Supporting Semantic Visual Feature Browsing in Content-based Video Retrieval

Xiangming Mu
University of Wisconsin-Milwaukee
3210 N. Maryland Ave., Milwaukee, WI53211
(414)-229-6039, mux@uwm.edu

ABSTRACT

A new shot level video retrieval system that supports semantic visual features (e.g., car, mountain, and fire) browsing is developed to facilitate content-based retrieval. The video's binary semantic feature vector is utilized to calculate the score of similarity between two shot keyframes. The score is then used to browse the "similar" keyframes in terms of semantic visual features.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Relevance feedback

General Terms

Algorithms, Design

Keywords

Content-based, video browsing, video retrieval, user interface

1. INTRODUCTION

In video retrieval, various browsing technologies are widely supported to augment text based query search, in particular when exact queries are hard to form. This may be because human beings are good at rapidly finding patterns, recognizing objects, generalizing or inferring information from limited data, and making relevance decisions. Browsing usually follows a search operation to pinpoint the correct matches.

For shot level content-based retrieval (where a shot represents a series of consecutive frames with no sudden transition), temporal neighbor browsing is the most common navigation method [1,2]. Temporal neighbor browsing allows users to navigate around the selected sample shot keyframe (a single frame that is representative of the content of a shot) from a text query returns. Potential relevant shots may appear just before or after the sample one due to the asynchronous of the visual content and its related transcript.

One limitation of the temporal neighbor browsing is the limited support for visual objects searching, such as people, car, map, etc.. We propose a video browsing system that supports both temporal neighbor and semantic visual feature browsing.

2. SYSTEM USER INTERFACE

Figure 1 is the main interface of the system. On the top part (part A) a traditional text input field is provided for text-based query. In our system the videos' transcripts were utilized for text-based retrieval.

Figure 1: User interface of the video browsing system

In the middle (part B and C) is the result panel. Video transcript of a selected shot is displayed in Part B, while Part C shows the matched shot keyframes in storyboard style.

At the bottom of the interface (Part D) is a browsing panel where two browsing methods are supported. After performing a text query, users can subsequently proceed with further navigation in this area to find more matches. A tabbed layout is adopted to facilitate users switching between the two browsing methods. "TEMPORAL" tag will lead to temporal neighbor browsing and "FEATURE" tag will go to the semantic visual feature browsing. All the neighboring or similar frames will be displayed in the same size as the sample, which is highlighted in the middle of the filmstrip.

3. REFERENCES

[1] Heesch, D., Howarth, P., Magalhaes, J., May, A., Pickering, M., Yavlinsky, A., and ruger, S. (2004). Video retrieval using search and browsing. In proceedings of TRECVID2004.

[2] Wildemuth, M. B., Yang, M., Hughes, A., Gruss, R., Geisler, G., and Marchionini, G. (2003). Access via features versus access via transcripts: user performance and satisfaction. In proceedings of TRECVID2003.