

# *Analysis of Slow Motion based on Adaptive Detection Method in Tennis Video*

Lijun Liu<sup>1, a</sup>, Jiahong Wang<sup>2</sup>

<sup>1</sup>*Sport College of Soochow University, Soochow, 215006, China, Physical Education Department of Heilongjiang Institute of Technology, Harbin, 150050, China*

<sup>2</sup>*Sport college of Soochow University, Soochow, 215006, China*

<sup>a</sup>*lijunliull@sina.com*

**Keywords:** Sports Video, Logo transition, Adaptive threshold

**Abstract:** This paper select the tennis video which has logo transition both at the beginning and the end of slow motion as research material. An adaptive threshold based slow motion replay detection method is proposed. This method solves the problem of adaptively selects threshold in the process of slow motion replay detection. Unlike previous approaches, this method has great improvement in both precision rate and positioning accuracy of the boundaries, and can be used for other sports videos which also have logo transition.

## 1. Introduction

Slow-motion shots are played by studio director to represent highlights in the form of slow motions as to cater audience's psychological needs [1-4]. They contain abundant semantic information. The exploration into slow motion shots has significant meanings: one other hand, it's helpful to create video abstract or generate wonderful collection of a match; on the other hand, the removal of replayed shots is good for professional researchers to analyze better sports video contents or solely make semantic analysis of slow motions [5-7]. Tennis competition is a kind of sports events with high attention. Video games on the Internet for thousands of hours of tennis games, the current users are most concerned about is no longer how to use the network to transmit and play all the games, but to be able to quickly find their own interesting fragments of interest. Before the advent of video content analysis technology in tennis, it is very difficult for the average user to finish the task, only by the professional television production department to invest a lot of manpower and material resources for manual editing. Therefore, the extraction of video abstraction from a large video database becomes a direction of video research. Because of the semantic information contained in the slow shots itself, the detection of the slow shot is an important way to detect the highlights. It can be used to calibrate the position of the highlights in the tennis match, to assist in the creation of a video summary or to create a wonderful collection of games.

In addition, for people engaged in sports video research, the extraction of the slow motion and the moving object in the game will help to analyze the sports video content better. The final purpose of

the research of sports video is to realize the automatic annotation of the high level semantics of the video, which includes the acquisition of the underlying features, the generation of the intermediate objects and the inference of the high-level semantics. The acquisition of the underlying features is the basis of the latter two tasks. The generation of intermediate objects includes target extraction and target recognition, which is based on the feature extraction from the video frame sequence. The main target of tennis video includes the court, the players, the racket and the ball. The generation of intermediate objects can provide clues and reasoning basis for high-level semantic annotation. In order to automatically label the high-level semantic of sports video. First of all, it is necessary to analyze the semantic level of the content, and the target extraction is the premise of semantic content analysis. As an important part of the sports video, slow motion detection is the basis of the semantic content analysis. Slow shots formation is divided into two categories: The first kind of slow shots is formed by the ordinary camera after the insertion frame, add effect, etc. The second kind of slow motion camera is formed by the high speed camera after the normal speed.

The detection of slow motion is the starting point and end point of the slow shot. Because of the way the slow shot form different, and different video directed editing methods are different in slow motion. Therefore, the detection of slow motion should be based on the actual situation. This paper mainly focuses on the extraction of moving objects in the slow motion of tennis video. In the slow motion, the focus of the camera is usually followed by the movement of the player or the ball, and the change of the background is obvious. This paper makes use of the characteristics of the frame of the tennis match and the knowledge of the relevant fields to compensate the movement background, the efficiency of the algorithm is improved.

## **2. Shot boundary detection technology**

Shot segmentation is an important content of video semantic content analysis, and the accuracy of segmentation directly affects the accuracy of semantic analysis. The premise of the shots segmentation is to detect the shot boundary. The transformation between the shots is mainly divided into two Species: mutation and gradient. A mutation is a direct switch from the current shots to the next; Gradient is defined from the current shots to the next shot transform, there will be a change such as dissolution, wipe, fade out the process.

