MediaMill: Semantic Video Search using the RotorBrowser

Ork de Rooij
Intelligent Systems Lab Amsterdam, University of Amsterdam
Kruislaan 403, 1098 SJ Amsterdam, The Netherlands
orooij@science.uva.nl

Cees G. M. Snoek
Intelligent Systems Lab Amsterdam, University of Amsterdam
Kruislaan 403, 1098 SJ Amsterdam, The Netherlands
cgmsnoek@science.uva.nl

Marcel Worring
Intelligent Systems Lab Amsterdam, University of Amsterdam
Kruislaan 403, 1098 SJ Amsterdam, The Netherlands
worring@science.uva.nl

ABSTRACT

In this technical demonstration we showcase the current version of the MediaMill system[1], a search engine that facilitates access to news video archives at a semantic level. The core of the system is a thesaurus of 500 automatically detected semantic concepts. To handle such a large thesaurus in retrieval, an engine is developed which automatically selects a set of relevant concepts based on a textual query, and an novel user interface which uses multi dimensional browsing to visualize the result set.

Categories and Subject Descriptors: H.5.1 Information Interfaces and Presentation: Multimedia Information Systems

General Terms: Algorithms, Design, Performance

Keywords: Video retrieval, information visualization

1. INTRODUCTION

Most commercial video search engines such as Google, Blinkx, and YouTube provide access to their repositories based on text as this is still the easiest way for a user to describe an information need. The indices of these search engines are based on the filename, surrounding text, social tagging, or a transcript. 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 querying the content becomes even harder as automatic speech recognition results are so much poorer. Additional visual analysis yields more robustness. Thus, in video retrieval a recent trend is to learn a lexicon of semantic concepts from multimedia examples and to employ these as entry points in querying the collection.

We present the *MediaMill* semantic video search engine, which combines visual and textual analysis techniques with a semantic lexicon of 500 concepts. These vary from pure format like a detected *split screen*, a style like an *interview*, objects such as a *horse* or a *car*, or events like *airplane take off* or *explosion*.

2. EXPLORATION USING THREADS

The basis for navigation in the MediaMill system is the thread, defined as a linked sequence of camera shots from various videos in some specified order. These automatically computed threads run through the entire dataset, creating a "web" of relatedness.

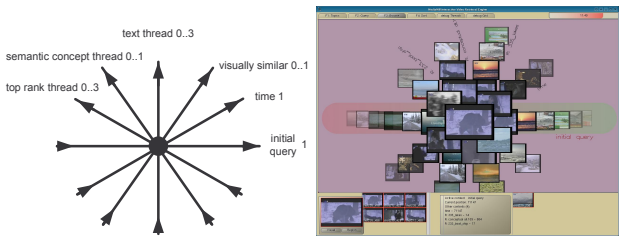

Figure 1: Left: multi thread dimensions for the RotorBrower. The number indicates the times this dimension can be present for any shot. Right: multi thread visualization in the demo.

Navigation using these threads is done as follows. The user specifies one initial query using both textual and semantic components. This generates a list of relevant shots for this query, with the most relevant shot selected as the starting point for exploration. The system computes relevant threads for this shot and displays these to the user as possible navigation directions. An example of this can be seen in figure 1. The user can browse into any of the shown directions by selecting any visible shot. By doing so the selected shot will become the new starting point, and the system will add relevant threads for that shot to the visualization.

We define the following threads. The textual thread contains shots with similar textual annotation. The visual thread contains visually similar shots constructed from low-level visual features. The timeline thread contains all shots ordered by their original timeline. The semantic thread contains semantically equivalent shots constructed from high-level textual and visual features.

3. DEMONSTRATION

We demonstrate exploration through video using a dataset of 180 hours of news video with the MediaMill system.

4. REFERENCES

[1] C. G. M. Snoek, M. Worring, D. C. Koelma, and A. W. M. Smeulders. A learned lexicon-driven paradigm for interactive video retrieval. *IEEE Trans. Multimedia*, 9(2):280–292, 2007.

CIVR'07, July 9–11, 2007, Amsterdam, The Netherlands.
ACM 978-1-59593-733-9/07/0007.