Shot Type and Replay Detection for Soccer Video Parsing

NGOC NGUYEN† ATSUO YOSHITAKA††

Parsing the structure of soccer video plays an important role in semantic analysis of soccer video. In this paper, we present a method to detect shot type based on the size of players. Besides, replay detection algorithm based on logo image recognition also proposed. First, the candidate logo images are selected based on contrast image and histogram difference. From candidate logo images, the contrast logo template is calculated. Logo images are found by comparing the similarity between the contrast images of candidate logo images and the contrast logo template. Finally, the replay is detected by pairing the logos and finding the beginning and the end of logo transition. Experiments on three soccer matches showed that our method is effective for soccer videos.

1. Introduction

In recent years, with the development of high speed Internet, high capacity storage, and the presence of low-cost cameras, video data is extremely popular in offline storage and sharing networks. The need of analysis, retrieval, and summarization of such rich information became an urgent issue, and has attracted a lot of researchers. Especially, with a large pool of audience worldwide, the sport videos analysis has become the hot topic due to its high commercial potential. One of the challenges of semantic content understanding is bridging the gap between high-level semantics and low-level features.

Shot classification and replay detection are fundamental work for video structure and semantic analysis; therefore, many approaches have been reported in literature. For shot type classification, Xu et al. [7] proposed a simple method that is only based on obtaining adaptive thresholds for grass area ratio feature. They assumed that there is a large amount of grass pixels in long view, some grass pixels in middle view and the lowest grass pixels in close view. Such assumption is not always reasonable since the number of grass pixels in some close views is higher than the number of grass pixels in some middle views. Ekin et al. [1] proposed a method based on the grass area ratio and the spatial information of the grass pixels. The grass region was divided in 3:5:3 proportions in both directions by using Golden section spatial composition rule. The shot type was classified based on the grass pixels ratio of middle part and the difference of grass ratio between the middle part and two neighboring parts. However, this method could not differentiate between middle view and close view as shown in Figure 3 (b) (c). Yang et al. [6] categorized shot type based on grass area ratio, non grass area ratio and location of top grass pixels. Top grass pixels were approximated by multiple lines, and the differences of x-coordinate and y-coordinate were used to recognize middle view. However, the middle views whose the whole body of player is in the middle of grass pixels as shown in Figure 3 (b) are misclassified as close views.

H. Pan et al. [2] extracted slow-motion segments for detecting replay segments. Hidden Markov model was used to model the structure of slow-motion segments, and a zero crossing measure was used to measure the frequency and amplitude of the fluctuations of adjacent frame difference. However, the algorithm could not detect the slow-motion segments which are shot by a high-speed camera. Besides, in some broadcasting programs, replay segment is only slowed down in small portion of segment; the rest of segment is played in normal speed. Therefore, H. Pan et al. [3] proposed an automatic algorithm by determining logo template from frames surrounding slow-motion segments. A replay segment was identified by grouping the detected logo frames and slow-motion segments. Xiaofeng et al. [4] proposed replay detection algorithm which is only based on logo detection. Logo transition was detected based on intensity mean square difference, and logo template was extracted from logo samples. Then, SVM classifier using shot and motion features was employed to identify replay segments. However, the logos of different broadcasters are different, and some broadcasting programs do not use logo transition to mark the beginning and the end of replay segments. Jinjun et al. [5] proposed the method that learned the pattern of shot types in replay segments. The pattern of shot types was used to detect replay segments. However, the result of this method is not high due to the dependency of shot classification result.

In this paper, we propose an effective shot classification method based on dominant color and the sizes of players. It is clearly to see that the sizes of players are different in different view types. The replay detection algorithm is also proposed which is based on contrast feature. Contrast feature is selected because most of logos of broadcasters whose colors is contrasted to the rest of frame. The rest of paper is organized as follows. Section 2 describes grass field extraction and the rule for shot classification. The detail of replay detection method is presented in Section 3. In Section 4, we analyze the effect of using player sizes in shot classification, and using contrast feature in replay detection. Finally, we conclude the paper and mention future work in Section 5.

