreting humans and human behaviour in videos). Emphasis was given to research related to recognition and interpretation of hand gestures for human/computer and human/robot interraction.

Publications MP-codes:

8.11 Contribution by INRIA-Vista

Researchers involved: Ivan Laptev, Patrick Bouthemy

Activities   Action recognition with kernel methods
We continued the work on action recognition with kernel methods in E-team *Kernel methods in dynamic problems*. Extension of the current Sequence Kernel SVM to multi-class problems has been considered. Jerome Louradour has performed initial experiments to compare performance of Sequence Kernel with the performance of Local Feature Kernel applied to the problem of recognizing human actions. Both theoretical and practical advantages of the Sequence Kernel are being investigated such as appropriate treatment of outliers within the framework.

Publications MP-codes:

8.12 Contribution by CNR-ISTI

Researchers involved: Ovidio Salvetti, Sara Colantonio, Davide Moroni

Activities   - Analysis and development of advanced NN models suitable to address effective and optimized categorization and recognition problems. - Image segmentation, based on a fuzzy-neural features classification. Tests have been performed on cytological images. Results are reported in a paper that will be presented at the Mass-Data Analysis Workshop, Leipzig, Germany, July 14-15, 2006. - Design of an advanced neural model for the recognition and categorization of the deformation pattern of elastic objects. Results will be presented at the Mass-Data Analysis Workshop, Leipzig, Germany, July 14-15, 2006. - Algorithms for active video surveillance, based on stereo and infrared imaging. Active tracking, using a CBR paradigm adopted for re-localizing lost target. Results will be presented at the MUSCLE Special Session Recognizing humans and human behavior in video of the 14th European Signal Processing Conference EUSIPCO 2006, 4-8 September, Florence Italy.

Publications MP-codes: 443-444-445

8.13 Contribution by INRIA-Ariana

Researchers involved: Ian Jermyn, Josiane Zerubia, Peter Horvath, Saloua Bouatia

Activities   Work has begun on the construction of a phase diagram (that is, a systematization of the various possible categories of global and local minima as a function of the parameter values) for the quadratic higher-order active contour prior energy using an interaction function that decreases monotonically with distance. Such a diagram will enable the choice of parameters to achieve any one of the geometries represented in the diagram, and will thus represent a full understanding of this class of models. Parameter ranges leading to two distinct phases, stable circles and stable periodic bars, are already known, but it is not clear whether irregular networks can ever be global, as opposed to local, minima, nor whether there are other phases for particular parameter ranges. Saloua Bouatia, an intern in INRIA-Ariana since April 2006, is contributing to this work by performing stability calculations on bar shapes. This will enable the selection of parameters leading to bars of a certain width and with prescribed stability properties, factors important for the extraction of road networks and for future applications. the previous work of Peter Horvath is also a contribution to this work.

Publications MP-codes: 455

8.14 Contribution by UTIA

Researchers involved: Michal Haindl, Petr Somol, Jana Novovicova

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

Feature Selection Concurrent advance has been taking place regarding Floating Hybrid Search methods. A non-standard application of feature selection algorithms has been proposed in the form of multi-subset search. This result was accepted for ICPR 2006 conference publication. Additionally the filter feature selection methods were investigated in cooperation with AUTH partners.

Publications MP-codes:

9   Overview activities in WP 9

9.1   Contribution by GET

Researchers involved: Beatrice Pesquet-Popescu, Christophe Tillier, Sebastien Brangoulo

Activities   In Sept. 2005, we participated to the 74th MPEG meeting (joint with ITU/JVT) in Nice, France. In the Vidwav Ad-Hoc Group, we continued to promote the scalable video coding format based on the motion-compensated wavelet technology. We contributed to several input and output documents of this standardization body.
In Dec. 2005, we participated to the 75th MPEG meeting (joint with ITU/JVT) in Bangkok, Thailand. The activity in Vidwav AHG continued, in particular with the set-up of new tests for HD sequences.
In April 2006 we participated to the Vidwav activities at the 76th MPEG meeting in Montreux, Switzerland.

Publications MP-codes:

9.2   Contribution by CNR-ISTI

Researchers involved: Ovidio Salvetti, Massimo Martinelli, Patrizia Asirelli

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: (1) aceMedia (Multimedia Ontology Group), (2) Distributed Systems Technology Centre - DSTC Australia, (3) E-team on "Integration of structural and semantic models for multimedia metadata management".

