MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11-13, 2002, Nara-ken New Public Hall, Nara, Japan

3—16 Tracking Soccer Player Using Multiple Views

Sachiko Iwase Hideo Saito
Department of Information and Computer Science,
Keio University *

Abstract

In this paper, we propose a method of tracking a soccer player using multiple cameras. For tracking soccer players, occlusion is always the big problem and tracking is often failed when only a single camera is used to take the scene. Therefore, we use multiple view images for avoiding the occlusion problem, so that we can obtain robustness in player tracking. In our method, inner-camera operation is performed independently in each camera for player tracking. If the player is not detected in the camera, inter-camera operation is performed and tracking information of all cameras are integrated by the geometrical relationship between cameras which is represented by fundamental matrices. Then the location of the player on each camera image is obtained in every frame. Experimental results show that robust player tracking is available by taking advantage of using multiple cameras.

1 Introduction

Soccer is one of the most popular sports around the world, and is often broadcasted on television. By using those movie data, various researches have done for soccer scene analysis. For soccer scene analysis, tracking player's location is significant technology, and it requires accuracy on getting player's location. Many researches have been focused upon player tracking.

There are researches of soccer scene analysis with the aim of strategy understanding and making digest TV programs of soccer games. One of the researches is to make use of some observing information such as texture, color, and shape, and motion estimation by Kalman fileter for player tracking[3]. Tracking player is mostly the base of soccer scene analysis. However, when only a single camera is used to take images of the scene, tracking is often failed when a player is occluded by others. Such occlusion is very common in a soccer scene, because a number of players participate in a soccer game.

Furthermore, there are researches of making images which help to understand the situation and have realistic sensations of the soccer game, such as generating intermediate view of soccer scene from multiple videos[1]. This enables to get images from the view which the user requests, though in fact there are no camera from that view.

For the researches of the camera control analysis, the goal is to achieve intelligent robot camera in TV programming production[4]. It is possible for the camera to move by a program, but it is still impossible to move and get the views by tracking the object in image data processing. However, optimized camera view point determination system is proposed[2], by filming the soccer game using multiple stable cameras and tracking the soccer ball. It gives an approach to an intelligent and automatic system of the camera control.

In our work, multiple cameras area used to get over the occlusion problem. Like above, the need of getting accurate location of the player is high for soccer scene analysis, therefore our work might help to have information of the player. Information as a location of the player in the camera image can be used in various researches; 3D-reconstruction of the soccer scene, intelligent and automatic system of the camera control, strategy understanding, some tools for soccer game broadcasting, and so on.

In addition, our method does not require strong calibration of multiple cameras which costs great time and effort. It is important to get the geometric relationship between cameras when integrating the information of multiple views. Usually camera calibration is difficult, because it is hard to see marker points with known 3D positions for calibration over a large area like a soccer ground. In our method, fundamental matrix is used which is computed from natural feature points commonly found in each camera, then the locations of multiple cameras are easily obtained.

2 System Environment

System environment is shown in Fig.1. Soccer scene is taken by 8 stable cameras, with 4 cameras, which are lined up relatively equal, set at both sides of the ground, aiming at the penalty area.

It is almost impossible to do camera calibration by using marker points with known 3D positions in a soccer stadium like this(Fig.1). In this paper, fundamental matrices between the cameras are computed by using about 20 natural feature points as corresponding points between the images.

In an initial frame, the location of the player to be tracked is given by hands in each camera. Then the player-location, x-y coordinate of its center, is found or estimated in each camera in every frame.

Flow of the method is shown in Fig.2. First, inner-camera operation is performed independently in each camera to track player. If occlusion does not occur, tracking is done by inner-camera operation alone, then the location of the player is saved. If the occlusion occurs, or the player is outside the angle of view, inter-camera operation is performed. Tracking information of all 8 cameras are integrated by the geometrical relationship between cameras which is represented by fundamental matrices, and the location of the occluded

*Address: 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, 223-8522 JAPAN. E-mail:{sachiko,saito}@ozawa.ics.keio.ac.jp

