Diminished Reality using Multiple Handheld Cameras

Akihito Enomoto* Hideo Saito†

Graduate School of Science and Technology, Keio University, Yokohama, Japan

Abstract

In this paper, we present a system for Diminished Reality with multiple handheld camera system. In this paper, we assume a situation such that the same scene is captured with multiple handheld cameras, but some objects occlude the scene. In such case, we propose a method for synthesizing image for each camera in which the occluding objects are diminished by the support of the different cameras that capture the same scene at the different viewpoints. In the proposed method, we use the AR-Tag marker to calibrate the multiple cameras, so that online process can be possible. By the use of AR tag, we compute homographies for between each camera's imaging plane and the objective scene that is approximated as a planar in this research. The homography is used for warping the planar area for synthesizing the viewer's image that includes only the objective scene without the occluding object that cannot be approximated as a planar area. We demonstrate that the viewer's image without the occluding object can be synthesized for every camera on-line, even though all the cameras are freely moving for all cameras each camera that relates.

Index Terms: Multiple Camera, Augmented Reality, Diminished Reality

1 Introduction

Using multiple cameras for the same scene has recently been popular in various applications. We can get the more information by using multiple cameras, comparing with using one camera. A lot of methods using multiple cameras have also been proposed in AR/MR research field. Sato et al.[1] proposed the head tracking method using a fixed camera for which the extrinsic parameters of the user camera is corrected by another camera that tracks the pose and position of the head. Pilet et al.[2] proposed the geometric and photometric calibration system for AR application using multiple fixed cameras. In this system, fixed cameras other than the viewer's camera for AR observation are used for providing the geometric and photometric calibration to the viewer's camera. Klein et al.[3] used a fixed camera to track the LEDs attached to a tabletPC. Tracking the LEDs provides high accurate pose estimation of tablet-mounted viewer's camera. Ichihara et al.[4] proposed driving support system called NaviView that uses roadside fixed cameras. As described above, most of the researches on AR observation using multiple camera systems are limited in terms of the following points; 1)some of cameras should be fixed 2)the fixed cameras are used only to support the viewer's camera. In this paper, we propose the system in which all cameras are not fixed but can be freely moved, and all cameras can be used as the viewer's camera. Because there is no limitation such that some fixed cameras should be prepared, our proposed system can be used for various applications in flexibility and the wide range. In this paper, we present a diminished reality system based on the proposed multiple camera system such that all camera do not need to be fixed.

*e-mail: akihito@ozawa.ics.keio.a.cjp

†e-mail:saito@ozawa.ics.keio.ac.jp

While Augmented Reality(AR) is a technique that overlays virtual objects onto the real world, Diminished Reality(DR) is a technique for removing the obstacle by overlaying an appropriate background image on obstacle area[5][6][7][8]. DR can be regarded as a part of AR. Related works of Diminished Reality can be divided into two approaches. One approach is based on a dense temporal sequence of video images. Lepetit et al.[7] proposed the method for removing the obstacle by tracking outline of obstacle over the sequence. Another approach is based on images taken by different viewpoints. Zokai et al.[8] proposed the method that uses images capturing from three viewpoints. Breakdown of three cameras is that center of three cameras is user camera and other two cameras are used for capturing the background images. Zokai et al used the proposed method for reconstruction of the industrial site. Zokai's method can deal with non-planar backgrounds by 3D reconstruction of backgrounds.

Our proposed method uses multiple images captured from different viewpoints. In Zokai's method, each camera has an individual role, used for observing the DR scene (viewer's camera) is user camera and other two cameras are used for capturing the backgrounds. In our method, every camera can be used as a viewer's camera and a camera for capturing the backgrounds in the same time. For realizing on-line performance of DR with such multiple cameras that freely move, we apply maker based on-line calibration using ARTag[9]. The advantage of on-line process is to be able to remove obstacles from objective scene even in the case of moving cameras, moving obstacles, and changing objective scene.

The remaining of this paper is organized as follows. Section 2 presents the overview of our proposed system. Section 3 is proposed method for online Diminished Reality. Section 4 shows experimental result. Section 5 describes conclusions.

