MUSCLE

Network of Excellence

Multimedia Understanding through Semantics, Computation and Learning

Project no. FP6-507752

Bi-monthly progress report # 16: Sept-Oct 2006

Due date of deliverable: 15.11.2006
Actual submission date: 15.04.2007

Start date of project: 1 March 2004   Duration: 48 months

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

Draft 1.0

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

Keyword List:

Contents

1 Overview activities in WP1

1.1 General scientific and administrative coordination

- The Scientific Coordination was transferred over from Eric Pauwels, CWI, to Nozha Boujemaa, INRIA.
- Administrative and financial coordination of the network by Remi Ronchaud, ERCIM. Patricia Ho-Hune, ERCIM, started to work for MUSCLE in October to reinforce the management team.
- Organisation of regular audio-conferences and organisation of the October Steering Committee meeting.
- A Face-to-face Steering Committee meeting was held on 2, 3 October 2006 in Vienna where MUSCLE integration activities (E-Teams, showcase, sponsorship), foreseen actions, dissemination and budget were discussed.
- Preparation for the organisation of the MUSCLE Plenary meeting.
- E-Teams re-organisation
- Launch of Showcase project call
- Discussions between ERCIM and CWI to prepare migration of MUSCLE website content from CWI to ERCIM
- Reimbursement of MUSCLE integration expenses (mobility support grant)

2 Overview activities in WP2

2.1 Contribution by partner 06 TU VIENNA-PRIP

Activities

Two MUSCLE-sponsored workshops related to evaluation have been organised. The CIS coin recognition competition result workshop took place on the 11th of September in Berlin, Germany (before the DAGM conference). The prize for the best performing program was won by Marco Reisert from the Chair of Pattern Recognition at the University of Freiburg (a MUSCLE member). He basically solved this rather difficult problem, obtaining a classification rate of 97.2% on the unseen dataset. All other coins were classified as unknown, meaning that he made no false classifications. Articles on the results of this competition have been submitted to ERCIM news and the IAPR newsletter. The outcome is also on the MUSCLE News webpage. The proceedings, ground truth and detailed results are online (http://muscle.prip.tuwien.ac.at/coin_proceedings_2006.php). A journal paper (Journal of Machine Vision Applications) is under preparation. The MUSCLE/ImageCLEF 2006 workshop took place on the 19th of September in Alicante, Spain (before the CLEF workshop). The invited speaker was Mark Everingham, co-organiser of the PASCAL Visual Object Challenge. We had representatives from three major evaluation campaigns present: PASCAL VOC, ImagEVAL and ImageCLEF. One of the main outcomes of the workshop is the possibility of greater cooperation between ImageCLEF and the PASCAL VOC. This will be discussed over the following months as we prepare evaluation campaigns for 2007. The proceedings are online: http://muscle.prip.tuwien.ac.at/ws_proceedings_2006.php

Achievements

A showcase proposal for a live evaluation event at the CIVR 2007 was written. It was later accepted to be financed as a MUSCLE showcase.

Problems
The full update of the WP2 website has been delayed so as to adapt them better to the changes in the MUSCLE main website.

Events
MUSCLE CIS Coin Classification Workshop, with following attendees: Michael Nölle, Allan Hanbury, Laurens van der Maaten, Marco Reisert

MUSCLE ImageCLEF workshop, with following attendees: Allan Hanbury, Christian Fluhr, William Hersh, Mark Everingham, Henning Müller, Lokesh Setia, Thomas Deselaers, Levente Kovacs, Nicolas Maillot, Philipp Daumke, Jeff Jensen, Mashashi Inoue, Jayashree Kalpathy-Cramer

Publications
P. Clough, M. Grubinger, T. Deselaers, A. Hanbury and H. Müller. Overview of the ImageCLEF 2006 photographic retrieval and object annotation tasks. In CLEF working notes, Alicante, Spain, September 2006.

2.2   Contribution by partner 09 CNR-ISTI

Activities
The fellowship on "Multimedia metadata: bridging the gap from low-level media specific features to high-level domain-specific semantic terms" started. Suzanne Little arrived in Italy and started working on multimedia metadata management, and on the integration of machine learning and ontologies. A presentation was given to the members at CNR-ISTI describing her previous work and outlining the fellowship. Initial, background reading and research was conducted including a presentation by Sara Colantonio on previous applications of machine-learning techniques at CNR-ISTI. A paper on the 4M architecture was prepared and submitted to the W3C Semantic Web Applications and Perspectives Workshop (SWAP). Suzanne Little was also involved in developing the proposal for the MUSCLE E-Team: Integration of structural and semantic models for multimedia metadata management.

Achievements
Tools for MPEG7 features extraction from audio, images and video

Events
Participation in the 11th World Congress on Internet in Medicine (Mednet 2006), October 14-19, 2006, Toronto (Canada), with the paper "Collaborative Virtual Working Environments for Medical Expertise Sharing".

Publications
S. Di Bona, P. Asirelli, O. Salvetti, M. Martinelli, D. Guerri "Collaborative Virtual Working Environments for Medical Expertise Sharing" in Proc. of the 11th World Congress on Internet in Medicine (Mednet 2006), October 14-19, 2006, Toronto (Canada)

2.3   Contribution by partner 35 UniS

Activities
The C++ library "RAVL" - Recognition and Vision Library - is available under the LGPL licence. The library includes basic container classes, as well as image/video processing,

pattern recognition, maths and other classes and functions. More information at http://www.ee.surrey.ac.uk/CVSSP/Ravl

2.4 Contribution by partner 41 UPMC

Publications
New criteria for evaluating image segmentation results S. Philipp-Foliguet, L. Guigues ICASSP 2006

2.5 Contribution by partner 05 BILKENT

Activities
We contributed some videos of fire and smoke into the MUSCLE Web site. We also contributed the results of our research on human activity detection; specifically some videos of falling person detection.

3 Overview activities in WP3

3.1 Contribution by partner 14 UTIA

Activities
The highest fidelity representations of realistic real-world materials currently used comprise Bidirectional Texture Functions (BTF). The huge size of such measurements, typically in the form of thousands of images covering all possible combinations of illumination and viewing angles, has prohibited their practical exploitation and obviously some compression and modelling method of these enormous BTF data spaces is inevitable. We developed an approach which combines BTF spatial clustering with cluster index modelling by means of an efficient Markov random field model. This method allows generating seamless cluster index of arbitrary size to cover large virtual 3D objects surfaces. BTF data compression using this method is about $1:100$ and their synthesis is very fast. We proposed in cooperation with SZTAKI (D. Chetverikov) a novel hybrid method for colour dynamic textures (DT) modelling. The method is based on eigen-analysis of DT images and subsequent preprocessing and modelling of temporal interpolation eigencoefficients using a causal auto-regressive model. The proposed method shows good performance for most of the tested DTs, which depends mainly on the properties of the original sequence. Moreover, this method compresses significantly the original data and enables extremely fast synthesis of artificial sequence, which can be easily performed by means of contemporary graphics hardware. A frequently used strategy for road sign classification is based on the normalized cross-correlation similarity to class prototypes followed by the nearest neighbor classifier. Because of the global nature of the cross-correlation similarity, this method suffers from presence of uninformative pixels (caused e.g. by occlusions), and is computationally demanding. A novel concept of a trainable similarity measure is introduced which alleviates these shortcomings. The similarity is based on individual matches in a set of local image regions. The set of regions, relevant for a particular similarity assessment, is refined by the training process. It is illustrated on a set of experiments with road sign classification problems that the trainable similarity yields high-performance data representations and classifiers. Apart from a multi-class classification accuracy, also non-sign rejection capability, and computational demands in execution are discussed. It appears that the trainable similarity representation alleviates some difficulties of other algorithms, currently used in road sign classification. A trainable similarity measure, which is specifficaly designed for the supervised classification of images was proposed. This measure is based on local matches in a set of regions within an image