The basic idea of the shot boundary detection technology is to compare some features of the adjacent frames. If there is a big change, then the shots transformation occurs, and a larger feature of the place where the camera is located in the boundary. At present, the common shot boundary detection technology mainly has the comparison method based on the pixel, the comparison method based on the block, Based on histogram comparison method, double threshold comparison method, etc [14].

### **2.1 The comparison method based on the pixel**

Comparison method based on the pixel is on between adjacent frames difference between pixels of summation, then will obtain the difference and same predetermined threshold, if the difference is greater than a threshold, shot change between the two frames, they belong to different shots. This method is simple in operation, small amount of calculation, but the shots and the moving object is very sensitive.

### **2.2 The block comparison method**

Based on block comparison method based on pixel comparison method is similar, but will use the global features of the image change using local features, image is divided into blocks, in the block

were compared, the effects it caused to the camera and moving object has certain robustness, but at the same time is also much higher than the substrate to like the comparative method of large amount of calculation.

### 2.3 The histogram comparison method

For the effects to further narrow the shots and moving objects on the performance of shot boundary detection, by comparing the histograms of adjacent images. This is because when the background region invariant, object movement caused by the histogram difference is very small and histogram comparison method for image rotation, observation angle and scale change has very good robustness. But this method does not take into account the spatial characteristics of the image, when different frames have the same histogram, the shot boundary cannot be detected.

Using  $H_i(j)$  to represent the  $i$ th frame of the histogram,  $n$  is the interval number of the histogram,  $c$  is the image of the color component, the difference between the  $i$  frame and the  $i+1$  frame can be expressed by the following formula:

$$D_{i,i+1} = \sum_{j=1}^n |H_i(j) - H_{i+1}(j)| \quad (1)$$

$$D_{i,i+1} = \sum_c \sum_{j=1}^n |H_{i,c}(j) - H_{i+1,c}(j)|$$

In addition, the histogram square difference comparison method can be used, which is the improvement of the histogram comparison method, which can enlarge the distance between adjacent frames:

$$D_{i,i+1} = \sum_i \frac{(H_i(j) - H_{i+1}(j))^2}{H_{i+1}(j)} \quad (2)$$

In shot boundary detection, were extracted first the characteristics of two adjacent frames, then according to which a formula of computing the difference between them, if the difference exceeds the preset threshold is considered a shot change between two frames, or continue to turn to the next frame to judge.

### 2.4 Double threshold comparison method

Shot boundary detection technology above to determine the final boundary points are required depends on the threshold and threshold selection is suitable or is not directly affect the accuracy of shot boundary detection. The usual practice is to frame the difference between the frame and the average value plus a small tolerance value as a threshold, which can be used for the detection of abrupt changes in the shots, but the effect is not ideal for gradual change detection.

## 3. Adaptive threshold detection method for slow-motion shots

When watching sports program, we can find in the beginning and ending of playing slow shots, there's shot cut with a Television station LOGO or some sports organization mark appearing in the middle of TV screen, which is called LOGO conversion shot. Since highlights are between two LOGO conversion shots, we can inspect LOGO transformation to realize detection of slow motion shots. Here we explore tennis video slow shots where LOGO conversion exists. For feature extraction, threshold selection and boundary location, we take new method, achieving adaptive

selection of threshold during our exploration and obtaining higher boundary location accuracy.