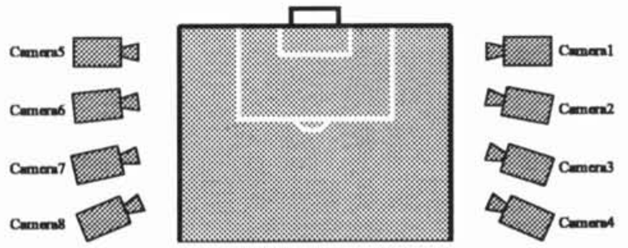
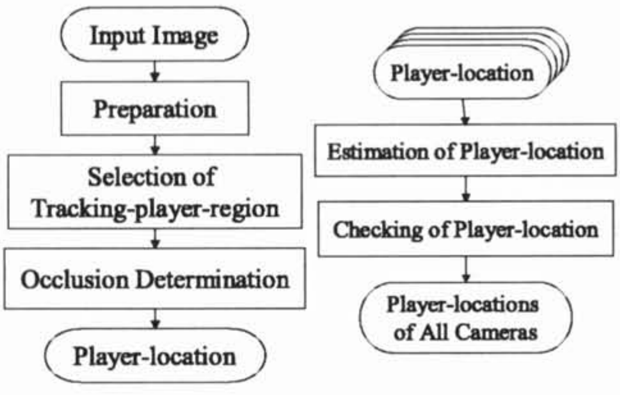


Figure 1: Locations of Cameras



(a) Inner-camera Operation (b) Inter-camera Operation

Figure 2: Flow of the Method

player is estimated. This method is based on the premise that occlusion determination is done with success in an inner-camera operation.

3 Inner-camera Operation

3.1 Preparation

Background subtraction is performed to extract player regions. Frame differenciation removes shadow effects, but makes it diffcult to detect those hardly move. As shadows do not appear in the input images, background subtraction is used. Then binary image is made by using thresholds of intensity and area with noise removal. Each region of the player is labeled and extracted, and its center and area are computed as feature values of the detected area.

3.2 Selection of Tracking-player-region

Player-region to be tracked is nominated from extracted player-regions of a current frame by using the player-location of a previous frame. If the player is inside the angle of view, it is nominated by calculating the moving distance from the player-location of a previous frame. If the player is outside, selection is not performed and the player-location is computed by inter-camera operation.

3.3 Occlusion Determination

In this section, whether the tracking player is occluded by other players in the nominated-player-region is determined. In occlusion determination, area of the nominated-player-region and number of the labels around the tracking player are compared between previous and current frame. For example, in the case that the area of the nominated-player-region increased and the number of the labels decreased, occlusion has occured.

Cameras determined not to be occluded in a nominated-player-region are able to track the player only by inner-camera operation. In this case, center of the nominated-player-region becomes the player-location to be tracked. Cameras determined to be occluded or in which the tracking player is not detected are not able to track the player by inner-camera operation alone. In that case, player-location is estimated by the next inter-camera-operation.

4 Inter-camera Operation

4.1 Estimation of Player-location

By epipolar geometry, computation from a pixel on the image of a certain camera to a corresponding epipolar line on the image of the other camera is possible. If the player-locations were obtained by inner-camera operation alone in more than 2 cameras, the player-location of the camera which is determined to be occluded could be estimated. Estimation is done by calculating an intersection of 2 epipolar lines, each corresponding to 2 of player-locations obtained in inner-camera operations. Fig.3 shows the player-location estimation when occlusion occurs. Same estimation is done when the player does not appear on the image.

Estimation requires more than 2 cameras which the player-locations are obtained by inner-camera operation alone. If the player-location is obtained in only one camera or none, it is impossible to do the estimation.

In the estimation, 2 cameras have to be selected from non-occluded cameras. Those 2 cameras are chosen in order that the distance between the cameras becomes the longest. If the distance is short, intersect angle of the 2 epipolar lines becomes small. In that case, error in the calculation of the intersection becomes big and it causes the tracking to fail. As geometrical relationship between the cameras and the soccer field is roughly known, relative distances between the cameras are easily obtained.

4.2 Checking of Player-location

By inner-camera operation and inter-camera operation, player-location is obtained in all cameras. However, after the occlusion occurs or when the player appears on the image in a mid frame, it tends to start tracking a wrong player. In order to achieve stable tracking, it is necessary to check the player-location using information of multiple views and epipolar geometry.

To check the player-location, epipolar lines are drawn to the cameras in which the player-location is obtained by the inner-camera operation alone, from the player-location which needs to be checked. Then in each camera which the epipolar line is drawn, distance between the epipolar line and the player-location of the camera is computed. If the distance is within the

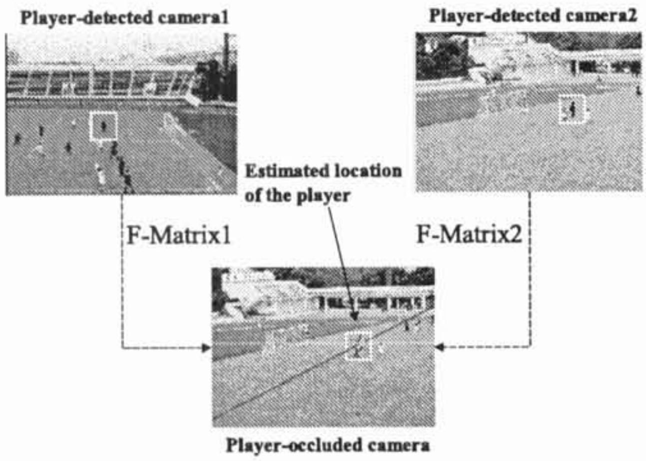


Figure 3: Estimation of the Player-location

threshhold, player-location is checked to be correct. If not, player-location is checked to be incorrect, so other player-location in the camera is selected and cheking is done in the same way.

