Immersive Observation of Virtualized Soccer Match at Real Stadium Model

Naho Inamoto† Hideo Saito †,††

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

††PRESTO, Japan Science and Technology Corporation(JST)

{nahotty, saito}@ozawa.ics.keio.ac.jp

Abstract

This paper presents a novel observation system for immersive soccer match taken by multiple video cameras at a real stadium. The user sees the soccer field model in front of his/her eyes from the viewpoint through head-mounted display, while the images of players and a soccer ball are also rendered onto the display. For geometric registration between the soccer field model in the real world and the dynamic soccer scene in the rendered images, the viewpoint position of the user is computed by using only natural feature lines in the HMD camera image. Since it is difficult to strongly calibrate the HMD camera and the multiple cameras that capture the real soccer scene, we employ projective geometry for the registration and the rendering. For demonstrating the efficacy of the proposed system, video images of soccer matches taken at real stadium are rendered onto the HMD camera images of a tabletop stadium model. This is a completely new challenge to apply augmented reality to a dynamic event in a large-space.

1. Introduction

One of the most exciting ways to enjoy entertainment events and sporting events is to watch on live. Therefore the media reporting such live events takes an important role for entertaining people who can not present at the live events. For enhancing the presence, many kinds of visual entertainment effect can recently be seen on TV broadcasts. One example of the application for such effects is "Eye Vision" system used at Super Bowl XXXV broadcast by CBS. In the Eye Vision system, the multiple video streams are captured with more than 30 cameras. The sequences of video images from different angles are then used to create a three-dimensional visual effect such that the viewpoint turns around the object event at a temporally freezed moment. This system makes use of visual effects of switching video images in order to give the impression of being immersed.

On the other hand, in the field of Virtual Reality (VR) or Mixed Reality (MR), special visual devices like head-mounted displays (HMD) or three-dimensional displays are often employed to realize a viewer to walk-through or fly-through in virtual space or virtual and real mixed space. Using such devices enable the user to have an experience in a virtual world, where CG objects are rendered, more effectively than using normal display or screen as represented by television.

As for dynamic events in a large space, especially sporting events, some approaches have been proposed for producing visual entertainment effects. One approach is reconstruction of a sporting match using CG animation. Data of the actions and positions of players are extracted from video camera images, and then the data derives CG model players. The viewer can watch an animation of the event from preferred viewpoint. In this approach, it is easy to render the object scene from various angles, and moreover, quality of images does neither depend on the number of cameras nor the quality of original video images. However reality of rendered video is not always sufficient.

Another approach is arbitrary view-synthesis by computer vision-based technologies. Arbitrary view of the event can be generated from 3D structure of the scene that is reconstructed via images captured with video cameras. For example, the methods targeting an certain local area, where a few players are exist, have proposed in [14, 15, 25]. These methods enable a viewer to watch a specific player in a large-space field from viewpoints around the player. As 3D models are reconstructed to render novel views, the accuracy of the models affect the quality of the rendered scene. Once the model has been constructed, however, the viewer can flexibly change the viewpoint in theory.

In the meanwhile, we have proposed a view-synthesis method for a soccer match captured by multiple cameras at a stadium [11, 12]. This method is targeting entire soc-

Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '03)
0-7695-2006-5/03 $17.00 © 2003 **IEEE**

cer field, that is, the object area is larger than that in the methods presented in [14, 15, 25]. In our method, view-interpolation between neighboring cameras is employed for adapting to view-synthesis in such a large object scene. The global soccer field is reconstructed from any arbitrary viewpoint between real cameras. The details of the shape information about each player may not be included, but the quality of the synthesized images sufficient for observation the appearance of dynamic event.

In this way, various approaches for producing visual effects have been proposed, however, immersive impression is not enough for entertainments. If these approaches are combined with VR or MR, more enjoyable application can be realized.

Hence, in this paper, we extend our view-synthesis algorithm for soccer scene to Augmented Reality (AR), so that we can introduce a new visualization system for immersive observation of an actual soccer match. The user sees a soccer field model in front of his/her eyes from the viewpoint, while the images of players and a soccer ball are overlaid onto the scene. Thus this system makes it possible to watch a soccer match on the soccer field model in front of the user with favorite viewpoint. The soccer match taken at a stadium is stored as video images. Then the images of the dynamic regions, which represent the event, are extracted from image sequences. Arbitrary views of players and a ball are synthesized by our existing method in accordance with the viewpoint position, and presented to the viewer through head-mounted display (HMD).

In the field of AR, typical approach includes overlaying virtual CG objects onto an images sequence captured by a video camera. In the proposed approach, instead of CG objects, the virtualized real soccer match is overlaid into the real world based on the concepts of image-based rendering.

Prince et al.[19] have proposed a related approach to our method. A system for live capture of 3-D content and simultaneous presentation in augmented reality is introduced. The user can watch the superimposed images of a remote collaborator in the real world, whose action is captured by fifteen cameras surround him/her. The difference from our method is that 3D models of the subjects are reconstructed using shape-from-silhouette in order to render the appropriate view. Furthermore, the subject is captured in the area, which is the volume of 3.3m diameters and 2.5m heights. One advantage of the proposed method is view-synthesis based on projective geometry between multiple cameras instead of reconstructing 3D models with strong camera calibration. Projective geometry of multiple cameras can easily be obtained by images themselves, while strong calibration [24] of multiple cameras is difficult to obtain, which is imperative in Prince 's method. Thus our method can be applied to even a dynamic event at a large-space such as a soccer match. This approach provides a new form of immersive observation system for entertainment or sporting events.

