Free Viewpoint Video Synthesis and Presentation from Multiple Sporting Videos

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

Abstract

This paper introduces two kinds of free viewpoint observation systems for sporting events captured with uncalibrated multiple cameras in a stadium. In the first system (Viewpoint on Demand System), a user can watch the realistic sporting scenes with the original stadium. In the second system (Mixed Reality Presentation System), a user can virtually watch the scenes overlaid on a desktop stadium model via a video see-through head mounted display (HMD). In the both systems, the user can observe sporting events from his/her favorite viewpoints, where the virtual view images are synthesized and presented by performing view interpolation. As outdoor scenes often have variations in lighting, we develop the systems to handle the changes of lighting condition. If the captured scene contains shadows, we synthesize the virtual view image of the shadows of the background and the foreground independently from real camera images using projective geometry between cameras. The shadows of the foreground objects are then overlaid on the synthesized background of the original stadium or the stadium model in front of the user respectively. The results indicate that the appearance of shadows can produce a realistic mixed reality presentation.

1. Introduction

Computer-generated visualization is increasingly used in sports broadcasting to enhance the viewer experience. Virtual objects such as virtual offside line in soccer and virtual record in field races are inserted in live video [7]. A system that produces virtual replays of tennis ball trajectories [6] or virtual camera motion of panning around at the moment of fine plays in American football has been developed [2].

Recently, interactive visualization is becoming more important with the ongoing convergence of television and Internet broadcasting. In previous works, we proposed two kinds of systems that can give immersive impression and interactivity in watching a sporting event, especially a soccer match [4, 5, 8]. One is the "Viewpoint on Demand System", which enables a user to observe entire soccer scenes from his/her favorite viewpoint through a GUI. The other is the "Mixed Reality (MR) Presentation System", which enables a user to watch a soccer match overlaid on a desktop stadium model via a video see-through HMD.

In this paper, we improve these systems to handle variations in lighting. The appearance of shadows is important for a realistic presentation of outdoor sporting scenes. If players, ball and stadium in a captured scene have shadows, we need to present their shadows correctly in virtual view images.

First, we propose our method of synthesizing shadows in a virtual view from real camera images. In the Viewpoint on Demand System, the shadows of the background and the foreground are synthesized independently for visualization of an entire soccer scene. We then introduce our method of representing the shadows of the players and the ball on the desktop stadium model for the MR Presentation System.

It is generally required for rendering shadows of an object to obtain the 3D shape of the object. Strong calibration [10] to obtain relationship between 3D object space and 2D image have much effort in the case of multiple cameras located in a large space such as a stadium. Instead of 3D reconstruction, we employ the projective geometry among cameras to synthesize shadows in virtual view images and represent them on the desktop stadium. Corresponding feature points in the captured images yield the projective geometry among cameras. Therefore the proposed systems can be applied to dynamic outdoor sporting scenes and enable their free viewpoint observation.

2. Overview

We explain an overview of the proposed systems. As shown in Figure 1, a soccer match is captured using uncalibrated multiple cameras in a real stadium. The soccer scene synthesized using view interpolation [1] is presented from the chosen viewpoint by a user.

In the Viewpoint on Demand System, virtual view images of the background including the stadium and its

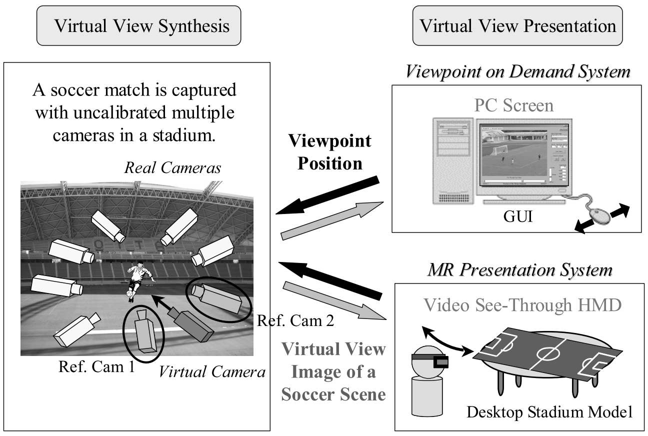

Figure 1: The overview of the proposed systems.

