

PROGRAM

Wednesday 6 June, 2007	
9:00-9:30	Registration
9:30-10:30	Invited talk I: Data-driven and Procedural Analysis and Synthesis of Multimedia, I. Essa
10:30-11:00	Coffee Break
11:00-12:00	Analysis I (Oral Session)
12:00-12:30	Coffee Break
12:30-13:30	Analysis II (Oral Session)
13:30-14:50	Lunch
14:50-16:10	COST 292 special session (Oral Session)
16:10-16:40	Coffee Break
16:40-18:00	COST 292 special session (Oral Session)

Thursday 7 June, 2007	
9:00-10:00	Invited Talk II: Recent Advances and Open Issues of Digital Image/Video Search, S.-F.Chang
10:00-10:30	Coffee Break
10:30-11:30	Tracking (Oral Session)
11:30-12:00	Coffee Break
12:00-13:20	Analysis III (Oral Session)
13:20-14:50	Lunch
14:50-16:20	Poster Session I
	Coffee Break
16:20-17:40	Coding (Oral Session)

Friday 8 June, 2007	
9:00-10:50	Industry Session
10:50-11:20	Coffee Break
11:20-12:30	CHORUS Panel on AudioVisual Search Engines
12:30-14:00	Lunch
14:00-15:30	Poster Session II
	Coffee Break
15:30-17:10	Applications (Oral Session)

CONFERENCE COMMITTEE

General Chair

Yiannis Kompatsiaris, Informatics and Telematics Institute, Greece
Yannis Avrithis, National Technical University of Athens, Greece

Technical Chairs

Noel O'Connor, Centre for Digital Video Processing, Dublin City University, Ireland
Vasileios Mezaris, Informatics and Telematics Institute, Greece
Milind R. Naphade, IBM T J Watson Research Center, USA

Special Sessions Chair

Murat Tekalp, College of Engineering, Koc University, Turkey

Industry Liasons

Adrian Matellanes, Motorola Labs, UK
Nikos Sarris, Athens Technology Center S.A., Greece

Webmaster

Stefanos Vrochidis, Informatics and Telematics Institute, Greece

GENERAL CHAIRS' MESSAGE

Welcome to the island of Santorini for the 8th International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS 2007.

WIAMIS is one of the main international fora for the presentation and discussion of the latest technological advances in interactive multimedia services. The objective of the workshop is to bring together researchers and developers from academia and industry working in all areas of image, video and audio applications, with a special focus on analysis.

After Louvain (1997), Berlin (1999), Tampere (2001), London (2003), Lisboa (2004), Montreux (2005), and Incheon, Korea (2006), WIAMIS 2007 is held in Santorini, Greece. The world famous volcanic island of Santorini is the most southern island of the Cyclades in the Aegean Sea. It is located 120 nautical miles south of Piraeus (the harbor of Athens) and 63 nautical miles north of Crete. With a breathtaking view of the caldera and the volcano, and its excellent conference facilities, Santorini provides an excellent setting for three days of exciting presentations and stimulating discussions on the most recent advances in multimedia services and analysis.

The technical program we put together for WIAMIS 2007 includes 84 accepted papers, organized in 6 regular oral sessions, one special session organized by the COST 292 action, and 2 poster sessions. It also includes two very exciting invited talks:

- *Recent Advances and Open Issues of Digital Image/Video Search*, by Prof. Shih-Fu Chang, Digital Video and Multimedia Lab, Department of Electrical Engineering, Columbia University
- *Data-driven and Procedural Analysis and Synthesis of Multimedia*, by Prof. Irfan Essa, Interactive Computing, College of Computing, Georgia Institute of Technology

Furthermore, to stimulate interaction between the workshop participants and strengthen the link between academia and industry, we have included in the workshop program an industry session, featuring talks by key industry players, and a panel, organized with the support of the CHORUS coordination action.

As General Chairs, we would like to thank all of those who have contributed to the success of WIAMIS 2007: the authors, the invited speakers and panel members, the Session Chairs, the Technical Program Committee and the additional reviewers, the Organizing Committee, and finally our sponsors: the COST 292 action, the aceMedia Integrated Project, the K-Space Network of Excellence, the MUSCLE Network of Excellence, the European Association for Signal Image Processing (EURASIP), the Institution of Engineering and Technology (IET), and the MPEG Industry Forum.

On behalf of the WIAMIS 2007 organizers, we hope that our carefully selected technical program, accompanied with the social events we have arranged for, will make your participation to WIAMIS 2007 worthwhile. We warmly welcome you and hope you will enjoy this workshop in Santorini.

WIAMIS 2007 General Chairs

Yiannis Kompatsiaris
Yannis Avrithis

INVITED TALKS

Wednesday, 9:30-10:30

TALK: "Data-driven and Procedural Analysis and Synthesis of Multimedia", Irfan Essa, Interactive Computing, College of Computing (RIM@GT & GVU Center), Georgia Institute of Technology

In this talk, I will outline the changes that have come about in the analysis and synthesis of multimedia, due to the availability of large amounts of data. I will present several of the recently successful methods that have been introduced in the last few years for example-based synthesis for animation and rendering of videos. I will also show how these methods have been extended to other modalities. I will also show how these approaches need to be extended by developing parametric and procedurals models to represent temporal variations. Using example from my groups work and also other efforts, I will discuss how video is becoming an accessible medium for all and I will also discuss some newer work on authoring of multimedia content.

Thursday, 9:00-10:00

TALK: "Recent Advances and Open Issues of Digital Image/Video Search", Shih-Fu Chang, Digital Video and Multimedia Lab, Department of Electrical Engineering, Columbia University

With the prevalent success of Internet search, researchers are facing new opportunities and challenges - developing next-generation multimedia search technologies that may reach a performance level similar to that of text search. Despite the grand scale of the challenges, some promising grounds have been revealed recently. In this talk, I will focus on two exciting areas - semantic annotation and multimedia document ranking. For the former, we are witnessing the significant developments in large-scale image/video collection for benchmarking, multimedia lexicons, and a large number of semantic classifiers. For example, we have developed classifiers for 374 semantic concepts with encouraging performance using more than 160 hours of digital videos in TRECVID 2006. The collective power of a large number of semantic models offers great potential - I will share recent results of this approach in video searching, topic threading, and temporal pattern mining. For the second area, I will present recent efforts to model video retrieval as a document ranking problem similar to that used for page ranking of web search. I will discuss how the semantic models and visual duplicate information may be used to approximate the information required in constructing document link graphs. I will conclude the talk with discussions of open issues in this dynamic and exciting area.

TECHNICAL PROGRAM

Wednesday, 11:00-12:00

Analysis I (Oral Session)

Title	Authors	Abstract
Composite Object Detection in Video Sequences: Application to Controlled Environments	Xavier Giro, Ferran Marques	This paper presents a set of techniques for the detection of composite objects in video recordings of a controlled environment. Firstly, a selective region-based analysis is performed by tuning the algorithm to the perceptual characteristics of the object in the environment. Secondly, the controlled perceptual and semantic variabilities of the object are addressed by the detection analysis thanks to a frame by frame update of the object models, and by allowing multiple models for a single object. The proposed techniques are illustrated in the detection of laptops from a zenithal view in a smart room.
Using Decision Trees for Knowledge-Assisted Topologically Structured Data Analysis	C. Simon, J. Meessen, D. Tzovaras, C. De Vleeschouwer	Supervised learning of an ensemble of randomized trees is considered to recognize classes of events in topologically structured data (e.g. images or time series). We are primarily interested in classification problems that are characterized by severe scarcity of the training samples. The main idea of our paper consists in favoring the selection of attributes that are known to efficiently discriminate the minority class in those nodes of the tree that are close to the leaves and where classes are represented by a small number of training examples. In practice, the knowledge about the ability of an attribute to discriminate the classes represented in a particular node is either provided by an expert or inferred based on a pre-analysis of the entire initial training set. The experimental validation of our approach considers sign language and human behavior recognition. It reveals that the proposed knowledge-assisted tree induction mechanism efficiently compensates for the shortage of the training samples, and significantly improves the tree classifier accuracy in such scenarios.
Event Alignment for Cross-Media Feature Extraction in the Football Domain	Jan Nemrava, Paul Buitelaar, Vojtech Svatek, Thierry Declerck	This paper describes an experiment in creating cross-media descriptors from football-related text and videos. We used video analysis results and combined them with several textual resources – both semi-structured (tabular match reports) and unstructured (textual minute-by-minute match reports). Our aim was to discover the relations among six video data detectors and their behavior during a time window that corresponds to an event described in the textual data. The experiment shows how football events extracted from text can be mapped to corresponding scenes in video and how this may help in extracting event-specific video detectors.

