Sports Scene Analysis and Visualization from Multiple-View Video

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

Abstract

In this paper, we introduce methods for sports scene analysis and visualization from multiple videos captured with multiple cameras. As the scene analysis, we present a method for tracking of multiple soccer players. 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. We also present a method for free-viewpoint visualization of soccer game as visualization technique of soccer scene. The free-viewpoint image is synthesized by view interpolation between actual cameras near the virtual viewpoint at each frame. Such free-viewpoint video can also be presented in a system of augmented reality (AR) for immersive visualization of soccer matches.

1. Introduction

Video images captured with multiple cameras provide a lot of information on the object scene. Shape of the object scene is one of such information that can be obtained by multiple cameras. One pioneering project is Virtualized Reality [6], in which 3D shape models of objects in target scene are reconstructed from multiple view images, and then colors on the real images are used to form texture of the 3D model. Localizing and tracking moving objects can also be enabled by using multiple cameras [1].

In this paper, we introduce methods for sports scene analysis and visualization from multiple videos captured with multiple cameras. First, tracking of multiple soccer players is presented, in which homographic transformation on the ground among the cameras is employed for merging locations of players estimated in different cameras [5].

Next, we introduce a method for free-viewpoint visualization of soccer game [3], in which virtual view image at the selected viewpoint is synthesized by view interpolation [2] between actual cameras near the viewpoint at each frame. We also introduce a system of augmented reality (AR) for immersive visualization of soccer matches, with which an user sees a desktop soccer stadium model in front of his/her eyes through a head-mounted display (HMD), while the images of players and a soccer ball are overlaid onto the display [4].

For using multiple camera images, it is important to get the geometric relationship between cameras, which is namely camera calibration. For strongly calibrating cameras, 3D positions of several points in the objective space and their 2D positions of them on every camera must be precisely measured. This measurement generally requires much effort.

In our methods, therefore, we do not require such strong camera calibration, but only apply projective geometry among multiple cameras, which can be obtained from correspondences among 2D positions of natural feature points in the images. One representation of the projective geometry is Fundamental matrix that relates one point in an image to the epipolar line in another image. We also employ Homography that provides relationship of 2D positions between two images for every point on a plane in the scene. The projective geometry among cameras is easily obtained by detecting several natural feature points in every image and giving correspondence of them, even in the case of large sports stadiums, because we do not need to measure 3D position of the feature points in the scene.

2. Tracking of Multiple Soccer Players

2.1. Algorithm

Cameras are located around the soccer ground to cover all the area.Figure 1 shows the flow of the player tracking. First, inner-camera operation is performed independently in each camera to detect the features of the

0-7803-8603-5/04/$20.00 ©2004 IEEE.

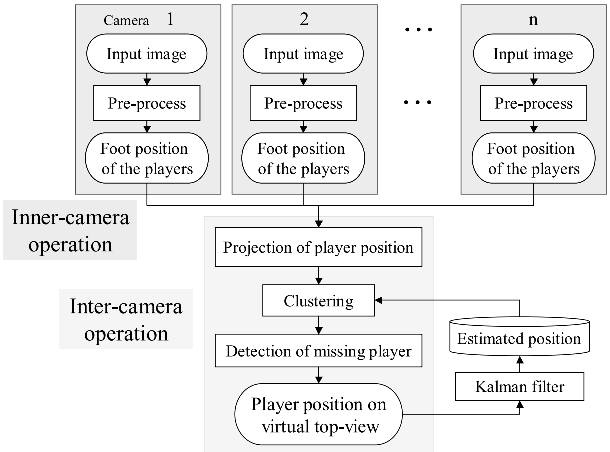


Figure 1. Flow of tracking

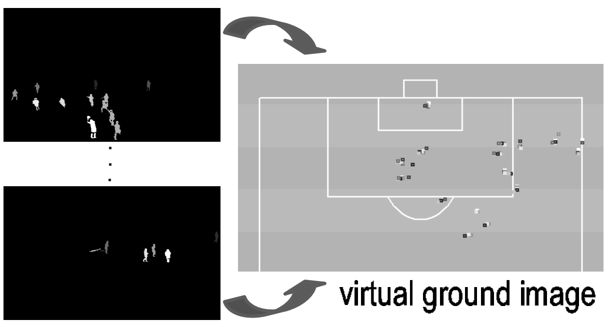


Figure 2. Virtual ground image.

extracted player regions in each camera image. First, background subtraction is done on an input image to extract player regions. For each extracted player-region, the center of gravity, the lowest vertical position as the foot position, vertical length, horizontal width, and area size are computed as player region's features.

