



Second Summer School on Multimedia Semantics

Analysis, Annotation, Retrieval and Applications

University of Glasgow
Scotland, UK

July 15th – 21st 2007

Abstract Booklet

Editors: Jana Urban, Frank Hopfgartner, Robert Villa, Joemon M. Jose, C.J. van Rijsbergen

Booklet Editor: Jana Urban, Frank Hopfgartner

Sponsors

K-Space, NoE (FP6-027026)
AceMedia, IP (FP6-001765)
Salero (FP6-027122)
IP-Racine (FP6/2002/IST-2-511316)
BCS-IRSG
Multimedia Knowledge Management Network
Yahoo! Research, Barcelona
University of Glasgow

Table of Contents

POSTERS	1
AHMED AZOUGH	1
LIANG BAI	2
TOBIAS BÜRGER	3
ANDRIUS BUTKUS	4
DARAGH BYRNE	5
KRISHNA CHANDRAMOULI AND EBROUL IZQUIERDO	6
PETR CHMELAR	7
ANDREAS COBET	8
YOLANDA COBOS, M.T. LINAZA	9
DANICA DAMLIANOVIC	10
KERSTIN DENECKE	11
JEAN-LOUIS DURRIEU	12
ALISTAIR J. EDWARDES	13
MEHDI ELLOUZE	14
ILKO GRIGOROV	15
CHRISTOPH GRÜN	16
MARTIN HALLER	17
MARTIN HALVEY	18
PEYMAN HEYDARIAN	19
MARK HUGHES	20
CHARLES INSKIP	21
HUANG DONG JUN	22
FATIMA ZAHRA KAGHAT AND MOURAD OUZIRI	23
PETROS KAPSALAS	24
CHANYUL KIM	25
ANDREAS KRUTZ	26
SEUNG-BUM LEE, GABRIEL-MIRO MUNTEAN, AND ALAN F. SMEATON	27
HAIMING LIU	28
AINHOA LLORENTE	29
ANTONINO LO BUE	30
MICHAEL MAY	31
DALILA MEKHALDI	32
AYMAN MOUNIR MOGHNIEH	33
DONN MORRISON, STÉPHANE MARCHAND-MAILLET, AND ERIC BRUNO	34
ZURINA MUDA	35
ALEV MUTLU	36
MAREK NEKVASIL	37
ANDREEA NICULESCU	38
STEFANIE NOWAK	39
XIMENA OLIVARES	40
DEISLAVA PANEVA	41
GIUSEPPE PASSINO	42
MARIOS PHINIKETTOS	43
SÉBASTIEN POUILLOT, OLIVIER BUISSON, AND MICHEL CRUCIANU	44
K. RAJKUMAR, B. RAMADOSS	45
RADHA KRISHNA RAMACHANDRUNI	46

MARK RESTALL	47
BERTRAND RICHARD	48
I. ROJAS AND F. BLANDÓN	49
MARIA M. RUXANDA	50
KRYSTIAN SAMP	51
PHILIPP SANDHAUS	52
VALIA SARAYDAROVA	53
ISABEL SEGURA BEDMAR	54
GIUSEPPE SERRA AND CARLO TORNIAI	55
MILAN STANKOVIC AND UROS KRCADINAC	56
JAVIER TEJEDOR	57
MARÍA TERESA VICENTE-DÍEZ	58
BEIMING WANG	59
JAN WEIL	60
LAILATUL QADRI ZAKARIA	61
HERWIG ZEINER	62
PROGRAM	64

A Representation Language for the Multimedia Semantic Web

Ahmed Azough

France Telecom R&D

ahmed.azough@orange-ftgroup.com

Abstract. While the low-level analysis of multimedia features was largely studied, that of the high-level descriptions extraction is still an open research field. As semantic web languages, based on Description Logics, are conceived for describing all types of resources, specific extensions for multimedia are needed. Descriptors offering enough expressivity to represent spatio-temporal constraints are the major needed tools to describe high-level video features. Such descriptors should allow for relying between different levels of descriptions and fill then the semantics gap between these descriptions.

In this poster we present a generic language for the representation and the description of Multimedia contents. Based on Finite Automatas and Description Logics, this formalism allows for expression of spatial and temporal constraints to define successions of visual events In addition to the description of hierarchic complex concepts, this formalism allows also for the video guided monitoring of process execution, the detection of important highlights in a video and the validation of MPEG7-based descriptions.

Video Semantic Content Analysis based on Ontology

Liang Bai

The Centre for Digital Video Processing, Dublin City University, Ireland
lbai@computing.dcu.ie

Abstract. The rapid increase in the available amount of video data is creating a growing demand for efficient methods for understanding and managing it at the semantic level. New multimedia standards, such as MPEG-4 and MPEG-7, provide the basic functionalities in order to manipulate and transmit objects and metadata. But importantly, most of the content of video data at a semantic level is out of the scope of the standards. MPEG-7 provides the rich functionalities to enable the generation of audiovisual descriptions and is expressed solely in XML Schema which provides little support for expressing semantic knowledge. In our work, a three-layer video semantic content analysis framework based on ontology is presented. Domain ontology is used to define high level semantic concepts and their relations in the context of the examined domain. In videos, similar and periodic actions often are used to express important semantic content. The similar and periodic-action patterns, maybe a frame, a group of frames, an object or a segmentation of audio track, share similar spatiotemporal behaviours which can be clustered and described with a linguistic concept. Perception Concepts are defined in our framework to represent these important semantic patterns for videos based on identifiable feature elements in middle level. In low level, we focus on how to combine MPEG-7 metadata terms of audiovisual descriptions into the ontology in OWL expressing to enrich video semantic analysis and how to recognize visual content, such as objects, in an image and associate it with a linguist concept defined in the domain ontology automatically. We try to propose an approach to include local contextual features and global features for visual objects detection, in which visual content is labelled with a finite set of objects labels at pixel level. The contextual features are incorporated into a Conditional Random Fields (CRFs) model trained by labelled image data to label objects. CRFs is a framework for building probabilistic models to segment and label sequence data, which offer several advantages over HMM and stochastic grammars including the ability to relax strong independence assumptions and avoid the bias problem made in those models. Web Ontology Language (OWL) is used for the ontology description. Rules in Description Logic are defined to describe how low-level features and algorithms for video analysis should be applied according to different perception content. Temporal Description Logic is used to describe the semantic events, and a reasoning algorithm is proposed for events detection. The proposed framework is demonstrated in sports video domain and shows promising results.

Towards Self-Producing Media

Tobias Bürger

Digital Enterprise Research Institute (DERI)
Innsbruck, Austria
tobias.buerger@deri.at

Abstract. The digital media and games industry is worldwide one of the biggest IT based industries. Recent observations therein showed that current production workflows may be potentially improved as multimedia objects are mostly created from scratch due to insufficient reusability capacities of existing tools. The amount of digital content that may be potentially reused is constantly growing. This is not only true for content in commercial databases but also for the vast amount of digital content on the Web. Therefore we work on a solution to increase the reusability potential of digital content, which aims to establish a semi-automatic adaptability framework for previously created digital content based on semantic technologies. With this solution producers will be able to face the huge amount of content available, i.e. they may be able to select which content to reuse and include it in their work. Rich semantic descriptions of the content will allow to select and handle content automatically in order to ease search, retrieval and automation of several processing steps in the media production chain.

Media Personalization using Multiple TV-Anytime Classification Schemes

Andrius Butkus

Technical University of Denmark
Lyngby, Denmark
ab@imm.dtu.dk

Abstract. Internet and broadcast platforms are swarming with content. The main problem is how to find what you like among all the available media. This is done by personalization. When it comes down to personalization, major research areas fall in one of the categories: recommender systems, content metadata and user profiles. In my PhD thesis I am analyzing media personalization from a metadata perspective. I think it is the key component that influences all the other ones. The most popular way to index content (create metadata) is to assign several keywords or write a short synopsis. Both of those methods fit perfectly for humans but are not so effective when it comes down to automatic content recommendation. Another way would be to somehow grasp the very essence of media item and use it to find similarities among media items. But what are the essential features and how to represent them in a metadata model? In my thesis I am using feature selection on the TV-Anytime metadata. TV-Anytime (TVA) is an industry driven specification for description and delivery of media both for broadcast and all kinds of new converged platforms. TVA uses controlled term vocabularies to describe media items on various axes: content, format, intention, atmosphere, etc. Since 2005 BBC (one of the main players to promote TVA) has made all of their program data descriptions available for public use (backstage.bbc.co.uk). I have used their database to test my hypothesis that it gives better results if we describe content using parallel simple classification schemes that supplement each other instead of one huge and very complex classification scheme. Using concepts from information theory (entropy, mutual information and information gain) and information economy (bundling and targeting) I have selected the most significant features and used them to find similarities among content items. Once the features are identified the semantic rules can be applied to infer the meaning of the content, what in my case would be expressed in RDF triplets.

Making Sense of SenseCam: Exploring Extremely Large Event Oriented Lifelog Image Sets

Daragh Byrne

Centre for Digital Video Processing (CDVP)
Dublin City University, Dublin 9, Ireland
daragh.byrne@computing.dcu.ie

Abstract. The SenseCam is a wearable device, which passively records a person's day-to-day activities as a series of photographs. It records approximately 3,000 images per day totaling over a million photographs within a year. The size of the photoset and likelihood of recording large volumes of mundane information are challenges to browsing and searching the photoset as well as engaging the user while performing a review of their collection. We present a novel rich and compelling visualisation for a SenseCam photoset which builds on previous work at the CDVP on the automatic segmentation of a SenseCam image collection into unique events. The visualisation adapts the established river metaphor and presents images as a series of animated streams. Each event is represented by a single representative keyframe as it moves across the screen and makes use of size, transparency and motion to indicate an events importance within the set. The visualisation successfully allows for passive review of a lifelog photo collection in a manner akin to a slideshow. With the visualisation established, we have extended and enhanced the application to enable directed interactive enquiry into the lifelog's events. Using additional sensors such as a Bluetooth and GPS, we are now annotating events with location information and the presence of Bluetooth-enabled devices, enabling people- and location-based search of a SenseCam lifelog. We are currently examining how best to provide this search functionality within the user interface.

