Robust Trajectory Estimation of Soccer Players by Using Two Cameras

Nozomu Kasuya, Itaru Kitahara, Yoshinari Kameda, Yuichi Ohta
University of Tsukuba, Graduate School of Systems and Information Engineering,
Tennodai 1-1-1, Tsukuba, Ibaraki, 305-8573, Japan
kasuya@image.eys.tsukuba.ac.jp
{kitahara, kameda, ohta}@ iit.tsukuba.ac.jp

Abstract

This paper proposes a method to estimate the trajectories of soccer players by using two cameras set in a large-scale outdoor space such as a soccer stadium, which is normally not advantageous for image processing. We estimate a player's position as the intersection of the primary axes of two body regions, corresponding to the two cameras, and a shadow region on the surface of the field. By projecting the foreground silhouette regions extracted from both cameras' images onto the soccer field, each player's body region and its shadow region are identified. Furthermore, our method utilizes color information of the players' uniforms to improve the accuracy of object tracking. We have applied our proposed method to a real soccer game held in a soccer stadium and demonstrated its effectiveness.

1. Introduction

3D video techniques that generate a video stream from arbitrary viewpoints form an active research field [1][2]. Sports broadcasting is expected to be an appropriate application for 3D video, since every audience member has his/her favorite way to watch a sports scene [3]. In ordinary 3D video systems, however, it is difficult to both choose desirable viewpoints and concentrate on the game simultaneously because you have to manipulate the pose of a virtual camera while watching the game.

Our approach to resolving this issue is to generate/display the video from a player's-eye view by using a 3D video technique. Observers can perceive more highly realistic sensations than possible with conventional 3D videos by watching sports through the exact view used by a soccer player. In order to generate the player's-eye view, it is necessary to estimate the time-series positional information, i.e. trajectory, of the target soccer player.

When we estimate the player's trajectory in a real soccer stadium, it is necessary to solve the following problems. (1) The region size in the captured image of each soccer player is small, due to the limited resolution of cameras. (2) Since soccer is a contact sport, many occlusions and interactions are observed in the images. (3) Sometimes, the lighting environment drastically changes due to weather conditions during a game. (4) Moreover, in order to provide 3D video broadcasting services, it is necessary to estimate the trajectory with short latency. This paper proposes a method to robustly estimate players' trajectories under such difficult conditions.

2. Trajectory Estimation in a Soccer Scene

Trajectory estimation methods for moving objects have been extensively researched. Methods based on template matching are widely used. However, because of their computational cost, they have difficulty with tracking many objects in real time. In addition, it is difficult to overcome the occlusion problem, since they depend on the appearance of objects [4]. A few methods that can estimate trajectories with short latency and that can overcome the occlusion problem have been developed [5][6]. However, they are not feasible for our video capturing condition because they require higher resolution of target appearance and cannot be applied to outdoor scenes, which are affected by shadow regions caused by sunshine. Chen et al. [1] proposed a method to solve the shadow problem by using HSV color space. However, this method also requires high-resolution images for target objects.

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

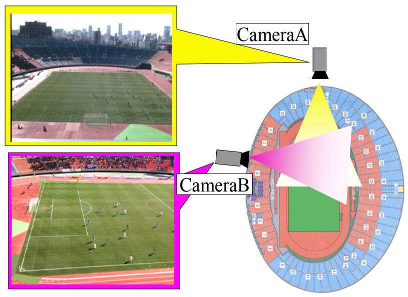

Figure 1. Layout of two cameras in a soccer stadium

We propose a robust method to estimate players' trajectories in a soccer scene of a real environment. We capture the soccer scene by using two fixed cameras to reduce the effects of occlusion and shadow problems. Figure 1 illustrates the layout of the two cameras in a soccer stadium. Our method reduces false tracking in "close play" by classifying the player regions based on the color of their uniforms.

3. Detection of Players

3.1. Candidate foreground regions

Since we capture the target scene by using fixed cameras, it is possible to extract candidate foreground regions by background subtraction. The size of player regions in captured images can be predicted by using the player's 3D location and size and known camera projection parameters. Regions caused by noise and the soccer ball, which are smaller than the predicted size, are removed. The remaining regions are treated as candidate foreground regions.

3.2. Estimation of a player's position

The candidate foreground regions include the body regions of soccer players and their shadow regions. We calculate the player's position by projecting the candidate foreground regions onto the field plane with 2D projective transformation.