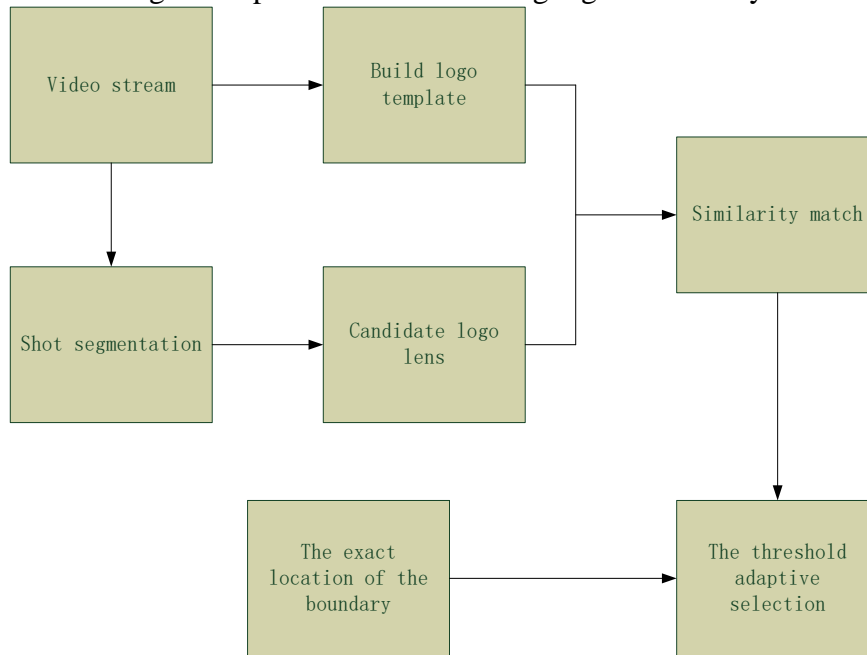


Fig 1 Flow chart of adaptive threshold detection method for slow motion camera

After shot segmentation, a tennis video is cut into numerous shots. Slow motion shot is actually a shot group formed by a set of shots. The LOGO conversion shots in the beginning and ending are separated apart. If we detect the two LOGO conversion shots and make respective boundary frames as beginning and ending of slow shot clip, it's possible to locate roughly the boundary of slow shots. The proposed show shot detection algorithm has following ideas: firstly, make shot segmentation of video streaming to find out candidate LOGO shots from therein; then, generate LOGO template; next make similarity matching of video frame and template in the candidate LOGO shots; according to matching result, determine adaptively threshold value size and locate roughly boundary of slow shots; finally to make detection result more accurate, utilize the change of brightness during LOGO conversion to further locate boundary of slow shots. Algorithm flow chart is shown in Figure1.

### 3.1 Shot segmentation results

According to the shot boundary detection technology, a double threshold comparison method is used for the shot segmentation.

Figure2 lists the three shot segmentation results, representing the end of front shot two frames. The start of after shot two frame.

### 3.2 Selection of candidate LOGO shots

Through observations on lots of tennis videos with LOGO conversion shots, LOGO shots last for 13-22 frames; in that range, shots included are LOGO shots and close-up, of which most are LOGO shots. So we use shots lasting for 13-22 frames after shot segmentation of video streaming as candidate LOGO shot.

### 3.3 Extraction of LOGO template

LOGO template is extracted usually in three ways: (i) artificial extracting method: pick up the

frame with obvious LOGO mark directly from video frame sequence; (ii) semi-automatic extraction method, such as choose by hand several frames which contain obvious LOGO sign [15]; then utilize frame average method to calculate their mean value and use as LOGO template; (iii) automatic extraction method: from 3-5 candidate LOGO shots, choose one candidate LOGO frame, then cluster those candidate frames; calculate the mean value of candidate LOGO frames around the clustering center and use as LOGO template [16]. In the paper, the extraction of LOGO template is realized by semi-automatic method.



(a) Previous shot



(b) Rear shot

Fig 2 Shot segmentation results

### 3.4 Similarity matching

Color correlogram. Color correlogram is a way to express image color distribution. The feature

not only portrays the proportion of one color's pixel quantity of the whole image and also reflects spatial correlation between different color pairs. It can realize higher retrieval efficiency than color histogram and color coherence vector, especially retrieving images whose spatial relation is consistent.

Suppose  $I$  means all pixels of the whole image;  $I_{c(j)}$  is all pixels of color as  $c(i)$ . Color correlogram can be expressed as:

$$\gamma_{i,j}^{(k)} = \frac{P_r}{p_1 \in I_{c(i)}, p_2 \in I} [p_2 \in I_{c(j)} \mid |p_1 - p_2| = k] \quad (3)$$

Color correlogram can be treated as a table indexing by color pair  $\langle i, j \rangle$ , where the  $k$ th component refers to the probability of the distance between pixel of color  $c(i)$  and that of color  $c(j)$  being less than  $k$ . Color correlogram can provide the best retrieving effect; however, if correlation between any colors is considered, color correlogram will get very complicated and huge, spatial complexity up to  $O(n^2)$ . One simplified variant is color autocorrelogram, which considers only spial relationship between images having the same color. So, spatial complexity is decreased to  $O(n)$ .

