Highlight Summarization in Soccer Video based on Goalmouth Detection

Zhao Zhao[1], Shuqiang Jiang[1], Qingming Huang[1], Qixiang Ye[2]

[1]Inst. of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China
[2]Graduate School of Chinese Academy of Sciences, Beijing 100080, China

{zzhao, sqjiang, qmhuang, qxye}@jdl.ac.cn

Abstract

Highlight summarization technology has been widely investigated in sports videos. In this paper, we propose a specific method for highlight summarization in soccer video based on goalmouth frames detection and audio energy ranking. Firstly, we extract the video clips which contain the goalmouth area as the highlight candidates. Then, based on the commonsense that the volume and the length of the commentators' comments and audiences' acclaim to different events are various according to their exciting level, we employ the audio energy feature as the criterion to rank the highlight candidates so as to generate the summarization. The method proposed in this paper is intuitive and robust to generate highlight summarization in soccer videos. Besides, it also provides references for other video summarization works.

Keywords: highlight summarization, video analysis, goalmouth, soccer video.

1 Introduction

Video analysis technology has been widely investigated for the purpose of providing the users a more rapid and convenient access to the interesting or important part of the videos. Especially, accompanying with the increasing amount of the videos in TV, internet, etc, the demands to automatically analyse and summarize videos becomes larger. With these technologies, the users can save their time by only watching a summarization of some concerned parts of a program. Besides, Internet or the wireless communication networks can also be facilitated from these technologies because their limited band width does not suitable for transferring large amount of data. For video, to transfer summarization in these networks is obviously an apt substitution.

There are many kinds of broadcast programs which have been investigated including sports games as mentioned by Bonzanini, Leonardi and Migliorati (2001) and Xu and Duan (2003), news as mentioned by Zhang and Gong (1994), etc. Among them, a widely investigated genre of video is sports game because it holds a large amount of audiences and has more regular features than some other videos. Specifically, for sports video, an effective abstract approach is the highlight summarization technology which focuses on how to generate a summarization of a match that includes all the interesting parts of it. Some former researchers have proposed many highlight summarization methods both for general sports game and for a specific kind of sports game. Ekin and Murut (2003) detected the play and break event in sports videos to generate the summary. Some other researchers summarize sports videos using slow motion replays as mentioned by Pan, Beek and Sezan (2001) and Pan, Li and Sezan (2002). However, analysing general sports game is still an open problem because of the variance and diversity of different games. Some researchers turn to study specific sports game such as soccer, basketball or diving. Ancona and Cicirelli (2001) proposed a goal detection method using SVM classifiers in soccer videos. Xu and Duan (2003) investigated the event detection method for basketball videos. Nepal and Srinivasan (2001) researched on the basketball game and proposed the representation of the "goal" event in basketball game. Xing and Ye (2005) investigated the highlight summarization method in racquet sports such as tennis and tabletennis.

Considering that soccer is a favourite game all over the world. In this paper, we propose a convenient and pointed method for highlight summarization in soccer game videos. We noticed that most of the exciting events in a soccer game occur near the goalmouth area such as the shooting, penalty and direct free kick, etc. Based on this observation, we first extract color and shape features. Then an SVM model is hired to detect out all the frames containing the goalmouth area as the highlight candidates. Furthermore, the exciting level of a clip can be deduced by observing the audiences' reaction to it. The more excited the clip is, the more fervid the audiences applause as well as the commentator comments. So we rank the highlight of the candidate clips according to their audio energy and generate the highlight summarization of the soccer video.

This paper contributes that a method for the highlight summarization of soccer game is proposed which is intuitive and robust.

2 System Overview

The system transacts with a soccer game video and finally generates a summary which sequentially consists of a set of highlight clips. The framework of the highlight summarization procedure is depicted in Fig. 1 which has 2 steps. First, visual features are hired to find the highlight candidates which are the goalmouth scenarios. Then, the audio features are utilized to rank the highlight level of the