Particle Swarm Optimisation in Image Classification

Krishna Chandramouli and Ebroul Izquierdo

Multimedia and Vision Research Group,
Queen Mary, University of London,
London, UK

{krishna.chandramouli, ebroul.izquierdo}@elec.qmul.ac.uk

Abstract. Recent developments in optimization techniques are inspired by complex biological systems. Such optimisation techniques are based on the behaviour of insects, birds and fish schooling. One such algorithm developed by Kennedy and Eberhart is Particle Swarm Optimisation, in which the bird flock movement is modelled to solve the optimization problem. Similar to other optimisation technique, the performance of particle swarm optimisation depends largely on the initial random distribution of particles in the solution space. Also, like other optimization techniques, particle swarm optimization suffers from local minima problem. In this work, two variants of particle swarm optimization, prime distribution based particle swarm optimization and chaotic particle swarm optimization will be presented addressing the random distribution problem and local minima respectively. The variants will be applied to optimize the weights of neurons in self organizing map for image classification problem based on MPEG – 7 low-level features. The performance comparison of the algorithms will be presented from the Corel dataset.

Mining Knowledge from Distributed Visual Surveillance Networks

Petr Chmelar

Brno University of Technology
Brno, Czech
chmelarp@fit.vutbr.cz

Abstract. Many visual surveillance systems continuously provide huge amounts of raw video data for security purposes all over the world. The consolidated data coming from a camera network is a potential source of useful information that cannot be discovered separately. The information may be used both for online event recognition and for offline querying, analyzing and mining. May be.

The problem is that computer vision techniques are still underdeveloped, they produce information that is supposed to be noisy, uncertain, and some states are missing. Although there are many problems concerning the visual sensor fusion (cameras with processing units), I believe that it is the solution. For instance, an object that cannot be identified in the rush hall, can be sometimes later marked as a dog. Of course, there are more serious applications wanted.

The first goal is to keep necessary information of an object (unique id) in large systems also with non-overlapping cameras. I proposed reversing the Kalman filter and using the reversed internal state together with the camera number as a set of features that simply classifies the previous locations – cameras in which the object appeared most probable. The classification is then validated using similarity search in the databases of the previous locations – a similar object had to be present at the time. The system training is performed using similarity search in different sensors and the operator only validates the proper pair of the same object in the training phase.

The database then contains a lot of useful information but there is a need to define a semantic. There are either environmental relations like enter, leave, cross, stay and bypass, or trajectory relationships like visit, together, merge and split. The last two might be used similarly to the left-luggage problem. However, the final goal of my research is concerning the knowledge about groups of objects, which cannot be recognized using computer vision based segmentation.

Text Detection in Videos

Andreas Cobet

Technische Universität Berlin
Germany
cobet@nue.tu-berlin.de

Abstract. Text can be found in the most videos. For example text is in news videos the topic or in soccer videos the score. This text carries a lot of information which describe the video. A lot of Optical Character Recognition (OCR) algorithm exists for images with black text and white background. This is well done. The next step is to find and recognize text in a coloured image or video with structured background. This task is divided in three steps. First step is finding the position of the text in the image. After the position of the text is known, the second step converted the coloured image of the text area to an image with black text and white background. In the last step OCR algorithms recognize the text. In the most cases the recognized text can describe the content of the video.

Multimedia Content Semantic Annotations based on MPEG-7

Yolanda Cobos, M.T. Linaza

VICOMtech (Visual Communication and Interaction Technologies Centre)
San Sebastian, Spain
ycobos@vicomtech.es, mtlinaza@vicomtech.es

Abstract. Multimedia content indexation is one of the major challenges for many organizations within different economics sectors. There are many applications and domains that can benefit from multimedia content semantic indexation and management, such as education, tourism, cultural heritage, entertainment, bio-medicine, architecture, shopping or eInclusion. Currently, semantic-based indexation and management techniques are mainly applied in digital libraries, multimedia directory services, broadcast media selection and multimedia editing. Our efforts are focused in designing, implementation and validation a mobile media-rich collaborative information exchange platform, scalable, based on multilingual approaches, accessible through a wide variety of networks, and therefore, interoperable and context aware of the domain, using an designed ontology with some concepts selected from the MPEG-7 multimedia standard. Due to the new technologies progress, consumers (or prosumers) increasingly generate content on any device, at anytime and anywhere. This motivation produces the existence of content-on-the-move term. Our platform gives us the chance to retrieve multimedia content and to create multimedia content-on-the-move.

Semantic Search for Enhanced Knowledge Access

Danica Damljanovic

University of Sheffield, Department of Computer Science
Sheffield, UK
danica@dcs.shef.ac.uk

Abstract. The idea of Semantic Web systems is to semantically enrich available content, so that it would be in machine-readable form. Querying semantically enriched data, where the emphasis is on searching by content and not by a keyword, requires improving existing search engines so that they support semantic search. Our research goal is to work on developing a semantic search engine and run it against the knowledge store to get meaningful answers. The User Interface will look as simple as common search engines, such as Google. The only difference should be that of finding the results that are more meaningful and more useful for user. Methods of assisting users while creating a query are also considered. Most of the available semantic search engines have very complex interface that implies that user is familiar with the concepts of Semantic Web, which doesn't have to be the case. Some applications allow searching the knowledge store using a controlled language, which means - they require syntactically correct sentences. The result of the specific query shouldn't depend on whether the user entered the exactly patternized question, or he missed few 'expected' words: the search engine should accept queries of any length and form. Additionally, search engine should be personalized, with a learning mechanism that provide improvement of the performance of the system over time. Such a system should also be portable, so that it can be easily used with different knowledge stores. Developing a semantic search engine that processes natural human language queries would be of a great contribution once the process of Semantic Web is established and applied at least at the applications, if not on the Web as a whole.

Information Extraction from Unstructured Medical Documents using Semantic Structures

Kerstin Denecke

University of Hannover, Research Center L3S,
Hannover, Germany
Technical University of Braunschweig,
Braunschweig, Germany
denecke@l3s.de

Abstract. In medicine, written text plays a significant role in documentation and communication. Automatically understanding document content can be a substantial aid for reusing information, for documenting as well as for decision making. For this reason, possibilities for automatically locating specific data in medical documents and for processing this data are necessary. A prerequisite for this however is a structured representation of the essential facts of a text.

Our research goal is to develop, analyse and evaluate a method for transforming sentences of a medical document into semantic representations using existing technologies. Based on these semantic structures, methods are developed for detecting specific data such as documented diagnosis and procedures. The goal of such methods is to overcome the limitations of existing medical language processing (MLP) tools. These tools are often limited to a certain medical domain and to a certain natural language. Their construction is fairly complicated because the acquisition of lexical knowledge is expensive and time-consuming.

The chosen approach in this work uses existing language engineering technology, i.e. a multiaxial medical terminology (Wingert Nomenclature), a concept-based morpheme lexicon, a word segmentation algorithm as well as semantic transformation rules for mapping syntactic information to semantic roles. Sets of indices are mapped to conceptual structures using information from semantic categories and word position. The extraction process itself is based on hand-crafted extraction rules and searches the semantic structures and the text for the desired information.

The procedure achieves a correctness of around 80% in generating semantic structures for a text. An evaluation of the system's performance for several information extraction subtasks in the medical domain was conducted. Values of 81-96% precision and 83-98% recall were obtained. These results suggest that the chosen methods can be used to accurately extract data from medical narratives. The methods provide two main benefits: by using existing language engineering methods the efforts for constructing a medical information extraction system are reduced and the system is not limited to a certain medical subdomain.

Currently, we are analysing whether the methods can also be based on the free available medical terminology UMLS. We also want to investigate the potentials of using relations of a medical semantic network to support the generation of semantic structures as well as the extraction of specific data from text.

Automatic Extraction of the Main Melody in Music Signals

Jean-Louis Durrieu

GET ENST Paris,
France
durrieu@enst.fr

Abstract. In order to mine a multimedia database in an efficient way, one needs more than the meta data that are usually available. Extracting relevant features directly from the source file is necessary and our research aims at providing some of these features. Extracting the melody from a song might benefit to several applications such as Query-By-Humming, Cover-Version Identification and any other application that involves a symbolic melody similarity step. In particular we might be able to build a database of monophonic melodies which can be useful to the above applications. We want to experiment source separation approaches for the task of automatic melody extraction. At first, we focus on songs with a predominant singer voice over a background music, which is the case for most of the popular pop songs. We use a Non-negative Matrix Factorization (NMF) approach to model the background music along with a specific modelling of the voice signal, with spectral Gaussian Mixture Models (GMM). We combine them by considering that the mixed signal is the sum of the voice signal and the music signal. The parameters of the models are estimated thanks to the Expectation-Maximization (EM) algorithm. Thanks to these parameters, we build the sequence of frequencies that most likely generated the voice signal, thus obtaining the wanted melody.

Describing Places: Building a Concept Ontology for Image Retrieval

Alistair J. Edwardes

Department of Geography, University of Zurich
Winterthurerstrasse 190, 8057 Zurich (Switzerland)
aje@geo.unizh.ch
<http://www.ProjectTripod.org>

Abstract. Current techniques for image retrieval lag far behind the state of the art in text search. For the most part this is due to the difficulty of describing an image, both in terms of what it is *of* and what it is *about*. One important factor related to this issue is how to describe the setting of *where* a photograph was taken. It is proposed in the Tripod project to achieve this by synthesising techniques from image and information retrieval with those from geographic information science. The work described here forms part of this project. It reports on experiments carried out to develop an ontology of geographical concepts used by people when describing photographs of landscapes. Two sets of experiments are presented. The first are empirical and based on a set of online tasks. The second are data-driven employing a database of image captions to mine descriptions of basic-level scene categories.

Toward an Intelligent TV based on Event Detection

Mehdi Ellouze

Research Group on Intelligent Machines (REGIM)
&
Sfax Superior Institute of Technological Studies
Sfax, Tunisia
mehdi.ellouze@ieee.org

Abstract. The digital television broadcast does not simply mean a mere shift from analogue to digital but provision of personalization and interactivity to its users. The number of spatial channels is increasing with a great speed. People are exposed to information overload due to the presence of several hundreds of alternative programs to watch. You can have more than one important information which may be broadcasted at the same time. The solution to this problem is to create an interactive and a personalized TV, a TV which stores the different profiles of the users and tries to filter in real-time the incoming video streams to insure that the current user (the current viewer) doesn't miss any important scene or program. Our research goal is to propose a real filtering strategy applied in real time on video streams. Our work is a kind of a real-time video indexing in which we will try to bridge the semantic gap between the broadcasted video signals and the users' preferences. The users' preferences are generally special events. They may be related structure events which are events related to the structure of the video, as goals in soccer video matches or predefined events which are events related to special circumstances, places and actors as stories in news video broadcast. To detect these two kinds of events we will base on a multimodal analysis of every video stream. We will analyze the visual, the auditory and the textual features and we will fuse them to decide if the current clip or scene (real time analysis) corresponds to one of the preferences the current user. The challenges are the determination of the limits of the current clip to analyze it, the features extraction and the fusion of the analysis results to decide through a comparison with the current user preferences if the current clip is interesting. To overcome all these challenges we thought that a multi agent system may solve this problem. There is a delimiter agent which will delimit the current clip, there is for every feature an agent in charge of extracting the feature vector from the current clip and there is also a fusion agent which will take in collaboration with the other agents a fuzzy decision if the current clip is interesting. In fact, it's impossible to be sure at 100% that a current clip is interesting or not, that's why the introduction a fuzzy decision and a confidence coefficient will be necessary. The fuzzy decision will be the result of a communication between the agents through fuzzy rules.

