# Goal event detection in soccer videos using multi-clues detection rules

SHI Ping, YU Xiao-qing
School of Information Engineering
Communication University of China
Beijing, China
shiping@cuc.edu.cn

***Abstract*—This paper proposed a framework for goal event detection in soccer videos by using multi-clues detection rules. In this framework, the visual clues including shot segmentation, shot classification and goal detection is extracted. Meanwhile, the audio clues including the audience's cheering and the commentator' excited speech are extracted. Then the goal event detection rules are defined by combining the visual clues, audio clues, the program composition and editing rules as well as the domain knowledge of soccer video. Experiments on real soccer video programs demonstrate encouraging results.**

***Keywords—goal event detection; video feature; shot segmentation; shot classification***

## I. INTRODUCTION

With the increasing amount of sports videos from broadcasters, the automatic analysis of sports video for efficient searching and browsing becomes more and more important. Soccer is one of the most popular sports in the world and the research on soccer video analysis has been addressed by a large number of researchers. In general, soccer video analysis includes shot boundary detection, shot classification, slow-motion replay detection, event detection, scene reconstruction, video annotation, and so on. Among them, goal event detection is probably the most active topic and has attracted a lot of attentions.

Gong et al. [1] use player, ball, line marks and motion features to parse TV soccer programs. Xie et al. [2] proposed a method to segment soccer video into "play" or "break" segments in a HMM (hidden Markov model) framework. Leonardi et al. [3] detect "goal" event in soccer video by camera motion analysis and shot boundary trigger in a Control Markov Chain framework. Assfalg et al. [4] use playfield zone classification, camera motion analysis and player's position to infer highlights of non-broadcast soccer video by FSM (finite state machine). Ding et al. [5] use Web-casting text as external knowledge to detect highlight boundary. In [6], goal events are detected by combining heuristic rules with unsupervised fuzzy c-means algorithm. Xie et al. [7] proposed a data mining framework which can greatly reduce the number of negative (non-event) instances and the feature dimension to facilitate the final event detection.

In this paper, we proposed a new framework for automatic goal event detection in soccer video. First, the video clues are extracted and this process consists of video feature extraction, shot segmentation, shot classification and goal detection. Meanwhile, the audio clues are extracted by detecting the audience's cheering and the commentator' excited speech. Then, the goal event detection rules are defined by combining the visual clues, audio clues, the program composition and editing rules as well as the domain knowledge of soccer video.

The rest of the paper is organized as follows. The video structure analysis is presented in section 2. Audio feature extraction is presented in section 3. The goal event detection rules is described in section 4. Experiment results are presented and analyzed in Section 5. Conclusions and future work are summarized in Section 6.

## II. VISUAL CLUES EXTRACTION

### *A. Video feature extraction*

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the A4 paper size. If you are using US letter-sized paper, please close this file and download the file for "MSW US ltr format".

The video features are extracted for subsequent structure analysis.

(1) The dominant color ratio $G_i$

$$G_i = P_g / P \tag{1}$$

where $P$ is the set of all pixels, and $P_g$ is the set of grass pixels.

(2) The absolute difference between two frames in their dominant color ratio $G_d$

$$G_d(i,k) = |G_i - G_{i-k}| \tag{2}$$

where the $G_i$ and $G_{i-k}$ are the dominant color ratio of the $i^{th}$ and $(i-k)^{th}$ frame respectively.

(3) The difference in color histogram similarity $H_d$

$$H_d(i,k) = |HI(i,k) - HI(i-k,k)| \tag{3}$$

This work is supported by program 211 of MOE China: Digital Media Technology and Art, No. 21103040115

$$HI(i,k) = \frac{1}{N}\sum_{m=1}^{N}\sum_{j=0}^{B_m-1}\min(H_i^m[j], H_{i-k}^m[j]) \quad (4)$$

The similarity between two histograms is measured by histogram intersection in Eq. 4, where the similarity between the $i^{th}$ and the $(i-k)^{th}$ frames, $HI(i,k)$, is computed. $N$ denotes the number of color components, and is three in this case, $B_m$ is the number of bins in the histogram of the $m^{th}$ color component, and $H_i^m$ is the normalized histogram of the $i^{th}$ frame for the same color component.

(4) The dominant color distribution ratio $R_{diff}$

$$R_{diff} = \frac{1}{2}\{|G_{R_1} - G_{R_2}| + |G_{R_2} - G_{R_3}|\} \quad (5)$$

The $R_{diff}$ is the mean value of the absolute differences of dominant color ratio between the block of $R_1$ and $R_2$, and between $R_2$ and $R_3$.

