A SOCCER FIELD TRACKING METHOD WITH WIRE FRAME MODEL FROM TV IMAGES

Tomoki Watanabe, Miki Haseyama, and Hideo Kitajima

School of Engineering, Hokkaido University
N-13 W-8, Kita-ku, Sapporo 060-8628, JAPAN
{watanabe, mikich, kitajima}@media.eng.hokudai.ac.jp

ABSTRACT

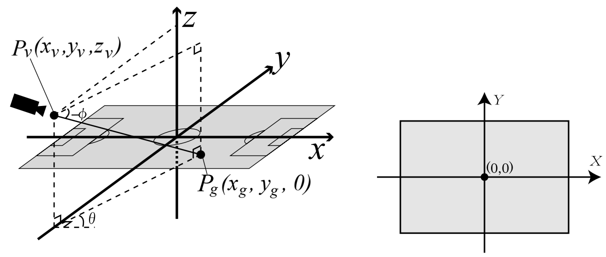
This paper proposes a tracking method of soccer field area in a soccer video captured from TV. The camera taking the soccer game video is controlled by three parameters: its mount position, the angle, and the magnification. In order to estimate these three parameters, the proposed method designs a wire frame model, which represents the official layout of the soccer field lines; and by the wire frame model matching with the field area in the video, the above three parameters can be estimated, even if the soccer video includes the camera panning, zooming, etc. By using the estimated parameters, we can accurately obtain where the field area corresponds in the actual soccer field. Some experiments in tracking the field area in actual soccer videos are performed and their results verify the high performance of the proposed method.

1. INTRODUCTION

The needs for analysis of field sports video are increasing. The main purpose is to understand the games or the tactics. Especially, in order to analyze the soccer games, several methods have been proposed [1]-[5]. In these methods, the soccer players can be tracked in images taken by a camera, such as TV images. However, for the analysis of tactics, players' positions not only in the image but also in the soccer field are necessary. In order to obtain the players' positions in the soccer field from these in the image, we must obtain where the field area in the images corresponds to in the soccer field.

Therefore, we propose a tracking method of the field area in a soccer video captured from TV. The camera taking the soccer game video is controlled by three parameters: its mount position, the angle, and the magnification. In order to estimate there parameters, the proposed method utilizes a wire frame model, which represents the official layout of the soccer field lines; and by the wire frame model matching with the field area in the video, the above three parameters can be estimated, even if the soccer video includes the camera panning, zooming, etc. By using the estimated parameters, we can accurately obtain where the field area corresponds in the soccer field.

This paper is organized as follows. In Section 2, the perspective transformation utilized in our method is summarized. A tracking method of a field area with a wire frame model is proposed in Section 3. Finally, Section 4 shows some experimental results to verify the effectiveness of the proposed method.



(a)World coordinates (b)Screen coordinates

Fig. 1. Coordinate systems: objects in the world coordinates are transformed to screen coordinates by the camera.

2. PERSPECTIVE TRANSFORMATION

In this section, the perspective transformation utilized in our method is summarized. This transformation is one of the projective transformations. In order to implement the perspective transformation, the world coordinates (x, y, z) and the screen coordinates (X, Y) are defined as shown in Fig.1. If the camera is set up at $P_v(x_v, y_v, z_v)$ and its lens is pointed at $Pg(x_g, y_g, 0)$, a transformation from the world coordinates (x, y, z) to the screen coordinates (X, Y) is given by the following equations.

$$X = -f\frac{x_1}{z_1}, \quad Y = -f\frac{y_1}{z_1}, \tag{1}$$

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} -\sin\theta & \cos\theta & 0 \\ -\sin\phi\cos\theta & -\sin\phi\sin\theta & \cos\phi \\ \cos\phi\cos\theta & \cos\phi\sin\theta & \sin\phi \end{bmatrix} \begin{bmatrix} x - x_v \\ y - y_v \\ z - z_v \end{bmatrix}, \tag{2}$$

