FIELD LINES AND PLAYERS DETECTION AND RECOGNITION IN SOCCER VIDEO

Li Sun, Guizhong Liu

School of Electronics and Information Engineering, Xi'an Jiaotong University, 710049, China

ABSTRACT

Objects like field lines and players are important for semantic analysis in soccer video. In this paper, we propose effective methods for field lines and players detection and recognition. Regions of field lines and players are first segmented from shot of wide angle view. Gray value top-hat transform is then performed on the segmented region to detect field lines. Mid-lines, end-lines and penalty-lines are then recognized based on their positions. Template matching method is adopted to detect and recognize players. Four types of templates for players are matched on the segmented region. Accurate position of player can be obtained after template matching, and the type of player is also identified. Experiment results on various data-sets with different production styles show effectiveness of our method.

***Index Terms*—** Image object detection, image object recognition, image region analysis.

1. INTRODUCTION

For decades, semantic analysis of soccer video has become an active research topic. Semantic events, such as goal, shoot, corner kick and penalty, are attractive for audiences and are useful for editors to generate highlights of a match. But these events are still difficult to detect for following two reasons. The main reason is the well-known semantic gap between low-level features and rich meanings of events [1]. Another reason is that broadcasted soccer video does not have fixed structures [2], which makes it difficult to decide when special events have started. Objects like field lines and players provide us important information for special events. They can fill the semantic gap to a certain extent. Field lines are solid cues to infer the active area on the playfield, while players are directly involved in many events.

There exist many works on analysis of soccer video. But only a few of them focus on detection and recognition of field lines and players. In [3], many features which can used for event detection in soccer video are described. But the author does not give details about how to extract these features. Luo *et al.* propose a soccer pyramid structure for highlight extraction [4]. In the pyramid, field lines are important feature for shot classification. A soccer field tracking method is proposed in [5]. Field lines are extracted by some thresholds in HSI color space. But this is not robust for color variation. Yu *et al.* detect many objects in soccer video [6]. Because they focus on tracking ball, so other detection algorithm is not accurate enough.

In this paper, we propose detection and recognition methods for field lines and players. First the method proposed in [7] is adopted to recognize wide angle shot in soccer video. As a by-product in this step, a binary field area mask and player mask are also obtained. Then combined with these two masks, field lines and players are detected and recognized in wide angle shot. Gray value top-hat transform is performed on the field region. The type of every line is recognized based on its position. Players are detected using template matching method. Four types of player templates on both teams are first manually segmented from soccer video. For each player candidate region, we can find a best matched template among the four types. Accurate player localization and player type recognition are accomplished at the same time. Compared with previous work, our work makes contribution on following two aspects: 1) Not only field lines are detected, but also types of field lines are recognized. 2) Accurate player position and player type are obtained efficiently. These achievements can provide us useful cues for semantic analysis of soccer video in the future.

The remainder of this paper is organized as follows: In section 2, we briefly describe the method of shot type classfication. In section 3 and section 4, the proposed detection and recognition methods for field lines and players are described in details. Experimental results and discussions are given in Section 5. Conclusions are finally drawn in section 6.

2. SHOT CLASSIFICATION

In this section, the method for shot type classification and object candidate region generation proposed in [7] is briefly explained. There are three basic types of shot in broadcasted soccer video, which are wide angle view, medium view and close up. We only focus on the first one because it provides much more information that the other two types. In order to recognize wide angle view shot, the first important step is

* This work is partially supported by China National Natural Science Foundations (No.60272072, No.60572045), Ministry of Education of China Trans- Century Elitists Project (TCEP, 2002), and Microsoft Research Asia.

to get the color of playfield. We hold the assumption that the color of playfield is the dominant color in whole soccer video. So accumulated histograms are adopted to learn the field color. It is accomplished in HSI color space. Cylindrical distance between color of testing pixel and dominant field color is calculated. We set a threshold for the distance and the binary image can be obtained as is shown in Fig.1(b). The ratio of field color in the image is used for categorizing the shot type, which is defined as follows:

$$F_d = \frac{N_d}{W \times H} \tag{1}$$

In (1), F_d is field color ratio of an image and N_d is the number of pixels with field color. W and H are width and height of the image respectively. Because frame in wide angle shot is usually of high ratio of field color, we can simply get wide angle frame by a suitable threshold. Then morphological filtering is performed on the binary image to remove the black blobs caused by players or field lines so that we get the mask of field area. The result is illustrated as Fig.1(c). The blob of player is usually of certain area and size, so we simply use the constraint of area and height-width ratio of the blob to coarsely get players mask as is shown in Fig.1(d). Note that the result of player mask is not perfect because of the failure of dominant color model in some region and the influence by white lines in the field area. We will show how to refine the result in section 4.

