MUSCLE

Network of Excellence

Multimedia Understanding through Semantics, Computation and Learning

Project no. FP6-507752

Bi-Monthly Progress Report: Sep – Oct 2005

Due date of deliverable: 15.12.2005
Actual submission date: 15.12.2005

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

2 Overview activities in WP 2

2.1 Contribution by CWI

Researchers involved: Eric Pauwels, Margriet Brouwer

Development of PHP/MySQL tools for administrative and scientific reporting in WPs.

3 Overview activities in WP 3

3.1 Work completed

- **CEA** finished the CLIC database and transfered it to the IRIT group (25/10/2005). The next steps are testing the downloading protocols (according to the capabilities of the IRIT - the database is quite huge, 60 GB), and checking some legal stuff with the CEA.
- **IBAI** has done some benchmarking on object matching algorithms.
- **UTIA** has further upgraded and tested its Prague Texture Segmentation Data Generator and Benchmark. Among its new features are:
  - Nineteen most frequented evaluation criteria categorized into three groups (region-based, pixel-wise and consistency measures) are available.
  - All benchmark mosaics can be corrupted with an additive noise to test the noise resistance of single classifiers or segmenters.
  - The benchmark was generalized for supervised classifiers evaluation. For all 100 textures used in the benchmark texture mosaics the corresponding separate trainee textures are available.
  - Resulting evaluation tables can be exported in the LaTeX form. Complete results for every mosaic are stored in the database (unrestricted mode).
- **ARCS** A major milestone of Seibersdorf activities in the given period is the successful completion of the first part of the Coin Image Seibersdorf, - CIS-, benchmark. The CIS benchmark is available through the MUSCLE benchmarking Web page. It contains images of 30.000 coins as well as training data for approximately 2500 pattern classes. Additionally, the ground truth for the test data is provided.

3.2 Work launched

- **ARCS and TU VIENNA-PRIP** The first steps for the MUSCLE coin benchmarking competition planed in 2006 have been taken.
- **TU VIENNA-PRIP** The planning for the 2006 benchmarking campaign in collaboration with the ImageCLEF campaign has been started. In addition to the existing tasks, we plan two which are more specific to MUSCLE:
  1. Photographic image retrieval task. In a database of 25000 images, around 25 images will be chosen as query images. An evaluation of the images returned in response to these query images will be done.
  2. Object recognition task. Around 20 objects from the LTUtech database (each represented by around 300 images) will be used for training. About 1000 images of these objects in a natural setting will be collected for the test set. Discussion with members of the PASCAL NoE on collaboration in this task has been started.

- **TU VIENNA-PRIP, CEA, ARMINES-CMM, TCD** In the framework of the E-team on Choosing Features for CBIR and automatic image annotation, we have begun creating the ground truth for an animal recognition task. So far, 30000 of the 60000 images in the Corel database have been manually classified as animal or non-animal. Manual segmentation on 900 of these images into the region corresponding to the animal and the background has been started. This is done using the SAIST tool developed at TU VIENNA-PRIP and available on the WP3 repository.

3.3 Events and Meetings

The *MUSCLE/ImageCLEF Workshop on Image and Video Retrieval Evaluation* took place on the 20th of September 2005. There were 27 participants. The proceedings and presentations are available online here: http://muscle.prip.tuwien.ac.at/ws_proceedings.php

4 Overview activities in WP 4

4.1 Contribution by FORTH

Researchers involved: Panos Trahanias

Design (in collaboration with the MUSCLE Steering Committee) and realisation of new MUSCLE poster.

4.2 Contribution by CNR-ISTI

Researchers involved: Ovidio Salvetti, Patrizia Asirelli, Sara Colantonio, Sergio Di Bona, Maria Grazia Di Bono, Massimo Martinelli, Davide Moroni, Gabriele Pieri

Presentation of WP9 activities at the following joint meetings:

- W3C Italian Office and Semedia Lab (Pisa, October 2005)
- Russian Academy of Science (Pisa, October 2005)

5 Overview activities in WP 5

5.1 Contribution by MTA-SZTAKI

Researchers involved: Tamas Sziranyi, Csaba Benedek, L?szl? Havasi, Levente Kov?cs

Task 2, Sub-task 1: Relative Focus Map Estimation Using Blind Deconvolution

An automatic focus map extraction method is presented using a modification of blind deconvolution for localised blurring function estimation. We use these local blurring functions (so called point spread functions, PSFs) for extraction of focus areas on ordinary images. In this inverse task our goal is not image reconstruction but the estimation of localised PSFs and the relative focus map. Thus, the method is less sensitive to noise and ill-posed deconvolution problems. The focus areas can be estimated without any knowledge about the shooting conditions or the used ptical system. The technique is suitable for main object selection and extraction, tracking in video and in surveillance applications, indexing of image databases.

Publications

- L. Kovacs, T. Sziranyi: Relative Focus Map Estimation Using Blind Deconvolution , Optics Letters, Vol.30, pp. 3021-3023, November 2005

Task 2, Sub-task 3: Spatial relations and geometrical configurations.