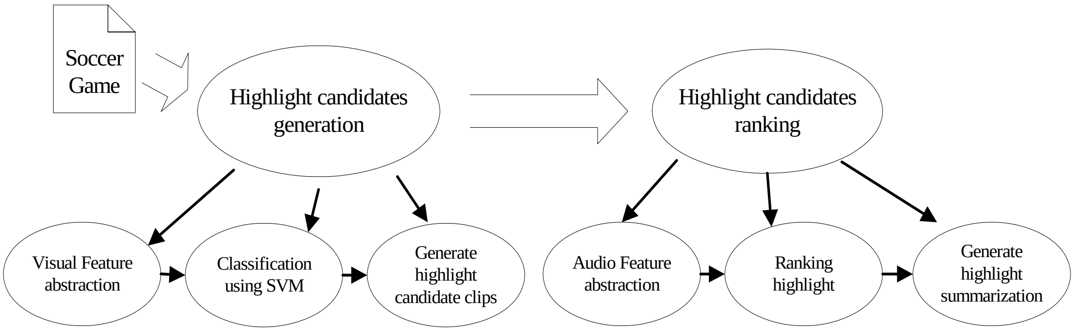

Figure 1: the soccer game highlights generation system

candidates and finally generate highlight summarization. The details will be elaborately narrated in the following Section 3 and Section 4.

3 Highlight Candidates Detection

As we have explained, for soccer video, the goalmouth scenarios can be selected as the highlight candidates, for the reason that most of the exciting events occurs in the goal mouth area such as the goal, shooting, penalty, direct freekick, etc. On the other hand, the non-goalmouth scenarios often consist of the dull passes in the midfield, defense and offense or some other shots to the audiences or coaches, etc, which are not as exciting as the former. So we managed to extract goalmouth scenarios from the soccer video as the highlight candidates. Fig. 2 shows two scenarios which illustrate the differences between a goalmouth and a non-goalmouth.

Figure 2: Two examples of goalmouth and non-goalmouth scenarios.

To detect goalmouth scenarios, we first classify all the frames of the video into two categories, the goalmouth frames and non-goalmouth frames. Then we obtain the goalmouth scenario clips which consist of sequential goalmouth frames.

It is apparent that the goalmouth frame has many differences with the non-goalmouth frame such as the shape and color ratio of the playfield. Here we use our previous color and shape features as mentioned in Ye and Huang (2005) to detect goalmouth frames as the following steps:

Detecting the playfield of a frame *(Section 3.1)*;

Extracting the shape and color features *(Section 3.2)*;

Classifying the frames into goalmouths or non goalmouths (*Section 3.3)*

3.1 Playfield detection

Here we employ our previous method as mentioned in Jiang and Ye (2004) to detect the playfield. We first accumulate the color of the training frames of the video to establish a Gaussian mixture model of the dominant color. Then the frames are mapped to a binary image according to the dominant color model. For each video, 100 frames are set to be the training samples and around 50000 pixels are extracted as training set. The GMM model can be denoted as follows:

$$p(\xi \mid \Phi) = \sum_{i=1}^{M} w_i b_i(\xi) \qquad (1)$$

where w_i is the mixture weight, $b_i(\xi)$ is the mixture component which is denoted as:

$$b_i(\xi) = \frac{1}{2\pi |\Sigma_i|^{1/2}} \exp\{-\frac{1}{2}(\xi - \mu_i)' \Sigma_i^{-1} (\xi - \mu_i)\} \qquad (2)$$

The GMM parameters are estimated by maximum likelihood learning with ordinary EM algorithm.

Then we use region growing algorithm to efface discrete points or small areas in the generated binary image as shown in Fig. 3.

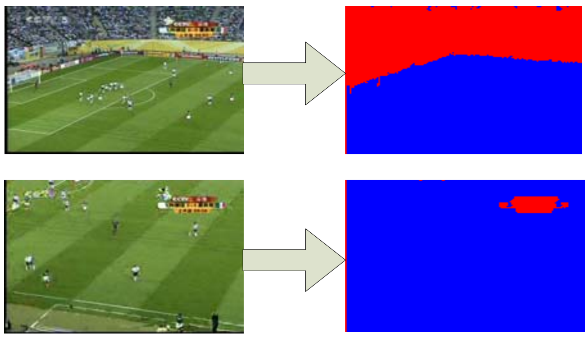

Figure 3: The playfield of frames

3.2 Feature Selection

In order to distinguish between goalmouth frames and non-goalmouth frames, we extract the shape and color features as the following narrated.

3.2.1 Color Features

We hire two color features which are the *Playfield Ratio(pr)* and *Projection Profile of non playfield (pp)*which is shown in Figure 4.

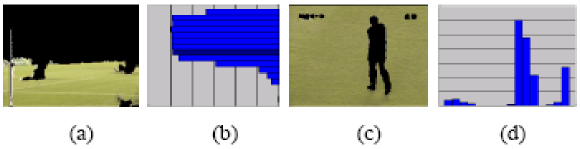

Figure 4:Non-playfield projection profile.