2 Proposed System

Now, a lot of cameras are used in everywhere. For example, everyone has a camera mounted on a cell-phone. In this paper, we would like to develop a Diminished Reality system that utilizes such cameras held by multiple people as shown in Fig.1. By sharing all images captured by all the cameras, an image without the obstacle can be synthesized at every viewer's position. Based on this system, everyone with the cell-phone camera can watch the objective scene image without any obstacle based on the proposed system, as shown in Fig.1. For instance, people can watch popular paintings in museums even in the case that the paintings are occluded by people standing in front of them.

Based on this system concept, we develop a prototype system as shown in Fig.2. The system consists of multiple cameras which can be freely moved. For example, every camera is held by different users and captures the same scene from different viewpoints. In this paper, we use multiple web cameras connected to different PCs. Each PC's is connected via network, so that they can share the each captured image data through the network. Each camera can be freely moved by a user as long as the ARTag marker placed in the environment can be detected, so that on-line calibration can be possible. By setting ARTag marker on the same plane as objective plane, no manual operation is required for Diminished Reality. Even if the object scene is not planar, it can be approximated as plane for the far-away scene from the cameras. Even when the

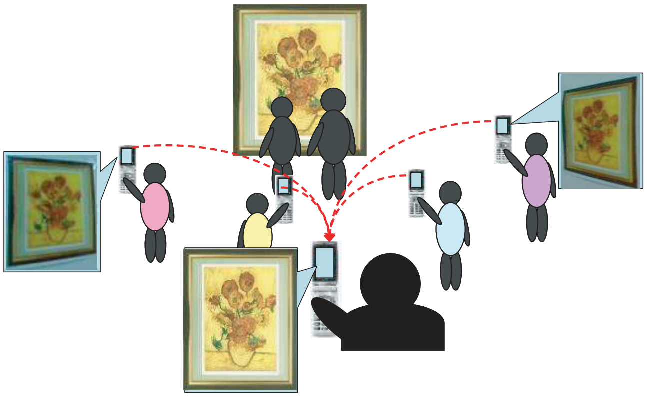


Figure 1: Proposed system

object scene is not on the same plane as ARTag, we can perform DR display by obtaining the 3D geometry of the object scene with off-line pre-process. Although it is possible to use more than three cameras in our proposed system, we will explain and demonstrate the proposed system by using three cameras as minimum system configuration as shown in Fig.2.

In this system, we can generate the image without obstacles at every camera by supporting of the other two cameras. We call the camera for generating the image without obstacle as viewer's camera, while the other two support cameras are called as support cameras. Every camera works as both viewer's camera and support camera.

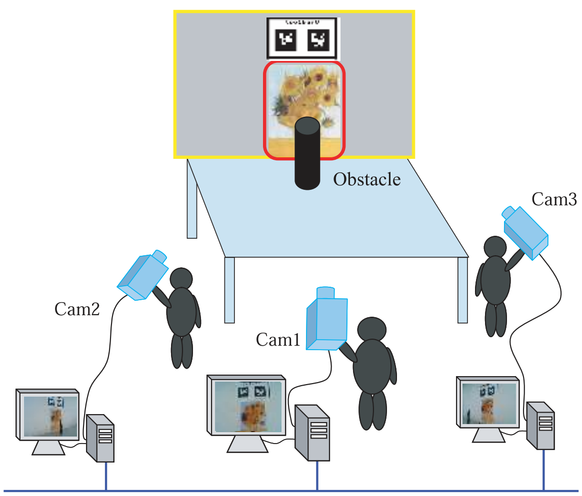


Figure 2: Prototype Experimental System

3 Diminishing Obstacles

As showing in Fig.3, our proposed method consists of the following two processes.

1. Generating warped image
2. Blending

In the first process, we warp the captured images to the other camera's viewpoints by assuming that the object scene can be approximated as a plane. For this warping, ARTag marker is employed for on-line calibration of the cameras. Second, in blending process, we merge the warped images from the other cameras and the captured image for synthesizing. In the blending process, we exclude areas which are not on the planar constraint between the different viewpoints of cameras, so that we can synthesize the image without obstacles. We describe each detail as follow.