Abstract: We introduce an algorithm for matching partially overlapping image-pairs where the object of interest is in motion, even if the motion is discontinuous and in an unstructured environment. We have shown that by using co-motion statistics matching of overlapping views can be done and then the projective geometry can be estimated. Here we will show how to optimize searching for concurrently moving pixels. The robust algorithm we describe here finds point correspondences in two images by using entropybased thresholding and without searching for any structures and without the need for tracking continuous motion. Our method makes it possible to (re)calibrate multicamera systems without human assistance.

Publications

- Z. Szlavik, T. Sziranyi, L. Havasi, C. Benedek, *Optimizing of Searching Co-Motion Point-Pairs for Statistical Camera Calibration* , IEEE International Conference on Image Processing (ICIP), Genoa, Italy, September 11-14, 2005.

5.2 Contribution by ACV

Researchers involved: Herbert Ramoser,

Preparations for the "Workshop on Applications of Computer Vision" held in conjunction with ECCV 2006 in Graz, Austria.

5.3 Contribution by TUVienna-PRIP

Researchers involved: Allan Hanbury,

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

5.4 Contribution by TUVienna-PRIP

Researchers involved: Allan Hanbury,

E-team on "Choosing Features for CBIR and Automated Image Annotation"

5.5 Contribution by TUVienna-PRIP

Researchers involved: Allan Hanbury, Lech Szumilas, Branisalv Micusik

Development of image segmentation algorithms which segment an image based on a sample of the texture to be found. This sample must be specified, for example by the user. The algorithm then attempts to mark all the regions in the image which correspond to the specified texture. 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 is being further developed to be fully automatic. Furthermore, attempts to automatically locate interesting textures in an image are underway.

5.6 Contribution by AUTH

Researchers involved: Constantine Kotropoulos, Ioannis Pitas, Athanasios Papaioannou

Task 4: Text and natural language processing

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

- D. Ververidis and C. Kotropoulos, *Sequential Forward Feature selections with low computational cost,* in Proc. XIII European Signal Processing Conf. Antalya, Turkey, September 2005. (Presented at EUSIPCO 2005).

5.7 Contribution by ARMINES

Researchers involved: Beatriz Marcotegui,

Kick-off meeting of e-team "Choosing Features for CBIR and Automated Image Annotation"

5.8 Contribution by CNR-ISTI

Researchers involved: Ovidio Salvetti, Luigi Bedini, Graziano Bertini, Massimo Magrini, Gabriele Pieri, Emanuele Salerno, Leonello Tarabella, Anna Tonazzini

Low-level feature extraction for visual content description Further extended the already proposed convolutive data model to cases where the convolution kernels are unknown (1). Developed a statistical electrophoresis signal model for basecalling in DNA sequencing. Contributed WP5 SOA (revised version).

1. A. Tonazzini, I. Gerace, *Bayesian MRF-based blind source separation of convolutive mixtures of images* , Proc. EUSIPCO 2005 (Antalya, Turkey, 4-8 September 2005).

Target tracking in video-sequeces The problem of object tracking in sequences of images acquired in visible and near-IR / thermal spectra has been faced. Motion detection and target recognition have been developed using morpho-densitometric and geometric features. To improve robustness and reliability of the approaches a multi-level neural network has been introduced able to establish whether the detected target is or not the right one.

1. Di Bono M.G., Colantonio S., Pieri G. Salvetti O., *Disease evolution monitoring based on multi-source signals and images* , Proc. AITTH 2005 (Advanced Information and Telemedicine Technologies for Health), Minsk, Belarus, 2005.
2. (3) Colantonio S., Benvenuti M, Pieri G., Salvetti O., *Object tracking in a stereo and infrared vision system,* Proc. Int. Ws. Advanced Infrared Technology and Applications?, 7-10 Sept. 2005,Rome.

Audio & speech Part of the activity has been devoted to organize in Pisa the 3rd Intl. Workshop on Computer Music Modeling and Retrival - CMMR-, 26-28 Sett. 2005, and presenting papers dealing with MUSCLE topics, (Tarabella L.: *The pureCMusic (pcM++) framework as open-source music language* Pre-proc. pp. 47-57; Bertini G., Magrini M., Tarabella L.: *An interactive musical exhibit based on infrared sensors* Pre-proc. pp. 16-25).

We have prepared the presentations of a paper on web-site (Bertini G., Magrini M.: *A prototype Lab Box With DSK C6711/13 for Rapid DSP Algorithms Development* sito web TechOnLine.com, section Technical Papers, Texas Instruments Audio and Video/Imaging Series. 19 Oct.05) and at these other conferences too: Bertini G., Gonzales D., Grassi L., Fontana F., Magrini M.,: *Voice Transformation Algorithms with Real Time DSP Rapid Prototyping Tools* , Proceed.on CD 13th EUSIPCO 2005 (European Signal Processing Conf.) Antalya, Turkey, 4-8 sept. 2005, Poster. Bertini G., Gonzales D., Grassi L., Fontana F., Magrini M.,: *Uso di tools per Rapid Prototyping nello sviluppo di algoritmi DSP complessi* Proc on CD of *DSP Application Day* , DICO, Univ. of Milan, Italy, 19 sept. 2005.

We have also planned the proposal of a MUSCLE wp5 task3 e-team, focused on real-time tools for audio/music signals processing/restoration and new systems to perform computer music.