Wednesday, 12:30-13:30

Analysis II (Oral Session)

Title	Authors	Abstract
Surveillance Event Interpretation Using Generalized Stochastic Petri Nets	Artyom Borzin, Ehud Rivlin, Michael Rudzsky	In this paper we present video event representation and recognition approaches that are based on Generalized Stochastic Petri Nets (GSPN). Along with the typical modelling capabilities of GSPN for video recognition, we propose to integrate the Petri net marking analysis for better scene understanding. This work focuses on behavior modelling and uses the results of an external module for object detection, tracking and classification. The proposed approach is evaluated using the developed surveillance system which can recognize events from videos and give a textual expression for the detected behavior. The experimental results illustrate the ability of the system to create complex spatiotemporal relations and to recognize the behavior of one or multiple objects in various video scenes.
A Framework for Ontology Enriched Semantic Annotation of CCTV Video	Bogdan Vrusias, Dimitrios Makris, John-Paul Renno, Neil Newbold, Khurshid Ahmad, Graeme Jones	This paper deals with the problem of semantic transcoding of CCTV video footage. A framework is proposed that combines Computer Vision algorithms that extract visual semantics, together with Natural Language Processing that automatically builds the domain ontology from unstructured text annotations. The final aim is a system that will link the visual and text semantics in order to routinely annotate video sequences with the appropriate keywords of the domain experts' terminology.
Hidden Markov Models for Video Skim Generation	Sergio Benini, Pierangelo Migliorati, Riccardo Leonardi	In this paper we present a statistical framework based on hidden Markov models (HMMs) for video skimming. A chain of HMMs is used to model subsequent story units: HMM states represent different visual-concepts, transitions model the temporal dependencies in each story unit, and stochastic observations are given by single shots. The skim is generated as an observation sequence, where, in order to privilege more informative segments for entering the skim, dynamic shots are assigned higher probability of observation. The effectiveness of the method is demonstrated on a video set from different kinds of programmes, and results are evaluated in terms of metrics that assess the content representational value of the obtained video skims.

Wednesday, 14:50-16:10

COST 292 special session (Oral Session)

Title	Authors	Abstract
A Topology Preserving Approach for Image Classification	Le Dong, Ebroul Izquierdo	In this paper, an approach for image analysis and classification is presented. It is based on a topology preserving approach to automatically create a relevance map from salient areas in natural images. It also derives a set of well-structured representations from low-level description to drive the final classification. The backbone of this approach is a distribution mapping strategy involving two basic modules: structured low-level feature extraction using convolution neural network and a topology preservation module based on a growing neural gas network. Classification is achieved by simulating the high-level top-down visual information perception in primates followed by incremental Bayesian parameter estimation. The proposed modular system architecture offers straightforward expansion to include user relevance feedback, contextual input, and multimodal information if available.
Orientation histogram-based matching for Region Tracking	David Marimon, Touradj Ebrahimi	A region tracking technique with particular emphasis on rotation robustness is presented. It is based on region matching divided in two consecutive steps, gradient orientation histogram matching and template matching through Normalised Cross Correlation (NCC). Given the orientation histograms of two image patches, a novel technique is used to estimate the rotation between them together with the similarity. This estimation enhances the performance and speeds-up the process of patch recognition. Fast computation of histograms using the integral histogram approach [9] is exploited. Experiments show a high accuracy in the estimation of location and orientation.
A Directional Texture Descriptor via 2D Walking Ant Histogram	Serkan Kiranyaz, Miguel Ferreira, Moncef Gabbouj	A novel texture descriptor, which can be extracted from the major object edges automatically and used for the content-based retrieval in multimedia databases, is presented. The proposed method is adopted from the 2D Walking Ant Histogram, which is in fact a generic shape descriptor recently developed for general purpose multimedia databases. 2D WAH shape descriptor is motivated from the imaginary scenario of a walking ant with a limited line of sight over the boundary of a particular object; eventually each sub-segment is traversed and the process keeps describing a certain line of sight, whether it is a continuous branch or a corner, using individual 2D histograms. In this paper we tuned this approach as an efficient texture descriptor, which achieves a superior performance especially for directional textures. Integrating the whole process as feature extraction module into MUVIS framework allows us to test the mutual performance of the proposed texture descriptor in the context of multimedia indexing and retrieval.
A Review of Different Object Recognition Methods for the Application in Driver Assistance Systems	Andreas Laika, Walter Stechele	Algorithms in the field of driver assistance have been limited by their requirement for real time in the initial phase of their development. However, as computing power is increasing steadily, new possibilities arise. With focus on this situation a review is presented not defined by its designated field of application in driver assistance systems, but rather by the methods in use, namely video-based object recognition using machine learning. Recent methods are compared in a highly summarized table using criteria such as recognition rate, computational requirements or number of training samples required. Concluding their potential use in driver assistance is discussed.

Wednesday, 16:40-18:00

COST 292 special session (Oral Session)

Title	Authors	Abstract
Automatic Detection and Classification of Traffic Signs	Carlos Filipe Paulo, Paulo Lobato Correia	This paper proposes algorithms for the automatic detection of traffic signs from photo or video images and their classification to provide a driver alert system. Several examples taken from Portuguese roads are used to demonstrate the effectiveness of the proposed system. Traffic signs are detected by analyzing color information, notably red and blue, contained on the images. The detected signs are then classified according to their shape characteristics, as triangular, squared and circular shapes. Combining color and shape information, traffic signs are classified into one of the following classes: danger, information, obligation or prohibition. Both the detection and classification algorithms include innovative components to improve the overall system performance.
Using Multiple Domain Visual Context in Image Analysis	Phivos Mylonas, Yannis Avrithis	In this paper we propose an algorithm to improve the results of knowledge-assisted image analysis, based on contextual information. In order to achieve this, we utilize fuzzy algebra, fuzzy sets and relations, towards efficient manipulation of image region concepts. We provide a novel context modelling, based on the OWL language, using RDF reification. Initial image analysis results are enhanced by the utilization of domain-independent, semantic knowledge in terms of concepts and relations between them. The novelty of the presented work is the context-driven re-adjustment of the degrees of confidence of the detected concepts produced by any image analysis technique, utilizing a domain-independent ontology infrastructure to handle the knowledge, as well as multiple application domains.
Discriminative Feature Selection for Applause Sounds Detection	Roman Jarina, Jan Olajec	The specific sounds such as applause, laughter, explosions, etc. are very helpful to understand high level semantic of audio/video content. The paper focuses on feature selection by evolutionary programming for an automatic detection of applause in audio stream. A set of the most discriminative features is selected by Genetic Algorithm and Simulated Annealing. The experiments are run on more than 9 hours of audio selected from various audio and video content. The results show that the applause sound recognition improves if only a few coefficients are selected from MFCC static and dynamic features. Further, the delta-delta coefficients (the 2nd time derivatives of

		MFCCs) highly outperform the delta coefficients.
Global Image Search vs. Regional Search in CBIR Systems	Stevan Rudinac, Marija Uscumlic, Maja Rudinac, Goran Zajic, Branimir Reljin	The global image search and regional image search are compared by using content-based image retrieval system with user relevance feedback. It was expectable that regional search can minimize the effect of the background to the image retrieval. Images from database are partitioned into regular rectangular regions: 4x4 non-overlapped (NOV) regions and 3x3 overlapped (OV) regions, and a feature vectors are determined for whole images and for regions. Four CBIR scenarios are considered: global search, search based on 4x4 NOV regions, based on 3x3 OV regions and based on arbitrary cropped part of a query image. System is tested over images from Corel 1K dataset.