(5) The edge distribution ratio $E_{diff}$

$$E_{diff}(i) = Pedge(i) / P(i) \quad (6)$$

A frame is divided into 16*16 blocks. $E_{diff}(i)$ is the edge distribution ratio of $i^{th}$ block. $Pedge(i)$ is the edge distribution of the $i^{th}$ block and $P(i)$ is the set of all pixels in $i^{th}$ frame.

### *B. Shot segmentation*

Shot segmentation is usually the first step in video analysis. Soccer video is different from a generic video in terms that there exists a single dominant color background, the grass field, in successive shots. Therefore, a shot change may not result in a significant difference in the frame histograms.

In this paper, we use the algorithm proposed by Ahmet Ekin [8] to detect shot boundary. This algorithm, which is based on dominant color region, is effective for soccer video. Moreover, it is robust to variations in the dominant color. The color of the grass field may vary from stadium to stadium, and also as a function of the time of the day in the same stadium. The flowchart of this algorithm is shown in figure 1.

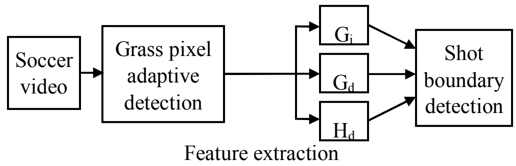


Figure 1 The flowchart of shot segmentation

### *C. Shot classification*

Based on the observation, there are four kinds of shots in soccer video as shown in figure 2. The first one is long shot, which is the global view of the field as shown in figure 2 (a). In this shot, the grass field covers a large area and players appear small. The second is in-field medium shot, which is a zoomed-in view of a specific part of the field as shown in figure 2 (b). There is still some area occupied by the grass in this shot. The third is close-up shot, which usually shows the above-waist view of one person as shown in figure 2 (c). In this kind of shot, there is very little area covered by grass area. The last one is out of field shot, which shows audience, coaches, referees or so on.

In this paper, we use the dominant color ratio, the dominant color distribution ratio and the edge distribution ratio to classify the shots. The flowchart of this algorithm is shown in figure 3.

### *D. Goal detection*

In general, the goal event is initiated by a long shot in which the goal usually appears. So goal detection is a crucial step in goal event detection. Although the shape of soccer goal is various in actual soccer video, the goal post always appear as two vertical lines. Therefore, the goal detection can be done by detecting the goal post. In this paper, we use the vertical line model, as shown in figure 4, to detect the goal post.

## III. AUDIO CLUES EXTRACTION

In soccer videos, the sound track mainly includes the foreground commentary and the background crowd noise. Based on the observation and prior knowledge, the commentator and crowd become excited during the goal event. Therefore, detection of the commentator' excited speech and the audience's cheering is helpful for goal events detection.

(a) (b) (c) (d)

Figure 2 Shot classification
(a) Long shot (b) in-field medium shot (c) close-up shot (d) out of field shot

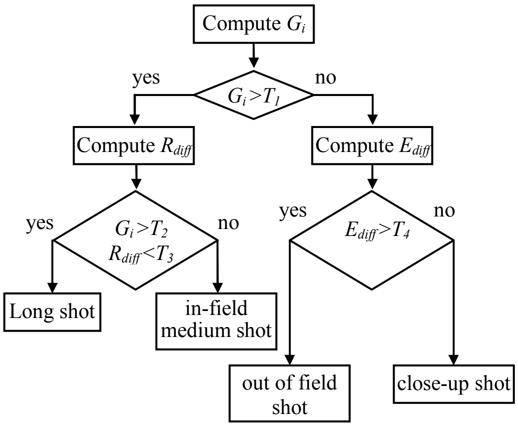


Figure 3 The flowchart of the shot classification algorithm

| | | |
|---|---|---|
| -1 | -2 | -1 |
| -1 | -2 | -1 |
| -1 | -2 | -1 |

Figure 4 Vertical lines detection model

### A. Detection of audience's cheering

We use the audio features of high volume rate and non-zero pitch ratio to detect audience's cheering.

The high volume rate ( $HRV$ ) is defined in (6).

$$HVR = \frac{1}{2N}\sum_{n=0}^{N-1}[\mathrm{sgn}(VOL(n) - 1.5avVOL) + 1] \tag{6}$$

The fundamental frequency can be determined from the average magnitude difference function (AMDF) which is defined in (7).

$$AMDF(m) = \frac{\sum_{n=0}^{N-l-1} | x(n+l) - x(n) |}{N-l} \quad l = 1,\cdots,N-1 \tag{7}$$

For an audio clip, if the fundamental frequency is around zero for a relative long period, and also the HVR meets the threshold, then we can judge that there is audience's cheering in this clip.

### B. Detection of commentator's excited speech

The audio features of high short time energy ratio and high auto correlation coefficients rate is used to detect the commentator's excited speech.

