Chapter 16
Probabilistic Graphical Models for Sports Video Mining

Guoliang Fan
Oklahoma State University, USA

Yi Ding
Oklahoma State University, USA

ABSTRACT

Semantic analysis is an active and interesting research topic in the field of sports video mining. In this chapter, the authors present a multi-level video semantic analysis framework that is featured by hybrid generative-discriminative probabilistic graphical models. A three-layer semantic space is proposed, by which the semantic video analysis is cast into two inter-related inference problems defined at different semantic levels. In the first stage, a multi-channel segmental hidden Markov model (MCSHMM) is developed to jointly detect multiple co-existent mid-level keywords from low-level visual features, which can serve as building blocks for high-level semantics. In the second stage, authors propose the auxiliary segmentation conditional random fields (ASCRFs) to discover the game flow from multi-channel key-words, which provides a unified semantic representation for both event and structure analysis. The use of hybrid generative-discriminative approaches in two different stages is proved to be effective and appropriate for multi-level semantic analysis in sports video. The experimental results from a set of American football video data demonstrate that the proposed framework offers superior results compared with other traditional machine learning-based video mining approaches.

DOI: 10.4018/978-1-4666-1806-0.ch016
INTRODUCTION

The goal of video mining is to discover knowledge, patterns, and events in the video data stored either in databases, data warehouses, or other online repositories (Chang, 2002; Mei, Ma, Zhou, Ma, & Zhang, 2005). Its benefits range from efficient video indexing, browsing and summarization to content-based or semantic-based video access and retrieval. Driven by the ever increasing need of numerous multimedia and online database applications, there is a growing need of efficient video search and indexing tools that allow us to quickly find the video data of particular interest from a large amount of video data. In this chapter, we are mainly focused on sports video mining due to its great commercial value and wide popularity among consumers. Current research on sports video analysis signifies two major trends, i.e., event-based analysis that aims at detecting events of special interests, e.g., goals and replays (Shyu, Xie, Chen, & Chen, 2008), and structure-based analysis where the goal to discover the overall semantic evolution of a game (Gupta, Srinivasan, Jianbo, & Davis, 2009). Just like the index and the table of contents (TOC) in a book both of which are useful tools for a reader, event-based analysis and structure-based analysis have complementary nature for semantic video understanding, but they are usually addressed separately in sport video mining research as two different aspects. Our goal is to develop a new video mining paradigm that is able to support both of them. Specifically, we consider the American football video as a case study in our research, where we assume that the video stream has been pre-segmented into a set of consecutive shots each of which corresponds to a specific play in a game.

The semantic evolution of a complete American football game has an interesting multi-level structure as “game flow → drives → plays ’’, as shown in Figure 1, which is also shared by many field-based sports, such as soccer and hockey. The game flow summarizes the overall game progression by a series of labeled drives and each drive includes a set of annotated plays. It supports both event-based (labeled drives and annotated plays) and structure-based (plays → drives → game flow) video analysis. Specifically, we consider three possible labels for a drive: scored, non-scored and turn-over, and we also specify three kinds of key-words for each play, play type (what happened?), camera view (where did it happen?), and possession (who was on offense/defense?) Consequently, we are interested two issues in this research: (1) how to annotate a play by a few mid-level key-words that can be used for high-level game flow analysis; (2) how to group a set of annotated plays into a series of labeled drives each of which has a variable number of plays. For simplicity, these two issues are referred to as keyword detection and game flow analysis in this chapter. Although the two issues are related, they are confronted by different challenges and require different approaches.

Machine learning, especially the probabilistic graphical modeling, is often considered as one of the most effective approaches for video mining (Cheng & Hsu, 2006; Liu, Carbonell, Weigele, & Gopalakarishnany, 2006). However, due to imperfect mid-level keyword detection or limited low-level feature extraction, the missing data problem (i.e., some plays with incomplete keywords) is prevalent in reality for high-level semantic analysis. Unfortunately, this issue is often under-treated in most video mining research. Our work is inspired by three machine learning techniques. One is the segmental hidden Markov models (SHMMs) (Gales & Young, 1993), which can handle short range temporal dependencies from observations that cannot be done in traditional HMMs, another is the Segmentation Condition Random field (SCRF) (Liu, et al., 2006) used in protein fold recognition to capture long range dependencies in both observations and latent states, and the other is the marginalization techniques for missing data handling in (Farhangfar, Kurgan, & Dy, 2008). Particularly, we extend the