10   Overview activities in WP 10

10.1   Contribution by UCL

Researchers involved: Fred Stentiford, Wole Oyekoya

Activities   Task 5: Multimedia Information Retrieval Using Novel Interfaces: Eye-tracking Work in this period has been directed at a deeper analysis of results produced to date and the production of a commercial evaluation of the prospects for eye tracking technology in the future. The most recent results are reported in a paper to be presented at the International Workshop on Intelligent Computing in Pattern Analysis/Synthesis 2006. Further experiments are underway to investigate the part that pre-attentive attention may play in eye gaze-driven search.

Publications MP-codes: 420

10.2   Contribution by INRIA-Texmex

Researchers involved: Patrick Gros, Manolis Delakis

Activities   We study the use of Hidden Markov Models and Segment Models for tennis video structure analysis. SMs is a generalization of HMMs that receives segmental features instead of frame-based observations. We argued that the use of segmental features can be beneficial for multimodal integration. The drawback of SMs is that we need to define segmental distributions, i.e., distributions over sequences of data. We have used so far HMMs as they can provide a probabilistic segmental score.

In the last two months we have explored the idea of using Recurrent Neural Networks to compute a segmental score, resulting thus to SM-RNN hybrid. RNNs receive as input a whole sequence of data and produce a unique output at the end that is used for classification purposes. The advantage of RNNs over HMMs is that they are discriminative models and thus may be more performant compared to HMMs. The disadvantage is that the Back-Propagation Through Time and Real-Time Recurrent Learning algorithms that are developed to train RNNs cannot efficiently propagate the gradient through long time lags, and thus cannot discover relative dependencies and features from the whole data sequence. To overcome this problem, Schmidhuber et al. introduced the Long-Short Term Memory, an RNN model that uses input, output and forget gates to efficiently distribute gradient over long time lags. We have implemented and tested LSTM in our data and compared it to BPTT-trained RNNs. BPTT failed to learn our segmental data as they exhibits time lags of order 5 to 20 time steps. On the contrary, LSTM learned quickly a good solution and clearly outperformed BPTT, despite its complexity. LSTM also slightly outperformed HMMs in a segmental classification task. To boost the generalization abilities of LSTM, a great deal of artificial examples were created by white noise addition. Nevertheless, LSTM was trained form the scratch, while HMMs were suitable initialized with a great deal of prior knowledge. We plan to incorporate LSTM to the SM.

Publications MP-codes: 457

11   Overview activities in WP 11

11.1   Contribution by GET

Researchers involved: Beatrice Pesquet-Popescu, Maria Trocan

Activities   In the framework of the E-team "3-D Texture Analysis and Detection" GET worked on using an adaptive LMS algorithm in the predict step of a lifting-based 3-D motion-compensated wavelet transform and applied this technique to scalable video coding. A common paper with Bilkent Univ. was presented to the IEEE ICASSP 2006 conference and another one, comparing linear and non-linear approaches, was accepted for publication in EUSIPCO2006.

Publications MP-codes:

11.2 Contribution by Bilkent University

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

Activities We continue to develop develope 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:

11.3 Contribution by Bilkent University

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

Activities We have developed 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 support vector machine (SVM) based classifiers. By means of this study on human face detection, we have been working on the same subject in omnidirectional cameras. We warp a virtual camera for region of interest (ROI), and correct the vision of ROI using signal processing techniques. Then, we determine possible face candidate regions using TSL colorspace information with Gaussian skin color model. Each region is filtered by a high-pass filter of a wavelet transform as our initial study, and so on.

Publications MP-codes: 301

11.4 Contribution by UPC

Researchers involved: Montse Pardas, Cristian Canton, Josep Ramon Casas

Activities Human Model and Motion Based 3D 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. A simple ellipsoid body model is fit to incoming 3D data to capture in which body part the gesture occurs thus increasing the recognition ratio of the overall system and generating a more informative classification output. 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 SmartRoom scenario.

Publications MP-codes: 426

11.5 Contribution by GET

Researchers involved: Beatrice Pesquet-Popescu, Christophe Tillier, Maria Trocan

Activities We have broadened the study of temporal analysis structures to a 5-band motion-compensated filter bank, which proved to be more efficient than the previous 2-band and 3-band schemes in applications such as video surveillance. A paper was accepted for presentation to IEEE Multimedia Signal Processing Workshop, Oct. 2006.

