
Multimedia Information Retrieval — Workshop Report

R Manmatha, CIIR, University of Massachusetts at Amherst, *manmatha@cs.umass.edu*
Stefan Ruger, Computing, Imperial College London, *s.rueger@imperial.ac.uk*
Alex Hauptmann, Computer Science, Carnegie Mellon University, *alex+@cs.cmu.edu*

Abstract. The Multimedia Information Retrieval workshop at SIGIR'05 was well attended with about 28 participants. The workshop format consisted of 8 papers on a variety of different topics from features and representations for image retrieval to a complete retrieval system for media monitoring. The list of papers may be seen at <http://mmis.doc.ic.ac.uk/mmir2005/>, where many of the papers are available online. A substantial amount of time was set aside for discussion of issues in multimedia retrieval, and this report will focus on the discussion.

Discussion on Multimedia Information Retrieval issues. It was recognized that multimedia retrieval is a hard problem. One question that came up was whether there were interesting narrow applications for current technology that could be useful commercially. Medical image retrieval was mentioned as a possible application and a specific system to recognize breast cancer was mentioned although doubts were expressed about how well such systems generalized. Andras Horti in his talk described a media monitoring system that is currently being buildt at the Joanneum in Graz. Specifically, it is a commercial system to track product advertisements and product placements. Companies find it useful to know where their advertisements appear and what kind of impact they are having. He showed some examples of logo and text detection in images to find such mentions of products in television broadcasts and claimed that it ran at 5 frames/sec. Sports videos were mentioned as another application but there was some question of whether one could create publicly available common datasets for such an application. There was considerable discussion of what people really want from multimedia retrieval and the fact that this had not been well studied. One problem is the difficulty in obtaining query logs for studying the needs of real users. The one study in this area for web search used query logs which are not publicly available. For example it was mentioned that the BBC has people annotating images/videos of “cute” animals because there is a demand for such video.

Associated text in the form of closed captions or speech transcripts has been very useful for retrieving video. Other examples include web image search by Google, MSN and Yahoo based on filenames and associated text on web pages. Mark Sanderson presented his paper with Paul Clough on hierarchical clustering of images using captions created by a small number of librarians. An approach that found favour with some researchers was the use of ontologies. One person mentioned that a lot of progress had been achieved by using controlled vocabularies in medicine and law. On the other hand it was argued that, while technical language was less ambiguous, for non-technical language one runs into the semantic primitive problem from the seventies. This lead naturally to discussions of what other kinds of information might be useful. Metadata was cited as one example: it was argued that different kinds of metadata about context about a video or the provenance of an image might be useful. It was pointed out that not all tasks or data had textual annotations or metadata available. Examples include amateur video footage or even raw television footage, for example from the BBC. Some participants discussed the use of specific detectors for some number of concepts which might advance the field of image/video retrieval. One suggestion was a well-defined vocabulary of 5000 concepts although others though 1000 was more realistic. There was mention of Flickr and whether a system like that could prove useful. Finding good annotated and publicly available datasets with well defined queries — beyond the news video domain — was recognized as a challenge given both the technical and legal (copyright) issues. It was also recognized that there are different tasks for which different datasets might be needed. The CLEF image track was suggested as a source of data although there were some objections to it because the images were monochrome. At least one researcher was willing to donate his photographs to the community for some tasks but it was not clear what would be good queries for such a

collection. One suggestion was to create thematic albums or alternatively albums for events. For example, it was suggested that the automatic organization of wedding photographs might be an interesting application which is challenging but still useful. However the photographic collections of individuals may lead to other issues. Another participant described a photo collection they had created as somewhat artificial. He argued that if you know whose photographs they are, then you can tell a lot without doing any analysis, and this is very different from searching web images.

Acknowledgements. Summarising, we feel that the workshop was a worthwhile and timely opportunity to exchange ideas on issues of multimedia retrieval. As organisers we wish to thank all those people who helped to make the workshop successful — including the participants, the authors of the papers, the program committee and the SIGIR'05 organizers. We would like to recognize Alexei Yavlinsky of Imperial College London for his many contributions including publicity and creating and running the workshop website and helping organize the reviewing process. We also wish to thank Jochen Leidner of Edinburgh University for taking notes on the discussion at the workshop. These notes formed the basis for this report.