

# High Diversity Transforms Multimedia Information Retrieval into a Cross-Cutting Field

- Report on the 8th Workshop on Multimedia Information Retrieval

James Z. Wang<sup>1</sup>, Nozha Boujemaa<sup>2</sup>, and Yixin Chen<sup>3</sup>

<sup>1</sup> The Pennsylvania State University, University Park, PA, USA.

<sup>2</sup> INRIA Rocquencourt, France. <sup>3</sup> University of Mississippi, University, MS, USA.

Indexing and retrieval of large quantity of multimedia data is a highly challenging and growingly important problem for the computer science research community. Researchers in multimedia, databases, computer vision, machine learning, signal and image processing and statistics have worked on multimedia information retrieval (MIR) for over a decade. A number of significant technological advances have been achieved in this field. Some of the techniques have been applied to application areas such as art image retrieval, biomedical image and video retrieval, education, sensor networks, large-scale online personal and professional photo sharing communities, classification and filtering of images on the Web, scientific content, computer forensics, threat assessment and security applications more generally.

Held in conjunction with the ACM Multimedia Conference in Santa Barbara on October 26-27, 2006, the purpose of the 8th ACM Workshop on MIR [2] was to bring together researchers, developers, and practitioners from academia and industry to showcase latest technological advances and to discuss future possibilities. The volume and quality of the papers submitted to the workshop confirmed the steady growth of the field and the status of MIR as the largest dedicated ACM meeting in this research area. The MIR 2006 received more than 70 completed submissions. The Program Committee selected 10 (14%) contributions to include in the program as regular papers and an additional 15 as poster papers. There were also two invited special sessions. The workshop had more than 70 registered participants. Two outstanding researchers, Donald Geman and Edward Chang, delivered keynote speeches. A highlight of the workshop was a panel consisting of distinguished researchers. They have discussed issues related to the great diversity in multimedia retrieval research.

## RESEARCH PAPERS

This workshop included 25 papers that were presented in two oral sessions and two poster sessions. These papers reported the recent progress on multimedia retrieval, annotation, and summarization.

The MIR community is extending into imagery of higher dimensions and is bridging with computer graphics. The paper ranked the highest by the program committee was authored by Berretti, Del Bimbo, and Pala from University of Florence [1]. They proposed a method for retrieval and recognition of 3D faces represented by iso-geodesic strips. Ohbuchi and Kobayashi investigated unsupervised learning of shape features and applied to the retrieval of 3D shape models.

Indexing, retrieval, and semantic annotation using learning and modeling are the foci of many papers this year. Xie et al. from NASA Jet Propulsion Laboratory and University of Southern California [4] introduced a kernel design approach that has natural applications in content-based image retrieval (CBIR) with relevance feedback. Nguyen et al. presented an active learning approach to identify a dissimilarity metric from user feedback. Bustos and Skopal introduced a method to dynamically combine similarity metric functions using M-tree. Urban and Jose described a framework to learn semantic relations between multimedia objects from user interactions. Kosinov et al. modeled CBIR as a diffusion process in response to a query. Yang and Hauptmann studied temporal consistency of semantic and visual content in video retrieval. Wilkins et al. introduced a fusion model to combine different features and multiple queries in video retrieval. Knees et al. presented a music playlist generation system that exploits artist similarity computed from web retrieval.

Liu et al. explored image annotation based on manifold ranking. Gao and Fan proposed a multi-level image annotation approach using salient objects and concept ontology. Setia et al. investigated image classification based on co-occurrence of local feature descriptors. Fleischman et al. proposed a method to model the temporal information of events in videos.

Video database retrieval continued to be an important area. Takeuchi and Sugimoto integrated personal photo libraries with a video summarization system to model users' preference. Ide et al. described a method to detect monologue scenes in news videos. Smeaton et al. introduced a method to automatically

select shots from an action movie to create a movie preview. Fan et al. explored an approach to match electronic slides with videos for distance learning.