The high short time energy ratio (HSTER) is defined as (8).

$$HSTER = \frac{1}{2N}\sum_{n=0}^{N-1}[\mathrm{sgn}(STE(n) - 1.5avSTE) + 1\}] \tag{8}$$

The high auto correlation coefficients rate (HACCR) is defined as (9).

$$ACC(m,l) = \frac{\sum_{n=0}^{N-l-1} x(n)x(n+l)}{2N+1} \quad l = 1,\cdots,N-1 \tag{9}$$

For an audio clip, we can judge there is commentator's excited speech in this clip if both HSTER and HACCR meet the thresholds.

## IV. GOAL EVENT DETECTION

In typical soccer video production and editing patterns, a goal event is constituted by three consecutive shots starting from the goal shot and followed by close-up shot of a player and an out of field shot of the audience. In the mean time, the commentator's speech and the audience's cheering become excited. Here the goal shot refers to the long shot with the goal post in pictures.

According to the above observations, two rules are defined in this paper for goal event detection.

Rule 1: A goal shot is followed by a close-up shot and an out of field shot. Meanwhile, the commentator's speech and the audience's cheering become excited.

Sometimes the goal post cannot be detected because of the light change and camera motion. The rule 1 is unsuitable for such case. Therefore the second rule is defined as follow:

Rule 2: A long shot is followed by a close-up shot and an out of field shot. Meanwhile, the commentator's speech and the audience's cheering become excited.

The occurrence of a goal event is judged if three consecutive shots and corresponding audio features satisfy the rule 1 or rule 2.

## V. EXPERIMENT RESULTS

Experiments have been carried out with videos of soccer games recorded off-air from broadcasters. The overall test set includes 4 halves of soccer game with typical goal events. Table 1 shows the experiment results.

TABLE I. EXPERIMENT RESULTS

| Sequence | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Total goal events | 40 | 36 | 21 | 24 |
| Identified goal events | 47 | 40 | 25 | 28 |
| False | 12 | 10 | 8 | 10 |
| Miss | 5 | 6 | 4 | 6 |
| Recall (%) | 87.5 | 83.3 | 81 | 75 |
| Precision (%) | 74.5 | 75 | 68 | 64.3 |

The experiment results are encouraging. The average recall value is 81.7% and the average precision value is 70.5%.

## VI. CONCLUSIONS

In this paper, a new framework for soccer goal events detection is proposed. In this framework, two multi-rules detection rules are defined by combining the visual clues, audio clues, the program composition and editing rules as well as the domain knowledge of soccer video.

Although the experiment results are encouraging, the algorithm leaves much room for improvement and extension. For example, the video features such as slow-motion replay and camera motion can be integrated into this framework to increase the accuracy of goal event detection. In addition, the framework can be extended to other sports, such as basketball and baseball.

## REFERENCES

[1] Y. Gong, T.S. Lim, and H.C. Chua, “Automatic Parsing of TV Soccer Programs”, IEEE International Conference on Multimedia Computing and Systems, May, 1995

[2] L. Xie, P. Xu, S.-F. Chang, A. Divakaran and H. Sun, “Structure Analysis of Soccer Video with Domain Knowledge and Hidden Markov Models,” Pattern Recognition Letters, 2004

[3] R. Leonardi, P. Migliorati, and M. Prandini,“Semantic indexing of soccer audio-visual sequences: a multimodal approach based on controlled Markov chains,” IEEE Trans on CSVT 2004 Vol.14, No.5.

[4] J. Assfalg, Marco Bertini,Carlo Colombo, Alberto Del Bimbo, Walter Nunziati, "Semantic annotation of soccer videos: automatic highlights identification", Computer Vision and Image Understanding, Special isssue on video retrieval and summarization, Volume 92, Issue 2/3, November/ December 2003.

[5] Xi-Feng Ding, Yin-Jun Miao, Fan Bu, Li-Feng Sun, Shi-Qiang Yang, “Highlight detection in soccer video using web-casting text”, IEEE 10th Workshop on Multimedia Signal Processing, Oct. 2008, pp. 604 - 609

[6] Yina Han, Guizhong Liu, Chollet, G., “Goal event detection in broadcast soccer videos by combining heuristic rules with unsupervised fuzzy c-means algorithm”, 10th International Conference on Control, Automation, Robotics and Vision, 2008. ICARCV 2008, Dec. 2008, pp. 888 - 891

[7] Zongxing Xie, Mei-Ling Shyu, Shu-Ching Chen, “Video Event Detection with Combined Distance-Based and Rule-Based Data Mining Techniques”, IEEE International Conference on Multimedia and Expo, July 2007, pp. 2026 - 2029

[8] Ahmet Ekin. “Automatic Soccer Video Analysis and Summarization”. IEEE Transactions on Image Processing, July 2003