GENERIC SLOW-MOTION REPLAY DETECTION IN SPORTS VIDEO

Lei Wang[†], *Xu Liu*[†], *Steve Lin*[‡], *Guangyou Xu*[†], *Heung-Yeung Shum*[‡]

[†]: Key Laboratory of Pervasive Computing, Computer Department, Tsinghua University, Beijing, 100084, P.R.China
[‡]:Microsoft Research Asia, No. 49, Zhichun Road, Haidian, Beijing, 100080, P.R.China.

ABSTRACT

Slow-motion replays highlight important and exciting events in sports videos. Previous works for slow-motion replay detection, however, are usually limited to only some specific production rules such as frame repetition, special video effects and transitions. In this paper, we present a generic method for detecting slow-motion replays in sports video based on the difference of motions between slow-motion replays and normal shots within the same shot class. Experiments on different types of sports videos have verified our approach and achieve reasonable results.

1. INTRODUCTION

Slow-motion replays provide quite important information in sports video, since they are usually related to highlights or key events. In the literature, many works have utilized this information for event detection [1], video summarization [2, 3], structure analysis [4] and highlight generation [5].

1.1. Relevant Work

Only a few papers have addressed the problem of automatic detection of slow-motion replays in video. In [4], slow-motion replay segments were detected according to human specified digital video effects. Similarly, logo patterns were detected in [6] to determine slow-motion replays. Obviously, these techniques are not general enough to detect all types of slow-motion replays in different sports video, especially when the production behaviour is not specifically modelled.

Another widely used technique is based on the observation that many slow-motion replays are produced by repeating frames recorded from standard cameras [7, 8, 5, 2]. As a result, the image difference between pairs of adjacent frames within slow-motion replay segments frequently alters between zero and non-zero differences. Subsequently, this fluctuation is the main feature used to detect slow-motion replays in these works.

In practice, there is however another common method for generating slow-motion replays. In this technique, video is recorded by high-speed super motion cameras, and the frames are played out at the normal speed to produce slow motion. In such cases, there is no fluctuation in the differences between adjacent frames that could be used for detecting slow-motion replay segments, so previous techniques cannot work on these videos. Actually, in our data set, we have found that in most popular sports events, such as World Cup (soccer), Union European Football Association Cup and Champions League (soccer), English Premier League (soccer) and NBA (basketball), slow-motion replays are generated with high-speed super motion cameras.

1.2. Our Approach

We approach this problem by first observing that humans can distinguish slow-motion replays from normal shots easily. Based on user studies, this detection is typically done by noticing the difference in human motion or ball motion between slow-motion replays and normal shots. While currently it is very difficult to extract human bodies or balls robustly from sports videos, this observation does give us some hints for detecting slow-motion replays directly by comparing the motions in shots.

A challenge is that the magnitude of motion varies among different types of shots. For example, in soccer video, global view shots usually have relatively much smaller motions than close-up shots, which are usually used to track motions of individual key players. It is possible that the motion of a normal global view shot may be smaller than a slow-motion replay of a close-up shot. One should first classify the shots, and then compare the motions in shots within the same category. Intuitively, slow-motion replay shots should have less motion than normal ones for the same shot type.

In our system, shot classification and motion comparison are performed in one pass. Features for differentiating shots such as grass ratio [9, 1] and features which are used to describe motion characteristics such as the mean block motion vector (which may also help shot classification) are put together into a SVM (support vector machine) based classifier to directly detect slow-motion replays. This procedure significantly reduces error in shot classification that would degrade slow-motion replay detection. Since shot segmentation is not the main contribution of this paper and has been the focus of much previous work [10], we assume it has

This research is supported by Chinese NSF grant No. 60273005.

been done by a preprocessing stage in this paper.

Based on this approach, our paper is organized as follows. Section 2 describes the features that we use for slow-motion replay detection. Our approach for classification is introduced in Section 3, with experimental results presented in Section 4. We discuss some relevant issues in Section 5 and conclude this paper in Section 6.

2. FEATURE EXTRACTION FOR SLOW-MOTION REPLAY DETECTION

2.1. Color Ratios

Color ratios such as grass ratios [9] and playground color ratios [1] have been widely used for shot classification. However, this feature is generic enough for general use only if: 1) The color extraction method works for different types of sports; 2) The color model can be learnt automatically; and 3) It is capable of handling multi-colored patterns such as striped uniforms and playing fields.