Publications MP-codes:

11.6   Contribution by INRIA-Ariana

Researchers involved: Ian Jermyn, Josiane Zerubia, Giuseppe Scarpa

Activities Advertisements have been placed to collect candidates for the Muscle postdoctoral fellowship awarded to INRIA-Ariana and TAU Visual. Several applications have already been received, and a final decision will be taken within the next few weeks. The successful candidate will start his or her postdoctoral work at TAU-Visual in September. This work will contribute to the Shape Modelling e-team. Anil Kokaram and Claire Gallagher from the Department of Electrical Engineering at Trinity College Dublin visited the Ariana project to give a seminar and to discuss possible ways of continuing previous collaborations, notably within EU FP5 project Moumir, whose aims were closely related to those of the Muscle project. Giuseppe Scarpa, joint ERCIM postdoc with UTIA, continues the second half of his fellowship at INRIA-Ariana.

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-72:** Dedeoglu, Yigithan; Toreyin, B. Ugur; Gudukbay, Ugur ; Cetin, A. Enis; *Real-time fire and flame detection in video*

- **MP-186:** Fauqueur, Julien; Kingsbury, Nick; Anderson, Ryan; *MULTISCALE KEYPOINT DETECTION USING THE DUAL-TREE COMPLEX WAVELET TRANSFORM*

- **MP-190:** Bamidele, Adetokunbo; *An attention-based model applied to colour histogram indexing*

- **MP-196:** Stentiford, Fred; *Attention based facial symmetry detection*

- **MP-199:** Stentiford, Fred; *Attention based symmetry detection in colour images*

- **MP-206:** Stentiford, Fred; *Attention based colour correction*

- **MP-207:** Davis, Marc; Smith, Michael; Stentiford, Fred; Bamidele, Adetokunbo; Canny, John; Good, Nathan; King, Simon; Janakiraman, Rajkumar; *Using Context and Similarity for Face and Location Identification*

- **MP-219:** Oyekoya, Oyewole; Stentiford, Fred; *An eye tracking interface for image search*

- **MP-237:** Cotsaces, Costas; Nikolaidis, Nikolaos; Pitas, Ioannis; *Video shot boundary detection and condensed representation: A review*

- **MP-244:** Rochery, Marie; Jermyn, Ian; Zerubia, Josiane; *Higher-order active contours*

- **MP-259:** Argyros, Antonis A. ; Lourakis, Manolis I.A.; *Real-Time Tracking of Multiple Skin-Colored Objects with a possibly Moving Camera*

- **MP-325:** Yan, Fei; Christmas, William; Kittler, Josef; *A Tennis Ball Tracking Algorithm for Automatic Annotation of Tennis Match*

- **MP-405:** Laszlo Havasi, Tamas Sziranyi and Michael Rudzsky *ADDING GEOMETRICAL TERMS TO SHADOW DETECTION PROCESS*

- **MP-406:** Asteriadis, Stelios; Nikolaidis, Nikos; Hajdu, Andras; Pitas, Ioannis; *A Novel Eye Detection Algorithm Utilizing Edge-Related Geometrical Information*

- **MP-410:** Dedeoglu, Yigithan; Toreyin, B. Ugur; Gudukbay, Ugur; Cetin, A. Enis; *Sillouette-Based Object Classification and Human Action Recognition in Video*

- **MP-411:** Grabner, Helmut; Roth, Peter; Grabner, Michael; Bischof, Horst; *Autonomous Learning a Robust Background Model for Change Detection*

- **MP-412:** Roth, Peter; Bischof, Horst ; *On-line Learning a Person Model from Video Data*

- **MP-413:** Roth, Peter; Fussenegger, Michael; Bischof, Horst; Pinz, Axel ; *Incremental Robust Learning an Active Shape Model*

- **MP-414:** Roth, Peter; Donoser, Michael; Bischof, Horst; *Tracking for Learning an Object Representation from Unlabeled Data*

- **MP-415:** Roth, Peter; Donoser, Micheal; Bischof, Horst; *On-line Learning of Unknown Hand Held Objects via Tracking*

- **MP-416:** Grabner, Michael ; Grabner, Helmut; Bischof, Horst; *Real-Time Tracking with On-line Feature Selection*

- **MP-418:** Szlavik, Zoltan; Sziranyi, Tamas; *Bayesian Estimation of Common Areas in Multi-Camera Systems*

- **MP-419:** Szlavik, Zoltan; Havasi, Laszlo; Sziranyi, Tamas; *Analysis of dynamic scenes by using co-motion statistics*

- **MP-420:** Oyekoya, Oyewole; Stentiford, Fred; *Perceptual Image Retrieval Using Eye Movements*

- **MP-421:** Stentiford, Fred; *Attention-based vanishing point detection*

- **MP-422:** Shilston, Rob; Stentiford, Fred; *An attention-based focus control system*