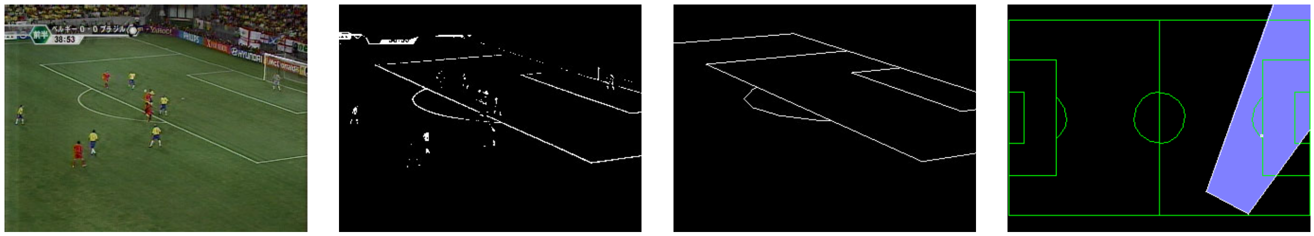
where x_v, y_v and z_v represent the coordinates of the camera position. The parameter f represents the magnification of the camera; θ and ϕ are defined as follows:

$$\theta = \tan^{-1}\left(\frac{y_v - y_g}{x_v - x_g}\right), \tag{3}$$

$$\phi = \tan^{-1}\left(\frac{z_v - z_g}{\sqrt{(x_v - x_g)^2 + (y_v - y_g)^2}}\right), \tag{4}$$

where the parameters x_g and y_g control the camera angle. This transformation projects the world coordinates to the screen coordinates. The parameters x_v, y_v, z_v, f, x_g, and y_g are called the camera parameters hereafter.

0-7803-8554-3/04/$20.00 ©2004 IEEE.



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

Fig. 2. Processes of our method: (a) A frame of a soccer video; (b) Extracted field lines from (a); (c) The wire frame model which transformed by using the optimal camera parameters; and (d) The field area obtained by the optimal camera parameters.

3. PROPOSED METHOD

The proposed method consists of the following three processes.

i) Process1 extracts the field lines from a soccer video.

ii) Process2 defines a wire frame model of the field lines which represents the official layout of the soccer field lines.

iii) Process3 tracks where the field area corresponds in the soccer field by utilizing the field lines and the wire frame model.

Each processes are explained in the following subsections.

3.1. Process1: Extraction of the field lines

The field lines are roughly extracted from a soccer video by using the colors of the field and the field line as follows:

1. Extract green colored pixels by using their hue range $\frac{2}{5}\pi < H < \pi$, which correspond to the field area candidates, where H represents the hue in the HSV color space. The values $\frac{2}{5}\pi$ and π are 5Y (yellow) and 5BG (blue green) in the Munsell color system, respectively.
2. Extract the largest field area candidate, which corresponds to the field area, because the field area occupies a large space in a soccer video.
3. Extract white colored pixels included in the extracted field area by using their intensity range $I > Th_I$, which correspond to the field lines, where I represents the intensity and Th_I is a predefined threshold.

A line-extraction result by the above procedures is shown in Fig. 2(b). Though the result includes not only the field lines but also other objects, such as the soccer ball, the players, etc; in our method explained later, the objects are not bad influences on the field-tracking accuracy.

3.2. Process2: Definition of the wire frame model

Our method designs a wire frame model based on the official layout of the soccer field lines. According to FIFA (Fédération Internationale de Football Association) regulations, the soccer field for the international matches should be 105×68 meters as shown in Fig. 3. Thus the model is defined as the same size as Fig. 3. By utilizing the wire frame model, the field lines taken by a camera from several angles, can be represented as shown in Fig. 4.

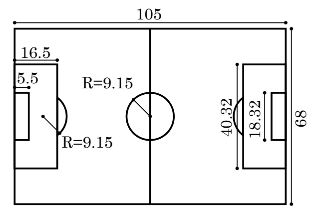


Fig. 3. The soccer field measurements regulated by FIFA: the size for the international matches should be 105×68 meters.

