Arbitrary Viewpoint Observation for Soccer Match Video

N. Inamoto and H. Saito

Keio University, Japan

Abstract This paper presents a new media production for sports event like a soccer match. Multiple cameras capture a soccer match at stadium. The soccer match can be observed at arbitrary viewpoints selected by each viewer. Video at the arbitrary viewpoints is synthesized from the videos captured with the actual cameras by image-based view interpolation technique. Our view interpolation method does not need full calibration of the multiple cameras, but only need projective geometry information such as fundamental matrices among the multiple cameras, which is easily measured even in a large space in an actual stadium. In this paper, two visualization systems are introduced for arbitrary viewpoint observation of soccer matches from multiple video images. The first is Viewpoint on Demand System that enables viewers to select their own viewpoints during observation through GUI. Photorealistic soccer scene, including players, ball and soccer field, is rendered at arbitrary viewpoints by view interpolation method from actual camera positions. The second is observation system via a video see-through HMD. Viewers see a desktop stadium model in front of their eyes through the HMD, while images of players and ball are overlaid on the display. This system enables viewers to virtually fly through over the soccer field model in the real world. The proposed systems lead to make a new type of immersive media for entertaining such events.

1. INTRODUCTION

Development of information and communications technology has enabled us to enjoy watching entertaining or sporting events taken place all over the world. Many kinds of visual entertainment effects can be seen in TV broadcasting or movies recently. For example, in "Eye Vision" system at Super Bowl XXXV, movie "The Matrix" series, viewpoint rotation around the object provides attractive visual effects. However, such viewpoint motion can only be controlled by the producer regardless of viewer's preferences. If the viewers can arbitrarily select their own viewpoints, the contents can be more enjoyable. This motivates us to develop interactive media production that can satisfy each viewer's demand.

If objective dynamic events are captured with multiple cameras, it becomes possible to produce any desired views of the events independent from the actual camera positions. Thus techniques for arbitrary view generation have an important role for such applications.

As for dynamic events in a large space, especially sporting events, some approaches have been proposed for arbitrary view generation. One approach is reconstruction of a sporting match using CG animation (7, 8). Data of the actions and positions of players are extracted from video images, and then the data derive CG model players. The viewer can watch an animation of the event from favorite viewpoint. In this approach, it is easy to render the object scene from various angles, and moreover, the quality of images does neither depend on the number of cameras nor the quality of the original video images. However, reality of the rendered video is not always sufficient.

Another approach is arbitrary view-synthesis with computer vision-based technologies. Arbitrary views of the event can be generated from 3D structure of the scene that is reconstructed via images captured by video cameras. For example, the methods for arbitrary view synthesis of a few players within an certain local area have been 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 affects the quality of the rendered scene. Once the model has been constructed, the viewer can flexibly change the viewpoint in theory. As the target area is larger, however, these methods have more difficulties in constructing accurate 3D models.

Accordingly, we have proposed view-synthesis method for entire dynamic events in a large-space. Instead of constructing 3D models, view interpolation based on projective geometry between actual cameras synthesizes novel views of the whole dynamic scene. Especially the technique for soccer match reconstruction from multiple video images captured at stadium is introduced in (11, 12). The soccer scene is classified into three regions, the dynamic regions, field regions and background region. Then appropriate projective transform is applied to each region for generating intermediate views. By superimposing every intermediate view image, the global

Printed and published by the IEE, Michael Faraday House, Six Hills Way, Stevenage, Herts SG1 2AY, UK

appearance of the soccer scene can be synthesized. Since projective geometry is easily obtained from images themselves by corresponding feature points, this method can be easily adapted to large targets such as events in a soccer stadium.

In this paper, we firstly introduce "Viewpoint on Demand System", which enables users to select their favorite viewpoints during observation through GUI. Both the dynamic regions of soccer scene, such as players and a ball, and the soccer stadium are presented according to our view-synthesis technique. Virtual view image at the selected viewpoint is generated by view interpolation between actual cameras near the viewpoint at each frame. The synthesized videos are photorealistic as if they are actual camera images from the viewpoint.

Secondly we extend our view-synthesis method to augmented reality (AR) for immersive visualization of soccer matches. 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. This system makes it possible to watch the soccer match on the desktop stadium model in front of the user with favorite viewpoints. Natural feature lines in the image captured by the camera attached to the HMD are used for geometric registration between the stadium model in the real world and the overlaid soccer scene from video images.

Both systems give a free choice of the viewpoint to users for observation of dynamic events in a large space. The proposed approach creates a new scheme of visual media that realize immersive observation of entertainments or sporting events.