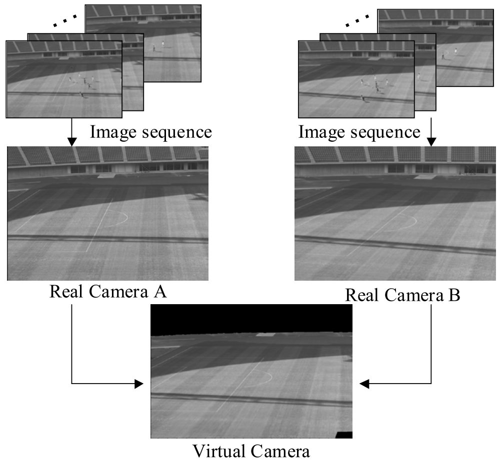

Figure 2: The example of synthesized backgrounds in the real cameras and the virtual camera.

shadow are generated in advance. Once the user selects the viewpoint position from the GUI, neighboring cameras near the chosen viewpoint, which are reference cameras, synthesize the virtual view image of the foreground objects such as a ball, players, and their shadows. The generated image is then overlaid on the virtual view image of the original soccer stadium at the corresponding viewpoint.

In the MR Presentation System, the user selects the viewpoint by moving his/her head from side to side. The viewpoint position is determined by the position and pose of the HMD. The virtual view images of the foreground objects, which are synthesized from reference cameras in the same way as in the Viewpoint on Demand System, are overlaid on the desktop stadium model through the HMD. If the captured scene contains shadows, the shadows of the players and the ball are represented on the stadium model using the relationship between the original stadium and the stadium model.

3. The Viewpoint on Demand System

Once a soccer match is captured, projective geometry between cameras is estimated. Corresponding natural features in the object space can yield the fundamental matrices [3] between the viewpoints of the cameras. Additionally, corresponding natural features on the planar objects provide the homographic matrices [3] between the planes in different views. The soccer scene is then classified into foreground objects, such as a ball and players, and background objects such as a ground, a goal and spectator's seats for performing view interpolation.

In the previous method [4, 8], the virtual view images of the background are synthesized only once because they are considered as static regions. However, if the captured scenes have variations in lighting, it is required to generate the virtual view images in every lighting condition. In our system, we perform view interpolation every 150 frames for the background. In regards to the foreground objects, view interpolation at each frame is executed since the shape or position of the foreground object changes quickly over time.

3.1. View Interpolation of the Background

In order to generate the virtual view images of the background, the image where neither the players nor the ball exists is required for each real camera in each lighting condition. We synthesize such background images by setting the mode value of every image sequence spanning several hundred frames (150 frames in our case) to each pixel. The generated images are then segmented into several plane regions that form the ground, goal and spectator's seats for performing view interpolation. The dense correspondence is obtained between neighboring views by applying the homographic transformation to each plane region. The virtual view image is synthesized using a morphing technique [9]. The shadows of the goal and spectator's seats are projected on the ground in each virtual view image through this process. Figure 2 depicts the background images at the positions of real cameras generated from the image sequence and that of a virtual camera located between the two real cameras.

3.2. View Interpolation of the Foreground

The foreground objects are extracted in each frame from the entire scene by subtracting the background of the corresponding camera and lighting condition. We segment the foreground into shadow regions and the player/ball regions for applying view interpolation respectively. Both the geometric information and the color information are used for this segmentation. It is assumed that the shadow is usu-

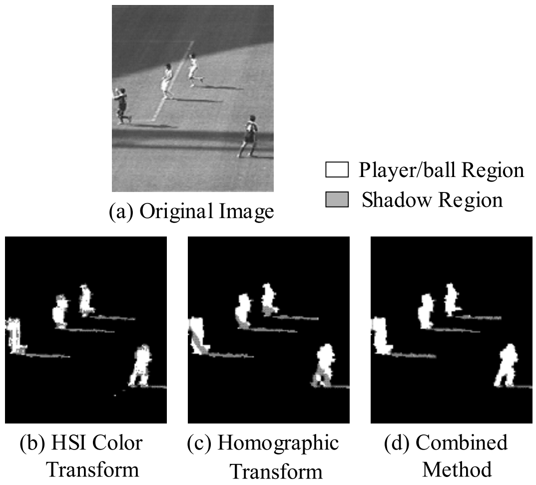

Figure 3: The comparison of segmentation results for foreground objects.