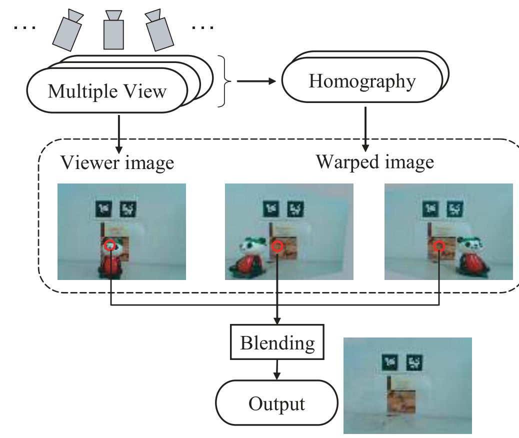


Figure 3: Overview of Proposed method

3.1 Generating the warped image

3.1.1 Computing the Projection matrix by ARTag

In eq.1 , (x,y) is 2D coordinates in the images , (X,Y,Z) is 3D coordinates in the world coordinate system and P is projection matrix that transform 3D coordinates to 2D coordinates. We use ARTag[9] to compute the camera projection matrix P. ARTag is the software that can implement on-line calibration with accuracy and high speed. Projection matrix computed by ARTag detection has the world coordinate system shown in Fig.4.

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \simeq P \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{1}$$

(a) marker on objective plane (b) marker on the perpendicular plane

Figure 4: World coordinate system

3.1.2 Computing the Homography of the objective plane

Using the P-matrix computed by marker detection, it is possible to compute the Homography of the objective plane. If Homography between objective plane and image planes is known, Homography of the objective plane between cameras is calculated as shown in eq.2. Here, we define homography between the objective plane and different cameras A and B as $\boldsymbol{H}_A$ and $\boldsymbol{H}_B$, respectively. Then, Homography $\boldsymbol{H}$ of the objective plane between image$A$ plane and image$B$ plane can be computed by the following equation.

$$\boldsymbol{H} = \boldsymbol{H}_B^{-1}\boldsymbol{H}_A \tag{2}$$

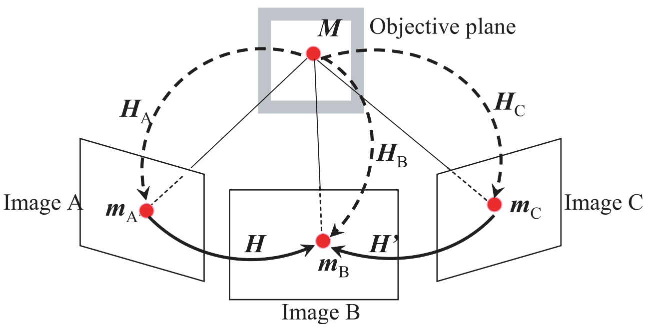


Figure 5: Relationship of each planes

Next, we show that how to compute the Homography between objective plane and image plane. As show in Fig.4(a), when marker is arranged on the objective plane, world coordinate Z=0 plane correspond to the objective plane. For Z=0 plane, P-matrix can be modified as the following.

$$\begin{aligned}\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \simeq P\begin{pmatrix} X \\ Y \\ 0 \\ 1 \end{pmatrix} \simeq \begin{pmatrix} p_{11} & p_{12} & p_{14} \\ p_{21} & p_{22} & p_{24} \\ p_{31} & p_{32} & p_{34} \end{pmatrix}\begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} \\ \simeq H\begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}\end{aligned} \tag{3}$$

As shown in eq.(3), P-matrix(3×4matrix) is expressed as Homography(3×3matrix), which is transformation from 2D coordinates to 2D coordinates. Therefore, Homography between the Z=0 plane and the image plane was able to be computed. As shown in Fig.6, two images were transformed to viewer image by using Homography of the objective plane, and two warped images were generated.