5.9 Contribution by MTA-SZTAKI

Researchers involved: Dmitry Chetverikov, S.Fazekas, L.Hajder

- A comparative study on dynamic texture recognition has been accepted and presented an Texture2005, a wokshop of ICCV 2005 (Beijing). Work on spatiotemporal periodicity of dynamic textures has been started.
- A novel method for 3D motion segmentation of multiple moving objects has been developed and presented the ICCV 2005 workshop on Dynamic Vision (Beijing). A novel method for 3D motion grouping has been developed and a related paper has been submitted to ECCV 2006 (Graz).

Publications Major related publications that acknowledge MUSCLE support:

1. S.Fazekas and D. Chetverikov, Normal versus complete flow in dynamic texture recognition: a comparative study, Texture 2005: 4th international workshop on texture analysis and synthesis (at ICCV 2005), Beijing, October 2005, pp.37-42, ISBN 1-904410-13-8. (Also on ICCV 2005 CD ROM.)
2. L.Hajder and D. Chetverikov, Robust 3D Segmentation of Multiple Moving Objects Under Weak Perspective, ICCV Workshop on Dynamical Vision, Beijing, October 2005, ICCV 2005 CD ROM.

5.10 Contribution by CEA

Researchers involved: Gregory Grefenstette, Svetlana Zinger, Pierre-Alain Moellic, Patrick Hede, Romaric Besancon, Christian Fluhr,

Updated the State-of-the-Art for natural language processing.

During this period, we submitted a workshop on creating Language resources for content-based image retrieval (to be called OntoImage'2006) to the LREC'2006 conference that will be held in Trento, May 24-26, 2006. This workshop has been accepted by the conference organisers. We also worked on extending NLP to Hungarian.

In Image processing, we continued working on implementing new techniques for point of interst indexing for our PiRia platform.

5.11 Contribution by ICCS

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

Spatio-temporal Visual Attention and Ontologies

Study of interactions between low- and high-level information in multimedia ontologies. Progress 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.

5.12 Contribution by ICCS

Researchers involved: Iasonas Kokkinos, P. Maragos

Synergy Between Image Segmentation and Object Recognition

We have developed a statistically motivated approach to the combination of object models of the Morphable/Active Appearance Models class with levelset-based image segmentation techniques , relying on the framework of the Expectation-Maximization (EM) algorithm. The variational formulation of the EM algorithm combines in a seamless manner these two methodologies, while lower-level cues like edge and intensity cues can be easily combined with higher level prior information about the shape and appearance statistics of the object class being segmented. Specifically, we allow shape information to be incorporated in curve evolution without introducing an external statistical shape force, but by directly using the top-down model's predictions about the object's appearance and position. Experimental results on two image categories, namely faces and cars, demonstrate the applicability of our approach to problems of practical interest.

5.13 Contribution by ICCS

Researchers involved: Anastassia Sofou, P. Maragos

Salient feature extraction for seeded region segmentation and image analysis

By the term salient image features we refer to those features that have properties which make them suitable for detecting/extracting regions of interest significantly different from their surroundings. The aforementioned features can be localized in spatial as well as in frequency domain and usually demonstrate tolerance to noise. A large number of features can be extracted using efficient algorithms considering the properties described above.

Following the traditional linear approach, the extraction of these features can be divided in the following stages: (1) Gaussian scale-space extrema detection where potential points of interest are identified, (2) keypoint localization, where keypoints are selected based on measures of their stability and (3) orientation assignment. The keypoints extracted correspond to blobs, edges, corner, junctions etc. With additional feature filtering and careful selection, a set of strong keypoints can be determined and then used as an initial marker set in a seeded segmentation scheme such as watershed- like segmentation, region growing segmentation, or in other image analysis tasks such as object detection, recognition, classification, matching etc.

From the nonlinear point of view salient feature extraction can be accomplished by employing ideas from the area of mathematical morphology. Morphological pyramids and scale spaces can be used to determine features invariant to scale changes. Other features can be obtained via a filtering procedure employing basic morphological operators such as open-closings and more sophisticated ones such as connected filters, alternating sequential filters, levelings and adaptive morphological operators. The aforementioned filters operate on the image structure with respect to a criterion or a combination of criteria such as contrast, area, luminance, texture, color, adaptive neighborhoods etc., in order to emphasize regions /features of interest and remove insignificant information.

Other approaches that are considered in the framework of efficient salient image features extraction include AM-FM image modeling and decomposition in order to extract texture-relevant features such as texture energy, orientation vectors, etc. Additionally, using ideas such as U + V image decomposition, images can be decomposed into two different signals one corresponding to texture/noise and the other corresponding to contrast. Each of these signals can be then treated separately in order to locate and extract appropriate features that can be used in other image analysis tasks.

5.14 Contribution by TUVienna-IFS

Researchers involved: Andreas Rauber, Thomas Lidy, Rawia Awadallah, Robert Neumayer, Georg P?lzlbauer

Task 1 Orientation and Roadmap Contributed to the MUSCLE State-of-the-art (SoA) report, specifically for audio feature extraction and classification as well as for cross language retrieval.

Task 3 Audio and speech processing

Sub-task 3.5. Events detection, segmentation and classification for audio streams

- Adaption of the music retrieval framework used to participate in the ISMIR 2004 audio description contest.
- Participated in the 2nd Annual Music Information Retrieval Evaluation eXchange (MIREX 2005), achieved 75.27 % accuracy on audio genre classification (5th rank, 3rd by group).
- Developed representation and interaction model based on clusters of audio data.