- **MP-423:** Chen, Li; Stentiford, Fred; *An attention based similarity measure for colour images*

- **MP-424:** Stentiford, Fred; *Attention based similarity*

- **MP-425:** Landabaso, Jose Luis; Pardas, Montse; *Cooperative Background Modelling using Multiple Cameras*

- **MP-426:** Canton, Cristian; Casas, Josep R.; Pardas, Montse; *Human Model and Motion Based 3D Action Recognition in Multiple View Scenarios*

- **MP-427:** Christophe, Millet; Gregory, Grefenstette; Isabelle, Bloch; Pierre-Alain, Mollic; Patrick, Hde; *Automatically populating an image ontology and semantic color filtering*

- **MP-428:** Picard, david; Cord, Matthieu; Revel, Arnaud; *CBIR in distributed databases using a multi-agent system*

- **MP-429:** Bassiou, Nikoletta; Kotropoulos, Constantine ; *Color histogram equalization using probability smoothing*

- **MP-430:** Kotti, Margarita; Kotropoulos, Constantine; Ziolko, Bartosz; Pitas, Ioannis; Moschou, Vassiliki; *Dialogue detection for movie content description*

- **MP-431:** Anderson, Ryan; Kingsbury, Nick; Fauqueur, Julien ; *Rotation-invariant object recognition using edge profile clusters*

- **MP-432:** Kingsbury, Nick; *Rotation-invariant local feature matching with complex wavelets*

- **MP-433:** Chetverikov, Dmitry; Fazekas, Sandor; *On motion periodicity of dynamic textures*

- **MP-434:** Asirelli, Patrizia; Martinelli, Massimo; Salvetti, Ovidio; *An Infrastructure for Multi-Media Metadata Management*

- **MP-435:** Asirelli, Partizia; Martinelli, Massimo; Salvetti, Ovidio; *Contribution for a Multimedia Ontology Framework*

- **MP-436:** Argyros, Antonis A.; Lourakis, Manolis I.A.; *Tracking Skin-colored Objects in Real-time*

- **MP-437:** Argyros, Antonis A.; Lourakis, Manolis I.A.; *Binocular Hand Tracking and Reconstruction Based on 2D Shape Matching*

- **MP-438:** Lourakis, Manolis I.A.; Argyros, Antonis A.; *Chaining Planar Homographies for Fast and Reliable 3D Plane Tracking*

- **MP-439:** Lourakis, Manolis I.A.; Argyros, Antonis A.; *Is Levenberg-Marquardt the Most Efficient Optimization Algorithm for Implementing Bundle Adjustment?*

- **MP-440:** Lourakis, Manolis I.A.; Argyros, Antonis A.; *Fast Trifocal Tensor Estimation Using Virtual Parallax*

- **MP-441:** Argyros, Antonis A.; Lourakis, Manolis I.A.; *Vision-based Interpretation of Hand Gestures for Remote Control of a Computer Mouse*

- **MP-442:** Lourakis, Manolis I.A.; Argyros, Antonis A.; *Exploiting the Sparseness of Bundle Adjustment for Efficient 3D Reconstruction*

- **MP-443:** Pieri, Gabriele; Salvetti, Ovidio; *ACTIVE VIDEO-SURVEILLANCE BASED ON STEREO AND INFRARED IMAGING*

- **MP-444:** Colantonio, Sara; Gurevich, Igor B.; Salvetti, Ovidio; *Automatic Fuzzy-Neural based Segmentation of Microscopic Cell Images*

- **MP-445:** Moroni, Davide; Perner, Petra; Salvetti, Ovidio; *A General Approach to Shape Characterization for Biomedical Problems*

- **MP-446:** Bertini, Graziano; Gonzalez, Diego; Grassi, lorenzo; Magrini, Massimo; Fontana, Federico; *Voice Transformation Algorithms with Real Time DSP Rapid Prototyping Tools*

- **MP-447:** Bertini, Graziano; Magrini, Massimo; *A Prototype Lab Box with DSK C6711/13 for Rapid DSP Algorithms Development*

- **MP-448:** Leonello *Tarabella*

- **MP-449:** Roh, Myung-Cheol; Christmas, William; Kittler, Josef; Lee, Seong-Whan; *Robust Player Gesture Spotting and Recognition in Low-Resolution Sports Video*

- **MP-452:** Roh, Myung-Cheol; Christmas, William; Lee, Seong-Whan; Kittler, Josef; *Robust Gesture Spotting in Low-Quality Sports Video*