which incrases its robustness. A novel color texture unsupervised segmentation algorithm was developed in cooperation with INRIA Ariana (G. Scarpa) which processes independently the spectral and spatial information. The algorithm is composed of two parts a color-based clustering (CBC) followed by a spatial-based clustering (SBC). The SBC, as well as the subsequent growing algorithm, make use of a characterization of the regions based on shape and context. Another unsupervised multispectral texture segmentation method with unknown number of classes was developed. Multispectral texture mosaics are locally represented by four causal multispectral random field models recursively evaluated for each pixel. The segmentation algorithm is based on the underlying Gaussian mixture model and starts with an over segmented initial estimation which is adaptively modified until the optimal number of homogeneous texture segments is reached. Two fast illumination invariant image retrieval methods for scenes comprising textured objects with variable illumination were published. Both methods are based on texture gradient modelled by efficient set of random field models. We developed the illumination insensitive measures for textured images representation and compared them favorably with steerable pyramid and Gabor features in the illumination invariant BTF texture recognition. Assuming local and shift-invariant texture properties we described the statistical dependencies between pixels by a joint probability density of gray-levels within a suitably chosen observation window. We estimate the unknown multivariate density in the form of a Gaussian mixture of product components from data obtained by shifting the observation window. By considering a subspace approach based on a structural mixture model we can increase the size of the observation window while keeping the computational complexity in reasonable bounds.

Events
Results presented on 18th ICPR, Hong Kong and IWICPAS Xian.

Publications
Filip, Jiri Haindl, Michal: BTF Modelling Using BRDF Texels, Lecture Notes in Computer Science 4153, Springer.
Filip, J. - Haindl, M. - Chetverikov, D. Fast Synthesis of Dynamic Colour Textures, 18th IAPR International Conference on Pattern Recognition, Hong Kong, 2006, ISBN 0-7695-2521-0, ISSN 1051-4651, vol. IV, pp. 25-28, IEEE Press.
Filip, J. - Haindl, M. - Chetverikov, D. Fast Synthesis of Dynamic Colour Textures, ERCIM News, 2006, no. 66, pp. 53-54, ISSN 0926-4981.
Paclík, P; Novovicova, Jana ; Duin, R.: Building Road-Sign Classifiers Using a Trainable Similarity Measure, IEEE Transactions on Intelligent Transportation Systems. 7, no. 3 (2006), s. 309-321. ISSN 1524-9050.
Paclík, P., Novovicova, Jana; A trainable similarity measure for image classification, Proceedings of the 18th Conference on Pattern Recognition. ICPR 2006. Los Alamitos : IEEE, 2006. s. 391-394. ISBN 0-7695-2521-0
Scarpa, G. Haindl, M. Unsupervised Texture Segmentation by Spectral-Spatial-Independent Clustering, 18th IAPR International Conference on Pattern Recognition, Hong Kong, 2006, ISBN 0-7695-2521-0, ISSN 1051-4651, vol. II, pp. 151-154, IEEE Press.
Haindl, M. - Mikes, S. Unsupervised Texture Segmentation Using Multispectral Modelling Approach, 18th IAPR International Conference on Pattern Recognition, Hong Kong, 2006, ISBN 0-7695-2521-0, ISSN 1051-4651, vol. II, pp. 203-206, IEEE Press.
Haindl, M. - Vacha, P. Illumination Invariant Texture Retrieval, 18th IAPR International Conference on Pattern Recognition, Hong Kong, 2006, ISBN 0-7695-2521-0, ISSN 1051-4651, vol. III, pp. 276-279, IEEE Press.

Grim, J. - Haindl, M. - Somol, P. - Pudil, P. A Subspace Approach to Texture Modelling by Using Gaussian Mixtures, 18th IAPR International Conference on Pattern Recognition, Hong Kong, 2006, ISBN 0-7695-2521-0, ISSN 1051-4651, vol. II, pp. 235-238, IEEE Press.

3.2 Contribution by partner 09 CNR-ISTI

Activities

Activity has been carried out for contributing the set up of the ET4 e-team. The method developed for 3D dynamic shape deformation analysis has been also improved. Tests have been performed on biomedical data.
An experimental tool, developed in the ITK environment, has been implemented to improve the real-time display of images.

Events

Anna Tonazzini has presented some results of the basecalling algorithm previously developed in the Workshop on Mass-Data Analysis of Images and Signals in Medicine.

Publications

Tonazzini, L. Bedini, "Statistical analysis of electrophoresis time series for improving basecalling in DNA sequencing", ICDM 2006, Workshop on Mass-Data Analysis of Images and Signals in Medicine, Biotechnology and Chemistry MDA´2006, Leipzig, Germany, July 2006.

3.3 Contribution by by partner TU VIENNA/PRIP

Activities

Coordination of the E-team on "Choosing Features for CBIR and Automated Image Annotation". Visit by Hanbury to E-team member UPC for two days - talk on morphological segmentation given. Development of image segmentation algorithms which segment an image based on a sample of the texture to be found continues. Improving the efficiency and efficacy of this method by creating a graph based on the results of a Maximally Stable Extremal Region (MSER) detector instead of directly on the pixels has been done and the results have been presented at the British Machine Vision Conference in September 2006. We have also been invited to submit it to the Image and Vision Computing special issue on the best of the BMVC. This technology will be included in the object recognition showcase. Previous work has used a two-class maximum flow/minimum cut optimisation algorithm to calculate the segmentations. This means that segmentations into many regions must be done iteratively. Work using multi-class optimisation algorithms continues. A conference paper is under preparation. The investigation of image keypoints based on a measure of symmetry combined with a new feature describing the shape of the area around keypoints is under development. A conference paper is under preparation.

Publications

Hanbury and A. Targhi, "A dataset of annotated animals", 2nd MUSCLE workshop on Image and Video Retrieval Evaluation, Alicante, 2006 (joint KTH-PRIP paper)
B. Micusik and A. Hanbury: Template Patch Driven Segmentation, British Machine Vision Conference (BMVC), Edinburgh, UK, 2006.

3.4 Contribution by partner 12 UPC

Activities

Coordination and planning of the e-team "Person Detection, Recognition and tracking" Presentation of work performed within Muscle WPs 3 and 7 in EUSIPCO 2006 (Muscle

special session) and ICIP 2006. Continuation of work on: -Facial features analysis -Body modelling and human gestures analysis -Video sequences segmentation.

Collaboration with PRIP-Vienna in the context of "Choosing features for CBIR and automated image annotation"
Segmentation of images combining texture gradient with watershed-based approaches. * Development of a boundary-based segmentation evaluation technique.
Preparation of a joint paper "Morphological segmentation on learned boundaries" to be submitted to Image and Vision Computing. Image Classification with Segmentation Graph Kernels: development of a family of kernels between two segmentation graphs, each obtained by watershed transforms from the original images. These kernels are based on soft matchings of subtree patterns of the respective graphs, leveraging the natural structure of images while remaining robust to the segmentation process uncertainty.

Achievements
Completion of a demo tool for facial expression analyis based on images from a web-cam. Results on 3D human body analysis using particle filters.

