

# A Statistical Framework for Replay Detection in Soccer Video

Ying Yang<sup>1,2</sup>, Shouxun Lin<sup>1</sup>, Yongdong Zhang<sup>1</sup> and Sheng Tang<sup>1</sup>

<sup>1</sup>Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences  
Beijing, China

<sup>2</sup>Graduate University of Chinese Academy of Sciences, Beijing 100085, China  
Email: {yyang, sxlin,zhyd,ts}@ict.ac.cn

**Abstract**—A novel statistical framework for replay detection is presented in this paper. Unlike current methods, the proposed framework exploits both inherent characters and transition relations of replay and non-replay scenes based on annotation of the video, which realizes segments and classifies video stream into replay and non-replay shots simultaneously. After annotation, the detected replay segment is further verified and its boundaries are adjusted to get more accurate replay segment considering probability distribution of lengths of replay and non-replay shots. Experimental results on soccer video are promising, demonstrating the effectiveness of the proposed framework.

## I. INTRODUCTION

In recent years, there has been increasing research interests in sports video analysis and summarization, such as important event detection and interesting highlight extraction [1], which facilitate browsing and retrieval of sports video. In sports video, replays are important video segments that emphasize key events by replaying the highlights at a slower motion to show the details of actions. Therefore, replay detection is of great help to sports video content analysis.

Many approaches of automatic replay detection have been reported in literatures. Earlier work was based on detecting still frames [1-4] or editing effects such as logo [5-6]. However, these methods are not effective on some kinds of sports video, for example, replays recorded by high-speed camera have no still frames, and not all sports videos have logos to indicate replay. At present, some general methods of replay detection are proposed, which use machine learning to detect replays. For example, Lei Wang et al. [7] used SVM and Jin et al. [8] used HMM to classify pre-segmented shots into replays and non-replays. However, the performance of these statistical methods are relatively poor, since the performance of replay detection is highly dependent on shot segmentation, and that replay and non-replay scenes are classified independently using their different inherent characters without considering transition relations between them. Wang et al. proposed a method of soccer replay detection using scene transition structure [9]. However, the method is difficult to be generalized since they used limited

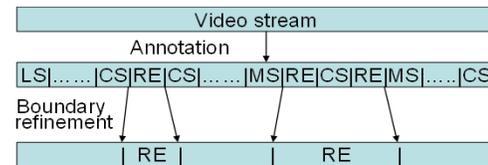


Figure 1. Framework of replay detection.

templates as transition rules and the templates are determined manually by observation.

In this paper, we proposed a novel statistical framework for replay detection. The main idea is to perform annotation of the video as segments of replays and non-replays by searching for global optimization, and then refine the results to get more accurate boundaries of replay. During the annotation stage, the characteristics and transition relations of replay and non-replay are integrated into a statistical model which is formulated by a conditional probability. Maximizing this probability results in the best scene sequence composed of replay and non-replay. As a result, replays and non-replays are detected and segmented simultaneously. Compared with other methods, replay is more accurately detected by exploiting both inherent characters of replays and transition connection relations between replay and non-replay scenes.

The procedure consists of two stages, as shown in Fig.1: 1) Video stream annotation, which segment video into 4 types of scene including Replay (RE), Long Shot (LS), Medium Shot (MS), and Close-up Shot (CS) by maximizing conditional probability. 2) Replay verification and boundary refinement, which is performed to get more reliable replay based on results got in video annotation and probability distribution of lengths of replay segment and non-replay segment.

## II. VIDEO STREAM ANNOTATION

In sports video, replay usually intervenes in video when a break occurs, while vanishes when a new play starts, so it often follows a close-up scene or medium scene indicating a break and is followed by a long scene indicating a new play. Hence, the whole video is segmented into RE, LS, MS and CS to better utilize the transition connection between replay and

non-replay segments.