Sub-task 3.6. High-level feature extraction for audio

Developed and expanded modules for audio feature extraction, specifically analysing the impact of psycho-acoustic transformations in the Rhythm Patterns features.

Developed new feature sets for music content description: Statistical Spectrum Descriptors and Rhythm Histograms.

Task 4 Text and natural language processing

Developed a question-answering system for english and specifically arabic questions using the web as a knowledge base evaluated on the TREC QA-track and english and arabic versions of the show *Who wants to be a millionaire* ?.

Investigated the extraction of semantic labels from free-form text documents.

5.15 Contribution by UniS

Researchers involved: Bill Christmas, Fei Yan, Ilias Kolonias, Myung Roh, Cyrille Hory

- Submitted paper to ECCV06 on gesture detection and identification.
- Submitted paper to CVPR06 on ball tracking in sports video.
- Submitted paper to ICASSP06 on audio cues for automatic tennis annotation.

5.16 Contribution by GET

Researchers involved: Beatrice Pesquet-Popescu, Christophe Tillier, Gregoire Pau

We continued the work on adaptive wavelet decompositions for images. Three papers have been presented in the IEEE ICIP'05 and EURASIP EUSIPCO'05 conferences based on this work. New seminorm combinations have been tested. An interface was created to allow JPEG2000 encoding of such decompositions.

5.17 Contribution by UniS

Researchers involved: Bill Christmas, Fei Yan, Ilias Kolonias, Myung Roh, Cyrille Hory

- Submitted paper to ECCV06 on gesture detection and identification.
- Submitted paper to CVPR06 on ball tracking in sports video.
- Submitted paper to ICASSP06 on audio cues for automatic tennis annotation.

6 Overview activities in WP 6

6.1 Contribution by AUTH

Researchers involved: Constantine Kotropoulos, Ioannis Pitas, Margarita Kotti, Emmanouil Mpenetos, Marios Kyperountas

Work on speaker change detection and speaker clustering continues. A paper describing a multiple pass algorithm for speaker change detection is submitted to ISCAS 2006.
A novel audiovisual scene change detection algorithm has been developed that is based on a set of eigen-audioframes that define an audio signal subspace associated to the audio background changes.

Publications

1. M. Kotti, E. Benetos, and C. Kotropoulos, *Automatic speaker change detection with the Bayesian Information Criterion using MPEG-7 features and a fusion scheme,* submitted to 2006 IEEE Symp. Circuits and Systems, Rhodes, Greece.
2. M. Kyperountas, C. Kotropoulos, and I. Pitas, *Enhanced eigen-audioframes for audiovisual scene change detection,* submitted to IEEE Trans. Multimedia.

6.2 Contribution by INRIA-Texmex

Researchers involved: Patrick Gros, Manolis Delakis, Pascale Sebillot, Stephane Huet

During the past two months we continued experimentation on the audiovisual fusion with Segment Models. We obtained a clear performance advantage over previous experimental setups by using a left-to-right HMM topology with 20 hidden states for modeling the audio content of a scene on top of the low-level audio frames (cepstral coefficients). This work is summarized in citetexmex05b. Furthermore, we are thinking of the use of stream state-dependent weights between the modalities, in a similar fashion as in multiband or audiovisual speech recognition. First results obtained by a simple grid search demonstrated a visible performance gaining in some setups. Algorithms like Minimum Classification Error could be used to provide a more consistent solution.

In addition to the fusion of video and audio, we also started experimentation in the fusion of them with text. We manually annotated the score indications superimposed in the video. This information is useful as a reliable indication that a point has marked. Indeed, using it as a binary descriptor that a score indication has appeared in the corresponding shot, we obtained +1.39% better results in the HMM framework. Secondly, these indications will give a reliable estimation on the number of scores totally marked in the tennis match. This number generally disagrees with that obtained from the Viterbi decoding. We can thus further improve the results if we prune all the paths which give a number of points different with that of the score indications. A solution to this problem could be given with an N-Best-based decoding algorithm which keeps track of the score according to the succession of states and prunes as soon as possible the non-allowable paths. This approach yields a +2.27% performance improvement.

We also began a work about the coupling of Natrual language processing techniques with speech transcription techniques. The aim of this work is to study whether using NLP techniques (tagging, topic detection) could improve speech detection by allowing a better choice between the most probable transcriptions. On the other way round, it is important to check whether NLP techniques are able to handle transcribed text which is a degraded text (no punctuation, many mistakes.)

Publications

- M. Delakis and G. Gravier and P. Gros, *Audiovisual Fusion with Segment Models for Video Structure Analysis,* (To appear in) Proceedings of the 2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies (EWIMT'05)

6.3 Contribution by ICCS

Researchers involved: George Papandreou, A. Katsamanis, V. Pitsikalis, P. Maragos

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.

In collaboration with TSI-TUC, 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. These visual features are combined with auditory features and enhance the performance of speech recognition systems 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), while the Vid-Timit audio-visual database (obtained from the University of Adelaide) might be used in the future for additional experiments.