Semantic Multimedia Annotation, Searching and Browsing of Public Resources and Events based on Combining of Administrative Aspect, Multilingual Ontology, Name Entity Databases, Gazetteers and Maps in Collaborative Environment

Ilko Grigorov

National University of Ireland Galway
Galway, Ireland
ilko.grigorov@deri.org

Abstract. There is a vast number of towns and villages all over the world varying from very small to huge metropolitan ones. In each of them there are different types and number of public resources like museums, theatres, universities, churches, parks, zoo gardens, sport fields, monuments and many others, as well as different types of public events happen all the time. For each of these resources and events public administration authorities are responsible for managing and storing of data for them. Unfortunately this data is not available on the Web in one place in a structured and machine readable format. People take pictures, audio recordings and video clips of such resources and events and publish them on the web providing little or no metadata, which disables the chance of retrieving them by the search engines. The author's research aims at developing of a framework where administrative authorities can publish metadata for public resources and events, and people can link their multimedia resources like image, audio and video ones to existing multilingual ontology like EuroWordNet and BalkaNet in one collaborative environment like Wikipedia. Ontology, name entity databases and gazetteers are used as a base for metadata resource annotation, searching and browsing. Ontology represents the vocabulary that describes objects and the relations between them in a formal way. The name entity databases consist of name instances of the ontology and the gazetteers geographical dictionaries. Additional tools like multi query system and download manager are intended to be developed for convenient multimedia resource browsing. This will give opportunity for people from different countries to browse resources from different locations in the world and specific topics they are interested in.

Framework for Developing Location-based Services in the Tourism Area

Christoph Grün

Vienna University of Technology, Austria
christoph.gruen@ec.tuwien.ac.at

Abstract. The mobile industry expects location-based services to be the cash-cow of the future as those services. This is because those services can exploit the user's context situation to provide relevant and selective information that is adapted to their current situation and preferences. Mobile users might be more willing to pay for such highly personal information. One major application area of those services is in tourism where they might assist tourists during their trip in retrieving personalized information about points of interests in a relaxing way. Many research projects have been done in this field, with most of them focusing on single aspects such as technical or business related ones without providing a comprehensive guideline how to develop location-based services in this area. In my PhD research I focus on the problem of establishing such a framework that consists of several aspects and views. a) the business view, comprising the different stakeholders in the value chain such as content provider, network operator or mobile customer; b) the technical view, that focuses on aspects concerning the location detection method and the infrastructure of the system; and c) interaction view, that sheds light on the possible ways of interaction and the flow of information, e.g., using a game concept to raise the entertaining value of information delivery and enhance the tourist's experience. This framework should be then tested within a prototypical example in the tourism area.

Automatic Video Summarization

Martin Haller

Communication Systems Group, Technische Universität Berlin
haller@nue.tu-berlin.de

Abstract. A summary of a video only contains ideally the essential audiovisual content. Reduction of details is obtained by selecting important parts of the original video to represent the content of the whole video clip. Such a summary can be a dynamic summary (video clip) or a static summary (comic-like strip of keyframes). Summaries are useful for navigation in video databases, search for a certain video and specific content (e.g. location/action/topic), and presentation of a shortened version of an original video. Firstly, this research work examines content-based audiovisual analysis techniques that are useful for a summarization of videos. Currently, methods for motion-based video parsing are developed and evaluated. Thereafter, promising video summarization approaches will be investigated, enhanced and evaluated.

Analysis of Online Video Search and Sharing

Martin Halvey

Information Retrieval Group, University of Glasgow
United Kingdom
Martin.halvey@gmail.com

Abstract. It is now feasible to view video at home as easily as text-based pages were viewed when the Web first appeared. This development has led to the emergence of video search engines providing hosting, indexing and access to large, online video repositories. A key question in this new context is whether users search for media in the same way that they search for text. This poster presents a first step towards answering this question by providing novel analyses of people's linking and search behaviour using a leading video search engine. Initial results show that page views in the video context deviate from the typical power-law relationships seen on the Web. However, more positively, there are clear indications that tagging and textual descriptions play a key role in making some video-pages more popular than others. This shows that many techniques based on text analysis could apply in the video context.

Analysis and Classification of the Persian Musical Modes

Peyman Heydarian

Queen Mary, University of London
UK
peyman.heydarian@elec.qmul.ac.uk

Abstract. Persian music is based on a modal system consisting of seven main modes and their five derivatives. They are characterised with the tuning, the modal tonic and the mood of a piece. A musical mode usually falls into one of the five different tuning systems.

In this research, algorithms based on the spectral averages, and pitch histograms are developed to identify the tuning and consequently the mode of an audio signal. Using the spectral average and pitch histograms and a similarity measure like the Manhattan distance all the samples of our database are identified correctly. Knowledge of the mode can be used in music information retrieval, audio snippet, music archiving and access to the musical content, music recommendation and playlist generation, audio compression and coding and music transcription.

Analysing Image-Text Relations between News Stories and Images

Mark Hughes

Centre for Digital Video Processing
Dublin City University
Ireland
mhughes@computing.dcu.ie

Abstract. The automatic analysis and structuring of semantic multimedia content is crucial in the development of multimedia information retrieval, browsing and summarization systems. The combined use of different media is a defining characteristic of multimedia but prior research has concentrated on analysing and structuring multimedia data types in isolation and integrating them, if at all, in ad-hoc ways. The fusion of multimedia data streams requires knowing how, in a particular instance, the meaning of one media type relates to the meaning of another. For example, a web search engine that indexes images using keywords from neighbouring text needs to know if and how the text relates to what is depicted in the image. A recent semiotic analysis of multimedia documents suggested that it is possible to systematically identify a finite set of text-image relations that describe the conventions in how linguistic and visual sign systems combine. It was hypothesised that the ways in which the meanings of images and texts relate to one another – the image-text relations – can be classified on the basis of low-level media features, i.e. without needing to understand the content of either the image or the text, but this idea has not previously been tested. In our research we aim to explore this hypothesis experimentally. Our work to date shows that news images of people can be successfully classified into “general” or “specific” based on low-level visual features. We are currently extending our analysis to the associated texts, and will then analyse the relationship between classified image and text types. These relations could be potentially relevant for a wide range of multimedia retrieval, browsing and summarization applications.

How Different Social Groups within the Music Industry Communicate Meaning in order to Satisfy their Information Needs

Charles Inskip

Centre for Interactive Systems Research
Department of Computing Science, City University
UK
c.inskip@city.ac.uk

Abstract. When searching for music, users follow a number of search strategies depending on whether they are looking for known items or unspecified items that suit certain contextual criteria. Digitisation has made Music Information Retrieval (MIR) increasingly important because users now have access to a large number of globally and locally situated documents. Complexity of facets (harmony, polyphony and timbre) and representation (symbolic, audio, visual and metadata) combines with cultural, disciplinary and experiential factors to make MIR a significant topic with an established global research base. While there is an emerging discipline there is not much existing research in user needs, particularly in the area of users needing music for work purposes.

The research aims to identify needs of music industry professionals, examine their information seeking behaviour and the way they communicate and resolve their needs, and recommend ways to improve their search results in the context of Information Retrieval (IR) theory and tools. A reflexive communication model derived from semiotic analysis suggests that both users and producers of music give it meaning, which is affected by their codes and competences, and Ingwersen and Järvelin's Interactive Information Seeking, Retrieval and Behavioural processes model suggests that context is a key influence in information seeking and retrieval. Using an inductive approach and qualitative analysis the writer plans to investigate information seeking and behaviour by questionnaire, interview, and observation using a framework based on semiotic analysis described by Tagg; evaluate some existing systems using traditional Information Retrieval techniques (precision and recall); and use the results to determine whether a general MIR system would be able to satisfy the needs of music industry professionals.

Super Resolution of Face Images Based on PCA and Markov Random Filed Model

Huang Dong Jun

University of Glasgow, UK
Central South University, China
djhuang@mail.csu.edu.cn

Abstract. The sharper an image is, the more details it shows. If it is blurred, it will be difficult for us to interpret it. When we enlarge an image, the necessary interpolation, such as bi-cubic, nearest neighbours, will generate a blurred image. As a matter of fact, it is impossible to restore the true information missed, which corresponds to the details we cannot see in the original image. However, it is conceivable to create a plausible super resolution image by guessing the missing data. This is called learning-based super resolution methods. Considering the importance of the face images, we study the face image super resolution techniques based on image examples. We propose a novel algorithm that uses Principal Component Analysis (PCA) to produce several best familiar face images from the training dataset for the input low-resolution face image to control the global analogue and employ Markov Random Network to learn the details for the final output image. Different from other familiar methods that also employ PCA, our algorithm creates a best mean face image by fusing the best familiar face images selected by PCA for the input image. Then, we model the output high-resolution image as a patch-based Markov Random Network. Because it is computationally infeasible to solve the Markov Random Network model, we approximately use the alignment position information and the compatibility between adjacent patches. This works well under certain conditions. Related experiments based on the well-known FERET image database are carried out and demonstrate the effectiveness and good performance of the proposed algorithm.

Design and Realization of a Tool of Semantic Description and Representation of Images

Fatima Zahra Kaghat and Mourad Ouziri

René Descartes University, Paris
France
fatimazahrakaghat@yahoo.fr

Abstract. The design of new information systems within the semantic Web framework is a very promoter research orientation. Indeed, the Web represents a large collection of sources heterogeneous multimedia resources. The problem of access to these data resources is primarily due to the semantic heterogeneity of their contents.

