Content-Based Multi-Functional Video Retrieval System

Huang-Chia Shih and Chung-Lin Huang
Institute of Electrical Engineering National Tsing Hua University HsinChu, Taiwan

ABSTRACT

This paper introduces a content scalable access platform (CSAP) for supportting video scalability and multi-functionality retrieval (MFR). First, we describe a hierarchical retriveal method and shows the availability by two cases study. Second, we represent the methodology of content analysis depending on the different content semantic layer may be applied for MFR. In the experiments, we demonstrate the effectiveness of our method in generating high scalability of video content.

INTORDUCTION

The research interest of the applications of digital videos browsing has increased enormously over the last decade, due to the rapid growth of video streaming on the World-Wide Web. Many scalable video coding technologies utilizing the spatial and temporal features have been developed [1][2]. To access the useful information and reduce the transmission cost, the development of video indexing and retrieval mechnism has become a popular research topic [3][4]. However, the current mechnisms do not provide sufficient analysis of the video content, so that it does not provide the correct information for the user. The user can only fetch what he need from the pre-categorized content through video indexing technology. The user cannot have the overview of the entire video, neither can he realize the representative of the fetched video.

CONTENT-SCALABLE MFR SYSTEM

We need to develop an effective video retrieve scheme that can be used to deliver the video of interests (VOIs) to the client based on the video content rather than the internet bandwidth only.

The hierarchical strategy describes the video by applying the hierarchically structured four-layer frames. As shown in Figure 1, each frame in each layer has a pointer to the related frames in the lower or upper layer. Each frame in the upper three layers has a pointer to the related video clip frame in the bottom layer. The pointers is formulated by the joint appearance probability, and the *ID* in each frame is determined by the occurrence function Ω. Assume an object *A* appears in certain video clip, it is denoted by $\Omega(ID_{o=A})=1$. For instance, the object frame in the object layer has the information of the player ID and a set of pointers to the related action frames in action layer representing his presence in the sport video program (as shown in Fig. 1, the pointer $P_{o\to a}$ indicates the relation between the object frame and the action frame). The action frames have different fidelity to represent the significance of the layer in the video. Since there are other players in the action frame, the fidelity represents its role in the video. The other pointer $P_{o\to c}$ connects the relation between the object frame and the conclusion frame. The conclusion frame records the outcome of the player's action in the video which is either a strike-out or a scoring.

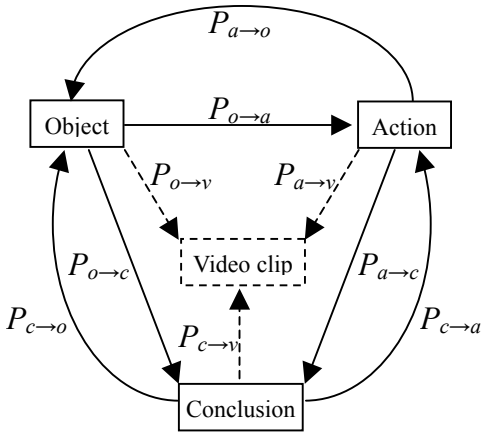


Figure 1. Pointing of MFR for four class elements

Content-scalable is most compatible for different scope's client user, when they want to retrieve the database based on their personal subjectivities. Firstly, to choose the desired video program from categorized video archive list, the key frame and the ID of different semantic layer will be presented to the client. Next, client selects the favorite retrieval conditions in multiple functionality items then feedback the interesting item to MFR system. Before retrieving the video sequence, the system performs two steps: the visual feature analysis and the content analysis (see Figure 2). The video program classification and shot boundary detection also be needed. In addition, the key frame detection id developed to extract the most significant frame that can help the client user to

decide his favorites video. Shot-boundary detection exploit the correlation between the frames belonging to the same shot by calculating consecutive frame dissimilarities in certainty features (e.q., colour, luminance histogram) and finding for major difference values that should hint the start of a new shots. In visual feature analysis, we evaluate some essential features from low-level visual domain including the VOP or foreground region extraction, inference model needed for semantic understanding [5] and NB region identification for conclusion analysis. For content analysis, we have developed three analyzers: role analyzer, action analyzer and conclusion analyzer, to get the content information (i.e., ID_o, ID_a and ID_c). In the following, we will illustrate how to extract the content information.

