12. Video Event Mining via Multimodal Content Analysis and Classification

Min Chen, Shu-Ching Chen, Mei-Ling Shyu, and Chengcui Zhang

Summary. As digital video data become more and more pervasive, the issue of mining information from video data becomes increasingly important. In this chapter, we present an effective multimedia data mining framework for event mining with its application in the automatic extraction of goal events in soccer videos. The extracted goal events can be used for high-level indexing and selective browsing of soccer videos. The proposed multimedia data mining framework first analyzes the soccer videos by using multimodal features (visual and audio features). Then the data prefiltering step is performed on raw video features with the aid of domain knowledge, and the cleaned data are used as the input data in the data mining process using the Nearest Neighbor with Generalization (NNG) scheme, a generalized Instance-Based Learning (IBL) mechanism. The proposed framework fully exploits the rich semantic information contained in visual and audio features for soccer video data, and incorporates a data mining process for effective detection of soccer goal events. This framework has been tested using soccer videos with different styles as produced by different broadcasters. The results are promising and can provide a good basis for analyzing the high-level structure of video content.

12.1 Introduction

With the increasing amount of digital video data, mining information from video data for efficient searching and content browsing in a time-efficient manner becomes increasingly important. Motivated by the strong interest of automatic annotation of the large amount of live or archived sports videos from broadcasters, research toward the automatic detection and recognition of events in sports video data has attracted a lot of attention in recent years. Soccer video analysis and events/highlights extraction are probably the most popular topics in this research area.

The major challenges in soccer event detection lie in the following four aspects. First, the value of sports video drops significantly after a short period of time [5], which poses the requirement of real-time (or close to real-time) processing. Second, unlike some of the other sports, such as baseball, tennis, etc., where the presence of canonical scenes (e.g., the pitching scene in baseball, the serve scene in tennis, etc.) could greatly simplify the technical challenges, soccer videos possess a relatively
loose structure. Third, the important video segments (events or highlights) in a sports video constitute only a minor portion of the whole data set. Consequently, the limited number of training data points increases the difficulties in detecting these so-called rare events, especially in the present of noisy data introduced during the production process. Last, but not least, the video data obtained from various sources might be inconsistent due to different production styles and postproduction effects. In other words, although some basic production rules might apply, the overall presentations vary greatly.

In the literature, many researches have been devoted to address these issues from the media content analysis [3, 12, 14, 25, 26, 33–35, 37] to the supervised classification techniques [1, 20, 21, 27, 30, 31]. An overview of the related work will be detailed in Section 12.2. However, few approaches possess the capabilities of tackling all the above-mentioned challenges. In response to these issues, in this paper, an effective multimedia data mining framework is proposed with its application on the soccer goal event detection, which seamlessly integrates the multimodal content analysis and the Nearest Neighbor with Generalization (NNG) scheme which is a generalized Instance-Based Learning (IBL) mechanism. Here, an event is defined in the shot level as the shot is widely regarded as a self-contained unit with an unbroken sequence of frames taken from one camera.

In our proposed framework, multiple cues from different modalities including audio and visual features are fully exploited and used to capture the semantic structure of soccer goal events. Then the NNG scheme is applied for event detection. Currently, most existing classification techniques adopted in the event detection area are called model-based approaches as they compute the global approximation (or called model) of the target classification function, which is then used to classify the unseen testing data. In contrast, the IBL mechanism is called lazy method in the sense that the generalization of the observed (training) data delays until each new query instance is encountered [24]. Therefore, it can use the query instance for selecting a local approximation to the target classification function [13] each time when a query instance is given. The importance of incorporating the query instance lies in the fact that the current production style is one of the key factors in determining the pattern of a targeted event. Therefore, by adopting the IBL mechanism, we direct our focus on the instances themselves rather than on the rules that govern their attribute values. In addition, in response to the requirements of real-time processing and rare event detection, a data prefiltering step is integrated in the IBL mechanism to perform on the raw video features with the aid of the specific domain knowledge. We have evaluated the performance of the proposed framework by using a large amount of soccer video data with different styles and different broadcasters. The experimental results demonstrate the effectiveness and the generality of our proposed framework.

The contributions of the proposed framework are summarized as follows:

- First, an advanced video shot detection method is adopted in this work, which can not only output the shot boundaries, but also generate some important visual