HIDDEN MARKOV MODEL BASED EVENTS DETECTION IN SOCCER VIDEO

Guoying Jin Linmi Tao Xinghua Sun Guangyou Xu

Department of Computer Science and Technology
Tsinghua University, Beijing 100084 P.R.China
jingy97@mails.tsinghua.edu.cn, {linmi, sun_linux, xgy-dcs}@tsinghua.edu.cn

ABSTRACT

This paper presents an algorithm based on hidden Markov models for cues fusion and events inference in soccer video. Four events, shoot, foul, offside and normal playing, are defined to be detected. The states of the events are employed to modeled the observations of five cues, which are extracted from shot sequences directly. The experimental results show the algorithm is effective and robust in inferring events from roughly extracted cues.

1. INTRODUCTION

One of the most challenging issues in content-based retrieval is to tackle the semantic gap between the low-level features and the high-level semantics. Event analysis has been considered as a promising approach to bridge this gap. In this approach, events with semantic meanings are inferred from fusing relevant cues, which are extracted from low level features under the guidance of context and domain knowledge.

Video events detection techniques have attracted more and more researchers in recent years. However, there is no uniform conception of "event" up to now, as the definition of "event" is related to the application domain. The events that users concern with are different under the different backgrounds even in the same video. Thus, the definitions of event and the methods of event detection and recognition are distinct from each other.

Whereas events hold relevantly clear semantics in sports video analysis. Run [1] and Zhang [2][3] presented methods for parsing baseball videos. Examples for analyzing basketball video are listed in [4]. Assfalg [5] decomposed each shot into its visual and graphic content elements and capture semantic content at a higher level of significance. Using a unique domain-specific feature grass-area-ratio, Xu [6] classified each soccer video frame into three kinds of view, namely global, zoom- in, and close-up. Then the play/break status of soccer game is obtained by heuristic rules. The dominant color region detection algorithm proposed by Ekin [7] can robustly handle the temporal variations in the dominant color due to field, weather and lighting conditions throughout a sports video. It can also be applied in shot boundary detection and shot type classification in sports video, referee and player of interest detection, penalty-box detection [8], and a fast algorithm for play-break event detection.

Hidden Markov Models (HMM) are a powerful tool that has been shown to be of great use in speech and signal processing [9]. The HMM theory is increasingly being applied to the areas of video analysis, such as detection of slow-motion replay segments in sports video for highlights generation [10], adaptive background estimation in surveillance applications [11], and classification of different types of video [12].

In this paper, we propose an HMM-based algorithm for video events detection based on cues fusion and inference. The video shot is regarded as a unit of video analysis. Under the guidance of context and domain knowledge, relevant cues are extracted and fused into the observations of HMM to study the relationship between the observations and the states (shots' contents) that could not be seen directly. Therefore we can achieve the modeling and detection of specific events in video. The experiments on soccer videos show that HMM-based method is effective to the modeling and detection of video events based on the well- selected cues.

In Section 2 we describe the HMM for inferring soccer video events. In Section 3 the cues extraction methods are presented, while in Section 4 we propose the HMM-based soccer events detection method. Experiment results on soccer video are shown in Section 5. Section 6 concludes this paper.

2. THE HIDDEN MARKOV MODEL FOR SOCCER VIDEO EVENTS

One of the keys to the success of applying HMM method on cues fusion and inference for video event modeling and detection is whether the selected cues are appropriate for detecting events. Good cues reflect the statistical relation to the states of HMM effectively, so that the relative models with suitable states and observations can be used in event detection.

First of all, it is important that the detected events should be defined on extractable cues. For example, the appearance of playback shots is selected as an extractable cue of specific events, since playback shots usually appear for providing more details of the events to audience when the significant events (such as shoot and foul) happened in

soccer game video. In this paper, we define **four events**: (1) the shoot event, (2) the foul event and (3) the offside event that contain playback shots, (4) the general game process without playback shots.

The general game process mainly includes the global shots of the soccer field, in which the actions of players could not be seen very clearly. Some zoom-in shots such as kick-off, pass and defense are intermixed, and there are some close-up shots of player's bust or even head portrait in addition. The shoot event usually starts with a global shot near the penalty box area, followed by playback of shoot that are mostly zoom-in shots. The shoot scene may be replayed multiple times from various viewpoints. There could also be several close-up shots of shoot player or goalkeeper. The foul event may start with global/zoom-in shot, and then succeed to the playback of foul. The close-up shots of relevant players or referee may appear before or after the playback shots. The occurrence of offside event is not as frequent as other events. It may consist of a global shot, playback shots that are commonly global shots, and possibly close-up shots of relevant players or referee as well.

Secondly, the states of HMM are defined in terms of the events to be detected. According to the above description of events, the time-dependent changes of event contents obey some particular rules when these events happen. In our hidden Markov model, these rules are described by means of the finite states of the model,

$$S = \{s_0, \ldots, s_{N-1}\}, \quad (1)$$

where $N = 6$ represents the number of states of the events (**Six states**, Fig. 1). Each state represents one of the contents as below: (0) playback; (1) kick-off; (2) pass near the goal area; (3) pass on midfield; (4) ball handling / scrambling / defense; (5) break / close-up.