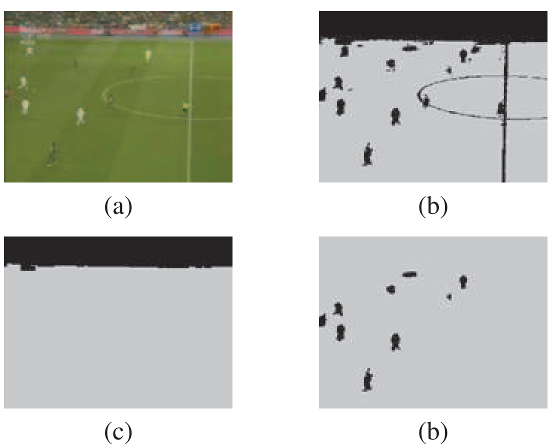


Fig. 1. Result of wide angle image segmented by dominant field color. (a) Original image; (b) Binary image segmented by dominant color; (c) Field area mask obtained by performing morphological operation on (b); (d) Coarse player mask.

3. FIELD LINE DETECTION AND LINE TYPE RECOGNITION

In this section, we describe our method for field line detection and line type recognition. It is based on field area mask and player mask obtained in previous section, which can easily eliminate noises coming from outside field.

3.1. Field line detection

The morphological top-hat operator for gray scale images is one of the basic tool for image processing. Its function is to detect contrasted objects on non-uniform backgrounds. For gray scale images, there are two types of top-hat: the white top-hat and the black top-hat. The previous one can extract bright structures while the latter one extracts dark structures. In this paper, only the white top-hat operator is used because field lines are normally white in video and they are much brighter than its background. The definition is like this:

$$T_w(f) = f - f \circ b \tag{2}$$

Here f is the original image. b is a grayscale structuring element. $\circ$ denotes the opening operation. From equation (2), we can see that the white top hat transform is the difference between original image and its opening.

The field line detection method is executed as follows: 1) Apply the top-hat operator on the gray level image of wide angle view shot and binarize the result with a suitable threshold, as shown in Fig.2(b). 2) Use field area mask and player mask to eliminate the noises caused by the out of field and players. The result is illustrated in Fig.2(c). 3) Thinning algorithm is performed on Fig.2(c) so that we get the final result of field lines, as is shown in Fig.2(d)

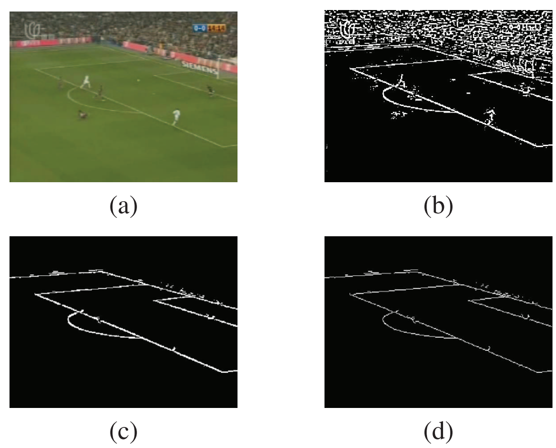


Fig. 2. Result of field line detection. (a) Original image; (b) Image after top-hat transform; (c) Image removed out of field area and players; (d) Image after performed thinning algorithm.

3.2. Field line type recognition

There exist five types of field lines in a soccer field, which are mid-lines, side-lines, end-lines, penalty-lines and some circles. In this paper, we only focus on mid-lines, end-lines and penalty-lines because they are useful cues for event detection. The production style of soccer video varies greatly in different country, so it is very difficult to develop a general algorithm to recognize every line in the soccer field at one time. Our strategy is to achieve line recognition step by step, from the easy one to the hard one.

