MediaMill: Exploring News Video Archives based on Learned Semantics

Cees G.M. Snoek, Marcel Worring, Jan van Gemert, Jan-Mark Geusebroek,
Dennis Koelma, Giang P. Nguyen, Ork de Rooij, and Frank Seinstra
ISLA, Informatics Institute
University of Amsterdam
Kruislaan 403, 1098 SJ
Amsterdam, The Netherlands
mediamill@science.uva.nl

ABSTRACT

In this technical demonstration we showcase the MediaMill system. A search engine that facilitates access to news video archives at a semantic level. The core of the system is an unprecedented lexicon of 100 automatically detected semantic concepts. Based on this lexicon we demonstrate how users can obtain highly relevant retrieval results using query-by-concept. In addition, we show how the lexicon of concepts can be exploited for novel applications using advanced semantic visualizations. Several aspects of the MediaMill system are evaluated as part of our TRECVID 2005 efforts.

Categories and Subject Descriptors: H.3.3 Information Storage and Retrieval: Information Search and Retrieval
General Terms: Algorithms, Experimentation, Performance

Keywords: Semantic indexing, video retrieval, information visualization

1. INTRODUCTION

Despite the emergence of commercial video search engines, such as Google [3] and Blinkx [1], multimedia retrieval is by no means a solved problem. In fact, present day video search engines rely mainly on text - in the form of closed captions [3] or transcribed speech [1] - for retrieval. This results in disappointing performance when the visual content is not reflected in the associated text. In addition, when the videos originate from non-English speaking countries, such as China or The Netherlands, querying the content becomes even harder as automatic speech recognition results are much poorer. For videos from these sources, an additional visual analysis potentially yields more robustness. For effective video retrieval there is a need for multimedia analysis.

In this demonstration we present the *MediaMill* system, a semantic video search engine fueled by multimedia analysis. The system lets users interact at a conceptual level, opposed to the traditional data level. We improve upon our previous system [6] in many ways: i.e. novel visual features [2], inclusion of camera motion properties, an extended semantic lexicon, a new browsing interface, support for active learning, and inclusion of a new archive containing 170 hours

of news video from English, Chinese, and Arabic sources. Moreover, all aspects of the system are evaluated on several tasks of the TRECVID 2005 benchmark [5].

2. THE MEDIAMILL SYSTEM

2.1 Semantic Indexing

The central assumption in our semantic indexing architecture is that any broadcast video is the result of an authoring process. When we want to extract semantics from a digital broadcast video this authoring process needs to be reversed. For authoring-driven analysis we proposed the semantic pathfinder, for details see [6].

As an author uses all modalities to convey meaning, the semantic pathfinder starts with multimodal content analysis. First, it extracts features from the visual [2], textual, and auditory modality. Then the architecture exploits supervised machine learning to automatically label segments with semantic concepts. In the second analysis step, the broadcast video is analyzed based on its style properties, again using supervised machine learning for semantic labelling. Finally, semantic concepts are analyzed in context, with the potential to boost index results further. The virtue of the semantic pathfinder is its ability for generic video indexing, by following the optimal path of analysis steps on a per concept basis.

2.2 Semantic Querying

We currently index news video archives with an unprecedented lexicon of 100 semantic concepts, based on the semantic pathfinder architecture. The lexicon contains concepts related to *vehicles*, *sports*, *objects*, and *settings*, as well as specific persons such as *Hu Jintao* and *Tony Blair*. Based on the automatically derived indexes users can explore the video archive using query-by-concept, e.g. retrieving all footage containing *people marching*. Naturally, users can also query on combinations of the 100 semantic concepts. Finally, it is also possible to include camera-specific operations in the query, e.g. find me all *zoom-ins* of *explosions*.

2.3 Semantic Visualizations

Traditional visualizations of video retrieval results are based on so-called storyboards, which essentially visualize query results as a grid of key frames. Within the MediaMill system

Figure 1: Semantic exploration with the MediaMill system. On the left query-by-concept results for US flag, on top the SphereBrowser, and on the right active learning using a semantic cluster-based visualization.

we explore advanced visualizations for novel video retrieval applications.

We take the lexicon of semantic concepts as a basis. It attaches to every video shot, represented as a key frame, a total of 100 concept probabilities. Thus every shot is represented in a 100-dimensional space. The mapping from this space to 2 dimensions optimally preserves semantic relations between shots, while maintaining visibility [4]. Based on the similarity based visualization user can quickly select the relevant examples. We exploit this observation for active learning of novel concepts, beyond the ones already in the lexicon.

Apart from active learning, we also offer a novel browsing environment. It allows to browse a news video in two dimensions. The horizontal dimension is the linear temporal dimension of a news episode. The vertical dimension offers video shots that are conceptually similar to the ones on the horizontal axis. When traversing with the resulting *Sphere-Browser*, users are offered a surprising exploration of news video archives.

3. DEMONSTRATION

We demonstrate semantic exploration of news video archives with the MediaMill system. We will show how a lexicon of 100 concepts can be exploited for effective access to video at a semantic level. In addition, we will exhibit novel appli-

cations that use the correlation of video shots in semantic space. Taken together, the search engine provides users with semantic access to news video archives.

4. ACKNOWLEDGMENTS

5. REFERENCES
[1] Blinkx Video Search, August 2005.
    http://www.blinkx.tv/.
[2] J. M. Geusebroek and A. W. M. Smeulders. A six-stimulus theory for stochastic texture. *Int. J. Comput. Vision*, 62(1/2):7–16, 2005.
[3] Google Video Search, August 2005.
    http://video.google.com/.
[4] G. Nguyen and M. Worring. Similarity based visualization of image collections. In *Int'l Worksh. on Audio-Visual Content and Information Visualization in Digital Libraries*, 2005.
[5] A. Smeaton. Large scale evaluations of multimedia information retrieval: The TRECVid experience. In *CIVR*, volume 3569 of *LNCS*, pages 19–27. Springer-Verlag, 2005.
[6] C. Snoek. *The Authoring Metaphor to Machine Understanding of Multimedia*. PhD thesis, Universiteit van Amsterdam, October 2005.