Taking in consideration the multimedia semantics of the contents on the Web is thus of primary importance. The semi-automatic annotation of multimedia contents using web semantic techniques is an efficient way to assure a semantic-based access to these resources.

Our work comes within the scope of the conception of semi-automatic annotation tools of fixed images. Based on Description Logics, this tool allow for using local or uploaded owl ontologies to annotate image contents into concepts and make relations between them. This tool is characterized by its particularity to annotate segments of images using multiple ontologies in the same time. The tool aims also to facilitate the segmentation of images using many geometric forms to fit with the contents forms. The annotation is generated using into rdf/owl files for the purpose of post use for multimedia information retrieval.

A Local Feature-Based Approach for Detecting Humans in Cluttered Images and Estimating their Topology

Petros Kapsalas

Electrical & Computer Engineering Department, National Technical University of Athens
pkaps@image.ece.ntua.gr

Abstract. The development of machine vision schemes for object detection has gathered the interest of many research groups during the last several years. The detection issue has been generally approached through different viewpoints each handling in a different way the background in-homogeneity and the object features. However, there are several aspects that limit the potential and the applicability of such approaches and they should be carefully considered through the design of an effective detection scheme. The wide variations in the objects scales and orientations (up-right, rotated) as well as the occlusion phenomena that may occur are the major factors inducing considerable limitations in the performance of machine vision-based detection systems. This work is geared towards building a robust feature set, capable of determining the occurrence of humans and extracting accurately their topology and extent. Reliable detection of the object's location under various illumination conditions and orientations is also an objective for this work. The selection of the specific object of interest is associated with the challenges that arise when trying to detect persons in images. More specifically, further to the classical aspects that should be considered such as those mentioned above we should also take into account further issues related to the wide range of articulation of the human body, clothing, and variations in poses.

The contribution of this work in accurately detecting humans in a cluttered background is two-fold. Thus, at first we introduce a robust feature set capable of estimating the image features associated with the presence of a human in an image sub-region. At a further stage, the exact location of human body occurrence is approached by selecting only susceptible regions according to a homogeneity predicate (we assume that the areas of human presence in an image cause discontinuities to the image background). To overcome noise artefacts introduced due to the background clutter and illumination changes, our detection scheme considers only local features. The method is based on evaluating well-normalized local histograms of image gradient orientations in a dense grid. The basic idea is that local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients and edge directions. Local feature evaluation is realized by dividing the image into square, small in size spatial regions (either overlapping or non-overlapping) and for each region accumulating a local 1-D histogram of gradient directions or edge orientations over the pixels of the region. The combined histogram entries form the representation. Contrast-normalization is also employed to reduce illumination changes, and shadowing. Finally, the detection scheme involves tiling the detection window with a dense grid of descriptors and using the combined feature in a conventional SVM-based window classifier.

A brief evaluation of our system's results verify that using locally normalized histogram of gradient orientations features in a dense grid gives quite satisfactory results for object detection. Its major advantage is that it does not rely on colour or texture alterations as the traditional approaches but is rather based on information extracted from the objects' boundaries directionality. Such features have proven to be robust to noise. Thus, false positives induction introduced due to noise or objects' orientation is eliminated. Further adaptation of the system's parameters such as fine orientation binning, or modification of the SVM kernel and its parameters may refine the systems performance when dealing with a specific image dataset. Our system can also be extended towards measuring human motion through evaluating optical flow and thus deriving more reliable information regarding the human body boundaries and the association between the measured motion and the human motion template. Further research on the feature extraction module so as to derive rotation and scale-invariant features would also aid the built of more robust features. The detection results derived by our system can be subsequently used to feed a data-mining system able to index images illustrating specific scenes and retrieve it by content.

Distributed Video Coding in Wireless Camera Sensor Network

Chanyul Kim

Dublin City University
Dublin, Ireland

Chanyul.Kim@eeng.dcu.ie

Abstract. Distributed video coding is a new paradigm for video compression, based on Slepian and Wolf's and Wyner and Ziv's information theoretic results from the 1970s. It enables low complexity video encoding where the bulk of the computation is shifted to the decoder. In a camera sensor network, multiple cameras will generate signal which need to be sampled, filtered, transmitted, processed, fused, stored, indexed, and summarized as semantic events to allow efficient and effective queries and mining. Camera sensor networks provide a formidable challenge to the underlying infrastructure due to the large computational requirements and the size of captured data. The amount of video generated can consume the same bandwidth as thousands of scalar camera sensors. Our researches goals are to reduce power required for image encoding and to address analysis issues from the hardware platform perspective will allow us to focus on real application requirements such as network deployment and real-time operation. Issues of sensor node self-localization and visual analysis for rudimentary scene understanding are thus required to be addressed.

Motion-based Object Segmentation using Sprites and Anisotropic Diffusion

Andreas Krutz

TU Berlin
Berlin, Germany
krutz@nue.tu-berlin.de

Abstract. Many algorithms have been developed to recognize regions, edges, colour, 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.

Performance-aware Intelligent Data Exchange for Heterogeneous Next Generation Network Applications

Seung-Bum Lee^{1,2}, Gabriel-Miro Muntean¹, and Alan F. Smeaton²

¹Performance Engineering Laboratory (PEL)

²Centre for Digital Video Processing (CDVP)

^{1,2}Dublin City University, Dublin, Ireland,
{sblee, munteang}@eeng.dcu.ie
alan.smeaton@computing.dcu.ie

Abstract. The next generation of networks will involve heterogeneous technologies such as WLAN, 3G, WiMAX, WPAN, etc. Even though they adopt the Internet Protocol (IP) as common infrastructure, the heterogeneity of connectivity provides various challenges in terms of performance, cost-effectiveness and quality of service. Furthermore, there are many computing and consumer electronic devices that enable acquisition and storage of information and most support user-triggered data transfer between them. This has supported user demands for ubiquitous access to data distributed across different devices. Applications on these systems require personalized services, user-centric approaches and simple user-device interaction. To support these, we propose Smart PIN - a novel performance and cost-oriented context-aware personal information network that enables efficient user access to information located on remote devices. It involves cost dependent context and content data pair delivery based on user interest in the content and on existing network constraints.

Comparing Dissimilarity Measures for Image Retrieval

Haiming Liu

Knowledge Media Institute
The Open University
Milton Keynes, UK
h.liu@open.ac.uk

Abstract. Dissimilarity measurement plays a crucial role in content-based image retrieval, where data objects and queries are represented as vectors in high-dimensional content feature spaces. There have been a large number of dissimilarity measures from computational geometry, statistics, information theory and operational research, which can be used to image dissimilarity search. However, only a limited number of them have been adapted to image search so far. Moreover, the performance of a dissimilarity measure may largely depend on different feature spaces. There is lack of a systematic investigation on the applicability and performance of different dissimilarity measures on different feature space for image retrieval. In this paper, we summarize nineteen core dissimilarity measures and theoretically classify them into four categories. A systematic performance comparison is carried out to test these dissimilarity measures with three typical feature spaces (i.e., colour, texture and shape) on the Corel image collection. From our extensive experimental results, we have drawn a number of remarkable observations and insights on distance measurement in image retrieval, which will lay a foundation for developing more effective image search technologies.

The use of Ontologies for Improving Image Retrieval & Annotation

Ainhoa Llorente

Knowledge Media Institute, The Open University,
Walton Hall, MK7 6AA, Milton Keynes,
United Kingdom
A.Llorente@open.ac.uk

Abstract. Nowadays, the number of digital pictures is increasing at an alarming rate both in personal collections and on the Internet due to the falling price of storage devices and the wide availability of digital cameras. Without efficient retrieval methods the search of pictures in large collections is becoming a painstaking work. Most of the traditional image search engines rely on keyword-based annotations because they lack the ability to examine image content. The main goal of our research is to improve image retrieval by means of an efficient annotation of pictures. However, we focus on automated image annotation which is the process of creating a model that assigns visual terms to pictures because manual annotation is a time consuming and inefficient task. Up to now, most of the automated image annotation systems are based on a combination of image analysis and statistical machine learning techniques. The intuition behind our research is whether the underlying information contained in an ontology created from the vocabulary of terms used for the annotation could be effectively used together with the extracted visual information in order to produce more accurate annotations.

Synthesis of Hypermedia Using Ontologies and Rules

Antonino Lo Bue

Medialab Laboratory, ICAR-CNR Sezione di Palermo,
Via Ugo La Malfa 153
90146 Palermo, Italy
ninokeys@hotmail.com
<http://medialab.pa.icar.cnr.it>

Abstract. A hypermedia is a spatio-temporal hypertext, namely a collection of media connected by synchronization and linking relations. Semantic Web technologies allow to associate metadata to media and multimedia documents and to describe univocally their contents as well as relations among them, so opening the door to semantic media retrieval. Integration of rule based expert systems with ontologies offer an enabling technology for intelligent semantic media retrieval and for smart reuse of retrieved contents in structured presentations like hypermedia. HyperJessSyn is a tool which uses logic programming and Semantic Web technologies for synthesizing hypermedia according to descriptions of discourse structure and of presentation layout. A rule-based system in Jess applies inference rules to interpret an OWL graph which expresses semantic and navigation relations among media instances, and production rules to turn media contents descriptions in XML/MPEG-7 format in an XMT-A/MPEG-4 hypermedia script via XSL transformations. The system has been used to produce a hyperguide for virtual visiting a museum. Present service prototype implementation allows synthesizing hypermedia according to just one passive navigation metaphor, namely the *virtual visit* of taxonomy of topics but the chosen architecture is scalable to other metaphors augmenting the KB with appropriate sets of rules.