Industrial research papers brought refreshing perspectives to the audience. Jaffe et al. from Yahoo! Research and The Open University of Israel introduced a framework to summarize geo-referenced photographs. The work is related to the popular Flickr online photo sharing community. Zhang et al. from Yahoo! and State University of New York at Binghamton tested the scalability of video search using a large set of video data on the Web. Sun et al. from NTT Labs in Japan proposed a method to identify visual patterns of semantic concepts using image resources provided by the Web. Rebai et al. of INRIA France proposed an interest point detector that produces stable radial symmetry centers. Ando et al. from Tokyo Institute of Technology and NHK in Japan described a scene recognition framework using the context information from consecutive video frames.

Work of international interests was also presented at the workshop. Bezerra and Lima from Brazil presented two feature descriptors for the analysis of soccer videos. Ataer and Duygulu from Turkey applied CBIR to the retrieval of Ottoman documents.

#### KEYNOTE SPEECHES AND SPECIAL SESSIONS

The first keynote speech was addressed by Donald Geman, Professor of Johns Hopkins University, on “Interactive Image Retrieval by Mental Matching”. This talk presents a new visual query paradigm called “mental image retrieval”. Unlike the “query-by-visual-example,” the typical scenario in image retrieval, there is no starting query image example belonging to the database and the matching is “mental” and related to the user. The similarity decision is entirely subjective and user-dependent. An interactive search engine which is based on information theory and statistical inference was discussed. The algorithm involves a Bayesian relevance feedback model and an optimality criterion based on conditional entropy. Performance is measured by the expected number of iterations necessary to match the identity (target search) or the class (category search) of the query. Designing metrics and response models which are consistent with human behavior is essential for achieving practical results with large databases illustrated with art and face collections.

The second keynote speech titled “Unified and Scalable Learning in Multimedia Information Retrieval”, was presented by Edward Chang, Director R&D in Google China and Professor at the University of California at Santa Barbara. Chang presented recent work on a unified learning paradigm (ULP). ULP is motivated by how human being acquires knowledge: learn by being taught (supervised learning), by self-study

(unsupervised learning), by asking questions (active learning), and by being examined for the ability to generalize (reinforcement learning). His work focused on substantially reducing the amount of training. He also presented their proposed algorithmic and data-processing techniques to speed up kernel-based learning.

Two special sessions were organized. First session, entitled “Query Systems for Data Retrieval in Large Personal Image and Video Databases” was organized by C. Dulong et al. Personal media database management needs to deal with complex and mostly person-oriented concepts and lack of extensive annotation. This raises complexity in this domain of algorithms at all levels: vision and low-level feature extraction, ranking and classification, multimodal fusion, long-term and short-term on-line learning and user interface design. The special session focused on query mechanisms capable of supporting image and video retrieval in large personal databases. It has address algorithms for region-based query systems with semantics localization, efficient queries by example, query by concepts, aspects of the on-line learning of user queries, propagation of semantic concepts and the related problems.

The second session addresses “Benchmarking Image and Video Retrieval” and was organized by S. Marchand-Maillet and M. Worring. We have seen the emergence of several image and video retrieval benchmarks. These benchmarks use fixed datasets with well defined units of retrieval and explicit tasks to solve. They allow objective comparison of different methods in a common framework. They also allow objective measures to follow the progress made. For image retrieval various benchmarks have been developed, each with their own merits. Creating a benchmark requires accepted well-defined purposes and sufficient commonalities between proposed systems which is not an easy objective. This session considered different aspects of benchmarking. The data sources range from pure visual information, to visual information with associated free text, and to visual information and associated structured XML data. The evaluation methods consider both task as well as user criteria. The session provided insight in the possibilities and the limitations of the various evaluation campaigns.

#### PANEL ON DIVERSITY

Multimedia information retrieval is a diverse field. Unlike many other research fields, a wide range of data types have been studied by multimedia researchers. These include text, hypertext, audio, graphics, animation, image, video, rich text, spread sheet, presentation slide, combinations of these, and user interaction. It is a challenge to categorize research and development activities in this field. Roughly, one can consider the

current activities residing in a multi-dimensional space, the axes of which can be: system, content, services, use, evaluation, implementation, social/business, and applications.

While continually developing new techniques for multimedia retrieval, researchers in the field have leveraged mature methodologies developed in related fields including databases, information retrieval, signal and image processing, graphics, vision, document imaging, design research, human-computer interaction, computational sciences, machine learning, statistical modeling, data mining, pattern analysis, artificial intelligence, data fusion, social sciences, and domain knowledge for applications. Researchers in the field are rarely graduates of a clearly defined multimedia major. Instead, most of them are from various disciplines: computer science, computer engineering, electrical engineering, mathematics, statistics, information sciences and technology, physics, business, humanities, biology, medicine, ...