2. RELATED WORKS

The methods for synthesizing arbitrary view images from a number of real camera images have been studied since the 1990s in the field of computer vision (13, 18). These techniques, called Image Based Rendering (IBR), can be categorized into two groups, Model Based Approach and Transfer Based Approach.

Model Based Approach constructs a 3D shape model of an object to generate the desired view (6, 20, 24). Since the quality of the virtual view image depends on accuracy of the 3D model, a large number of video cameras surrounding the object or range scanners are used for construction of an accurate model. Also strong camera calibration (23), which is carried out to relate 2D coordinates in images to 3D coordinates in object space, is usually required. As 3D positions of several points in the object space must be measured, this calibration becomes difficult especially in a large space. For such reasons, the object area is generally limited within a few cubic meters in this approach.

On the other hand, Transfer Based Approach synthesizes arbitrary view images without an explicit 3D model (2, 4, 19). Instead, image warping such as transfer of correspondence is employed for synthesizing new view images. The dense correspondence between the original images, which is required for view-synthesis, is often obtained manually or by the use of optical-flow, so almost all targets are static images or slightly varying images such as facial expression.

Thus we have proposed view-synthesis method targeting dynamic events in a large-space such as a soccer match captured at stadium, which is included in Transfer Based Approach.

In the field of AR, typical approach includes overlaying virtual CG objects onto images sequences captured by video camera (1). In the proposed approach, instead of CG objects, the soccer match, which is reconstructed by multiple cameras, is overlaid onto the real world based on the concepts of IBR.

Prince et al. (17) 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 15 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. The subject is captured in the limited 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 between cameras can easily be obtained by just images, while strong calibration of multiple cameras is difficult to obtain, which is imperative in Prince 's method. Therefore our method can be applied to even a dynamic event at a large-space such as a soccer match. The proposed method is completely new challenge to apply AR to large-scale dynamic events.

3. OVERVIEW

Fig. 1 shows overview of the proposed method. The top row of Fig.1 indicates the flow of the observation system with a normal video display, while the bottom row indicates the system with a video see-through HMD. The soccer matches taken at stadium by multiple cameras are stored as video images. Then the projective geometry used for view-synthesis is estimated between neighboring cameras. The proposed systems employ fundamental matrices between the viewpoints of the cameras, and ho-

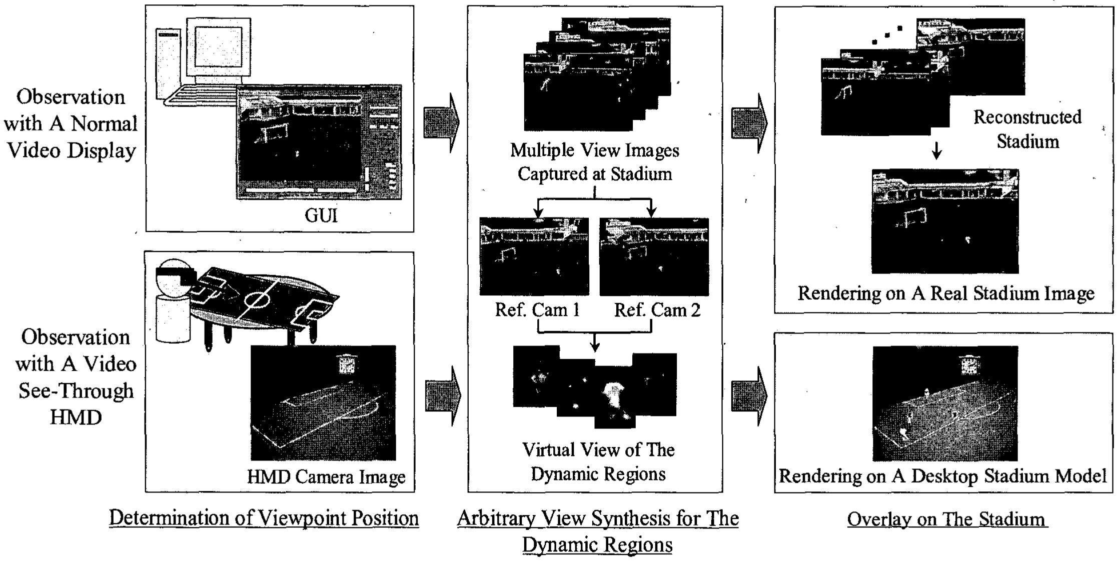


Fig. 1 Overview of the proposed method.

mographic matrices for the planes of the ground, soccer goal and background among neighboring views.

The process consists of three stages in both of the systems, determination of viewpoint position, arbitrary view synthesis 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 camera 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.