Next, features of all the cameras that are detected in the inner-camera operation are integrated in inter-camera operation for obtaining trajectory of players. Homography of the ground is used for integrating the features.

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 2 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.

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. 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 $\mathbf{p_e}$.

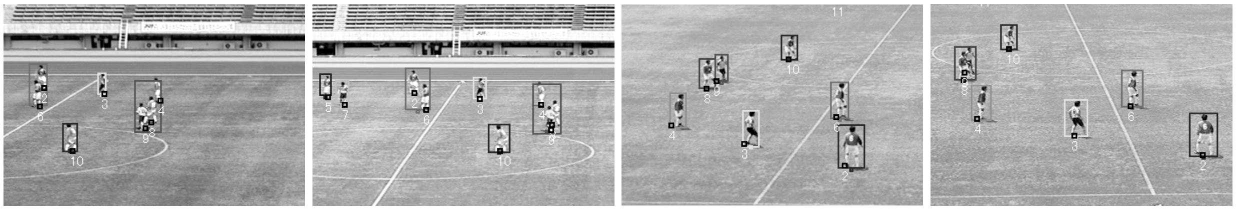
3 Update the estimated positions $\mathbf{p_e}$ with the centroids, and repeat **1**,**2** as the mean strain converges.

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

2.2. 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, 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.



(a) camera 3 (b) camera 4 (c) camera 12 (d) camera 13

Figure 3. Tracking in each camera image

Figure 3 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.

3. Free-Viewpoint Visualization

In this secion, we present two systems for free-viewpoint visualization of soccer matches, which are taken at stadium by multiple cameras are stored as video images. Figure 4 describes overview of the proposed systems. The observation system with a normal video display is shown at the top, and the system with a video see-through HMD is shown at the bottom.

For the both systems, the process consists of three stages: determination of viewpoint position, arbitrary view generation for the dynamic regions of soccer scene, and overlay the soccer scene on the stadium. At the first stage, a virtual viewpoint position is given on GUI directly for the observation system with a normal display, or determined by the position and pose of HMD cameras for the system with a video see-through HMD. At the second stage, neighboring cameras of the virtual viewpoint, which are reference cameras, synthesize the virtual viewpoint image of the dynamic regions by view interpolation. At the final stage, synthesized soccer scene is overlaid onto the real stadium images through a normal display or the desktop stadium model through the HMD.

In these systems, we employ view interpolation method as a key method for free-viewpoint visualization. For view interpolation of dynamic objects such as players and a ball, all dynamic regions are first extracted by subtracting the background image from the current frame image. After the silhouettes have been generated by binarization, every silhouette region of a player is

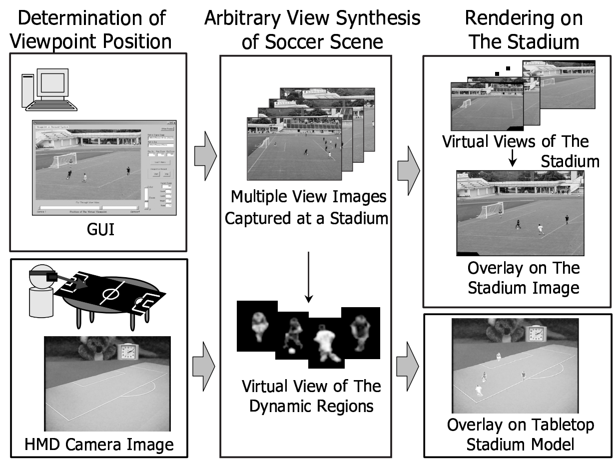


Figure 4. Overview of the proposed method.

corresponded by using homography of the ground plane between the reference cameras. This is based on the fact that all feet of the players are attached on the ground.

For each silhouette, epipolar lines are drawn between neighboring views by using a fundamental matrix. On each epipolar line, the edge points are corresponded and the pixels inside the silhouette are then corresponded by linear interpolation of the edge points. Thus the pixel-wise correspondence within the silhouette is obtained between each pair of silhouettes. Such pixel-wise correpondence is used for view interpolation of dynamic regions.

3.1. Visualization with Normal Display

In order to present the soccer scenes at free-viewpoint video on a normal video display, arbitrary viewpoint images for the soccer stadium are also generated by view interpolation. As the stadium including ground, soccer goal, and stadium background scenes are considered as static regions, which have little or no changes over time, view synthesis for all possible intermediate viewpoints is implemented just one time. The static regions are approximated as sets of planes.

