Tracking Soccer Players Based on Homography among Multiple Views

Sachiko Iwase and Hideo Saito

Department of Information and Computer Science, Keio University, Yokohama, Japan

ABSTRACT

In this paper, we propose a method of tracking soccer players using multiple views. As many researches have done on soccer scene analysis by using trajectories of the players and the soccer ball, it is desirable to track soccer players robustly. Soccer player tracking enables strategy analysis, scene recovery, making scenes for broadcasting, and automatic system of the camera control. However, soccer is a sport that occlusion occurs in many cases, and tracking often fails by the occlusion of the players. It is difficult to track the players by using a single camera alone. Therefore, we use multiple view images to avoid the occlusion problem, so that we can obtain robustness in player tracking. As a first step, inner-camera operation is performed independently in each camera to track the players. In any case that players can not be tracked in the camera, inter-camera operation is performed as a second step. Tracking information of all cameras are integrated by using the geometrical relationship between cameras called homography. Inter-camera operation makes it possible to get the location of the player who is not detected in the image, who is occluded by the other player, and who is outside the angle of view. Experimental results show that robust player tracking is available by taking advantage of using multiple cameras.

Keywords: tracking, soccer, homography, multiple cameras

1. INTRODUCTION

Soccer is one of the most popular sports around the world, and the games are often broadcasted on television. By using those imaging data, various researches have done for soccer scene analysis. One of the applications of soccer scene analysis is strategy understanding. Collecting locus of every player is very important for understanding a team sport like soccer. Soccer scene analysis can also be applied to broadcasting of soccer games, such as making images which help to understand the situation and experience exciting scenes of the soccer game. Generating intermediate view of soccer scene from multiple videos enables to get images from the view which the user requests, though there are no camera from that view.[3]

There are other related works reported focusing on soccer. For example there is a research[2] that converts a video sequence of soccer game into an animated virtual 3D scene. The players and the ball are tracked and a 3D virtual scene is constructed with Open GL, in which users can walk freely. As an example, other researches are the one that estimates the distance of the ball in a soccer game,[1] and the one that proposes image mosaicing method of soccer video taken from a rotating and zooming camera using line tracking and self-calibration.[6]

There are researches of the camera control analysis for the use in the sport casting. The goal is to achieve intelligent robot camera in TV programming production.[5] It is possible to move the camera by a program, but it is still impossible to move the camera by tracking the object in image processing. However, optimized camera view point determination system is proposed, by filming the soccer game using multiple stable cameras and tracking the soccer ball.[7] It gives an approach to intelligent and automatic system of the camera control.

In soccer scene analysis by image processing, robust tracking of the player is important, because it provides essential information on the events happening in the soccer game. Tracking player is mostly the base of soccer

Further author information:
E-mail: {sachiko,saito}@ozawa.ics.keio.ac.jp,
Telephone: +81-45-563-1141, FAX: +81-45-566-1747
Address: 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, 223-8522, Kanagawa, Japan

scene analysis. In most of the researches of player tracking, only a single camera is used so far. There are researches[8–12] that make use of some observing information such as texture, color, and shape, and motion estimation by Kalman filter for player tracking. In the work of Misu and et al.,[8] at first they observe the texture, color, and shape of the target to detect its position, size and velocity. Then Kalman filters are used to adaptively fuse the detection results depending on the occlusion status and on the reliability of the detection. In the work of Utsumi and et al.,[11] they propose a object detection and tracking method refering to color rarity and local edge property, and which integrally evaluate them by a fuzzy function. Their tracking method can track objects occluded each other using a color based template matching.

When only a single camera is used to take images of the scene, tracking is often failed when occlusion occurs. Such occlusion is very common in a soccer scene, because a number of players participate in a soccer game. In this paper, we will propose a method of tracking soccer players using multiple cameras. By using multiple view images, we can avoid the occlusion problem, and obtain robustness in player tracking. The location of the player on each camera image is obtained in every frame. The need of getting accurate location of the player is high for soccer scene analysis, therefore this work might help to have information of the player.

In addition, it is important to get the geometric relationship between cameras when integrating the information of multiple views like the proposed method. Usually camera calibration is done, which is difficult, as it is hard to state accurate points for calibration over a large area like a soccer ground. However this method does not require strong calibration of multiple cameras which costs great time and effort. In this method, homography is used which is computed from marker points commonly found in each camera, then the locations of multiple cameras are easily obtained.

