International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.1 (2016), pp.33-42
http://dx.doi.org/10.14257/ijmue.2016.11.1.04

The Key Events Extraction Algorithm Based on Shot Events in Soccer Video

Xiang Zhai

Department of Physical Education, Northeast Forestry University,
Harbin 150040, China
597390959@qq.com

Abstract

Football is one of the most popular sports in the world. It is deeply loved by massive ball fans. Generally one complete football game lasts around two hours, but only a very little part of it has wonderful views which attract audiences. Although in sport matches, episodes which people care about are of high subjectivity, for football games, they share some in common. . So how to fetch out concerned events from tremendous football videos has become a very important and typical problem. Here we propose a new algorithm for extracting effectively key events from football videos, which is named key event extraction method based on attributes of shots. The method proves higher recall and precision rate.

Keywords*: Video retrieval, Shot segmentation, Soccer video, MPEG, Key events extraction*

1. Introduction

Y.Gong *et al.* [1] based on the field knowledge about football matches to infer video contents by recognizing white lines in the pitch, detecting camera movements and analyzing football and players [2-3]. Such contents include the position in the field, shooting, corner kicks *etc*. To be specific, if the scenario is close to the goal area and the rubber ball is moving towards the gate [4-5], it's believed a shot. Experimental results show that the system recognizes the pitch position more accurately, up to 90% against 53% for recognizing shot and corner kick. That is mainly due to hi-speed motions and blocking, which makes it difficult to detect the ball [6-7].

Ming Luo *et al.* [8] in Tsinghua University presented one sport video analysis system in the case of football. Their system divides lens to long-distance and close lens according to the field color rate and object size in the key frames [9-10]. Besides, for long-distance lens, they found that for shooting or long passing, quick camera motions would make images vague. Hence they suggested detecting key events in the football match according to the obscurity of frame images [11-12]. In experiments, the algorithm reached 73.7% and 79.2% respectively for precision rate and recall ratio in detecting shooting at the goal and long passing events.

Takagi [13] and Snoek [14] used different classification algorithms, including machine learning methods like hidden Markov model (HMM) [15-16], decision-making tree *etc.*, to implement the fetch of key incidents in football videos. Despite recall and precision rate were both improved, their systems didn't perform well in practice because of complexities and huge computational amount [17].

2. Goal Event Extraction Method Based on Attributes of Shots

In one football match, there're lots of key events, such as corner kick, free kick and scoring a goal as shown in Figure 1. But what's concerned the most is goal event. So here

ISSN: 1975-0080 IJMUE

we center around goal events in football videos to describe the goal event extraction algorithm based on attributes of shots and its implementation.

(a) The Goal Event

(b) The Corner Event

(c) FreeKick Events

Figure 1. The Typical Event in Soccer Video

2.1 Creation of Rules

In general, goal event is abstract idea of semantics in a high level. But how to match well the retrieval of features in lower levels with abstract semantics in a higher level is rather challenging. In football match rules, when the ball passes the goal line as a whole, it's defined a goal. But it's pretty difficult to detect directly lens regarding the ball going over the goal line as lots of motion information exist in football videos, easily leading to false and missing detection. After analyzing plentiful football videos we learnt that:

(1) In the occurrence of a goal event, cameras would move along with the ball towards the goalmouth and take close-up shots of it;

(2) Cameras follow the man who hits home and take photos of celebration;

(3) With TV editing methods, slow-motion replay is broadcasted regarding scoring a goal. No laws exist in the three shots if viewed separately. But put in the football video, and for the reason of temporal continuity, they can be used as the foundation for goal event extraction. Figure 2 displays the whole procedure and specific regularity among those shots.

Based on the above considerations, we choose gate view detection (potential goal shot), player celebration shot detection and slow-motion replay detection of scoring a goal all in close distance as one rule for goal event extraction to build one screening algorithm, which is goal event extraction approach based on shot attributes, to eventually realize the fetch of goal episode.

2.2. Determination of Attributes of Relevant Shots

Firstly, we let football video data for the variable step shot detection based on DC images and treat with key frames of each shot as follows:

(1) Utilize equation (1) to transform colorful frame images to gray images;

$$Y = 0.299R + 0.587G + 0.114B \tag{1}$$

Where: Y is the gray value. R, G, B respectively shows each pixel of the red, green, and blue component.

Figure 2. The Law of the Goal Event Lens

(2) Select appropriate threshold T for the segmentation and shift images to black-white binary image $I(x, y)$

The segmentation of threshold value is to split image space into some meaningful areas. Suppose one image has the histogram as seen in Figure 3. From the picture, we know that most pixel grey scale of image $f(x, y)$ is very low. From that we deduce the image is formed by grey-scale objects overlapping one dark background. We make threshold value T as to cut the histogram into two sections, as shown in the figure. The selection of T should stick to the following principle: B_1 should contain the grey level relevant to the backdrop; B_2 should contain all grey levels of objects.