Thursday, 10:30-11:30

Tracking (Oral Session)

Title	Authors	Abstract
Regression-Based Template Tracking in Presence of Occlusions	Ioannis Patras, Edwin Hancock	This paper addresses the problem of efficient visual 2D template tracking in the presence of large motions and partial occlusions. We adopt a learning approach, in our case using a Bayesian Mixture of Experts (BME), in which observations at each frame yield direct predictions of the state (e.g. position / scale) of the tracked target. In contrast to other methods in the literature, we explicitly address the problem that the prediction accuracy can deteriorate drastically for observations that are not similar to the ones in the training set; such observations are common in case of partial occlusions or of fast motion. To do so, we couple the BME with a probabilistic kernel-based classifier which, when trained, can determine the probability that a new/unseen observation can accurately predict the state of the target (the 'relevance' of the observation in question). In addition, in the particle filtering framework, we derive a recursive scheme for maintaining an approximation of the posterior probability of the target's state in which the probabilistic predictions of multiple observations are moderated by their corresponding relevance. We apply the algorithm in the problem of 2D template tracking and demonstrate that the proposed scheme outperforms classical methods for discriminative tracking in case of motions large in magnitude and of partial occlusions.
kNN-based high-dimensional Kullback-Leibler distance for tracking	Sylvain Boltz, Eric Debreuve, Michel Barlaud	This paper deals with region-of-interest (ROI) tracking in video sequences. The goal is to determine in successive frames the region which best matches, in terms of a similarity measure, a ROI defined in a reference frame. Two aspects of such a measure between the reference region and a candidate region can be distinguished: radiometry which indicates if the regions have similar colors and geometry which correlates where these colors are present in the regions. If not using geometry, the number of potential matches increases. A soft geometric constraint can be added in the form of a joint radiometric-geometric PDF. High-dimensional PDF estimation being a difficult problem, measures based on these PDF distances may lead to an incorrect match. Instead, we propose to compute the Kullback-Leibler distance between high-dimensional PDFs without explicit estimation of the PDFs, i.e., directly from the samples using the kth-nearest neighbor (kNN) framework. Results showed accurate tracking.
Probabilistic Matching Algorithm for Keypoint Based Object Tracking Using a Delaunay Triangulation	Trichet Remi, Merialdo Bernard	This article presents a matching algorithm developed for a generic object tracking system. Matching is a critical part for the effectiveness of tracking. The proposed method is a probabilistic algorithm inspired from the emerging "discriminative random fields". Points are associated according to their visual similarity and to spatial relations in their neighborhood, based on a Delaunay triangulation. Experimental results are presented to validate this contribution.

Thursday, 12:00-13:20

Analysis III (Oral Session)

Title	Authors	Abstract
Using Graphics Processor Units (GPUs) for Automatic Video Structuring	Peter Kehoe, Alan F. Smeaton	The rapid pace of development of Graphic Processor Units (GPUs) in recent years in terms of performance and programmability has attracted the attention of those seeking to leverage alternative architectures for better performance than that which commodity CPUs can provide. In this paper, the potential of the GPU in automatically structuring video is examined, specifically in shot boundary detection and representative keyframe selection techniques. We first introduce the programming model of the GPU and outline the implementation of techniques for shot boundary detection and representative keyframe selection on both the CPU and GPU, using histogram comparisons. We compare the approaches and present performance results for both the CPU and GPU. Overall these results demonstrate the significant potential for the GPU in this domain.
Fast and Robust Graphic Character Verification System for TV Sets	Antonio Rama Calvo, Ramon Alujas Tejada, Francesc Tarres Ruiz	In this paper, we present a flexible system to verify any kind of graphic TV character or symbol in television charts. The proposed approach is very robust towards translations, rotations, changes in scale and illumination variations of the graphic characters. Another important attribute of the proposed system is its speed. The system should be very fast since the characters have to be verified in a real production line of TV sets. The proposed system is composed of fast and accurate preprocessing modules in order to normalize the characters, and a PCA-based verification method. Results show a very high performance with verification rates over the 99.8%.
Bayesian Region Growing and MRF-based Minimization for Texture and Colour Segmentation	Ilias Grinias, Nikos Komodakis, Georgios Tziritas	We propose a generic, unsupervised feature classification and image segmentation framework, where only the number of classes is assumed as known. Image segmentation is treated as an optimization problem. The framework involves block-based unsupervised clustering using k-means, followed by region growing in spatial domain. High confidence statistical criteria are used to compute a map of initial labelled pixels. A new region growing algorithm is introduced, which is

		named Independent Flooding Algorithm and computes a height per label for each one of the unlabeled pixels, using Bayesian dissimilarity criteria. Finally, a MRF model is used to incorporate the local pixel interactions of label heights and a graph cuts algorithm performs the final labelling by minimizing the underlying energy. Segmentation results using texture, intensity and color features are presented.
Image Denoising Method Using Diffusion Equation and Edge Map Estimated with K-Means Clustering Algorithm	Woong Hee Kim, Thomas Sikora	The purpose of noise filtering is to preserve features such as edge or corners, and to reduce noise. However, it is difficult to preserve them after noise filtering, because most of noise filtering algorithms have the characteristic of low pass filter, which results in the loss of signal components with high frequencies like edge or corners. Noise filtering methods using diffusion equation try to solve this problem, and improve it to some extent. Edge detector has been used in filtering methods with diffusion equation to determine the speed of diffusion term, so the performance of them depends on the edge detection algorithm to some extent. Previous noise filtering approaches of diffusion equation use a simple gradient edge detector to determine the speed of diffusion. Also, each channel of color images are filtered separately, which result in the filtered image with distortion or blurring. In this paper, we propose a new noise filtering method for color images using diffusion equation and edge map estimated with K-means clustering algorithm together.

Thursday, 14:50-16:20

Poster session I

Title	Authors	Abstract
A New Method for Infrared Weak Small Target Enhancement Based on Image Fusion	Lei Tang, Zong-gui Zhao	A new method for infrared weak small target enhancement based on image fusion was presented. The structure of this method is two times fusion. In first fusion stage, the simple image fusion method is used to fuse the continuous frames. In second fusion stage, the multiscale decomposition method is used to decompose the result from first fusion stage and then fuse them. The fusion effect with different multiscale decomposition methods such as the Wavelet transform, the Contourlet transform and the Wavelet-based Contourlet transform (WBCT) is contrasted. The experimental results show that the method is effective and can enhance infrared weak small targets.
Visual Data Processing and Action Control Using Binary Neural Networks	A.V. Kazantsev	A new model of the brain-like neural network for visual data processing and action control is proposed. The neural network is built on discrete elements with binary input and output with memory cells. Optimization of the network is conducted using a natural selection process within the framework of an artificial life paradigm. The theoretical principles of the neuron and network structure construction have been tested and assured by real experiment using a computer program which models a population of virtual bacteria living and evolving in a restricted 2D area. Virtual bacteria act using binary visual information as input. Given the rules of survival and neural network mutation, new generations of bacteria form their brain using the neural networks of their successful predecessors. The proposed approach demonstrates the possibility of constructing a brain-like neural network based only on binary data processing.
Using Condensation Tracking to Recover Stroke Order of Chinese Calligraphic Handwritings with CCM	Ying-fei Wu, Yue-ting Zhuang, Yun-he Pan, Fei Wu	Stroke order is an intrinsic writing feature which is missing in off-line handwriting. Recovering stroke order from off-line handwriting is an important task by which the visualization of writing process is possible. Chinese Calligraphic handwriting is some different with western handwriting. Traditional method with assumption of the minimization of curvature between two adjacent line segments is not applicable for Chinese Calligraphic handwriting. We propose a new framework named Chinese Calligraphic Model (CCM). With this framework CONDENSATION algorithm - conditional density propagation over time is performed to find the best matching model in model database with the input pattern to recover stroke order. Experiments results show our method is a promising way to recover stroke order from Chinese Calligraphic handwriting based on CCM by CONDENSATION tracking.
Spatio-temporal constraints for matching view-based descriptions of 3D objects	E. Delponte, N. Noceti, F. Odone, A. Verri	We propose a 3D object recognition method based on first extracting a compact description of image sequences, and then matching these descriptions with a two-steps strategy. The compactness of the description, made of a set of time-invariant local features, allows us to use a simple nearest neighbour matching to obtain an initial set of recognition hypothesis; spatio-temporal constraints help us to confirm or to reject these hypotheses. We carried out an extensive experimental analysis to assess our method against changes in illumination, abrupt scale changes, occlusions, clutter, view-point variations. Since our approach is based on the use of image sequences we also tested it with different camera motions. Our description strategy may be also applied to relatively simple poor-textured objects, since few stable features are enough for recognition.
Accurate Breast region Detection in Digital Mammograms Using a Local Adaptive Thresholding Method	Maysam Shahedi B K, Rassoul Amirfattahi, Farah Torkamani Azar, Saeed Sadri	Recently Computer Aided Diagnosis (CAD) systems have improved diagnosis of abnormalities in mammogram images. To improve the accuracy and reliability of such systems, the exact breast region as the region of interest (ROI), has to be segmented. Furthermore, focus on ROI can eliminates the effect of image background noise. Consequently it can reduce the detection algorithms execution time. Also, the accurate breast region detection ends to the accurate breast border detection, which can improve clinical diagnosis of abnormalities. Currently, new methods have been presented that achieve the accurate breast border detection. Nevertheless most of these methods are complicated and have undesirable effects on the execution time. In this paper we proposed a novel approach for accurate breast region segmentation in digital mammograms based on local thresholds. The suggested method can extract the breast region accurately.
Recognition of Botanical Bloom Characteristics from Visual Features	Margit Lang, Harald Kosch, Stella Stars, Carsten Kettner,	This paper introduces multimedia indexing and retrieval techniques to enable the extraction of botanical characteristics for semi-automatic plant recognition. The identification of plants requires recognition and determination of plant species by using specific features such as bloom color, inflorescences, shape of blooms, number of petals and shape of leaves. Our indexing and retrieval