After the stage of feature extraction, a video stream can be taken as a sequence of feature vectors, denoted by  $O = o_1 o_2 \dots o_T$ , which indicates a potential shot sequence, denoted by  $H = h_1 h_2 \dots h_T$ . So the task of video annotation can be interpreted as finding a shot sequence that maximize the conditional probability of  $H$  under condition  $O$ , i.e. finding

$$\hat{H} = \arg \max_H \{P(H | O)\} = \arg \max_H \{P(H) \cdot P(O | H) / P(O)\} \quad (1)$$

The above equation is transformed by applying Bayes' theorem. Obviously,  $P(O)$  is constant for  $O$  is a known sequence. So the problem can be simplified as following

$$\hat{H} = \arg \max_H \{P(H) \cdot P(O | H)\} \quad (2)$$

$P(H)$  indicates probability of shot sequence without effects of features, which be computed by transition relationship of shots called as Bi-gram.  $P(O|H)$  is the probability of features sequence under a given shot sequence, which can be calculated according to the adopted shots and replay model. Since HMM (Hidden Markov Model) is good at temporal signal analysis, we build 4 HMMs to model replay and 3 non-replay shots, and each of them is called as shot HMM for simplicity in the following description. Hence,  $P(H)$  and  $P(O|H)$  are determined by transition relationship and characters of shots. In other words, results of video annotation are based on characters and transitions of shots.

#### A. Bi-gram Construction

Shot sequence  $H$  is composed of successive shots of different categories, so  $P(H)$  is dependent on the transition probability between adjacent shots. Since shot sequence is a temporal sequence, we suppose the shot sequence is a 1D Markov, i.e. the appearance of present shot is only related to the last shot, which can be formulated by

$$P(h_m | h_1 h_2 \dots h_{m-1}) = P(h_m | h_{m-1}) \quad (3)$$

So  $P(H)$  can be calculated by

$$P(H) = P(h_1 h_2 \dots h_T) = P(h_1) \cdot P(h_2 | h_1) \cdot \dots \cdot P(h_T | h_{T-1}) \quad (4)$$

This assumption is reasonable since various types of shots alternate to exhibit certain semantic clues and don't appear randomly. For example, LS is shown to exhibit the global game status, MS or CS is often shown to track the player or ball. So we use Bi-gram to model this transition connection between each pair of shot  $h_i$  and  $h_j$ , which is represented by a probability model  $P(h_j | h_i)$ .  $P(h_j | h_i)$  can be derived from the statistics of train data using

$$\begin{cases} P(h_j | h_i) = \alpha N(h_i, h_j) / N(h_i) & \text{if } N(h_i) \neq 0 \\ P(h_j | h_i) = 1/l & \text{otherwise} \end{cases} \quad (5)$$

where  $N(h_i, h_j)$  is the number of times shot  $h_j$  follows shot  $h_i$  and  $N(h_i)$  is the number of times that shot  $h_i$  appears.  $l$  is the total number of distinct shot models, and  $\alpha$  is chosen to ensure that  $\sum_{j=1}^l P(h_j | h_i) = 1$ . Hence, for each shot sequence

$H (H = h_1 h_2 \dots h_T)$ ,  $P(H)$  can be calculated given Bi-gram, i.e. transition probability between each pair of shots.

#### B. Feature Extraction

As described above,  $P(O|H)$  is dependent on the feature sequence which is also a temporal sequence indicating a certain shot sequence. Since we use HMM to model shot, feature sequence of a shot is an observed sequence of a given shot HMM  $h_i$ , and each emitting state  $s$  of shot HMM  $H$  produces a feature vector  $o$  in feature sequence [10]. So  $P(O|H)$  can be transformed into

$$P(O | H) = \sum_S P(O, S | H) \quad (6)$$

Where  $S = s_1 s_2 \dots s_T$  is the state sequence which emits feature sequence  $O = o_1 o_2 \dots o_T$  through the link of all the shot HMMs which we called a super HMM. A super HMM is obtained by concatenating the corresponding shot HMMs using a pronunciation lexicon [11]. So  $P(O, S | H)$  can be derived from

$$\begin{aligned} P(O, S | H) \\ = \prod_{t=1}^T P(o_t, s_t | s_{t-1}, H) = \prod_{t=1}^T [P(s_t | s_{t-1}, H) \cdot P(o_t | s_t)] \end{aligned} \quad (7)$$

$P(o_t | s)$  is the emission probability distribution of states  $s$ , and  $P(s_t | s_{t-1}, H)$  is the transition probability between two states, which can be derived from parameters of shot HMM.