Events
Event: EUSIPCO 2006, including Muscle Special Session Attendees: Cristian Canton, Ferran Marques Event: ICIP 2006 Attendees: Ferran Marques, Veronica Vilaplana, Camilo Dorea

Publications
C.Canton-Ferrer, J.R.Casas, M.Pardàs, "Human Model and Motion Based 3D Action Recognition in Multiple View Scenarios", European Signal Processing Conference (EUSIPCO), Firenze (Italy), September 4-8 2006.
J. L. Landabaso, M. Pardàs, "Cooperative background modelling using multiple cameras towards human detection in smart-room", European Signal Processing Conference (EUSIPCO), Firenze (Italy), September 4-8 2006.
V. Vilaplana, C. Martínez, J. Cruz, F Marques, "Face Recognition using groups of images in smart room scenario", IEEE Int'l Conference on Image Processing, ICIP'06, Atlanta, USA, October 2006.
C.C. Dórea, M. Pardàs, F. Marqués, Generation of long-term color and motion coherent partitions, IEEE Int'l Conference on Image Processing, ICIP'06, Atlanta, USA, October 2006.
J.L. Landabaso, M.Pardàs, J.R.Casas, "Reconstruction of 3D shapes considering inconsistent 2D silhouettes", IEEE Int'l Conference on Image Processing, ICIP'06, Atlanta, USA, October 2006.

3.5 Contribution by partner 16 AUTH
Activities
Research on facial features detection, namely eyes and mouth detection has continued in AUTH. Within the reporting period, a conference paper on this topic that has been jointly authored by AUTH and UPC and is the result of a joint research effort that took place during the visit of AUTH student S. Asteriadis to UPC has been accepted for publication in the International Conference on Computer Vision Theory and Applications (VISAPP 2007).

Achievements
 The proposed facial feature detection technique achieves results that are above the state of the art in this area.

Publications
S. Asteriadis, N. Nikolaidis, I. Pitas, M. Pardas, Detection of Facial Characteristics Based on Edge Information, International Conference on Computer Vision Theory and Applications (VISAPP 2007), Barcelona, Spain (accepted for publication).

3.6 Contribution by partner 37 ENSEA

Publications
FReBIR : Fuzzy Region-Based Image Retrieval S. Philipp-Foliguet, J. Gony, IPMU 2006
- Multimedia indexing and fast retrieval based on a vote system S. Philipp-Foliguet, Guillaume Logerot, Patrick Constant, P.-H. Gosselin, Christian Lahanier, ICME 2006

3.7 Contribution by partner 21 TECHNION-MM

Activities
Paper on Role of optics in the accuracy of depth from defocus systems

Publications
- Blayvas, E. Rivlin, and R. Kimmel, Role of optics in the accuracy of depth from defocus systems, Accepted to Journal of Optical Society of America A (JOSA).

3.8 Contribution by partner 15 UVA

Activities
E-team on "Visual Saliency" collaboration with INRIA-Imedia - Content-analysis showcase - collaboration with TUWien-IFS

Achievements
Preparation of the initial set of data for the Content-analysis showcase - The research done in collaboration with INRIA-Imedia on context-Based object-class recognition and retrieval using generalized correlograms has yielded a paper accepted for the IEEE Transactions on Pattern Analysis and Machine Intelligence - We investigated the challenges and current trends in human-centered computing and the results were published in ACM Multimedia conference.

Problems
Organization of 1st International Workshop on Human-centered Multimedia, in conjunction with ACM MM

Publications
Human-Centered Computing: A Multimedia Perspective A. Jaimes, N. Sebe, D. Gatica-Perez, ACM Multimedia, Santa Barbara, October 2006.

3.9 Contribution by partner 07 MTA SZTAKI

Activities
Chetverikov (SZTAKI) visited UTIA to give a talk on new results related to MUSCLE activities and discuss joint research in dynamic texture analysis. The partners agreed to test the dynamic texture synthesis method developed by UTIA using the evaluation algorithms developed by SZTAKI.

Focus Area Extraction by Blind Deconvolution for Defining Regions of Interest in Image and Video Indexing Estimation of Vanishing Point in Camera-Mirror Scenes Using Video

Achievements

A joint paper by SZTAKI and TAU-Visual has been written and submitted describing the novel method for dynamic texture detection developed by the partners.

We present an automatic focus area estimation method, working with a single image without a priori information about the image, the camera or the scene. It produces relative focus maps by localized blind deconvolution and a new residual error based classification. Evaluation and comparison is performed, and applicability is shown through image indexing. Knowledge of the vanishing-point position is the key for the geometrical modeling of reflective surfaces or cast shadows. An automatic method is presented using motion statistics to determine correspondences, and an improved fitting function for final parameter estimation which takes into account the statistical properties of image-points. The experiments show that our approach gives robust results in the context of widely different environments especially in cases where the correspondences are corrupted with considerable amounts of noise. We presented new results on using localized blind deconvolution for automatic relative focus map extraction from ordinary images with no explicit knowledge about the image or exposure conditions. We have introduced a new robust error measure based on the orthogonality principle. Multiscale relative maps can be extracted from images and/or video frames which can be used for focus based feature extraction. Proposed applications include image indexing and retrieval, focus tracking in videos, main actor selection, news anchor detection, closeup detection in scenes, object extraction or tracking. In general, focus maps could be well used as a complementary indexing feature in image databases.

Events

Chetverikov attended British Machine Vision Conference (September 2006, Edinburgh) and presented MUSCLE work on temporal periodicity of dynamic texture.
ICIP 2006, MUSCLE & DELOS Image indexing workshop

Publications

D. Chetverikov and S. Fazekas "On motion periodicity of dynamic textures" Proc. British Machine Vision Conference, vol. 1, pp.167-176, 2006. (http://visual.ipan.sztaki.hu/publ/bmvc2006.pdf)
L. Kovács , T. Szirányi, "Focus Area Extraction by Blind Deconvolution for Defining Regions of Interest", IEEE Tr. Pattern Analysis and Machine Intelligence, accepted, 2006
L. Havasi, T. Sziranyi: "Use of Motion Statistics for Vanishing Point Estimation in Camera-Mirror Scenes", International Conference on Image Processing (ICIP), IEEE, Atlanta, 2006
L. Havasi, T. Szirányi, "Estimation of Vanishing Point in Camera-Mirror Scenes Using Video", Optics Letters, Vol. 31, No. 10, pp: 1367-1566, 2006
L. Havasi, Z. Szlávik, T. Szirányi: "Higher order symmetry for non-linear classification of human walk detection", Pattern Recognition Letters, Vol.27, pp.822-829, 2006

3.10 Contribution by partner 19 ACV

Activities
Analysis of the AdaBoost training algorithm for person detection in general surveillance scenarios.

Problems
Adaboost analysis results show that a comprehensive training set that covers all types of possible positive and negative samples is crucial for acceptable performance. Solution: acquire even more image samples.

3.11 Contribution by partner 30 FORTH

Activities

FORTH has continued work on symbolic object detection and cognition, an more specifically on visual detection and 2d/3d tracking of objects according to colour information. FORTH has also continued work 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.

3.12 Contribution by partner 27 TAU-VISUAL

Activities

Prior-based segmentation, mutual segmentation and symmetry-based segmentation research.

Other

Shape-modelling e-team activity with INRIA-ARIANA

3.13 Contribution by partner 03 UCL

Activities

Attention-based similarity has been applied to the problem of copy detection. The precision and recall results are encouraging. Some of this work has been reported at ICANN 2006 in the Special Session on Visual Attention Algorithms and Architectures for Perceptual Understanding and Video Coding. A further paper has been submitted to IEEE Trans on Circuits and Systems for Video Technology.

Investigations of attention based 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. A paper entitled "An attention based method for motion detection and estimation" has been published.