	Janine Lachner, Doris Oborny	techniques will be integrated into an existing medicinal plant database, MedPhyt®. An evaluation study finishes the paper.
3D Motion Recognition based on Ensemble Learning	HongLi Zhu, PengYing Du, Jian Xiang	In this paper, a novel method is presented for 3D motion recognition based on Motion Capture Database. We use 3D features and their keyspaces of each human joint to represent human motion. After features extraction, Ensemble HMM learners are used to train data. Then each action class is learned with one HMM and bagging algorithm is used to ensemble all learners. Since ensemble learning can effectively enhance supervised learners, ensembles of weak HMM learners are built. It is obvious that the proposed methods are effective by experimental results.
Equivalent Key Frames Selection Based on Iso-Content Distance and Iso-Distortion Principles	Costas Panagiotakis, Anastasios Doulamis, Georgios Tziritas	We present a key frames selection algorithm, which is very flexible on any changes of content descriptors, based on Iso-Content Distance and Iso-Distortion principles. In both of the cases, the equality principle provides to the selected key frames the property to be equivalent on content video summarization. The estimated key frames properties and the experimental results indicate the good performance of the proposed schemata.
A Semantic Event Detection Approach for Soccer Video Based on Perception Concepts and Finite State Machines	Liang Bai, Songyang Lao, Weiming Zhang, Gareth J.F. Jones, Alan F. Smeaton	A significant application area for automated video analysis technology is the generation of personalized highlights of sports events. Sports games are always composed of a range of significant events. Automatically detecting these events in a sports video can enable users to interactively select their own highlights. In this paper we propose a semantic event detection approach based on Perception Concepts and Finite State Machines to automatically detect significant events within soccer video. Firstly we define a Perception Concept set for soccer videos based on identifiable feature elements within a soccer video. Secondly we design PC-FSM models to describe semantic events in soccer videos. A particular strength of this approach is that users are able to design their own semantic events and transfer event detection into graph matching. Experimental results based on recorded soccer broadcasts are used to illustrate the potential of this approach.
Voice Query Transcription and Expansion Scheme for Efficient Music Retrieval	Seungmin Rho, Byeong-jun Han, Eenjun Hwang, Minkoo Kim	In this paper, we present a scheme for efficient humming-based music retrieval. For that purpose, we first describe how to extract a sequence of pitch and duration pairs as its feature information from sung or hummed query accurately and automatically. And then, we propose a novel scheme for reformulating user query to improve retrieval performance. The scheme is based on user relevance feedback with genetic algorithm. We implemented a prototype system based on these scheme and performed various experiments. Experimental result shows that our proposed scheme achieves an impressive performance.
Motorcycle Detection and Tracking System with Occlusion Segmentation	Chung-Cheng Chiu, Min-Yu Ku, Hung-Tsung Chen	This paper proposes a vision-based motorcycle monitoring system to detect and tracking motorcycles. The system proposes an occlusion detection and segmentation method. The method uses the visual length, visual width, and Pixel Ration to detect the classes of the motorcycle occlusions and segment the motorcycle from each occlusive class. Because the motorcycle riders must put on their helmets, the helmet detection or search method is used to make sure whether the helmet/motorcycle exits or not. Experiments obtained by using complex road scenes are reported, which demonstrate the validity of the method in terms of robustness, accuracy, and time responses.
A New Content-based Image Retrieval Using Color Correlogram and Inner Product Metric	Thurdsak Leauhatong, Kiyooki Atsuta, Shozo Kondo	The color correlogram is a simple statistical descriptor of a color image that has been widely used for content-based image retrieval (CBIR) systems. To measure similarity between two images using the correlogram, the traditional approaches use the relative distance. In this paper, to improve performance of the CBIR systems, the inner product metric is used to measure similarity of images instead of the relative distance. Results of experiments proved that the CBIR using the inner product metric has better performance than the one using the relative distance.
A Visual Tracking Algorithm for Real Time People Detection	T. D' Orazio, M. Leo, P. Spagnolo, P.L. Mazzeo, N. Mosca, M. Nitti	In this paper we present a multi-people-tracking algorithm which is able to detect and track humans in complex situations with varying light conditions, high frame rate, and real time processing. We propose a stochastic approach for foreground people tracking based on the evaluation of the maximum a posteriori probability (MAP). The algorithm evaluates geometrical information on the blob overlapping and does not require the feature extraction to track the single object. Experimental tests have been carried out on soccer image sequence in which some players enter into the camera view and remain for some time.
Motion-based Object Segmentation using Sprites and Anisotropic Diffusion	Andreas Krutz, Matthias Kunter, Thomas Sikora, Mrinal Mandal, Michael Frater	Many algorithms have been developed to recognize regions, edges, color, and objects in images and videos. For applications like surveillance or object-based video coding, it is important to segment the foreground objects from the background. The task is very challenging in the case of a moving camera. We present a foreground segmentation approach that is designed for sprite coding as well as other applications, e.g. video surveillance. Accurate frame-to-frame image registration and sprite generation build the pre-processing step. The segmentation algorithm operates on error images, which are produced by the image registration and subtraction from reconstructed background frames. It is processed in several steps including low-pass filtering using anisotropic diffusion. Experiments show excellent results with single- and multi-view test sequences.
Learner's Activity During the Understanding of an Interactive Animated Mechanical System: Eye Tracking Investigation	Emmanuel Schneider, Jean-Michel Boucheix	In two experiments we examined how learners understand an animation about a mechanical complex three pulleys system. In the first experiment were tested the animation user-control and the orientation of the attention of the learners. The second experiment tested the potential effect of adding signaling cues like arrows on the comprehension of the mechanical device. In the two experimental designs eye tracking was registered before a comprehension test. The first experiment showed a positive effect of a full controllable animation and also a verbal explicit orientation of attention on the understanding of the pulley system. The eye tracking data indicated that the learners process strongly the areas of the animations where a great amount of motion is involved along the casual chain of events. The second experiment did not show any effect of adding implicit signaling cues.
Tracking based on local motion estimation	Vincent Garcia, Eric Debreuve,	The extraction of a video object contour, called "rotoscoping" in cinematographic post-production, is usually performed manually and frame by frame. Semi-automatic algorithms have been proposed to