6.4 Contribution by TUVienna-IFS

Researchers involved: Andreas Rauber, Robert Neumayer, Thomas Lidy

Task 3 Cross-Modal Integration for Multimedia Analysis and Recognition

Subtask 3.3: Integrated Multimedia Content Analysis

Commenced work on automatic semantic concept detection in textual documents to be integrated with audio data, analyzing song lyrics. The goal is to obtain additional genre information from textual data, to be combined with features etracted from song signals.

7 Overview activities in WP 7

7.1 Contribution by CNR-ISTI

Researchers involved: Ovidio Salvetti, Ercan Kuruoglu

A new filtering technique has been developed based on sequential Monte Carlo for eliminating speckle noise in SAR images. A new technique based on revesible jump MCMC has been developed for the estimation of components in a mixture of impulsive signals.

Publications

1. Kuruoglu E.E., Baccigalupi C. - Special issue on applications of signal processing in astrophysics and cosmology. Editorial activity: Special Issue on Applications of Signal Processing in Astrophysics and Cosmology . In 'EURASIP Journal on Applied Signal Processing, Vol. 2005 n. 15 (2005), pp.2397-2399. Hindawi Publishing Corporation, 2005.

2. Bedini L., Herranz D., Salerno E., Baccigalupi C., Kuruoglu E.E., Tonazzini A. - Separation of correlated astrophysical sources using multiple-lag data covariance matrices. In: EURASIP Journal on Applied Signal Processing, Vol. 2005 n. 15 (2005), pp. 2400-2412. Hindawi Publishing Corporation, 2005.

3. A. Achim, E. E. Kuruoglu and J. Zerubia, "SAR Image Filtering Based on the Heavy-tailed Rayleigh Model," EUSIPCO 2005, September 2005, Antalya, Turkey.

4. D. Gencaga, E. E. Kuruoglu and A. Ertuzun, "Estimation of Time-varying Autoregressive symmetric alpha-stable Processes using particle filters", EUSIPCO 2005, September 2005, Antalya, Turkey. 5. D. Gencaga, E. E. Kuruoglu and A. Ertuzun, "SAR Image Enhancement Using Particle Filters", ESA-EUSC Workshop, October 5-7 2005, Frascati, Italy.

7.2 Contribution by TCD

Researchers involved: Simon Wilson,

Final version of Deliverables 7.3 and 7.4 completed. Webpage and call for papers for MUSCLE workshop on computation intensive methods in computer vision (to be held at ECCV06) written and published. WP7 focus meeting organised in conjunction with WP5. Organisation of ECCV workshop on computational methods. E-team on Choosing Features for CBIR and Automated Image Annotation (part WP7). E-team on Choosing Features for CBIR and Automated Image Annotation (part WP7), set up meeting, September 12th 2005, Ecole des Mines, Paris.

8 Overview activities in WP 8

8.1 Contribution by TUG

Researchers involved: Horst Bischof, other TUG members

The research on learning has focused on three issues:

1. The conservative learning framework was adapted to an online version. This was achieved by employing an on-line AdaBoost classifier developed at our institute. The framework has been used to develop a person detector and a car detector.

2. Work on incremental LDA started. We have designed an algorithm based on the combination of reconstructive and discriminative information. Currently this algorithm is analyzed theoretically and in several applications.

3. Work on incremental Adaboost is continuing. New applications for background modelling and tracking are currently developed.

Publications

- G. Langs, P. Peloscheck, R. Donner, and H. Bischof. A clique of active appearance models by minimum description length. In W.F. Clocksin, A.W. Fitzgibbon, and P.H.S. Torr, editors, Proc. of British Machine Vision Conference (BMVC). BMVA, 2005.
- P. Roth, H. Grabner, D. Skocaj, H. Bischof, and A. Leonardis. On- line conservative learning for person detection. In Proc. 2nd Joint IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance. IEEE Computer Society, 2005.
- P. Roth, H. Grabner, D. Skocaj, H. Bischof, and A. Leonardis. Conservative visual learning for object detetection with minimal hand labeling e ort. In W.G. Kropatsch, R. Sablatning, and A. Hanburry, editors, Pattern Recognition 27th DAGM Symposium, volume LNCS 3663, pages 293 300. Spinger, 2005.
- R. Leitner and H. Bischof. Recognition of 3d objects by learning from correspondences in a sequence of unlabeled training images. In W.G. Kropatsch, R. Sablatning, and A. Hanburry, editors, Pattern Recognition 27th DAGM Symposium, volume LNCS 3663, pages 369 376. Spinger, 2005.

8.2 Contribution by AUTH

Researchers involved: Constantine Kotropoulos, Vassiliki Moschou

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. Such patterns form N clusters on the (N-1)-dimensional simplex. 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.

Publications

1. C. Kotropoulos and V. Moschou, *Self-organizing maps for reducing the number of clusters by one on simplex subspaces,* submitted to 2006 IEEE Int. Conf. Acoustics, Speech, and Signal Processing.

8.3 Contribution by ARMINES

Researchers involved: Beatriz Marcotegui, Francis Bach

Collaboration on kernel methods for speaker verification: design of new theoretical framework for the design of sequence kernels.