The selection of cues is the balance between the correspondence to the event content and the feasibility of extracting cues in practice. **Five cues** are chosen based on our previews work:

(1) Shooting scale: global / zoom-in / close-up. It's an important attribute and is also the base for producing other cues.

(2) Playback: yes/no. This cue is essential to infer the three special events.

(3) Goalmouth appearance: yes/no. It's helpful to identify if the events happen near the penalty box.

(4) The occurrence of football trajectory: yes/no. It is usually extracted only in zoom-in shots to judge whether relate to shoot or kick-off.

(5) Blur: yes/no. The blurred frame may imply that there are fast motions in shots. It is extracted in zoom-in shots as well.

Finally, the observation of HMM is defined as the combination of these five cues. There is a finite set of possible observations relative to the cues,

$$\Theta = \{\theta_1, \ldots, \theta_M\}. \quad (2)$$

In our case, the cardinality is

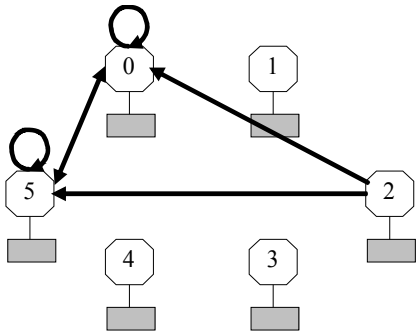

Figure 1. The HMM of events

$$M = 3 * 2 * 2 * 2 * 2 = 48, \quad (3)$$

where five digitals represent the conditions of the cues in the order of the above list.

3. CUES EXTRACTION

In general, model design and cues extraction should be compromised, so as to make each in its proper place. A major advantage of our method is there is no need of high accurate cues extraction algorithms, which are difficult problems to be handled. Our purpose is to validate the feasibility of video event detection by using a hidden Markov model to fuse cues for inferring events.

3.1. Global / Zoom-in / Close-up Shot Detection

Significant color information except the field color is obtained among the frames whose backgrounds are mostly on the soccer field by means of learning the hue of the soccer field. Several color models has been chosen as the models of player uniforms' color. Thus we can determine the shooting scale by combining the field scale and relative size of person (Figure2_a).

3.2. Playback Detection

Playback shots are determined by the methods of defining the midpoint of each shot as the starting point and running Viterbi algorithm forward / backward, based on the detection method of slow-motion replay segments presented in [10] (Figure2_b).

3.3. Goalmouth Detection

The goalposts are detected by searching vertical white strip by Hough Transform, with the help of the net texture detection. In global shots, both the position and the orientation of the soccer field are useful in finding the goalmouth. It is also helpful to detect the field lines in searching for goalmouth (Figure2_c).

3.4. Football Detection

(a) Shooting Scale

(b) Playback

(c) Goalmouth

(d) Football

(e) Blur

Figure 2. Five cues of soccer events

Because of the small size of football in global shots and the few appearance of football in close-up shots, we only detect and trace football in zoom-in shots by a simple approximate method: detecting white round area in certain amount of successive frames (Figure2_d).

3.5. Blurred Shot Detection

Blur value is the reciprocal of the number of edge points. Two kinds of blurred shot should be detected: (1) the blurring of the frame, which represents the fast motion of camera. (2) The blurring of players, which reflects the fast motion of players (Figure2_e).

4. HMM-BASED VIDEO EVENTS DETECTION

After the shot detection and cues extraction, we can obtain the observation sequences by fusing these cues. The resulting shot content sequences seem to comply with the Markov property, so that we apply the HMM method on event inference. The HMM are used to model the temporal evolution of the features of shots. The system framework of video event detection is shown in figure 3.

A hidden Markov model [9] is described by a finite set of possible states S, observations θ, and relative parameters $\lambda = (A, B, \pi)$, where A is the transition probabilities matrix, B is the observation probabilities matrix, and π is the initial state distribution.

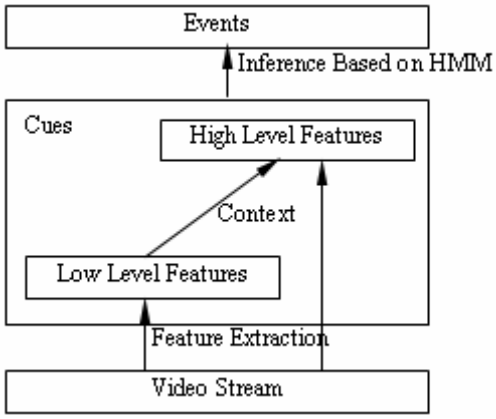

Figure 3. System framework

4.1. A Hidden Markov model of events

Assume there are N_e events to be detected. Each $Event_i$ $(i = 0, \ldots, N_e - 1)$ could be denoted by a HMM parameter λ_i. The hidden Markov modeling of the events is the training or learning problem: Given L_i observation sequences $O^{(l)}\ (l = 1, \ldots, L_i)$ produced from the training data of $Event_i$ where

$$O^{(l)} = O_1^{(l)}, O_2^{(l)}, \ldots, O_{T_l}^{(l)}, \tag{4}$$

to maximize $P(O|\lambda_i)$ by adjusting the model parameters A, B and π. The HMM training method is based on the Baum-Welch algorism.

