HIGHLIGHTS EXTRACTION IN SOCCER VIDEOS BASED ON GOAL-MOUTH DETECTION

Ying Yang[1, 2], *Shouxun Lin*[1], *Yongdong Zhang*[1], *and Sheng Tang*[1]
[1]Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China
[2]Graduate University of Chinese Academy of Sciences, Beijing 100085, China
E-mail: {yyang, sxlin, zhyd, ts}@ict.ac.cn,

ABSTRACT

A new method is proposed for highlight extraction in soccer videos based on goal-mouth detection. This approach is based on the observations that the appearance of goal-mouth points to a high likelihood of exciting action in soccer videos and that highlight is composed of certain types of scene views which exhibit certain transition rules. To exploit those observations, first goal-mouth are detected and segmented in soccer videos by Top-Hat Transform and some domain rules, and then with scene transition rules highlight is extracted based on goal-mouth detection. The effectiveness and efficiency of this approach are demonstrated by the experimental results on shot detection.

1. INTRODUCTION

To better serve the audiences and sports professionals, there is a great demand on automatic highlight extraction effectively and efficiently from sports videos, especially for the most popular sports, soccer. Many works have been done in highlight extraction in soccer videos. Replay based approaches assume that the replayed actions are usually highlights [1]. The work presented by Assfalg et al. used Hidden Markov Models to detect and recognize highlight in videos [2]. Gong et al. exploited the play field and player position information for the purpose of highlight extraction [3]. This approach appears to be fairly expensive. Wang et al. proposed a method based on Dynamic Bayesian Networks (DBNs) [4]. However, the feature extraction is the key problem to the DBNs.

In this paper, we present a novel and more generic approach for highlight extraction in soccer videos, which exploits the observations that the appearance of goal-mouth indicates a high likelihood of exciting action in soccer videos and that highlights are composed of certain types of scene views. A novel algorithm is proposed to detect and segment goal-mouth in videos based on Top-Hat Transform and some domain knowledge. Based on the results of goal-mouth detection, highlight is extracted from soccer videos with frame views transition rules.

The rest of the paper is organized as follows. In Section 2, the algorithm for goal-mouth detection and segmentation is described. In Section 3, we discuss the application of goal-mouth detection in highlights extraction. Experiment results for goal-mouth detection and highlight extraction are given in Section 4. Finally, conclusions are drawn and future work is discussed in Section 5.

2. GOAL-MOUTH DETECTION AND SEGMENTATION

Despite the importance of goal-mouth appearance in soccer game, little work has been done in goal-mouth detection for the complicated background of goal-mouth and the various lines in playfield. Yow et al. used a 3×3 mask and Hough Transform to detect goal-mouth [5]. However, this algorithm become invalidated when the video has low quality or the backgrounds of goal-mouth is complicated. The algorithm of goal-mouth detection proposed by Wan et al. is based on the localization of goal-line, which needs to compute CSR according to the grass coverage and dominant grass ratio [6]. The performance of the algorithm is highly dependent on the accuracy of grass extraction, and the CSR definition, which is crucial to the goal line localization.

Our method exploits some unique characteristics of goal-mouth in soccer videos to detect goal-mouth appearance based on Top-Hat Transform. The algorithm is composed of three stages: 1) Edge detection, which can detect and segment the goal-mouth edges in the image. 2) Potential goalposts detection, which can get the vertical lines satisfying certain conditions. 3) Goalposts determination, which determines whether the two goalposts are present by some heuristic rules. The details of the algorithm are described as follows. The data we used below are MPEG-1 compressed videos, and the resolution of the video is CIF Resolution (352×288).

2.1. Edge Detection

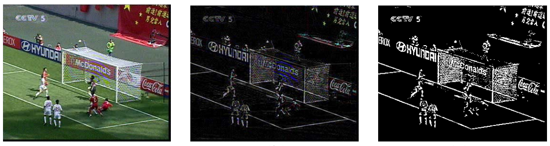
Top-Hat transform (THT) is one of the most important mathematical morphology operations, which can segment the wave crest in gray scale images. THT can be performed on color images to intensify white pixels. The color of the goal-mouth is always white whatever the video quality is, also the weather and lighting condition doesn't affect the color. Therefore the two goalposts can be clearly extracted after the THT operation, avoiding the irrelevant edges and noises caused by billboard or auditoria compared to the method proposed by Yow et al. [5]. The formula of THT is as follows:

$$h(x, y) = f(x, y) - (f(x, y) \circ b) \quad (1)$$

1-4244-0779-6/07/$20.00

Where $f(x,y)$ is the input image, $h(x,y)$ is the output image after THT operation, b is the structuring element, and $f(x,y)\circ b$ indicates the Open transform in the mathematical morphology operation. 5×5 structuring element is used in this paper.

