Parallel Tracking of All Soccer Players
by Integrating Detected Positions in Multiple View Images

Sachiko Iwase   Hideo Saito
Department of Information and Computer Science
Keio University
3-14-1 Hiyoshi, Kohoku-ku, Yokohama, 223-8522, Japan
{sachiko,saito}@ozawa.ics.keio.ac.jp

Abstract

Soccer, one of the popular sports around the world, is often broadcasted on TV, and various researches have done on soccer scene images such as strategy analysis, scene recovery, automatic indexing of soccer scenes, and automatic intelligent sports casting. As robust player tracking is fundamental to those researches, there is a demand for an automatic player tracking system using soccer imaging data. In this paper, we propose a method of tracking soccer players using multiple views. Tracking is done by integrating the tracking data from all cameras, using the geometrical relationship between cameras called homography. Integrating information from all cameras enables stable tracking on the scene, where the tracking by a single camera often fails in the case of occlusion.

1. Introduction

Soccer is one of the sports that draw a lot of interest around the world, and sports programs targeting soccer are often broadcasted on television. Various researches have done on soccer scene analysis such as strategy understanding, intelligent robot camera in TV programming production[4], and making 3D images which help to understand the situation and experience exciting scenes of the soccer game.

In soccer scene analysis by image processing, it is desirable to get accurate positions of the players that changes on the ground with time. In most of the researches of player tracking[1, 2, 5, 6, 7, 8], some observing information such as texture, color, and shape are used for tracking, and there are researches that estimate motion of the players by Kalman filter. Color based template matching is one of the common methods to deal with the occlusion problem.

In many of those researches, tracking is done by using a single camera[1, 5, 6, 8]. Multiple cameras will be effective for tracking in crowded scene [2, 3]. Mutiple cameras can also be used to show the player positions in 3D space and player information from other cameras are not used for tracking[7].

In this paper, we will propose a method for parallel tracking of all soccer players on the ground using multiple view images taken by fixed cameras. Positions of players are detected in each camera, and then the detected positions are projected onto the virtual top-view image via homography of the ground between the camera image and the top-view image. By integrating the detected positions in all cameras, all players can robustly be tracked.

2. System Environment

Cameras are located around the soccer ground to cover all the area, and are lined up relatively equal, set at both sides of the ground. Figure 1 is an example of the system environment. Theoretically it is possible to track players if the tracking area is taken by any camera, however system environment like Figure 1 is recommended.

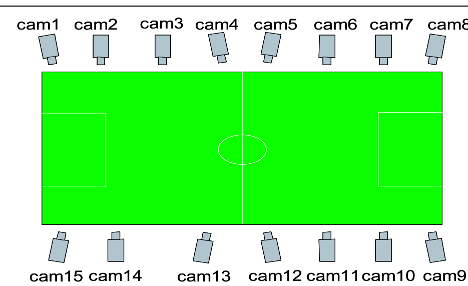


Figure 1. System environment

0-7695-2128-2/04 $20.00 (C) 2004 IEEE

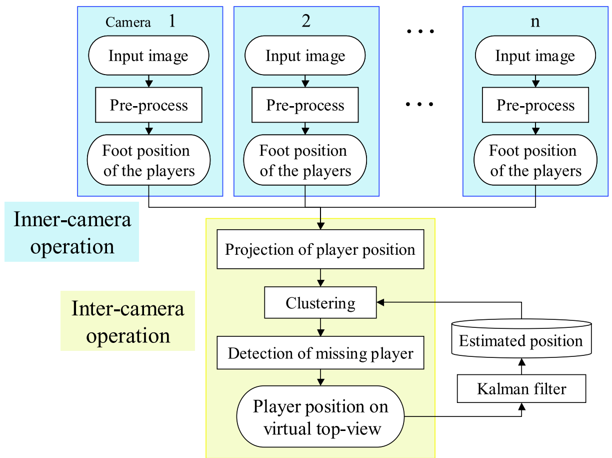


Figure 2. Flow of tracking

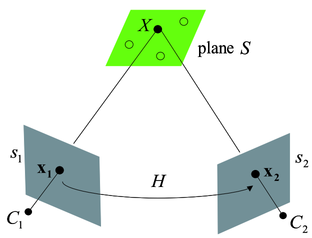


Figure 3. Homography

3. Tracking

Figure 2 shows the flow of the player tracking. First, inner-camera operation is performed independently in each camera to detect the features of the extracted player regions in each camera image. Then, those features of all the cameras are integrated in inter-camera operation for obtaining trajectory of players.