2. View synthesis

2.1. Related works

In general, view synthesis techniques can be categorized into two groups. In the first group, a 3D shape model of an object is reconstructed to generate the desired view [13, 20, 25]. Since the quality of the virtual view image depends on accuracy of the 3D model, a large number of video cameras or range scanners are typically used to reconstruct an accurate model. In the second group, arbitrary view image is synthesized without an explicit 3D model. Instead, image warping such as transfer of correspondences is employed for synthesizing new view images [1, 6, 7, 18, 21]. The view synthesis employed in this paper belongs to the latter group. We describe our algorithm roughly in the next section.

2.2. Intermediate view generation for soccer scenes

We have proposed view-synthesis method for a soccer match in [11, 12]. We give brief explanation of our algorithm here, especially about the method of view interpolation for the dynamic regions. Soccer scene can be classified into dynamic regions and static regions. The former corresponds to players and the ball of which the shape or position change over time. The latter corresponds to the ground, goal, and background in the scene. Though the methods of view interpolation for both regions are explained in [11, 12], we are interested in only dynamic regions to be overlaid

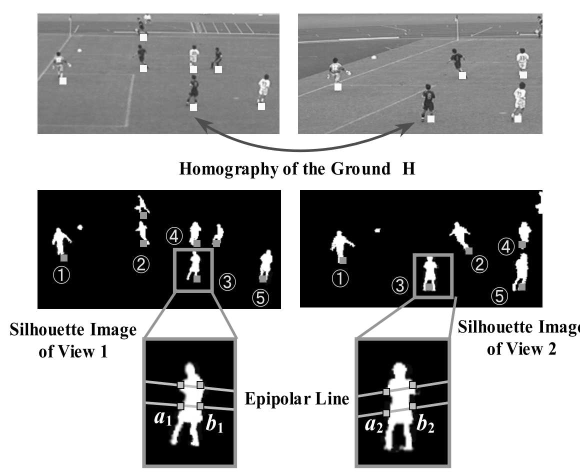


Figure 1. Correspondence for the dynamic regions in soccer scene.

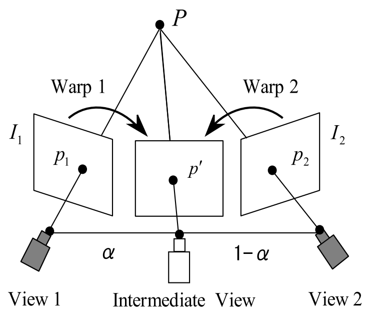


Figure 2. Transfer of the correspondence.

onto the images of the real stadium model.

Hence the view synthesis method for only dynamic regions is described below. A single scene usually contains several players and a ball, so we deal with these objects separately. Firstly, all dynamic regions are extracted by subtracting the background from the image. After the silhouettes have been generated by binarization, every silhouette region is segmented with different label. If occlusion is detected, segmented silhouettes of the previous frame are used for dividing the silhouette regions of the current frame. The silhouette of a player between the different cameras is corresponded by using the homography [10] of the ground plane as shown in Figure 1. This is based on the fact that all feet of the players are attached on the ground. Even when a player is jumping off the ground, the error caused by the jumping is sufficiently small, so the homographic matrix of the plane that represents the ground can still find corresponding silhouettes.

Next, each pair of silhouettes is extracted to obtain the pixel-wise correspondence within the silhouette. Epipolar lines are drawn between images in the different views, view 1 and view 2, by using a fundamental matrix [10]. On each epipolar line, the points at the intersections with boundaries, such as a_1 and a_2, b_1 and b_2 of Figure 1, are corresponded. The correspondences between the pixels inside the silhouette are obtained by linear interpolation of the points at the intersection. After a dense correspondence for the whole silhouette is obtained, the pixel positions and values are transferred from the source images of view 1 and view 2 to the destination image by image morphing [4] as described by the following equations.

$$\acute{\boldsymbol{p}} = (1-\alpha)\{(\boldsymbol{p}_1-\boldsymbol{c}_1)\frac{\acute{f}}{f_1}+\boldsymbol{c}_1\}+\alpha\{(\boldsymbol{p}_2-\boldsymbol{c}_2)\frac{\acute{f}}{f_2}+\boldsymbol{c}_2\} \quad (1)$$

$$I(\acute{\boldsymbol{p}}) = (1-\alpha)I(\boldsymbol{p}_1) + \alpha I(\boldsymbol{p}_2) \quad (2)$$

where $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ are the coordinates of the matching points in images I_1 and I_2. $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ are the coordinates of the principal points in images I_1 and I_2 as well. f_1 and f_2 are the focal lengths of camera 1 and 2. $\acute{f}$ is the focal length of the virtual camera. $I(\boldsymbol{p}_1)$ and $I(\boldsymbol{p}_2)$ are the value of the matching points in images I_1 and I_2. $\acute{\boldsymbol{p}}$ is the interpolated coordinates and $I(\acute{\boldsymbol{p}})$ is the interpolated value. α defines the interpolating weights given to the respective actual viewpoints.

All correspondences are used in the transfer to generate a warped image. Here two transfers are required, one from view 1 and the other from view 2 as shown in Figure 2. Two generated warped images are then blended to complete the image of the virtual view. If the color of a pixel is different in the two images, the corresponding pixel in the virtual view is rendered with the average of the colors; otherwise the rendered color is taken from either actual image. Equation (1) makes it possible to change the zoom by the ratio of the focal length of the real camera to the focal length of the virtual camera. The above algorithm is applied to every pair of silhouettes and synthesized in order of distance from the viewpoint completes view interpolation for dynamic regions of soccer scenes.