- **MP-453:** Bertini, Graziano; Magrini, Massimo; Tarabella, Leonello; *An interactive musical exhibit based on infrared sensor*

- **MP-454:** Scarpa, Giuseppe; Haindl, Michal; *Unsupervised Texture Segmentation by Spectral-Spatial-Independent Clustering*

- **MP-455:** Horvath, Peter; Jermyn, Ian; Kato, Zoltan; Zerubia, Josiane; *A Higher-Order Active Contour Model for Tree Detection*

- **MP-456:** Poggi, Giovanni; Scarpa, Giuseppe; Zerubia, Josiane; *Supervised Segmentation of Remote Sensing Images Based on a Tree-Structured MRF Model*

- **MP-457:** Delakis, Manolis; Gravier, Guillaume; Gros, Patrick; *Score oriented Viterbi search in sport video structuring using HMM and segment models*

- **MP-458:** Pitsikalis, Vassilis; Katsamanis, Athanassios; Papandreou, George; Maragos, Petros; *Adaptive Multimodal Fusion by Uncertainty Compensation*

13   Resource Tables

See next page. Notice that the WP-numbering refers to new WP-organisation as detailed in JPA3.

MUSCLE - Effort Table 13 March/April 2006
(Person Month effort per Partner per Workpackage)

Participant	WP01a	WP01b	WP02	WP03	WP04	WP05	WP06	WP07	TOTAL
01- ERCIM	2.07							0.4	2.47
02 - CWI		0.9	0.3	1.07	0.84		1.3		4.40
03 - UCL				0.35	0.25	1			1.60
04 - KTH			0.1	0.85	0.66				1.60
05 - BILKENT		0.2	0.4	0.35	0.25		0.4	1.8	3.40
06 - VIENNA PRIP			0.2	2.00	1.60				3.80
07 - MTA SZTAKI			0.5	0.19	0.12		1	0.3	2.10
08 - UU							1.00		1.00
09 - CNR-ISTI		0.1	0.4	0.05	0.05		0.2		0.80
10 - FT									
11 - TUG		0.19	0.4	0.35	0.25		1.3		2.49
12 - UPC			0.1	0.19	0.12				0.40
13 - UFR					0.05		0.25		0.75
14 - UTIA			1				1.3	0.2	2.50
15 - UVA						1.1			1.10
16 - AUTH				0.22	0.14			0.3	0.66
17 - CEA			0.2	0.35	0.25	0.2			1.00
18 - TU VIENNA IFS				0.90	0.70	0.5	0.5	0.25	2.85
19 - ACV				0.10	0.04				0.14
20 - TECHNION-ML							0.1		0.10
21 - TECHNION-MM				0.16	0.10	0.38		0.65	1.29
22 - IBAI		0.11	0.33				0.67	0.22	1.33
23 - ICCS		0.1	0.1	0.02	0.05	1.2		0.2	1.67
24 - TSI-TUC		0.09	0.03			1.69	0.03	0.37	2.21
25 - ARMINES		0.1		0.35	0.25				0.70
26 - TAU-SPEECH			1.25	0.21	0.14		0.5		2.10
27 - TAU-VISUAL				1.45	1.15				2.60
28 - SEIBERSDORF			0.6						0.60
29 - TCD		0.2	2.8	0.30	0.21		4.7		8.20
30 - FORTH				0.26	0.17	0.83	0.22	1.3	2.78
31 - VTT		0.2				0.33			0.53
32 - INRIA Ariana		0.32	0.23	0.01	0.04		0.35	0.35	1.30
32 - INRIA Imedia				0.02	0.05	1	0.3	1	2.37
32- INRIA Parole		0.1	0.7	0.08	0.03			0.2	1.10
32- INRIA Tex Mex						0.93			0.93
32 - INRIA Vista		0.06	0.2	0.48	0.36	0.21	0.22	0.44	1.97
33 - GET				2.50	2.01		1.30		5.81
36 - UCAM-DENG			0.6	0.13	0.07		0.60		1.40
34 - LTU			1.5						1.50
35 - UNIS			0.19	0.04	0.05	0.45	0.19	0.73	1.65
37 - ENSEA		0.2		0.05	0.05		0.4		0.70
38 - CNRS			0.25	0.04	0.04				0.33
39 - UPS – IRIT									
40 - EC3									
41 - UPMC									
42 - NUID / UCD									
Total	2.07	2.87	12.38	13.04	10.06	9.82	16.83	8.71	76.24