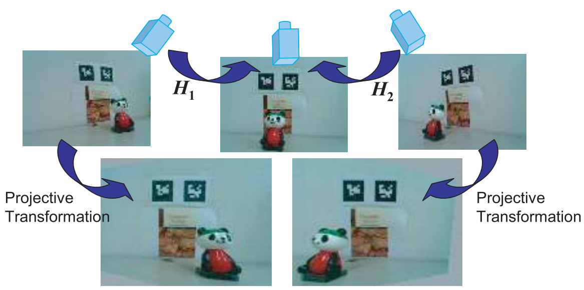


Figure 6: Projective Transformation

3.2 Blending process

Fig.7 is the result that three images are mixed in a equal ratio. In the Fig.7, the obstacles of three images are not diminished. If pixel values are not blended in an appropriate ratio every pixel, obstacle is left. In this section, we explain about the blend processing that is performed to remove an obstacle in detail. In our approach, the blending process is used for removing the obstacles. The blending is the process of mixing the pixel value according to the difference of the pixel values of the captured image and the warped images at the same position. For the warping, there are three pixel values at the same pixel. If they are almost same value, we select the pixel value of the viewer image. If pixel value of viewer image are different others, we select two similar pixel value and ignore one different pixel value because it may be warped from the obstacle area. For convenience, we define pixel value of the viewer's camera at the position of image coordinates $(\boldsymbol{i},\boldsymbol{j})$ as $\boldsymbol{p}_V$, pixel value of the warped image n as $\boldsymbol{p}_n$. Here, pixel value of the output image at the position of image coordinates$(\boldsymbol{i},\boldsymbol{j})$ is defined as $\boldsymbol{q}$. The blending function is shown as follows.

Figure 7: miss blending

$$Case1 \quad if(p_V \approx p_1 \approx p_2) \qquad q = p_V \tag{4}$$

$$Case2 \quad if(p_V \approx p_1 \neq p_2) \qquad q = p_V \tag{5}$$

$$Case3 \quad if(p_V \approx p_2 \neq p_1) \qquad q = p_V \tag{6}$$

$$Case4 \quad if(p_V \neq p_1 \approx p_2) \qquad q = (p_1 + p_2)/2 \tag{7}$$

The Case 2 is situation that an attention pixel is in the obstacle area of the warped image 2. In this case, pixel value of viewer image is chosen in output image. The Case 3 is situation that an attention pixel is in the obstacle area of the warped image 1. In this case, pixel value of viewer image is chosen as well as Case 2. The Case 4 is situation that an attention pixel is in the obstacle area of the viewer image. In this case, pixel value of output image is consist of the warped image 1 and the warped image 2.

Figure 8: Case1

Figure 9: Case2

Warped image1 Viewer image Warped image2

Figure 10: Case3

Warped image1 Viewer image Warped image2

Figure 11: Case4

In obstacle area of viewer's image, pixel value of viewer image is different from two warped images. By using these equations, we use the pixel value of two warped images in obstacle area of viewer's image. So, the obstacle is removed in output image.

3.3 ARTag marker on the perpendicular plane

As described in the previous section, the Homography of the objective plane between all the combinations of cameras should be computed in the on-line process. That is why we place the ARTag on the objective plane. However, even if the ARTag cannot be placed on the objective plane, we can perform the DR display by obtaining the projective relationship between the objective plane and the ARTag marker by off-line process beforehand. In the case of marker on not the objective plane, we have to specify 3D location of objective plane in off-line.

Especially when marker is arranged in the perpendicular plane to the objective plane, world coordinate X = d plane corresponds to the objective plane. In this paper, we simply specify the 3D location as $X = d$ by setting the marker on the perpendicular plane to objective plane. In this case, Homography for the objective plane can be represented as

$$
\begin{aligned}
\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \simeq P \begin{pmatrix} d \\ Y \\ Z \\ 1 \end{pmatrix} &\simeq \begin{pmatrix} p_{12} & p_{13} & p_{14}+dp_{11} \\ p_{22} & p_{23} & p_{24}+dp_{21} \\ p_{32} & p_{33} & p_{34}+dp_{31} \end{pmatrix} \begin{pmatrix} Y \\ Z \\ 1 \end{pmatrix} \\
&\simeq H \begin{pmatrix} Y \\ Z \\ 1 \end{pmatrix}
\end{aligned}
\tag{8}
$$

Here, P-matrix is modified to represent the transformation from 2D coordinate to 2D coordinate for computing Homography between the X=d plane and the image plane. d is calculated by specifying the correspondence points between three images by manual in the off-line pre-process.