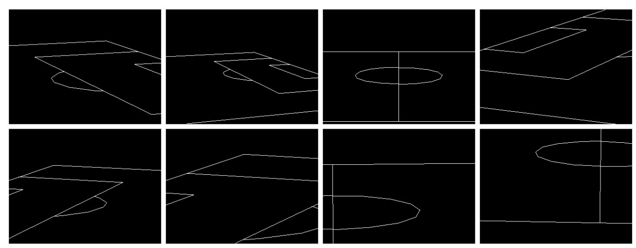


Fig. 4. Some examples of the field lines taken by a camera from several angles: the wire frame model can represent all field lines by applying the perspective transformation.

3.3. Process3: Tracking of the field area

The field area is tracked by utilizing the camera parameters. The optimal camera parameters are obtained by searching. The ranges of the possible camera parameters are as follows:

i) The camera position (x_v, y_v, z_v) is

$$x_v = 0 \text{ [m]}, \quad y_v \in [-50, -150] \text{ [m]}, \quad \text{and} \quad z_v \in [10, 30] \text{ [m]}; \tag{5}$$

because the camera taking the soccer game video is set up at the center of the auditorium.

ii) The camera angle $(x_g, y_g, 0)$ is

$$x_g \in [-52.5, 52.5] \text{ [m]} \quad \text{and} \quad y_g \in [-34, 34] \text{ [m]}; \tag{6}$$

because the camera takes only inside of the field.

iii) The camera magnification f is

$$f_{min} < f < f_{max}, \tag{7}$$

where f_{min} and f_{max} represent the lowest magnification and the highest magnification, which can be obtained by the distance between the camera position and the soccer field.

If only one frame of the soccer video is given, we have to search the above ranges for the optimal camera parameters. However, after obtaining the optimal parameters in the frame, we do not need to exhaustively search the ranges in the frames after it; consequently we can obtain the camera parameters as follows:

i)$'$ The parameters y_v and z_v are the same ones obtained in the previous frame. Therefore, we can obtain the optimal y_v and z_v without searching.

ii)$'$and iii)$'$ The parameters x_g, y_g, and f can be obtained by searching the following ranges.

$$|x_g - x_g^{frame_{t-1}}| < \epsilon_{x_g}, \quad (8)$$

$$|y_g - y_g^{frame_{t-1}}| < \epsilon_{y_g}, \quad (9)$$

$$|f - f^{frame_{t-1}}| < \epsilon_f, \quad (10)$$

where $x_g^{frame_{t-1}}$, $y_g^{frame_{t-1}}$, and $f^{frame_{t-1}}$ denote the optimal values of x_g, y_g, and z_g in the previous frame, respectively; and ϵ_{x_g}, ϵ_{y_g}, and ϵ_f are predefined thresholds; because the time interval between the current frame and the previous one is too short for the camera to be operated.

By using i)$'$, ii)$'$, and iii)$'$, we can reduce the computational costs for searching. The camera parameters are determined by matching the transformed wire frame model by the camera parameters in the ranges described above to the field lines in the soccer video. In order to evaluate how much the wire frame model fits to the field lines, we define a fitness criterion as follows: the fitness criterion is calculated as follows:

$$Fitness = \sum_{i=1}^{w_I} \sum_{j=1}^{h_I} f(i,j), \quad (11)$$

$$f(i,j) = \begin{cases} 1 & \text{if } I_{model}(i,j) > 0 \\ & \text{and } \displaystyle\sum_{m=-D}^{D} \sum_{n=-D}^{D} I_{line}(i+m, j+n) > 0 \\ 0 & \text{otherwise} \end{cases}, \quad (12)$$

where I_{line} is the binary image obtained by Process1; I_{model} is also the binary image obtained by Process2; w_I and h_I are the width of I_{line} and the height of I_{line}, respectively; and D is a predefined parameter. In order to obtain the optimal camera parameters, which present the highest fitness value, we compute the fitness values by using the all possible camera parameters y_v, z_v, f, x_g, and y_g in the ranges explained before with the quantization steps δ_{y_v},δ_{z_v}, δ_f,δ_{x_g}, and δ_{y_g}, respectively. By this searching, we can obtain the optimal camera parameters. A wire frame model, which transformed by the optimal camera parameters, is shown in Fig. 2(c).