First, Hough transform is performed on the binary image of line detection result like Fig.2(d). Every line is represented by two numbers (ρ, θ) in Hough transform, where ρ is a distance between $(0, 0)$ point and the line, and θ is the angle between x-axis and the normal to the line. Length of the line L is also important for following work.

Then mid-lines are first recognized, because mid-lines are long, noticeable and almost vertical no matter in which kind of production style. Here we set two conditions for mid-lines, which are:

$$\theta < TH_\theta \quad L > TH_l \tag{3}$$

TH_θ and TH_l are two thresholds for the angle of the line and the length of the line. The first condition is to ensure that the line is almost vertical and the second condition is to guarantee that it is long enough. If there are no mid-lines in current image, we continue to detect whether there exist end-lines and penalty lines. The conditions below are for end-lines:

$$\theta_{min} < |\theta| < \theta_{max} \quad Dis < TDis_{min} \tag{4}$$

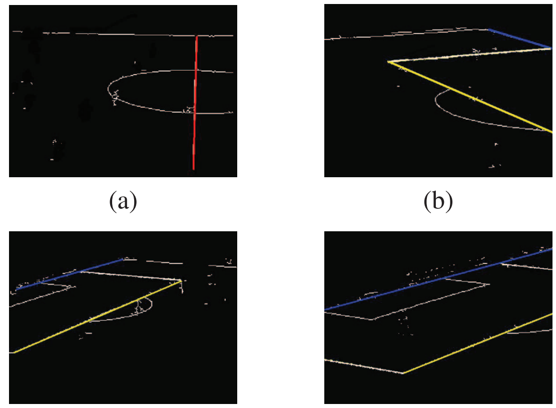
The first one is still the angle condition. It can eliminate side-lines which are almost horizontal. In the second condition Dis is the distance between the midpoint on the line and the border of field area. It must be small enough because end-lines are always near the border of field area. If end-lines are recognized, we continue to detect penalty-lines. Penalty-lines paralleled with end-lines and are distant from the border of field area, so we use the condition below to recognize it. Here θ_{end} is the angle of the detected end-line. TH_θ and $TDis_{max}$ are two thresholds for the angle and the distance respectively.

$$|\theta - \theta_{end}| < TH_\theta \quad Dis > TDis_{max} \tag{5}$$

We show the recognition result of the type of field lines in Fig.3, where red, blue and yellow lines are mid-lines, end-lines and penalty-lines respectively.

4. PLAYER DETECTION AND PLAYER TYPE RECOGNITION

In this section we introduce our method of player detection in details. As illustrated in Fig.1(d), a binary mask of player candidates can be obtained from simple dominant field color segmentation. But because of the influence of field lines in soccer field and the failure of dominant color model, there are maybe still some blobs which are actually not players. In section 3, we have already detected field lines. So we can use field line mask illustrated in Fig.2(c) to eliminate line blobs



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

Fig. 3. Result of field line type recognition in typical wide angle view. Mid line, end-line, and penalty-line are in red, blue, and yellow respectively.

and then we dilate the remaining blobs. In the rest of the blobs, there may be one player, more than one player, or none of player. The last two cases are caused by noises and player overlapping.

In order to further determine the exact number and position of players, we use a template matching method. The method is based on the fact that the dresses of the players are designed to be visually distinctive so that viewers can see them clearly. Template images are manually selected from soccer video. Four templates are used for four types of players, which are normal player and goal keeper on both sides. Players in different position of playfield are often of different size, so the size of template should also vary within image. Here we used the method proposed in [6] to estimate the player size at image coordinate (i, j). The method is based on the pinhole camera model and calculates the size variation matrix for each position in the image. Then we scale the templates to the estimated size according to positions of candidate blobs. The scaled templates are moved across every pixel in the blob to get the best position with highest similarity. Normal square differences are used to measure the similarity between templates and candidate blobs, which is defined as follows:

$$R(x, y) = \frac{\sum_{x',y'}[T(x', y') - I(x + x', y + y')]^2}{\sqrt{\sum_{x',y'} T(x', y') \cdot \sum_{x',y'} I(x + x', y + y')}} \tag{6}$$

In equation(6), $R(x, y)$ is the normal square difference. $T(x, y)$ is the template image, and $I(x, y)$ is the player candidate image. (x, y) and (x', y') are coordinates in player candidate image and template image respectively. After template matching, we get four similarity measures for each blob in the image. If the biggest one is bigger than a threshold, we consider the blob is not actual player. To reduce com-

