SLOW-MOTION REPLAY DETECTION IN SOCCER VIDEOS BASED ON MULTI-LEVEL HMM INTEGRATED WITH SHOT DETECTION

Guoying Jin, Linmi Tao, and Guangyou Xu

Key Laboratory of Pervasive Computing, Ministry of Education
Department of Computer Science and Technology
Tsinghua University, Beijing 100084 P.R.China

ABSTRACT

Slow-motion replay is a key cue for events detection in sport videos since it is the immediate consequence of highlights or important events happened in sports. In this paper, a general approach for slow-motion replay detection in soccer videos is proposed based on multi-level hidden Markov model integrated with shot detection. Experiment results on soccer videos show that our method is effective for detecting slow-motion replays come from both standard and high-speed cameras, with multiple types of transitions including fade, wipe and logo.

1. INTRODUCTION

Slow-motion shots are usually the replays of highlights and exciting events in sports videos. Detail information is provided in slow-motion replays so that people could review these events from closer and different viewpoints. Therefore the slow-motion replay information is always used for video analysis and summarization [1][2], highlights generalization [3], and event detection [4][5].

Two kinds of information are usually used in slow-motion detection: one is the inherent character of slow-motion frames such as frame repetition number; the other is the transition between slow-motion replays and normal frames. Both kinds of information are widely applied in slow-motion replay detection in previous works. Although the method proposed by Wang et al [6] is based on motion information without either of them, the detection results are not so good as other methods below.

Kobla et al [7] adopt the first kind of information in their method. The number of both forward predicted and backward predicted macro-blocks in B frames is used to determine if a slow motion replay come from standard cameras is present in MPEG video. This method is employed in identifying sports videos without giving the boundary of slow-motion replays.

The method proposed by Babaguchi et al [8] utilized transition information of a specific DVE (Digital Video Effect) to determine the boundary of the slow-motion replays. Thus only the slow-motion replays that use this specific DVE as boundary could be detected.

Farn et al [9] detected all gradual transitions and regarded all shots between two gradual transitions as the candidate shots. Then, the candidate shots are checked according to the properties of slow-motion replays come from standard cameras (such as numbers of repeating or inserted frames), and from high-speed cameras (such as larger variation between two consecutive frames and dominant color). This method is not effective in the case of logo and wipe transitions, which are also very common in soccer video.

Pan et al [3] applied HMM method on slow-motion detection using both two kinds of information. Since the HMM features are based on inter-frame difference and color histogram, the method could only detect the slow-motion replays that come from standard cameras with the gradual transition boundary. As an extended of this method, logo transition could also be determined in [10]. The inference algorithm must start at a frame in the slow-motion replay, and feed the given length L "forward/backward pass" frames into the HMM to determine the optimal state sequence. Two problems are not be mentioned and solved in this method. One is how to pick up one frame in each slow-motion replay when there are multiple replays in a video stream. The other is how to make sure that L is long enough to contain the boundaries.

Since the slow-motion replays may come from standard or high-speed cameras and there are multiple different types of transitions between slow-motion replays and normal shots, we proposed a framework that could deal with all these cases. Multi-level hidden Markov model is introduced, and more features are extracted to overcome the small detection scope of the previous methods. Furthermore, shot detection is integrated to avoid the two problems mentioned above.

After the initial shot detection, the input video stream is segmented into clips according to the shot detection results (Fig. 1). Each clip is fed into the low-level HMMs so that the optimal state sequences and the probabilities

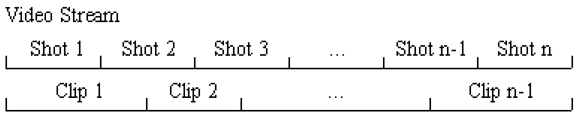


Fig. 1 Video Stream Segmentation Based on Shot Detection Results

$P(O|\lambda_i)$ $i = 1, \cdots, N_{hmm}$ are obtained, where N_{hmm} is the number of low-level HMMs. The observation sequence of global HMM for video stream is composed of these probabilities. Finally the slow-motion replays could be determined by the optimal state sequence of video clips.

The rest of this paper is organized as follows. Section 2 describes the shot detection method for initial segmentation of video stream. The HMMs integrated with shot detection results are presented in Section 3, and the global HMM for video stream is introduced in Section 4. We show the experiment results in Section 5 and conclude this paper in Section 6.

