**Proc. 2002 Int. Conf. on Pattern Recognition, pp. 303-306, Aug. 2002.**

# Tracking Players and a Ball in Video Image Sequence and Estimating Camera Parameters for 3D Interpretation of Soccer Games

Akihito Yamada, Yoshiaki Shirai and Jun Miura
Dept. of Computer-Controlled Mechanical Systems, Osaka University
2-1, Yamadaoka, Suita, Osaka 565-0871, JAPAN
E-mail:{yamada,shirai,jun}@cv.mech.eng.osaka-u.ac.jp

## Abstract

*To recognize and retrieve soccer game scenes, the movement of the players and the ball must be analyzed. This paper describes a method of tracking the players and the ball and estimating their 3D positions from video images recorded from TV broadcasting. Our system detects the players and tracks them by predicting their motions. Our system determines the ball from among ball candidates by considering the motion continuity of the ball. The camera parameters (pan, tilt, and zoom) are estimated by detecting line regions on the field in every frame and by matching them to the model. Because our method uses not only straight lines but also circular ones for the matching, it can be applied to the whole area of the field. Our system also estimates the 3D position of the ball by fitting a 3D physical model to the observed ball trajectory. The effectiveness of the method is shown by experiments with actual image sequences.*

## 1 Introduction

Recently, the need for efficient video retrieval systems from large databases is increasing because the amount of available image information has been increasing rapidly. For example, in TV programs, viewers may wish to retrieve certain contents. In soccer games, viewers may want to watch only exciting scenes in a long game. To retrieve such scenes automatically, it is essential to know the movement of the players and the ball and their positional relationship. Most previous methods use template matching [1, 2], or snake [3] for tracking. In these methods, however, the human operator often has to modify the position of players manually during occlusion. Ohno et al. [4] detect occlusion between players and distinguish them by their color and position after occlusion. They also estimate the position of the players and the ball; the camera parameters are calibrated by manually selecting feature points. This calibration has to be done only once because the camera parameters are all fixed. In our case, however, since we deal with video images taken by a rotating and zooming camera, we must estimate the camera parameters (pan, tilt, and zoom) in each frame automatically to know the position of the players and the ball. D. Yow et al. [2] proposed a method of estimating camera parameters in each frame. However, they estimate only the pan angle and the zoom. Kim and Hong [5] use line features on the field to estimate the camera parameters in each frame. However, the method can be used only for the scenes which include many straight lines such as the goal areas.

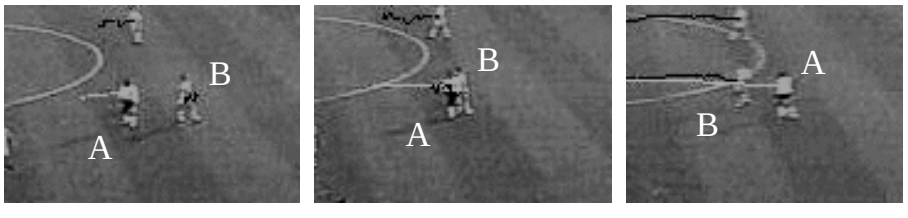
Our system can estimate the camera parameters (pan, tilt, and zoom) in a wide area by matching not only the straight lines but also the circular ones to the model.

## 2 Finding and tracking the players

We use color information and positional relationship between shirts and pants to extract the player candidate regions [4]. We then extract the shirts and pants region whose sizes are within a certain range in the player candidate regions. Since the size of the player changes according to its position and the zooming of the camera, we change the range using the estimated focal length and the player's position on the field. We search for pairs of the shirts and the pants region which align vertically in the image by assuming that the players stand up. Fig. 1 shows an original image and the result of the player extraction.

(a) Original image (b) Result of extraction

**Figure 1. Extraction of players**




(a) 0th frame (b) 25th frame (c) 50th frame

**Figure 2. Tracking overlapping players**

In tracking a player, we predict the position of the player on the field by a simple linear extrapolation from the previous two positions because the player moves at almost constant speed on the field during a short period of time. We project the predicted position to the image by using the estimated camera parameter and search a neighboring area of the projected position for the uniform regions.