4. ARBITRARY VIEW SYNTHESIS FOR THE DYNAMIC REGIONS

In this section, we explain the algorithm of arbitrary view synthesis for the dynamic regions of soccer scene. As the shapes or positions change over time, so view interpolation between selected neighboring cameras generates virtual view image at each frame. A single scene usually contains several dynamic objects such as players and a ball, so we deal with these objects separately. Firstly, all dynamic regions are extracted by subtracting the background image from the current frame image. If the background image, which includes neither players nor ball, can not be captured, it can be made by setting mode value of image sequence to each pixel. After the silhouettes have been generated by binarization, every silhouette region is segmented with different label. The silhouette of a player between neighboring cameras is corresponded by using homography (10) of the ground plane as shown in Fig. 2. This is based on the fact that all feet of the players usually contact with the ground. If a player is occluded by other players, however, the above algorithm may not work well. In this case, segmented silhouettes of the previous frame are used for dividing and corresponding the silhouette regions of the current frame. As Fig. 3 shows, foot position of the occluded player in the current frame is calculated with homography of the ground plane, and then bounding box(surrounding rectangle for each player) is projected from the previous frame. In this way, silhouette is corresponded even when players occlude each other.

Next, each pair of silhouettes is extracted to obtain the pixel-wise correspondence within the silhouette. Epipolar lines are drawn between images in neighboring 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 Fig. 2, are corresponded. The correspondences between the pixels inside the silhouette are obtained by linear interpolation of the intersection points. 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,

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

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

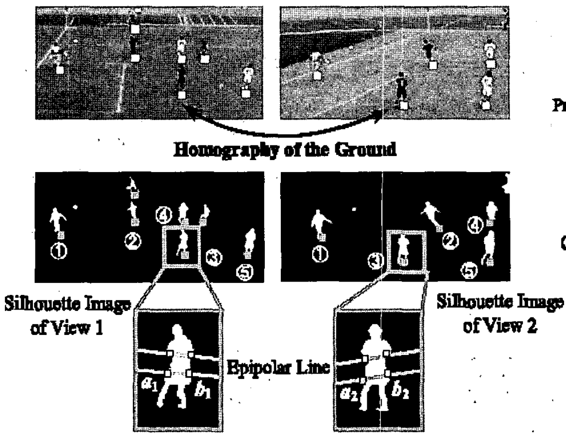


Fig. 2 Correspondence for the dynamic regions.

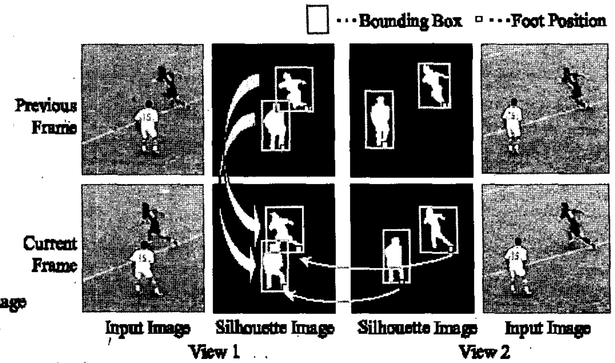


Fig. 3 Correspondence in case of occlusion.

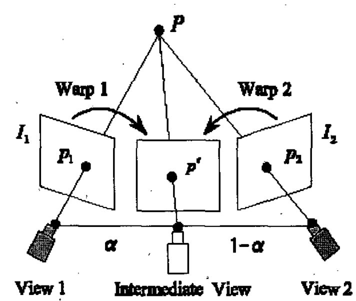


Fig. 4 Transfer of the correspondence.

where p_1 and p_2 are the coordinates of the matching points in images I_1 and I_2. c_1 and 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. $I(p_1)$ and $I(p_2)$ are the value of p_1 and p_2. $\acute{p}$ is the interpolated coordinates and $I(\acute{p})$ 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 Fig. 4. Two generated warped images are then blended to complete the image of the virtual view. If the color of a pixel is different in 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. Zooming can be achieved by changing 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. Synthesizing them in order of distance from the viewpoint completes view interpolation for dynamic regions of soccer scene.

5. ARBITRARY VIEW OBSERVATION WITH A NORMAL VIDEO DISPLAY

While the previous section describes view synthesis for the dynamic regions, this section explains the methods for reconstructing the soccer stadium and presenting entire soccer scene with a normal video display.