of spatio-temporally weighted salient points	Michel Barlaud	reduce the load of this task. However, they classically use region information and are usually based on a notion of homogeneity of the object. This homogeneity description might be difficult to establish and, consequently, the tracking may be not precise enough. The proposed method relies on the analysis of some temporal trajectories of salient points, or keypoints, called tracks. The main contribution of this paper is the local estimation, both spatially and temporally, of the contour motion from these tracks. The proposed method seems accurate, robust to outliers, and allows local deformation. Moreover, it can deal with partial occlusions.
Recognition of Hand Raising Gestures for a remote Learning Application	Bill Kapralos, Andrew Hague, Hamed Sabri	A central technical issue in developing synchronous distance learning technology is enabling the remote class and the instructor to interact with each other. Issues such as "how does a student capture the instructor's attention?", "how can the instructor select one student to converse with?", and "how can the instructor attend to the student once (s)he has been selected?" are complex problems that must be addressed if the class and instructor are to interact in an effective manner. This paper describes the use of Hidden Markov Models for the recognition of students signaling their intent to interact with the instructor using "traditional" classroom hand gestures such as raising and waving hand motions. Hand raising gestures are detected using motion cues over a sequence of omni-directional images using a set of pre-defined Hidden Markov Models.
Scene Identification Using Invariant Radial Feature Descriptors	Laura Worthy, Eric Sinzinger	This paper addresses the challenge of identifying and retrieving related scenes from image databases with a focus on low-level feature descriptor construction. A set of affine covariant regions are identified via a radial segmentation algorithm. Local descriptors are built using two different types of histograms: (i) Polar Image Gradient (PIG) orientation histogram, and (ii) saturation-weighted hue histogram. The combination of geometric and photometric information yields a significant improvement in a feature's discriminative power. A cascading matching algorithm is used for feature matching and evaluation. To demonstrate the descriptor's image matching capabilities, a voting algorithm for similar scene retrieval is implemented utilizing results from the feature matches. Challenging images of buildings with inherent replicative feature regions due to common edificial texture are used to test the robustness and applicability of the radial-based methodology.
A Secure and Privacy Friendly 2D+3D Face Authentication System Robust under Pose and Illumination Variations	Filareti Tsalakanidou, Christos Dimitriadis, Sotiris Malassiotis	An end-to-end face authentication system integrating 2D color images and depth data is presented in this paper, based on a low-cost sensor capable of real-time acquisition of 3D images and associated color images. Depth data is used for robust face detection, localization and 3D pose estimation, as well as for compensating pose and illumination variations of facial images prior to classification. The performance of the proposed system is tested on an extended face database recorded in real-world conditions, while a complete security and privacy analysis is conducted in order to ensure that all necessary countermeasures are built into the system.
Vision-based human-computer interface using hand gestures	Janez Zaletelj, Jernej Perhac, Jurij F. Tasic	The proliferation of interactive digital TV applications will depend on their complexity and ease of use. The investigation of simple, intuitive human-machine interfaces is therefore crucial for their success. While the operation of most home devices is today based on using remote control, in this paper we investigate an alternative solution using a visual input of the user's gestures. The proposed system recognises point and click gestures of the user's hand, which is enough to control simple iconic user interfaces of the interactive TV applications. Position of the user's hand is tracked using Kalman filtering, and shape analysis is performed to detect the pose of the forefinger. A finger bending gesture is recognized and transformed into a click action on the screen. Simulations show that the system performs well in uncontrolled home environment under different lighting conditions and cluttered backgrounds and thus could be used in future smart home devices.
Camera Motion Analysis in On-line MPEG Sequences	Fabricio Tiburzi, Jesus Bescos	This work describes a two stages algorithm for camera motion analysis. First, it detects frames defining changes in the camera motion pattern, and then characterises the motion pattern for each video segment. The proposed scheme is motivated by hierarchical content-aware adaptation of on-line video. Its main characteristics are on-line operation, direct analysis on the compressed domain, efficiency and robustness. Results are similar to those achieved by state-of-the-art camera motion classification schemes based on iterative fitting to a motion model, despite involving a significantly smaller number of operations.
Semantic Mapping of Image Databases using Perceptual Similarity	Christos Theoharatos, George Economou, Spiros Fotopoulos	Semantic mapping of image databases is explored based on a human-centered approach for judging visual similarity. The management is pursued using a standard psychophysical experiment followed by a well-suited data analysis methodology. The end-result is a cognitive discriminative biplot, which is a visualization of the intrinsic organization of the image database best reflecting the user's perception. The discriminating power of the introduced biplot constitutes it an appealing tool for image retrieval and a flexible interface for visual data mining tasks.
Image Retrieval based on Wavelet Computation and Neural Network Classification	A. C. Gonzalez-Garcia, J. H. Sossa-Azuela, E. M. Felipe-Riveron	We describe a simple way to retrieve images from a database. During training a wavelet-based description of the histogram of circular window of each image is first obtained using Daubechies 4-wavelet transformation. Resulting coefficients are used to train a neural network (NN). For image retrieval an image is presented to the system. The system responds with the most similar images. Results are given two different databases. With first database a 98.44% of efficiency with training images was obtained; a 50 % was obtained with images different from those used for training. With the second database corresponding efficiencies were of 88% and 64%.
Cartoon Detection Using Fuzzy Integral	Antonio Rama, Francesc Tarres, Laura Sanchez	With the growth of digital television TV program classification has become a major research topic. Recent classification techniques have reported acceptable results for specific genre detection. Cartoon is one of these genres which has deceived some attention because of its importance in push scenarios where parents want to control their children's access to television. In this paper a flexible scheme based on a non-linear classifier called Fuzzy Integral is presented. This operator is supposed not only to classify but also to give a relevance measure to all the features involved in the classification. Preliminary results using this operator for cartoon detection are presented and compared with other well known statistical classification methods like PCA, LDA or K-NN.
Edge Pixel Histograms	Antonio M. G.	Edge Histograms are extensively used as an image descriptor for image retrieval and recognition

Characterization with Neural Networks for an Improved Semantic Description	Pinehiro	applications. Edges represent textures and are also representative of the shapes in an image. In this work a histogram of the pixel edge directions is defined for image description. The edges detected with the Canny algorithm will be described in 4 directions. Images are divided into 16 sub-images, and a descriptor with 64 bins results. The descriptor ability for comparing images based in the Euclidean distance between histograms is going to be tested. Although the measure of the images similarity is important, it is also important to define new methods for high level description of images. The level of description can grow by defining image classes related with the image content. In this work, a neural network is used for the decision process of assigning each image to a set of defined image classes.
A framework for ontology-driven multimedia analysis and composition	Stamatia Dasiopoulou, Ioannis Kompatsiaris, Lyndon J.B. Nixon	The task of resource retrieval and composition is particularly challenging in the case of multimedia resources. Usual approaches focus on the use of textual multimedia annotations to enable computer-based management of multimedia repositories. In this paper, we introduce two approaches which make use of a logical formalism called an ontology to improve on the state of the art in multimedia retrieval and composition. Through the use of ontologies, these approaches form a part of the emerging Semantic Web, and offer important contributions to enabling multimedia-based Semantic Web applications.
Exact and Speedy Computation of Legendre Moments on Binary Images	G.A. Papakostas, E.G. Karakasis, D.E. Koulouriotis	A novel algorithm for the computation of Legendre moments, which ensures high accuracy and low computation time, is presented in this paper. The presented method exploits the compression capabilities, the image block representation exhibits. By describing the images with a set of homogenous blocks, keeping only the upper left and down right pixel coordinates, the long time consuming formulas used to compute Legendre moments, can be converted to more efficient ones. A typical comparison of the proposed method with the conventional one concludes to very impressive and promising results.

Thursday 16:20-17:40

Coding (Oral Session)