Matching method. Since color autocorrelogram has advantages on image retrieval and calculation amount, after LOGO template is got, we employ color autocorrelation characteristics to perform similarity matching of video frame and template in candidate LOGO shots. LOGO pattern often lies in the center of video frames; color autocorrelation characteristics of close-up shot video frame central area are apparently discernible from LOGO shot frames. So by color autocorrelation characteristics, we can separate LOGO shots and close-up. Meanwhile, in order to reduce computation, we segment video frames and LOGO template in candidate LOGO shots both by 1:2:1 in two directions, only analyzing and processing contents in central region. At first, we fetch color autocorrelation information of template and that of each frame in candidate LOGO shots; insert them in MySQL database; then take euclidean distance method to compute respectively the distance between each frame and template; insert calculated distance result into video frame related lines in database table, as to use later for locating boundary of slow shots.

$$Distance = \sum_{i=1}^n (\mu_i - v_i)^2 \quad (4)$$

### 3.5 Adaptive threshold selection

The selection of threshold is a very critical issue. Since video contents differ greatly from each other, a threshold may be effective for one video clip, but not for another video, so it's very difficult to pre-define a suitable threshold value. The threshold here in the proposed slow-motion detection algorithm being too big or small will affect the final precision ratio and boundary locating accuracy. To avoid the influence by artificial specified threshold or experical threshold, we proposed a threshold adaptive selection method.

After similarity matching with template through Euclidean distance formula, the distance of LOGO frames is smaller than non-LOGO frames. The objective here is to adaptively choose a threshold and ensure frames to which the distance being less than the threshold corresponds are all LOGO frames. We sort all frames in order by Euclidean distance from small to big after similarity matching. The result is LOGO frames stay ahead and non-LOGO frames stay behind. Finding out a breaking point, we can separate correctly two kinds of frame. The value of distance at the point is threshold value. Considered from the perspective of variance, if other-type frames in one class after classification are fewer, meaning the variance of the class is smaller. If LOGO frames and non-

LOGO frames are correctly divided, the variance of two classes is the smallest, thus the variance sum of them is the smallest. The chosen threshold in the paper is the smallest point which makes variance summation the minimum of two classes after classification. The algorithm is carried out in following steps:

(1) Use calculated Distance= $\{dis_1, dis_2, \dots, dis_k\}$  result by Euclidean distance formula to read from database and sort in ascending order; get the array Dis\_sort= $\{D_1, D_2, \dots, D_k\}$  after ranking, where k is the number of video frames;

(2) Make cyclic variable i valuing from 1 to k; calculate respectively variance of  $\{D_1, \dots, D_i\}$  and  $\{D_{i+1}, D_{i+2}, \dots, D_k\}$  to get summation  $A_i$  of them; compute variable i from 1 to k, one-dimensional variance array  $\{A_1, A_2, \dots, A_k\}$  is obtained;

(3) Search from one-dimensional variance array  $\{A_1, A_2, \dots, A_k\}$  the smallest  $A_i$ ; write down subscript i, corresponding to  $D_i$  in the array Dis\_sort, which is acquired adaptive threshold value.

The adaptive threshold selection method proposed here is suitable for all sports videos with LOGO conversion, because after template matching, the selection of threshold is not relative with video contents. Instead, dividing all frames into LOGO frames and non-LOGO frames are from the perspective of minimum inter-class variance after correct classification. The flexibility of the algorithm is strengthened, able to optimize by itself along with different video contents.

### 3.6 Location of slow shot boundary

Initial location of slow shot boundary. After the threshold is calculated adaptively, fetch all frames to which the distance in Distance being smaller than  $D_i$  corresponds; the outcome is many sets of continuous frame collections. Arrange those frame collects as  $(S_1, S_2, \dots, S_n)$  according to frame order, where n is number of frame collection. Because slow shots appear between two LOGO conversion shots, for frame collection in odd number of segments, which is believed to be LOGO shot starting from slow-motion shot, we choose from it the most preceding frame as starting point of slow shots; for frame collection in even number of segments, which is believed to be LOGO shot ending in slow-motion shot, we choose from it the most posterior frame as ending point of slow shot; till this moment, boundaries of a segment of slow shot are initially located.