2. Shot type classification

2.1 Grass field extraction

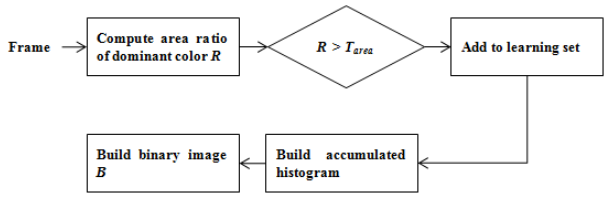


Figure 1 The flowchart of the dominant color detection

University of Information Science – Japan Advanced Institute of Science and Technology

algorithm

The soccer field surface is colored with a tone of green, which varies according to stadium, weather, and lighting conditions. In our work, we assume that the soccer field has one distinct dominant color and it covers a large area of frame. In order to detect grass field color, there are two main steps which consist of the organization of learning set and extraction of dominant color. The flowchart of the dominant color detection algorithm is shown in Figure 1.

First, to be able to deal with grass field color variations due to view angle, weather, the L equally spaced frames are selected to compute the area ratio of dominant color R, which is defined by the number of dominant color pixels divided by the number of pixels in a frame. For each frame i, 2D histogram H in rg color space is computed, and the peak i_{peak} of histogram is determined. An interval $[i_{min}, i_{max}]$ is defined for each component where i_{min} and i_{max} satisfy the following conditions.

$$i_{min} \leq i_{peak} \leq i_{max}$$
$$H[i_{min}] \geq k * H[i_{peak}]$$
$$H[i_{min} - 1] < k * H[i_{peak}]$$
$$H[i_{max}] \geq k * H[i_{peak}]$$
$$H[i_{max} + 1] < k * H[i_{peak}]$$

The conditions define the minimum (maximum) index i_{min} as the smallest index to the left (right) of the peak i_{peak} that has the predefined number of pixels ($k = 0.1$). The pixel that is included in the bin index in the range of $[i_{min}, i_{max}]$ is called dominant color pixel. The frames which ratio of dominant color is larger than the threshold T_{area} are added into learning set.

Second, N frames in the learning set are used to compute accumulated histogram. Dominant color pixels are also defined so as to follow the above conditions, and most of grass field areas are identified by the dominant color. However, in many cases, there remain some areas of soccer field, which have different color due to the lightning condition. Most of these areas are usually close to dominant color. Moreover, because the color of uniform of player is prominent compared to soccer field, the range of dominant color is safely expanded without losing the players. In this paper, the following procedure is applied to detect dominant color pixels:

- Build binary image B whose width and height are bin sizes of r and g respectively. Set zero values for all pixels.
- Find the peak (r_{peak}, g_{peak}) of histogram, H_{peak} (the value of histogram at peak) and set $B[r_{peak}, g_{peak}] = 255$
- Find all pairs (r, g) that the values of histogram at (r, g) are larger than the predefined number of pixels $k * H_{peak}$, and set $B[r, g] = 255$
- For each pair (r, g), set 255 for 8 neighbors of $B[r, g]$: $B[r-1, g]$, $B[r+1, g]$, $B[r, g-1]$, $B[r, g+1]$, $B[r-1, g]$, $B[r-1, g-1]$, $B[r-1, g+1]$, $B[r+1, g-1]$, $B[r+1, g+1]$
- Each pixel has the corresponding bin index (r, g). The dominant color pixel is the pixel whose the value of B at (r, g) is 255

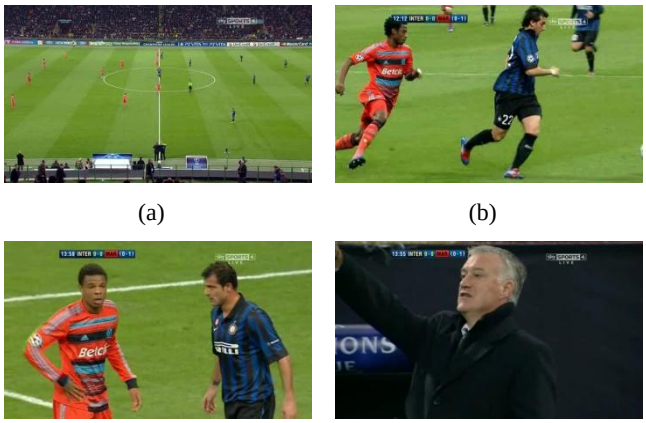
The playfield mask image in which white pixels represent field region and black pixels represent out of field region is created based on dominant color. However, the playfield mask can contain noise because of the following reasons: some audiences have the similar color to dominant color, or field regions are cut into small area due to white lines or players. The field region is determined by joining the close connected component. **Error! Reference source not found.** shows the result of grass field extraction.