By using the optimal camera parameters, we can obtain the correspondence between the field area on the official layout soccer field and the field area in the soccer video. In order to avoid readers confusion, we call the former field area the actual field area and the latter field area the filmed field area hereafter. From Eqs. (1) and (2) with $z = 0$, the relationship between the pixel at (x, y) in the actual field area and the pixel at (X, Y) in the filmed field area is described below.

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -\sin\theta & \cos\theta & 0 \\ \sin\phi\cos\theta & -\sin\phi\sin\theta & \cos\phi \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} + \begin{bmatrix} x_v \\ y_v \end{bmatrix}, \quad (13)$$

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} \cos\phi\cos\theta & -\cos\phi\sin\theta & -\sin\phi \\ f & 0 & X \\ 0 & f & Y \end{bmatrix}^{-1} \begin{bmatrix} -z_v \\ 0 \\ 0 \end{bmatrix}. \quad (14)$$

By executing the processes described above in each frame, we can track the actual field area from the soccer video, and an example is shown in Fig. 2(d).

4. EXPERIMENTAL RESULTS

The effectiveness of the proposed method is confirmed by some experiments. For the experiments, the soccer videos are captured at 10 fps from TV. The size of the captured images are 320 × 240 pixels. In the experiments, Th_I=0.65, D=3, f_{min}=1500, f_{max}= 3000, δ_{y_v}=1[m], δ_{z_v}=1[m], δ_{x_g}=0.25[m], δ_{y_g}=0.5[m], δ_f=10, ϵ_{x_g}=5[m], ϵ_{y_g}=5[m], and ϵ_f=100.

The proposed method is applied to the soccer video of 10 minute length. A part of the result is shown in Fig. 5. As shown in Fig. 5 (i) and (ii), the wire frame model is successfully matched to the lines in the soccer video by searching, and thus the camera parameters are accurately estimated by the proposed method. As shown in Fig. 5 (iii), we can verify that the field areas are correctly tracked by the proposed method, even if the camera are (a)panning or (b)zooming.

5. CONCLUSIONS

This paper has proposed a tracking method of the field area in a soccer video captured from TV. In the proposed method, the field area is tracked by matching the wire frame model to the field lines in the soccer video. The effectiveness of the proposed method has been confirmed by the experiments with the soccer videos captured from TV, which include the camera panning and zooming.

6. REFERENCES

[1] T. Taki, J. Hasegawa, and T. Fukumura, “Development of motion analysis system for quantitative evaluation of teamwork in soccer games,” Proc. Int. Conf. on Image Processing, vol. 3, pp. 16-19, Sep. 1996.

[2] Y. Ohno, J. Miura, and Y. Shirai, “Tracking players and estimation of the 3D position of a ball in soccer games,” Proc. 15th Int. Conf. on Pattern Recognition, vol. 1, pp. 145-148, Sep. 2000.

[3] O. Utsumi, K. Miura, I. Ide, S. Sakai, and H. Tanaka, “An object detection method for describing soccer games from video,” Proc. Int. Conf. on Multimedia and Expo, vol. 1, pp. 26-29, Aug. 2002.

[4] S. Leféyre, J. Gérand, A. Piron, and N. Vincent, “An extended snake model for real-time multiple object tracking,” Int. Workshop on Advanced Concepts for Intelligent Vision Systems, pp. 268-275, Sep. 2002.

[5] H. Ok, Y. Seo, and K. Hong, “Multiple soccer players tracking by condensation with occlusion alarm probability,” Int. Workshop on Statistical Methods for Vision Processing, pp. 1-6, 2002.