3. Immersive observation of virtualized soccer match

3.1. Overview

This section explains the algorithm for immersive observation of a virtualized soccer match at a real stadium model. We extend the view-synthesis method introduced Sec. 2.2 to the field of AR and enable an user to watch a sporting event on the field model in front of him/her. The user sees the real world via video-see-through HMD and observes dynamic objects of soccer scenes overlaid onto the real stadium model. Figure 3 describes an overview of the proposed method. The procedure of the proposed method is as follows.

1. Capture a soccer match at a stadium by uncalibrated multiple cameras.
2. Estimate projective geometry between neighboring cameras.
3. Extract profile about the soccer scenes of image sequences.
4. Capture a soccer field model by uncalibrated HMD camera.
5. Detect feature lines (at least 4 lines) in the HMD camera image.
6. Calculate the parameters for deciding the position of the HMD camera.
7. Generate virtual view of the dynamic objects based on the parameters.
8. Render the dynamic objects onto the HMD.

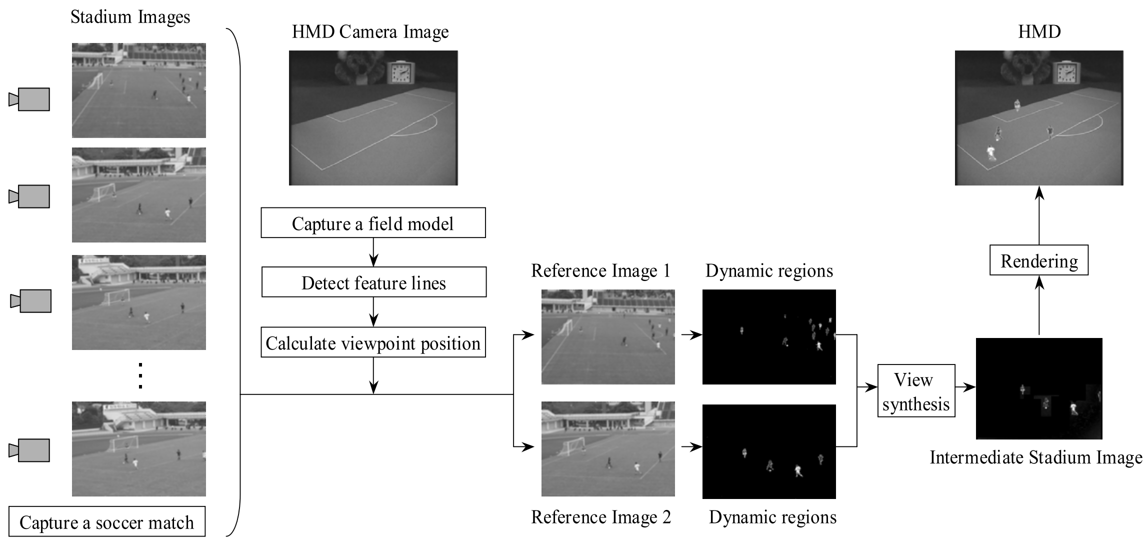


Figure 3. Overview of the proposed method.

The operations 1∼3 are performed once as preprocessing. The other operations 4∼8 are implemented at every frame of image-sequence online. In distinction from HMD camera images, we refer to captured images at a stadium as stadium images, and interpolated view images between neighboring cameras as intermediate stadium images. In operation 7, virtual view of the dynamic objects is generated in the following order. An intermediate stadium image is synthesized from two original stadium images based on the position of the viewpoint in the way explained in Sec. 2.2. A virtual view image is then modified from the intermediate stadium image in consideration of geometric registration. Lastly, virtual view of the dynamic objects is rendered onto the HMD.

3.2. Geometric registration

One of the most important issues for AR is geometric registration between the real and the virtual world. This generates a correct view of a virtual object, and overlays it onto a view of the real world. In order to achieve geometric registration, positioning sensors and/or captured AR images are generally used. The positioning sensors are used for measuring the pose and position of the virtual camera (correspond user's view) to the object scene, so that the relationship between the virtual world, the real world, and the viewpoint can be obtained [2]. For such registration, strong camera calibration of the AR camera is required. The method using positioning sensors is stable against the change of light condition, however, the viewpoint can move only within a limited area and the system is expensive because of the necessity of special devices. On the other hand, in case of using only captured AR images, registration is performed based on the relationship with two or more basis images. Typically artificial markers placed in the real world are used [16, 22] or natural feature points are used [8] for the registration. Recently the methods using both of sensor and the AR camera images are also proposed for accurate estimation of the pose and position of the camera [3, 17, 23]. In such method, strong camera calibration is necessary as well and some artificial markers are also required.

In the proposed method, we employ only captured AR images for the geometric registration between a soccer field model in the real world and the dynamic objects of the soccer scene. The cameras for capturing the real world are attached to the HMD, so we refer to AR camera image as HMD camera image below.

In the related work [19], it is imperative to calibrate both HMD camera and multi-camera which capture the subject for constructing 3D models. On the contrary, the proposed method does not need the strong camera calibration. Even though the HMD camera can be calibrated, the view-synthesis method explained in Sec. 2.2 is based on the concepts of using uncalibrated multiple cameras. Our target is a dynamic event in a large-space like a soccer match. Since the movements of each player are complex, it's almost impossible to reconstruct an accurate 3D model of the scene. Furthermore, the strong camera calibration is very difficult in a real stadium. In this method, therefore, we use the projective geometry for the registration between real images taken with the HMD and virtualized soccer images synthesized by our method.