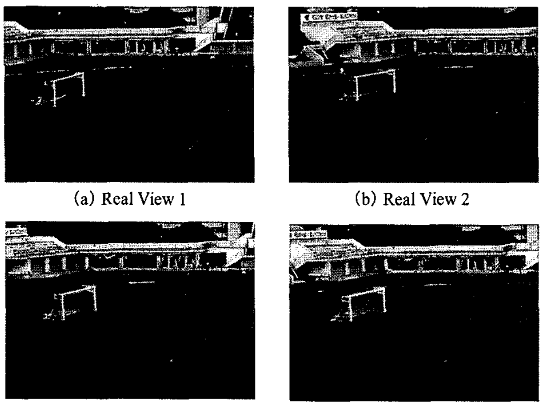
5.1. Reconstruction of The Stadium

The regions of the stadium including ground, soccer goal and spectators' seats are considered as static regions, which have little or no changes over time. View interpolation for all possible intermediate viewpoints is implemented just one time to the background image, where neither players nor ball exist. The static regions can be classified into two regions. One is field regions, which is approximated as sets of planes. The other is a background region, which is approximated as an infinitely distant plane. View interpolation for each region generates intermediate view images, and synthesizing them completes stadium image from another viewpoint.

1) Field regions. In a soccer scene, the ground and soccer goal can be considered as a single plane and a set of planes respectively. We then apply homography to each plane to obtain the correspondences required for the generation of intermediate view images. The following equation gives the pixel-wise correspondence for two views of a plane.

$$\tilde{p}_2 \cong H\tilde{p}_1 \quad (3)$$

where H is the homographic matrix that represents the transformation between the planes, and also, $\tilde{p}_1$ and $\tilde{p}_2$ are homogenous coordinates on the images I_1 and I_2 of different views. The homographic matrices of the plane



(a) Real View 1 (b) Real View 2

(c) Intermediate View (Weight 4:6) (d) Intermediate View (Weight 6:4)

Fig. 5 Example images of the reconstructed stadium.

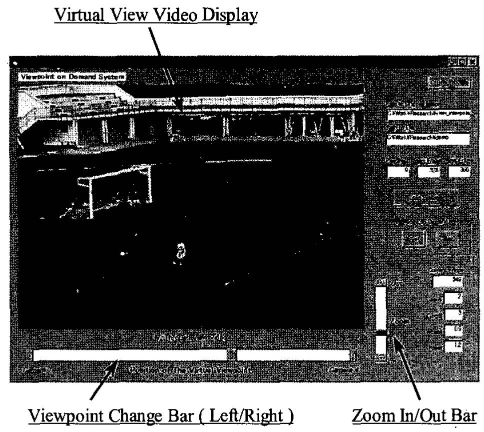


Fig. 6 The interface of Viewpoint on Demand System.

that represents the ground and the planes of the soccer goal provide the dense correspondence within these regions. Image morphing then transfers the position and the value of the pixels in the same way as the dynamic regions. Blending two warped images completes the virtual view image for the field regions.

2) Background region. The background is the region far from camera positions enough to be considered as a single infinitely distant plane. We compose images from each of two real viewpoints in order to make mosaics, which are the respective panoramic images of the background. Intermediate views images are extracted from these panoramic images.

In composition, we start by integrating the coordinate systems of two views through the homographic matrix $\boldsymbol{H}_b$ for the background. Next, blending the pixel values of the overlapping area so that pixel colors at junction areas can be smoothed connects the two backgrounds. The pixel value in the mosaic image is given by the following equation,

$$\acute{v} = \begin{cases} v_1 & (x < x_1) \\ (1-\beta)v_1 + \beta v_2 & (x_1 \leq x \leq x_2) \\ v_2 & (x > x_2) \end{cases} \tag{4}$$

where

$$\beta = \frac{x - x_1}{x_2 - x_1},$$

v_1 and v_2 are the pixel values on the image I_1 and I_2, x_1 and x_2 are the x coordinates of the left side and the right side of the overlapping area. Then partial area that is necessary for each virtual view is cut out from the panoramic image. The following homographic matrix $\acute{\boldsymbol{H}}_b$ is then used in the transformation of coordinates to complete the intermediate view of the background region.

$$\acute{\boldsymbol{H}}_b = (1-\alpha)\boldsymbol{E} + \alpha \boldsymbol{H}_b^{-1} \tag{5}$$

where α is the interpolating weight and $\boldsymbol{E}$ is a 3×3 unit matrix.

Since it is obvious that background region is further from camera positions than field regions, the entire soccer stadium is generated by superimposing the field regions on the background region. Fig. 5 presents example images of the reconstructed stadium. Fig. 5 (a) and (b) show real camera images, and also (c) and (d) show interpolated images from (a) and (b). The interpolating weight of the virtual view to the real views is 4 to 6 in (c) and 6 to 4 in (d).