8.4 Contribution by CNR-ISTI

Researchers involved: Ovidio Salvetti, Sara Colantonio, Davide Moroni

Image Categorisation (IC) has been studied as decomposition of specialised tasks organised in a Multilevel Artificial Neural Network (MANN) model. A MANN has been defined as a modular and hierarchical combination of sub-networks having different topologies and typologies.

We defined a multilevel system architecture capable to process multi-source data according to a coarse-to-fine paradigm. Applications to neuro-signals and image categorization were considered as cases study.

We presented also a methodology, based on neural paradigms, suitable to analyse periodically deforming structures and recognize their state. Preliminary tests have been performed on the heart dynamics study.

The work done in the reference period has led to the preparation of papers accepted at AITTH 2005 and IEEE ISSPIT 2005 International Conferences.

8.5 Contribution by TUVienna-IFS

Researchers involved: Andreas Rauber, Georg P?lzlbauer, Thomas Lidy, Michael Dittenbach, Rudolf Mayer, Dieter Merkl

Task 2 - Supervised learning Evaluated the application of supervised machine learning techniques, particularly SVM.
Performed a series of experiments for classification of audio data into different semantic categories, specifically genre, artist and rhythm categories

Task 3 - Unsupervised learning Developed Matlab prototype modules for visualization of cluster structures on high-dimensional maps, with a set of different visualization approaches aiming at visualizing concurrently the impact of different sub-spaces of a high-dimensional feature space. Utilized Self Organizing Maps (SOMs) for unsupervised organisation of multimedia content, creating a topology preserving mapping of audio data by perceived sound similarity.

Publications

- Georg P?lzlbauer, Michael Dittenbach, Andreas Rauber. A visualization technique for Self-Organizing Maps with vector fields to obtain the cluster structure at desired levels of detail. Proceedings of the International Joint Conference on Neural Networks (IJCNN 2005), pp. 1558-1563, July 31-August 4, 2005, Montreal, Canada.
- Michael Dittenbach, Andreas Rauber, and Georg P?lzlbauer. Investigation of Alternative Strategies and Quality Measures for Controlling the Growth Process of the Growing Hierarchical Self-Organizing Map. Proceedings of the International Joint Conference on Neural Networks (IJCNN 2005), July 31-August 4, 2005, Montreal, Canada.
- Georg P?lzlbauer, Michael Dittenbach, Andreas Rauber. Gradient visualization of grouped component planes on the SOM lattice. Proceedings of the 5th Workshop On Self-Organizing Maps Paris (WSOM 2005), September 5-8 2005, Paris, France.
- Rudolf Mayer, Dieter Merkl, Andreas Rauber. Mnemonic SOMs: Recognizable Shapes for Self-Organizing Maps. Proceedings of the 5th Workshop On Self-Organizing Maps Paris (WSOM 2005), September 5-8 2005, Paris, France.

8.6 Contribution by TCD

Researchers involved: Padraig Cunningham, UTIA, CNR, UU,SZTAKI, TU Graz, UvA, CWI, UCL, INRIA IMEDIA, INRIA VISTA

Six WP8 deliverables were produced in September 2005. These were:

- ML Software Repository: a web page linking to ML software that has been made available by Muscle partners.
- 5 Technical Deliverables: The deliverables were edited by Rozenn Dahyot, Derek Greene and Padraig Cunningham from TCD.

A kick-off meeting was held in TCD on 28th and 29th of September to establish e-teams related to WP8 activities. In all, 5 e-teams were discussed.

It was agreed that Padraig Cunningham (TCD) will coordinate an e-team on Dimension Reduction.

8.7 Contribution by UU

Researchers involved: Niall Rooney, Petra Perner (IBAI)

UU and IBAI have written an submitted a joint research fellowship proposal titled "A conversational case-based reasoning approach for the multimodal retrieval of documents consisting of image and text" for review by the Muscle consortium.

8.8 Contribution by FT

Researchers involved: Christophe Garcia, Sid-Ahmed Berrani

Our research activities are related to task 2 (Supervised Learning) and task 5 (Applications of state-of-the-art techniques to current problems of multimedia understanding).

We have been carrying on our work on supervised ML techniques for object detection and recognition in images and for interest point localization in face images.

Convolutional Neural Networks

We have been carrying on our work on the development of Convolutional Neural Networks for object detection and recognition in images:

- Learning rate adaptation for learning speed acceleration and network well-conditioning. Different Hessian-based techniques have been evaluated on the problem of face /nonface classification problem, including on-line computation of the principal eigenvalue of the neural weight Hessian matrix.
- Started study and development of methods for automatic building of convolutional neural network architecture, including growing and pruning techniques.

Facial Feature Detection

We have carrying on our work on facial feature detection (eyes, nose, mouth). The current neural-based facial feature detection scheme was designed to precisely locate 4 facial features in faces of variable size and appearance, rotated up to 30 degrees in image plane and turned up to 60 degrees, in complex real world images. Based on a specific architecture of convolutional and hetero-associative neural layers, the proposed system automatically synthesized simple problem-specific feature extractors and classifiers from a training set of faces with annotated facial features, without making any assumptions or using any hand-made design concerning the features to extract or the areas of the face pattern to analyze. Moreover, global constraints encoding the face model were automatically learnt and used implicitly and directly in the detection process. After training, the facial feature detection procedure acted like a pipeline of simple convolution and subsampling modules that treated the raw input face image as a whole and built facial feature maps where facial feature positions were easily retrieved by a simple global maxima search.