For the registration, we need to generate the virtualized soccer image at the same viewpoint of the HMD, and over-

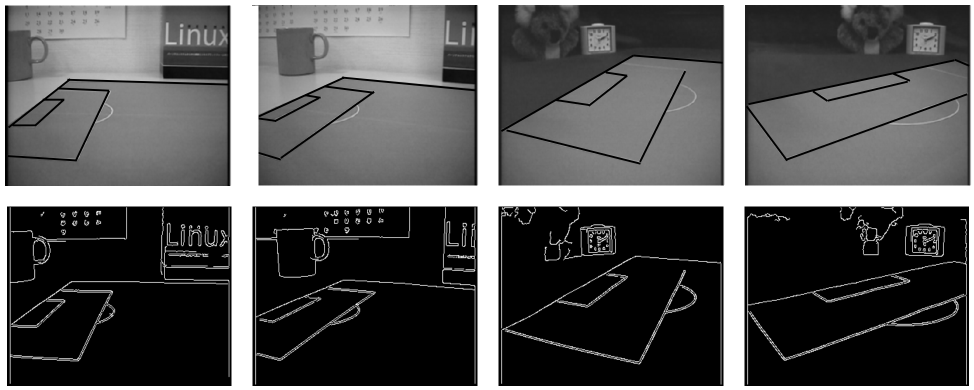


Figure 4. Detected natural feature lines in HMD camera images (top) and edge images (bottom).

lay the image onto the HMD camera image. In our view-synthesis algorithm, the position of the viewpoint of the virtualized soccer image is specified by three elements, which are (1) selection of neighboring two reference cameras, (2) interpolating weight value between two reference cameras, and (3) zoom ratio between the real camera and the virtual camera. Therefore, we need to determine those elements from the HMD camera image, so that the HMD viewpoint can be same as the virtualized soccer image viewpoint. In addition to those elements, one more element is also required for geometric registration, which is homography represents transformation between the soccer ground plane in the real soccer match scene, and the ground plane of the real stadium model taken in the HMD camera image. This homography determines the positions of the players and the ball in the virtualized soccer image that is overlaid onto the HMD camera image.

(a) Feature line detection

As the image of soccer field model captured by the HMD cameras contains natural feature lines, which are easy to track such as the lines of the penalty area or the goal area, we employ these natural feature lines instead of using any artificial markers. Therefore the efforts for locating artificial markers can be reduced. We use Hough transform for line detection. The Canny operator [5] is first applied for edge detection and all edge points mapped into the Hough space. The strong peaks which form the lines of penalty area and goal area are then found in the space. The results of line detection are shown in Figure 4 (top) with the edge images (bottom). All elements for specifying virtual viewpoint position can be determined based on those natural feature lines, which must be tracked in the HMD camera image at every frame, as described in the follows.

(b) Viewpoint position calculation

We apply a vanishing point for selection of neighboring reference cameras and interpolating weight between the cameras. Vanishing point is the point to which the extensions of parallel lines appear to converge in a perspective projection. The vanishing point geometry is often used to assess the orientation of the photographic image. In the proposed method, angle of the user's view are estimated by the position of the vanishing point. Two cameras whose angles are close to the user's view angle are selected as the reference cameras.

In advance, locations of vanishing points in all stadium images are measured by extending lines of the goal area and the penalty area. Whenever HMD camera image is captured, the feature lines are detected in the image and the location is measured in the same way (See Figure 5). Just a horizontal component of the location of a vanishing point is used for calculation because we assume that a user moves the viewpoint almost horizontally from side to side. We also assume that all cameras capturing a soccer match at a stadium are placed at the almost same height. According to such assumptions, we select two stadium images in which the location of the vanishing point is closest to the vanishing point in the HMD camera image as reference camera images. Then relative distance between the vanishing points of reference cameras and that of the HMD camera determines interpolating weight w as the following equation,

$$w = \frac{x_{hmd} - x_{stL}}{x_{stR} - x_{stL}} \tag{3}$$

where $V_{stL}(x_{stL}, y_{stL})$ and $V_{stR}(x_{stR}, y_{stR})$ are the vanishing points in two reference camera images, and also $V_{hmd}(x_{hmd}, y_{hmd})$ is the vanishing point in HMD camera image.

(c) Zoom ratio correction

It is assumed that a user observes a virtualized soccer match near the location of stadium cameras, but the size of soccer field is sometimes different between in HMD camera

Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '03)
0-7695-2006-5/03 $17.00 © 2003 **IEEE**

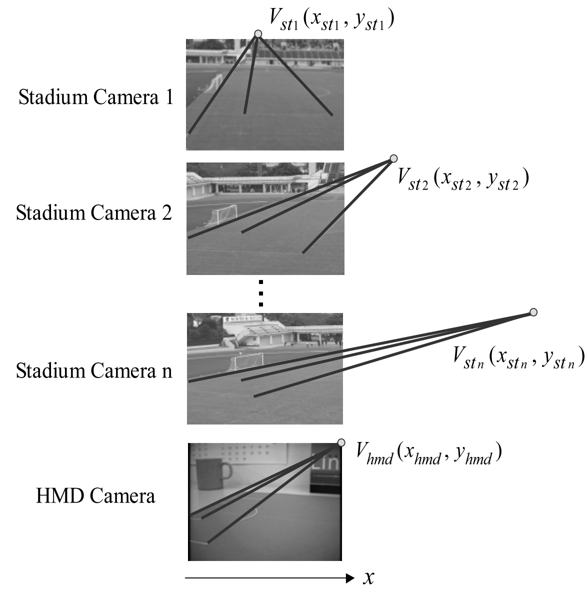


Figure 5. Location of the vanishing points.