5.2. Rendering on A Real Stadium Image

As players and a ball always exist on/over the soccer field, the whole soccer scene is completed by rendering the dynamic regions on the image of reconstructed stadium for the desired viewpoint. The global appearance including players, ball and soccer stadium are displayed in window of the GUI.

Fig. 6 describes the interface of Viewpoint on Demand System. At the center of the window, synthesized virtual view images are drawn in accordance with the viewpoint position. In our method, viewpoint position is determined by the position of the virtual camera and the room ratio of the virtual camera to the actual camera. The horizontal slide bar at the bottom of the window decides the position of the virtual camera, which means two reference cameras and the interpolating weight. The vertical slide bar on the right of the window decides the zoom ratio.

Viewers can change the viewpoint using two slide bars anytime during observation, and so watch the soccer match from their favorite viewpoints with a normal video display.

6. ARBITRARY VIEW OBSERVATION WITH A VIDEO SEE-THROUGH HMD

This section explains the algorithm for immersive observation of soccer match on a desktop stadium model. We extend the view-synthesis method to an AR application and enable a user to watch sporting events on the stadium model in front of him/her. The user sees the real world via a video see-through HMD and observes dynamic objects of soccer scenes overlaid onto the stadium model. The procedure is as follows.

(a) Capture a soccer stadium model by an uncalibrated HMD camera.
(b) Detect natural feature lines in the HMD camera image.
(c) Calculate the viewpoint position of the HMD camera.
(d) Generate virtual view of the dynamic objects based on the viewpoint position.
(e) Render the dynamic objects onto the HMD.

In Viewpoint on Demand System, viewpoint position is determined directly by users through GUI, but it is calculated from the image of the HMD camera in this system. In distinction from HMD camera images, we refer to images captured at a stadium as stadium images, and interpolated view images between neighboring cameras as intermediate stadium images. Detail explanation of each process is described below.

6.1. 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 (3,9,16,21). The proposed method employ only captured AR images for the registration between a soccer stadium 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 (17), it is imperative to calibrate both HMD camera and multiple cameras 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, our view-synthesis method is based on the concepts of using uncalibrated multiple cameras. We use the projective geometry for the registration between real images taken with the HMD and soccer images synthesized by our method.

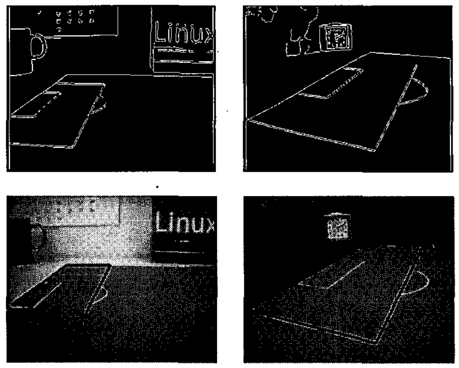


Fig. 7 Examples of edge images (top) and detected natural feature lines (bottom).

For the registration, we need to generate the soccer scene at the same viewpoint of the HMD, and overlay the image onto the HMD camera image. In our view-synthesis algorithm, the position of the viewpoint of the generated soccer image is specified by three elements, which are (a) neighboring two reference cameras, (b) interpolating weight value between two reference cameras and (c) 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 generated 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 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.

1) Detecting natural feature lines. The image of the soccer stadium model captured by the HMD camera 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 are mapped into the Hough space. The strong peaks that form the lines of penalty area and goal area are then found in the Hough space. The results of line detection are shown in Fig. 7 (bottom) with the edge images (top). 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. In the examples of Fig. 7, 4 lines are tracked and used for determination of the viewpoint.

2) Calculating viewpoint position. We apply a vanishing point for selection of neighboring reference cameras

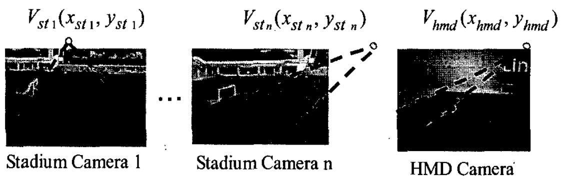


Fig. 8 Location of the vanishing points.

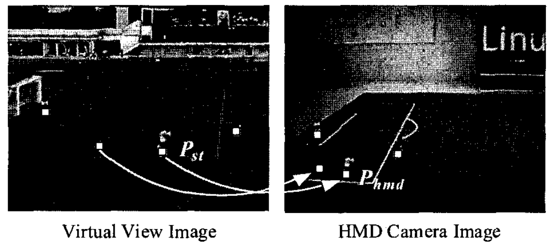


Fig. 9 Determination of the rendering positions on the HMD.

and determination of 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 the closest 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 Fig. 8). Just a horizontal component of the location of a vanishing point is used for calculation of viewpoint position 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,

$$w = \frac{x_{hmd} - x_{stL}}{x_{stR} - x_{stL}} \tag{6}$$

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 the HMD camera image.