2. SYSTEM ENVIRONMENT

Our system environment is shown in Figure 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. In an initial frame, the location, corresponding index number, and the state of the players are given for each camera. Then the player-location, x-y coordinate of its foot, is found or estimated in each camera in every frame.

It is almost impossible to perform camera calibration by using marker points with known 3D positions in a soccer stadium. In this paper, homography between the cameras are computed by using about 7 natural feature points as corresponding points between the images. Those natural feature points are such as the corners of the penalty area and the goal area.

Figure 2 shows the flow of the proposed method. First, inner-camera operation is performed independently in each camera for player tracking. In inner-camera operation, player-region is detected by background subtraction, and the player is tracked by using the information of the player-region such as area, color and foot coordinate. If the player is not detected in the camera, or if the player is occluded by the other player, or if the player is outside the angle of view, inter-camera operation is performed and tracking information of all cameras are integrated by using homography, the geometrical relationship between cameras.

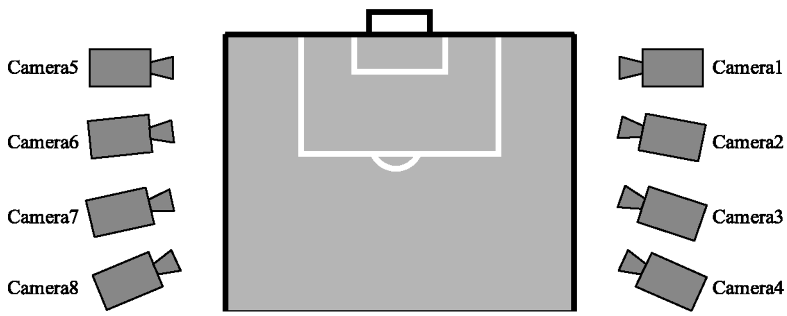
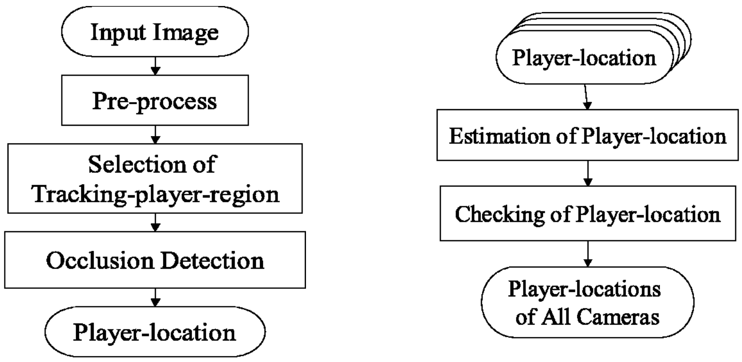


Figure 1. Locations of Cameras



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

Figure 2. Flow of the Method

3. INNER-CAMERA OPERATION

3.1. Pre-process

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. Temporal diffrential image removes shadow effects, but makes it difficult to detect those hardly move. As shadows do not appear in the input images, background subtraction is used. For each extracted player-region, its center, foot coordinate, length, width, and area are computed as player features.

3.2. Selection of Tracking-player-region

Second, matching of the player-region and the player-index is done. If the player is not occluded in the previous frame, player-region to be tracked is nominated from extracted player-regions of a current frame, calculating moving distance from the tracking-player-location of a previous frame. Also the area and the color of the extracted player-regions of a current frame are compared to those of the previous frame. If the player is outside the angle of view, if the player-region is not detected, or if the player was occluded in the previous frame, matching is not done in this selection of tracking-player-region.

3.3. Occlusion Detection

Third, it is determined whether the tracking player is occluded by other players in the player-region. In occlusion determination, area of the player-region is compared between previous and current frame. If the area of the player-region increases drastically, it is considered that the occlusion has occurred in the player-region in that frame. If the occlusion has occurred or the player was occluded in the previous frame, matching of the player-region and the player-indexes is done by comparing the center of the player-region instead of using the foot coordinate. This matching of the player-region and the player-indexes makes it possible to get which players are included in the player-region.

Non-occluded player is able to be tracked only by inner-camera operation. In this case, foot coordinate of the player-region becomes the player-location to be tracked and it is saved. The player determined to be occluded, not detected, or outside the angle of view are not able to be tracked by inner-camera operation alone. It is impossible to get the accurate location of the player. In that case, player-location is estimated by the following inter-camera-operation.