Further exchanges have taken place between INRIA and UCL continuing the collaboration on copy detection. A joint paper describing objective performance comparisons is in preparation.

The detection of perspective in colour images has yielded good results and a paper has been presented at ICIP.

Achievements

The photo colour correction technique presented at SPIE in January 2006 forms the basis of an award of an NCGE fellowship to Ade Bamidele. This will enable him to travel the US for 6 months and establish a case for a spin out company based on this technology.

An on-line internet demonstration of attention-based focusing has been set up using both local cameras and remote PTZ cameras across the world. Focusing commands are issued depending on the results of image analysis. A paper has been presented at ICIP.

Events

ICIP 2006 - F W M Stentiford ICANN 2006 - L Chen LCS 2006 - F W M Stentiford, S Zhang

Publications

F W M Stentiford, "Attention-Based Vanishing point detection," ICIP, 8-11 October, Atlanta, 2006.

R Shilston and F W M Stentiford, "An attention-based focus control system," ICIP, 8-11 October, Atlanta, 2006.

L Chen and F W M Stentiford, "An attention based similarity measure for colour images," ICANN 2006, Special Session on Visual Attention Algorithms and Architectures for Perceptual Understanding and Video Coding, 10-14 September, Athens, 2006.
S Zhang and F W M Stentiford, "An attention based method for motion detection and estimation," London Communications Symposium, 14-15th September, 2006.

3.14 Contribution by partner 11 TUG

Activities
We are working on a new, computationally efficient and rather sparse object representation based on a very simple specificity of spatial relations.
Object tracking was formulated as a matching problem of detected keypoints between successive frames. Classifiers for local features are learned whereas samples are collected over time.

3.15 Contribution by partner 32 INRIA VISTA

Activities
We concerned the problem of object class recognition and localization in natural images. Building upon recent advances in the field we showed how histogram based descriptors can be combined with a boosting classifier to provide a state of the art object detector. Among the improvements we introduced a weak learner for multi-valued histogram features and showed how to overcome problems of limited training sets. We also addressed the computational aspects analyzing the tradeoff between the speed and the accuracy of the detector. Validation of the method on recent benchmarks for object recognition showed its superior performance. In particular, using a single set of parameters our approach outperforms all the methods reported in VOC05 Challenge for 7 out of 8 detection tasks and four object classes while providing close to real-time performance.
Segmentation of transparent layers in video. We developed a method for segmenting moving transparent layers in video sequences. We make assumptions that the images can be divided into areas containing at most two moving transparent layers. We call this configuration (which is the mostly encountered one) bi-distributed transparency. The proposed method involves three steps: initial block matching for two-layer transparent motion estimation, motion clustering with 3D Hough transform, and joint transparent layer segmentation and parametric motion estimation. The last step is solved by the iterative minimization of a MRF-based energy function. The segmentation is improved by a mechanism detecting areas containing one single layer. The framework is applied to various image sequences with satisfactory results.

Publications
Laptev, "Improvements of Object Detection Using Boosted Histograms" (2006) in Proc. BMVC'06, Edinburgh, UK, pp. III:949-958.
Vincent Auvray, Patrick Bouthemy, and Jean Lienard, "Motion-based segmentation of transparent layers in video sequences" (2006) International Workshop on Multimedia Content Representation, Classification and Security (IWMRCS), Springer LNCS 4105, pp. 298-305.

4 Overview activities in WP4

4.1 Contribution by partner 09 CNR-ISTI

Activities
Definition and planning of activity and costs for the e-Team on "Semantic fron audio: features, perception and synthesis" (G. Bertini)

Achievements
Developed a platform suitable for audiology specialists in order to implement not-standard gain curves for wide band compensation of presbycusic people for a better quality hearing of music.

Events
Results presented in the Audio Processing/Enhancement Poster Session (chair G. Bertini) at 14th EUSIPCO 2006 Florence, Italy, 4-8 Sept. 2006.
Joint paper ISTI/CNR - TU Wien at 7th DSP Applications Day Milan, Italy, 20 Sept. 2006.

Publications
Bibliography of MUSCLE Papers :
MP 517 (EUSIPCO 2006),
MP 518 (EUSIPCO 2006),
MP 519 (joint paper, DSP Appl. Day 2006)

Other
Partnership of ISTI audio team in a new EC Leonardo Project - M.O.D.E.M (Music Open Distance Exchange Model- Contract n. I/05/B/F/PP-154059) regarding a web-based environment for exchanging multi-track audio/musical signals.

4.2 Contribution by partner 18 TU Vienna IFS

Activities
- Participated in the ISMIR 2006 conference with a paper presentation on "Visually Pro\ufb01ling Radio Stations" and a poster presentation of "The Map of Mozart"
- Present at the annual Music Information Retrieval Evaluation eXchange (MIREX) benchmarking forum. Presented a poster about results.
- Wrote report about contribution to MIREX 2006 on "Computing Statistical Spectrum Descriptors for Audio Music Similarity and Retrieval"

Achievements
- The first large-scale human listening test for Music Similarity and Retrieval in MIREX 2006 showed, that TU VIENNA IFS approach is competing with state-of-the-art algorithms - no significant difference in performance has been found between the top 5 algorithms. It was also one of the two fastest algorithms, with by far the most efficient distance calculation.

Events
- ISMIR 2006: Andreas Rauber and Thomas Lidy
- MIREX 2006: Andreas Rauber and Thomas Lidy

Publications

Thomas Lidy, Andreas Rauber: Visually Profiling Radio Stations. In: Proceedings of the 7th International Conference on Music Information Retrieval, Victoria, Canada, October 8-12, 2006

Rudolf Mayer, Thomas Lidy, Andreas Rauber: The Map of Mozart. In: Proceedings of the 7th International Conference on Music Information Retrieval, Victoria, Canada, October 8-12, 2006

Thomas Lidy, Andreas Rauber: Computing Statistical Spectrum Descriptors for Audio Music Similarity and Retrieval. MIREX 2006 - Music Information Retrieval Evaluation eXchange, Victoria, Canada, October 8-12, 2006

4.3 Contribution by partner 05 BILKENT

Activities

We developed a natural language based interface for the BilVideo Video
Database System. The NLP-based interface allows users to formulate
queries as sentences in English.

Publications

Onur Kucuktunc, Ugur Gudukbay, Ozgur Ulusoy, "A Natural Language Based Interface for Query Specification in a Video Database Management System", IEEE Multimedia, Multimedia-at-Work, to appear.

4.4 Contribution by partner 38 CNRS

Activities

Automatic versus interactive analysis of Arabic;
Proposal of metrics for evaluating the interactive analysis;
Design and implementation of software for interactive vowellization, lemmatisation and POS-tagging of Arabic.

4.5 Contribution by partner 40 EC3

Activities

Research on 3D virtual worlds for visualization of music information spaces. In particular, we are investigating the use of game engines as an interface to Self-Organizing Maps (SOMs). The spatial arrangement of data items based on their similarity will be transformed into a 3D representation allowing multiple users interactive exploration.

4.6 Contribution by partner 21 TECHNION-MM

Activities

Research on a short time Beltrami kernel for smoothing images and manifolds

Publications

Spira, R. Kimmel, and N. Sochen, A short time Beltrami kernel for smoothing images and manifolds, Accepted to IEEE Trans. Image Processing.

4.7 Contribution by partner 16 AUTH

Activities

Supervised classifiers based on non-negative matrix factorization for musical instrument classification have been developed. Joint research has been performed with Andreas Rauber and Thomas Lidy of Technical University of Vienna. The related research is part of Eteam 8.