(a) (b)

Figure 2 Grass field extraction (a) Original image (b) Playfield mask image *

2.2 Shot type classification

Shot type conveys interesting semantic information, and plays important role in analysis and summarization problem. We classify the shot type into three as defined in [1]: long view, medium view, and close view. Long views are usually shot so as to include the soccer field which aim to show the global view of the field to catch team tactics. Middle views consist of several players in a zoom-in view or a whole body of single player to catch the detailed movement of player. Close views or out of field views display only half body of players or the scenes of audience. Since the close views and out of field views capture the emotion of players or audience, and usually appear when special events occur, such as foul or getting a goal, we classify out of field views and close views as "close view". Figure 3 describes all shot types in soccer video.



(a) (b)

(c) (d)

Figure 3 Shot types in soccer (a) Long view (b) Middle view (c) Close view (d) Close view *

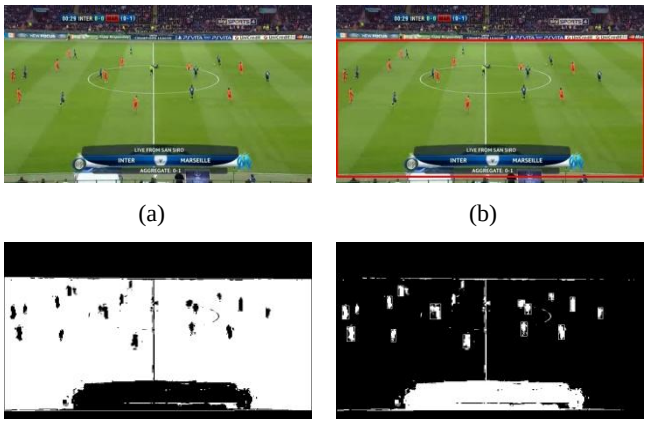
It is clearly to see that the sizes of players in different shot types are different, for example, the sizes of players in long shot

* Inter Milan vs. Marseille on March 13, 2012 http://www.fullmatches.net/inter-milan-vs-marseille-13-mar-2012-full-match-dawnload/

** Getafe vs. Valencia on March 24, 2012 http://www.fullmatches.net/getafe-vs-valencia-24-mar-2012-full-match-dawnload-la-liga/

are smallest. In this paper, we classify shot view type based on the following classification procedure:

- If grass colored pixel ratio in the i^{th} frame, G_i, is less than the threshold T_{Close} (in our experiment $T_{Close} = 0.1$), it is easily to classify these frames as "close view". G_i is the ratio of grass colored pixel to total number of pixels.
- Player mask image in which white pixels represent player region in the soccer field is extracted from playfield mask image. In player mask image, we find the contours which have sizes in the range of (T_{width}, T_{height}). Then we erase the contours that have small sizes, and find the largest contour whose the area of white pixels is larger than a threshold (for example, 0.4). The frame is classified as long view if it satisfies the following conditions: 1) The grass colored pixel ratio is larger than the threshold T_G 2) The number of contours whose sizes are in range of (T_{width}, T_{height}) is larger then T_{count} 3) Do not exist the large contour whose size is larger than $(3 * T_{width}, 3 * T_{height})$. Figure 4 illustrates the procedure to classify long view.
- We find the approximate rectangle for grass field region, and the largest contour *C*. The frame is classified as close view if it satisfies the following conditions: 1) The approximately rectangle of grass field region is close to the size of frame (95% of frame size) 2) The width and the height of largest contour are larger than the thresholds 3) The y-coordinate of the largest contour is larger than 95% of the height of frame size.
- The other cases are considered as middle views.



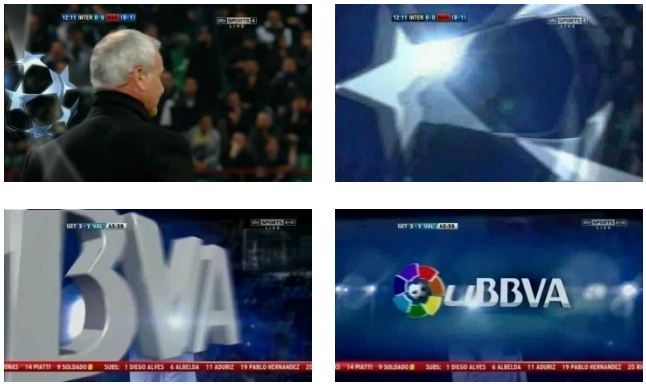
(a) (b)

(c) (d)