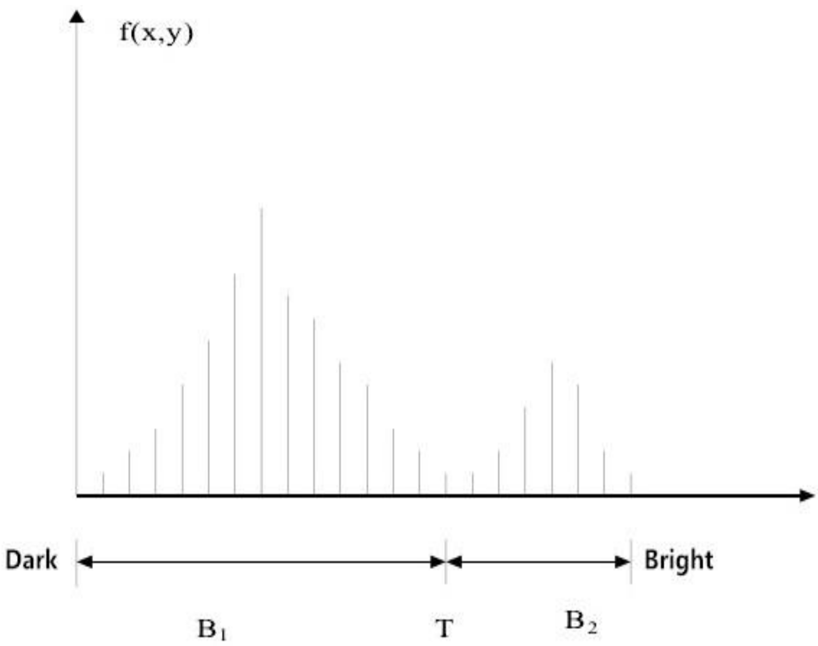


Figure 3. The Gray Level Histogram of Image f (x, y)

In order to complete the detection of goals here, we need to regard the gate as an object and others as background for threshold segmentation. Hence, the selection of threshold T is critical. By threshold segmentation, the image is changed to binary image $I(x, y)$. It shown in the formula (2).

$$I(x, y) = \begin{cases} 255, & f(x, y) \geq T \\ 0, & f(x, y) < T \end{cases} \tag{2}$$

(3) Perform Laplacian sharpening of image $I(x, y)$ to get image $D(x, y)$

Employ Laplacian edge detection operator $\Delta^2 G$ to enhance the image's edge parts; smooth the image by Gaussian filter; two-dimensional Gaussian filter's response function is:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{\frac{x^2+y^2}{2\sigma^2}} \tag{3}$$

$I(x, y)$ is image function; by the exchangeability of convolution and differential in linear systems, we have:

$$\nabla^2\{G(x, y) * I(x, y)\} = \{\nabla^2 G(x, y)\} * I(x, y) \tag{4}$$

By integrating the image's Gaussian smoothing filter and Laplacian differential operation, we can get one convolution operator like:

$$\begin{aligned}\nabla^2 G(x, y) &= \frac{1}{2\pi\sigma^4}\left(\frac{x^2+y^2}{\sigma^2} - 2\right)e^{-\frac{x^2+y^2}{2\sigma^2}} \\ &= A^2\left(\frac{x^2}{\sigma^2} - 1\right)e^{-\frac{x^2}{2\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} + A^2\left(\frac{y^2}{\sigma^2} - 1\right)e^{-\frac{x^2}{2\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} \\ &= K_1(x)K_2(y) + K_1(y)K_2(x)\end{aligned} \tag{5}$$

The operator of image got image $I(x, y)$, after Laplasse sharpening the image $D(x, y)$

(4) Make vertical projection of image $D(x, y)$ as per equation (6).

$$V(j) = \begin{cases} 0, & 0 < j < b \\ 255, & b \leq j < N \end{cases} \tag{6}$$

Where B is the number of black pixels in each column, N is the image height.

Figure 4 shows results of the two different types of shot key frames by the image processing.

(a) The Non-Goal Lens

(b) The Goal Lens

Figure 4. Vertical Projection Detection

In Figure 4, we see that Figure 4(b) shows the clear vertical detection effect of the gate as when the goal event happens, cameras take close-up shots of it. Figure 4(a) displays very un-apparently the detection of goalmouth as they're not key frames in the goal event. Scan the image $D(x, y)$ from the left to right, up to bottom and count statistically the proportion $P_y = W / N$ of white pixel in each column. If $P_y > 16$, then shots regarding potential scoring a goal exist in it.