images and in stadium camera images. Then the size of virtual view of players and a ball is modified by zoom ratio. The focal length of the stadium camera f_{st} and the HMD camera f_{hmd} decide the zoom ratio as

$$z = f_{hmd}/f_{st} \ . \tag{4}$$

As uncalibrated cameras are used in our approach, however, the intrinsic parameters of the cameras are unknown. Therefore the focal length is computed with two vanishing points $V_1(x_{v1}, y_{v1})$ and $V_2(x_{v2}, y_{v2})$ by the following equation,

$$x_{v1}x_{v2} + y_{v1}y_{v2} + f^2 = 0 \ . \tag{5}$$

Here it is supposed the skew of the camera is 0, aspect ratio is 1, and principal point is the center of the image. The detailed explanation is found in [9].

(d) Homography transformation

We use homography transformation for the registration between the HMD camera images and the virtualized soccer images. In order to generate natural views of a virtualized soccer match, the dynamic objects need to be rendered correctly onto the real field model. Homographic matrix determines the positions of the players and the ball on the HMD camera images.

The homography represents transformation between the soccer ground plane in the intermediate stadium image and the plane of the real stadium model captured in 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. The feet position of every player in a virtual view is determined by the following equation.

$$\boldsymbol{p}_{hmd} \cong \boldsymbol{H}\boldsymbol{p}_{st} \tag{6}$$

where $\boldsymbol{H}$ is the homographic matrix that represents the transformation between the planes, and p_{st}, p_{hmd} are homogenous coordinates of the feet position in the intermediate stadium image, in the HMD camera image accordingly.

This is the same theory as well as segmented players are corresponded between neighboring views in Sec 2.2. All feet of the players are attached on the ground so that the homographic matrix of the ground plane can replace the position of the player.

3.3. Rendering

After intermediate stadium image is synthesized from the stadium images with three parameters reference cameras, interpolating weight and zoom ratio, dynamic objects are rendered into the real world by applying homography. Since it is obvious that players and a ball exist on/over a soccer field, it's not necessary to take into account the problem of occlusion in the rendering process.

If dynamic objects are occluded each other, the relationship have already been determined during the process of view-synthesis by considering the location of the feet of each player. Hence overlaying the dynamic objects onto the AR images completes the destination view image.

4. Experimental results

We implemented the immersive observation system based using only a PC (OS: Red Hat Linux 7.3, CPU: Intel Pentium IV 2.8 GHz) and a Canon Video-See-Through HMD VH-2002 (with front-mounted cameras captures images at a resolution of 640×480) (See Figure 6). Also we used MR Platform SDK library and OpenGL library for capturing a real world and rendering virtual objects.

The process of implementation is described below. Firstly, scenes of an actual soccer match were taken by multiple uncalibrated video cameras at two kinds of soccer stadiums. One is Oita stadium in Oita city, which is one of the stadiums the 2002 FIFA World Cup was held, and the other is Edogawa athletics stadium in Tokyo, Japan. As Figure 7 shows, 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 720×480 pixels, 24-bit-RGB color images. Secondly, fundamental matrices between the viewpoints of the cameras and homographic matrices between the ground planes in neighboring views were computed from images themselves by manually selection of

Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '03)
0-7695-2006-5/03 $17.00 © 2003 **IEEE**

Figure 6. Canon video-see-through HMD.

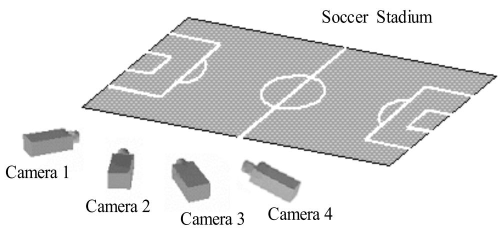


Figure 7. Camera configuration at the soccer stadium.

20 corresponding feature points in the images. Next, the position of a vanishing point was measured in each viewpoint image with the lines of the goal/penalty area. Moreover, profiles of the dynamic regions about each frame are extracted for online process. After every player region was segmented and labeled, the regions of the same player in the neighboring view images were corresponded by using homography of the ground plane between the views. The above process is preprocessing.

At the beginning of the online process, an HMD camera images was captured by uncalibrated HMD camera. The lines of the goal area and the penalty area were detected as natural feature lines in HMD camera image. Then, selection of two reference cameras, interpolating weight, and zoom ratio were determined by the method described in the previous section. An intermediate stadium image was synthesized based on the viewpoint defined above three parameters. Next, homography of the field plane between intermediate stadium image and the HMD camera image was calculated and virtual view modified with the homography. Finally the virtual view of dynamic objects was rendered onto the HMD. Online process was iterated until viewer stopped observation of the soccer match.