Figure 4 Long view classification (a) Original image (b) Grass field area (c) Playfield mask image (d) Player mask image *

3. Replay detection

Replay is one of important video editing ways in broadcasting program in order to let the audiences explore and view the details of important and interesting segments. Replay is usually played with a slow motion. Besides, in most cases, there exist the gradual transitions that precede and follow replay segments. These gradual transitions often contain the logos of broadcasters which are inserted special effects as shown in Figure 5. We call these transitions as logo transitions. As described above, there are advantage and disadvantage in the two approaches: one is based on slow motion, and another is based on logo transition. The difficulty of logo detection method is that the logos of different broadcasting programs are different. Besides, some programs do not use logo transition to mark the beginning and the end of replay segments. In this paper, we use logo to detect replay segments, and there are four main steps: finding the set of candidate logo images based on contrast image and histogram difference, calculating the contrast of logo template image, finding the beginning and the end of logo transition, and matching logo pairs.



(a) (b)

Figure 5 Logo transition in Champion League and La liga * **

3.1 Find the set of candidate logo images

In common, logo transitions contain logos of broadcasters whose color is contrasted to the rest of frame. Therefore, logo images are images whose contrast values are high enough to be separated from the rest of video frames. First, to calculate the contrast value, each frame is converted into binary image. The binary image contains noise when some field pixels are brighter than the threshold. Therefore, to reduce noise, all field pixels are changed into black values. The contrast value of frame is the ratio of white pixels in binary image, and is denoted R_w. If R_w satisfies $R_w \geq T_{contrast}$, this frame is added into the set of logo candidate images.

However, the set of logo candidate images may contain many errors due to the appearance of players, goalpost. We reduce the wrong logo candidate images based on the following heuristics: if the i^{th} frame belongs to the segment of a logo transition, this frame must be different from both of the $(i-k)^{th}$ frame and $(i+k)^{th}$ frame. The procedure of discarding wrong logo candidate images is described as follows

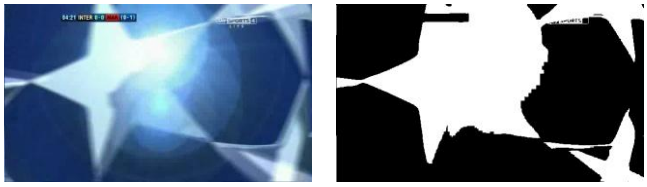
- Compute the histogram of i^{th} candidate logo image
- Compute the histogram of $(i-k)^{th}$ frame and histogram of $(i+k)^{th}$ frame. (In the experiment, $k = 15$)
- Compute the difference of histogram of i^{th} candidate logo image and the histogram of $(i-k)^{th}$, and that of i^{th} candidate logo image and the histogram of $(i+k)^{th}$ frames as the following formula, where H is the intensity histogram, i and j are frame indices

$$Diff(i,j) = \sum_{l=0}^{BIN} \frac{\left(H_i[l] - H_j[l]\right)^2}{\max(H_i[l], H_j[l])}$$

- If $Diff(i, i-k) < T_{Diff}$ or $Diff(i, i+k) < T_{Diff}$, this

image is discarded from the set of logo candidate images.

Logo transitions may contain a sequence of consecutive images that also have large contrast values. However, only one image in the sequence has full logo of broadcasting programs, and it is often the case where full logo is displayed in the image that has the largest contrast value. Based on this observation, with each i^{th} candidate logo image, if there exist the candidate logo images in the next 15 frames, we keep the candidate logo image that has the largest contrast value, and discard the remaining candidate logo images.



(a) (b)

Figure 6 (a) Logo image (b) Contrast logo template image *

3.2 Contrast logo template image

We randomly chose 20 candidate logo images, and calculate the average contrast image. This average contrast image is converted into binary image, which is called contrast logo template image as shown in Figure 6. Each image in the set of candidate logo images is compared with the contrast logo template by the following formula, where I_i is the contrast image of i^{th} candidate logo image, I_j is the contrast logo template, M and N are the width and height of images. The candidate logo images whose $Diff_{contrast}$ is larger than threshold (10% of total pixels) will be removed.

$$Diff_{contrast}\ (i,j) = \frac{1}{M * N} \sum_{m=0}^{M} \sum_{n=0}^{N} \frac{\left| I_i(m,n) - I_j(m,n) \right|}{255}$$

3.3 Find the beginning and the end of logo transition

Once the features of logo images are captured, the next procedure is to find the beginning and the end of logo transition. To find the beginning of logo transition, for each i^{th} candidate logo image, we will examine the successive frames from $(i-k)^{th}$ frame to i^{th} frame by the following procedure.