A frame image taken from a soccer video is used to explain the edge detection procedure. It comprises the following two steps: 1) Performing THT operation on the input image and getting a color edge image, as shown in Figure.1 (b). 2) Converting the color edge image to a binary image and getting a black/white edge image, using a self-adaptive threshold: The histogram of the color edge image are split into two classes using K-means clustering (k = 2), as shown in Figure.1 (c).



(a) (b) (c)

Figure 1. The edge detection of an input image: (a) the input frame (the original image), (b) the image after THT operation (c) the binary image (the final edge image).

2.2. Potential Goalposts Detection

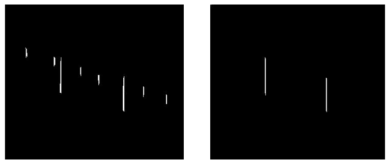
In spite of various video view types and camera motions, goalposts always appear as vertical lines $VL_l(x,y)$, unlike line-markings in the playfield, whose skew angle is not fixed. Therefore, satisfying some certain conditions are searched in binary edge image as potential goalposts $PGP_l(x,y)$.

The shortest length of the goalposts is set as 20 pixels according to the frame size and the discrimination of human. Potential goal pixel $PP(x,y)$ is belongs to the corresponding potential goalpost, and $PP(x,y)=1$ means point (x, y) is a potential goal pixels. So $PGP_l(x,y)$ are defined as follows:

$$PGP_l(x,y)=\begin{cases}1, & \forall (x,y)\in VL_l(x,y), \\ & \textit{if}\ \ PP(x,\ y)\ =\ 1 \\ 0, & \textit{otherwise}\end{cases} \tag{2}$$

$$PP(x,y)=\begin{cases}1, & \exists z\in[y\text{-}10,y+10], \forall i\in[\text{-}10,\ 10], \\ & \textit{if}\ Color(x,\ z+i)\ =\ White \\ 0, & \textit{otherwise}\end{cases} \tag{3}$$

The pixel color of point is (x, y) is denoted by $Color(x, y)$. In this way, all the potential goalposts are got by combining all the potential goal pixels in the vertical direction and these potential goalposts may have some gaps produced by Eq.2-3. The potential goalposts of Figure.1 (c) are shown in Figure.2 (a).



(a) (b)

Figure 2. (a) potential goalposts, (b) the final goalposts.

2.3. Goalposts Determination

The two longest vertical lines are chosen as goalposts among the potential goalposts by means of Hough transform. The goalposts and their distance in a frame must satisfy some rules, which are consistent with the laws established by authority such as FIFA and can be repeated for most soccer videos. Several heuristic and experiential domain rules derived from FIFA laws are applied to the two lines and their distances to determine whether a goal-mouth is present in a frame, where $Goal =$ 1 means a goal-mouth is present in current frame:

$$Goal=\begin{cases}1,\ \text{if} \begin{cases}20\ \text{pixels} \le L_i \le 0.3*H \ \ \text{and} \\ 20\ \text{pixels} \le D \le 0.4*W \ \ \text{and} \\ 0.5 \le D/\max(L_1,L_2) \le 2.5 \ \ \text{and} \\ 0 \le |(T_i - B_i) - L_i| \le 20\ \text{pixels}\ \ \text{and} \\ 1\ \text{pixel} \le O_v \le 0.3*H \ \ \text{and} \\ 0.5*H \le Y_c \le H \end{cases} \\ \qquad\qquad \text{where}\ i\ =1,2 \\ 0,\ \text{otherwise}\end{cases} \tag{4}$$

Where H and W is the pixel numbers of the height and width of a video frame respectively. L_i is the absolute length of each line, namely the actual pixel numbers, and D is the horizontal distance between the two lines. Where B_i and T_i is the lowest position and the highest position of each line respectively, and O_v is the overlap between the span of each line in vertical direction, the restrictions for these variables are avoid taking other white objects as goal-mouth. Since goal-mouth rarely occurs in the bottom part of video frame, the center pixel C (X_c , Y_c) of the minimal rectangle constructed by the two lines must in an appropriate area. According to above rules, the finally result of goal-mouth detection from Figure.1 (a) is shown in Figure.2 (b). Some goal-mouth detection results are shown in Figure.3.

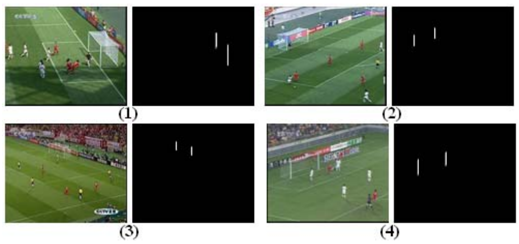


Figure 3. Various types of playfield in soccer videos and there goal-mouth.