4.2. Events Detection based on HMM

Given an observation sequence $O = O_1, O_2, \ldots, O_T$ produced from a shot sequence, the probability, $P(O|\lambda_i), i = 0, \ldots, N_e - 1$, relative to each model, λ_i, is calculated by using Viterbi algorithm. Thus this shot sequence belongs to the $Event_j$ where

$$j = \arg\max_{i=0,\ldots,N_e-1} P(Q, O|\lambda_i). \tag{5}$$

5. EXPERIMENT RESULTS

We have experimented on 913 shots in 16 video files recorded on two soccer games, by stochastically selecting half files as the training data and the rest as the test data. The average accuracy of 100 times of experiments is 85.7%.

Table 1 shows one of the events detection results. 8 video files (include 59 events) are selected to train the HMM parameters of four events, and the other 8 video files (include 67 events) are used as test data.

There are only 2 offside events in the training video files. Hence the learned parameters cannot create the proper model of offside event. As we can see in Table 1, the accuracy of offside event detection is very low. If more video files are added in training data, the event models trained should be better.

Table 1. Events detection results

Detected Events	Amount	Results	Correct in Results	Accuracy
General Game Process	32	33	31	93.9%
Shoot Event	21	19	18	94.7%
Foul Event	14	13	12	92.3%
Offside Event	0	2	0	0.0%
Total	67	67	61	91.0%

6. CONCLUSION

In soccer game video, the transition of shot contents in time domain obeys certain regular pattern. The regularity is more distinct when the special events (such as shoot events) happen.

According the above properties of video contents in soccer game, we applied the HMM theory on cues fusion and inference for the video event modeling and detection. Each video shot is viewed as a unit. Cues are extracted to form the observation of HMM. The video content sequence is defined as the state sequence of HMM model. The aim of video event modeling is to learn the parameters of HMM for the events detected. The video event detection is the match problem between the observation sequence and the HMM. Experimental results show that the HMM-based method can effectively detect events in soccer video. It is also clear that this method can be applied to the video event detection in other application domains similarly.

7. REFERENCES

[1] Y. Rui, A. Gupta, and A. Acero, "Automatically extracting highlights for TV baseball programs", *ACM Multimedia*, Los Angeles, CA, pp. 105–115, 2000

[2] D. Zhang and S.F. Chang, "Structure analysis of sports video using domain models", *IEEE Conference on Multimedia and Exhibition (ICME'01)*, Tokyo, Japan, pp. 920–923, August 2001

[3] D. Zhang and S.F. Chang, "Event detection in baseball video using superimposed caption recognition", *Proceeding of the tenth ACM international conference on Multimedia*, Juan-les-Pins, France, December 2002

[4] S. Nepal, U. Srinivasan, G. Reynolds, "Automatic detection of 'Goal' segments in basketball videos", *Proceedings of the ninth ACM international conference on Multimedia*, Ottawa, Canada, September 2001

[5] J. Assfalg, M. Bertini, C. Colombo, and A.D. Bimbo, "Semantic annotation of sports videos", *IEEE Multimedia*, 9(2), pp. 52–60, April/June 2002

[6] P. Xu, L. Xie, S.F. Chang, A. Divakaran, A. Vetro, and H. Sun, "Algorithms and systems for segmentation and structure analysis in soccer video", *IEEE International Conference on Multimedia and Expo (ICME'01)*, Tokyo, Japan, pp. 928–931, August 2001

[7] A. Ekin and A. M. Tekalp, "Robust dominant color region detection with applications to sports video", *Proc. IEEE ICIP (invited paper to Special Session on Sports Video Processing)*, Barcelona, Spain, Oct. 2003

[8] A. Ekin, A. M. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization", *IEEE Trans. on Image Processing*, vol. 12, no. 7, pp. 796-807, July 2003

[9] L.R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", *Proc.of the IEEE*, 77(2), pp. 257-285, 1989

[10] H. Pan, P. Van Beek, and M.I. Sezan, "Detection of slow-motion replay segments in sports video for highlights generation", *IEEE International Conference on Acoustic, Speech and Signal Processing*, 2001

[11] B. Stenger, V. Ramesh, N. Paragios, F. Coetzee, and J. M. Buhmann, "Topology Free Hidden Markov Models: Application to Background Modeling", *Proc. 8th IEEE International Conference on Computer Vision (ICCV 2001)*, Vancouver, Canada, vol. I, pp. 294-301, July 2001

[12] Y. Haoran, D. Rajan, and C.L. Tien, "An Efficient Video Classification System Based On HMM In Compressed Domain", *IEEE Pacific-Rim Conference on Multimedia (2003)*, Singapore, December 2003