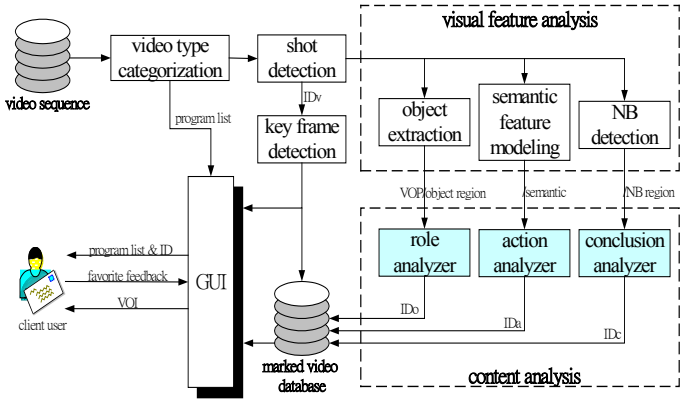


Figure 2. Overview of content-scalable MFR system

EXPERIMENTAL RESULTS

In finally of our experiments, we tried to implement a MFR platform which can be retrieval from one item of any semantic layer by user interesting. In Figure 3, we represented a user interface of our retrieval mechanism. In case, user only choosed a player "*Chen (53)"* as well as they don't care about what is the action and conclusion concept for the *Chen* in 1st program list (i.e., World Cup 2001). After user feedback, the system will response some corresponding results to client user such as right block (i.e., Candidate Shots). In the other hand, user also can choose the left-bottom the list to access the VOI based on the content semantic informations of marked video database. Query by who (player) they want? Where (field) they need? What (action) they mind?

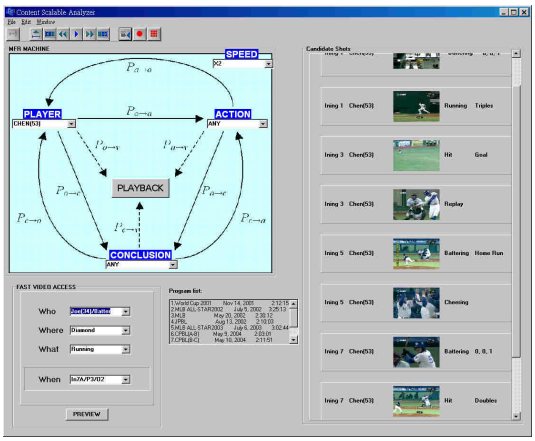


Figure 3. Video retrieval and access GUI for baseball game.

CONCLUSIONS

We are proposed a content-based video retrieval system, not only come off the content-scalable scheme that could be adaptive for different subjective client user but also supporting the video scalability in retrieval application.

REFERENCE

[1] H. Katata, N. Ito and H. Kusao, "Temporal-Scalable Coding Based on Image Content," IEEE Trans. Circuit and System Video Technology, vol. 7, pp. 52-59, Feb. 1997.

[2] M. Domanski, A. Luczak and S Mackowiak, "Spatio-Temporal Scalability for MPEG Video Coding," IEEE Trans. Circuit and System Video Technology, vol. 10, no. 7, Oct. 2000.

[3] D. Zhong and S. F. Chang, "Spatio-temporal video search using the object-based video representation," IEEE ICIP Santa Barbara, Oct. 1997.

[4] M. R. Naphade, I. Kozintsev, and T. S. Huang, "A Factor Graph frame work for demantic indexing," IEEE Trans. on CAS for VT, pp.40-52, Jan. 2002.

[5] Huang-Chia Shih and Chung-Lin Huang, "DETECTION OF THE HIGHLIGHTS IN BASEBALL VIDEO PROGRAM," in *Proc.* IEEE-ICME 2004.