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

$$z = f_{hmd}/f_{st}\ . \tag{7}$$

The focal length of the HMD camera is actually fixed, but the zoom ratio can be calculated by changing it virtually when we consider zooming as the change of the focal length. As uncalibrated cameras are used in our approach, the intrinsic parameters of the cameras are unknown. 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,

$$x_{v1}x_{v2} + y_{v1}y_{v2} + f^2 = 0\ . \tag{8}$$

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 (22). Thus the soccer scene at the same viewpoint of the HMD is generated from stadium images.

6.2. Rendering on A Desktop Stadium Model

We use homography transformation for the registration between the HMD camera images and the synthesized soccer scenes. In order to generate natural views of a soccer match, the dynamic objects need to be rendered correctly onto the stadium 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 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 as shown in Fig. 9,

$$\tilde{\boldsymbol{p}}_{hmd} \cong \boldsymbol{H}\tilde{\boldsymbol{p}}_{st} \tag{9}$$

where $\boldsymbol{H}$ is the homographic matrix that represents the transformation between the planes, and $\tilde{\boldsymbol{p}}_{st}$, $\tilde{\boldsymbol{p}}_{hmd}$ 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 Section 4, however, we need to consider about players and a ball which do not contact with the ground. In order to calculate correct positions of the dynamic objects, we apply homogrpahy transformation while taking into account the distance between each object and the ground. Since it is obvious that players and ball exist on/over a soccer field, overlaying them onto the HMD camera images completes the destination image.

7. EXPERIMENTAL RESULTS

We have implemented two types of observation systems for an actual soccer match. Firstly, scenes of the soccer matches were taken by multiple uncalibrated video cameras at two 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 Fig. 10 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 are converted to BMP format image sequences, which are composed of 720×480 pixels, 24-bit-RGB color images. Secondly, fundamental matrices between the viewpoints of the cameras and homographic matrices between the planes form the ground, goal and background in neighboring views are computed from images by manual selection of 50 corresponding feature points in the images.

Next, silhouettes of the dynamic regions in each frame are extracted. After every player region is segmented and labeled, the regions of the same player in the neighboring view images are corresponded by using homography of the ground plane between the views. Furthermore, the images of reconstructed stadium are generated for all possible viewpoints. Then the positions of vanishing points are calculated in each image of actual camera positions with the lines of the goal/penalty area. The above process is applied to the input multiple viewpoint videos as preprocessing of arbitrary viewpoint observation.

In Viewpoint on Demand System, online observation can be performed when users selects favorite scene and viewpoint on GUI. Virtual view image of the dynamic regions are synthesized in accordance with the viewpoint position decided two slide bars on the interface, and then superimposed onto the image of the stadium from the viewpoint. The global appearance of the soccer match is displayed

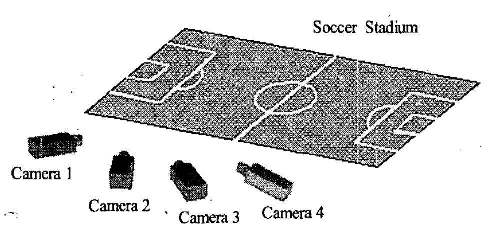


Fig. 10 Camera configuration at stadium.

in the window.