Table 1. experiment results of mid-line detection

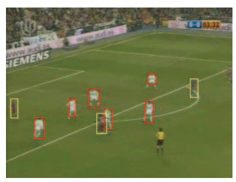
Game Type	Number of lines	Correct	False	Miss	Precision	Recall
Spain	63	63	3	0	95.4%	100%
Italy	57	57	7	0	89.1%	100%
World Cup	92	90	6	2	93.8%	97.8%

Table 2. experiment results of end-line detection

Game Type	Number of lines	Correct	False	Miss	Precision	Recall
Spain	121	87	9	34	93.1%	71.9%
Italy	105	83	4	22	95.4%	79.0%
World Cup	108	89	9	19	90.8%	82.4%

putational complexity, we further restrict that 80% area of the template should be in the blob. To deal with player overlapping, we calculate the intersection area between the best matching template and the candidate blob. If it is less than 40% of the total area of candidate blob, we consider that the blob has more than one player. So we continue matching templates on uncovered region. Fig.4 shows some results of player detection and player type recognition.

(a)



(b)

Fig. 4. Result of field line type recognition in typical wide angle view. Two types of player are shown in yellow and red rectangles

5. EXPERIMENTAL RESULT

We have tested our methods based on three soccer videos which are from Spanish Premier Division, Italy Serial A and FIFA 2006 world cup. All the testing data are TV broadcasted videos, but the production styles are quite different. We randomly select 200 frames from each video to evaluate our method of object detection. The ground truths for object detection are obtained by manual inspection. The results of line detection and recognition are shown as TABLE 1, 2, and 3. Note that we consider it is correct only if a line can be both correctly detected and recognized. For the case of player detection and player type recognition, the result is shown as TABLE 4.

From TABLE 1, 2 and 3, it is clear that our proposed method is useful for every line type. The recall and precision for mid-lines are the highest for each video because mid-lines are noticeable and almost not variant no matter in what kind of production style. From table 4, we can see that the precision of our player detection and recognition is high but recall is not so good because of the occlusion between players.

Table 3. experiment results of penalty-line detection

Game Type	Number of lines	Correct	False	Miss	Precision	Recall
Spain	124	87	9	37	90.6%	70.2%
Italy	117	79	8	38	90.8%	67.5%
World Cup	108	82	16	26	83.7%	75.9%

Table 4. experiment results of player detection and type recognition

Game Type	Number of players	Correct	False	Miss	Precision	Recall
Spain	1832	1597	89	235	94.7%	87.2%
Italy	1315	1202	102	113	92.2%	91.4%
World Cup	1657	1400	147	257	90.5%	84.5%

6. CONCLUSION

In this paper, we propose effective methods for field lines and players detection and recognition in soccer video. Morphological top-hat transform is performed on the wide angle image to detect field lines. Mid-lines, end-lines and penalty-lines are then recognized based on their positions. Template matching method is used to detect and classify players. Experiments show provide good performances of our methods.

7. REFERENCES

[1] Yijuan Lu et al., "What are the high-level concepts with small semantic gaps?," in *Proc. of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8.

[2] Yi Ding and Guoliang Fan, "Segmental hidden markov models for view-based sport video analysis," in *Proc. of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007, pp. 1–8.

[3] Mei Han et al., "Feature design in soccer video indexing," in *PCM*, 2003, pp. 950–954.

[4] Bo Yan et al., "Pyramidwise structuring for soccer highlight extraction," in *PCM*, 2003, pp. 945–949.

[5] M. Haseyama T. Wantanabe and H. Kitajima, "A soccer field tracking method with wire frame model from tv images," in *Proc. of the IEEE International Conference on Image Processing (ICIP)*, 2004, pp. 1633–1636.

[6] Xinguo Yu et al., "Trajectory-based ball detection and tracking in broadcast soccer video," *IEEE Trans. Multimedia*, vol. 8, no. 6, pp. 1164–1178, 2006.

[7] A. Murat Tekalp Ahmet Ekin and Rajiv Mehrotra, "Automatic soccer video analysis and summarization," *IEEE Trans. Image Processing*, vol. 12, no. 7, pp. 796–807, 2003.