5 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 2 scenes(350 frame and 190 frame).

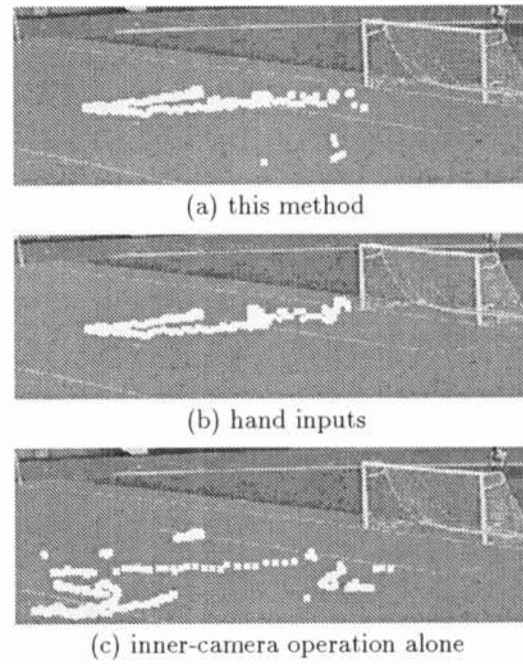
5.1 Tracking Results

Fig.4 shows an experimental result(a), real trajectory of the player gained by hand inputs(b), and trajectory of the player obtained from only inner-camera operation(c), of camera4 in the scene of 350 frames. Comparing (a) and (b), accurate and stable player tracking is seen by the proposed method. In some frames, tracking seems to be failed. This is because fundamental matrices contains error in some degree, so it is thought that this error give some effect to the estimation of the player-location.

Fig.5 shows the trajectory of the player that is represented in the virtual top-view camera. For obtaining such virtual top-view trajectory, the fundamental matrices of the top-view image and the images of the 8 cameras are calculated, so that the location of the tracking player can be easily computed. To calculate fundamental matrices, about 10 feature points are used such as corners of the penalty area and the goal area.

Table.1 shows the rate of some cameras of which the player tracking is succeeded in the scene of 350 frames. For example, in camera2, tracking player became occluded with other players and then separated from them for 4 times. However there are no frame which the wrong player is tracked. Tracking Failed is counted as frames. Table.1 shows that tracking is highly succeeded in any camera although occlusion has occured a few times.

On some cameras in the initial frame of the scene of 190 frames, tracking-player is outside the angle of view and do not appear on the image. However on



(a) this method

(b) hand inputs

(c) inner-camera operation alone

Figure 4: Trajectory of the Player

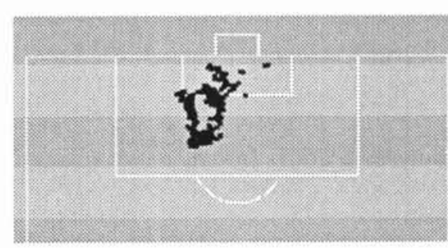


Figure 5: Trajectory from the Upper View

other cameras, tracking-player appears, and his location is able to obtain on that camera. Thus it is possible to start tracking the player by using information from other cameras even if he appears on the image in a mid frame. In single camera tracking, tracking can not be started if the player does not appear on the image and its location is not given in a initial frame. In this proposed method, estimation of player-location is possible using player-locations from other cameras, so tracking can be started in a mid frame.

In exceptional cases, tracking fails. If the scene is too crowded and the tracking-player is occluded by not only one player but two or three players, it tends to mistake in occlusion determination or in checking player-location. Then the wrong player begins to be tracked as a tracking-player. However there are cases that tracking is corrected by using information of the other cameras.

5.2 Comparison with tracking by a single camera

Comparing (a) and (c) of Fig.4, it is obvious that tracking is more robust in the use of multiple cameras than of a single camera. As it is not able to get information from other cameras by a single camera tracking,

Table 1: Rate of Tracking Succeeded

Camera Number	Occlusion Count	Tracking Failed (frame)	Rate
2	4	0	100%
4	6	8	97.7%
6	6	18	94.9%
8	6	3	99.1%

location of the occluded player can not be estimated. Furthermore, once the tracking has failed and start to track other player, it is impossible to track the right player again.