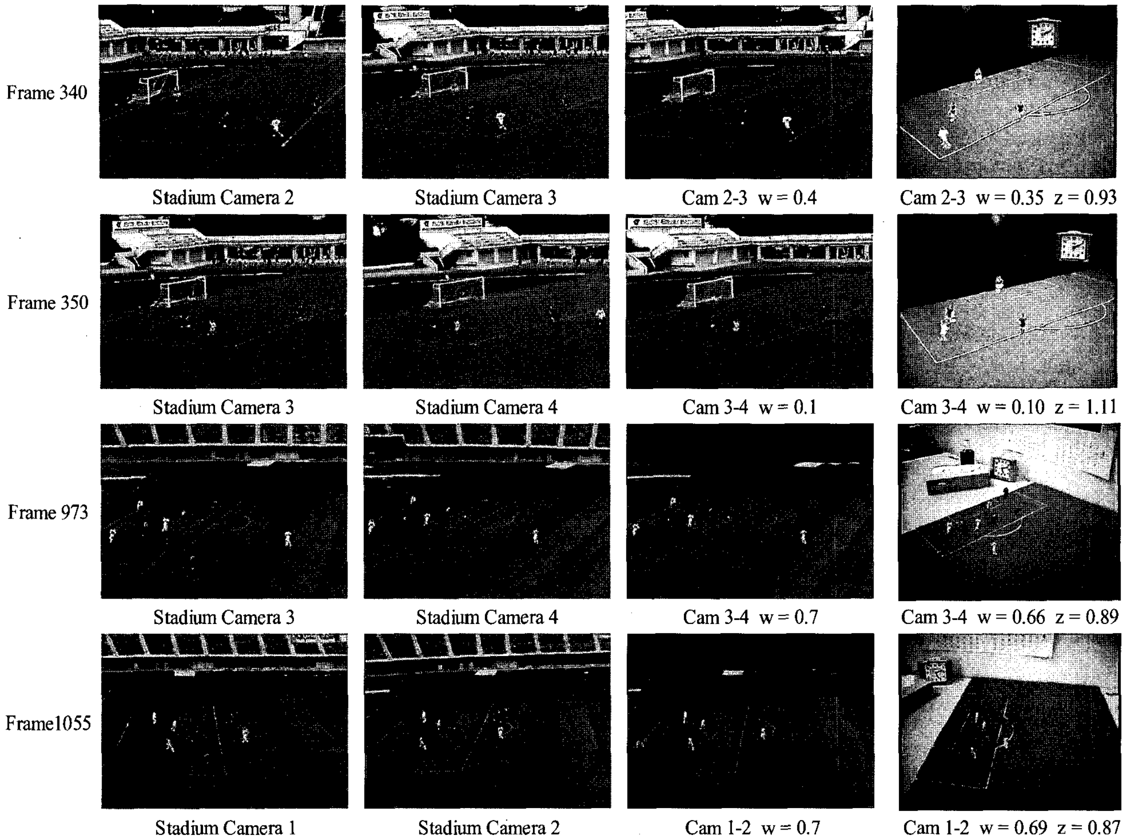
On the other hand, in observation system with a HMD, viewpoint position is calculated from HMD camera images. When observation starts, the lines of the goal area and the penalty area are detected as natural feature lines in the HMD camera image at each frame. Then virtual view image is generated in the same way as Viewpoint on Demand System. Next, homography of the field plane between intermediate stadium image and the HMD camera image is calculated, and virtual view is modified with the homography. Finally the synthesized soccer match is overlaid onto the real stadium model through the HMD. Online process is iterated until viewer stops observation of the soccer match.

Fig. 11 shows the results of displayed soccer scenes at two types of observation systems. The upper two rows describe the soccer match captured at Oita stadium and the lower two rows at Edogawa Stadium. The first and the second columns are reference stadium images used for virtual view generation. The third and fourth columns are overlaid soccer scenes on the real stadium images and on the desktop stadium model respectively. Interpolating weight and zoom ratio are indicated as w and z at the bottom of each image. For example, the image on the top of the last columns was generated based on the parameters that interpolating ratio is 0.35 between camera 2 and camera 3, and zoom ratio is 0.93.

Fig. 12 presents different views of the same frame of the soccer scene. On the left is close-up view displayed on viewpoint on demand system. Synthesized images are photorealistic as if they are actual camera images from the viewpoint. Then on the right is HMD image, where we can see that the dynamic objects of soccer scene are overlaid correctly onto the stadium model. 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 appearance of the objects sometimes has a small error. We are currently improving the method of determination of viewpoint position and zoom ratio stably.

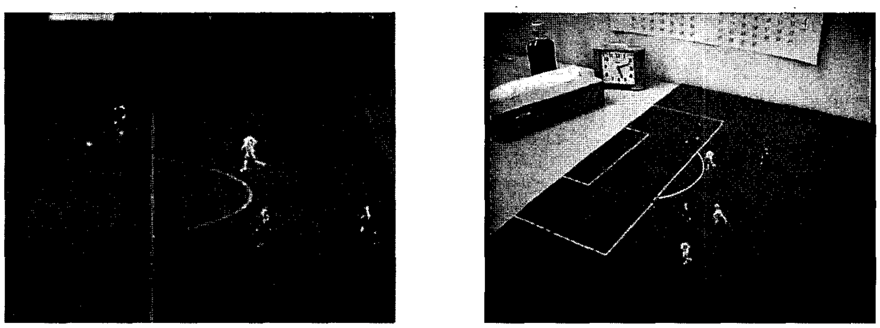
8. CONCLUSIONS

This paper has presented immersive visualization systems for a soccer match from multiple video images captured at stadium. In Viewpoint on Demand System, soccer match can be observed at arbitrary viewpoints with a normal video display. The system has enabled audiences to select favorite angle and zoom ratio, and to change them anytime while watching the soccer match. Furthermore, view-synthesis algorithm for a dynamic event in a large space has been extended to the AR application so that the soccer match can be observed on a desktop stadium model in the real world. The proposed registration



(a) Reference Camera 1 (b) Reference Camera 2 (c) on Real Stadium Image (d) on Desktop Stadium Model

Fig. 11 Displayed soccer scenes on the real stadium images and on the desktop stadium model.



