Multimedia Information Retrieval

"New Challenges in Audio Visual Search"

Roelof van Zwol
Yahoo! Research, Spain
roelof@yahoo-inc.com

Stefan Rüger
The Open University, UK
s.rueger@open.ac.uk

Mark Sanderson
University of Sheffield, UK
m.sanderson@shef.ac.uk

Yosi Mass
IBM Research, Israel
yosimass@il.ibm.com

1 Introduction

With the rising popularity of rich media services such as Flickr, YouTube, and Jumpcut, new challenges in large scale multimedia information retrieval have emerged that not only rely on meta-data but on content-based information retrieval combined with the collective knowledge of users and geo-referenced meta-data that is captured during the creation process. For the future, it is envisioned that multimedia search in mobile environments or on P2P networks will take off on a large scale.

This workshop followed four previous SIGIR workshops on multimedia information retrieval (1998, 1999, 2003, 2005), and aimed to address and explore new challenges in multimedia information retrieval by bringing both researchers and practitioners together. We encouraged submission and participation in this workshop not only from the core information retrieval community but also from researchers in databases, multimedia and image processing thus cross-fertilizing to information retrieval research.

In response to the call for papers, 16 submissions were received. Each submission was reviewed by at least three members of the program committee. At this point, we would like to thank the members of the Program Committee for their contribution to this workshop. Based on their recommendations the workshop organizers have selected the 7 best scoring articles for publication and presentation at the workshop. The articles are grouped around three themes: *Querying multimedia*, *Content-based multimedia retrieval*, and *Social media mining and meta-data*. For presenting the article, each author had a time-slot allocated of 25 minutes, of which 5 minutes were reserved for discussion.

Prior to the regular paper sessions, a keynote speech was given by *Dr Wei-Ying Ma* on the topic of "The Challenges and Opportunities of Mining Billions of Web Images for Search and Online Applications".

To further stimulate the interactive character of the workshop we decided to allow each participant of the workshop to propose a subject about which he/she could talk for a maximum of

10 minutes during the *Speakers' Corner*. The output of the Speakers' Corner was used to organize the direction of the *brainstorm session*, which allowed for the discussion of new challenges in Multimedia Information Retrieval.

In the remainder of this report, we summarize the event following the order of the workshop program. All articles, and the full proceedings are available online at `http://www.yr-bcn.es/events/mir2007-workshop/`.

2 Keynote Speech by Dr Wei-Ying Ma, Microsoft Research, Beijing

The keynote speech, entitled "The Challenges and Opportunities of Mining Billions of Web Images for Search and Online Applications", was delivered by Dr Wei-Ying Ma. He is a principal researcher at Microsoft Research Asia, and is based in Beijing. As a Research Area Manager, he leads a team of talented, passionate researchers to advance the state-of-the-art in Web search and data mining.

In his speech, he argued that although content-based image retrieval has been studied for decades, most commercial search engines still rely on text information to index Web images. This is because of the many fundamental limitations in current content-based image retrieval technologies when applied to Web-scale data and a lack of business incentives to rely on image content for online advertising. In this talk, he discussed the technical hurdles that exist when we attempt to build systems to analyze and index billions of Web images based on content. He presented several important Internet applications that have the potential to take off and make significant impacts if we find ways to overcome existing technical hurdles.

3 Paper Session: Querying Multimedia

A Semantic Vector Space for Query by Image Example

Joao Magalhaes, Simon Overell, and Stefan Rüger

Abstract: Content-based image retrieval enables the user to search a database for visually similar images. In these scenarios, the user submits an example that is compared to the images in the database by their low-level characteristics such as colour, texture and shape. While visual similarity is essential for a vast number of applications, there are cases where a user needs to search for semantically similar images. For example, the user might want to find all images depicting bears on a river. This might be quite difficult using only low-level features, but using concept detectors for "bear" and "river" will produce results that are semantically closer to what the user requested. Following this idea, this paper studies a novel paradigm: query by semantic multimedia example. In this setting the user's query is processed at a semantic level: a vector of concept probabilities is inferred for each image and a similarity metric computes the distance between the concept vector of the query and of the concept vectors of the images in database. The system is evaluated with a COREL Stock Photo collection.

