



Multimedia Search Engines

October 8th and 9th, 2007

Location: [Perolles II](#) in the [building 21](#) (see map next page)

With the emergent constitution of multimedia archives, the expansion of infotainment and the democratisation of multimedia acquisition devices, multimedia search engines are becoming crucial for managing both personal data and large multimedia archives. Further, medias (images, videos, sounds, documents, etc.) are often related and thus cross-media mining methods are fundamental not only to enrich multimedia indexing methods but also to extend current querying and search strategies. The major goal of this seminar is to give a state-of-the-art of the current research trends and to bring out the major challenges related to the creation of multimedia search engines.

Monday 8th (room G140)

09.00 – 09.30	Registration (G140)
09.30 - 10.30	TUTORIAL. George Quenot, CLIPS-IMAG, Grenoble, F
	<i>Coffee break</i>
11.00 – 12.00	TUTORIAL. George Quenot, CLIPS-IMAG, Grenoble, F
	<i>Lunch</i>
13.30 – 15.30	TUTORIAL. Marcel Worring, University of Amsterdam, NL
	<i>Coffee break</i>
16.00 – 17.00	TUTORIAL. Marcel Worring, University of Amsterdam, NL
17.00 – 18.00	TALK. Fabio Crestani, University Swiss Italian, CH

Tuesday 9th (002 Joseph Deiss)

09.00 - 10.30	TUTORIAL. Andreas Rauber, Vienna University of Technology, A
	<i>Coffee break</i>
11.00 – 12.00	TUTORIAL. Andreas Rauber, Vienna University of Technology, A
	<i>Lunch</i>
13.30 – 14.30	TALK. Alessandro Vinciarelli, IDIAP, CH
14.30 – 15.30	TALK. Robert van Kommer, Swisscom Innovation, CH
	<i>Coffee break</i>
16.00 – 16.45	TALK. Alex Jaimes, IDIAP, CH
16.45 – 17.15	TALK. Maurizio Rigamonti, University of Fribourg, CH

Organization

- [Denis Lalanne](#), University of Fribourg, Switzerland,
- Co-organizers: [Rolf Ingold](#), University of Fribourg, [Stephane Marchand-Maillet](#), University of Geneva.

Tutorial 1: Video processing for indexing and retrieval (Dr. Georges Quénot, Laboratoire d'Informatique de Grenoble - CNRS)

Indexing the contents of video documents requires a lot of consecutive steps to go from the raw binary contents up to its semantic interpretation. In this course, we will focus on the first stages of the contents analysis which correspond to the extraction of low to intermediate level information. This includes: low level visual feature extraction; shot Boundary Detection, key frame selection and shot or keyframe clustering; camera motion indexing and mobile object segmentation and tracking; story segmentation and video structuring or summarization; speaker identification and emotion indexing. The question of the evaluation of video indexing and retrieval systems will also be addressed and illustrated in the context of the TREC/TRECVID campaigns.



Georges Quénot is Researcher at CNRS (French National Centre for Scientific Research). He has an engineer diploma of the French Polytechnic School (1983) and a PhD in computer science (1988) from the University of Orsay. He is currently with the Multimedia Information Indexing and Retrieval group (MRIM) of the Laboratoire d'Informatique de Grenoble (LIG) where he is responsible for the activities on video indexing and retrieval. His current research activity is about semantic indexing of image and video documents using supervised learning, networks of classifiers and multimodal fusion. He participated since 2001 in the NIST TRECVID evaluations on shot segmentation, story segmentation, concept indexing and search tasks.