4. INTER-CAMERA OPERATION

4.1. Estimation of Player-location

By the epipolar geometry, computation from image points on a plane of a certain camera to a corresponding image points of the other camera is possible using homography as shown in Figure 3. There is a perspective, $x_2 = Hx_1$, between the two image planes. Homography is determined uniquely between the two images.

Every player is on the a world plane which is called as a soccer ground. If the player-locations were obtained by inner-camera operation alone in more than a single camera, the player-location of the occluded player could be estimated using homography. Figure 4 illustrates the estimation done by calculating a corresponding x-y coordinate from foot coordinates of the tracking player obtained in inner-camera operations. Same estimation is done when the player-region is not detected or when the player is outside the angle of view. Homography matrices between the cameras are calculated beforehand by using more than 4 natural feature points on the soccer ground as corresponding points between the images.

In the case when the tracking-player-locations were obtained by inner-camera operation in more than a single camera, x-y coordinate of the player to be tracked in each camera is added up and its average is computed at first. For each camera, difference between the average and the x-y coordinate of the player-location is calculated, and for the second time, x-y coordinate is added up only if the difference is within the threshold which is determined in advance. The second average is considered as an estimated player-location. The second x-y coordinate average is saved as an estimated player-location in the camera in which the player-location could not be obtained by inner-camera operation alone.

Estimation of player-location can not be performed if the player-location is not obtained by inner-camera operation alone in any camera. This means the player is occluded by the other player, is not detected, or is outside the angle of view in all cameras. Then the player-location can not be estimated as there are no information from the other cameras. In this case it is impossible to continue tracking of the player for the rest of the scene.

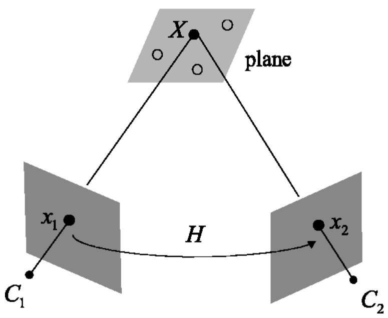


Figure 3. Homography

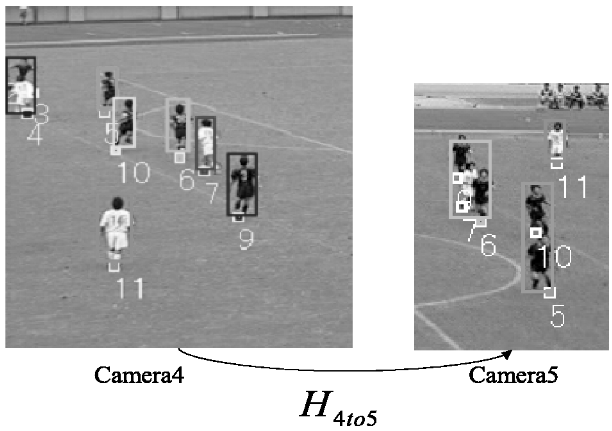


Figure 4. Estimation of Player-location

4.2. Checking of Player-location

By inner-camera operation and the estimation of player-location, player-locations are 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.

In order to check the player-location, difference between the player-location which is saved and the foot coordinate of the extracted player-region is computed, and the nearest player-region is found for all the players appeared on the image. Then the index number of the player region is checked. This process is done in each camera. If the player-region is not indexed, new index number is given by using the player information of other cameras. If the index number is checked as incorrect, player-location can not be obtained or estimated in that frame. However there is a chance to start tracking the player again if the player is kept tracking in the other cameras.

In the proposed method, all the players appeared in all camera images are tracked. Then the information of other players in all cameras can be used to check the player's location. Checking of player-location can not be done if only one player is tracked or if the players are tracked by using only a single camera. Checking the player-location by using the consistency of player information in all cameras enables high precision in tracking.

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, scene1 of 450 frames, and scene2 of 150 frames, around the penalty area.

The state of the initial frame of scene1 in all cameras are shown in Figure 5. The location which is the x-y coordinate of the player's foot, corresponding index number, and the state are given for each player in each camera as an initial information. Players are tracked with the initial information and estimatin of player-location is also done in the initial frame as shown in Figure 5. The index number is appeared at each player's feet and the xy-coordinate of its foot is marked white. Each player-region is surrounded with the boundary box. The state of scene1, frame100 in all cameras are shown in Figure 6 as a tracking result. Even after 100 frames from start and the occlusion has occurred a number of times, tracking is done like Figure 6. Tracking is done by obtaining the location of the player on each camera image in every frame like Figure 5 and 6.