As shown in Figure 2, a line is fitted to each projected region by calculating the orientation of the primary axis and gravity point (x_{0i}, y_{0i}):

$$a_i x_{0i} + b_i y_{0i} + c_i = 0 .$$

We estimate the intersection point (x, y) of the three lines as the player's position by using a least squares

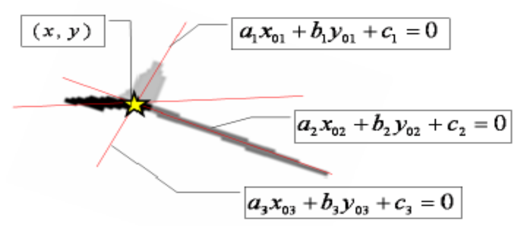

Figure 2. Player's position estimated using a shadow region and two body regions

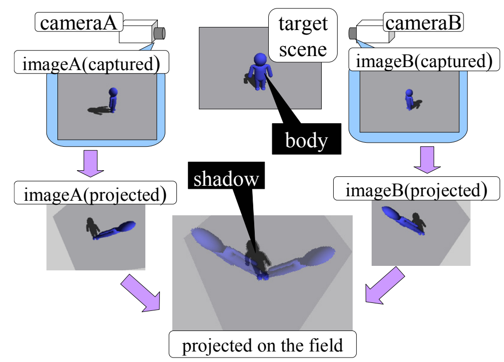

Figure 3. Distinction of body and shadow regions

method. This estimation method works correctly even when the shadow region does not exist on cloudy days. Because it is possible to calculate the intersection point (x, y) by using at least two lines. In night game illumination this method also works well, since weak shadows are removed in the foreground subtraction process described in section 3.1.

3.3. Distinction of body and shadow regions

As described in the above, our method needs to distinguish body regions from shadow ones, they usually touch with each other as illustrated in Figure 3. It is not easy to separate the two regions robustly in a single image. We distinguish the body and the shadow regions by utilizing the images from two cameras based on the property that a shadow exists on the surface of the soccer field while a body does not.

We project the candidate foreground regions in each of the images onto the field plane by using 2D projective transformation as illustrated in Figure 3. The parameters of the 2D projective transformation can be derived from the parameters of the cameras' 3D projection matrix. The shadow regions from the two cameras exactly coincide with each other on the field plane, while the body regions make two different elongated regions, as in the bottom of Figure 3. It is easy to separate and distinguish the body and the shadow regions based on this characteristic.

4. Classification of Player's Team

4.1. Clustering the color of uniforms

When multiple players are located closely to each other, it becomes difficult to estimate the correct trajectories of the players because of the confusion in tracking. In a soccer game, every player wears his own team's uniform, and the uniforms of the two teams have different colors. We try to improve the stability of the tracking process by classifying the players into each team by using color information. The number of color clusters can be previously known in a soccer game. Then we employ the k-means algorithm for the clustering process.

The L*a*b* color system is used to represent the color, since its distance measure is similar to human color perception. The Mahalanobis distance is used for the distance measure in each cluster to deal with the various color distributions.

If we try to update the color data and execute k-means clustering in every frame, the processing cost is prohibitive for real-time processing. Thus we perform k-means clustering only once by using the learning data obtained from the first 30 frames. After the initial phase, the body regions in each frame are classified based on the seed points. When the Mahalanobis distance is larger than a threshold value, we label the target region as 'suspicious' because the reliability for correct classification is low.

4.2. Integration of clustering results of two images from different cameras

In order to increase the accuracy of the clustering process, we integrate the results estimated in two images from two different cameras. When a player is classified to a same class in both images, the class is accepted. When a player is classified into different classes, it is labeled as 'no-decision.' When a player is classified into a class in one image but is labeled as 'suspicious' in the other image, the class determined in one image is accepted.

5. Tracking the Trajectory of a Player

The player's trajectory is estimated by connecting the positions of a player in a sequence of frames.

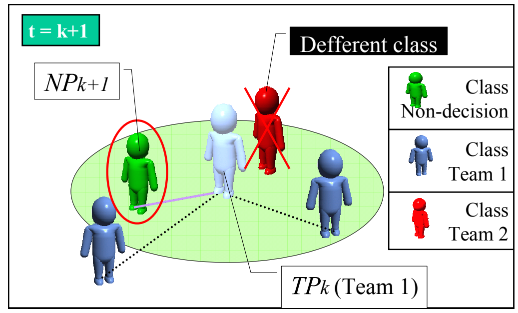

Figure 4. Tracking the trajectory of a player

Since we capture the video at 30 frames per second (fps), the inter-frame motion of a player is small. As illustrated in Figure 4, we first detect the nearest player NP_{k+1} in the $k+1$th frame from the target player TP_k in the kth frame. If the class of NP_{k+1} is the same as that of TP_k or the class is 'no-decision,' and the distance between NP_{k+1} and TP_k is less than a threshold, we determine NP_{k+1} as TP_{k+1} and move to the next fame. If the class of NP_{k+1} is different from that of TP_k, the nearest player NP_{k+1} is discarded and the next-nearest one is examined.

When the distance between NP_{k+1} and TP_k is larger than the threshold, we suppose that the target player is lost in the $k+1$th frame for some reason. Then, we move to the next frame to search for TP_{k+2} with a larger threshold value. When we detect the lost player again by repeating the searching process n times, the lost positions are estimated by linear interpolation using the positions TP_{k+n} and TP_k.