Title	Authors	Abstract
XML-driven Bit Rate Adaptation of SVC Bitstreams	Tom Paridaens, Davy De Schrijver, Wesley De Neve, Rik Van de Walle	Thanks to technological evolutions, the number of devices capable of playing video bitstreams is growing. The heterogeneity in these devices grows in terms of screen resolution, processing power, and available band width. In this paper, we describe an MPEG-21 Bitstream Syntax Description Language-based (BSDL-based) adaptation framework that allows providers to easily adapt scalable bitstreams without having to recode the original bitstream. We describe the steps necessary to adapt the bitstreams through BSDL. The main contribution of this paper is an optimized adaptation framework using a Bitstream Syntax Schema developed to minimize the size of the Bitstream Syntax Descriptions (BSDs). Furthermore, we created a Streaming Transformations for XML Stylesheet (STX-stylesheet) to exploit the advantages of Fine Grain Scalability, this to adapt the bitrate of Scalable Video Coding bitstreams in the most accurate way possible. Our results show that BSDL-based adaptation is able to compete with binary adaptation tools. The target bitrates can be reached within a margin of 2%, which is comparable to the reference software which uses binary adaptation.
Optimized Motion Estimation for Robust Video Coding in Packet Loss Environment	Fuzheng Yang, Shuai Wan, Ebroul Izquierdo	An improved motion estimation method for robust video coding is proposed. It is designed to enhance packet loss resilience in lossy environments. The proposed motion estimation is based on an end-to-end rate-distortion optimisation framework. Initially, the end-to-end reconstructed distortion is efficiently estimated considering the distortion caused by quantisation, error propagation, and error concealment. Then, the total bit rate for coding the residual is estimated using a quadratic rate-distortion model. The results are incorporated into the Lagrangian optimisation during motion estimation. Optimised motion vectors are selected in the sense of coding efficiency and robustness. A comparative evaluation with conventional robust video coding techniques was conducted. The experimental results demonstrate a superior performance of the proposed method.
Video coding based on distributed source coding	Cristian Perra, Daniele Giusto	Distributed video coding is a new paradigm for video compression. Practical distributed video coding are promising solutions for emerging applications. Nevertheless, more study and exploration is needed. This paper proposes an application of distributed source coding for video coding. While most of the solutions proposed in literature rely on an off-line model training the proposed solution is based on an on-line model training that exploits information from conventionally encoded frames.
Towards A Model Based Paradigm for Efficient Coding of Context Dependent Video Material	Sam J.C. Davies, Dimitris Agrafiotis, Nishan Canagarajah, David R. Bull	This paper proposes a generalised framework for model based, context dependent video coding, based on exploitation of characteristics of the human visual system. The system utilises variable quality coding, based on a map which is created using context dependent rules. The technique is demonstrated for a specific video context, namely open signed content. Consequently a model for gaze prediction in open signed content is developed, based upon motion and shot changes. The framework is shown to achieve a considerable improvement in coding efficiency for the given context.

Friday 14:00-15:30

Poster Session II

Title	Authors	Abstract
How to architect an IPTV system	Yuzo Iano, Marcus Pereira	Today's Internet Protocol Television (IPTV) is being offered and taking off, driving radical changes in the way telecom networks are designed. Future consumers will look for unique, personalized

	Magri	services through multiple devices, including the television. On demand content, Interactive TV, Video and voice communications, photo, music, and home video sharing, and online gaming, all delivered seamlessly, are a few examples of the services that consumers could receive. This requires innovative changes in the entire end-to-end service delivery architect, quality of service and customer management, reliability and security.
Channel Adaptive Multi-User Scalable Video Streaming with Unequal Erasure Protection	Hassan Mansour, Vikram Krishnamurthy, Panos Nasiopoulos	In this paper, we propose a solution to the problem of streaming multiple video sequences in a capacity constrained, error-prone network. This scenario arises in situations such as multiple mobile computers accessing streamed video through a single access point, or multiple cell phone users serviced by a single base-station. Our approach constitutes a rate-constrained distortion-minimized streaming scheme to allocate the appropriate rate for each video stream that achieves minimal distortion without violating the capacity constraint. Moreover, we developed an unequal erasure protection (UXP) scheme that minimizes the protected bitstream rate and the expected packet loss probability for each video frame. Performance evaluations show that this approach delivers an improved decoded video quality and fairer rate-allocation when compared with standard UXP and rate-allocation schemes, specifically in limited capacity and congested channels.
AVC Intraprediction Mode Decision Based on 4x4 Integer Transform Coefficients	Florian Obermeier, Marko Durkovic, Michael Zwick, Klaus Diepold	In order to achieve AVCs full potential in intracoding efficiency an encoder must choose an appropriate coding option for each block in every frame. This leads to a significant computational effort. In this paper, we investigate employing the AVC 4x4 integer transform in order to decide intraprediction size and mode for AVC encoding in an efficient manner. The transform is used to constrain the encoding options with a feasible overhead, as it can be implemented using shifts and additions only.
XML-based Exploitation of Region of Interest Scalability in Scalable Video Coding	Davy De Schrijver, Wesley De Neve, Davy Van Deursen, Yves Dhondt, Rik Van de Walle	The use of Regions Of Interest (ROIs) is a useful concept for many application scenarios, especially for those applications that are deployed in heterogeneous multimedia environments. In this paper, we show how Flexible Macroblock Ordering can be used in the scalable extension of the H.264/AVC specification in order to define the ROIs in the coded bitstream. Furthermore, we introduce an XML-driven adaptation framework based on the MPEG-21 Bitstream Syntax Description Language in order to implement the ROI extraction process. This framework gives us the opportunity to adapt scalable bitstreams by using an engine that has no knowledge of the underlying coding format. From the performance analysis of our adaptation framework, we can conclude that the ROIs can be extracted in the XML domain and that the ROIs in the adapted bitstream are still intact without quality degradation. Furthermore, the traditional drifting problem caused by the ROI extraction can be neglected. Finally, we show that the adaptation process in the XML domain can be executed in real time.
An Alternative Scattered Pattern for Flexible Macroblock Ordering in H.264/AVC	Yves Dhondt, Stefaan Mys, Saar De Zutter, Rik Van de Walle	Spatial error concealment methods in block-based video codecs use information from surrounding macroblocks to repair lost macroblocks. Using the slice group tool of H.264/AVC, it is possible to encode the macroblocks of a frame based on a scattered pattern, hereby allowing an encoder to actively help concealment methods at the decoder side. In this paper, the requirements for such a pattern are presented. Two commonly used patterns are verified against those requirements and an alternative scattered pattern is presented. Experiments show that our alternative pattern outperforms the old scattered patterns by up to 1.5 dB in case two or more slice groups are lost.
Online Collage CAPTCHA	Mohammad Shirali-Shahreza, Sajad Shirali-Shahreza	In some websites it is necessary to tell apart human users from computer programs which is known as CAPTCHA (Completely Automated Public Turing test to tell Computers and Human Apart). CAPTCHA methods are mainly based on the weak points of OCR systems while using them are undesirable to human users. In this paper a method has been presented for telling the human users and computer softwares apart on the basis of choice of an object shown on the screen. In this method some objects are chosen randomly and the pictures about these topics are downloaded from Internet. Then after effecting some changes such as rotating, all of the pictures are shown on the screen. Then we ask the user to choose a specific object. If the user chooses the right object we can guess that the user has been a human being and not a computer software. The main advantage of this method is its simplicity because the user does not have to type anything.
A Novel Geometrical Robust Image Data Hiding Scheme	Dan Wang, Peizhong Lu	Robustness to geometrical distortions is essential to the data hiding scheme. However, most of existing methods mainly focus on the geometrical invariant watermark, and data hiding techniques with high robustness to geometrical attacks is still a challenge. In this paper ¹ , we propose a novel geometrical invariant image data-hiding scheme based on fast correlation attacks of stream ciphers. The scheme can resynchronize the attacked image on the basis of the robustness of feature points and the powerful error-correcting characteristic of fast correlation attacks technique. The paper explicitly analyses the channel noise model, which provides a basis on which the parameters of fast correlation attacks is designed. Simulation results demonstrate that our data-hiding scheme is not only robust to geometrical distortions including RST and general affine transformations, but also robust to common signal processing such as JPEG compression.
Fast Protection of the Color of High Dimension Digital Painting Images	M. Chaumont, W. Puech	In this paper, we propose a method to embed the color information of an high dimension image in its corresponding grey-level image. The objective of this work is to allow a free and rapid access to the grey-level image and give color image access to secret key owners. This method is made of two steps which are the fast high-dimension color image decomposition (in a grey-level image and its associated color information) and the data-hiding. The two main contributions of this paper are the energy function proposed to model the decomposition of the color image and the fast optimization. The optimization of the proposed energy function leads to the achievement of an index image and a color palette. The good properties of that decomposition are an index image which is similar to the luminance of the color image and a color palette which is well suit for the data-hiding. The obtained results confirm the model quality and the real-time property.
Low Complexity Multiple Description Coding for H.264/AVC	Kenneth Vermeirsch, Yves Dhondt, Stefaan	Today's wireless networks suffer from high error rates in real-life conditions. Moreover errors tend to occur in large bursts. Multiple Description Coding (MDC) of digital video is a promising technique to overcome these problems in wireless environments by taking advantage of frequency or channel