4 Experimental Result

We demonstrate the effectiveness of the proposed system via experiments. In these experiments, we removed the obstacle by using the proposed method in both the case to arrange the marker on the object plane and to arrange the marker on the perpendicular plane to the object plane. These experiments were performed on a PC with a 2.8-GHz CPU and 1024 MB of RAM. Windows XP was installed on the PC. We use the three web cameras as minimum system configuration.

4.1 marker on the object plane

When ARTag marker arranged on an objective plane in Fig.12, we remove an obstacle put in front of the objective plane by using proposed method.

Figure 12: marker position

As experiment 1, we show result that the obstacle is diminished when viewer's camera is center among three cameras in Fig.13. Although in Fig.13(a) the objective plane is occluded by an obstacle, in Fig.13(b) the obstacle is removed and we can see the occlusion area on the objective plane.

(a) input (b) output

Figure 13: marker on objective plane

(a) input (b) output

Figure 14: Multiple Cameras Result

Fig.14 shows a result of removing the obstacle in each camera. Fig.14(a) is input images of right camera, center camera, and left camera. In this result, we show that obstacle can be diminished in all cameras. This result represent that all cameras can be used as viewer's camera in our proposed system.

Fig.15(a) shows a situation that the user is drawing the picture(ball and star) on the white paper as the object plane. User's hand becomes the obstacle that occluded the drawn picture in the user camera. However, in Fig15(b), the user's hand can be diminished, and the written picture can be seen by using the proposed method. A part of the area cannot completely be diminished in the output image. This is because that the obstacle area in the warped image overlaps with obstacle area in viewer image.

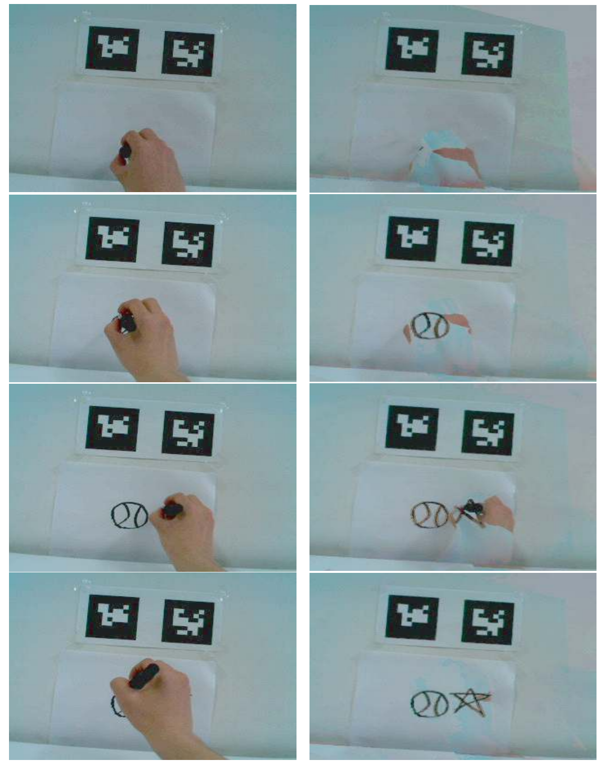
Fig.16 represents the case of solid scene. To compare with the case of plane scene, the blending image is blurred. Homography is plane-to-plane transformation, so if the scene is not a plane, the obstacle can not be diminished as shown in Fig.16(b).

4.2 marker on the perpendicular plane to the object plane

When the marker is arranged in the perpendicular plane to the object plane, it is necessary to estimate depth information from the marker to the object plane. In this experiment, we specified the two pairs of corresponding points between two cameras for computing 3D positions and pose of the marker in off-line processing. As shown in Fig.17, we show that the obstacle can be removed even when the marker is arranged in a perpendicular plane to the objective plane.