Parameters of HMM are estimated by EM algorithm using the feature vector sequence [12], so features are vital to construction of HMM and description of a shot. Thus,  $P(O|H)$  is dependent on the adopted shot model which is derived from the inherent characters of shots.

Each shot is partitioned into segments to extract features from each segment. Shot segment can be one or more consecutive frames, which is called as Shot Segment Unit (SSU). Features are extracted from each SSU to construct a feature sequence. Features are firstly extracted from each frame in a SSU, and feature values of a SSU are the mean of corresponding feature values of all the frames in it. The following 2 types of features are used for they are generic.

Since different types of shots have different playfield and player size, these differences are expressed in color distribution. Therefore, 3 color features are extracted from each frame by computing the mean of L, U, V components of all the pixels in it for CIE LUV color space is approximately perceptual uniform.

Different types of shots have different motion magnitudes. Hence, 3 motion features are extracted from each frame to differentiate replay shots from other types of shots, which are frame difference, compensated frame difference and motion magnitude. Compensated frame difference denotes the frame difference based on global motion compensated block. Motion magnitude of a frame is obtained by computing the mean of motion magnitude of all the blocks in it.

#### C. Procedure of Video Annotation

With Bi-gram and shot HMM constructed, sports video can be segmented and classified into 4 types. Since log



manually. The implementation of video annotation is based on HTK 3.3 which is a HMM Toolkit [13].

TABLE I. TEST VIDEOS

Name	Match (2006)	Length (min:sec)
Soccer1	ENG-POR (07-01)	46:38
Soccer2	GER-ARG (06-30)	46:05
Soccer3	JPN-BRA (06-22)	47:04
Soccer4	POR-MEX (06-21)	47:52

Table II shows the final experimental results, where Correct denotes the overlap of the ranges of detected replay and ground-truth is more than 60% of the length of ground-truth. As we can see, the performance of replay detection achieves 81.2% recall and 83.1% precision rate on the average.

TABLE II. EXPERIMENTAL RESULTS OF PROPOSED FRAMEWORK.

Match	Soccer1	Soccer2	Soccer3	Soccer4
Ground-truth	20	22	23	20
Correct	16	21	18	14
False	0	2	6	6
Recall	80.0%	95.4%	78.3%	70%
Precision	100%	91.3%	75.0%	70%

For comparison, we applied the method proposed in [7] on the test data, which detects replay segment based on SVM classifier. According to [7], RBF kernel is used for SVM and the features are the same as described in the paper. The experimental results are shown in Table III.

TABLE III. EXPERIMENTAL RESULTS OF SVM CLASSIFICATION.

Match	Soccer1	Soccer2	Soccer3	Soccer4
Correct	11	16	12	14
False	5	13	9	9
Recall	55.0%	72.7%	52.2%	70%
Precision	68.8%	55.2%	57.1%	60.9%

The total recall and precision rates achieved by SVM are 62.4% and 60%, respectively, which approximate to the results given in [7]. However, they are far lower than the results obtained by our method.