If two players are overlapping, we detect the overlap and track both players as one region until the players are separated. We distinguish them by color and positional relationship in the image. Fig. 2 shows the result of tracking overlapping players. In this figure, Fig. 2(a) is the beginning of the overlapping of player A and B, B is occluded by A in Fig. 2(b), Fig. 2(c) is the end of occlusion.

## 3 Estimation of camera parameters

We must calibrate the camera in each frame to know the position of the players and the ball because video images from TV broadcasting are taken by a rotating and zooming camera. Because the position of the broadcasting camera is usually fixed with respect to the world coordinates, we have only to estimate the pan and the tilt angles and the focal length. These three parameters are called the camera parameters here after.

We estimate the camera parameters by extracting lines including the straight lines and the circular ones (the center circle and the penalty arcs) and matching them to the model which is made by following the official rule. Fig. 3 shows the model. To extract line regions, we extract white regions on the field excluding the player regions. Fig. 4 shows an example of extracted line regions.

We project the line regions to the model by transforming the image coordinates $(x, y)$ to the world coordinates $(X, Y, Z)$ via the camera coordinates $(X_c, Y_c, Z_c)$. Fig. 5 shows the positional relationships of the coordinate systems.

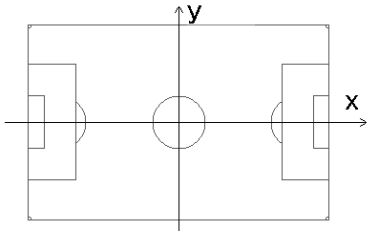
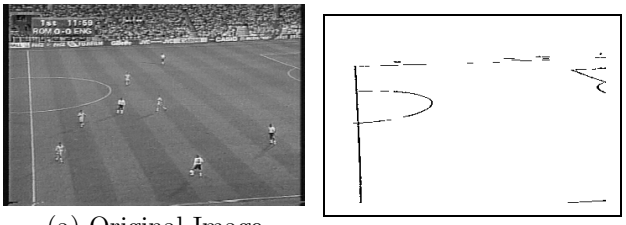



**Figure 3. Model**



(a) Original Image (b) Extracted line regions

**Figure 4. Result of extracting line region**

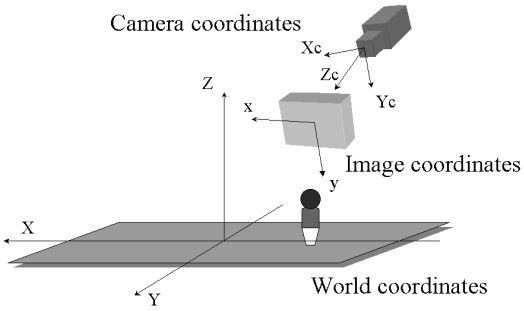



**Figure 5. Relation of the coordinate systems**

The rotation matrix $\mathbf{R}$ in the transformation from the world coordinates to the camera coordinates is defined by the following equation:

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\beta & sin\beta \\ 0 & -sin\beta & cos\beta \end{bmatrix} \begin{bmatrix} cos\alpha & 0 & -sin\alpha \\ 0 & 1 & 0 \\ sin\alpha & 0 & cos\alpha \end{bmatrix} \mathbf{R}_{wc}, \quad (1)$$

where $\alpha$ and $\beta$ denote the pan angle (rotation about $Y_c$-axis) and the tilt angle (rotation about $X_c$-axis) respectively. $\mathbf{R}_{wc}$ denotes the rotation matrix from the world coordinates to the camera coordinates with the panning and tilting rotation being equal to zero. Using $\mathbf{R}$, the transformation from the world coordinates to the camera coordinates is given by

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \mathbf{R} \begin{bmatrix} X - X_{cp} \\ Y - Y_{cp} \\ Z - Z_{cp} \end{bmatrix}, \quad (2)$$

where $(X_{cp}, Y_{cp}, Z_{cp})$ denotes the position of the camera in the world coordinates. Note that $Z = 0$ for the

line region on the ground. We also have the following relationships between the image coordinates and the camera coordinates from a perspective projection model:

$$\begin{cases} X_c &= Z_c \frac{x}{f}, \\ Y_c &= Z_c \frac{y}{f}, \end{cases} \tag{3}$$

where $f$ denotes the focal length. If $\mathbf{R}$ is known, we can obtain three equations with three unknowns $(X, Y, Z_c)$ by substituting $(X_c, Y_c)$ in eq. (3) for those in eq. (2). By solving the equations, we can calculate the position $(X, Y)$ of a point on the ground from its position $(x, y)$ in the image.