3. HIGHLIGHT EXTRACTION BASED ON GOAL-MOUTH DETECTION

The main camera positioned along the long sides of the playfield always tracks the position of the ball and the action of the game. There exists a strong correlation between the ball positioned around the goal-mouth and highlights' happen. Therefore, the appearance of goal-mouth in a frame taken by the main camera indicates potential goal attempts, which are the most important highlights in soccer videos [5]. Furthermore, there are some production rules that sports video-makers usually follow, which aim to globally convey the status of the game and closely track the action [7]. After observed large amount of videos, some rules which meet above objectives are derived from various types of soccer videos to assist automatic highlight extraction. Shot event is used as example to discuss highlight extraction. Meanwhile view type information is important for highlight extraction, certain views sequence indicate a highlight. Each frame are classified into four views[8], named as long view, medium view, close-up view and out of field view according to the grass ratio and the player size, as shown in Figure.4.

(a) (b) (c)

Figure 4. (a) Long view, (b) medium view, (c) close-up view and out of field view.

When a shot event happens, these various types of views exhibit certain transition and editing rules. First shown are some long or medium views, which mostly include a goal-mouth, occasionally interrupted by short close-up views. These long or medium views will last until the end of a shot event, and then some close-up and out of field views will appear, whose content usually includes players who shot the ball.

Based on these rules and characteristics, we design the algorithm for highlight extraction as follows: For each input video frame, if it is long view LV or medium view MV, decide whether a goal-mouth is present in the current frame f, where GV denote to view including goal-mouth and NV denote to the view not including goal-mouth. If GV continually show for at least M frames and the number of NV don't exceed a threshold P, a shot event may exist, and then the view type of the following frames is investigated. If the total number of close-up CV and out of field views OV exceeds a threshold N, we assume a shot event happens. We set the threshold $M = 30$, $N = 20$, and $P = 100$. Figure.5 shows the pseudo-code for highlights extraction.

Other types of highlights can also be extracted by similar means. Such as corner kick and free kick event, usually a medium view within the player who kicks the ball shows first, then followed by a long view including a goal-mouth, which track the motion of ball towards the direction of goal-mouth. Therefore, goal-mouth appearance detection and scene views transition are the bases of highlights extraction.

```
set all counts to Zero
read the first frame to f
While f is not the end of the video
  while the type of f is not GV
      read the next frame to f
  while the type of f is LV or the type of f is MV and the count of NV<=P
      if the type of f is GV
        add 1 to the count of GV
      else add 1 to the count of NV
      read the next frame to f
  while the type of f is CV or the type of f is OV
      add 1 to the count of CV and OV
      read the next frame to f
  if the count of CV and OV >= N and the count of GV >= M
     A shot event
  set all counts to Zero
end While
```

Figure 5. The pseudo-code for highlight extraction.

4. EXPERIMENTAL RESULTS

The experiment is composed of two parts. We first evaluated the performance of our proposed goal-mouth detection algorithm, and then evaluated the highlights extraction. Experimental data comprised about 11 hours of MPEG-1 soccer videos without half-time break, including 7 full soccer games taken from different places and broadcasters, each is CIF-352×288×25fps. The parameters of each game video are shown in Table1. Where EB and HC have shadowed playfield, AB and BA are played in night, and HS has complicated field background.

Table.1 Test Videos

Name	Match	Duration	Goal-Mouth	Shot	Goal
EB	2002 FIFA World Cup England vs Brazil	1:39:43	94	15	3
CG	2002 FIFA World Cup Cameroon vs Germany	1:33:18	88	19	2
ES	2002 FIFA World Cup England vs Sweden	1:31:32	91	26	2
HC	2004 European Soccer Championship Holland vs Czech	1:35:10	96	33	5
AB	2006 UEFA Champions League ACMilan vs Barcelona (2006-4-19)	1:34:46	103	27	1
BA	2006 UEFA Champions League Barcelona vs Arsenal (2006-5-18)	1:38:01	102	30	3
HS	05/06 Germany Bundesliga I Hannover96 vs Stuttgart (2006-4-16)	1:31:51	113	28	6

The ground truth of goal-mouth occurrences in each game video is labeled manually by an independent person for each frame under the following directions:1) the whole goal-mouth be clearly seen, 2) at least shown for 30 frames, 3)each goal-mouth occurrence are labeled with a start frame and an end frame. The ground truth of goal-mouth occurrences may be larger than the work in [6] since we label the goal-mouth occurrence whatever the goal-mouth size is in a frame to avoid the differences in subjective judgments from various persons. The ground truth of shots including the wide shots and goals are from OPTA.