2. SHOT DETECTION

Our shot detection method provides original results for video stream segmentation. There are various types of transition in soccer videos, including **cut**, **fade**, **wipe** and **logo**. The detection method for each type of transition is introduced simply as follows.

The ratio of image difference to average image difference in a window $R_{avrg}(t) = [(2*w+1)d(t)] \Big/ \sum_{t'=t-w}^{t+w} d(t')$ is used for **cut** detection. If the ratio is larger than a given threshold, then a cut transition is detected.

The color histogram difference and image difference between w frames $d_w(t) = \|I(t+w) - I(t)\|$ are calculated to detect **fade** transitions.

The ratio of wipe edge point numbers to the whole image pixel numbers is the measurement for **wipe** detection. The criteria to determine if the pixel (u,v) is a wipe edge point is:

$$b(u,v,t) = \begin{cases} 1 & d(u,v,t) > K * \underset{L}{median}\ d(u,v,t) \\ 0 & otherwise \end{cases},$$

where $\underset{L}{median}\, d(u,v,t)$ is the median of pixel intensity difference in temporal window with length L.

Logo transitions have fixed duration, similar color histogram and similar image in one soccer game. Our method for logo detection has four steps:

Step 1. Select all segments in which the color histogram of each frame is far from the average color histogram of video stream. Cut each segment into several clips according the similarity of color histogram.

Step 2. Cluster all clips into groups based on the average color histograms of clips.

Step 3. In each group, calculate the image difference between frames in each pair of clips. If the number of small image differences in a group is larger than the given threshold, we consider this group as the logo group.

Step 4. Detect logo transitions based on the logo color model created by the logo group.

3. LOW-LEVEL HMMS INTEGRATED WITH SHOT DETECTION RESULTS

According to the shot detection results, video stream is segmented into clips (Fig. 1). Each clip C_i consists of the second half of shot S_i, the first half of shot S_{i+1} and the transition between these two shots. The features extracted from clip frames compose the observation sequences that are fed into the low-level HMMs. These HMMs model the segments that translate from one shot to another.

3.1. Features for Low-level HMMs

Both the inherent character of slow-motion frames and the transition information are adopted as the low-level HMM features.

Slow-motion related features are the features for detecting replays come from standard cameras and high-speed cameras. The former [3] are based on the inter-frame difference $D(t)$, including (a) a measure of zero-crossings in a sliding window; (b) the lowest value of $D(t)$ in the sliding window; and (c) the differences of every two adjacent values of $D(t)$. The latter involve the inter-frame difference and the grass ratio, since the replays come from high-speed have larger variation between frames and always review the scenes on grass field.

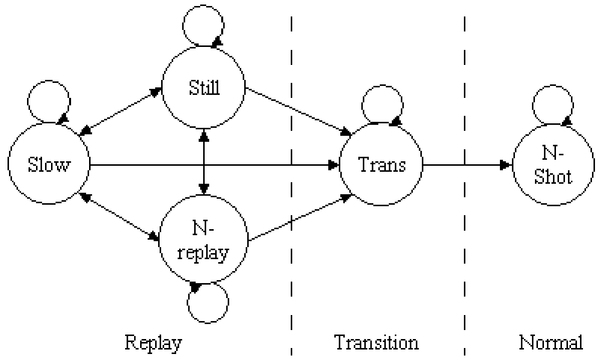
Transition related features are the edge change fraction [11], the color histogram difference, and the wipe edge point ratio as mentioned in Section 2.

All features are normalized and then compose the observation vectors for low-level HMMs. The mixture Gaussian HMM method [12] is used since the features are all continuous variables.

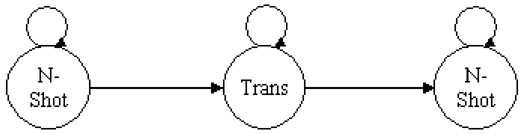
3.2. Hidden Markov Models in Low-level

There are four types of transition segments: slow-motion shot to normal shot (**S-N**), normal shot to slow-motion shot (**N-S**), slow-motion shot to slow-motion shot (**S-S**), and normal shot to normal shot (**N-N**). Since the first two types are symmetrical, we define one HMM [3] to model these two types. The HMM models of S-N segments and N-N segments are shown in Fig. 2. The HMM model of S-S segments has the similar structure of S-N HMM model, in which the normal part is replaced by the symmetrical structure of replay part.