Figure 8 shows the results when the proposed method was applied to the soccer match captured at Oita stadium. The first and the second row are reference stadium images used for virtual view generation. The third row is intermediate stadium image synthesized from the two reference images by previously proposed method. The last low is HMD image modified from the third-row image by the proposed method in this paper. The user sees this image through HMD. Interpolating weight and zoom ratio is indicated as w and z respectively. For example, the image on the top of the last low was generated based on the parameters that interpolating ratio is 0.49 between camera 1 and camera 2, and zoom ratio is 0.89.

Figure 9 presents the close-up view of the HMD images of frame 255 and 266. We see that the dynamic objects of soccer scene were overlaid correctly into the real world. The rendered scenes look so natural that user does not feel any discomfort. However, the way to decide the viewpoint positions and zoom ratios is not stable enough. Therefore appearence of the objects sometimes have a small error. We are currently investigating for improving the method of determination of viewpoint position and zoom ratio stably.

Figure 10 shows the results when the proposed method was applied to another soccer match captured at Edogawa athletic stadium. The order of the images are the same as in the Figure 8. In the images of the third row, soccer match is reconstructed with background and soccer ground based on captured images like another camera view. On the other hand, in the images of the last row, virtualized soccer match is reconstructed over/on the field model in front of the viewer.

Since our method is based on view-interpolation between neighboring cameras, it is only possible to provide convincing renderings from a range of viewpoints limited by the placement of the cameras used for capturing. If we increase the number of cameras surrounding the soccer field, the range of viewpoints can be extended. However, we ensure that the quality of generated images does not depend on the number of the cameras, but only depends on the geometrical relationship between two neighboring cameras. This is because that the intermediate view image is rendered by using only two neighboring cameras. If the appearence of the objects is not much different between two camera views, our approach can synthesize good quality image of virtualized soccer match successfully.

5. Conclusion

This paper has presented an immersive observation system for a soccer match captured by multiple video cameras at a stadium. View-synthesis algorithm for a dynamic event in a large space was extended to the field of AR and it became possible to give a user stronger impression to be immersed in exciting scenes while watching the event from the favorite viewpoint. The proposed method uses only natural feature lines in HMD camera images for calculation of the viewpoint positions and geometric registration, without any artificial markers or sensor devices. Thus the strong calibration of the HMD camera and multiple cameras that capture the subject is not necessary. Therefore it can be applied to

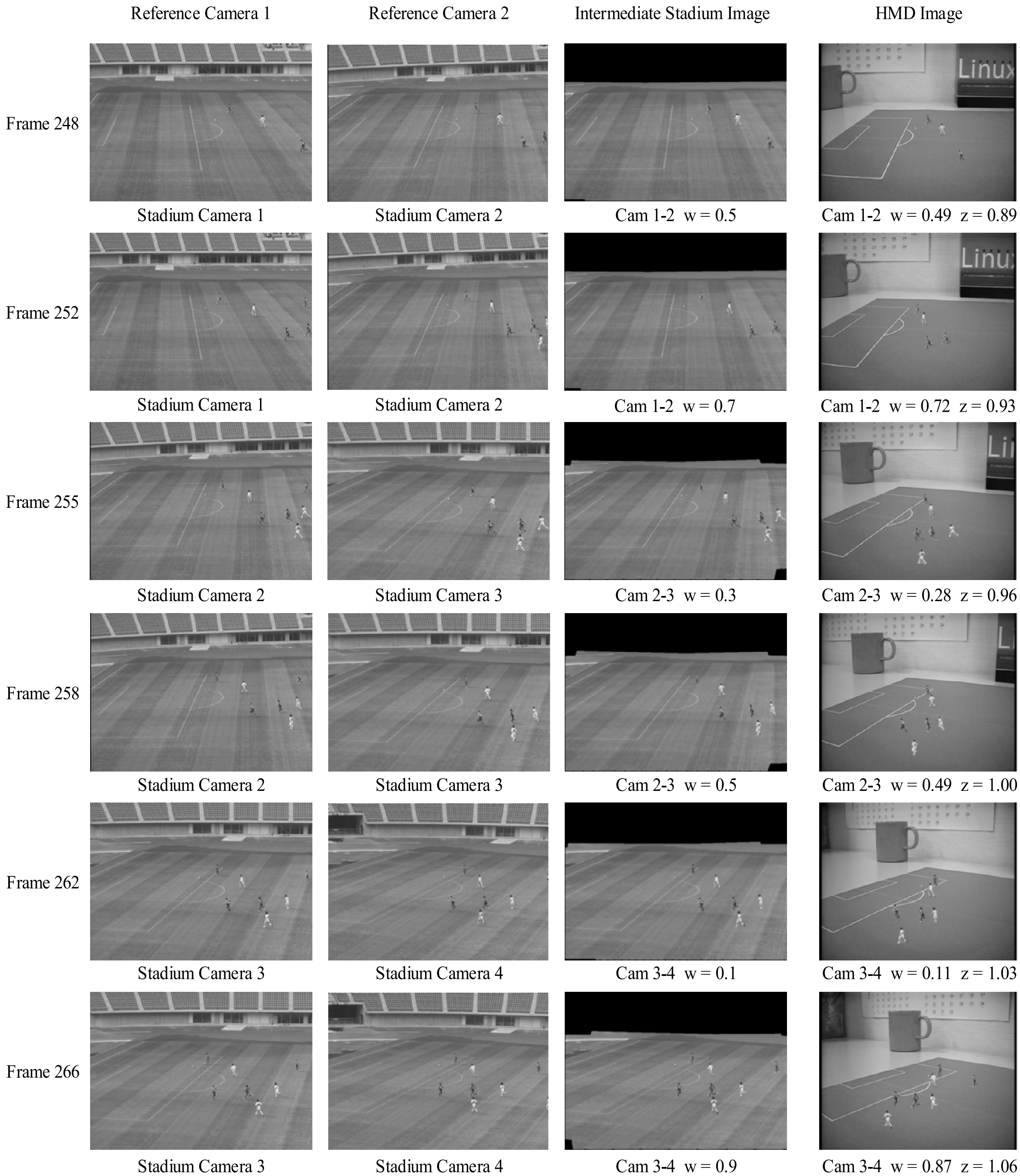
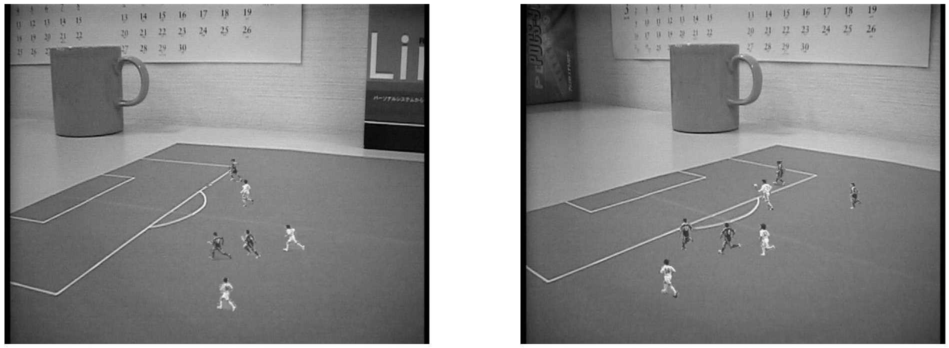