In our previous work [11], we proposed an algorithm to estimate the dominant color (which is usually the playground color in sports video) with Gaussian Mixture Models automatically in general sports videos. The model is capable of dealing with multi-colored patterns and the extracted color model is further refined based on domain constraints of sports videos. In addition, player uniform colors could be extracted automatically as well by predicting the body region with face detection. Some of the results are shown in Section 4.1. We refer the readers to [11] for the technical details. This approach provides several important cues for shot classification, and in this paper, those features are related to the extracted color model:

Playground Color Ratio: The ratio between the number of playground pixels and the number of pixels in a frame;

Player Uniform A Color Ratio: The ratio between the number of player uniform A pixels and the number of pixels in a frame;

Player Uniform B Color Ratio: The ratio between the number of player uniform B pixels and the number of pixels in a frame;

Note that as features of a shot, all three of these ratios are the mean values of their corresponding color ratios for all the frames in the shot.

2.2. Motion Related Features

Motion related features are also employed in our method. The **frame difference** is defined as

$$D_f = \sum_{all\ pixels} (I_j(x,y) - I_{j-1}(x,y))$$

where $I_j(x,y)$ is the color of pixel (x,y) in frame j. The mean D_f of all the consecutive frames in a shot is taken as the first motion related feature of a shot. Obviously, slow-motion replays usually consist of small frame differences due to their relatively slight scene changes. In our system, all the frames are subsampled by 4 times in both row and column to reduce the calculation as well as to smooth the noise in the video signal.

The second motion related feature is the mean **motion compensated block based frame difference** D_m of all the consecutive frames in a shot. The frame is divided into blocks (in our system, the size of each block is 4×4).

$$D_m = \sum_{B_k \in G} (\min_{(u,v)} \sum_{(x,y)\in B_k} (I_j(x,y) - I_{j-1}(x+u, y+v)))$$

where G is the set of all blocks, B_k is a block, and (u,v) is within a search range of motion vector (similar to MPEG encoding). This feature is robust to camera and object motions, and it is different from the previous feature, especially for different shot types. For example, while global view shots usually have a small D_f and even smaller D_m, the D_f for an in-field medium-range shot or a close-up shot could be very large while corresponding D_m could be very small.

In the previous calculation, (u,v) corresponding to the minimum block difference is the motion vector of the block. **magnitude of motions** $M = \sum_{B_k \in G} \sqrt{u_k^2 + v_k^2}$, where (u_k, v_k) is the motion vector of B_k. The mean M for all the frames in a shot is our third motion related feature. This feature is a direct representation for detecting slow-motion replays.

2.3. Other Features

The mean L,U,V colors of all the pixels in a shot compose other three features. We work in the CIE LUV (CIE XYZ) color space in this paper, since it is "perceptually uniform", meaning that two colors of equal cartesian distance in the color space are also equally distant perceptually.

The length (in frames) of a shot is the last feature. Slow-motion replays usually reviews an entire event in the sports game and are produced when the game is paused. Consequently they are not as long as some global view shots which may last one minute, but should have enough length to clearly demonstrate a whole activity. The feature is quite helpful to the problem, since most slow-motion replays are relatively long in-field medium shots [2] about several players, while most normal medium shots are very short.

Obviously, all these ten features are robust to different methods for generating slow-motion replays, and they can be extracted from sports video data automatically. Now each shot is represented as a vector of ten real numbers and is applicable to SVM based training and classification.

3. SUPPORT VECTOR MACHINE BASED CLASSIFICATION

50 slow-motion replays and 250 normal shots from 2 video clips are taken as our training data. We reduced the samples for normal shots to avoid over-fitting in training. The implementation of our SVM classifier is based on the LIBSVM, which is a C++ library for SVM classifiers [1]. And we followed the procedures proposed in [12].

Each dimension of the feature vector space was linearly normalized to the range [-1,+1] before applying SVM. The RBF (radial basis function) kernel was first considered, since it is capable of producing non-linear classifiers without encountering numerical instabilities, and it has fewer parameters to tune. (In practice, we have tried four basic kernels: RBF, linear, polynomial and sigmoid. RBF was slightly better than other three kernels in our case.) There are two parameters while using RBF kernels: C and γ. Cross-validation and grid-search [12] were applied to find the optimal parameters.

4. EXPERIMENTS

Our system was tested on eight sports video clips (two soccer clips from World Cup 2002, two soccer clips from the English Premier League 2003, two soccer clips from Union European Football Association Champions League 2003, and two basketball clips from the NBA 2002). Two of them (World Cup 2002 video A1 and the English Premier League 2003 video B2) were used as training data for the SVM based classifier. After training, it was tested on all the data sets without any change.

4.1. Extracting Color Models

Figure 1 and Figure 2 show the results of detecting playground color pixels and player uniform color pixels. As can be seen, our approach works well for both single colored and two-colored playgrounds and player uniforms, as well as for different sports. The color models were successfully estimated and provide very robust color ratio features.