Events

C. Kotropoulos attended XIV European Signal Processing Conference.

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.

4.8 Contribution by partner 26 TAU SPEECH

Activities

Continued experimentation and study with single-channel audio separation of speech and music.
Algorithms for support vector machine rescoring of hidden Markov models with applications to speech recognition.

Achievements

Design of adaptive sparsifying transformations for improved separation.
Design of improved transformations for support vector machine rescoring of hidden Markov models, and the testing of these algorithms on speech recognition tasks.

Problems

Speech / Music Separation quality is still below expectations.

Events

Participation in EUSIPCO 2006, Florence.

Publications

Submitted the following papers:
Alba Sloin and David Burshtein, ``Support Vector Machine Training for Improved Hidden Markov Modeling'', Submitted to IEEE Transactions on Signal Processing.
Alba Sloin and David Burshtein, ``Support vector machine re-scoring of hidden Markov models,'' Proceedings of the 24th IEEE Conference of Electrical and Electronics Engineers in Israel, Eilat, Israel, (presented on November 2006).

4.9 Contribution by partner 27 TAU VISUAL

Activities

1) Continued experimentation and study with single-channel audio separation of speech and music.
2) Algorithms for support vector machine rescoring of hidden Markov models with applications to speech recognition.

Achievements

1) Design of adaptive sparsifying transformations for improved separation.

2) Design of improved transformations for support vector machine rescoring of hidden Markov models, and the testing of these algorithms on speech recognition tasks.

Problems
Speech / Music Separation quality still below expectations.

Events
Participation in EUSIPCO 2006, Florence

Publications
Submitted the following papers (partially supported by Muscle):
Alba Sloin and David Burshtein, ``Support Vector Machine Training for Improved Hidden Markov Modeling'', Submitted to IEEE Transactions on Signal Processing.
Alba Sloin and David Burshtein, ``Support vector machine re-scoring of hidden Markov models,'' Proceedings of the 24th IEEE Conference of Electrical and Electronics Engineers in Israel, Eilat, Israel, (presented on November 2006).

5   Overview activities in WP5

5.1   Contribution by partner 05 BILKENT

Activities
1. Current CCTV surveillance systems are mostly based on video. It is now possible to install cameras monitoring sensitive areas but it may not be possible to assign a security guard to each camera or a set of cameras. In addition, security guards may get tired and watch the monitor in a blank manner without noticing important events taking place in front of their eyes. Recently, intelligent video analysis systems capable of detecting humans, cars, etc. were developed. Such systems mostly use HMMs or SVMs to reach decisions. They detect important events but they also produce false alarms. It is possible to take advantage of other low cost sensors including audio to reduce the number of false alarms. Most video recording systems have the capability of recording audio as well. The analysis of audio for intelligent information extraction is a relatively new area. Automatic detection of broken glass sounds, car crash sounds, screams, increasing sound level at the background are indicators of important events. By combining the information coming from the audio channel with the information from the video channels, reliable surveillance systems can be achieved. In this work, a surveillance system that utilizes both video and audio to detect fight among people at unattended places is under-development.

2. We continued our work on the implementation of hand gesture-based interaction for the specification of spatial relations between objects of interest in videos, especially the 3D relations. The implementation is based on taking hand movements as input with the help of a camera. Different query types require different interaction modalities, a multimodal interface, including visual (mouse-based), natural language-based, and gesture-based components, is the most appropriate way of query specification.

3. We worked on fighting people detection using both sound and visual information. We compared various sound feature vectors for shouting detection. We haven't published our results.

5.2   Contribution by partner 14 UTIA

Activities

A fast range image segmentation method 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. The exploitation of multimodal data significantly improves the segmentation quality.

Events

Results presented on 18th ICPR, Hong Kong.

Publications

Haindl, M. -Zid, P. Multimodal Range Image Segmentation by Curve Grouping, 18th IAPR International Conference on Pattern Recognition, Hong Kong, 2006, ISBN 0-7695-2521-0, ISSN 1051-4651, vol. IV, pp. 9-12, IEEE Press.

5.3   Contribution by partner 23 ICCS-NTUA (also TSI-TUC, INRIA-Parole, INRIA-Texmex)

Activities

Research Activity #1: Audio-Visual Interaction for Speech Recognition
MUSCLE members involved ICCS-NTUA, TSI-TUC
Researchers involved in WP activity G. Papandreou, A. Katsamanis, V. Pitsikalis, P. Maragos (ICCS-NTUA) A. Potamianos, E. Sanchez-Soto (TSI-TUC)
Concise description of WP activity 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 speaker's 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 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) and Asynchronous HMMs (A-HMM), obtaining further improvements. While in our earlier research we had focused in the incorporation of feature uncertainty in decoding/classification rules only, in our latest research

during the reporting period we have extended our approach to statistical multimodal learning, as well. Specifically, we have shown how conventional multimodal learning rules based on the Expectation-Maximization (EM) algortihm for Gaussian Mixture Models (GMMs) and HMMs should be modified when features in the training set are only known with limited precision. This introduces new interesting feature regularization and cross-modal interaction phenomena during model training and leads to more robust and better able to generalize models, particularly in the case of few available training data. During the reporting period, we have launched an effort to build a real-time audio-visual automatic speech recognition demonstrator, as part of the Muscle Showcasing initiative. Participating partners are TSI-TUC (leader), ICCS-NTUA, and INRIA-TEXMEX. More specifically, experience with AV-ASR so far has been confined to research-level experimental setups: typically, videos of the speakers are shot under carefully controlled conditions, acoustic noise is artificially added, and processing is performed off-line. Towards practically deployable AV-ASR, we have started building a proof-of-concept laptop-based AV-ASR prototype that will: (i) use consumer microphone and camera to capture the speaker; (ii) perform visual/audio feature extraction, as well as speech recognition on the laptop in real-time; (iii) will be robust to failures of a single modality, such as visual occlusion of the speaker's face; and (iv) automatically adapt to changing acoustic noise levels. The system is planned to be delivered by June 2007.

Research Activity #2: Audiovisual Attention Modeling and Salient Event Detection
MUSCLE members involved ICCS-NTUA
Researchers involved in WP activity G. Evangelopoulos, K. Rapantzikos, and P. Maragos
Concise description of WP activity 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. During the reporting period, we have launched an effort to build a Movie Summarization and Skimming Demonstrator, as part of the Muscle Showcasing initiative Participating partners are ICCS-NTUA (leader), TSI-TUC, AUTH, and INRIA-TEXMEX. As the amount of video data available (movie, TV programs, clips) in a personal recorder or computer are becoming increasingly large (100h in VCRs or hundreds of hours on

a PC) intelligent algorithms for efficiently representing video data and presenting them to the user are becoming important. Video summarization, movie summarization and movie skimming are increasingly popular research areas with immediate applications. In this showcasing project we will: (i) use combined audio and video saliency detectors to identify the importance of movie content to the user and (ii) design an interface that presents the audio and video information to the user in a compressed form, thus saving time with little or no loss of information. The demonstrator will have the ability to render a movie from its typical 2h duration down to 30' by skimming over (fast forwarding or omitting) non-salient movie scenes while playback at regular speed parts of the movie with salient audio and video information. The interface will also have the ability to break the synchrony of the audio/video streams and selectively present audio or video information. The system is planned to be delivered by June 2007.

