

A Framework for Customizable Sports Video Management and Retrieval

Dian Tjondronegoro, Yi-Ping Phoebe Chen, and Binh Pham

Centre for Information Technology Innovation
Faculty of Information Technology
Queensland University of Technology, Brisbane, Australia
{ d.tjondronegoro, p.chen, b.pham }@qut.edu.au

Abstract. Several domain specific approaches for sports video management have shown the benefits of integrating low- and high- level video contents in supporting more robust retrieval. However, there are very little work has shown how to integrate them in order to support different types of sports. In this paper, we firstly propose a framework for customizable video management system which allows the system to detect the type of video to be indexed, so that appropriate tools can be used to extract the key segments. It is also customizable because the system manages user preferences and usage history to make the system supports specific requirements. Secondly, we will show how the extracted key segments can be summarised using standard descriptions of MPEG-7 in a hierarchical scheme which is potentially easy to share between users. Thirdly, we have developed and tested some queries which show that XQuery provides a powerful language for our video management's retrieval.

1 Introduction

With the ongoing advance of computer technology, the requirements for using video data, such as large storage capacity, fast processing power, and broadband networking, have become inexpensive. This phenomenon triggers fast growth of video data for a vast range of applications, such as sport, news and security. Unfortunately, the effectiveness of video usage is still very limited due to the unavailability of a single and complete technology which can fully capture the semantic content of video and index video parts according to their contents, so that users can intuitively retrieve specific video segments. Thus, content-based Video Management Systems (VDMS) has become an active research topic for many researchers from both academia and industries.

To enable effective usage of video, the key segments of a video stream need to be indexed (i.e. extracted and labelled according to their contents) to support content-based retrieval. Since video data is typically long and any of its arbitrary frames may contain important information which can be of users' interest, the process of developing content-based VDMS requires the following set of complex procedures. The first step is *video segmentation* where the key-segments (or highlights) of the video are identified. At the same time, *content analysis* is used to analyse the content of each segments, which works in conjunction with *video labelling*, where video segments are labelled according to their contents. The last step is to *structure the*

video descriptions for efficient retrieval, such as by constructing some hierarchical summaries of the exciting segments of sport videos.

These processes rely heavily on accurate and complete understanding of the semantic contents of video data. However, computers are yet as intelligent as human to fully understand the complexity of semantic concepts, while manual work is too expensive and often biased by subjectivity. Thus, computers depend on the computerised semantic of video, such as by examining the characteristics of its audio and moving pictures via signal analysis algorithms. This process is complex since different video applications have specific features and characteristics. For example the audio components of sport video typically consist of background noise, crowd cheers and commentator's voice. To overcome these limitations, we aim to demonstrate in this paper the significance of integrating *domain specific approach* to improve the accuracy and effectiveness of content extraction process, as well as allowing robust retrievals according user/application requirements. The integrated system can be customised to suit other domains, particularly for the video domains that have similar characteristics. This paper will present a framework for customizable VDMS by focusing on the three most important components: key-segment extraction, content descriptions, and retrieval. Our experimental work on segmentation will be based on indexing goal segments in soccer videos, while using the currently active-evolving technologies of MPEG-7 for our video content descriptions and XQuery for retrieval. Due to their rapid evolutions, this paper only reflects upon the current status of MPEG-7 and XQuery.

2 Related Work

State of the art content-based VDMS can be categorized into two main approaches based on the types of data which can be extracted (see [1-3] for review). The first approach allows users to retrieve video according to its high-level semantic contents. However, the main limitation is that manual work for annotating the video data is expensive and often subjective. The second approach utilizes low level features comparison, which allows users to query video by example (i.e. using audio or visual samples). Although this approach uses automatically extracted features, users would not be able to retrieve video based on the high-level concepts. For this reason, many researchers have tried to show the benefits of domain specific approach in closing up the main gaps between the high-level and low-level features of video. The followings are some examples of domain-specific video management approaches which are applied in sport videos, such as tennis, soccer, and basketball.

In tennis domain, Zhong and Chang [4] summarised the temporal structure of live broadcasted tennis video by detecting the recurrent event boundaries, such as serving views which consist of unique visual cues, such as colour, motion and object layout. Such views can be automatically detected by using supervised learning and domain-specific rules. Thus, users can browse and query based on these recurrent events. In addition to serve views, Miyamori et al [5] segmented tennis video by extracting: court-net lines, player-ball position, and player behaviour. Each court-net line is detected by connecting two feature points at both sides, while the positions of players and ball are captured by adaptive template matching. Players' behaviours are