Tutorial 2: Semantic Indexing & Retrieval of Video (Dr. Marcel Worring, University of Amsterdam)

The semantic gap between the low level information that can be derived from the visual data and the conceptual view the user has of the same data is a major bottleneck in video retrieval systems. It has dictated that solutions to image and video indexing could only be applied in narrow domains using specific concept detectors, e.g., “sunset” or “face”. This leads to lexica of at most 10-20 concepts. The use of multimodal indexing, advances in machine learning, and the availability of some large, annotated information sources, e.g., the TRECVID benchmark, has paved the way to increase lexicon size by orders of magnitude (now 100 concepts, in a few years 1,000). This brings it within reach of research in ontology engineering, i.e. creating and maintaining large, typically 10,000+ structured sets of shared concepts. When this goal is reached we could search for videos in our home collection or on the web based on their semantic content, we could develop semantic video editing tools, or develop tools that monitor various video sources and trigger alerts based on semantic events. This tutorial lays the foundation for these exciting new horizons. It will cover: Different methods for semantic video indexing; Semantic retrieval; Interactive access to the data; Evaluation of indexing and interactive access in TRECVID; The challenges ahead and how to meet them.



Marcel Worring received the MSc degree (honors) and PhD degree, both in computer science, from the Vrije Universiteit, Amsterdam, The Netherlands, in 1988 and the Universiteit van Amsterdam in 1993, respectively. He is currently an associate professor at the University of Amsterdam. His interests are in multimedia search and systems. He has published over 100 scientific papers and serves on the program committee of several international conferences. He is the chair of the IAPR TC12 on Multimedia and Visual Information Systems. He is general chair of the 2007 ACM International Conference on Image and Video Retrieval in Amsterdam.

Tutorial 3: Music Information Retrieval (Prof. Andreas Rauber, Vienna University of Technology)

In this course we will take a closer look at the various areas, tasks, and methods that together form the field of music information retrieval (MIR). We will start by considering the various types of data that are relevant for MIR activities, ranging from both symbolic as well as acoustic music data, via textual, up to image and video data. This will be followed by a brief overview of the overwhelming number of tasks and challenges in MIR to provide a thorough understanding of the problem domain and the interdisciplinary nature of this domain. The core part of the course will then address a number of selected topics. Specifically, we will focus on various techniques for feature extraction from music, and their utilization for tasks such as retrieval, genre classification, chord detection, and others. We will also analyze and discuss the benefits of combining different modalities, such as textual and acoustic information, as well as the utilization of web information for these tasks. Last, but not least, we will take a closer look at a few applications, such as the PlaySOM and PocketSOM, that assist users in organizing their music collections, creating playlists on desktop computers as well as mobile phones.



Andreas Rauber is Associate Professor at the Department of Software Technology and Interactive Systems (ifs) at the Vienna University of Technology (TU-Wien). He furthermore is head of the iSpaces research group at the eCommerce Competence Center (ec3). He received his MSc and PhD in Computer Science from the Vienna University of Technology in 1997 and 2000, respectively. In 2001 he joined the National Research Council of Italy (CNR) in Pisa as an ERCIM Research Fellow, followed by an ERCIM Research position at the French National Institute for Research in Computer Science and Control (INRIA), at Rocquencourt, France, in 2002. In 1998 he received the ÖGAI Award of the Austrian Society for Artificial Intelligence (ÖGAI), and the Cor-Baayen Award of the European Research Consortium for Informatics and Mathematics (ERCIM) in 2002. His research interests cover the broad scope of digital libraries, including specifically text and music information retrieval and organization, information visualization, as well as data analysis and neural computation.

Content of Talks

"How solid are the foundations of speech-driven information retrieval?" (Pr. Fabio Crestani, Faculty of Informatics, University of Lugano, CH)

Mobility is changing information access applications and information retrieval is not immune to that. In this talk I will discuss some of the issues related to research on mobile information retrieval and in particular to speech-driven information retrieval. I will report on some work at the foundations of this area of research, studying the differences between spoken and written queries and the differences in the perceptions of relevance of spoken and written documents retrieved in response to a query.