Figure 7 and 8 focus on the players whose index numbers are 8 and 9, in scene1, frame94 and frame95, of 4 cameras. In Figure 7, player8 and 9 are occluded with each other. However as either player is tracked and its location is obtained by inner-camera operation alone in cameras 2 to 4, estimation of player-location can be done by using the geometrical relationship between the cameras, the homography. Thus the xy-coordinate of the foot of player8 and 9 could be computed by using the information of the other cameras. Just mentioning about camera5 to 8, both players are tracked by inner-camera operation alone in camera5, they are occluded with each other in camera6, and both players are outside the angle of view and not appeared on the images in camera7 and 8. Although in a theoretical sense, player-location obtained by inner-camera operation alone in a single camera is enough to do the estimation, player-location more than 3 cameras might be needed for a stable tracking.

As shown in Figure 8, player8 and 9 are not occluded anymore and they are separated. However either player is tracked right by using the player information of other cameras. Thus players are tracked accurately even if the occlusion occurs.

In camera3 in Figure 7, one player with a white uniform, who must be player9, is not tracked. This is because the mismatching of the player-region and the index number has occurred in the other place of the camera image, and player9 could not be tracked in frame94. However player9 is tracked in camera2 and 4, so the tracking of player9 is corrected again.

In camera3 in Figure 8, taking a look at player8 in camera2 and player9 in camera3, only the xy-coordinates of the foot are marked. This means the corresponding player-region could not be detected in some reason such as the player-region did not appear in the background subtraction. However as the player is tracked in inner-camera operation alone in other cameras, estimation of the player-location can be done. Similarly tracking is also succeeded when the player came inside the angle of view in a mid frame, if the player is being tracked in other cameras.

In Figure 9, there are 4 players occluded with each other, but only two players are tracked. For the other two players, estimation can not be done or mismatching of the player-region and the player-index has occurred. In the proposed method, tracking fails in exceptional cases. If the scene is too crowded and the tracking-player is occluded not only one player but two or three players, it tends to mistake in occlusion detection and checking of player-location. Then the wrong player begins to be tracked, which means the correspondense between the player-index and the player-region is wrong, and sometimes the error continues for quite a few frames. However there are cases that tracking is corrected by using information of the other cameras.

Figure 10 shows the trajectory of the players that is represented in the virtual top-view camera. This is the trajectory of scene2, from frame0 to 350. Such virtual top-view trajectory is easily obtained by computing the location of the tracking players using the homography between the top-view image and the camera images. By showing the trajectory of all the players on the virtual top-view image from the tracking results, we can see how the players move through the scene.

6. DISCUSSION

6.1. Epipolar and Projective Geometry

In this paper, planar homography in projective geometry is used, which is the relationship in more restricted condition than epipolar constraint. In our previous research,[4] fundamental matrix in epipolar geometry was used to track the player. The result was much better than that of the research which was using fundamental matrix instead of homography. As homography gives one-to-one relationship of the points between two images while fundamental matrix gives relationship of the point on one image and the line on the other image, error in estimation of player-location became smaller. Computed fundamental matrices naturally contained some error, and gave negative effects on the intersection calculation of the epipolar lines which was needed to estimate the player-location. Also, player-location had to be obtained by inner-camera operation alone in more than two cameras in order to estimate the location in other cameras when using the fundamental matrix. However player-location of more than a single camera is needed when using the homography. In addition, only a single player was tracked in the research using the fundamental matrix. In the research using the homography, all the players appeared on the images are tracked. As information of the other players can be used to determine whether the player-region is occluded or not, success rate of occlusion determination would have increased when using the homography. Thus the robustness in tracking is available.

6.2. Comparing with Tracking by a Single Camera

If only a single camera is used to do the tracking, tracking may often fail in the case of occlusion. Moreover, as the player information of the other cameras could not be used, there is no way to estimate the player-location when the player-region could not be detected or when the player came inside the angle of view in a mid frame. Tracking can not be started if the player does not appear on the image and its location, its state and its player-index are not given in a initial frame. Above all, once the tracking has failed and start to track the wrong player, it is impossible to track the right player again. In the tracking of the proposed 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. Thus tracking is done more robustly compared to the single camera tracking.