A Query Language for Multimedia Content

Jonathan Mamou, Yosi Mass, Michal Shmueli-Scheuer and Benjamin Sznajder
Abstract: The growing amount of digital multimedia data available today and the de-facto MPEG-7 standard for multimedia content description has lead to the requirement of a query language for multimedia content. MPEG-7 is expressed in XML and it defines descriptors of the multimedia content such as audio-visual descriptors, location and time attributes as well as other metadata such as media author, media URI and more. While most search solutions for multimedia today are based on text annotations, having the MPEG-7 standard opens an opportunity for real multimedia content based retrieval. In this paper we propose an IR-style query language for such multimedia content based retrieval that exploits the XML representation of MPEG-7. The query language is an extension of the "XML Fragments" query language that was originally designed as a Query-By-Example for text-only XML collections. We mainly focus on the unique characteristics of Multimedia content which needs to support similarity search query (range search and K-nearest neighbors) and queries on spatio-temporal attributes.

4 Paper Session: Content-based Multimedia Retrieval

Sharing Personal Experiences while Navigating in Physical Spaces

Rui Jesus, Ricardo Dias, Rute Frias, Arnaldo Abrantes, Nuno Correia
Abstract: Social network popularity has been increasing on the Web over the last years. These web sites provide an easy way to share digital data between people with common interests. With this way of exchanging memories of personal experiences, large repositories of data are being created. Indeed personal media, like photos and videos, are excellent vehicles to exchange experiences, however to manage large multimedia databases suitable tools are needed. This paper describes a multimedia retrieval system to access the personal memories shared by the community that visits a point of interest. The multimedia retrieval system uses multimodal information: visual content, GPS data and audio information annotated at capture time. This system can be used via a Web site or when people are visiting the place. The multimedia retrieval system was evaluated using a mobile interface in a cultural heritage site where the personal media can be shared by visitors and can be used to guide the visit. Experimental results are presented to illustrate the effectiveness of the system.

An Optimal Set of Indices for Dynamic Combinations of Metric Spaces

Benjamin Bustos, Nelson Morales
Abstract: A recent trend to improve the effectiveness of similarity queries in multimedia databases is based on dynamic combinations of metric spaces. The efficiency issue when using these dynamic combinations is still an open problem, especially in the case of binary weights. Our solution resorts to the use of a set of indices. We describe a binary linear program that finds the optimal set of indices given space constraints. Because binary linear programming is NP-hard in general, we also develop greedy algorithms that find good set of indices quickly. The solutions returned by the approximation algorithms are very close to the optimal value for the instances where these can be calculated.

Building Detectors to Support Searches on Combined Semantic Concepts

Robin Aly, Djoerd Hiemstra, Roeland Ordelman
Abstract: Bridging the semantic gap is one of the big challenges in multimedia information retrieval. It exists between the extraction of low-level features of a video and its conceptual contents. In order to understand the conceptual content of a video a common approach is building concept detectors. A problem of this approach is that the number of detectors is impossible to determine. This paper presents a set of 8 methods on how to combine two existing concepts into a new one, which occurs when both concepts appear at the same time. The scores for each shot of a video for the combined concept are computed from the output of the underlying detectors. The findings are evaluated on basis of the output of the 101 detectors including a comparison to the theoretical possibility to train a classifier on each combined concept. The precision gains are significant, specially for methods which also consider the chronological surrounding of a shot promising.

5 Paper Session: Social Media Mining and Meta-data

Learning Taxonomies in Large Image Databases

Lokesh Setia, Hans Burkhardt
Abstract: Growing image collections have created a need for effective retrieval mechanisms. Although content-based image retrieval systems have made huge strides in the last decade, they often are not sufficient by themselves. Many databases, such as those at Flickr are augmented by keywords supplied by its users. A big stumbling block however lies in the fact that many keywords are actually similar or occur in common combinations which is not captured by the linear metadata system employed in the databases. This paper proposes a novel algorithm to learn a visual taxonomy for an image database, given only a set of labels and a set of extracted feature vectors for each image. The taxonomy tree could be used to enhance the user search experience in several ways. Encouraging results are reported with experiments performed on a subset of the well known Corel Database.