Fabio Crestani is a Full Professor at the Faculty of Informatics of the University of Lugano in Switzerland since January 2007. Before that he was Professor of Computer Sciences of the University of Strathclyde (UK) in 2000-06 and Assistant Professor at the University of Padova (Italy) in 1992-97. In between he held research fellowships at the Rutherford Appleton Laboratory (UK), the International Computer Science Institute in Berkeley (USA), and the University of Glasgow (UK). He is an internationally recognised researcher in Information Retrieval, Text Mining and Digital Libraries having published over one hundred refereed publications on both theoretical and experimental aspects.

Making sense of people in audio recordings: social sciences in multimedia analysis. (Dr. Alessandro Vinciarelli, IDIAP)

Multimedia data rarely contain something else than people involved in social interactions. This presentation shows how algorithms inspired by social sciences, namely sociology and social psychology, can be used to extract from multimedia data information that can be difficult, if not impossible, to extract by other means.

A.Vinciarelli is a researcher at IDIAP where he has been active in handwriting recognition, Information Retrieval and Multimedia Analysis. His current activity focuses on the analysis of social interactions in multimedia data. He is author and co-author of one book and more than 30 scientific papers.

Interactive Media Retrieval in Mobile Communication (Dr. Robert van Kommer, Swisscom Innovation)

All-in-one mobile phones have changed our social communication behaviors and infotainment habits. For people on the move, accessing media content represents new challenges and different use cases: on the one hand, mobile phones' display and keyboard are much smaller than those on regular PCs; however, on the other hand, these devices are always on, personalized and carried around. In this context, the following topics are addressed: how to enhance user's access by cross-media indexing and, furthermore, how could the search/retrieval performance be improved with a "human in the loop" personalization algorithm? Both topics will be illustrated through an interactive media application tailored towards mobile user experience.

Besides a PhD in robust speech recognition, Robert van Kommer has filed more than 25 patents in the field of speech recognition, intelligent services and enhanced multimodal interaction. Since 2002, he has initiated and contributed to several multimodal interaction projects. Moreover, he has played a key role in the market launch of Swisscom's 1812 service, the first and fully automated directory assistance service of Switzerland.

Human-Centered Perspectives in Image Retrieval (Dr. Alex Jaimes)

Image retrieval is a human-centered task: images are created by people and ultimately accessed and used by people for human-related activities. In designing image retrieval systems and algorithms, or measuring their performance, it is therefore imperative to consider the conditions that surround both the indexing of image content and the retrieval (e.g., different levels of interpretation, possible search strategies, image uses, etc.). It is also important to consider the role of culture, memory, and personal context. In this talk I will outline important factors in image retrieval from a human-centered perspective, in particular, I will discuss different levels of description, types of users, search strategies, image uses, and issues such as human memory, context, and subjectivity, and their role in image retrieval system development and evaluation.

Alejandro Jaimes is scientific manager and a senior researcher at IDIAP Research Institute, Martigny, Switzerland. His research focuses on developing computer vision techniques that use machine learning, involve humans directly, and are rooted in principles, theories, or techniques from cognitive psychology, the arts, and information sciences. Jaimes received a PhD in electrical engineering from Columbia University. Contact him at alex.jaimes_at_idiap.ch.

Indexing and visual browsing of multimedia collections thanks to links between documents (Maurizio Rigamonti, Fribourg)

The indexing of multimedia data requires semantic information that is sometimes difficult to extract from isolated media in an automatic manner. In this talk, we present a technology that tries to overcome this problem using indexes based on the explicit and implicit correlations existing between multimedia documents. More precisely, the talk presents how the system elicits these relationships and how users can browse a collection of multimedia documents using the discovered links.

Maurizio is a PhD student in the DIVA group of the Department of Informatics of the University of Fribourg. His thesis deals with PDF document analysis, multimedia data indexing and retrieval, and information visualization. He has published several scientific papers in the domains of document engineering, multimedia and HCI.

This workshop is part of the 3eme cycle romand in Informatics
For more information, please refer to the conference website at:
<http://diuf.unifr.ch/3e-cycle/> or <http://www.cuso.ch/3e-cycle/bienvenue.html>