6.3. Future Works

For further research, we will try to get more stable player tracking. However there are two main problems in this proposed method. One is the matching of the player-region between the camera images. Sometimes mismatching occurs, so it must be figured out accurately that this player-region on this camera image corresponds to that player-region on that camera image. The other is the matching of the player-index and the player-regions in the camera images, which is the matter of tracking. It is needed to know who this player-region is in order to track players, so the mismatching induces error in player tracking. Concerning these problems for further research, we will try to track players on a virtual top-view camera image. By converting players' locations of all cameras to those of virtual top-view camera using homography, all of the data could be integrated on a virtual top-view image. Then the matching of player-regions among different camera images and the tracking could be done only on a virtual top-view image. Furthermore, improvement in tracking method is needed. In inner-camera operation, tracking, which is the selection of tracking-player-region, is done only by using the distance that the player has moved and the information like area and color. There is still way to improve the tracking such as using the motion estimation.

7. CONCLUSION

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

Experimental results show that the robust player tracking is available by the use of multiple cameras. As the need to have accurate location of the players is high in soccer scene analysis, proposed method might help to get the information of the players.

By using the tracking result, it is able to provide entertaining images to see soccer games. For example, it is able to cut out the image of the player and around the player, and see the player on demand. Users can request whoever they want, and see the image of the player from all views of the cameras. Information of the players enables strategy analysis, scene recovery, and control of the camera when making a TV program and so on. It is considered that application of this method is various.

REFERENCES

1. A. Branca, E. Stella, N. Ancona, A. Distante, "Goal Distance Estimation in Soccer Game," in *Proceedings of 11th International Conference on Image Analysis and Processing*, pp. 565–569, 2001.
2. Wei Du, Hua Li, André Gagalowicz, "Video Based 3D Soccer Scene Reconstruction," in *Proceedings of Mirage 2003*, INRIA Rocquencourt, pp. 70–75, 2003.
3. Naho Inamoto, Hideo Saito, "Intermediate View Generation of Soccer Scene from Multiple Videos," in *Proc. of International Conference on Pattern Recognition (ICPR 2002)*, 2002.
4. Sachiko Iwase, Hideo Saito, "Tracking Soccer Player Using Multiple Views," in *IAPR Workshop on Machine Vision Applications (MVA2002)*, pp. 102–105, 2002.
5. Daiichiro Kato, "New Technologies for Supporting Program Producing. Application to Broadcast Programs and Intelligent Robot Cameras," in *NHK Science and Technical Research Laboratories R&D*, Vol.30, No.10, pp. 24–31, 1998 (in Japanese).
6. Hyunwoon Kim, Ki Sang Hong, "Robust Image Mosaicing of Soccer Videos using Self-Calibration and Line Tracking," in *Pattern Analysis & Applications*, Vol.4, pp. 9–19, 2001.
7. Keisuke Matsumoto, Satoshi Sudo, Hideo Saito, Shinji Ozawa, "Optimized Camera View Point Determination System for Soccer Game Broadcasting," in *IAPR Workshop on Machine Vision Applications (MVA2000)*, 3-22, 2000.
8. Toshihiko Misu, Masahide Naemura, Shin'ichi Sakaida, Wentao Zheng, Yasuaki Kanatsugu, "Robust Tracking Method of Soccer Players Based on Data Fusion of Heterogeneous information," in *Technical Report of IEICE, PRMU2001-67*, 2001 (in Japanese).
9. Chris J. Needham, Roger D. Boyle, "Tracking multiple sports players through occlusion, congestion and scale," in *British Machine Vision Conference*, Vol.1, pp. 93–102, 2001.
10. Yoshinori Ohno, Jun Miura, Yoshiaki Shirai, "Tracking Players and Estimation of the 3D Position of a Ball in Soccer Games," in *Proceedings of 15th International Conference on Pattern Recognition*, Vol.1, pp. 145–148, 2000.
11. Okihisa Utsumi, Koichi Miura, Ichiro Ide, Shuichi Sakai, Hidehiko Tanaka, "An Object Detection Method for Describing Soccer Games from Video," in *Proceedings of IEEE International Conference on Multimedia and Expo.*, Vol.1, pp. 45–48, 2002.
12. Nicolas Vandenbroucke, Ludovic Macaire, Jack-Gérard Postaire, "Color image segmentation by supervised pixel classification in a color texture feature space. Application to soccer image segmentation.," in *Proceedings of 15th International Conference on Pattern Recognition*, Vol.3, pp. 621–624, 2000.