	Mys, Rik Van de Walle	diversity. This paper proposes two low complexity MDC techniques for H.264/AVC at the Network Abstraction Layer level, aiming at conversational (video teleconferencing) and sports Video on Demand (VoD) application scenarios, respectively. For conversational applications, an MDC scheme based on H.264/AVC data partitioning is used, offering low additional delay and good performance for video with static backgrounds. For the sports VoD application, a scheme based on redundant slices is proposed. The latter scheme performs better for high-motion sequences than the former, but also introduces more delay.
An Improved Wavelet-based Watermarking Method Using the Mathematical Morphology	Kasemsuk Sepsirisuk, Kiyooki Atsuta, Shozo Kondo	In this paper, we propose an improved wavelet-based watermarking method using the mathematical morphology. The wavelet coefficients of host image are firstly selected as the significant coefficients by a statistical predetermined threshold. The mathematical operations are then performed to tame some small perturbations on the selected coefficients caused by the manipulations or attacks. Moreover, the pixel-wise masking model which exploits the characteristics of Human Visual System is employed for improving watermark invisibility and robustness. The experimental results verify that our method is more robust and effective than the existing method.
Viewer authentication for personalized iTV services	Tilen Mlakar, Janez Zaletej, Jurij F. Tasic	The proliferation of interactive digital TV services will depend on their complexity and ease of use. The investigation of simple, intuitive human-machine interfaces is therefore crucial for their success. Personalized iTV services require consumer authentication and consumer input, which is traditionally done by using the remote control. In this paper, we investigate alternative ways to collect consumer input based on computer vision methods. The goal of the proposed system is to enable consumer authentication and automatic collection of consumer feedback on the TV program. The image processing system on the STB continually monitors the TV viewer(s). The face detection is employed to detect and track the number of people watching the TV program. Face recognition is employed to identify specific TV viewers, which is necessary in order to enable personalized user interface and personalized services. Eye detection and gaze recognition is used to continually track the interest level of the specific viewer towards the TV content, which is important feedback information.
An XML-based Synchronization of Audio and Graphical Representations of Music Scores	Adriano Barate, Luca A. Ludovico	This paper presents an overview on a future IEEE standard aimed at providing an overall description of music. This format, known as IEEE PAR1599, is based on the XML meta-language. Its purpose is taking into account the heterogeneous multimedia representations of music, such as audio tracks, video clips and graphical instances of score. The key characteristics of the format are the possibility to enjoy such heterogeneous contents in a synchronized way and the possibility to switch from a particular representation to another in real-time. After a short description of the particular XML language we have adopted, a case study will be presented: an application installed at the exhibition "Celeste Aida" held at Teatro alla Scala in November 2006 - January 2007.
An Innocuous Visual Cryptography Scheme	Hyoung Joong Kim, Vasily Sachnev, Su-Jeong Choi, Shijun Xiang	An innocuous visual secret sharing scheme over natural images is presented in this paper. Secret sharing scheme allows a group of participants to share a secret (i.e., an image) among them. In case of (k, n) visual cryptography any group of k qualified participants among n (where $k \leq n$) can reconstruct the secret. This paper presents (n, n) visual cryptography scheme. This scheme does not apply dithering techniques to hide a secret image. Thus, it does not degrade quality of the secret image and in particular, this scheme is far from negative photo effect. Instead of dithering, this scheme spreads data by applying simple arithmetic operations. The rationale of arithmetic operations is presented.
Personal Video Coding for Very Low Bit Rate Applications	Cataldo Guaragnella	In personal video communications, the postures assumed by the speaker during the connection are very similar. In past works the possible exploitation of inherent voice and video correlation has been proposed for the generation of a likely animation of a speaker, driven by audio information only. In this paper, a novel video coding technique is presented, based on the decomposition of the sequence spanned hyperspace in orthogonal components. Good quality video can be reconstructed at the receiver, particularly for very low bit rate applications, with low residual error and smooth appearance both in motion reconstruction and in the complete absence of crisp coding artifacts such as blocking structures, basing on a few proper selected examples of the speaker and a multidimensional interpolation algorithm. Principal Component Analysis exploited by Singular Value Decomposition has been used to decompose the video sequence into eigen-images. A limited set of eigen-images is chosen to span the whole video sequence subspace. Only coefficients to be used in the interpolation procedure are transmitted along the communication channel. Preliminary results are presented for Akyio QCIF video sequence.
Media-Driven Dynamic Scene Adaptation	B. Pellán, C. Concolato	This paper presents a method to complement existing video, audio and image adaptation techniques in order to improve the perceived quality of interactive rich-media applications in streaming scenarios. This approach consists in dynamically adapting multimedia services at the presentation level according to media-level adaptation decisions. First, an enhanced decision-taking engine decides on the best presentation form based on author constraints. Then, presentation updates dynamically modify the user display according to fluctuating environment constraints. A fine-grain control of the synchronization of these updates with adapted media streams is used to enable seamless switching of media bitstreams with different coding (transcoding) or switching to media streams with different modality (transmoding). This paper finally describes an application test-bed using MPEG-4 and MPEG-21 to implement the proposed adaptation approach and illustrates the user experience improvement.
Applications of a Neural Network to Estimate Watermark Embedding Strength	Cong Jin, Shihui Wang	Digital watermarking technology has the advantages of easy implementation and capability of providing wide security services such as copyright protection, authentication and secret communication. In this paper, a digital watermarking scheme using neural network is proposed. Firstly, the original image is divided into some 8×8 blocks, and then neural networks according to different textural features and luminance of each block decide adaptively different embedding strengths. The watermark detection adopts correlation technology. Experimental results show that

		the proposed scheme has good imperceptibility and high robustness to common image processing operators.
Motion Compensation and Reconstruction of H.264/AVC Video Bitstreams using the GPU	Bart Pieters, Dieter Van Rijsselbergen, Wesley De Neve, Rik Van de Walle	Most modern computers are equipped with powerful yet cost-effective Graphics Processing Units (GPUs) to accelerate graphics operations. Although programmable shaders on these GPUs were designed for the creation of 3-D rendering effects, they can also be used as generic processing units for vector data. This paper proposes a hardware renderer capable of executing motion compensation, reconstruction, and visualization entirely on the GPU by the use of vertex and pixel shaders. Our measurements show that a speedup of 297% can be achieved by relying on the processing power of the GPU, relative to the CPU. As an example, real-time playback of high-definition video (1080p) was achieved at 62.0 frames per second, consuming only 68.2% of all CPU cycles on a modern machine.
Multiple Digital Watermarking Scheme Based on ICA	Cong Jin, Ting Su, Liang-Gang Pan	This paper proposes a novel approach to multiple digital watermarking of images based on independent component analysis (ICA). Firstly, discrete wavelet transform (DWT) was applied to one sub-image extracted from the original image orderly, at the same time, multiple independent watermark signals mixed with ICA and generated mixed signals. Then a signal was chosen from the mixed signals as watermarking embedding into the intermediate frequency by mixing with ICA. At extracting stage, using the same rule as embedding stage, extracting the sub-image from watermarked image, then every embedded watermark signal can be extracted through twice demixed with ICA by inverse embedding process. Using the watermarking standard testing tool StirMark Benchmark 4.0 to attack the watermarked images. The simulation results show that multiple watermarks are extracted perfectly, and also demonstrate robustness and invisibility of the method.
Analysis of the Impact of Demosaicking on JPEG Image Compression	Colin Doutre, Panos Nasiopoulos	Most consumer digital cameras capture colour data using a single light sensor with a colour filter array (CFA). This results in a mosaic image where each pixel location contains either a red, green or blue sample, and the two missing colours must be interpolated from the surrounding samples in a process called demosaicking. After demosaicking the image is typically compressed with lossy methods. In this paper we examine the impact demosaicking has on subsequent JPEG compression of still images. Simulations indicate the demosaicking method chosen affects the amount of error introduced in the compression stage and that the amount of compression applied affects the benefit from using more advanced demosaicking methods.
Adapting Surveillance Video to Small Displays via Object-Based Cropping	Luis Herranz, Jose M. Martinez	The increasing availability of small mobile devices capable of playing video, such as PDAs and phones, makes possible their use for new video applications. In these small resolution devices, adaptation is essential to maximize the utility of the delivered content improving the user experience. Content analysis usually helps to detect the interesting parts trying to focus the adaptation on them. In this paper, we propose adaptive cropping as a useful method to deliver adapted video with the objects of interest for the user. Window filtering is used to cope with detection errors, and also to remove annoying artefacts as flickering and unnatural motion in the adaptation window. Two cases are studied: fixed size adaptation window and fixed aspect ratio adaptation window. Although the algorithms presented are generic, in this paper we focus on video surveillance sequences.
A New Spatio-Temporal Predictor for Motion Estimation in H.264 Video Coding	Spyridon K. Kapotas, Athanassios N. Skodras	The new JVT H.264 video coding standard is becoming the dominant standard, mainly due to its high coding efficiency. Among all modules in the encoder, motion estimation plays a key role since it can significantly affect the output quality of an encoded sequence. However, motion estimation requires a significant part of the encoding time. Thus, various fast motion estimation algorithms have been proposed. The H.264 reference encoder has currently adopted three fast motion estimation algorithms. These algorithms aim at reducing the search area by predicting the search points using suitable predictors. In this paper we propose a new spatio-temporal predictor and we study the effectiveness of different predictors in order to find out whether these predictors truly contribute to the decrease of motion estimation time without affecting video quality and adding undesirable complexity to the encoder.
Error resilient H.264/AVC video over satellite for low packet loss rates	Shankar Manuel Aghito, Soren Forchhammer, Jakob D. Andersen	The performance of video over satellite is simulated. The error resilience tools of intra macroblock refresh and slicing are optimized for live broadcast video over satellite. The improved performance using feedback, using a cross-layer approach, over the satellite link is also simulated. The new Inmarsat BGAN system at 256 kbit/s is used as test case. This system operates at low loss rates guaranteeing a packet loss rate of not more than 10 ⁻³ . For high-end applications as 'reporter-in-the-field' live broadcast, it is crucial to obtain high quality without increasing delay.
Improvement of Image Retrieval by Fusing Different Descriptors	A. Macedonas, S. Fotopoulos, G. Economou	A method for optimized image similarity measurement based on fusion of image descriptors is introduced. Descriptors of every kind can be used. The method is based on the nonparametric Wald – Wolfowitz (WW) statistical test, and the corresponding similarity measure, which is very well suited for that purpose. A simple addition of the above WW based similarity measures is the fusion rule proposed here. Experimental results in image retrieval task are very promising.
Empirical Analysis of Libs Images for Ovarian Cancer Detection	Claude Tameze, Robert Vincelette, Nouredine Melikechi, Vesna Zeljkovic, Ebroul Izquierdo	To develop a way to detect early epithelial ovarian cancer we expose blood samples to a laser. Using Laser induced breakdown spectroscopy, (LIBS) plasma images of the blood samples are generated and analyzed. In this paper we compare the images from blood specimens of cancer free mice to those of transgenic mice; that is, mice that are currently developing epithelial ovarian cancer and are approximately the same age as the disease free mice. Our ultimate goal is to look for changes in the shape and edges of the image that might indicate the presence of cancer. To analyze the images we use a nonlinear diffusion filter for enhancement of relevant image edges and removal of noise and irrelevant texture.
Analysis and Cryptanalysis of a Selective Encryption	W. Puech, J.M. Rodrigues, A.G Bors	This paper addresses the protection of images. We address the problem of simultaneous selective encryption (SE) and image compression. The SE is done by using the Advanced Encryption Standard (AES) algorithm with the Cipher Feedback (CFB) mode one a part of the Huffman coefficients