Research Activity #3: 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 a.Multimodal Processing, Interaction and Understanding multimedia content i.Audio-Visual ASR ii.Feature fusion iii.Video Analysis and Integration of Asynchronous Modalities b.Audio-Visual Saliency i.Audio-Visual Scene Change and Dialogue Detection ii.Audio-Visual Attention and Salient Event Detection c.Searching multimedia content i.Annotation of multimedia databases ii.Information retrieval for video or other multimedia databases iii.Integration of Vision + Text or Audio + Text d.Interfaces to multimedia content i.Multimodal dialogue interfaces ii.Eye-tracking interfaces for information retrieval iii.Mobile interfaces

----

Research Activity #4: Audio-Visual Speech Inversion

MUSCLE members involved ICCS-NTUA, INRIA-LORIA

Researchers involved in WP activity A. Katsamanis, A. Roussos, G. Papandreou, P. Maragos (ICCS-NTUA) Y. Laprie (INRIA-LORIA)

Concise description of WP activity This research activity concerns the development of audiovisual-to-articulatory speech inversion methods. Acoustic-to-articulatory inversion is an acoustical problem that relies onto a physical model and consists of recovering geometrical data from acoustic parameters. The main difficulty is that fundamentally important aspects of the physical system of speech production, i.e. the geometry of the vocal tract, the physical characteristics of the wall, and the dependencies of the acoustic signal from the physics of speech production, cannot be measured precisely. In addition, some of the natural data of this problem, i.e. the resonance frequencies of the vocal tract (called formants) cannot be extracted easily from speech. Alternative inversions from time-varying multiscale spectral data and/or audiovisual data are central objectives of our research because they would enable the exploitation of inversion in real applications. The acoustic data are either recorded speech or measured formant frequencies. In both cases the scientific difficulty is the infinite number of articulatory solutions. Thus, we need to deal with the incorporation of constraints that reduce the under-determination of this problem. Several aspects are investigated: the nature of information that can be incorporated (standard phonetic knowledge, audiovisual data), the most appropriate algorithmic framework and how visible constraints can be derived from video images of speaker's face. Indeed, there is strong evidence that the view of visible articulators, i.e. jaw and lips, is an important source of information for inversion. The original

contributions include novel methods for automatic formant tracking, novel methods for inversion without formants' knowledge (e.g. from multiscale time-frequency information, improved articulatory speech modelling, audiovisual data), and the introduction and optimal exploitation of constraints, particularly those derived from speaker's face via appropriate video processing.

Publications
Katsamanis, G. Papandreou, V. Pitsikalis, and P. Maragos, "Multimodal Fusion by Adaptive Compensation for Feature Uncertainty with Application to Audiovisual Speech Recognition", in Proc. of European Signal Proc. Conf. (EUSIPCO-2006), Florence, Italy, Sep. 2006.
V. Pitsikalis, A. Katsamanis, G. Papandreou, and P. Maragos, "Adaptive Multimodal Fusion by Uncertainty Compensation", in Proc. of Int'l Conf. Speech and Language Processing (ICSLP/INTERSPEECH-2006), Pittsburgh, PA, USA, Sep. 2006.

5.4   Contribution by partner 32 INRIA-Texmex

Achievements
Our work in the frame of the PhD thesis of Manolis Delakis enwt to an end with the defense of the Thesis. Petros Maragos, head of WP5, was invited as a reviewer of the thesis.

5.5   Contribution by partner 24 TSI-TUC

Activities
Research on multimodal dialogue interfaces investigates the optimal combination of various modalities mainly speech and tactile (keyboard, mouse). In our experiments we analyze user behavior as we modify the relative efficiency of the various input modes and well as the behavior of users towards adaptive interfaces.

Achievements
Results from our recent evaluations on desktop and PDA environments for our speech+tactile input interfaces shows that (in addition to our previous results reported in summer 2006): (1) users tend to use the more efficient modality but are biased toward speech (further research is needed to understand if this bias holds over time or is just a novelty effect), (2) our adaptive interface (modality tracking) performs better in terms of interaction time but worse in terms of inactivity time (probably adaptivity confuses users and imposes higher cognitive load), (3) most of the gain in performance of multimodal dialogue interfaces (vs speech-only interfaces) comes from visual feedback. These results are especially relevant for multimodal interface design.

Events
Presentation in MUSCLE plenary meeting Dec 2007.

Publications
M. Perakakis, M. Toutoudakis, and A. Potamianos, "Blending speech and visual input in multimodal dialogue systems," in IEEE/ACM Workshop on Spoken Language Technology, (Aruba), Dec. 2006.

5.6   Contribution by partner 21 Technion MM

Activities
paper on Geometric Curve Flows on Parametric Manifolds.

Achievements

We have developed efficient numerical schemes for the implementation of the classical geodesic curve evolution equations on parametric manifolds.

Publications

A. Spira, and R. Kimmel, Geometric curve flows on parametric manifolds, Accepted to Journal of Computational Physics.

5.7   Contribution by partner 16 AUTH

Activities

Speaker clustering algorithms are studied within the framework of dialogue detection in movies. Lip activity detection using mouth region intensity information is investigated in the same framework. A framework for audio-assisted two-person dialogue detection based on indicator functions and neural networks has been developed. The work is part of Eteam 11.

Publications

M. Kyperountas, C. Kotropoulos, and I. Pitas, "Enhanced eigen-audioframes for audiovisual scene change detection," IEEE Trans. Multimedia, accepted for publication, October 2006.

5.8   Contribution by partner 30 FORTH

Activities

During the reporting period, FORTH was activein fields related to Multimodal Processing and Interaction. Emphasis was given to research related to recognition and interpretation of hand gestures for human/computer and human/robot interraction.

5.9   Contribution by partner 03 UCL

Activities

WP 5.4.2 Eye-Tracking Interfaces for Information Retrieval This period has been devoted to consolidation and analysis. A significant conclusion from the most recent results confirms that pre-attentive vision is used in visual search. Participants are able to find target images in sessions where images were viewed for very short periods of 200ms and 300ms. In both cases participants reported that they felt that they had little control over the interface, but in spite of this the targets were located faster than a random search. Research results are continuing to be analysed and written up in a PhD thesis entitled Eye Tracking: A Perceptual Interface for Content Based Image Retrieval.

6   Overview activities in WP6

6.1   Contribution by partner 02 CWI

Activities

Marie van Liesehout. Invited talk 17 - 21 September: Deutsche Mathematiker-Vereinigung Jahrestagung 2006, Bonn, Germany. Perfect simulation for length-interacting polygonal Markov fields in the plane.

Marie van Liesehout (CWI) Invited talk 4 October: Signals and Images Seminar, CWI Image segmentation by polygonal Markov fields

6.2 Contribution by partner 03 UCL

Activities

Trip of Dr P.-H. Gosselin and Prof. M. Cord to visit Prof. J. Shawe-Taylor at the University College of London, in order to extend the collaboration between ENSEA and UPMC on the video browsing and retrieval, and to initiate a collaboration on machine learning for object class recognition in supervised and semi-supervised framework, with this partner in the PASCAL NoE.

6.3 Contribution by partner 09 CNR-ISTI

Activities

Definition of a new plan of activity for the e-Team on unsupervised segmentation. Research on blind separation methods for dependent sources.

Achievements

Extension of the MaxNG separation procedure to cases with nonnegligible noise. Definition of a strategy to overcome the scaling ambiguity in the sources reconstructed blindly from remote-sensed images.
Evaluation tests of a laboratory procedure for segmentation of colour images.

Problems

The number of goups participating in the e-Team on unsupervised segmentation was reduced. Consequently, the research plan has been limited to MCMC strategies for blind source separation from astrophysical data.