ally projected on the ground in a soccer scene. We detect a candidate for shadow regions by applying the homography of the ground plane to all the extracted dynamic regions in neighboring two view images. This detection based on the homography often includes a part of the player's feet. Therefore, we also use the pixel color for shadow extraction by applying HSI color transform to the candidate in each view image. The hue of the pixel is almost identical in the shadow regions between the current frame image and the background image, while it is different in the player/ball regions. Figure 3 exhibits the segmentation results by comparing the combined method (d), which uses both geometric transform; homography transform and color transform; HSI transform, with the methods using only either of them (b) and (c). It is evident that the combined method is better than the independent methods at segmenting the dynamic regions into shadows and players/ball.

After segmentation, view interpolation is performed on the shadow and player/ball regions independently. The dense correspondence within shadow regions is obtained using the homography of the ground plane. Epipolar geometry is employed for the player/ball regions. The interpolated view images are then synthesized with a morphing technique. Finally, superimposing the virtual view image of the foreground onto the background completes the image of the entire soccer scene from the chosen viewpoint by the user.

4. The MR Presentation System

In the MR Presentation System, a user watches soccer scenes overlaid on a stadium model via a video see-through HMD. In order to present the scene on the desktop stadium

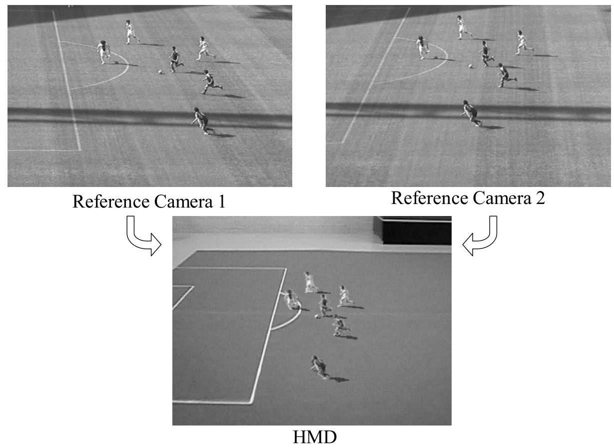

Figure 4: Overlaid shadows onto the desktop stadium model.

model, it is necessary to synthesize the scene from the viewpoint of the HMD. Additionally, the registration between the original stadium and the desktop stadium is required for overlaying the players, the ball and their shadows.

As a soccer field contains some lines, which indicates such as a penalty area and a goal area, we employ these lines for viewpoint determination and the registration. We calculate the viewpoint position from HMD camera image where feature lines are tracked. The virtual view images of the players and the ball are synthesized from reference cameras according to the viewpoint position. Their rendering positions on the desktop stadium are determined using the homography of the ground between the original stadium and the desktop stadium.

When the captured scene in the real stadium has shadow regions, the shadows of players and the ball need to be overlaid on the stadium model. As shadows are projected on the ground plane, a simple homographic transform between the ground plane of the soccer stadium and the desktop stadium model determines the shadow regions on the stadium model. We assume that the color value of the shadow regions become half as much as that of the original stadium model. Since it is obvious that a ball, players and their shadows exist on/over the soccer ground, overlaying them onto the stadium model completes the visualization in MR. Figure 4 shows an example of the represented shadows with the players and the ball on the stadium model from two reference camera images.