If LOGO conversion shots are missed from detection, there will make mistakes to follow the above method to match LOGO conversion shot as to locate slow shots. Through plenty of experimental observations, we realize slow-motion shot contain the most 10 shots. If the number of shots between two LOGO conversion shots is above 10, it's thought that missing detection exists in LOGO shots. In this case, we need give up the last one LOGO shot, letting next one LOGO shot matching with its subsequent LOGO shots, which is likely to solve effectively mismatching problem.

Precise positioning of slow-motion shot boundary. Since LOGO conversion is a gradual changing process. LOGO sign is not prominent when it appears and disappears. The slow-motion shot boundaries are not often precise if determined with adaptive threshold the method, generally 3-4 frames of LOGO frame can't be detected. However after observations, we see the brightness in those video frames' central region changes too big. Get 5-6 frames ahead / backwards from the starting /ending point of calibrated slow motion; use posterior frame to subtract anterior frame; then extract its brightness value and calculate its cumulative sum. The place where luminance sum is the biggest suggests LOGO begins to appear (about to disappear); choose frame at the place as boundary of slow-motion shot to realize accurate positioning of slow shot boundary. Figure3 is effect picture of slow shot boundary after precise positioning, where the first line is boundary of anterior LOGO shot; the second line is partial middle slow shot frame; the third line is boundary of posterior LOGO shot.



Fig 3 After an accurate positioning of the slow shot boundary effect chart

## 4. Experimental analysis and results

### 4.1 Experiment results

Two criteria usually used to evaluate goodness or badness of one algorithm's detection effect is recall rate and precision ratio. Here we utilized boundary positioning accuracy rate to measure precision of the algorithm. Of it, missing frame number means those belong to frame containing LOGO but not detected.

$$\text{Recall} = \frac{\text{Detect the correct number of slow shot}}{\text{Total number of slow shot in video}} \quad (5)$$

$$\text{Precision} = \frac{\text{Detect the correct number of slow shot}}{\text{Detect the total number of slow shot}} \quad (6)$$

$$\text{Accuracy of boundary location} = \frac{\text{Number of frames}}{\text{The logo area total frames conversion}} \quad (7)$$

We choose five tennis match videos which contain LOGO transformation as experimental data, validating the proposed algorithm and slow shot detection technique in [17-18]. It is shown in Table1 and Table2.

Table 1 Experimental results of the slow shot detection algorithm are presented in this paper

Video sequence number	Actual number	Correct number	Error number	Recall (%)	Precision (%)	Accuracy of boundary location (%)
1	18	17	0	93.5	100	0.4
2	16	17	0	95.8	100	0.5
3	24	22	1	91.2	96.3	0.5
4	31	27	0	93.5	100	0.4
5	35	29	1	93.6	96.6	0.6

Experiment results reveal that the proposed slow shot detection algorithm can detect well slow motion shots in videos and achieve higher precision rate and boundary positioning accuracy. Recall



ratio is affected because errors exist during shot segmentation, causing that a few LOGO shots are not detected when candidate LOGO shots are being detected. The algorithm has better precision ratio and boundary positioning accuracy than existing slow motion shot detection approaches, because the utilization of color autocorrelation guarantees image spatial correlation during matching, and adaptive selection of threshold assures that after initial location, what's all acquired are shots containing LOGO conversion, and LOGO shot mismatching problem is corrected, making the algorithm with higher precision ratio. Lastly, on the basis of initially positioning slow shot boundary, the use of brightness changes during LOGO conversion to position boundary accurately increases remarkably the precision of boundary location.