Events

Organization of the MUSCLE Special Session "Recognizing humans and human behavior in video" of the 14th European Signal Processing Conference – EUSIPCO 2006, September 4-8, Florence Italy. Presentation at the same conference of the results achieved in the field of active video surveillance, based on stereo and infrared imaging. Further work on the unsupervised/supervised analysis of complex colour images. The fuzzy-neural segmentation of cytological images has been improved adding a colour space transformation for better exploit the colorimetric information.

Publications

Authors: Caiafa, C.F., E. Salerno, A. N. Proto, L. Fiumi , Title: Dependent component analysis as a tool for blind spectral unmixing, Published in: Proceedings of the European Signal Processing Conference, Firenze, Italy, 4-8 September 2006.
Authors: Salerno, E. , Title: A noisy data model for MaxNG, Published in: ISTI-CNR Technical Report, Pisa, Italy, NI ISTI-B4-09, September 2006.
Authors: Salerno, E., Title: Scale disambiguation of blindly unmixed endmember abundances, Published in: ISTI-CNR Technical Report, Pisa, Italy, NI ISTI-B4-10, September 2006.
Authors: Salerno, E., A. Tonazzini, L. Bedini , Title: Digital image analysis to enhance underwritten text in the Archimedes palimpsest , Published in: Int. J. Doc. Anal. & Recogn., DOI: 10.1007/s10032-006-0028-7.
Authors: G. Pieri, O. Salvetti , Title: Active Video-Surveillance Based on Stereo and Infrared Imaging
Published in: Proc. of the 14th European Signal Processing Conference – EUSIPCO 2006, September 4-8, Florence Italy

6.4 Contribution by partner 14 UTIA

Activities

D. Ververidis (NTUA) visited UTIA, Prague between September 2 and 8, 2006. He discussed with Prof. M.Haindl, Prof. J. Grimm, Prof. P. Somol, and Prof. J. Novovicova his work on speeding-up and increasing the accuracy of feature selection methods.

A novel filter feature selection method based on mutual correlation and targeted for huge feature spaces was developed in cooperation with AUTH (D. Ververidis, C. Kotropoulos). The classification performance of the proposed filter method was assessed by using the selected features to the Bayes classifier. Alternative filter feature selection methods that optimize either the Bhattacharrya distance or the divergence were 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 this comparative study. A framework that enables the use of traditional feature selection algorithms in a new context - for building a set of subsets of specified properties was developed. During the course of search individual items are added/removed to/from one of the subsets in the subset system one at a time to maximize an overall criterion. Different tasks of prototype search type can be solved in this alternative way depending on suitable criterion definition. The usability of the concept is shown on keyword extraction example. Among recent topics studied in context of feature selection the hybrid algorithms seem to receive particular attention. We proposed a new hybrid algorithm, the flexible hybrid floating sequential search algorithm, that combines both the filter and wrapper search principles. The main benefit of the proposed algorithm is its ability to deal flexibly with the quality-of-result versus computational time trade-off and to enable wrapper based feature selection in problems of higher dimensionality than before.

Problems

A deviation of the person-month breakdown included in JPA3 has occurred. In particular, although AUTH did not forecast any contribution to WP6, we think that the research actually performed fits best this WP.

Events

Results presented in the 11th Iberoamerican Congress on Pattern Recognition, Cancun, Mexico, in the 18th ICPR, Hong Kong and in S+SSPR, Hong Kong.

Publications

Authors: Haindl, M., Somol, P., Ververidis, D., Kotropoulos, C., Title Feature Selection Based on Mutual Correlation , Published in: Lecture Notes in Computer Science 4225, ISSN 0302-9743, Springer-Verlag, Berlin, pp. 569–577, 2006.
Authors: Somol, Petr; Pudil, Petr , Title: Multi-Subset Selection for Keyword Extraction and Other Prototype Search Tasks Using Feature Selection Algorithms , Published in: Proc. ICPR 2006, Hong Kong, IEEE Press
Authors: Somol, Petr; Novovicova, Jana; Pudil, Pavel, Title Flexible-Hybrid Sequential Floating Search in Statistical Feature Selection , Published in: Lecture Notes in Computer Science, 4109, Springer

6.5 Contribution by partner 24 TSI-TUC

Activities

Unsupervised combination of metrics with application to automatic semantic class induction.

Achievements

We have devised an algorithm for unsupervised computation of weights to individual metrics of different lexical scopes for the problem of automatic semantic class induction from text. The various metrics are linearly combined into a hybrid metric. The proposed algorithm monitors the efficiency of each individual metric and attempts to assign greater weight to the ``best-performing'' metric. The algorith is based on the inter- and intra-class distance of the classes resulting from each metric. Experiments on two different corpora showed that the adaptively computed weights outperform the fixed weight computation scheme. Also three metric combination (wide-, mid- and narrow-context size) significantly outperformed each one of the individual metrics in terms of precision and recall of generated classes. The proposed unsupervised metric combination algorithm makes it possible to employ a corpus-independent semantic similarity metric for semantic class induction. It can also be generalized to other problems where the fusion of various metrics is required.

6.6 Contribution by partner 29 TCD

Activities

Rozenn Dahyot. Conference management On programme committee of IMVIP 2006
Simon Wilson. Contributed talk Improving CBIR by modelling the search process: a Bayesian approach, September 2006. Proceedings of the European Signal Processing Conference

Achievements

Simon Wilson. Start of Science Foundation Ireland research grant in "Variational Methods" (in collaboration with Anthony Quinn at TCD and TAU-Visual).

Events

Dr. Anthony Quinn joins the MUSCLE group at TCD. His interests are in image processing and computational methods in Bayesian signal processing.

Publications

Authors: Dahyot, R., Title: Unsupervised Camera Motion Estimation and Moving Object Detection in Videos, Published inL Irish Machine Vision and Image Processing conference, Dublin Ireland, Sept. 2006.
Authors: Rea, N., G. Lacey, C. Lambe and R. Dahyot , Title: Multimodal Periodicity Analysis for Illicit Content Detection in Videos , Published in: 3rd European Conference on Visual Media Production (IET CVMP 2006), London November 2006
Authors: Wilson, Simon , Title: Improving CBIR by modelling the search process: a Bayesian approach, Published in: Proceedings of the European Signal Processing Conference, Firenze, Italy.
Authors: Grimaldi, M., Cunningham, P., Kokaram, A., Title: Discrete wavelet packet transform and ensembles of lazy and eager learners for music genre classification, Published in: Multimedia Systems, Apr 2006, Pages 1 - 16, DOI 10.1007/s00530-006-0027-z.
Authors: Nugent, C., Cunningham, P., Kirwan, P., Title: Object Recognition and Active Learning in Microscope Images, Published in: Proceedings of 17th Irish Conference on Artificial Intelligence and Cognitive Science. D. A. Bell, P. Milligan, and P. P. Sage, (eds.) pp182-191.
Authors: Zheng H., Cunningham P., Tsymbal A., Title: Adaptive offset subspace self-organizing map with an application to handwritten digit recognition, Published in: 7th Multimedia Data Mining Workshop MDM/KDD '06 (in conjunction with KDD 06), ACM Press, 2006.

Authors: Greene, D., Cunningham, P., Title: Efficient Prediction-Based Validation for Document Clustering, Published in: 17th European Conference on Machine Learning (ECML 2006), J. Fürnkranz, T. Scheffer, M. Spiliopoulou (Eds.) LNCS 4212, pp663-670, Springer Berlin , 2006 dx.doi.org/10.1007/11871842_65.

6.7 Contribution by partner 32 INRIA-VISTA

Activities

Patrick Bouthemy and Ivan Laptev organized a special session "Content Analysis and Representation" within International Workshop on Multimedia Content Representation, Classification and Security (IWMRCS) that was held in Istanbul, September 2006. Five papers have been selected and presented during this session from which three papers were from MUSCLE partners.