View interpolation for each region generates intermediate view images, and synthesizing them completes stadium image from another viewpoint.

Free-viewpoint images of the whole soccer scene is sythesized by rendering the dynamic regions on the image of reconstructed stadium for the desired virtual viewpoint. The global appearance including players, ball, field and stadium are displayed in window of the GUI.

3.2. AR Visualization

In this system, a user sees the real world via a video see-through HMD and observes dynamic objects of soccer scenes overlaid onto the stadium model. In order to overlay the dynamic objects of soccer scenes on the stadium model captured with the HMD camera, geometric registration between overlaid soccer scene and the stadium model is necessary. This means that the virtual viewpoint should be determined by computing the position and pose of the HMD camera, while the viewpoint position is determined directly by users through GUI in the free-viewpoint video on a normal display. As the image of the soccer stadium model captured by the HMD camera contains natural feature lines, such as the lines of the penalty area or the goal area, we track these natural feature lines for the geometric registration. For geometric registration, we use homography transformation between the soccer ground plane in the real soccer match scene, and the ground plane of the stadium model taken by the HMD camera. This homography determines the positions of the players and the ball in the generated soccer image that is overlaid onto the HMD camera image.The homography is computed from more than 4 corner points of goal area or penalty area. The intersection points of detected feature lines are used as corner points of each area.

3.3. Experimental Results

We implemented two types of observation systems for an actual soccer match. Soccer matches were captured by multiple uncalibrated video cameras at actual soccer stadiums. A set of 4 fixed cameras was placed to one side of the soccer field to capture the penalty area mainly. The captured videos were converted to BMP format image sequences, which were composed of 720x480 pixels, 24-bit-RGB color images. Figure 5 shows example results of free-viewpoint visualization of a soccer scene. The upper row shows reference stadium images used for virtual view generation. The lower row indicates overlaid soccer scenes on the real stadium images and on the desktop stadium model. Interpolating weight and zoom ratio are indicated as w and z at the bottom of each image. Synthesized scenes

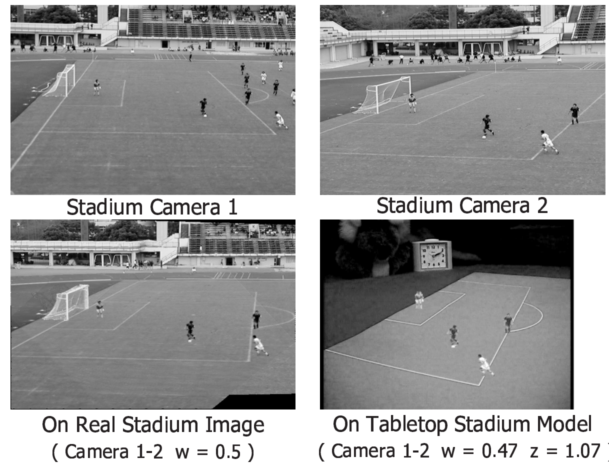


Figure 5. Determination of the rendering positions on the HMD.

on the real stadium images are photorealistic as if they are actual camera images from the viewpoint. In HMD images, we can see that the dynamic objects of soccer scene were overlaid correctly onto the real stadium model.

4. Conclusion

We present a method for tracking of multiple soccer players as soccer scene analysis, which is performed by integrating the tracking data from all cameras based on the geometrical relationship between cameras. We also present two systems for free-viewpoint visualization of soccer game with a normal display and a HMD for AR presentation.

References

[1] T. Chang, S. Gong, “Tracking Multiple People with a Multi-Camera System,” Proc. IEEE Workshop on Multi-Object Tracking (WOMOT’01), pp.19-28, 2001.

[2] S.E.Chen and L.Williams, “View interpolation for image synthesis,” Proc. of SIGGRAPH ’93, pp.279-288, 1993.

[3] N.Inamoto and H.Saito, “Intermediate view generation of soccer scene from multiple videos,” Proc. ICPR’02, vol.2, pp.713-716, 2002.

[4] N.Inamoto and H.Saito, “Immersive Observation of Virtualized Soccer Match at Real Stadium Model,” Proc.ISMAR03, pp.188-197, 2003.

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

[6] T.Kanade, P.J.Narayanan and P.W.Rander, “Virtualised reality:concepts and early results,” Proc. of IEEE Workshop on Representation of Visual Scenes, pp.69-76, 1995.