5 Conclusion

We proposed multiple handheld camera system used for Diminished Reality. In our system, all cameras should not be fixed and provide DR display. We performed experiments for removing the obstacle in the two cases of marker location. By using the ARTag marker, our proposed method is possible to work in on-line. The experimental results demonstrate that the proposed method can remove the obstacle in the cases of moving obstacle, moving camera, and the changing objective scene.



(a) input (b) output

Figure 15: Removing the Hand

References

[1] Kiyohide Sato, Shinji Uchiyama, and Hiroyuki Yamamoto, A Head Tracking Method Using Bird ' s-Eye View Camera and Gyroscope, Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality(2004)

[2] Julien Pilet, Andreas Geiger, Pascal Lagger, Vincent Lepetit, Pascal Fua, An All-In-OneSolution to Geometric and Photometric Calibration, Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality(2006)

[3] Georg Klein, Tom Drummond, Sensor Fusion and Occlusion Refinement for Tablet-based AR, Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality(2004)

[4] Eitaro Ichihara, Yuichi Ohta, NaviView: Visual Assistance Using Roadside Cameras- Evaluation of Virtual Views -, IEEE Intelligent Transportation Systems Conference Proceedings 2000 Dearborn

[5] S. Mann and J. Fung, Videoorbits on eye tap devices for deliberately diminished reality or altering the isual perception of rigid planar patches of a real world scene, International Symposium on Mixed Reality (ISMR2001)

[6] Adelson and Wang, Representing moving images with layers, IEEE Transactions on Image Processing Special Issue: Image Sequence Compression, September 1994.

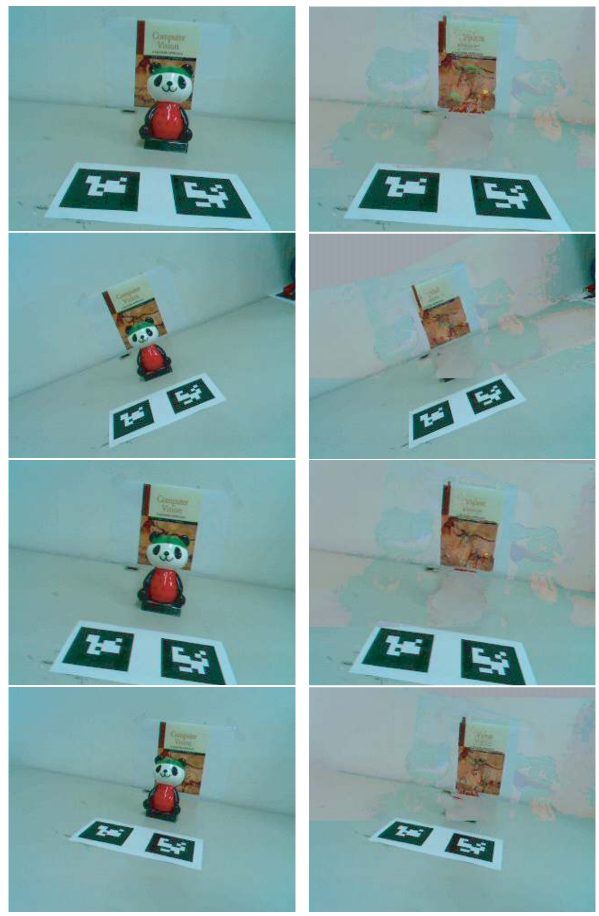
[7] V. Lepetit and M.-O. Berger, A semi-automatic method for resolving occlusion in augmented reality, In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR ' 00), June 2000.

[8] Siavash Zokai, Julien Esteve, Yakup Genc and Nassir Navab, Multiview Paraperspective Projection Model for Diminished Reality, Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2003).

[9] Mark Fiala, ARTag, a fiducial marker system using digital techniques, IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2005

(a) input (b) output

Figure 16: Experimental of solid scene



(a) input (b) output

Figure 17: marker on perpendicular plane