When the observation sequence O_i of clip C_i is fed



(a) The hidden Markov model of S-N segments



(b) The hidden Markov model of N-N segments

Fig.2 The hidden Markov models of S-N, N-N segments

into the three HMM models, $P(O_i|\lambda_{S-N})$, $P(O_i|\lambda_{S-S})$ and $P(O_i|\lambda_{N-N})$ could be obtained directly by Viterbi algorithm. The fourth probability $P(O_i|\lambda_{N-S}) = P(O_i'|\lambda_{S-N})$ should be calculated by feeding O_i' the into S-N HMM model, where O_i' is the time reversed sequence of O_i. The observation vector of high-level HMM model is composed of these four probabilities.

3.3. The Integration of Shot Detection with Low-level HMMs

There must be error-segmentation and miss-segmentation in shot detection, so we apply merge and separate method based on the low-level HMMs results to update the shot detection results.

Merge method is utilized to overcome the error-segmentations. If there are two transitions that are very close, for example, no longer than 15 frames between the two transitions, we merge the two transitions to one. Besides, if there is no transition state in the optimal state sequence of the best matched HMM that has highest probability $P(O_j|\lambda_i)$ of clip C_j, we consider this clip as the result of an error-segmentation. These neighbor clips should be merged into one clip.

Separate method feed the partial of observation sequence after transition state into the HMMs to detect new transition. The clip will be separated into two clips, if new transition states appear in the result sequence.

According to the updated shot detection results, we could re-segment the video stream into new clips. The observation sequences of new clips will be fed into HMMs again, and new optimal state sequences and new

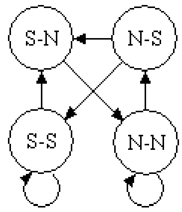


Fig. 3 The global hidden Markov model of video stream
S stands for slow-motion replay, and N for normal shot

probabilities are obtained to update shot segmentation. This procedure could run iteratively until there are no merge and separate operation, and the last updated shot detection results are considered as the final shot detection results.

4. GLOBAL HMM OF VIDEO STREAM

The observation vector of clip C_i consists of probabilities $P(O_i|\lambda_{S-N})$, $P(O_i|\lambda_{N-S})$, $P(O_i|\lambda_{S-S})$ and $P(O_i|\lambda_{N-N})$ from low-level HMMs. These probabilities are continuous variables so that a mixture Gaussian HMM is used as global HMM (shown is Fig. 3) to model the whole video stream. As we can see, a slow-motion replay with one shot contains only two states: N-S and S-N, while a replay with multiple shots has several S-S states between N-S and S-N.

Given the observation sequence produced from the video clips sequence, the optimal state sequence could be obtained by Viterbi algorithm. Therefore, all the shots that involve the clips in S-S state are determined as the slow-motion replays.

5. EXPERIMENTS

Our method is tested on the soccer videos with the duration of 17 hours and 27 minutes. 67 video clips from six classes of soccer videos are used as the experiment data set. These video classes come from World Cup (WC), Union European Football Association Champions League (UEFA), Italian Football League Serie A (IFLSA), and English Premier League (EPL). Detail information of each video class is listed in Table 1.

As we can see, there are multiple types of transition correspond to each transition segment type, but only one (and two in some N-N segments) of them will appear in the videos of same class. Thus 6 video clips are used in HMM training, and each low-level HMM parameters of different transition types are learned respectively, and then combine by a weight vector. In the iterative procedure of low-level HMM inferences integrated with shot detection, previous shot detection results could give the guidance to initialize the weight vectors, while the could be used for re-estimating the low-level parameters.

In the six classes of soccer videos, WC2 is the only

		WC1	WC2	UEFA	IFLSA	EPL1	EPL2
Transition types for each transition segment type	S-N, N-S	Wipe	Logo	Wipe	Fade	Logo	Fade
	S-S	Cut	Fade	Fade	Fade	Fade	Fade
	N-N	Cut, fade	Cut, fade	Cut, fade	Cut, fade	Cut	Cut
Camera type		High-speed	High-speed	Standard	Standard	Standard	Standard
Duration and clips number		0:55:18 (6)	6:35:10 (17)	1:33:33 (10)	4:27:32 (18)	1:36:23 (6)	2:19:24 (10)
Normal shots detected correctly		294	5866	521	3993	529	914
Normal shots detected falsely		6	109	11	67	14	21
Replay shots detected correctly		23	485	37	322	60	108
Replay shots detected falsely		4	81	8	45	11	13
Recall (%)		85.19%	85.69%	82.22%	87.74%	84.51%	89.26%
Precision (%)		79.31%	81.65%	77.08%	82.78%	81.08%	83.72%