3.1. Inner-camera operation

First, background subtraction is done on an input image to extract player regions, and binary image is made by using thresholds of intensity and area with noise removal and smoothing. Then each region of the player is labeled and extracted. For each extracted player-region, its center, (x, y) coordinate of which the y-value is the largest, which is equal to the foot coordinate, length, width, and area are computed as player features.

3.2. Inter-camera operation

Tracking is done in inter-camera operation by integrating the foot position of the extracted players in inner-camera operation. Player positions on the virtual ground image from the upper-view are computed every frame.

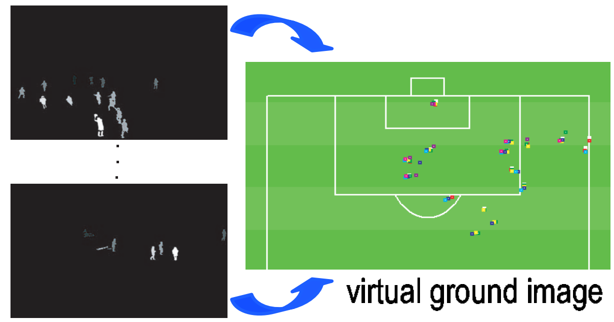


Figure 4. Virtual ground image.

3.2.1. Homography Looking at the point $\mathbf{X}$ on the plane $\mathbf{S}$ in 3D space from 2 cameras $\mathbf{C_1}$ and $\mathbf{C_2}$ like Figure 3, relationship between the point $\mathbf{x_1}$ and the point $\mathbf{x_2}$ on the camera images s_1 and s_2 is shown as 1. 3×3 matrix $\mathbf{H}$ is called Homography, which means the planar projective transformation, and is unique for each plane in 3D space. Homography is computed using more than 4 corresponding pairs of points between 2 camera images, and gives one-to-one correspondence of points.

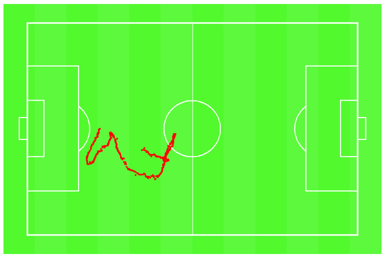
$$\begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix} \simeq \begin{pmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} \qquad (1)$$

3.2.2. Projection of player position Foot positions of all the players detected in inner-camera operation are projected on the virtual ground image from the upper-view using the relationship of homography as soccer ground is a plane. Figure 4 shows an example of the projected points on the virtual ground image. The projected positions on the virtual ground image are represented as $\mathbf{p_p}$. The number of $\mathbf{p_p}$ is equal to total number of detected players in all cameras.

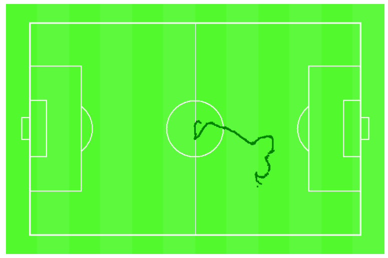
Homography matrices between the virtual ground image and the camera images are computed beforehand, using 10 to 20 corresponding points such as corners of the penalty area and foot positions of the players.

3.2.3. Clustering Since the detected position by the inner-camera operation includes some errors, we correct the player trajectory by integrating the detected position in all cameras by clustering the projected positions $\mathbf{p_p}$ on the virtual ground image.