Figure 8. Results of virtualized soccer match taken at Oita stadium.

Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR ’03)
0-7695-2006-5/03 $17.00 © 2003 **IEEE**



Close-up View of Frame 255 | Close-up View of Frame 266

Figure 9. Close-up view of the HMD images.

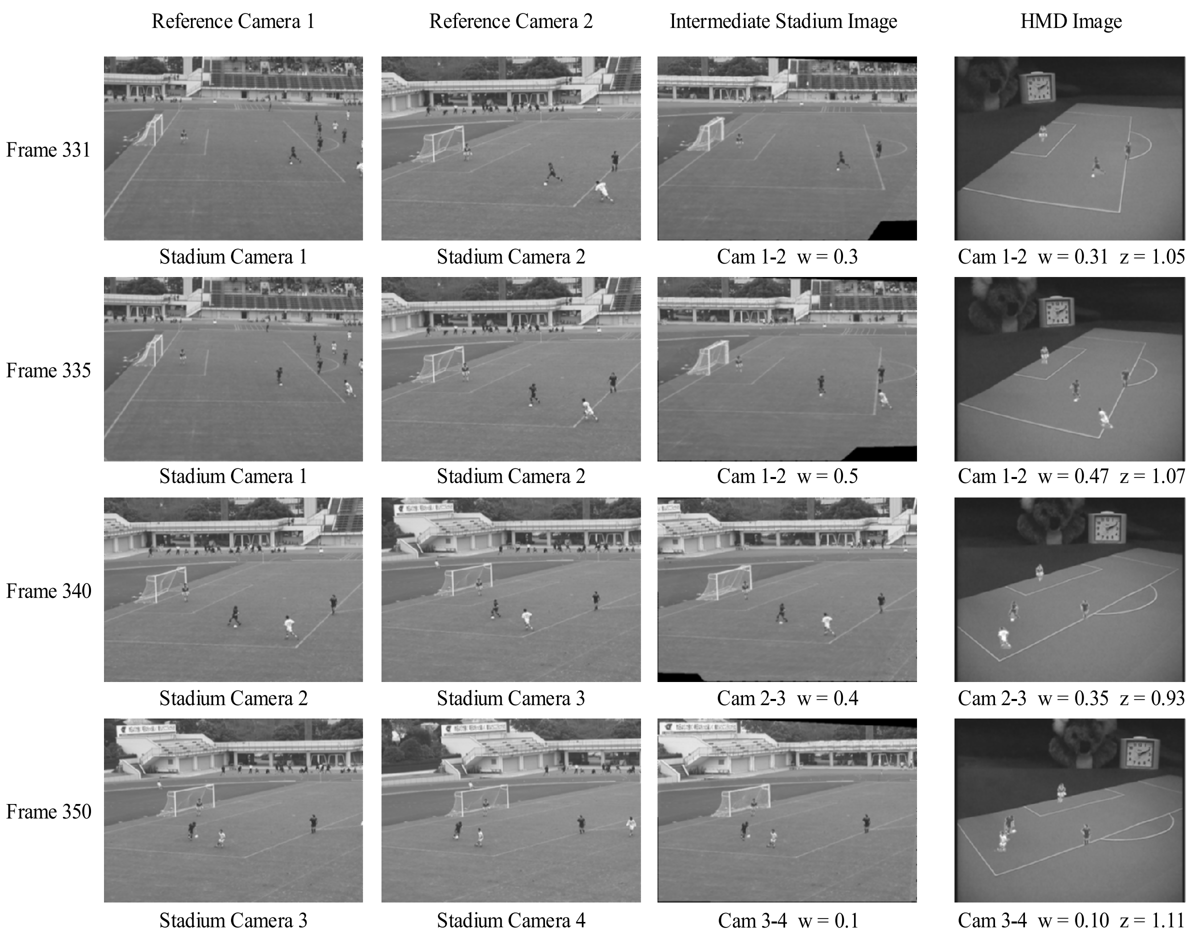


Figure 10. Results of virtualized soccer match taken at Edogawa athletic stadium.

observations not only on the tabletop stadium model in front of the user, but also at an empty real stadium. For example, audience will be able to watch a World Cup game held in a foreign country at a domestic stadium via HMD, that is, the images of the virtualized World Cup Game is overlaid onto the soccer field of domestic stadium where the audience exists. Our system suggests a new way for enjoying entertainment events or sporting events wherever they want. We believe the proposed system in this paper must be a prototype of such attempt.