We have developed a new technique (published in [1], Best Young Researcher Paper Award) for robustly and automatically detect a finer set of facial features, based on a hierarchical approach that comprises three successive stages: face detection, coarse feature detection (the 4. facial features: eye centres, nose tip and mouth center) and fine feature detection (eye centers and corners, mouth corners). At each stage, the detection results of the preceding stage are used to focus the search of facial features on a restricted image region. In the coarse and fine detection stages, several specialised feature detectors based on the same specific neural architecture are applied. This architecture consists of several heterogeneous neural layers, automatically synthesising simple problem-specific feature extractors and classifiers from a training set with annotated facial features. After training, each feature detector acts like a pipeline of simple filters that treats the raw input image as a whole and builds global facial feature maps, where facial feature positions can easily be retrieved by a simple search for global maxima. We experimentally show that our method is very precise and robust to lighting and pose variations.

Publications

1. Everingham M., Zisserman A., Williams C. K. I., Van Gool L., Moray A., Bishop C. M., Chapelle O., Dalal N., Deselaers T., Dork? G., Duffner S., Eichhorn J., Farquhar J. D. R., Fritz M., Garcia C., Griffiths T., Jurie F., Keysers T., Koskela M., Laaksonen J., Larlus D., Leibe B., Meng H., Ney H., Schiele B., Schmid C., Seemann E., Shawe-Taylor J., Storkey A., Szedmak S., Triggs B., Ulusoy I., Viitaniemi V., and Zhang J., *The 2005 PASCAL Visual Object Classes Challenge* , Selected Proceedings of the first PASCAL Challenges Workshop, (eds) F. d Alche-Buc, I. Dagan, J. Quinonero, LNAI, Springer, 2006.

2. Duffner S., Garcia C., *A Hierarchical Multi-Stage Approach to Precise Facial Feature Detection* , Proceedings of Compression et Representation des Signaux Audiovisuels (CORESA 05), Rennes, France, November 2005. Best Young Researcher Paper Award

8.9 Contribution by CWI

Researchers involved: Mark Huiskes,

We have implemented a retrieval engine based on aspect-based relevance learning principles as described in Huiskes (2005). The engine is linked up to a new interface for relevance feedback that is particularly suited for retrieval applications where example selection may be to an important extent be based on partial relevance. Additionally we are exploring various strategies for active relevance learning, amongst others leading to new approaches for dealing with negative examples.

Publication

1. Huiskes (2005). Aspect-based relevance learning for image retrieval. International Conference on Video and Image Retrieval (CIVR'05), Singapore, Lecture Notes in Computer Science 3568, W.-K. Leow et at. (Eds), pp. 639-649.

8.10 Contribution by ENSEA

Researchers involved: Matthieu Cord, Guillermo Camara Chavez, Philippe Gosselin

- Video processing: learning for cut detection;
- Active learning: introduction of new utility criteria dedicated for information retrieval different from the ones optimized for classification process. Weakly supervised learning to optimize Gram matrices for image database feature representation.

Publications/dissemination:

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

8.11 Contribution by GET

Researchers involved: Beatrice Pesquet-Popescu, Marine Campdel

GET continued the activity of satellite image (non-supervised) classification based on adaptive wavelet transforms. A larger image database was build (3600 small images extracted from much larger satellite images) and different feature selection and classification algorithms were tested on it. The EUSIPCO paper was presented in Antalya. It was ranked among the "top papers" at this conference and we received an invitation from Bjorn Ottersten, the Editor-in-Chief of the EURASIP Signal Processing journal to submit a long version to this journal.

9 Overview activities in WP 9

9.1 Contribution by CNR-ISTI

Researchers involved: Ovidio Salvetti, Patrizia Asirelli, Maria Grazia Di Bono, Massimo Martinelli

Activities

Overview of the work completed in the reporting period

In the reported period, a research activity has been performed in the frame of a collaboration among ISTI-CNR and SEMEDIA Lab, DEIT of the Polytechnic University of Marche. In particular, an original approach towards multimedia metadata handling and a solution for massive metadata processing, based on distributed computing, has been approached. As well, a tool kit that glues low level metadata with higher-level representation capabilities of RDF has been also approached. Actions have been focused on the study, design and development of MetaMedia, a computational and storage infrastructure that is being developed specifically for the need of scalable, high performance centralized or semi centralized multimedia metadata services. It would provide the necessary processing power and scalability to keep and maintain (add, compare, remove duplicate) a large collection of MPEG-7 low level descriptions and to provide efficiency in query execution thanks to the intelligent scheduling and routing of queries to the most appropriate servers. Development is in progress at this time, and is carried out as Open Source in public development repositories. Final goal will be to make available to the NoE partners a tool for implementing real interoperability.