Events

International Workshop on Multimedia Content Representation, Classification and Security (IWMRCS), Istanbul, September 2006. Attended by Patrick Bouthemy, Vincent Auvray and Ivan Laptev.

Publications

Laptev, Ivan; Title: Improvements of Object Detection Using Boosted Histograms URL: http://www.irisa.fr/vista/Equipe/People/Laptev/objectdetection.html, Published in: in Proc. of British Machine Vision Conference, September 4-7, Edinburgh, UK
Authors: Law-To, Julien; Buisson, Olivier; Gouet, Valerie, Title: Robust Voting Algorithm Based on Labels of Behavior for Video Copy Detection , Published in: Proceedings of the ACM Multimedia, Santa Barbara

6.8 Contribution by partner 37 ENSEA

Activities

Trip of Dr P.-H. Gosselin and Prof. M. Cord to visit Prof. J. Shawe-Taylor at the University College of London, in order to extend the collaboration between ENSEA and UPMC on the video browsing and retrieval, and to initiate a collaboration on machine learning for object class recognition in supervised and semi-supervised framework, with this partner in the PASCAL NoE.

6.9 Contribution by partner 41 UPMC

Activities
Participation to TRECVID 2006

Events
Results and Paper submission: SHOT BOUNDARY DETECTION AT TRECVID 2006 G. Camara Chavez, F. Precioso, M. Cord, S. Philipp-Foliguet

Publications
Camara Chavez, Guillermo; Cord, Matthieu, Title: Video Segmentation by Supervised Learning , Published in: The Brazilian Symposium on Computer Graphics and Image Processing, SIBGRAPI, held in October 8-11, 2006, at Manaus

6.10 Contribution by partner 42 NUID/UCD

Activities
Preparing contributions for the book on Machine Learning and Multimedia Processing to be published by Springer Verlag. Assisting Matthieu Cord with editing this book.

Problems
There was a delay in signing a contract with Springer confirming that they would publish the book.

7 Overview activities in WP7

7.1 Contribution by partner 05 BILKENT

Activities
A call for "showcase" proposals was issued to all partners in October 2006. A "showcase" proposal must be submitted by at least two partners. At the end of each showcase project, a software or a demonstration must be produced. These will be demonstrated in the "International Conference on Image and Video Retrieval (CIVR 2007), Amsterdam, 9-11 July 2007".

Achievements
A total of 17 showcase proposals was submitted to the call. This high number of proposals shows that this call is widely accepted by partners and it promotes joint research and development within MUSCLE. The proposals will be evaluated by the Scientific Committee during the General Research Meeting that will be held in Paris in December.

Other
Nozha Boujemaa is a member of the Steering Board of NEM ETP (Networked and Electronic Media European Technology Platform). Dissemination of MUSCLE research to NEM partners is encouraged. There will be a special session on MUSCLE-related research in "The Fifth International Workshop on Content-Based Multimedia Indexing CBMI-2007 will take place in Bordeaux, France on June 25-27, 2007."

7.2 Contribution by partner 20 Technion MM

Activities
 BMVC 06 conference

Achievements
oral presentation and poster presentation

Publications
Oral presentation: Pose and Motion from Omnidirectional Optical Flow and a Digital Terrain Map Authors: Ronen Lerner, Oleg Kupervasser, Ehud Rivlin
Poster: Functional 3D Object Classification Using Simulation of Embodied Agent by Bar Aviv and Rivlin, BMVC 06 proceedings, ISBN: 1-904410-14-6

7.3   Contribution by partner 07 MTA SZTAKI

Activities
AJAX based folk song search and retrieval system

Achievements
A digital folk song search and retrieval system with a hand gesture based interface is presented. Tillarom is a comprehensive collection of original Hungarian folk songs recorded using different technologies such as phonographs and/or stereo DAT cassettes. This digital archive contains professional quality metadata records as well as MIDI recordings for presenting the different types of clustered folk songs. An AJAX based search and retrieval interface was developed that can be used together with optically recognized Kodály's hand signs to formulate queries through a web browser. The appearance based recognition of hand gestures utilizes contour analysis and SVM based classification. We evaluated the performance of the recognition of hand signs and investigated the main problems of their usage in our system.

Problems
with gesture interface based on Kodály hand signs

Events
International Workhsop on Human-Centered Multimedia

Publications
 Attila Licsar; Tamás Szirányi; László Kovács; Balazs Pataki> "Tillarom: an AJAX based folk song search and retrieval system with gesture interface based on Kodály hand signs", First International Workhsop on Human-Centered Multimedia, Santa Barbara, USA, 2006

7.4   Contribution by partner 27 TAU VISUAL

Activities
Dynamic texture segmentation and analysis.

Achievements
Submission of a joint (TAU-VISUAL and MTA SZTAKI) article to a conference.

Other
Collaboration between TAU-VISUAL and MTA-SZTAKI.

7.5   Contribution by partner 12 UPC

Activities
Coordination and planning of the e-team "Person Detection, Recognition and tracking" Presentation of work performed within Muscle WPs 3 and 7 in EUSIPCO 2006 (Muscle special session) and ICIP 2006.

Achievements
Completion of a demo tool for facial expression analyis based on images from a web-cam.

8 Resource Table

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

Institute	WP1a	WP1b	WP2	WP3	WP4	WP5	WP6	WP7	Total
01 - ERCIM	2.4	0	0	0	0	0	0	0	2.4
03 - UCL	0	0	0	3	0	0	0	0	3
05 - BILKENT	0	0.2	0.4	0	0.4	0.65	0	1.6	3.25
06 - VIENNA PRIP	0	0	0.3	3	0	0	0	0	3.3
07 - MTA SZTAKI	0	0	0	1	0	1	1.5	1.5	5
09 - CNR-ISTI	0	0.32	0.64	1.6	0.5	0	1.64	0	4.7
11 - TUG	0	0.13	0	1.8	0	0	2.1	0	4.03
12 - UPC	0	0	0	1	0	0	0	1	2
14 - UTIA	0.1	0	0.3	0.4	0.5	0.4	0.5	0	2.2
15 - UVA	0	0	0	2.4	0	0	0	0	2.4
16 - AUTH	0	0	0	0	0	0	1.214	0	1.214
18 - TU VIENNA IFS	0	0.05	0	0	2.5	1	0	0	3.55
19 - ACV	0	0	0	0.25	0	0	0	0	0.25
21 - TECHNION-MM	0	0	0	0.16	0.1	0.38	0	0.65	1.29
25 - ARMINES	0	0	0	1.5	0	0	0	0	1.5
26 - TAU-SPEECH	0	0	0	0	3	0	1.4	0	4.4
27 - TAU-VISUAL	0	0	0	2	0	0	0	1	3
29 - TCD	0	0.2	0.6	0	0	0	2	0	2.8
30 - FORTH	0	0	0.22	0.11	0.11	0.22	0.22	0.67	1.55
31 - VTT	0	0	0	0	0	0.01	0	0	0.01
32 - INRIA Ariana	0	0.17	0	0.23	0	0.55	0.17	0	1.12
32 - INRIA Vista	0	0	0	1.7	0	0	1.3	0	3
37 - ENSEA	0	0.22	0.05	0.15	0	0	0.3	0	0.72
40 - EC3	0	0	0	0	2.6	0	0	0	2.6
42 - NUID / UCD	0	0	0	0	0	0	0.2	0	0.2
Total	**0.1**	**1.29**	**2.51**	**20.3**	**10.71**	**4.21**	**12.544**	**6.42**	**58.084**