To estimate $\mathbf{R}$, we match extracted line regions to the model. If a pixel in the extracted line regions exists a neighboring area of a model line, the pixel is regarded as a part of the model line. We assume that the matching point of the pixel is the nearest point on the model line. Fig. 6 shows an example of pairs of matching points. We search for the camera parameters $(\alpha, \beta, f)$ which maximize the number of pairs of the matching points. If the numbers which are obtained by other camera parameters are the same, we choose the camera parameters which minimize the sum of the squared distance between each pair of matching points in the original image.

In the initial frame, we search a large parameter space for the optimal parameters. In the second frame, we search around the initial parameter. In subsequent frame, we first predict the parameters by using a simple linear extrapolation of the parameters in the two previous frames because the motion of the broadcasting cameras are usually smooth. We then search a neighbor of the predicted parameters for the optimal camera parameters. Fig. 7 shows the projected image to the model by the estimated camera parameters. In this figure, the black points are detected pixels as the line regions in the image. Fig. 8 shows a mosaicing result of a video image sequence (267 consecutive images). This figure shows that we can make a wide mosaic image by matching not only straight lines but also circular ones to the model.

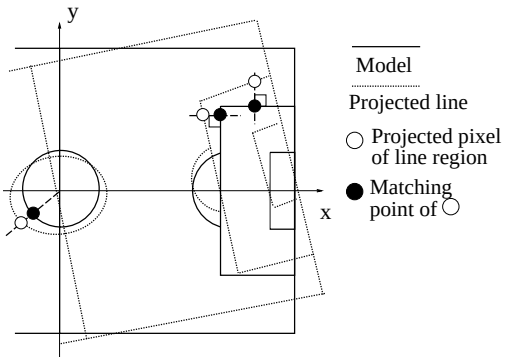



**Figure 6. Examples of pairs of matching points**

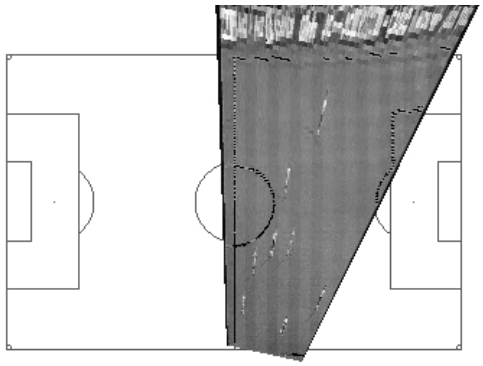


**Figure 7. Result of projecting an original image to the model**

## 4 Finding and tracking a ball

It is difficult to find the ball only by using color because the ball is small and moving fast in the image. We extract white regions excluding the player regions and the line regions, and we regard them as ball candidates. In subsequent frames, we track each candidate by searching a neighboring area of the predicted position for the ball candidates. If the candidates are found, only the one which is the nearest to the predicted position is retained. If they are not found, tracking of the candidate is terminated. These steps are repeated until all but one candidates are deleted; the remaining candidate is determined as the ball region.