- Find the average contrast difference of the shot preceding replay segment by the following formula (in the experiment $m = 5$ and $k = 20$)

$$Diff_{avg} = \frac{\sum_{l=i-k}^{i-k+m} |R_w(l+1) - R_w(l)|}{m}$$

  where $R_w(l)$ is the contrast value of l^{th} image, m is the number of frames which we run to find the average contrast difference.
- Calculate the contrast difference of two successive frames, and if this difference changes more than 50% of average contrast difference, this frame will be marked as the beginning of logo transition

The same procedure is carried out for finding the end of logo transition, but the frames will be examined from $(i+k)^{th}$ frame back to i^{th} frame.

3.4 Match logo pairs

A typical replay segments is approximately less than 1000 frames. According this evidence, with each logo image, if the number of frames from current logo image to next logo image is larger than 1000 frames, current logo image is considered as faulty detection, and it will be discarded.

4. Experiments

In order to evaluate our algorithm, several representative soccer videos are picked up, including three half-time long (45 minutes) soccer videos, i.e., 135 minutes in total. Table 1 shows the result of shot type classification for three matches. The misclassification occurred between long view – middle view and middle view-close view. The long view is misclassified as middle view when the sizes of players in long view are larger than average size. Some middle views which have small grass colored pixels ratio is misclassified as close views.

Table 2 shows the results of replay detection. The wrong cases occurred in the images that have large contrast value. Our method couldn't detect replay segments where the producers don't use logo to mark at the beginning and the end of replay segments. However, our method works well for popular broadcasting programs, and the precision of replay detection is high enough to be applied for soccer video summarization.

Game	Numbers of shot	Precision
Inter Milan vs. Marseille	380	82.1%
Real Madrid vs. CSKA Moscow	100	84%
Getafe vs. Valencia	180	81.1%

Table 1 The result of shot classification

Game	Numbers of replay	Precision
Inter Milan vs. Marseille	29	100%
Real Madrid vs. CSKA Moscow	25	96%
Getafe vs. Valencia	25	88%

Table 2 The result of replay detection

5. Conclusion and Future work

Shot type classification and replay detection are fundamental work of semantic analysis for sports video. In this paper, we proposed a shot type classification algorithm that is based on dominant color and the sizes of players. Besides, replay detection algorithm based on contrast feature is also proposed. Our experimental results showed that our method is effective enough to use for high level analysis. However, our method is currently applicable for soccer videos analysis. In the future, we apply the results of shot type classification and replay detection to build the summarization framework for soccer videos.

Reference

[1] A. Ekin, and A.M. Tekalp, “Shot type classification by dominant color for sports video segmentation and summarization,” *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. III-173-6, 2003.

[2] H. Pan, P. van Beek, and M.I. Sezan, “Detection of slow-motion replay segments in sports video for highlights generation,” *IEEE International Conference on Acoustic, Speech, and Signal Processing*, pp. 1649-1652, 2001.

[3] H. Pan, B. Li, and M.I. Sezan, “Automatic detection of replay segments in broadcast sports programs by detection of logos in scene transitions,” *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 3385-3388, 2002.

[4] T. Xiaofeng, L. Hanqing, L. Qingshan, J. Hongliang, “Replay detection in broadcasting sports video,” *Proceedings of Third International Conference on Image and Graphics*, pp. 337-340, 2004.

[5] W. Jinjun, C. Engsiong, X. Changsheng, “Soccer replay detection using scene transition structure analysis,” *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2005.

[6] Bo Yang, Li-Feng Sun, Fei Wand, Peng Wang, and Shi-Qiang Yang, “Mid-Level Descriptors Extraction of Soccer Video with Domain Knowledge,” *IEEE International Conference on Systems, Man and Cybernetics*, pp. 4937-4941, 2006.

[7] P. Xu, L. Xie, S.F. Chang, A. Divakaran, A. Vetro, and H.F. Sun, “Algorithms and system for segmentation and structure analysis in soccer video,” *IEEE Conference on Multimedia and Expo*, pp. 928-931, 2001.

[8] W. Lei, L. Xu, S. Lin, X. Guangyou, and S. Heung-Yeung, “Generic slow-motion replay detection in sports video,” *International Conference on Image Processing*, pp. 1585-1588, 2004.