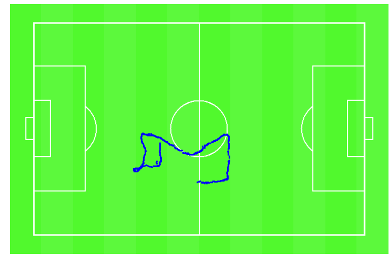
In our method, clustering means to sort the projected player positions $\mathbf{p_p}$ into the player positions estimated from the trajectory in the previous frames by Kalman filtering, which are represented as $\mathbf{p_e}$ on the virtual ground image.



(a) player ID no.8



(b) player ID no.19

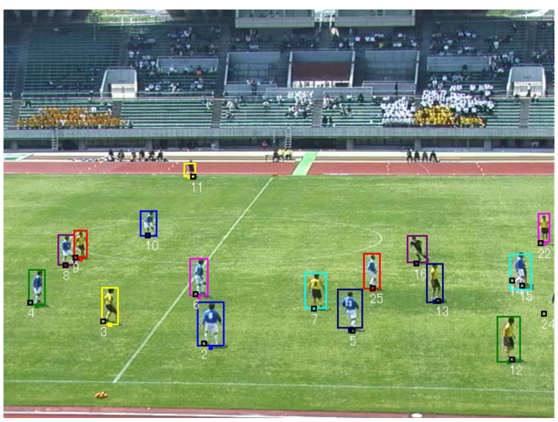


(c) player ID no.25

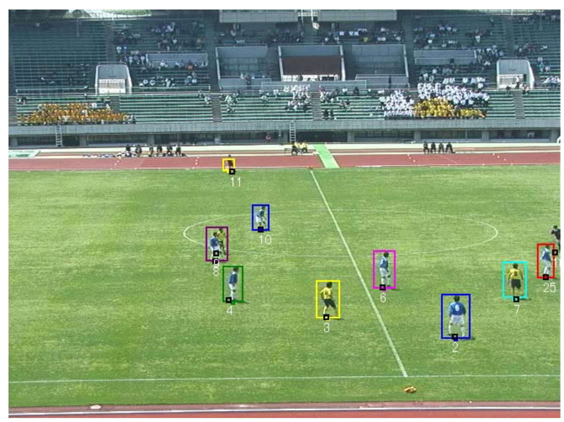
Figure 5. Trajectory of the players

(a) camera 3

(b) camera 4



(c) camera 12



(d) camera 13

Figure 6. tracking in each camera image

This clustering is equal to find which player the projected point belongs to. The number of the estimated positions $\mathbf{p_e}$ depends on that of players started to track in the initial frame. The player positions in the initial frame are given manually on the virtual ground image. The following is how the clustering works in each frame.

1 For each projected position $\mathbf{p_p}$, find the closest estimated position $\mathbf{p_e}$ with the following two conditions, and then make the projected point $\mathbf{p_p}$ belong to the closest estimated position $\mathbf{p_e}$. The first condition is that only one projected point $\mathbf{p_p}$ from one camera belongs to one estimated position $\mathbf{p_e}$ of the player. The second one is that if the distance is above the threshold, the projected point $\mathbf{p_p}$ does not belong to any estimated position $\mathbf{p_e}$.

2 For each estimated position $\mathbf{p_e}$, calculate the centroid of the projected points $\mathbf{p_p}$ which belong to the estimated position $\mathbf{p_e}$.

3 Update the estimated positions $\mathbf{p_e}$ with the centroids, and repeat **1**,**2** as the mean strain converges.

Finally, the centroids of the projected points are the player positions in the current frame.

3.2.4. Detection of missing player When the scene is too crowded with the players, foot positions of some players are not detected in any of the camera images, and are not projected to a virtual ground image. Then no projected point would belong to the estimated position of such player, so the estimated position is not updated. However the occlusion passes off as the player moves, and the projected point of the player appears on the virtual ground image which enables to track the player again.