camera1

camera2

camera3

camera4

camera5

camera6

camera7

camera8

Figure 5. scene1 frame0

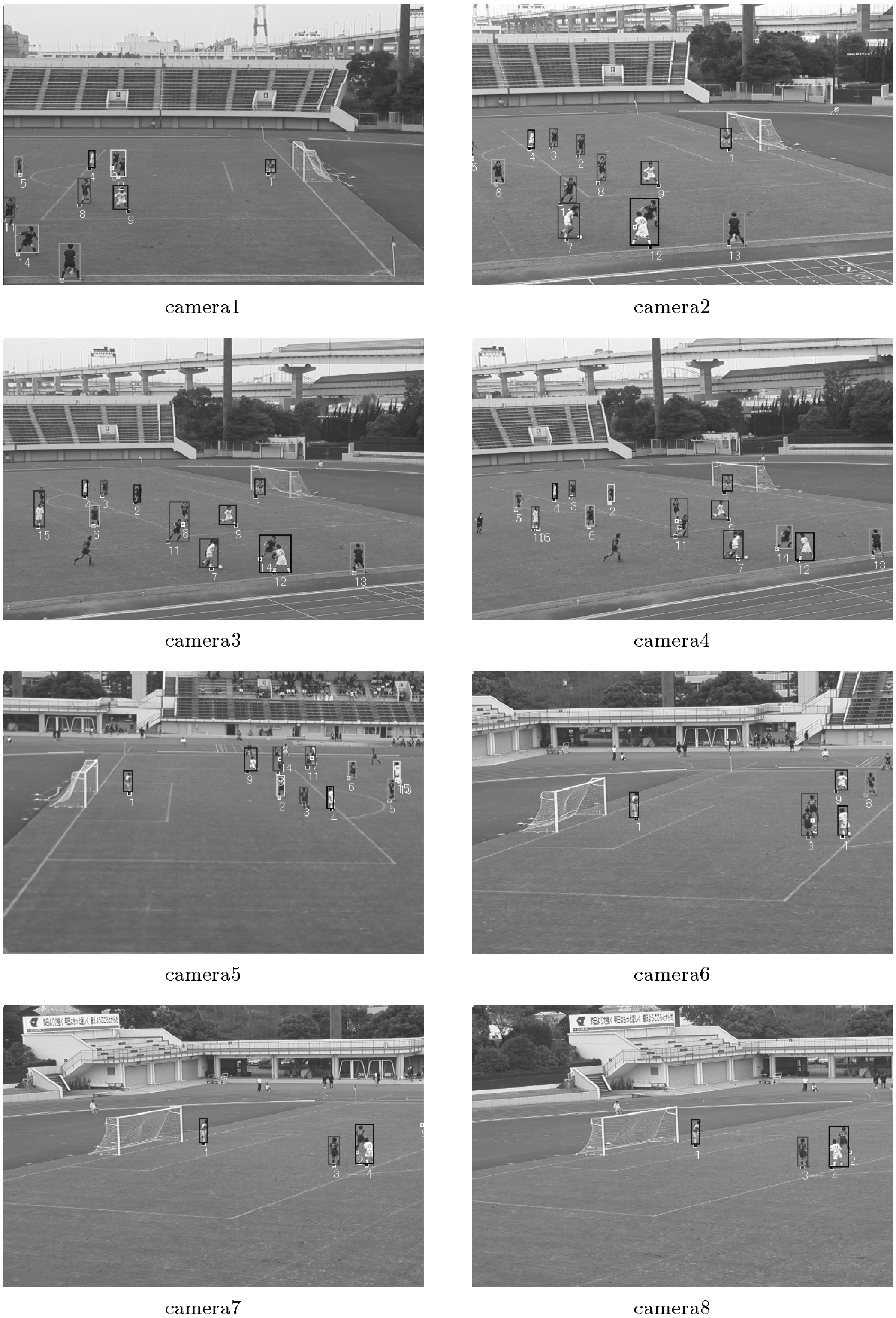
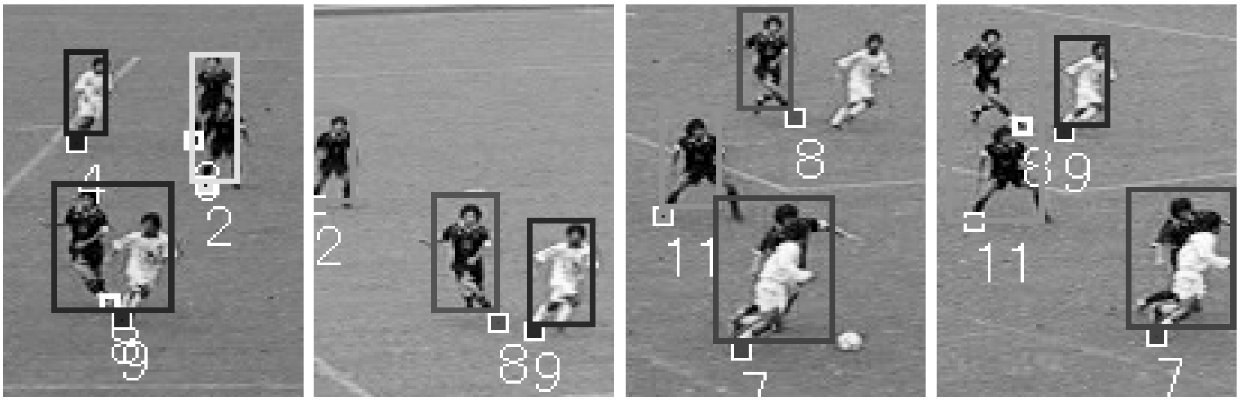
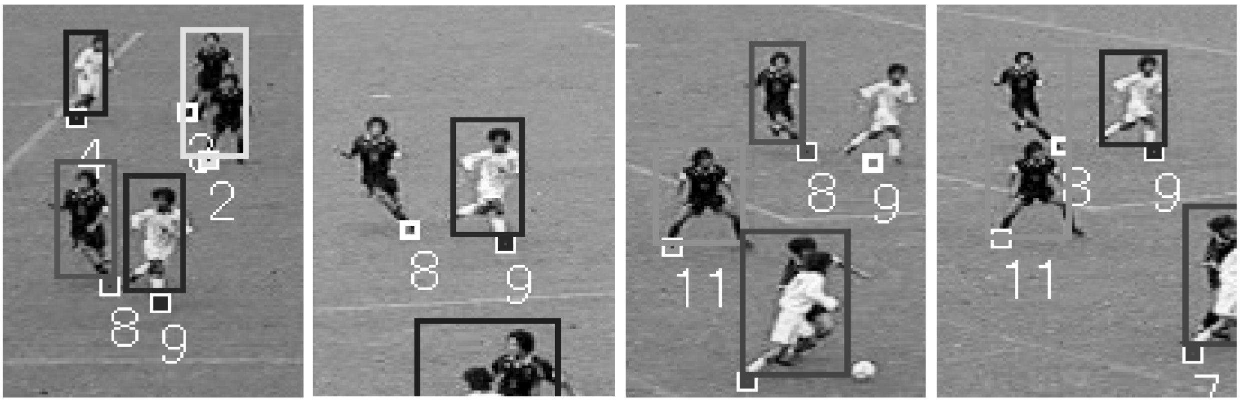