Camera 1-2 weight 0.3 zoom 1.80 Camera 2-3 weight 0.77 zoom 0.92

Fig. 12 Observation from different angle on real stadium image (Left) and desktop stadium model (Right).

method in the AR application, which is based on projective geometry between cameras, uses only natural feature lines in HMD camera images for geometric registration, without any artificial markers or sensor devices. The strong calibration of the HMD camera and multiple cameras that capture the subject is not necessary. Therefore it can be applied to observations not only on the desktop stadium model, but also at any place where user likes. The introduced systems create a new scheme of visual media for entertaining observation of large-scale events.

Acknowledgements

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.

References

1. Azuma,R.,T., 1997, Presence, 6, 4, 355-385
2. Avidan,S., and Shashua,A., 1998, IEEE Trans. on Visualization and Computer Graphics, 4, 4, 293-306
3. Bajura,M., Fuchs,H., and Ohbuchi,R., 1993, Computer Graphics, 26, 2, 203-210
4. Beier,T., and Neely,S., 1992, Proc. of SIGGRAPH '92, 35-42
5. Canny,J., 1986, IEEE Trans. on Pattern Analysis and Machine Intelligence, 8, 6, 679-698
6. Chen,S.,E., and Williams,L., 1993, Proc. of SIGGRAPH '93, 279-288
7. CyberPlay, http://www.orad.co.il
8. Du,W., Li,H., and Gagalowicz,A., 2003, Proc. of Mirage, 70-75
9. Ferrari,V., Tuytelaars,T., and Bool,L.,V., 2001, Proc. of the IEEE and ACM Intl. Symposium on Augmented Reality, 87-96
10. Hartley,R., and Zisserman,A., 2000, "Multiple View Geometry in Computer Vision", Cambridge University Press
11. Inamoto,N., and Saito,H., 2002, International Workshop on Entertainment Computing (IWEC2002) Workshop Note, 94-101
12. Inamoto,N., and Saito,H., 2002, Proc. of International Conference on Pattern Recognition (ICPR2002), 2, 713-716
13. Kanade,T., and Narayanan,P.,J., and Rander,P.,W., 1995, Proc. of IEEE Workshop on Representation of Visual Scenes, 69-76
14. Kitahara,I., Ohta,Y., Saito,H., Akimichi,S., Ono,T., and Kanade,T., 2001, Proc. of International Display Workshops (AD/IDW'01), 1377-1380
15. Kitahara,I., and Ohta,Y., 2003, Proc. of the IEEE Virtual Reality 2003, 45-52
16. Kutulakos,K.,N., and Vallino,J., 1996, Proc. IEEE Virtual Reality Ann. Int. Symp.(VRAIS'96), 25-36
17. Prince,S., Cheok,A.,D., Farbiz,F., Williamson,T., Johnson,N., Billinghurst,M., and Kato,H., 2002, Proc. of the International Symposium on Mixed and Augmented Reality (ISMAR'02), 7-13
18. Saito,H., Baba,S., Kimura,M., Vedula,S., and Kanade,T., 1999, Proc. of the Second International Conference on 3-D Imaging and Modeling (3DIM99), 516-525
19. Seitz,S.,M., and Dyer,C.,R., 1996, Proc. of SIGGRAPH '96", 21-30
20. Seitz,S.,M., and Dyer,C.,R., 1997, Proc. Computer Vision and Pattern Recognition (CVPR1997), 1067-1073
21. Seo,Y. and Hong,K., 2000, IEEE Trans. on Visualization and Computer Graphics, 6, 4, 346-359
22. Simon,G., Fitzgibbob,A.,W., and Zisserman,A., 2000, Proc. of the International Symposium on Augmented Reality, 120-128,
23. Tsai,R.,Y., 1987, IEEE Journal of Robotics and Automation, RA-3, 4, 323-344
24. Wheeler,M.,D., Sato,Y., and Ikeuchi,K., 1998 Proc. of ICCV '98, 917-924
25. Yaguchi,S., and Saito,H., 2001, Proc. of SPIE Vol.4309 (Videometrics and Optical Methods for 3D Shape Measurement), 294-304