Overview of the work launched

- Collaboration activity with theW3C consortium / SEMEDIA Lab, DEIT Univ. Polit. Marche (MM metadata infrastructure)
- Experimentation and interaction activities in MPEG environment (automatic metadata generation)
- Workshop preparation on *Data Mining in Analysis of Microscopic Images* for the Industrial Conference on Data Mining (ICDM http://www.data-mining-forum.de) that will take place in Leipzig/Germany, July 14-15, 2006
- Preparation of E-Team on "Integration of structural and semantic models for multimedia metadata management"
- Preparation of Fellowship Programme, titled "Multimedia metadata: bridging the gap from low-level media specific features to high-level domain-specific semantic terms"

Major results (achievements)

In progress development of a project (MetaMedia) oriented to realise: tools for the MPEG-7 MM metadata automatic generation from image, audio and video; tools and interfaces for remote MM metadata retrieval; tools for remote querying of MM metadata distributed database.

Events

- MM Metadata and Semantic Web technologies, W3C/SEMEDIA working meeting
- Audio Conferences

Publications

1. P. Asirelli, M.G. Di Bono, M. Martinelli, O. Salvetti, M. Catasta, C. Morbidoni, F. Piazza, G. Tummarello, O. Signore, *Toward a Scalable Multimedia Metadata Infrastructure using Distributed Computing and Semantic Web Technologies* , accepted to the 2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies, 30 November - 1 December 2005, London.

9.2 Contribution by GET

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

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.

10 Overview activities in WP 10

11 Overview activities in WP 11

11.1 Research Activities

In the framework of the E-team "3-D Texture Analysis and Detection" Beatrice Pesquet-Popescu and Maria Trocan (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 submitted to the IEEE ICASSP 2006 conference

SzTAKI has been working on real-time video analysis. Sztaki group has a new model for foreground-background-shadow separation [1]. Their method extracts the faithful silhouettes of foreground objects even if they have partly background like colors and shadows are observable on the image. It does not need any a priori information about the shapes of the objects, it assumes only they are not point-wise. The method exploits temporal statistics to characterize the background and shadow, and spatial statistics for the foreground. A Markov Random Field model is used to enhance the accuracy of the separation. They validated their method on outdoor and indoor video sequences captured by the surveillance system of the university campus, and they also tested it on well-known benchmark videos. BILKENT now has real-time working versions of falling person detection [2] and fire and smoke detection methods [3].

BILKENT Bilvideo database group developed an interface that uses natural language interaction to input queries from the user. The interface is now available through the Web. The interface for BilVideo to formulate queries in natural language and examples of such queries can be accessed from http://pcvideo.cs.bilkent.edu.tr/querying.html [4-6]. Currently, there are also plans for implementing a hand based gesture interaction system for BilVideo. The idea is to use hand-gestures as a complement to the mouse interaction. The most promising area that we can utilize this kind of interaction is to input spatio-temporal relations between objects of interest in video, especially 3D relations between the objects in video frames. Collaboration between SzTAKI, BILKENT and FORTH on this subject is planned. Another aspect of the BilVideo research is content-based access to surveillance videos. Siginificant research work dealing with automated access to visual surveillance has appeared in the literature. However, the event models and the content-based querying and retrieval components have significant gaps remaining unfilled. BilVideo group proposes a database model for querying surveillance videos by integrating semantic and low level features [6].

11.2 Collaboration

Ugur Toreyin, a Ph.D. student at Bilkent University visited ENST in Paris in September 2005. A joint paper was submitted to IEEE Conference, ICASSP to be held in France in 2006. Members of WP-11 will participate in E-team meetings in Paris. The first meeting will be held in INRIA on December 1 and 2. The second meeting will be held in ENST on December 3 and organized by B. Pesquet-Popescu and A. Enis Cetin. Topics covered in E-team meetings will include face detection in video, human body detection in video, multimedia databases, 3-D texture detection and applications of 3-D wavelet transforms.

11.3 Publications

- Benedek C, Sziranyi T: "Markovian Framework for Foreground-Background-Shadow Segmentation of Real World Video Scenes", accepted for publication in ACCV 2006, Hyderabad, India, to be published in Lecture Notes in Computer Science.

- B. Ugur Toreyin, Yigithan Dedeoglu, A. Enis Cetin, Flame Detection in Video Using Hidden Markov Models, in Proc. of IEEE International Conf. Image Processing, Sept. 2005, Genova, Italy.

- B. Ugur Toreyin, Yigithan Dedeoglu, A. Enis Cetin, HMM Based Falling Person Detection Using Both Audio and Video", presented in HCI-2005: IEEE Intl. Workshop on Human-Computer Interaction. The article was published in Lecture Notes in Computer Science (LNCS), Oct. 2005.

- Mehmet Emin Dnderler, Ediz Saykol, Umut Arslan, zgr Ulusoy, Ugur Gdkbay, "BilVideo: Design and Implementation of a Video Database Management System", Multimedia Tools and Applications, Vol. 27, pp. 79-104, September 2005. 5-Tarkan Sevilmis, Automatic Detection of Salient Objects for a Video Database System, M.S. Thesis, Department of Computer Engineering, Bilkent University, Ankara Turkey, November 2005.

- Ediz Saykol, Ugur Gdkbay, zgr Ulusoy, "A Database Model for Querying Visual Surveillance by Integrating Semantic and Low-Level Features", in Lecture Notes in Computer Science (LNCS) , (Proc. of 11th International Workshop on Multimedia Information Systems (MIS'05)), Vol. 3665, pp. 163-176, Edited by K. Selcuk Candan and Augusto Celentano, Sorrento, Italy, September 2005.