6. Experiments

6.1. Environment

We have conducted experiments to evaluate the effectiveness of our proposed method at a real soccer stadium, the National Stadium in Tokyo. As illustrated in figure 1, half of the soccer field is captured by two cameras that look down from overhead positions, such as the top of the electric signboard and the roof of the main stadium. The cameras are Sony DXC9000s and they are synchronized by GPS signals. The size of the captured image is VGA (640 x 480 pixels) and the frame rate is 30 fps. The relationship between the camera and the stadium coordinate systems is calibrated by using 3D laser-based surveying equipment. We obtained a set of video sequences from a real soccer game at the stadium.

6.2. Results

Figure 5 (a) shows the player's trajectory obtained by using the proposed method. Figure 5 (b) shows that obtained without the team classification. In both results, the player with a white uniform is correctly tracked for frames t=1 to t=8. In frame t=8, the white-uniformed player crosses the blue-uniformed player in front of the goal. As the result, the method without the team classification fails to track the player in the white uniform after the frame. On the other hand, it is possible to prevent false tracking by using our proposed method with its team classification process.

Table 1 shows the accuracy of player detection. A sequence of 600 frames is processed to obtain the results in Table 1. The false negative rate increases when multiple players become close, since we do not consider the number of players in a region in the current system. We can decrease this error by estimating the number of players in a region based on the correspondence between frames.

Table 2 shows the accuracy of team classification. A sequence of 600 frames is processed to obtain the results in Table 2. Although many players are classified as 'no-decision' because the lighting condition drastically changed during the game and our image resolution is low, it is possible to detect the class of a 'no-decision' player by using the correspondence between frames. Here, false classification results from failure in the foreground region segmentation, e.g., only the foot area is detected as a player region.

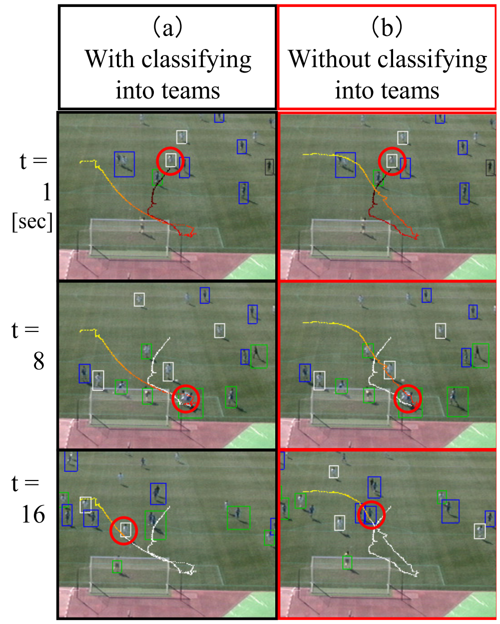

Figure 5. Estimated trajectories of players.

Table 1. Accuracy of player detection.

Total players #	Detected players #	False detection #	False negative rate	False positive rate
11810	9896	38	16.20%	0.30%

Table 2. Accuracy of team classification

Dected players #	Correctly classified #	No-decision #	Falsely classified #	Successfully classified #
9896	6768	3130	9	68.30%

7. Conclusions

In this paper, we have proposed a method to estimate the trajectories of objects robustly under difficult conditions such as an outdoor soccer stadium. By using the images from two cameras, we can correctly detect the region of a soccer player's body by identifying the shadow region. The 3D locations of players on the soccer field can thus be estimated. The false tracking caused by contact plays is reduced by classifying each player based on the uniform's color. We have applied our proposed method to a set of video sequences of a real soccer game captured at National Stadium in Tokyo. The results demonstrate the validity of our method.

References

[1] T. Kanade, P. Rander, P. J. Narayanan, "Vitualized Reality: Constructing Virtual Worlds from Real Scenes," IEEE Multimedia 1997, Vol. 4, No. 1, pp. 34-47.

[2] W. Matusik, C. Buehler, R. Rasker, S. J. Gortler, L. McMillan, "Image-Based Visual Hulls," ACM SIGGRAPH 2000, pp. 369-374

[3] T. Koyama, I. Kitahara, Y. Ohta, "Live Mixed-Reality 3D Video in Soccer Stadium," ISMAR 2003, pp. 178-187

[4] K. Uchida, J. Miura, Y Shirai, "Tracking Multiple Pedestrians in Crowd," Proc. Workshop of Machine Vision and its Applications, pp. 533-536, Nov. 2000.

[5] T. Misu, M. Naemura, W. Zheng, Y. Izumi, K. Fukui, "Robust Tracking of Soccer Players Based on Data Fusion," In Proc. 15th International Conference on Pattern Recognition, 2002, pp. 556-561.

[6] Bruno Muller Junior, Ricardo de Oliveira Anido, "Distributed Real-Time Soccer Tracking," VSN'04, October 15

[7] CHEN Baisheng, LEI Yungi, "Indoor and Outdoor People Detection and Shadow Suppression by Exploiting HSV Color Information," CIT'04, 0-7695-2216-5/04