2.3. Implementation of the Algorithm

Make the chosen lens of three types constitute one decision rule according to particular attribute order and then determine with it the football video key frames which are obtained after detection of lens.

It includes the following steps:

(1) Discern potential goal shots in key frame groups of all shots; if at least one frame there satisfies the decision condition, mark the shot and go to step (2); or think it's a goal shot;

(2) Inspect the shot of key frame groups of two shots in temporal succession and determine it based on main color rate; if one of the two shots qualifies for a close-up shot, i.e. celebration shot, keep the mark in step (3) and further to the next step; otherwise, cancel the mark in the above step (1) and determine it's not a goal shot;

(3) Feed shots which meet both (1) and (2) into slow-motion lens and replay them for detection; detect the frame differential histograms of two shots which are temporally successive and are defined as field's observable shots because they meet with main color rate as per equation (3-6); if at least one shot qualifies for the review of slow-motion lens, keep the mark and determine it's goal shot and what happened is goal incident; otherwise, cancel the mark of shots and return step (1) for the new detection.

3. Experimental Analysis and Results

This experiment is in the Windows XP operating system, Visual C++ 6 development environment to complete. The experiment used 5 field compression for soccer video clips of MPEG-1 as the test data, respectively denoted as test1, test2, test3, test4, test5.

3.1. Detection Results of Shots in Football Videos

The performance of shot boundary detection algorithm is often used to Recall and Precision to represent:

$$Recall = \frac{N_c}{N_c + N_m} \times 100\% \qquad (7)$$

$$Precision = \frac{N_c}{N_c + N_f} \times 100\% \qquad (8)$$

Where, N_c is the number of correctly detected lens, N_m is lens of the missing number, N_f is lens number of false detection

3.2. Determination of Potential Goal Views

The first step for the proposed strategy is to judge potential goal shots because all subsequent detections are done based on it. In light of fewer goals in one football match, any false or missing detection will affect directly the detection results of extraction system. So in this part we set a higher recall ratio for determining potential goal views. That is, setting loose decision condition, which can guarantee all goal views are included. In the experiment, we only judge if only one of key frames suffices the decision condition, it’s determined that the frame’s shot is potential goal shot. After analysis of abundant goal incidents, in the experiment, we make T value 180. Figure5 depicts the procedure of judging one potential goal shot. Results of determining the potential goal shots in video data test1-test5 are listed in Table 1.

(a) The Original Image (b) Gray Image

(c) Threshold Segmentation, the Sharpening (d) The Vertical Projection Image

Figure 5. Decision Process of Underlying Goal Lens

From Table1 we note that the proposed algorithm reached higher recall ratio 96.7% with lower precision rate 80.3%. That is because the determinative standard chosen by the method is too loose. Being the first step to extract goal events, it requires higher recall ratio. The proposed algorithm can fully meet that requirement. Experimental results are used as original data for further selection and extraction.

Table 1. Decision Outcome of Underlying Goal Lens

	Lens number	Key frames	Potential goal shot key frames	Missing number detection	Wrong detection number	Recall (%)	Precision (%)
Test1	127	420	32	2	5	93.7	84.4
Test2	117	372	28	0	4	100	85.7
Test3	130	398	23	1	6	95.7	73.9
Test4	138	406	19	0	4	100	78.9
Test5	103	304	20	1	5	95.0	75.0
Total	615	1900	122	4	24	96.7	80.3

3.3. Detection Results of Slow-Lens Replay

Slow-lens replay is an important part in the extraction method for goal events. In the experiment, we chose five groups of football videos as testing data (test1-test5) to detect slow-motion replay. Results are put in Table2.

Table 2. Testing Result of Slow-Motion Replay

	slow-lens replay number(N_c)	Missing number Detection(N_m)	Wrong detection number(N_f)	Recall (%)	Precision (%)
Test1	18	1	3	94.4	83.3
Test2	15	1	2	93.3	86.7
Test3	22	2	3	90.9	86.4
Test4	14	2	1	85.7	92.8
Test5	10	1	2	90.0	80.0
Total	79	7	11	91.1	86.1

We see from the Table2, 72 views of 79 slow-motion replay are detected. The recall ratio reaches 91.1%. 11 general shots are falsely detected as slow-motion replay. The accuracy rate is 86.1%. We analyzed that in some long-distance shots, object movements are not so noticeable. The histogram difference between neighboring frames is smaller than decision threshold, giving rise to the wrong detection.