4. Experimental results

Inputs are the image sequences of the soccer game in the multiple-view points. They are digitized of 720×480 pixels, RGB 24 bit and 15 frame/sec. Experiment is done in a scene of 500 frames, taken by 15 cameras like in Figure 1, and 22 players and 3 referees are tracked.

14 out of 25 are tracked perfectly through the scene of 500 frames. Each player has its own ID number given in the initial frame. The ID number is sometimes replaced by the other ID number when the scene is too crowded with the players. The replacement has occurred about 10 times for the other 11 players, however the trajectory of the players in the scene can be obtained by correcting the ID numbers manually.

As shown in Figure 5, players are tracked robustly for 500 frames. Figure 6 shows the player tracking in 4 camera images in a certain frame. Looking at the player IDs 4, 8, and 9, occlusion has occurred in camera 3 and 4. However as these players stay apart in camera 12, it is possible to detect the foot positions in inner-camera operation. In camera 13, foot position of the player of ID no.4 is detected. Therefore although the occlusion occurs in some cameras, tracking is stably done on the virtual ground image by using the player information from the cameras in which the occlusion is not occurred. Much the same is true for the players of ID no.2 and 6 in Figure 5.

We also compare the performance of the tracking of the proposed method with the method presented in [2], in which the trajectory of the players are corrected by checking the consistency of the detected position between two cameras, rather than clustering the detected positions in all cameras on the virtual ground image. In this comparison, we try to track 16 players from 8 camera images of 100 frames. While the proposed method in this paper perfectly tracks all 16 players in all 100 frames, the method presented in [2] fails tracking of 5 players. This suggests that the clustering on the virtual ground image is effective to robust tracking of players.

5. Conclusion

Experimental results show that the robust player tracking is done, and the trajectory of all the players on the virtual ground image is available. The results show the usefulness of a new method for player tracking, which uses homography that represents the geometrical relationship between cameras calculated from only a weak calibration, and integrates the player information of multiple cameras on the virtual ground image.

By having the trajectory of all the players in the soccer scene, it could be used in some applications such as strategy analysis, scene recovery, and control of the camera when making a TV program and so on. As the need to track players automatically and to get their accurate locations is high in soccer scene analysis, it is considered that application of this method is various.

References

[1] S. Choi, Y. Seo, H. Kim, K.S.Hong, “Where Are the Ball and Players? : Soccer Gamer Analysis with Color-based Tracking and Image Mosaick,” Proc. International Conference on Image Analysis and Processing, pp.196-203,1997.

[2] S.Iwase, H.Saito, “Tracking Soccer Players based on Homography among Multiple Views,” Proceedings of SPIE, Vol.5150 (VCIP2003), pp. 283-292, 2003.

[3] J. Kang, I. Cohen, G.Medioni, “Tracking People in Crowded Scenes Across Multiple Cameras,” Proc.ACCV04, pp.390-395, 2004.

[4] D. Kato, K. Abe, A. Ishikawa, M. Yamada, S. Kuwashima, “Camera Control System Detecting 3D Positions of Objects: Study of the Intelligent Robot Camera,” Proceedings of SPIE, Vol.3018, 1997.

[5] T. Misu, M. Naemura, W. Zheng, Y. Izurni, K. Fukui, “Robust Tracking of Soccer Players based on Data Fusion,” Proc. 15th International Conference on Pattern Recognition, Vol.I, pp.556-561, 2002.

[6] C. J. Needham, R. D. Boyle, “Tracking Multiple Sports Players Through Occlusion, Congestion and Scale,” Proc. British Machine Vision Conference 2001, Vol.1, pp.93-102, 2001.

[7] Y. Ohno, J. Miura, Y. Shirai, “Tracking Players and Estimation of the 3D Position of a Ball in Soccer Games,” Proc. 15th International Conference on Pattern Recognition, Vol.I, pp.303-306, 2002.

[8] O. Utsumi, K. Miura, I. Ide, S. Sakai, H.Tanaka, “An Object Detection Method for Describing Soccer Games from Video,” Proc. IEEE International Conference on Multimedia and Expo.,Vol.1, pp.45-48, 2002.