4.2. Slow-motion Replay Detection

Table 1 shows the results for slow-motion replay detection (WC - World Cup, PL - the English Premier League, UEFA - Union European Football Association Champions League, data sets marked with an asterisk are training data). Note that it is very important to show the results for normal shots, since a good slow-motion replay detector should not only find the replays, but also not take too many normal shots as slow-motion replays. Although the results are not so perfect, they are still comparable with other works [2], and it

[1] Available at http://www.csie.ntu.edu.tw/~cjlin/libsvm/

Fig. 1. Results for extracting the playground color. Top row: original frame; Bottom row: playground color pixels. From left to right: results for World Cup video A1, results for English Premier League video B2, results for NBA video D1.

Fig. 2. Results for extracting player uniforms using the recovered color models. Top row: original frame; Bottom row: player uniform color pixels. From left to right: results for World Cup video A2, results for England Premier League video B2, results for NBA video D1.

should be noted that they were achieved using a generic algorithm, and made no use of any special patterns of producing slow-motion replays, such as frame repetition or logo transitions.

While the results for the two NBA videos are acceptable, it suggests that it is better to have separate training for different types of sports videos, since they usually have very different shot characteristics and production rules. For example, there are many more slow-motion replays in basketball sports video, since there is much more scoring.

5. DISCUSSION

This paper presents a very important problem: how to generically detect slow-motion replays in sports videos, especially for different production methods and sports types. While a solution is presented as well, the results are far from perfect.

It seems that better features are needed to solve this prob-

Sequence	WC A1*	WC A2	PL B1	PL B2*
Normal shots correctly detected	280	253	197	211
Normal shots falsely detected	9	14	12	8
Slow-motion replays correctly detected	21	19	17	20
Slow-motion replays falsely detected	4	6	6	3
recall(%) vs. precision(%)	84.0:70.0	76.0:57.6	73.9:58.6	87.0:71.4
Sequence	UEFA C1	UEFA C2	NBA D1	NBA D2
Normal shots correctly detected	227	219	91	87
Normal shots falsely detected	13	12	12	14
Slow-motion replays correctly detected	21	19	15	16
Slow-motion replays falsely detected	7	7	9	8
recall(%) vs. precision(%)	75.0:61.8	73.1:61.3	62.5:55.6	66.7:53.3

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

lem. Color is a very important cue, but it is unstable since it changes a lot with different cameras (sensors) and lightings. It is also difficult to accurately extract colors from cluttered backgrounds, e.g., some billboard advertisements and audiences may have colors very similar to player uniforms and are visually connected to the players. In this case, they would greatly affect the player uniform color ratios. Frame difference and motion are highly dependent on the camera motion and the state of the sports game, thus they are not ideal features. Global view shots for a fast-paced offense may have a very large frame difference and motion, while other global view shots for some passing in the midfield may be nearly still.

A possible extension of our work would be incorporating human body detection and tracking. Humans have inherently nearly constant frequency for walking, running, and other human behaviors. This "physical clock" helps us to tell which motions are slower than "normal" speed, and is a very useful and robust cue for distinguishing slow-motion shots from slowly moving players due to camera positions and orientations.

6. CONCLUSIONS

A generic method for detecting slow-motion replays in sports video is presented in this paper. This method works for different types of sports video, without any prior knowledge of special production rules such as frame repetition or logo transitions. It based on the observation that human can tell some shots are slower in the same shot type according to motion, and it performs shot classification and slow-motion replay detection in one pass. Experimental results on various sports videos have validated our approach and further extensions are proposed to elevate performance.

7. REFERENCES

[1] 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, June 2003, pp. 17–24.

[2] 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.

[3] 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, April 2003, vol. 3, pp. 169–172.

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

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

[6] 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, May 2002, vol. 4, pp. 3385–3388.

[7] V. Kobla, D. DeMenthon, and D. Doermann, "Detection of slow-motion replays for identifying sports videos," in *IEEE Third Workshop on Multimedia Signal Processing (MMSP'99)*, Copenhagen, Denmark, September 1999, pp. 135–140.

[8] V. Kobla, D. DeMenthon, and D. Doermann, "Identification of sports videos using replay, text, and camera motion features," in *SPIE Conference on Storage and Retrieval for Media Databases*, San Jose, California, USA, January 2000, vol. 3972, pp. 332–343.

[9] 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," in *ICME'01*, pp. 928–931.

[10] A. Hanjalic, "Shot-boundary detection: Unraveled and resolved?," *CSVT*, vol. 12, no. 2, pp. 90–105, February 2002.

[11] L. Wang, B. Zeng, S. Lin, G. Xu, and H. Shum, "Automatic extraction of semantic colors in sports video," in *ICASSP'04*, Montreal, Canada, May 2004.

[12] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," Tech. Rep., 2003.