5. Experimental Results

We have developed free viewpoint observation systems and applied them to actual soccer scenes having variations in lighting. Soccer matches were captured using four uncalibrated cameras placed on one side of the soccer field. The

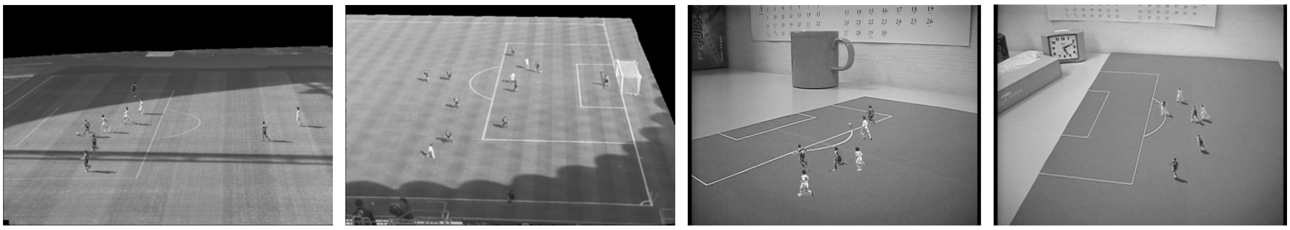

Figure 5: The result of free viewpoint observation of soccer scenes including variations in lighting.
(Left: Viewpoint on Demand System; Right: MR Presentation System.)

projective geometry such as fundamental matrices between the viewpoints of the cameras and homographic matrices between the planes in neighboring views were calculated by manual selection of 50 corresponding feature points in the images.

Figure 5 shows example results of free viewpoint observation of soccer scenes including variations in lighting. The left two images present the results of the Viewpoint on Demand System. The shadows of the stadium and the players are well synthesized in the virtual view images. We obtained a video replay from arbitrary viewpoints where positions of the shadow of the stadium slightly changes according to the lighting condition and shadows of the players and the ball changes according to their movements.

The right two images in Figure 5 present the results of the MR Presentation System. By comparing these two images, we see that the appearance of shadows can produce a more realistic MR presentation. The virtual shadows are naturally overlaid on the desktop stadium for representation of the scene including shadows in the rightmost image.

In the current work, errors sometimes occur in the rendering positions of the shadows. When the segmentation process for the dynamic region into player region and shadow region has mistakes, we see a gap between a foot of the player and the shadow. This is a problem to be solved in the future.

6. Conclusion

We have proposed two kinds of free viewpoint observation systems that can represent soccer scenes including variations in lighting. The method of synthesizing shadows of both the background and the foreground in the dynamic scene was introduced for the Viewpoint on Demand System. Moreover, the method of representing the shadow of foreground objects on the desktop stadium model was described for the MR Presentation System. Visualization of outdoor sporting scenes from arbitrary viewpoints became possible from just multiple images captured by uncalibrated cameras. Projective geometry between cameras is employed in the synthesis and representation of shadows. The proposed system could be applied to other sporting events such as baseball games, football games, and tennis matches.

Acknowledgments

This work is supported in part by a Grant in Aid for the 21st century Center of Excellence for Optical and Electronic Device Technology for Access Network from the Ministry of Education, Culture, Sport, Science, and Technology in Japan. The first author is a JSPS Research Fellow and partly supported by JSPS Research Fellowship for Young Scientists.

References

[1] S. E. Chen and L. Williams, "View Interpolation for Image Synthesis," *Proc. of SIGGRAPH '93*, pp. 279-288, 1993.

[2] Eye Vision,
http://www.ri.cmu.edu/events/sb35/tksuperbowl.html

[3] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.

[4] N. Inamoto and H. Saito, "Fly-Through Viewpoint Video System for Multi-View Soccer Movie Using Viewpoint Interpolation," *Proc. of Visual Communications and Image Processing (VCIP '03)*, SPIE, Vol. 5150, 122, Jul. 2003.

[5] N. Inamoto and H. Saito, "Immersive Observation of Virtualized Soccer Match at Real Stadium Model," *Proc. of International Symposium on Mixed and Augmented Reality (ISMAR '03)*, pp. 188-197, Oct. 2003.

[6] Lucent, http://www.lucent.com

[7] Orad, http://www.orad.co.il

[8] H. Saito, N. Inamoto, and S. Iwase, "Sports Scene Analysis and Visualization from Multiple-View Video," *Proc. of International Conference on Multimedia and Expo (ICME'04)*, June 2004.

[9] S. M. Seitz and C. R. Dyer, "View Morphing," *Proc. of SIGGRAPH '96*, pp. 21-30, 1996.

[10] R. Y. Tsai, "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses," *IEEE Journal of Robotics and Automation*, Vol. RA-3, No. 4, pp. 323-344, Aug. 1987.