Table.2 Experimental results

Name	Goal-Mouth Detection		Highlights extraction	
	Recall	Precision	Shot Recall	Goal Recall
EB	91.49%	88.66%	86.67%	100%
CG	92.05%	87.10%	94.74%	100%
ES	96.70%	78.57%	88.46%	100%
HC	94.79%	81.98%	90.91%	100%
AB	87.38%	94.74%	96.30%	100%
BA	81.37%	90.22%	83.33%	100%
HS	81.42%	84.40%	82.14%	100%

Table2 shows the results of goal-mouth detection and highlights extraction. The results of goal-mouth detection by our proposed algorithm are satisfactory. The proposed algorithm detected 611 goal-mouths out of 687 in the test data. Thus, on the average, it achieved 88.94% recall and 86.18% precision rates. The misses are mainly from the long view frame, which includes very small goal-mouth and is the view of goalkeeper kick off or the break time during the game. Therefore, the misses are not critical goal-mouth occurrence, which don't affect the performance of its application in highlights extraction and can be overlooked. False alarms are accepted since audiences don't want to miss any highlights. The false alarms are due to the auditoria which may mixed white color objects. The area of edge detection can be limited to reduce the false alarms and further enhance the accuracy. For the first 3 videos, the recall rates of our algorithm is close to that of the algorithm proposed in [6] which is only tested on FIFA world cup videos. For the last 4 videos, both of the work in [5] and [6] failed to detect and segment goal-mouth accurately for the low quality of the videos. Overall, the algorithm achieved good performance on various videos with different stadium, weather conditions and broadcasting stations. Especially, the results of BA is good despite of the low video quality for the game played in a rainy day and the camera lens is blurred, which strongly demonstrate the robustness of our algorithm.

The results of highlights extraction are also encouraging with 100% recall rate of goal detection and average 88.76% recall rate of shot detection. The value of M, N and P are set just as in section3. The algorithm of highlights extraction took 17 hours to run the total 11 hours videos on a P4-1800MHz 512 MB PC, which achieved about 2 times real-time performances. The good results also further demonstrate the effectiveness and efficiency of goal-mouth detection algorithm and its application to highlight extraction. The misses are due to the wide shot which don't have entire goal-mouth and can't be detected, which don't belong to highlights for some audiences. Furthermore, above algorithms are successfully applied to the analysis of 2006 FIFA World Cup match, and all the highlights can be retrieved at http://souqiu.ict.ac.cn, which is our online soccer highlights retrieval system.

5. CONCLUSIONS

In this paper, we proposed a novel method for goal-mouth detection and its application in highlight extraction. The proposed goal-mouth detection methods are robust, effective and efficient for various types of soccer videos despite of the stadium, weather conditions and broadcasting stations according to the experimental results. The application of the proposed algorithm in highlights extraction is concise and efficient for the encouraging experimental results and achieving 2 times real-time performance. The application can be extended to video content analysis and soccer video summarization.

In the future work, we will further develop the application of goal-mouth detection with other features, such as text or audio information for highlights extraction, and use statistics model such as SVM to deduce more accurate and robust rules for advanced applications.

REFERENCES

[1] H.Pan, van Beek, M.I.Sezan, "Detection of slow-motion replay segment in sports video for highlights generation," in *IEEE Proc.ICASSP*, vol.3, pp.1649–1652, 2001.

[2] J.Assfalg, M.Bertini, A. Del Bimbo, W.Nunziati, P.Pala, "Soccer highlights detection and recognition using hmms," In *IEEE Proc. ICME*, pp. 825–828, 2002.

[3] Yihong Gong, Lim Tech Sin et al., "Automatic Parsing of TV Soccer Programs," in *IEEE Proc. Multimedia Computing and System*, pp. 167-174, 1995.

[4] Fei Wang, Yu-Fei Ma et al., "Dynamic Bayesian network based event detection for soccer highlights extraction," In *IEEE Proc.ICIP*, p.633–636, 2004.

[5] D.Yow, B.L.Yeo et al., "Analysis and presentation of soccer highlights from digital video," in *Proc. ACCV*, 1995.

[6] Kongwah Wan, Xin Yan et al., "Real-Time Goal-Mouth detection in MPEG Soccer Video," in *ACM Multimedia*, pp.311-314, 2003.

[7] P.Xu, L.Xie, S.-F.Chang et al., "Algorithms and system for segmentation and structure analysis in soccer video," In *IEEE Proc. ICME*, pp.928-931, 2001.

[8] Ahmet Ekin, A.Murat Tekalp et al., "Automatic Soccer Video Analysis and Summarization," in *IEEE Trans. Image Processing*, vol.12, pp.796-807, 2003.