Figure 6. scene1 frame100



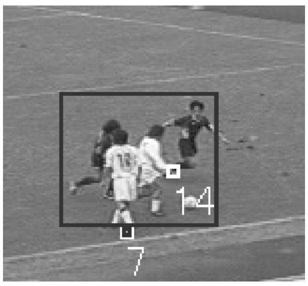
camera1 camera2 camera3 camera4

Figure 7. scene1 frame94



camera1 camera2 camera3 camera4

Figure 8. scene1 frame95



camera4, scene1, frame117

Figure 9. Failure in tracking

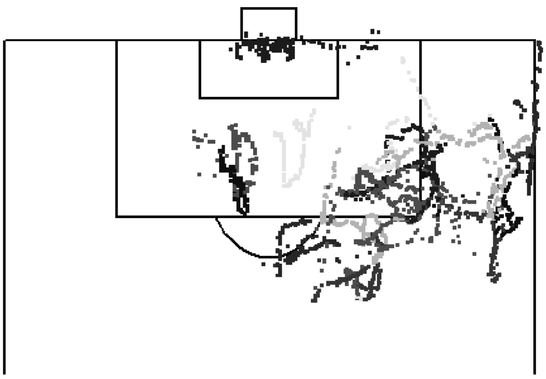


Figure 10. Trajectory from the Upper View