Semiotics for Flexible Component-Based Multimedia

Michael May

LearningLab DTU,
Technical University of Denmark
mma@dtv.dk

Abstract. Component-based flexibility of interfaces and content presentation is an important issue in multimedia analysis and design in different contexts, e.g. from the personalization of multimedia content over different devices (e.g. mobile), over the deployment of modular learning objects for web-assisted teaching and learning, to the support for flexible (adaptive or tailorable) forms of information presentation in complex safety-critical work domains. Despite the differences in the context-of-use for multimedia content for entertainment, education or work, there are also similar problems with regard to the classification and metadata description of multimedia multi-representational content if it is to support flexibility of use. The aim of my current research is to explore the applicability of semiotics and cognitive semantics (i.e. theories of meaning and signification) as well as the methodology of Formal Concept Analysis (conceptual modelling based on lattice theory) for describing the compositional semantics of multimedia multi-representational objects, and furthermore to suggest extensions to existing multimedia description schemes in order to support flexibility. In the educational context of learning object repositories the relevance of an extended indexing, retrieval and flexibility of multimedia objects is given by the importance for students and educators to know the didactic potential of learning objects, i.e. beyond traditional library metadata and describing the didactic uses of flexible combinations of media types, forms of representation and interactivity as supported by the objects. In the industrial context of information presentation in control rooms (e.g. power plants or chemical process plants) the relevance of flexible component-based multimedia is given by the importance for operators of safe and efficient support for tailorable interfaces, as well as adaptive forms of information presentation and documentation supporting different plant modes and tasks (e.g. monitoring, maintenance, diagnosis). Like the flexible modular “learning objects” of future educational repositories, “smart instruments” in the future control room (or on mobile devices for distributed control) would contain metadata models describing they own contexts-of-use as well as their potential for being redesigned and transformed (“transcoded”) with regard to scale types, media types,

A Study on Multimodal Document Alignment: Bridging the Gap between Textual Documents and Spoken Language

Dalila Mekhaldi

University of Wolverhampton
Wolverhampton, UK
Dalila.mekhaldi@wlv.ac.uk

Abstract. The multimodal alignment framework aims mainly at linking static documents with temporal data, in order to exploit the multi-level structure of documents for indexing multimedia recordings of events. This novel multimodal alignment method is applied on two particular case studies, meetings and lectures. Aligning static documents with the speech transcript of meeting dialogs consists in establishing relationships between them, according to their textual content, at various levels of granularity. The main relationships studied in this work are based on shared thematic content, quotations and references made by speakers to the static documents used during the meeting. In addition to this novel multimodal document alignment method, a complementary research axis has been investigated in this work: the bimodal thematic structuring of meetings. Based on a spatial and temporal clustering of the thematic alignment results, the proposed bimodal segmentation method generates simultaneously the thematic segmentation of the discussed static documents and the meeting dialogs. The satisfactory results obtained within this work prove the performance of our proposed multimodal document alignment solution, and that it supports meetings structuring, searching and browsing. These results highlight also the document usability for accessing multimedia data and its role in multimodal applications.

Continuous Interest-based Exploration of Semantic Information Spaces

Ayman Mounir Moghnieh

UPF, Spain
Ayman.moghnie@upf.edu

Abstract. The spreading of web 2.0 applications and services has revolutionized the structure and content of online information spaces. They effectively support online social interaction and empower users to publish, comment, classify, and share their knowledge and impressions with their peers. The available tools for navigating dynamic information spaces are mainly based on query engines and link lists which support a discrete convergent exploration. As contact with information spaces becomes more dynamic, users tilt toward a continuous and divergent navigation, and justify the need for new exploration paradigms that provide a wider interaction with information. Basing on web 2.0, we attempt to develop an automated digital engine that finds, collects, and structures information spaces and visualize them in accordance with the user's interests and desired actions. Three fundamental objectives currently form the basis of our reasoning: identifying user interests and modeling information spaces accordingly, developing a framework for pleasant and continuous exploration of information spaces, and providing the necessary services to support social interaction and collaboration in the constructivist process of knowledge emergence.

Inter-query Learning for Automatic Incremental Content Abstraction

Donn Morrison, Stéphane Marchand-Maillet, and Eric Bruno

Centre Universitaire Informatique
University of Geneva, Geneva, Switzerland
{donn.morrison, marchand, eric.bruno}@cui.unige.ch
<http://viper.unige.ch/>

Abstract. The semantic gap, recognised as the major hurdle in image retrieval, can be narrowed by tracking patterns of user interaction during queries. This important source of information has largely been ignored, as previous research tends to focus on fully automatic methods using low-level features such as colour, texture, and shape or structured augmentation using ontologies. As with patterns in web traffic analysis, users of image retrieval systems exhibit useful information via their browsing and searching habits.

The inherent limitations of relying solely on low-level features in image retrieval become apparent after a brief appraisal of the available literature. Retrieval systems cannot reliably glean high-level semantics from low-level features due to a lack of image understanding in computer vision. There are many facets to semantic meaning and images can be described in many ways. Subjectivity and intent in photography as well as in retrieval play a critical role. Therefore, we feel it is necessary to place more focus on user interaction in image retrieval and annotation.

In this research, the goal is to semantically describe the images users are searching for, thus facilitating subsequent queries. This involves the propagation of keywords across partially annotated databases using a mixture of long-term learning and low-level image features. By abstracting semantics from images as queries are made, the database can be incrementally annotated. Although the disadvantage is that this requires a cold start, where the database contains no semantic information, the information can accumulate as interaction grows.

If each query session can be described as a concept, then inter-query learning takes examples of relevance feedback over many query sessions and uses these inter-image similarity assessments to learn image-concept relationship. This information can be taken by users making queries through a retrieval system or even simply browsing images.

Classification and Image Annotation for Bridging the Semantic Gap

Zurina Muda

University of Southampton
Southampton, United Kingdom
zm06r@ecs.soton.ac.uk

Abstract. The use of digital images is rapidly increasing in digital archives, community databases, as well as on the Web. This creates new challenges for image management and retrieval and promotes the importance of automatic image classification and annotation research. In general current content-based image retrieval methods are still struggling to deal with the semantic gap between low-level visual features and the high-level abstractions perceived by humans. Manual annotation is typically a difficult and tedious task involving a process of describing the content and context of an image to provide direct access to the semantics. Automatic classification can allocate images or image regions to specific object classes and automatic annotation also aims to add descriptive labels to images. This paper will explore classification and image annotation in bridging the semantic gap and present some related projects which illustrate the advantages of these techniques for image retrieval in the medical and cultural heritage domains.

Personal Video Information Management System

Alev Mutlu

Intelligent Systems Laboratory
Department of Computer Engineering
Middle East Technical University
Ankara, Turkey
mutlu@ceng.metu.edu.tr
<http://www.isl.ceng.metu.edu.tr>

Abstract. In this project, a personal video information management system will be developed. The features that distinguish this project from others are; storing video data according to the content semantics, querying according to the video content, recording new video data from TV broadcasts or from Internet in accordance with user preferences. Users will be able to define their preferences and also the system will learn user preferences from the previously recorded or watched videos. According to these preferences, TV broadcasts and Internet services will be automatically searched, and similar content will be recorded and user will be informed. User will be able to store or delete the new content. Also, the archive will be searched according to video content through intelligent user interfaces.

The main feature of the proposed personal video information management system is that it will be providing user-specific content. Video archive may contain TV recordings, videos obtained from the Internet or personally recorded videos. In addition to be updated by the users, archive content would be updated and personalized by the system agent. New content should be found from various sources and should be personalized. By interacting with the system, user helps his/her profile to be formed. Besides specifying personal preferences (specific kind of movies or songs, favourite artists, etc.) explicitly, the user profile evolves by keeping track of user actions.

One of the main sources for finding new and personal content is the Internet. The web services that provide content will be searched by the proposed system's agent automatically. Without doubt, the system should understand what the multimedia content is about and its semantic characteristics. At this point, semantic web becomes important. With semantic web, the content in the Internet becomes understandable by machines.

Many Internet sites are giving services within the framework of semantic web and personalization. Personalization is especially important for e-commerce sites. Suggestion services (movie or music suggestions; Amazon.com, last.fm) can be used as sources for finding favourable new content according to the user profile. Many video sites (youtube.com, Google video, Yahoo video etc.) provide metadata about the content that they host. Although this is not in the form of advanced metadata standards (MPEG-7, MPEG-21) currently, the information is provided via semantic tags. Setting off from the user profiles, the proposed personal video information system will search the appropriate tags and find new and gripping content.

For videos without semantic content tags, it is needed to extract the metadata automatically. In such cases, the metadata will be extracted from sources like the HTML documents which include the videos or the electronic program guides (EPG) of TV broadcasts, by using natural language processing and text mining methods.

It is necessary to query the built video database efficiently and effectively. Consequently, a user-interface will be developed which allows users to search with natural language (English). Also, the retrieved videos will be summarized with the help of metadata and then the summaries will be presented to the users.

The use of Ontologies in Wrapper Induction

Marek Nekvasil

University of Economics Prague
Czech Republic
nekvasim@vse.cz

Abstract. The vast majority of information resources available on the internet is designed for processing by humans. Formats that are used for visualizing data (e.g. HTML) or handling the users input (e.g. HTML forms) fully agree with this purpose. However, it isn't always appropriate to handle information from available resources by hand. There is an alternative to the manual handling, an automatic handling. The purpose of our work is to bring in an extension of advanced knowledge models, known as ontologies, so that they can be utilized in the process of automated information extraction from the web documents. Major part of it is dedicated to a proposition and derivation of an inference model, based on principles of fuzzy logic, for evaluation of the pattern matches and their combination into templates of typical values of datatype properties of ontology classes. Further we propose a simple method of wrapper induction which is able to utilize the results of automatic document annotation with the proposed templates.

Multilingual Aspects of Spoken Document Retrieval

Andreea Niculescu

University of Twente,
The Netherlands

a.i.niculescu@ewi.utwente.nl
<http://hmi.ewi.utwente.nl/>

Abstract. This research aims to improve multilingual access to multimedia document collections by overcoming several constraints emerging on various levels in the multimedia information retrieval framework, such as speech processing, indexing and user query processing. According to each one of these levels we can identify several areas of interest: at the speech processing level we distinguish between mono-lingual ASR (one model to one language) used for general speech processing and multi-lingual ASR (one model to multiple languages) used for applications like language detection, mixed and/or non-native speech processing; at the indexing level machine translation techniques can be applied either to translate the index table in other languages or to convert the ASR transcript in language used for indexing; at the user query processing machine translation and cross languages information retrieval techniques help to get information across documents in several languages. The context of this research is the MESH project which intends to develop system able to extract, compare and combine content from multiple multimedia news sources, providing end users with a “multimedia mesh” news navigation system (<http://www.mesh-ip.eu>).

Segmentation of Semantic Regions in Personal Image Collections

Stefanie Nowak

Fraunhofer Institute for Digital Mediatechnology,
Langewiesener Str. 22,
98693 Ilmenau,
Germany
nwk@idmt.fraunhofer.de
<http://www.idmt.fraunhofer.de>

Abstract. One challenging goal in the context of content-based image retrieval and semantic metadata enrichment is the automatic extraction of semantics in images. Most problematic is the mapping of the image inherent attributes to a semantic interpretation of the image, the human brain performs autonomously. Recent research proposes two different methods, either solving the problem implicit by applying machine learning algorithms or explicit by utilizing ontologies or knowledge infrastructures to model rules and concepts. Both approaches only achieve satisfying results adapted to narrow domains. Frameworks for semantic image retrieval, combining both techniques, offer a promising perspective regarding broader domains, but still deliver insufficient results.