Table. 1 Results for our slow-motion replay detection algorithm

one that could be used in slow-motion replay detection by the algorithm in [3]. As a solution to the two problems in [3], shot detection method is introduced and the results could be used to bring out the input sequences for low-level HMMs. Since the shot detection is integrated, the effect of shot detection accuracy on the performance of our algorithm should be considered. If we manually verify shot detection results to raise the accuracy from 98.7% to 100% after the iterative procedure, the recalls and precisions of our algorithm would be improved about 4~6%.

The experiment results with manually verification of shot detection results are shown in Table 1. Note that the results of EPL1 are worse than those of WC2 since there are two kinds of different logo transitions in EPL1.

6. CONCLUSIONS

Most previous works on slow-motion detection are available in small scope of sports video. We propose a general approach for slow-motion replay detection based on multi-level HMM. Low-level HMMs integrated with shot detection provide the probabilities $P(O|\lambda_i)$ as observations of high-level HMM. Experiment results on soccer videos show that our method is effective for detecting slow-motion replays come from both standard and high-speed cameras, with multiple types of transitions between slow-motion shots and normal shots, including fade, wipe and logo.

Acknowledgement: this work is supported by the NSFC (National Science Foundation of China) grant 60273005.

7. REFERENCES

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

[2] B.X. Li, H. Pan, and M.I. Sezan, “A general framework for sports video summarization with its application to soccer”, in *ICASSP'03*, Hong Kong, vol. 3, pp. 169-172, April 2003.

[3] H. Pan, P. van Beek, and M.I. Sezan, “Detection of slow-motion replay segments in sports video for highlights generation”, in *ICASSP'01*, vol. 3, pp. 1649-1652, May 2001.

[4] A. Ekin and A.M. Tekalp, “Generic event detection in sports video using cinematic features”, in *Second IEEE Workshop on Event Mining: Detection and Recognition of Events in Video (EVENT'03)*, Madison, Wisconsin, USA, pp. 17-24, June 2003.

[5] Guoying Jin, Linmi Tao, and Guangyou Xu, “Hidden Markov Model Based Events Detection in Soccer Video”, *ICIAR'04*, *LNCS 3211*, Porto, Portugal, pp. 605-612, September /October 2004.

[6] Lei Wang, Xu Liu, Steve Lin, Guangyou Xu, and Heung-Yeung Shum, “Generic Slowmotion Replay Detection in Sports Video”, in *Proceedings of IEEE International Conference on Image Processing (ICIP’04)*, Singapore, October, 2004

[7] V. Kobla, D. DeMenthon, and D. Doermann, “Detection of slow-motion replays for identifying sports videos”, in *Proc. IEEE Workshop on Multimedia Signal Processing*, Copenhagen, Denmark, pp.135-140, September 1999.

[8] N. Babaguchi, Y. Kawai, Y. Yasugi, and T.Kitahashi, “Linking live and replay scenes in broadcasted sports video”, in *International Workshop on Multimedia Information Retriecal (MIR'00)*, Los Angeles, USA, pp. 205-208, November 2000.

[9] EN-JUNG FARN, LING-HWEI CHEN and JENN-HANN LIOU, “A new slow-motion replay extractor for soccer game videos”, *International Journal of Pattern Recognition and Artifcial Intelligence*, Vol. 17, No. 8, pp. 1467-1481, 2003.

[10] H. Pan, B.X. Li, and M.I. Sezan, “Automatic detection of replay segments in broadcast sports programs by detection of logos in scene transitions”, in *ICASSP'02*, Orlando, Florida, vol. 4, pp. 3385-3388, May 2002.

[11] Ramin Zabih, Justin Miller, and Kevin Mai, “A feature-based algorithm for detecting and classifying production effects”, *Multimedia Systems*, vol. 7, pp. 119–128, March 1999.

[12] L.R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition”, *Proceedings of the IEEE*, Vol. 77, No. 2, pp. 257-286, February 1989.