3.2.2 Shape Features

Five dimensions shape features are extracted which are:

ts - slope of top boundary;

ls - slope of top left boundary;

rs - slope of top right boundary;

cp - corner position;

bs -slope of bottom boundary.

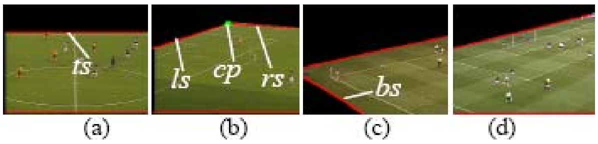

Figure 5: Shape features of a frame.

These shape features can provide information to describe where the main camera is focus on the field. For more detail about feature selection, see in Ye and Huang (2005).

3.3 Classifying

Here we employ an SVM model to classify between the goalmouths and non-goalmouths.

After the categorizing procedure of the frames is accomplished, we obtain the goalmouth clips which consist of sequential goalmouth frames. The results can be illustrated in Fig. 6.

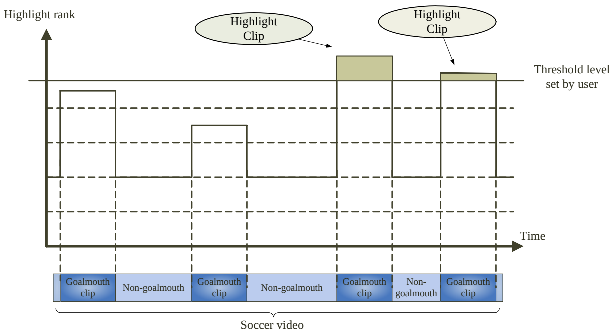

Figure 6: Goalmouth and non-goalmouth clips in a video

In Fig. 6, the soccer video is divided into goalmouth clips and non-goalmouth clips where the columns represent goalmouth clips, height of which denotes the highlight level of each clip which is explained in the next section.

4 Highlight Ranking

In soccer games, most of the exciting events are accompanied with the agitated comments of commentators and fervid applauses of the audiences. In addition, the more exciting the event is, the more intense the voice is and the longer it lasts. So we selected two audio features to represent the highlight level which are the average energy of the audio in the determined durations and the total length of them. Here we use a linear model to represent highlight level of clip *i*:

$$H_i = w_e E_i + w_t \sum_m T_{im} \tag{3}$$

where E_i and T_{im} is the average energy and length of the duration of the audio segment in a goalmouth clip. We determine the position and length of the audio segment by the following algorithm:

a. Set a threshold which can detect the fervid comments and applauses from the general comment and background noise.

b. Lengthen the goalmouth clip by t seconds from its end so that a new clip is constructed.

c. Step through the new clip to determine an audio segment whose short time average audio energy value is over the threshold. Then define the end of the audio clip to be the end of the new goalmouth clip named Cross Point.

d. Obtain the average energy feature and duration feature of the new goalmouth clip and go to the next clip.

The procedure can be illustrated in Fig. 7.

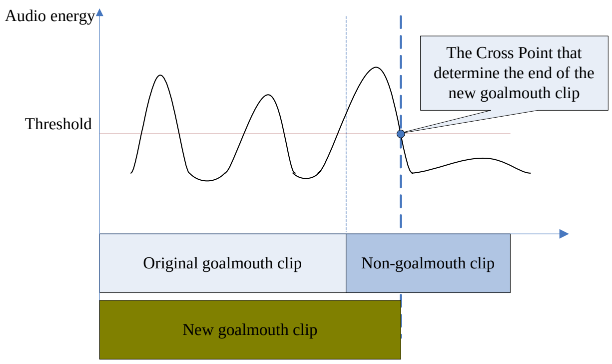

Figure 7: New goalmouth clips generation considering audio effect

This new goalmouth clip generation procedure guarantees that the comments or the applause of every highlight clip is entire thus makes the effect of the summary comfortable for the users. These new clips can be units of the highlight summary of the soccer game.

5 Experiments

We test our method on a set of soccer games to evaluate the validity of the goalmouth frame detection algorithm. The SVM model is generated from our former experiments and the test set is chosen from the EuroCup 2004 and Olympic 2004. Totally about 15000 frames have been tested. As the result shows, the performance of the classifier is feasible but not very high because the different camera position and editors' compiling habits in various games may cause some declination of the performance of classification.