TABLE IV. EXPERIMENTAL RESULTS WITHOUT VERIFICATION AND BOUNDARY REFINEMENT

	Match	Soccer1	Soccer2	Soccer3	Soccer4
<b>With Bi-gram</b>	<i>Correct</i>	11	16	11	11
	<i>False</i>	6	9	12	13
<b>Without Bi-gram</b>	<i>Correct</i>	10	16	13	10
	<i>False</i>	10	10	14	15

The results directly obtained from video annotation without replay verification and boundary refinement are also given in the top part of Table IV. As we can see, the performance is not satisfactory due to many false alarms. In addition, further investigation shows that the boundaries of detected replays are not accurate, resulting in lower recall rate. Comparing Table IV with Table II, it can be seen that the proposed statistical method of replay refinement greatly reduces false alarms and effectively enhances performance by eliminating incorrect replays and adjusting inaccurate boundaries. In the bottom part of Table IV, we can also see that false alarms decreased with the application of Bi-gram since

Bi-gram considers transition probability between each pair of shots and eliminates the illogical replay segments.

## V. CONCLUSIONS

A novel statistical framework of replay detection is presented in this paper. Replay and non-replay shots are firstly detected and segmented simultaneously based on statistical inference using HMM and Bi-gram, and then replay segments are further verified and its boundaries adjusted based on the probability distribution of length of replay segment and non-replay segments to get improved performance. Experimental results on soccer video are promising, which demonstrates the effectiveness of the proposed statistical framework.

## ACKNOWLEDGMENT

This work was supported in part by the National Basic Research Program of China (973 Program, 2007CB311100), the National Hi-Tech and Research Development Program of China (863 Program, 2007AA01Z416), and the Beijing New Star Project on Science & Technology (2007B071).

## REFERENCES

- [1] Ahmet Ekin, A.Murat Tekalp, R.Mehrotra. Automatic soccer video analysis and summarization. In IEEE Trans. Image Processing, vol. 12, 2003, pp.796-807.
- [2] V. Kobla, D. DeMenthon, D. Doermann. Detection of slow-motion replays for identifying sports videos. In IEEE Third Workshop on Multimedia Signal Processing (MSP'99). Copenhagen, Denmark, September 1999, pp.135-140.
- [3] H. Pan, P. van Beek, and M.I. Sezan. Detection of slow-motion replay segments in sports video for highlights generation. In ICASSP2001, vol. 3, pp.1649-1652.
- [4] H. Pan, B.X. Li, Sezan M I. Automatic detection of replay segments in broadcast Sports programs by detection of logos in scene transitions. In ICASSP2002, vol. 4, pp.3385-3388.
- [5] Lingyu Duan, Min Xu, Qi Tian, Changsheng Xu. Mean Shift based video segment representation and applications to replay detection. In ICASSP2004, vol. 5, pp.709-712.
- [6] Xiaofeng Tong, Hanqing Lu, et al. Replay detection in broadcasting sports video. In IEEE Conference on Image and Graphics, 2004.
- [7] Lei Wang, Xu Liu, S. Lin, Guang-You Xu and Heung-Yeung Shum. Generic slow-motion replay detection in sports video. In IEEE ICIP2004, pp.1585-1588.
- [8] Guoying Jin, Linmi Tao, and Guangyou Xu. Hidden markov model based events detection in soccer video. In International Conference on Image Analysis and Recognition, 2004 vol. 3211, pp.605-612.
- [9] Jinjun Wang, Engsiong Chng, Changsheng Xu. Soccer replay detection using scene transition structure analysis. In ICASSP 2005, vol.2, pp.433-436.
- [10] Rabiner, L.R. A tutorial on hidden markov models and selected applications in speech recognition. In Proceeding of the IEEE, 1989, vol.77, pp.257-286.
- [11] Hermann Ney, Stefan Ortmanns. Progress in dynamic programming search for LVCSR. In Proceeding of the IEEE, 2000, vol.88, pp.1224-1240.
- [12] Bilmes, J. A gentle tutorial on the EM algorithm and its application to parameter estimation for gaussian mixture and Hidden Markov Models. Technical Report of University of Berkeley, ICSI-TR-97-021, 1998.
- [13] Steve Young, et al. The HTK book (for HTK version 3.3). Cambridge University Tech Services Ltd, 2005.
- [14] Larry Wasserman. All of statistics: A concise course in statistical inference. Springer, 2004.