FXPAL *MediaMagic* Video Search System

John Adcock
FX Palo Alto Laboratory Inc
3400 Hillview Ave, Bldg 4
Palo Alto, CA 94304
+1 650 813 7374
adcock@fxpal.com

Matthew Cooper
FX Palo Alto Laboratory Inc
3400 Hillview Ave, Bldg 4
Palo Alto, CA 94304
+1 650 813 7145
cooper@fxpal.com

Francine Chen
FX Palo Alto Laboratory Inc
3400 Hillview Ave, Bldg 4
Palo Alto, CA 94304
+1 650 813 7327
chen@fxpal.com

ABSTRACT
This paper describes FXPAL's interactive video search application, "MediaMagic" [1].

Categories and Subject Descriptors
H.5.1. Information interfaces and presentation: Multimedia information systems – *video.*

General Terms
Algorithms, Design, Experimentation, Human Factors.

Keywords
TRECVID, video search, interactive

1. INTRODUCTION
FXPAL has participated in the TRECVID [2] interactive search task since 2004. In our search application we employ a rich set of redundant visual cues to help the searcher quickly sift through the video collection. A central element of the interface and underlying search engine is a segmentation of the video into stories, which allows the user to quickly navigate and evaluate the relevance of moderately-sized, semantically-related chunks.

2. MULTILEVEL INDEXING
Videos and query results are handled at 3 levels of granularity in our system. In order from longest to shortest they are: program, story, and shot. In a wide variety of video genres, including news and documentaries, shots with wildly varying visual content may be contained within a single, semantically coherent, story. We combine latent semantic analysis [3] of the text transcripts and a similarity-based segmentation technique [4] with the video-based shot boundaries to automatically create a story-level segmentation.

3. RICH PRESENTATION
Figure 1 shows the search interface. Stories are the main units for presenting query results to the user and are summarized and sized according to their retrieval scores on the left. The visual indication of query relevance is repeated in the coloring of the video timeline and individual shot thumbnails on the right side. Shots and stories garner a gray (visited), green (relevant), or red (non-relevant) overlay as they are visited so that the searcher can avoid retracing the same path. In addition to the extensive use of visual cues, the interface extensively employs drag-and-drop for convenience and hot keys for power searchers.

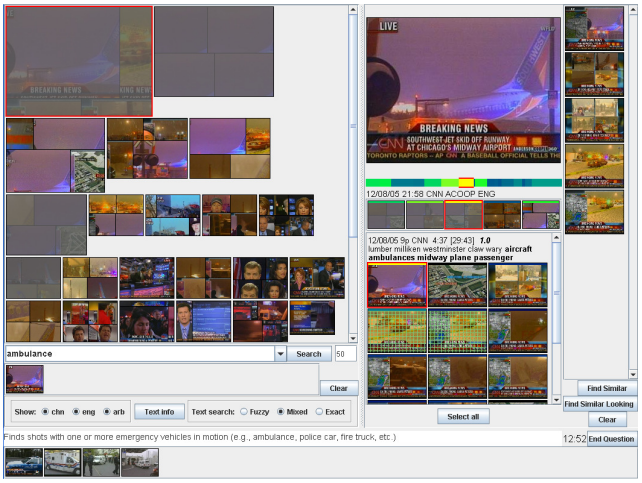

Figure 1: *MediaMagic* interface. Retrieved stories on the left are expanded in the player and shot viewer on the right. The TRECVID query and example media are at the bottom.

4. SEARCH ENGINE
Initial queries are specified in the form of keywords and example images. The searcher-entered text is used for fuzzy text search within a latent semantic space that takes advantage of the automatic story segmentation of the corpus. In addition, any selection of stories or shots can be used as the basis of a new "*find similar*" query based upon the textual, visual, and conceptual content of the selection. In this way the searcher can explore a region of the corpus without explicitly forming a query that defines it.

5. REFERENCES
[1] J. Adcock, M. Cooper, A. Girgensohn, L. Wilcox. Interactive Video Search Using Multilevel Indexing. CIVR 2005, 205-214

[2] TRECVID: TREC video retrieval evaluation Workshop (2001-2006). http://www-nlpir.nist.gov/projects/trecvid/

[3] Berry, M.W., Dumais, S.T., O'Brien, G.W. Using linear algebra for intelligent information retrieval. SIAM Rev. 37 (1995) 573–595

[4] Cooper, M., Foote, J. Scene boundary detection via video self-similarity analysis. In: IEEE Intl. Conf. on Image Processing. (2001) 378–381

CIVR'07, July 9-11, 2007, Amsterdam, The Netherlands.