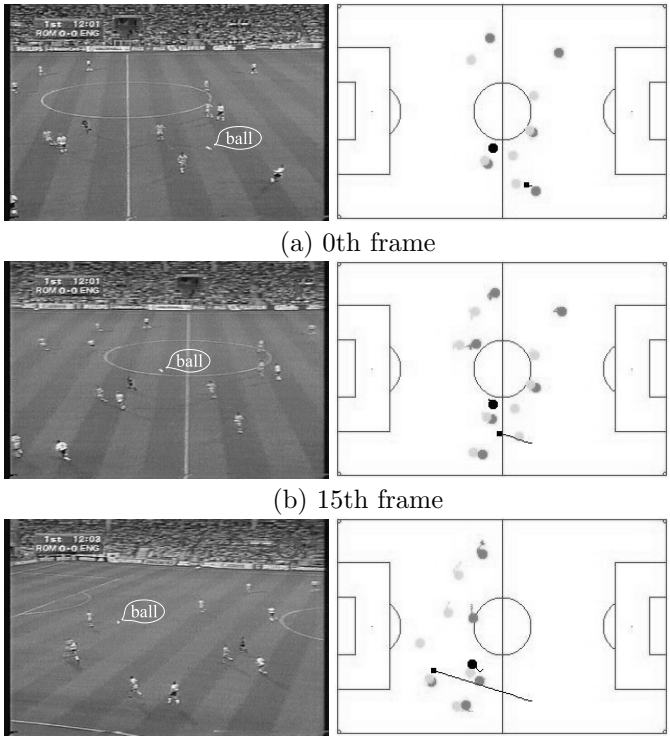
## 5 Estimation of the position of the players and the ball

To know the position of the players on the field, we detect the footing points of the players and project them to the model by using the camera parameter. To estimate the ball position in the 3D space, we estimate the 3D trajectory of the ball by assuming that the ball motion is determined by the gravity and the air friction. The motion is expressed as follow:

$$\begin{aligned} X(t) &= X(0) + tv_X(0) - a\int_0^t t\, v_X(t)\, dt, \\ Y(t) &= Y(0) + tv_Y(0) - a\int_0^t t\, v_Y(t)\, dt, \\ Z(t) &= Z(0) + tv_Z(0) - \int_0^t t\,\{g + av_Z(t)\}\, dt \end{aligned} \tag{4}$$

where $(X(t), X(t), Z(t))$ and $(v_X(t), v_Y(t), v_Z(t))$ denote the position and the velocity of the ball at time $t$ respectively, $g$ denotes the acceleration of gravity, and $a$ denotes the friction coefficient. We estimate the $(X(0), Y(0), Z(0))$ and $(v_X(0), v_Y(0), v_Z(0))$ by matching the observed position of the ball to the physical model in the world coordinates.

**Figure 8. Mosaicing image**




(a) 0th frame

(b) 15th frame

(c) 56th frame

**Figure 9. Estimation of the 3D position of the ball and the players**

Fig. 9 shows the result of estimating the 3D position of the players and the ball. In this figure, a big black dot, a small black dot, and light and dark gray dots denote the referee, the ball, and the two different team's players, respectively. The right side player in Fig. 9(a) kicks the ball to the left side player in Fig. 9(c). We can estimate the 3D position of the ball and the players in the scenes which do not include many straight lines such as Fig. 9(b). When the number of data of the observed ball position in the image is small, the estimated ball trajectory may not be reliable from our experience. To reliably estimate the trajectory of the ball in 3D space, we need at least about 30 frame (1 second).

## 6 Conclusion

In this paper, we have described a system to track the players and the ball and to estimate their 3D position for 3D interpretation in TV broadcasting video images. Our system estimates the camera parameters by fitting the line image to the model at each frame. It can work in the wide area by using not only straight lines but also circular ones. It can also determine the 3D position of the ball by using the camera parameters and a physical model in the 3D space.

This idea can be applied to other sports such as American football and basketball.

Future works include an improvement of the color model to adapt to various illumination conditions.

## References

[1] Y. Gong, C. Hock-Chuan, L.T. Sin "An Automatic Video Parser for TV Soccer Games ", Proc. of ACCV, pp. 509-512, 1995.

[2] D. Yow, B.L. Yeo, M. Yeung, B. Liu "Analysis and Presentation of Soccer Highlights from Digital Video ", Proc. of ACCV, pp. 499-502, 1995.

[3] S. Lefèvre, C. Fluck, B. Maillard, N. Vincent "A Fast Snake-based Method to Track Football Players ", Proc. of MVA, pp. 501-504, 2000.

[4] Y. Ohno, J. Miura, Y. Shirai. "Tracking Players and Estimation of the 3D Position of a Ball in Soccer Games", Proc. of ICPR, pp. 145-148, 2000.

[5] H. Kim, K.S. Hong. "Soccer Video Mosaicing using Self-Calibration and Line Tracking", Proc. of ICPR, pp. 592-596, 2000.