Multimodal Document Alignment: towards a Fully-indexed Multimedia Archive

Dalila Mekhaldi
Abstract: This paper presents a multimodal document alignment framework, which aims to a cross-indexing of the various documents within multimodal applications (eg, meetings and lectures). These documents might be either present during the event (eg, static documents) or generated after the event (eg, speech transcript). In the current study, three documents types have been considered, static documents, speech transcript and slideshows. The framework presented in this paper is an extension of a previous bimodal alignment framework, between static documents and speech transcript of meetings. Furthermore, several features have been considered and studied in our enhanced alignment framework, such as effect of the TF.IDF metric, the WordNet thesaurus and the noisy speech on our alignment process. The obtained results in this enhanced framework prove that the slideshows are a considerable alternative for speech recordings to temporally index static documents. Moreover, it is shown that the integration of slideshows has improved and reinforced the bimodal alignment of static documents with speech transcript.

6 Speakers' Corner

In response to the call for contributions for the speakers' corner 6 participants stood up and took the opportunity to share their thoughts in front of the workshop participants.

Jussi Karlgren - Use cases in multi-media information access

Jussi Karlgren presented the CHORUS coordination action and argumented for the informed creation of USE CASES as a focus for service design, system development, and algorithm evaluation. He stated that evaluation is dear to our hearts, all of ours, really, if we come to a SIGIR conference. We know about relevance assessment, precision, and recall... but how these arguably bloodless target metrics are translated to not-only-reliable-but-valid-system evaluation exercises is non-trivial. Especially so for systems in new areas such as multimedia retrieval, where the usage scenarios may be less obviously patternable to previous tradition.

Mor Naaman - Social Media: Changing the image of multimedia

Mor Naaman pleaded that the advent of media-sharing sites like Flickr and YouTube has drastically increased the volume of community-contributed multimedia resources available on the web. These collections have a previously unimagined depth and breadth, and have generated new opportunities – and new challenges – to multimedia research. How do we analyze, understand and extract patterns from these new collections? How do we use this analysis to improve current applications and introduce new ones? These questions were discussed and a demo was shown.

Laura Hollink - Bringing semantics into multimedia retrieval

As stated by Laura Hollink, two very different research fields are now slowly growing towards each other: semantic web and (content-based) image retrieval. This combination has the potential to improve retrieval performance and open up new ways of searching. Large bodies of structured knowledge are available on the web. Annotations of images, eg, tags, manual annotations, automatically detected concepts etc, can be linked to these 'ontologies' and take advantage of its semantics: queries can be answered even though no directly matching annotation was found; relations between tags can be used for browsing.

Yosi Mass - Search in audio-visual content using P2P information retrieval

As the coordinator of the SAPIR european project, Yosi Mass discussed that today, Web searches are dominated by search giants such as Google, Yahoo, or MSN that deploy a centralized approach to indexing and utilize text-only indexes enriched by page rank algorithms. Consequently, while it is possible to search for audio-visual content, the search is limited to associated text and metadata annotations. Supporting real content-based, audio-visual search requires media-specific understanding and extremely high CPU utilization, which would not scale in today's centralized solutions. He claimed that large-scale, distributed P2P architectures will make it possible to search audio-visual content using the query-by-example paradigm.

Stefan Rüger - Advertisement

Stefan announced that the Knowledge Media Institute of The Open University is hiring talented researchers to address new challenges in audiovisual search.

Mark Sanderson - "External" meta-data to enhance multimedia search

Mark Sanderson pointed out that the problems of multimedia search can be mitigated by use of externally gathered metadata. Images geo-referenced with GPS data for example can lead to a number of additional improvements. For example, researchers at Berkeley have queried weather databases to determine the conditions at the time and place where a photograph was taken to guess quite accurately what weather conditions are pictured in the photograph.

7 Brainstorm Session

Following the trend that was set during the speakers' corner, the brainstorm session was used to clarify many of the open issues and controversial statements that were brought in by the participants. The brainstorm was highly interactive and according to the participants, a good way to wrap up the workshop, which was perceived by the participants as a highly successful event.

8 Acknowledgments

We would like to thank the members of the program committee for their contributions to the material. We also extend our sincere thanks to SIGIR, to the keynote speaker, all the paper presenters, to the speakers of the Speakers' Corner and all 36 participants, who jointly made this workshop an outstanding workshop.

This workshop was supported by the European Community under the Information Society Technologies (IST) priority of the 6th Framework Programme for R&D projects: SEMEDIA (IST-FP6-045032), PHAROS (IST-FP6-045035), Tripod (IST-FP6-045335), and SAPIR (IST-FP6-045128).