Because of the inherent heterogeneity of the field, it is natural that different people take very different approaches in their research. If leveraged, this diversity is a powerful advantage because it helps the field maneuver quickly and smoothly in today's highly dynamic environment. The diversity helps to transform the field into a cross-cutting field. On the other hand, however, this high degree of diversity tends to make breeding grounds for prejudices, biases, or even bigotry which in turn harm the creative potential of the field.

The workshop participants enthusiastically engaged in the discussion during the panel chaired by James Wang of Penn State. The panelists are (1) Nozha Boujemaa of INRIA, Rocquencourt, France; (2) Alberto Del Bimbo of University of Florence, Florence, Italy; (3) Donald Geman, Johns Hopkins University, Baltimore, MD, USA; (4) Alexander G. Hauptmann, Carnegie Mellon University, Pittsburgh, PA, USA; and (5) Jelena Tešić of IBM T. J. Watson Research Center, Hawthorne, NY, USA. These highly visible active researchers from both the academia and the industry opened a discussion on the importance of diversity to the healthy growth of the field. An 8-page paper [3] records their opinions expressed at the panel as well as the personal views of the chair. Five main issues discussed at the panel are:

- *Benchmarking or Not:* There are a number of benchmarking initiatives (e.g., TRECVID for video retrieval) in the field of multimedia information retrieval. Does benchmarking kill innovation? Do existing benchmark initiatives cover diverse multimedia retrieval tasks?
- *Image or Video:* Some may argue that video retrieval is now the focus of the field of multimedia information retrieval. Are there still needs for image

retrieval research? What are the interesting and open research problems in image retrieval? What about other multimedia retrieval problems?

- *Tools or Systems:* Is there a need for a wide spectrum of research areas (e.g., from developing core mathematical tools to developing application systems, from users in the loop to fully automatic) in the field of multimedia retrieval?
- *Trends and Creativity:* What advice would you offer young researchers in the field on choosing research topics? Should they follow current trends? Or should they try to be creative on selecting problems?
- *Academia and Industry:* Academic researchers and industrial researchers often have different goals. In your opinion, what would be the ideal interaction between the two sides?

The workshop participants asked a number of questions after the presentations of the panelists. The chair had to allocate extra time for the panel. Still, some participants were not able to ask their questions.

Clearly, there is no right or wrong answers for some of these questions. For example, on the question regarding benchmarking, some panelists felt that benchmarking was necessary to compare work in the field while others felt that benchmarking could potentially kill diversity and hurt the field. As put in writing by panelist Donald Geman, "I sometimes feel that the celebrated CalTech 101 database, though constructed with the best intentions, has in fact had some negative consequences. Many questions have arisen about whether or not it is 'representative', results are generally poor, papers report incremental gains, and, perhaps most importantly, researchers are criticized for not using it." [3]

The panelists did not agree to each others on some of these questions and no consensus was reached. But in the opinion of the chair, this exactly shows the beauty of the great diversity of a field. The purpose of the panel was exactly to encourage everyone to work together on preserving this diversity and transforming the field into a cross-cutting field.

## REFERENCES

- [1] S. Berretti, A. Del Bimbo, and P. Pala. "Description and Retrieval of 3D Face Models Using Iso-Geodesic Stripes." *Proc. of the 8th Multimedia Information Retrieval Workshop*, pp. 13-22, 2006.
- [2] MIR 2006, <http://riemann.ist.psu.edu/mir2006>.
- [3] J. Z. Wang, N. Boujemaa, A. Del Bimbo, D. Geman, A. Hauptmann, and J. Tešić, "Diversity in Multimedia Information Retrieval Research," *Proc. of the 8th Multimedia Information Retrieval Workshop*, pp. 5-12, 2006.
- [4] H. Xie, V. Andreu, and A. Ortega. "Quantization-Based Probabilistic Feature Modeling for Kernel Design in Content-Based Image Retrieval." *Proc. of the 8th Multimedia Information Retrieval Workshop*, pp. 23-32, 2006.