Our research focuses on expanding a low-level feature based image retrieval application to a higher conceptual stage. Therefore, in the first instance, we apply a graph-based segmentation approach to gather information about regions in images. Through employing concepts borrowed from disciplines like cognitive sciences and cognitive psychology, e.g. salient points or gestalt laws, the retrieved information will be enlarged step by step. Additionally, techniques like latent semantic indexing should help to develop a mapping between image features and semantics. The goal is to achieve a generic framework, allowing for semantic segmentation and labeling of objects in images.

Object Recognition in Large Scale Image Databases Using User Information

Ximena Olivares

Universitat Pompeu Fabra, Spain
ximena.olivares@upf.edu

Abstract. At present time, several online image resources are freely available to every day user where they can upload their own pictures and perform search over pictures. Examples of these online services are Flickr, PicasaWeb Album and Zoomr, among others. These services are used to share photos as albums and also as an online storage media. Since the available capacity that each user has is almost unlimited, the amount of photos available is increasing rapidly. Besides the uploaded images, the users also can insert into the system several amount of data related with the content of the image, like title, description, tags and notes. Other users are free to add more content to the images. This data can be used to give a semantic meaning to the pictures, and use this information to learn about the objects that appear on the pictures. Using all this information, the question will be: How can we learn about the information submitted by the users? And, can we improve the retrieval of the picture system using this knowledge? In this work we combined the information provided by the users and existent techniques used in object recognition to give an idea of how the retrieval can be improved by using the user information.

Service-based Architecture for Personalized and Adaptive Access to the Knowledge in Iconographic Digital Library

Desislava Paneva

Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences
Sofia, Bulgaria
dessi@cc.bas.bg

Abstract. The most essential functionalities for the knowledge delivery systems such as digital libraries are personalization, content adaptation, and changeability of the interface according to the users' individual characteristics, preferences and behaviour in the environment. Our main tasks are users' acquaintance, capturing data about them, user modelling, and development special services for environment customization and quick access to objects/collections of users' interest and need. We develop service-based architecture for realization of personalized and adaptive access to the knowledge in the *Virtual encyclopaedia of Bulgarian iconography* multimedia digital library. It is based on a special IEEE PAPI and IMS LIP-oriented user modelling ontological structure, also called user ontology. The main architectural services aims to provide user's customized access, browsing, searching, grouping and recommending digitised objects and collections in order to realize personalized and adaptive content flow. The services require and trace out data about the preliminary user's knowledge level in the iconographic domain, their object observation style, cognitive goals and interests, preferences about the objects/collections presentation and grouping, physical limitations, used knowledge delivery channels (Web, mobile phone), *etc.*, after that they transform the available iconographical objects into a new personalized form and deliver them to the user. Other services attend to the profile management, user behaviour tracking, *etc.* We use special usage scenarios/instructions defining a wide range of service actions dependent on the user's background, events and non-formal learning situations, knowledge delivery channels, *etc.*

Conditional Random Fields for High-Level Part Correlation Analysis in Images

Giuseppe Passino

Queen Mary, University of London
Mile End Rd
London, E1 4NS, UK
giuseppe.passino@elec.qmul.ac.uk

Abstract. A novel approach to model the semantic knowledge associated to objects detected in images is presented. The model is aimed at the classification of such objects according to contextual information combined to extracted features. The system is based on Conditional Random Fields (CRF), a probabilistic graphical model used in the image classification field to perform inference in problems characterised by a high number of structurally simple features. The advantage of CRF is that the conditional a posteriori probability of the object classes is modelled, thus avoiding problems related to the source modelling and to features independence constraints. The model has been successfully applied for the classification of images based on low-level local features analysis. The challenge is to exploit the advantage of such conditional model to handle high-level, semantically rich object interrelationships among image parts. The difference in the graph structure and in the role of the descriptors discourages a straightforward application of the model to this new type of problem. In this work a first implementation of the system is presented

Graph-based Spatiotemporal Video Segmentation

Marios Phinikettos

Image, Video and Multimedia Systems Lab,
National Technical University of Athens, Greece
finik@image.ntua.gr

Abstract. Segmentation is a key step in video analysis and its results are extensively used for interpreting objects and video scenes. In this white research area, we have started working on a graph-based approach to extract the spatiotemporal regions from a video sequence.

In this approach, each frame is segmented using a fast segmentation algorithm, in particular recursive shortest spanning tree (RSST) and watershed. The segmentation result is then used to create a graph depicting the regions and the relations between them. The method performs iteratively on pairs of frames, while a spatiotemporal graph is used to keep information on previous frames. Then, temporal merging is performed based on overlap criteria between regions, leading to a spatiotemporal graph for the two frames. As more frames are processed, the graph is extended to keep track of all regions in

A future extension is to use spatiotemporal segmentation in order to extract semantic objects. For this purpose, a further step of spatiotemporal region merging will be applied on the resulting graph. We will attempt to recognize or classify objects corresponding to each region during this merging step. Another future extensions is to recognize complex objects described by sub-graphs and consisting of a number of parts of different visual features.

Z-grid-based Probabilistic Retrieval for Copy Detection in a Very Large Video Database

Sébastien Poullot^{1,2}, Olivier Buisson¹, and Michel Crucianu²

¹Institut National de l'Audiovisuel
spoullot(a)ina.fr, obuisson(a)ina.fr,
<http://www.ina.fr>

²Conservatoire National des Arts et Métiers
CEDRIC - Vertigo, 292 rue St Martin
75141 Paris Cedex 03, France
Michel.Crucianu(a)cnam.fr,
<http://cedric.cnam.fr/~crucianm/>

Abstract. Scalability is the key issue in making content-based copy detection (CBCD) methods practical for very large image and video databases. Since a copy is a transformed version of an original document, CBCD involves some form of similarity-based retrieval using as query the description of a potential copy. To outperform an existing competitive method, we introduce here three improvements of this retrieval process: use of a Z-grid for building the index, uniformity-based sorting and adapted partitioning of the components.

We also construct a best deformation model of the descriptors. In this way retrieval becomes significantly faster, enabling us to monitor with a single computer two TV channels against a database of 120,000 hours of video.

Modeling and Annotation of the Dance Media Semantics

K. Rajkumar, B. Ramadoss

Department of Computer Applications
National Institute of Technology, Tiruchirappalli 620015, India
{rajcumarkannan@yahoo.co.in, brama@nitt.edu}

Abstract. Dance data is essentially multimedia by nature consisting of visual (dance pieces), audio (music, tempo, intonation) and textual (lyrics of songs) information. This research represents the semantic based approach to the modelling, annotation, authoring and retrieval of dance media objects. The aim is to develop semantic models to incorporate the various dance video semantics of both Indian and generic dances, to allow dance experts to annotate the semantics, to perform semi-automatic authoring of MPEG-7 and XML instances from these annotations and to handle the dance users' semantic queries. This research introduces representational structures for the semantic (such as dancers, dance movements, context, culture, emotions etc), spatiotemporal (like relationships between the characters and their body parts) and relational features of the dance video; representations for narrative structures such as action and events; and representations and strategies for dance video queries. These representations and strategies form the basis for the prototype systems, *IndVideo* (Indian dance videos), *DanVideo* (generic Dance Videos) and *DMAR* (Dance Media Annotation and Retrieval) which provide facilities for the semi-automatic annotation, authoring and dance semantics retrieval.

Visual Object Detection in Multimedia Sensor Networks

Radha Krishna Ramachandrani

Centre for Digital Video Processing (CDVP)
Dublin City University, Ireland
krishnar@eeng.dcu.ie

Abstract. The availability of inexpensive CMOS cameras, power efficient portable computing platforms and communication modules has fostered significant research in Wireless Multimedia Sensor Networks (WMSNs). Example applications for WMSN's are surveillance and tracking, object identification and event detection. The primary goals of our research are to develop and evaluate algorithms for such applications and also to identify other potential application areas for this emerging technology. We will start by implementing a well defined set of visual analysis algorithms on a reconfigurable platform such as a FPGA and highlight various performance and power constraints. This will enable us to effectively determine the architectural requirements of a prototypical reconfigurable "Media- Mote". Various other issues surrounding the design of a Media-Mote such as ease of application development and deployment, meeting real time constraints for time critical applications, power efficiency and robust transmission will also be addressed.

Targeted Content from Automated Feature Extraction from Video

Mark Restall

Hewlett Packard Laboratories
mark.restall@hp.com

Abstract. HP is continually interested in developing new areas of research with the long term aim of improving existing technologies and developing new market opportunities. With the increase in user generated content, it is becoming increasingly difficult to locate people, objects and activities without re-playing entire length of the material. The analysis and interpretation of features from within video content allows the search and retrieval of the metadata on a frame by frame basis. The focus of the work is around the analysis of the video content to derive a level of understanding so that additional targeted services can be delivered to the end consumer without significant manual intervention.

Active Reading of Audiovisual Documents: From Annotations to Hypervideos

Bertrand Richard

University of Lyon
LIRIS UMR 5205
43 Boulevard du 11 Novembre 1918
69622 VILLEURBANNE CEDEX
FRANCE
Bertrand.richard@liris.cnrs.fr

Abstract. Taking notes or preparing parts of a document for later use, for example in a future presentation, or reflecting about the content while reading a document, is what is called active reading. Applied to audiovisual documents, active reading is often associated with annotating and visualizing a movie or hypervideo. Actually, the process of active reading with respect to audiovisual documents is an important and frequently performed action. Unfortunately, its technological support so far mainly addresses the aspect of annotation ignoring the other aspects, as organization and visualization of the product resulting from active reading. Our work aims at addressing the whole activity of active reading, being situated in the context of the Advène project [<http://liris.cnrs.fr/advène>], which mainly addresses the problem of annotating audiovisual material for creative distribution on the internet. Our main goal is to develop models and tools adapted to the user, that help the active reader during his activity, based on thorough analysis of the actual practices of active reading. This requires an understanding of the different phases of and the operations and mechanisms involved in the different practices of active reading. This analysis results in the definition of criteria to support the reader appropriately. For all the identified processes involved in active reading, we need a sufficiently flexible annotation model (description schema and structure) without restrictions regarding the practices of the reader. Other important aspects regarding this model are the potential share and reuse of such schemas, so that users may easily share their points of view and the description they built on an audiovisual document. The developed model is the basis for our definition and implementation of interfaces that not only allows manipulating this model, but also allows creating, modifying or visualizing new instantiations of the annotation schemas by the reader, during its activity.