Method for JPEG Images		corresponding to the AC frequencies. For the compression we consider the JPEG algorithm. Our approach is done without affecting the compression rate and by keeping the JPEG bitstream compliance. In the proposed method, the SE is performed in the Huffman coding stage of the JPEG algorithm without affecting the size of the compressed image. We provide an experimental analysis of the proposed method when applied on still images as well as the results when applying SE in JPEG compressed images.
Optimal Estimation of Descriptor Scales for Multimedia Retrieval	Nikolaos Doulamis	Relevance feedback schemes use a given set of descriptors and at each feedback iteration they just modify the importance of each descriptor to the retrieval performance. This implies that the same type of descriptors is used for all feedback iterations. Therefore, such approaches cannot improve the retrieval performance beyond the efficiency of the descriptors used to represent the multimedia content. In this paper, an alternative concept is to adopt a progressive descriptor organization and to allow dynamic modification of the type of descriptors (expansion or shrink) at each retrieval iteration by exploiting the current user's information needs. In other words, descriptors that best fit the current user's preferences are expanded into more details, in constant to the descriptors that are far way from the user's needs which are shrinking. In this case, instead of using a large set of descriptors, most of them inappropriate with respect to the current user's information needs, a small number of dominant descriptors is selected for the retrieval process. Thus, the proposed approach significantly increases the precision performance by discarding descriptors that yield noise in the retrieval.
A Feature Extraction Method for Multimedia Data Analysis in Robot Wireless Sensor Networks	Heng Huang, Fillia Makedon, Dan Popab, Harry Stephanou, James Ford, Yurong Xu	Multimedia sensor networking technologies have developed very rapidly in the last ten years. Robot wireless sensor networks are an increasingly capable platform and show great potential in environments from future battlefields to disaster discovery. Robot wireless sensor networks can enhance multimedia surveillance and provide the foundation for strategies based on multi-modal sensor integration. We illustrate this concept with an example feature extraction approach for image or video data: lines are extracted as image features from images transformed using the Hough transform technique. This approach enables features to be combined and constrained by their structure, allowing for example region correspondence mapping between images from different sensors or the same sensor in different time phases.

Friday 15:30-17:10

Applications (Oral Session)

Title	Authors	Abstract
Time-Compression of Speech in Informational Talks using Spectral Entropy	Mohammed Ajmal, Azadeh Kushki, Konstantinos N. Plataniotis	Speech is one of the richest sources of information in human communication. In several situations, for example, classroom lectures, speech data alone conveys the majority of pertinent information [1]. Large speech archives of lecture material are increasingly available on the Internet, but the lack of effective audio skimming tools complicates the interaction between users and audio material. In this paper, we describe a novel method for time-compression of speech based on the information theoretic measure of entropy. We demonstrate the applicability of spectral entropy in identifying important segments in speech, enabling time-compression of speech for skimming. We present results from a user study which confirms the viability of our approach.
Cassandra Framework: A Service Oriented Distributed Multimedia Content Analysis Engine	Fons de Lange, Jan Nesvadba, Johan Lukkien	Connected Consumer Electronics (CE) devices face an exponential growth in storage, processing capabilities and connectivity bandwidth. This increased complexity calls for new software architectures that naturally admit self-organization, resource management for efficient workload distribution, and a transparent cooperation of connected CE-terminals. The interconnected nature makes advanced applications feasible such as the smart distribution of complex multimedia content analysis algorithms, resulting in automatic semantic content-awareness creation. In this paper we describe a modular, self-aware, self-organizing, real-time and distributed multimedia content analysis system. Based on a Service Oriented Architecture it is, furthermore, capable of efficiently and dynamically using the available resources in a grid-like manner.
Multimedia Content Adaptation Modelled as a Constraints Matching Problem with optimisation	Fernando Lopez, Jose Maria Martinez	Within a multimedia adaptation engine working with different adaptation tools and parameters, an important function is the selection of the most appropriate parameters to execute these adaptation tools. MPEG-21 compliant multimedia adaptation engines make use of MPEG-21 description tools to drive the decisions targeting to adapt multimedia content to the usage environment (terminal capabilities, network conditions, user preferences, etc). In this paper, a simple but effective method to model the decision phase of the adaptation engine is presented. This method is based on constraints matching with a final optimisation phase. The proposed system works over MPEG-21 compliant usage environment descriptions, also making use of MPEG-7 content descriptions when available.
Design and Evaluation of a Metadata-Driven Adaptation Node	Michael Ransburg, Christian Timmerer, Hermann Hellwagner, Sylvain Devillers	MPEG-21 Digital Item Adaptation (DIA) allows for a media codec agnostic multimedia adaptation approach which enables the implementation of generic adaptation engines. However, DIA is optimized for static, server-based adaptation. In this paper we introduce novel mechanisms to extend the DIA approach towards dynamic and distributed scenarios. This facilitates the placement of generic adaptation nodes which perform media codec agnostic and dynamic adaptation anywhere along the content delivery path. To validate our work we implemented such an adaptation node and evaluate its performance.
Recursive Partitional-Hierarchical Clustering for Navigation in Large Media Databases	Angelica Lozano, Paulo Villegas	We present in this document a tool for automatic clustering of repositories of tagged media items (images, videos) and a 3D visualization interface to enable interactive browsing of repository contents based on the generated clusters.

NOTES

NOTES