References

[1] S. Avidan and A. Shashua. Novel view synthesis by cascading trilinear tensors. *IEEE Trans. on Visualization and Computer Graphics*, 4(4):293–306, 1998.

[2] M. Bajura, H. Fuchs, and R. Ohbuchi. Merging virtual objects with the real world: Seeing utlrasound. *Commun of the ACM*, 36(7):52–62, 1993.

[3] M. Bajura and U. Neumann. Dynamic registration correction in video-based augumented reality system. *IEEE Computer Graphics and Applications*, 15(5):52–60, 1995.

[4] T. Beier and S. Neely. Feature-based image metamorphosis. *Proc. of SIGGRAPH '92*, pages 35–42, 1992.

[5] J. Canny. Computational approach to edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[6] S. E. Chen and L. Williams. View interpolation for image synthesis. *Proc. of SIGGRAPH '93*, pages 279–288, 1993.

[7] O. Faugeras and S. Laveau. Representing three-dimensional data as a collection of images and fundamental matrices for image synthesis. *Proc. of International Conference on Pattern Recognition (ICPR1994)*, pages 689–691.

[8] V. Ferrari, T. Tuytelaars, and L. V. Bool. Markerless augumented reality with a real-time affine region tracker. *Proc. of the IEEE and ACM Intl. Symposium on Augmented Reality*, pages 87–96, 2001.

[9] A. Z. Gilles Simon, Andrew W Fitzgibbob. Markerless tracking using planar structures in the scene. *Proc. of the International Symposium on Augmented Reality*, pages 120–128, Oct 2000.

[10] R. Hartley and A. Zisserman. Multiple view geometry in computer vision. *Cambridge University Press*, 2000.

[11] N. Inamoto and H. Saito. Fly through view video generation of soccer scene. *International Workshop on Entertainment Computing (IWEC2002) Workshop Note*, pages 94–101, May 2002.

[12] N. Inamoto and H. Saito. Intermediate view generation of soccer scene from multiple videos. *Proc. of International Conference on Pattern Recognition (ICPR2002)*, 2:713–716, August 2002.

[13] T. Kanade, P. J. Narayanan, and P. W. Rander. Virtualised reality:concepts and early results. *Proc. of IEEE Workshop on Representation of Visual Scenes*, pages 69–76, 1995.

[14] I. Kitahara and Y. Ohta. Scalable 3d representation for 3d video display in a large-scale space. *Proc. of the IEEE Virtual Reality 2003*, pages 45–52.

[15] I. Kitahara, Y. Ohta, H. Saito, S. Akimichi, T. Ono, and T. Kanade. Recording multiple videos in a large-scale space for large-scale virtualized reality. *Proc. of International Display Workshops (AD/IDW'01)s*, pages 1377–1380, 2001.

[16] K. N. Kutulakos and J. Vallino. Affine object representations for calibration-free augmented reality. *Proc. IEEE Virtual Reality Ann. Int. Symp.(VRAIS'96)*, 1996.

[17] U. Neumann, S. You, J. Hu, B. Jiang, and J. W. Lee. Augmented virtual environments (ave): Dynamic fusion of imagery and 3d models. *Proc. of the IEEE Virtual Reality 2003*, pages 61–67.

[18] S. Pollard, M. Pilu, S. Hayes, and A. Lorusso. View synthesis by trinocular edge matching and transfer. *Image and Vision Computing*, 18:749–757, 2000.

[19] S. Prince, A. D. Cheok, F. Farbiz, T. Williamson, N. Johnson, M. Billinghurst, and H. kato. 3d live: Real time captured content formixed reality. *Proc. of the International Symposium on Mixed and Augmented Reality (ISMAR'02)*, pages 7–13, September 2002.

[20] H. Saito, S. Baba, M. Kimura, S. Vedula, and T. Kanade. Appearance-based virtual view generation of temporally-varying events from multi-camera images in 3d room. *Proc. of the Second International Conference on 3-D Imaging and Modeling (3DIM99)*, pages 516–525, 1999.

[21] S. M. Seitz and C. R. Dyer. View morphing. *Proc. of SIGGRAPH '96*, pages 21–30, 1996.

[22] Y. Seo and K. Hong. Calibration-free augumented reality in perspective. *IEEE Trans. on Visualization and Computer Graphics*, 6(4):346–359, 2000.

[23] A. State, G. Hirota, D. Chen, W. Garrett, and M. Livingston. Superior augmented reality registration by integrating landmark tracking and magnetic tracking. *Proc. SIGGRAPH'96*, pages 429–438, 1996.

[24] 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*, RA-3(4):323–344, August 1987.

[25] S. Yaguchi and H. Saito. Arbitrary view image generation from multiple silhouette images in projective grid space. *Proc. of SPIE Vol.4309 (Videometrics and Optical Methods for 3D Shape Measurement)*, pages 294–304, 2001.