Fig.6 is the graph which shows how the tracking by a single camera and by multiple cameras differ in ditance from the real trajectory, in camera 4 of the scene of 350 frames. In single camera tracking, player is well tracked for first 40 frames, but wrong player is tracked after that. It never starts to track the correct player again. In the tracking by this method, although wrong player is tracked in some frames, it starts to track the correct player by the use of multiple cameras and tracking is succeeded after all. In this way, tracking is done more robustly compared to the single camera tracking.

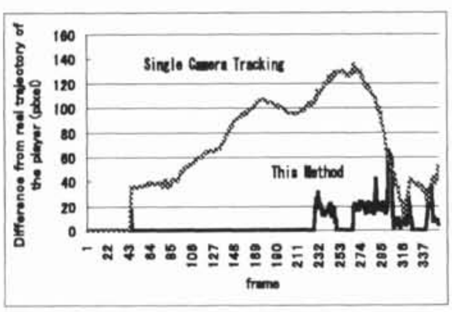


Figure 6: Comparison by real trajectory

6 Conclusion and Future Work

In this paper, a new method for player tracking is proposed, using multiple views to avoid an occlusion problem. Fundamental matrix which represents the geometrical relationship between cameras is calculated from only a weak calibration, then the information of multiple cameras are intergrated.

In this work, robust player tracking is available by using multiple cameras. Soccer scene analysis needs to have accurate loation of the player because it becomes the base of the research. This work might help to have information of the player.

For further research, we are now trying to track more than two players applying the proposed method. Main methods, the inner-camera operation and the inter-camera operation, are mostly the same as the proposed method except the occlusion determination. In the proposed method, only the area and the number of the players around the tracking player is compared as only a single player is tracked. It happens to fail in occlusion determinatin when the scene is too crowded. However, if all players appeared on the image are tracked, information of the other players also can be used to determine whether the tracking-player is occluded or not. By adding information of other players to the conditions of occlusion determination, it is considered that success rate of occlusion determination would increase and thus the robustness in tracking is available.

Furthermore, use of homography is taken account of instead of the fundamental matrix. Computed fundamental matrices used in this method naturally contain some error, and give negative effects on the intersection calculation of the epipolar lines. Also, player-location has to be obtained by inner-camera operation alone in more than 2 cameras in order to estimate the location in other cameras. However, while fundamental matrix gives relationship of the point on one image and the line on the other image, homography gives one-to-one relationship of the points between 2 images. Thus the error in the location estimation might be small, and player-location of more than a single camera is needed to estimate the location in other cameras. Homography can be used in the same way as fundamental matrix, and the tracking might be done more stable than the method which is using fundamental matrix.

We will work on to get more stable player tracking. The goal in the future is to track all players appeared on the image throughout the game. Information of the players enables strategy analysis, reconstruction of soccer scenes, and control of the camera when making a TV program. If the soccer ball is also tracked, automatic judgement of the offside rule is available. It is considered that application of this proposed method is wide and various.

References

[1] Naho Inamoto, Hideo Saito, "Intermediate View Generation of Soccer Scene from Multiple Videos", Proc. of International Conference on Pattern Recognition (ICPR 2002), Aug.2002

[2] Keisuke Matsumoto, Satoshi Sudo, Hideo Saito, Shinji Ozawa, "Optimized Camera View Point Determination System for Soccer Game Broadcasting", IAPR Workshop on Machine Vision Applications, 3-22, 2000

[3] Toshihiko Misu, Masahide Naemura, Shin'ichi Sakaida, Wentao Zheng, Yasuaki Kanatsugu, "Robust Tracking Method of Soccer Players Based on Data Fusion of Heterogeneous Information", Technical Report of IEICE, PRMU2001-67(2001-07)(in Japanese)

[4] Daiichiro Kato, "New Technologies for Supporting Program Producing. Application to Broadcast Programs and Intelligent Robot Cameras", NHK Science and Technical Research Laboratories R&D, Vol.30, No.10, pp.24-31, 1998(in Japanese)

[5] Vera Kettnaker, Ramin Zabih, "Counting People from Multiple Cameras", IEEE Int Conf on Multimedia Computing and Systems, VOL.2th, pp.267-271(1999)

[6] Paulo Peixoto, Jorge Batista, Helder Araújo, "Integration of information from several vision systems for a common task of surveillance", Robotics and Autonomous Systems, Vol.31, Issues 1-2, Page.99-108, 2000

[7] Yiming Ye, John K. Tsotsos, Karen Bennet, Eric Harley, "Tracking a Person with Pre-recorded Image Databases and a Pan, Tilt, and Zoom Camera", IEEE Workshop on Visual Surveillance, Page.10-17, 1998