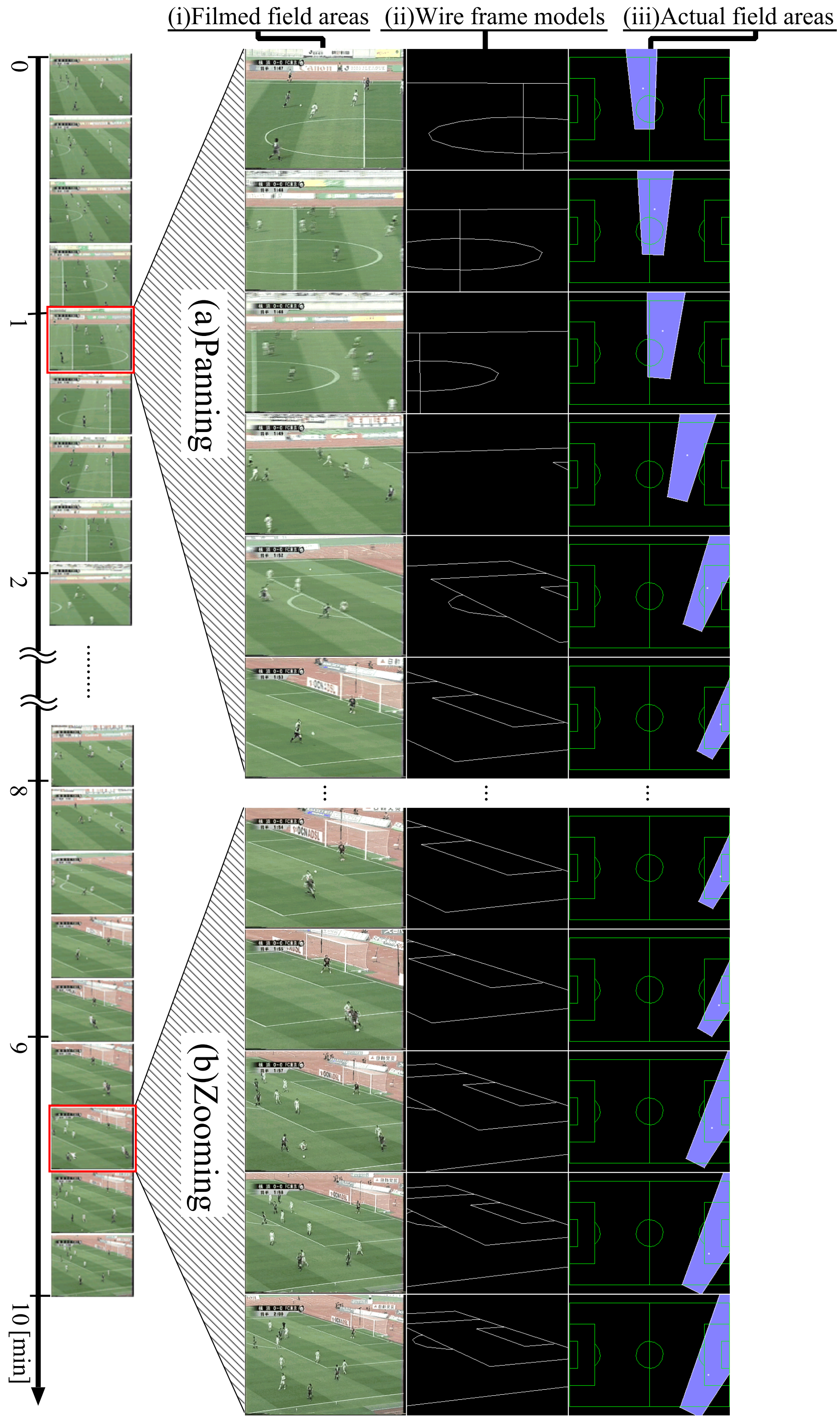


Fig. 5. Results of the field area tracking: (a) A scene taken by the panning camera; (b) A scene taken by the zooming camera; (i) The filmed field areas; (ii) The transformed wire frame models; and (iii) The actual field areas obtained by the proposed method.