Table 2 Experimental results of slow shot detection algorithm in [17][18]

Video sequence number	Slow shot detection algorithm in [17]			Slow shot detection algorithm in [18]		
	Recall (%)	Precision (%)	Accuracy of boundary location (%)	Recall (%)	Precision (%)	Accuracy of boundary location (%)
1	90	94.5	2.1	90	94.3	2.5
2	93.9	94.5	2.3	89.6	95.7	2.3
3	87	92.7	2.5	88	91.3	3.2
4	92.5	90.4	2.8	89.6	92.6	2.8
5	93.6	90.9	2.6	90.6	95.6	2.9

## 4.2 Algorithm complexity analysis

The adaptive threshold detection method introduced here for slow-motion shots in tennis videos includes shot segmentation, candidate LOGO shot selection, LOGO template extraction, similarity matching, adaptive threshold selection and slow-motion boundary positioning. Of them, the core is adaptive threshold selection, because the choice of threshold matters directly the accuracy of algorithm detecting slow motion shots. For the selection of threshold, we used one-dimension array as sample space, getting step by step the variance of each component array and sum them up; then seek for the smallest variance sum; the time complexity is  $O(n)$  and spatial complexity is likewise  $O(n)$ . Besides, in other steps of the algorithm, the biggest complexity of time and space is  $O(n)$ . Accordingly, the time and space complexity of slow-motion shot detection algorithm here is both  $O(n)$ .

## 5. Conclusion

In this paper, we firstly introduce the shot in tennis video shot boundary detection and classification of related knowledge; Secondly according to the algorithm process is introduced in detail the slow motion detection method; in the end of this paper, put forward the method of the experiment results and algorithm complexity analysis, and comparing the results with the other two methods. Experimental results show that this method has better detection results.

## References

- [1] Zhang Xiaoli, Chi min, "An adaptive threshold detection method for slow motion in tennis video", *Computer engineering and science*, 2013,04, pp.99-103.
- [2] Liu Dian, "Soccer video slow motion detection and video summary generation", *North China University of*

Technology, 2008

[3] Liu Kun, "Research of soccer video retrieval based on MPEG", Jilin University, 2007

[4] Zhang jujub, "Research and implementation of the slow motion detection technology in sports video", University of national defense science and technology, 2004

[5] Gu Shuwen, "Classification and retrieval methods of soccer videos", Huazhong University of Science and Technology, 2012

[6] Li Huifang, "Multi level analysis and retrieval of tennis video", Nanjing University of Science and Technology, 2009

[7] Wang Haibo, "Research on the detection technology of sports video events", North China Electric Power University (Beijing), 2009

[8] Pan Ruifang, Zhongxiang Ming, "Research on the application techniques of long shot in documentary film", Press, 2012,01, pp.24-27.

[9] Lin Bin, Liu Qun, Wang Qun, Nie Yanliu, "Classification method of soccer video camera", Computer engineering and design, 2012,04, pp.1467-1471.

[10]Zhang Yuzhen, Ding Sijie, Wang Jianyu, Dai Yuwei, Chen Qian, "Multi modality event detection based on HMM fusion", Journal of system simulation, 2012,08:1638-1642.

[11]Wang Haoran, Bai Liang, Yang Song Yang, "Semantic modeling method for soccer video based on graph model", Computer science, 2011,06, pp.266-269.

[12]Liu Haitao, Song Yang, "Video semantic content analysis based on ontology", Computer science, 2009,07, pp.170-174.

[13]Yu Junqing, Wang Ning, "Classification of soccer video shots based on sub window region", Chinese Journal of image and graphics, 2008,07, pp.1347-1352.

[14]Zhang H J,Kankanhalli A,Smoliar S W, "Automatic partitioning of full motion video",Multimedia System,1993,1(1), pp.10-28

[15]Xiaofeng Tong,Hanqing Lu,Qingshan Liu,and Hongliang Jin, "Replay Detection in Broadcasting Sports Video" ,In:International Conference on Image and Graphics(ICIG).2004, pp.337-340.

[16]Luo Xiaowei, "Sports video analysis and personalized customization. Wuhan", Wuhan University of Technology, 2010

[17]Zhang jujube. "Research and implementation of the slow motion detection technology in sports video", Changsha: National University of Defense Science and technology, 2004

[18]Zhang Yuzhen, Wei Daidi, Wang Jianyu, Dai Yuwei, "Semantic analysis of soccer video based on multimodal fusion", Computer science, 2010,37 (7), pp.273-276.