3.4 Extraction Results of Goal Events

Table 3 gave the experimental results of the goal event extraction algorithm based on lens attributes for the testing data test1-test5. Obviously, the recall and precision rate of the method here both reached above 80%. Compared with the method by Ming Luo, the new algorithm has improvements in terms of recall and precision rate. Also we learn that goal event is one of key events in football videos. The goal event can't often occur, a very few times in one match or even impossibly takes place in one whole competition. So the wrong or missing detection only once can bring about serious consequences too. In the proposed algorithm, since errors are inevitable for the detection of football video shots and determination of each process in the goal event extraction, there must be differences in the final extraction of such events. To improve the algorithm, it's necessary to enhance recall ratio and precision rate in each process.

Table 3. The Goal Event Extraction Results

	Goals number (N_c)	Missing number detection(N_m)	Wrong detection number (N_f)	Recall (%)	Precision (%)
Test1	4	1	0	75.0	100
Test2	3	0	1	100	66.7
Test3	3	0	0	100	100
Test4	2	1	0	50.0	100
Test5	3	0	1	100	66.7
Total	15	2	2	86.7	86.7

4. Conclusion

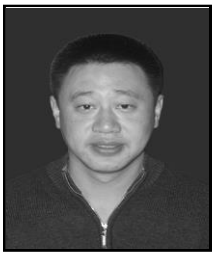
After analyzing the key episodes in football videos, we found that there's strong temporal correlation between camera shots. As long as related shots are detected, it's likely to implement the extraction of key events. So we proposed the extraction method based on attributes of shots and selected typically goal event to illustrate the implementation of it. At last it proved the feasibility through experiments. After analysis of experimental data, we concluded its merits and what's to be improved.

References

[1] G. Yihong, S. L. Teck and C. C. Hock, "Automatic Parsing of TV Soccer Programs", Proceedings of the International Conference on Multimedia Computing and Systems (ICMCS'95), **(1995)**, pp. 167-174.
[2] G. Shuwen, "Research on the retrieval method and classification of soccer video scoring events", Huazhong University of Science and Technology, **(2012)**.
[3] W. Qianhu, "Soccer video information capturing and copy detection", Huazhong University of Science and Technology, **(2012)**.
[4] X. Rong, "Study on the fusion of audio and video features of soccer video retrieval", Jilin University, **(2008)**.
[5] Z. Yanjiao, "Regional map object detection of video abstract under Gauss", Hebei Normal University, **(2014)**.
[6] Y. Ming, "The extraction of football video shot based on exciting degree", Huazhong University of Science and Technology, **(2007)**.
[7] N. Z. Xing, "Soccer video modeling and content analysis method research", Xi'an Electronic and Science University, **(2012)**.
[8] M. Luo, X. Bai and G. Xu, "Content-based Analysis and indexing of Sports Video", in Proc. SPIE Storage and Retrieval for Media Databases 2002, Bellingham, Wash., **(2002)**.
[9] X. Tao, "Research on soccer video indexing algorithm based on multiple features", Huazhong University of Science and Technology, **(2013)**.
[10] Y. Yaguang, Y. Min and M. Shuang, "The key scene detection in soccer video", Cable TV technology, vol. 1, **(2008)**, pp. 84-87.
[11] Z. Yumei and G. Jian, "A key frame extraction system based on shot attribute", TV technology, vol. 1, **(2007)**, pp. 73-75.
[12] Z. Pixi, W. Xiukun and W. Wei, "Li starting point, Li Guohui, Classification method of close shot in soccer video", Journal of South China University of Technology (Natural Science Edition), vol. 9, **(2007)**, pp. 70-73.
[13] Takagi S., Hattori S. and Yokoyama K., "Sports video categorizing method using camera motion parameters", Proceedings of the Conference on Multimedia and Expo 2003. Baltimore, Maryland: IEEE Computer Society, **(2003)**, pp. 461-464.
[14] Snoek C. G. M. and Worring M., "A state-of-the-art review on multimodal video indexing", Proceedings of the 8th Annual Conference of the Advanced School for Computing and Imaging. Lochem, Netherlands：Springer Verlag, **(2002)**, pp. 21-24.

[15] L. Bin, L. Qun, W. Qun and N. Yanliu, "The soccer video shot classification method", Computer engineering and design, vol. 4, **(2012)**, pp. 1467-1471.
[16] W. Haoran, "The white light, the old song yang, Soccer video semantic modeling method based on graph model", Computer science, vol. 6, **(2011)**, pp. 266-269.
[17] H. Yucheng, Y. Junqing, H. Xianqiang, H. Yunfeng and G. Tao, "Mining user preferences in soccer video search engine", Chinese Journal of image and graphics, vol. 4, **(2014)**, pp. 622-629.

Author



Xiang Zhai, He received his B.S degree in physical education from Sports Science Institute of Harbin Normal University. He got his M.S degree in physical education from Shenyang Sport University. He is a lecturer in Department of Physical Education, Northeast Forestry University. His research interests include sports sociology.