	Precision	Recall
EuroCup 2004	87.2%	73.2%
Olympic 2004	86.7%	81.3%

Table 1: Experiment of the goalmouth frames detection

Then some of the soccer videos are extracted on which the highlight ranking algorithm is applied. Fig. 8 illustrates one experiment of the highlight ranking on part of a soccer video in EuroCup 2004 which is more than ten minutes. The temporal distribution of the goalmouth clips is not even, but for visual convenience, they are evenly deployed in Fig 8. The highlight levels distributes from 0 to 1 which are represented by the lengths of the columns. It is apparent that the highlight ranks of the goalmouth clips denote the exciting levels of the events occurring in these moments.

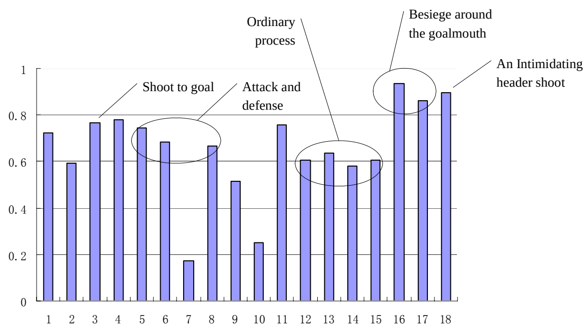

Figure 8: The highlight rank of the goalmouth clips in part of a soccer video.

The result corresponds well with the human's subjective emotion. From the ranking result we can summarize the video appropriately according to the highlight ranks of the goalmouth clips. In Fig.8, for example, if we set the highlight threshold at 0.8, the most exciting clips of this part will be extracted which contain several besieges around the goalmouth concluding with an intimidating header shoot to the goal.

6 Conclusion and Discussion

In this paper, we present a specific method for automated highlight summarization in soccer video. This method is based on the commonsense that most of the exciting events in soccer video occur in the area of goalmouth. So we first utilize our former field classification algorithm to extract goalmouth clips from the video. Then considering that the more exciting the events are, the more intensively the audiences and commentators react, we employ audio energy and applause duration as the features to calculate the highlight rank. After the goalmouth clips are ranked, the highlight summarization can be generated according to the individual demand.

The experiments show the feasibility of this method though the performance is not good enough such as for goalmouth detection and need to be improved in future works. Besides, how to model the highlight with better features is another issue that should be investigated in future works.

Acknowledgement

This work is partly supported by NEC Research China on "Context-based Multimedia Analysis and Retrieval Program", "Science100 Plan" of Chinese Academy of Sciences and "Beijing Natural Science Foundation: 4063041".

References

Bonzanini, A., Leonardi, R., Migliorati, P. (2001): Event recognition in sport programs using low-level motion indices. *IEEE International Conference on Multimedia and Expo.*

Min Xu, Ling-Yu Duan, Changsheng Xu, Kankanhalli, M., Qi Tian (2003): Event detection in basketball video using multiple modalities. *Joint Conference of the Fourth International Conference on the Fourth Pacific Rim Conference on Multimedia*, Volume: 3 , 15-18

HJ Zhang, Y Gong, SW Smoliar, SY Tan (1994): Automatic Parsing of News Video. *ICMCS*

Ahmet Ekin' and A. Murut Tekalp (2003): Generic play-break event detection for summarization and hierachycal sports video analysis. *IEEE International Conference on Multimedia and Expo*

H.Pan, P.Beek, M.Sezan (2001): Detection of slow-motion replay segments in sports video for highlights generation. *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*.

H. Pan, B. Li, and M. I. Sezan (2002): Automatic detection of replay segments in broadcast sports programs by detection of logos in scene transitions. *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*.

Ancona, N., Cicirelli, G., Branca, A., Distante, A. (2001): Goal detection in football by using support vector machines for classification, *International Joint Conference on Neural Networks*, Volume: 1, 15-19

Surya Nepal, Uma Srinivasan, Graham Reynolds (2001): Automatic detection of "Goal" segments in basketball videos. *Proceedings of the ninth ACM international conference on Multimedia.*

Shuqiang Jiang, Qixiang Ye, Wen Gao, Tiejun Huang (2004): A New Method to Segment Playfield and its Applications in Match Analysis in Sports Video. *ACM international conference on Multimedia.*

Qixiang Ye, Qingming Huang, Wen Gao, Shuqiang Jiang (2005): Exciting Event Detection in Broadcast Soccer Video with Mid-level Description and Incremental Learning. *ACM international conference on Multimedia.*

LY, Xing., Qixiang, Ye., Weigang, Zhang., Qingming, Huang. and Hua, Yu.(2005): A scheme for racquet sports video analysis with the combination of audio-visual information, *Visual Communications and Image Processing.*