Piano Sound Characterization in the Wavelet Domain: Different Properties and Transformations

I. Rojas and F. Blandón

Universidad de Costa Rica
San José, CR
isarojhe@ulatina.ac.cr

Abstract. Daily, the structure of audio signals is being studied by means of different characterizations in a basis of both frequency domain and time domain. A fundamental element that has to be taken into account when characterizing a fixed tuning instrument, as in the case of the piano, is the observation of the harmonic components that build the sound. When applying the Fast Discrete Wavelet Transform (FDWT) one can obtain a better focus when dealing with certain specific frequencies, particularly high ones that are considered key components when characterizing a sound.

Different properties for piano signals have been retrieved in the Wavelet domain, it is possible to characterize an audio signal by using these properties and determine if the signal is constructed by a piano or a different instrument. The most important properties that have been determined in this work are the spectral centroid and the spectral power; they have been determined for different Wavelet transforms, for different intensities and for different ranges among the piano sound.

Modeling of Playlists when Querying for Similar Music

Maria M. Ruxanda

University of Aalborg
Aalborg, Denmark
mmr@cs.aau.dk

Abstract. Due to the explosion of the volumes of multimedia content being broadcasted over the Internet, there is a clear need for systems and techniques that enable so-called similarity search in multimedia content. It is no surprise that systems such as Last.fm, Pandora, and MusicIP that provide similarity search for music are becoming increasingly popular. While systems such as these compute the similarity of music differently, they all deliver playlists to their users. The concept of a playlist encompasses personalized streaming radio stations (as in Last.fm and Pandora) and customized lists of music (as in MusicIP), and it can be regarded as a type of object that captures a sequence of pieces of music such as songs by different artists. Although playlists are now commonly used as the units for exchange of music among Internet users, the notion of a music playlist has yet to be formalized. We thus offer a formalization of the concept of a playlist. The formalization is flexible and works well with the implementation of various music search strategies. As search for similar music implies the querying of a music database, we provide a framework that integrates playlists with database concepts.

Atom Interface – an interactive exploration of tree structures based on the metaphor of electrons, atoms and molecules

Krystian Samp

Digital Enterprise Research Institute
Galway, Ireland
krystian.samp@deri.org

Abstract. Tree structures appear everywhere in many flavours. It can be a hierarchy of species, a classification tree of articles, a directory tree, a well defined taxonomy like DMOZ, facets and their values for a set of resources, categories and sub-categories for a bunch of photos, a menu with options and commands like in many applications etc. Traditional user interfaces built on top of tree structures use all kinds of lists where the items are organised in a vertical or horizontal manner (e.g. MS Windows start menu). They are difficult to navigate and explore. Some of the reasons are non-optimal lengths from starting to destination point and high degree of navigation accuracy required from the user. Moreover, traditional approaches usually obscure some space on the screen even if not used. I propose a novel ATOM interface suitable for tree structures based on the metaphor of electrons, atoms, and molecules. The approach uses so-called compact radial layout to organise items around their parents in a circular fashion. This shortens the distances between nodes and highlights parent-child relationships making it easier to navigate and explore. The layout is capable of handling small and large trees (up to about 700 000 nodes). The metaphor of electrons, atoms and molecules pushes the idea further. It is possible to have several tree structures displayed on the screen at once. The nodes from different trees can be connected like atoms. This facilitates finding relationships between nodes coming from different hierarchies or just different parts of the same hierarchy. Atom Interface is rich in features and supports zooming, rotating, transparency and many others making it more convenient, engaging and appealing than existing solutions.

Intelligent Image Retrieval for personal Photo Books

Philipp Sandhaus

OFFIS - Institute for Information Technology
Oldenburg, Germany
philipp.sandhaus@offis.de

Abstract. Photo books have ever been a means to capture and preserve one's personal experiences and memories. The creation was done within the limits of the individual experiences and knowledge of the user designing the photo book and his or her personal media collection. Nowadays it is possible to digitally design a photo book on a home PC and let it be printed by commercial photo finisher companies. In these days of digital photo book authoring, the task became digital but remained the same in its general idea. The user still has to manually decide for a meaningful selection of photos and arrange them over the pages. With the help of an existing software architecture for metadata extraction and derivation for photos we derive metadata like time, place, person, sharpness and in-/ or outdoor determination by analysing the photo's content and context. On the basis of these metadata we try to determine, what a "good" selection for a photo book is. For this, we analyze photos in existing photo books and try to examine, what kind of pictures are chosen and how and which metadata can be used computationally derive a similar selection. We also believe that such selections are individual for every person. That is why we are, with the help of Machine Learning techniques like artificial neuronal networks or Support Vector Machines, aiming at learning from the personal selection process and iteratively enhance the automatic retrieval of photos for personal photo books.

Image Annotation on a Mobile Camera Phone

Valia Saraydarova

Institute of Information Technologies - BAS
Sofia, Bulgaria
saraydarova@iinf.bas.bg

Abstract. Digital image annotations on a camera phones has an advantage of platform features like easy for user interaction, network connection, communication with other devices or centralized system, programmable processing that can be interpreted at the moment of image capture. The aim of the annotations is to facilitate the image management and search. To use the user interaction 10 groups are selected for rough classification (some of them can be user defined), groups can be combined and each group has domain ontological description. The groups can be assigned with numbers from 0-9 to be easy selected. Other annotation information is retrieved from Bluetooth scanning for surrounding information sources. The Bluetooth information exporters can be mobile devices that present personal public information identification or any culture object identification for example. With this information and other properties like date-time, location, etc. image albums can be automatically generated with supporting textual description of images. Images can be navigated in spatial order or in other relations based on predefined groups semantic.

Support Vector Machines for Classification of Semantic Relations between Nominals

Isabel Segura Bedmar

University Carlos III of Madrid
Madrid, Spain
isegura@inf.uc3m.es

Abstract. In the last years, World Wide Web is becoming a universal and essential knowledge source accessible to everybody. Daily, billions of queries are submitted to web search engines. An excess of information is available and to find concise answers has become a difficult task. The QA research deals with a wide range of questions types such as factoid (asking for the name of a person, a location, the day on which something happened, etc), definition (What / Who is X?), list (one answer containing a list of items), question more complex such as How, Why, and so on. Based on the idea that the answer is just a reformulation of the question, the use of lexical-syntactic patterns could be sufficient to answer simple questions such as factoid questions. However, in some occasions expressions semantically equivalent could not be detected by these lexical-syntactic patterns. In these cases, a deeper understanding of the question and the documents is necessary. Several techniques have been used such as named-entity recognition, coreference resolution, syntactic alternations, word sense disambiguation, detecting relations and so on. Within the topic of the detection of relations, until recently, research has focused primarily on the detection of relations between entities. In addition to this, we believe that the semantic relations between concepts as well as between the verb and its arguments could be beneficial to improve the performance of the QA systems. This work describes an automatic system for classification of the semantic relations between nominals. In this initial version of the system, a set of lexical and semantic features is used to train a support vector machine classifier. The main goal of the system is to classify seven kinds of semantic relations: Cause-Effect, Instrument-Agency, Product-Producer, Origin-Entity, Part-Whole, Theme-Tool and Content-Container, which are deeply described in [1]. This system was evaluated in the 4th International Workshop on Semantic Evaluations, SEMEVAL. For each relation, SEMEVAL provides complete datasets (training and testing) consisting of annotated sentences. In the sentences, the nominal boundaries were explicitly identified and for each nominal, its WordNet sense key was also provided. The best results were achieved in the relations: Instrument-Agency (F=73.7%), Product-Producer (F=73.9%), Part-Whole (F=76.4%). Final scores obtained are comparable to state-of-the-art baselines. The present system represents a part of the on-going research in UC3M (Universidad Carlos III de Madrid) aiming at improving techniques for Information Retrieval and Extraction. Thus, our final goal is to include the results of semantic classification into an integrated architecture IR.

- [1] GIRJU, R., BADULESCU, A. AND MOLDOVAN, D. Learning Semantic Constraints for the Automatic Discovery of Part-Whole Relations, In *Proceedings of the Human Language Technology Conference (HLT-NAACL)*, Edmonton, Canada, May-June (2003).

Semantic Annotation and Retrieval of Video Events using Multimedia Ontologies

Giuseppe Serra and Carlo Torniai

University of Florence
Florence, Italy
{serra,torniai}@dsi.unifi.it

Abstract. Effective usage of multimedia digital libraries has to cope with the problems of building efficient content annotation and retrieval tools. In this paper we describe an architecture for building Multimedia Ontologies that include both linguistic and dynamic visual components. We illustrate the implementation and use of such a multimedia ontology for the soccer video domain is shown. The structure of the ontology itself, together with reasoning to infer knowledge not explicitly asserted in the ontology, can be used to perform higher-level annotation of the clips, generate complex queries that comprise actions and their temporal evolutions and relations, and create extended text commentaries of video sequences.

Using Agent-Based Technologies for Information Gathering

Milan Stankovic and Uros Krcadinac

University of Belgrade
Serbia

milan.stankovic@gmail.com, uros@krcadinac.com

Abstract. Web users express need for getting relevant information from multiple, distributed, and heterogeneous information sources available, and ever-increasing quantities of available data make this task very complicated and annoying. We believe that systems which aim to solve this problem share the common need for agent-based technologies. TALARIA System (The Autonomous Lookup and Report Internet Agent System) is a multi-agent system we developed for academic purposes at the University of Belgrade, Serbia, FON - School of Business Administration (<http://iis.fon.bg.ac.yu/talaria/>). It was built as a solution to the common problem of gathering information from diverse Web sites that do not provide RSS feeds for news tracking. TALARIA integrates information gathering and filtering in the context of supporting a user to manage her/his Web interests. The system provides each user with a personal agent, which periodically monitors the Web sites that the user expressed interest in. The agent informs its user about relevant changes, filtered by assumed user preferences and default relevance factors. Human-agent communication is implemented via email, so that a user can converse with her/his agent in natural language, whereas the agent heuristically interprets concrete instructions from the mail text. Human-like interaction, autonomy-related aspects of this system, and acting on behalf of the user emphasize the usability advantages of this agent-based software, and, we believe, present a new way of using agent-based technologies in order to solve everyday information gathering problems.

