Chapter 9
Event Detection in Sports Video Based on Generative–Discriminative Models

Guoliang Fan
Oklahoma State University, USA

Yi Ding
Oklahoma State University, USA

ABSTRACT
Semantic event detection is an active and interesting research topic in the field of video mining. The major challenge is the semantic gap between low-level features and high-level semantics. In this chapter, we will advance a new sports video mining framework where a hybrid generative-discriminative approach is used for event detection. Specifically, we propose a three-layer semantic space by which event detection is converted into two inter-related statistical inference procedures that involve semantic analysis at different levels. The first is to infer the mid-level semantic structures from the low-level visual features via generative models, which can serve as building blocks of high-level semantic analysis. The second is to detect high-level semantics from mid-level semantic structures using discriminative models, which are of direct interests to users. In this framework we can explicitly represent and detect semantics at different levels. The use of generative and discriminative approaches in two different stages is proved to be effective and appropriate for event detection in sports video. The experimental results from a set of American football video data demonstrate that the proposed framework offers promising results compared with traditional approaches.

DOI: 10.4018/978-1-60960-024-2.ch009
Event Detection in Sports Video Based on Generative-Discriminative Models
classified into two major categories: scripted and non-scripted (Xiong, Zhou, Tian, Rui, & Huang, 2006), which are usually associated with different video mining tasks. Scripted videos (e.g., news and movies) are produced or edited according to a pre-defined script or plan, for which we can build a Table-of-Content (TOC) to facilitate the viewing or editing of the video data (Rui, Huang, & Mehrotra, 1998). In Non-scripted videos (e.g., sports), events usually occur spontaneously in a relatively fixed setting, such as meetings, sports, and surveillances. Therefore, how to detecting the highlights or events of interests is of great interest for non-scripted videos.

In our research, we focus on sports video and use the American football video as a case study. Sports video mining has been widely studied due to its great commercial value (L. Duan, Xu, Tian, Xu, & Jin, 2005; Gong, Sin, Chuan, Zhang, & Sakauchi, 1995; Kokaram, et al., 2006; Xie, Chang, Divakaran, & Sun, 2002). Although the sports video is considered non-scripted, they usually have a relatively well-defined structure (such as the field scene) or repetitive patterns (such as a certain play type), which could help us enhance its “scriptedness” and develop effective tools for retrieval, searching, browsing and indexing. Currently, there are two kinds of approaches for sports video mining: structure-based (Kokaram, et al., 2006; Xie, Chang, Divakaran, & Sun, 2004) and event-based (Assfalg, Bertini, Colombo, Bimbo, & Nunziati, 2003; T. Wang, et al., 2006). The former one uses either supervised or unsupervised learning methods to recognize some basic semantic structures (such as the canonical view in a baseball game or the play/break in a soccer game). This can serve as an intermediate representation to support semantics-oriented video retrieval, but usually cannot deliver high-level semantics directly. The latter one provides a better understanding of the video content by detecting and extracting the events-of-interest or highlights, which could be very specific and task-dependent and usually requires sufficient and representative training data.

Because these two approaches are complementary in nature, researchers have investigated how to integrate both of them in one unified computational framework. For example, a mid-level representation framework was proposed for semantic sports video analysis involving both temporal structures and events hierarchy (L. Y. Duan, Xu, Chua, Q. Tian, & Xu, 2003) and a mosaic-based generic scene representation was developed from video shots and used to mine both events and structures (Mei, et al., 2005). The advantage of this kind of video representation lies in its expandability and openness to support versatile and flexible video mining tasks.

Inspired by previous efforts, the goal of our research is to develop an integrated video mining framework that offers both structure analysis (at the mid-level) and event analysis (at the high-level) in a systematic way. Specifically, we promote a collaborative use of two different statistical models, i.e., generative and discriminative, for event detection in sports video. The major challenge of our research is the semantic gap between the understandable high-level semantics and computable low-level features. In this chapter, we will propose a hybrid generative-discriminative model based approach that supports explicit semantic modeling and direct semantic computing and bridges the semantic gap via two inference processes. Specifically, we addressed the following three technical issues: (1) to represent sports video data by a three-layer semantic space that involves low-level visual features, mid-level semantic structures, and high-level semantics; (2) to formulate the issue of semantic event detection as two related statistical inference problems, i.e., the inference from low-level visual features to mid-level semantic structures and the inference from the mid-level semantic structure to the high-level semantic events; (3) to employ appropriate statistical tools to tackle the two inference problems of different nature, i.e., the generative models for the first one and the discriminative models for the second one.
Related Content

Grid-Based Visualization and its Medical Applications
Yingjun Qiu, Youbing Zhao and Jiaoying Shi (2008). *User Centered Design for Medical Visualization* (pp. 25-57).
[www.igi-global.com/chapter/grid-based-visualization-its-medical/30623?camid=4v1a](www.igi-global.com/chapter/grid-based-visualization-its-medical/30623?camid=4v1a)

2D and 3D Visual Attention for Computer Vision: Concepts, Measurement, and Modeling
[www.igi-global.com/chapter/2d-and-3d-visual-attention-for-computer-vision/139000?camid=4v1a](www.igi-global.com/chapter/2d-and-3d-visual-attention-for-computer-vision/139000?camid=4v1a)

A Recurrent Probabilistic Neural Network for EMG Pattern Recognition
[www.igi-global.com/chapter/recurrent-probabilistic-neural-network-emg/28039?camid=4v1a](www.igi-global.com/chapter/recurrent-probabilistic-neural-network-emg/28039?camid=4v1a)

Some Fuzzy Tools for Evaluation of Computer Vision Algorithms
[www.igi-global.com/article/some-fuzzy-tools-for-evaluation-of-computer-vision-algorithms/201460?camid=4v1a](www.igi-global.com/article/some-fuzzy-tools-for-evaluation-of-computer-vision-algorithms/201460?camid=4v1a)