Graphemes vs. Phones in a Spanish Keyword Spotting System

Javier Tejedor

HCTLab-Universidad Autonoma de Madrid
Madrid, Spain
javier.tejedor@uam.es

Abstract. Keyword Spotting deals with the search of a reduced set of keywords within large audio repositories. Commonly, the keyword models are composed of sub-word units such as phonemes or phones, but rarely are composed of graphemes due to they do not represent the real sound of the words. In addition to this, in languages such as English, the correspondence between graphemes and phones is very far. However, it is not the case of the Spanish language, where there is a high dependence between graphemes and phones. Modelling graphemes in Keyword Spotting systems avoids the use of grapheme-to-phone conversion rules which many times are very complicated to define and even can vary and also avoid having a big knowledge about the Spanish language. Experiments performed over a Spanish keyword spotting system from a hybrid word/grapheme and word/phone architecture show that the system performance does not change significantly when graphemes are used as sub-word units instead of phones. The word/phone approach gets a Figure-of-Merit (FOM) of 78.34 and the word/grapheme gets a FOM of 77.31. The hybrid architecture consists of an HMM-based keyword spotting module to extract the keywords and a confidence measure based on a lexical access module which, from the monophones or monographemes string retrieved by a phonetic decoding or a grapheme decoding in the time intervals the keyword spotting module proposes a keyword, and from a confusion matrix trained previously from this string, proposes a keyword which best matches with this string. The score retrieved by this module allows us to define a threshold with which the keywords proposed by the keyword spotting module are rejected or accepted. Both the threshold and the confusion matrix were trained from the development set, different from the evaluation set.

Improving IR Applications through the Extraction of Chronological Information: Temporal Expressions Recognition and Normalization

María Teresa Vicente-Díez

Universidad Carlos III de Madrid
Madrid, Spain
teresa.vicente@uc3m.es

Abstract. Dating of contents is crucial in the Information Retrieval (IR) field. Extraction of chronological information allows improving the results obtained by typical IR applications, such as Question Answering (QA) or Automatic Text Summarization. In QA, it is fundamental to solve references that can help finding the answer to temporary questions (“*What year did Cervantes die?*”) or to questions with time restrictions (“*Who was the president of the French Republic in 2005?*”). In order to achieve this goal, techniques considering a temporal reasoning both in query formulation and in answer extraction should be researched. Thus, an accurate recognition of temporal expressions in data sources should be carried out, handling them in an appropriate standard format that allows reasoning without ambiguity. This poster presents a system for temporal expressions recognition and normalization in Spanish to be integrated in a QA system. This approach was presented by the Universidad Carlos III de Madrid to the NIST ACE07 evaluation. A detailed description of the system architecture and modules is presented, as well as the rule-based mechanisms used in recognition and normalization stages, together with its evaluation results.

Single Channel Musical Audio Separation

Beiming Wang

Queen Mary, University of London
London, UK
beiming.wang@elec.qmul.ac.uk

Abstract. In digital music processing, musical recordings we deal with are often composed of multiple instrumental sounds. The aim of our research is to develop a methodology for separating musical audio into streams of individual sound sources. Applications like automatic music transcription, instrument identification can thus be performed on each single sources and greatly improve its accuracy. It also provides the possibilities to manipulate the music in a more flexible way, such as remixing, object coding, special sound effects and so on. Here we propose a novel approach for separating musical instruments in a single-channel audio recording using the Non-negative Matrix Factorization (NMF) algorithm and human interaction. The system has been tested on both artificially mixed audio and real musical recordings with encouraging results. And a Graphical User Interface is also developed to aid this processing.

On the Effect of Cepstral Deconvolution on Chord Identification

Jan Weil

Technische Universität Berlin
Berlin, Germany
weil@nue.tu-berlin.de
<http://www.nue.tu-berlin.de>

Abstract. Automatic identification of chords in music audio signals has been a problem of interest in the context of music information retrieval for several years. The sequence of chords provides an effective way to summarize a piece of music and can, for example, also be used for indexing and retrieval purposes. Instrumental audio signals may be modelled as an excitation that is input to a system of resonators. Cepstral deconvolution (CD) provides a means to separate the excitation from the impulse response of the instrument. It has been shown that, for mono-timbral MIDI-generated audio signals, CD improves the automatic identification of chords. Whether this is also true for real music recordings, however, remains to be explored. We present a system which uses a chromagram representation and an HMM to estimate chord sequences from non-synthesized polyphonic and multi-timbral music audio signals. Moreover, we examine the effectiveness of CD as a pre-processing step within this system.

Generation of Semantic Annotation in form of Natural Language for Multimedia Documents (Images) using Multimedia Ontology and Spatial Relationships Analysis

Lailatul Qadri Zakaria

Southampton University
United Kingdom
lqz06r@ecs.soton.ac.uk

Abstract. A web contains tremendous amount of multimedia contents from television series, film archives, and pictorial collections to sound and music. The use of web to publish those items in great numbers are supported by the improvement in storage and network technologies which allow easy and safe method to deliver multimedia contents in high resolutions and a better steaming manner. These technologies have drawn attentions to web users to develop websites for publishing and sharing multimedia contents among communities such as videos in youtube.com and veoh.com; images in fotopages.com and facebook.com. These websites allow its communities to share, tag and describe their multimedia contents in a simple approach. Multimedia items should be properly annotated and tagged using standard technologies which are already suggested in the semantic web languages. Proper tagging is required in the semantic web layer, whereby it will allow agents or intelligent applications to reuse multimedia metadata to enhance multimedia content based retrieval (CBIR). Therefore, what needed is a tool that can aid users to participate in the semantic annotation and description of multimedia documents. Two main aims in this research are therefore; 1) to propose an approach to assist in analysing multimedia documents (images) using multimedia ontology (such as MPEG7 and VRA) and a specific domain ontology that describe the media domain and 2) to provide a proper annotation by recording result in each multimedia document analysis stage in form of semantic web technologies language. The proposed approach mainly consist of three stages; 1) basic image analysis which will involves processes such as segmentation and finding region of interest based on low level features 2). Identifying object in the image by mapping to the domain ontology and multimedia ontology 3) Describing the content of the image in form of natural language sentence(s) by analysing the spatial relationship between regions or objects. Multimedia ontology (MPEG7) will provide standardization for information representation, while domain specific ontology will cover information within the image domain. At the end of this research, we are hoping to bridge the semantic gap by producing a natural language description of images and a standard representation for image annotation

MediaCampaign¹ – A Multimodal Semantic Analysis System for Advertisement Campaign Detection

Herwig Zeiner

JOANNEUM RESEARCH
Graz, Austria
herwig.zeiner@joanneum.at

Abstract. MediaCampaign's scope is on discovering and inter-relating advertisements and campaigns, i.e. to relate advertisements semantically belonging together, across different countries and different media. The project's main goal is to automate to a large degree the detection and tracking of advertisement campaigns on television, and press. For this purpose we introduce a first prototype of a fully integrated semantic analysis system for detecting new creatives and campaigns. The main scientific challenges in this process involve the detection of new advertisements (called 'creatives' in business jargon) and the interrelation of advertisements belonging semantically together to campaigns. The new creative detection is based on the fusion of a multi-modal semantic analysis for the modalities image, video, audio and text. The knowledge fusion and campaign detection is based on relating metadata, initially acquired during the processing of spots (incoming new advertisements not classified yet). Hence the main workflow steps involved in the MediaCampaign project are data acquisition, multi-modal analysis, creative detection, knowledge fusion and campaign detection, and finally the delivery system for querying and displaying results. This poster describes the system architecture and the main technical workflow of the MediaCampaign prototype system.

¹ MediaCampaign, URL: <http://www.media-campaign.eu/>

Program

Monday (16th of July)

08:00 — 08:45	Registration/Breakfast
08:45 — 09:00	Opening of SSMS 2007
09:00 — 10:30	Audio processing — Gaël Richard
10:30 — 11:00	Tea/Coffee break
11:00 — 12:30	Semantic Web 1— Steffen Staab
12:30 — 13:30	Lunch
13:30 — 15:00	Semantic Web 2 — Steffen Staab
15:00 — 15:30	Tea/Coffee break
15:30 — 17:00	Human Language Technology for Multimedia Analysis — Thierry Declerck
17:00 — 19:00	Reception (partially sponsored by Yahoo! Research Barcelona) & poster session

Tuesday (17th of July)

08:00 — 09:00	Breakfast
09:00 — 10:30	Speech processing — Steve Renals
10:30 — 11:00	Tea/Coffee break
11:00 — 12:30	Image processing — Noel O'Connor
12:30 — 13:30	Lunch
13:30 — 15:00	Video processing — Noel O'Connor
15:00 — 15:30	Tea/Coffee break
15:30 — 17:00	Cross-modal analysis 1 — Alex Hauptmann
17:00 — 18:00	Poster session
19:00 — 23:00	Dinner

Wednesday (18th of July)

08:00 — 09:00	Breakfast
09:00 — 10:30	Cross-modal analysis 2 — Alex Hauptmann
10:30 — 11:00	Tea/Coffee break
11:00 — 12:30	Concept-based video indexing — Cees Snoek
12:30 — 13:30	Lunch
14:00 — 19:00	Boat Cruise, Loch Lomond

Thursday (19th of July)

08:00 — 09:00	Breakfast
09:00 — 10:30	Concept-based video retrieval — Cees Snoek
10:30 — 11:00	Tea/Coffee break
11:00 — 12:30	IR foundations — Keith van Rijsbergen
12:30 — 13:30	Lunch
13:30 — 15:00	Multimedia IR evaluation initiatives — Alan Smeaton
15:00 — 15:30	Tea/Coffee break
15:30 — 17:00	Character retrieval and annotation in multimedia - or "How to find Buffy" — Andrew Zisserman
17:00 — 18:00	Poster session & reception

Friday (20th of July)

08:00 — 09:00	Breakfast
09:00 — 10:30	Music Information Retrieval — Stephen Downie
10:30 — 11:00	Tea/Coffee break
11:00 — 12:30	Multimodal interaction 1— Lynda Hardman
12:30 — 13:30	Lunch
13:30 — 15:00	Multimodal interaction 2 — Lynda Hardman
15:00 — 15:30	Feedback session
15:30 — 16:00	Tea/Coffee break
16:00 — 17:00	Panel discussion
17:00 — 18:00	Reception
