Correct Occlusion Effect in the Optical See-through Immersive Augmented Reality Display System.

Kaori Murase
Tsukuba Univ.
1012, Laboratory of Advanced Research-B,
1-1-1, Tennoudai,
Tsukuba, Ibaraki, Japan
murase@gil.cs.tsukuba.ac.jp

Tetsuro Ogi
Keio Univ.
4-1-1, Hiyoshi,
Kouhoku-ku,
Yokohama, Kanagawa,
Japan
ogi@sdm.keio.ac.jp

Kota Saito
Spin Inc.
Kudan Square Biru,
3-5-11, Kudanminami,
Chiyoda, Tokyo, Japan
http://www.spin-inc.co.jp/viss/index.html

Takahide Koyama
Spin Inc.
Kudan Square Biru,
3-5-11, Kudanminami,
Chiyoda, Tokyo, Japan
http://www.spin-inc.co.jp/viss/index.html

Abstract

This paper proposes the technique for creating the correct occlusion effect in an immersive augmented reality environment with wide field of view. Generally, it is difficult to represent the correct occlusion between virtual objects and real objects in augmented reality displays, based on a semi-transparent mirror. This problem also occurs in the immersive augmented reality display system, named ARView, which we have developed in previous work. To address this problem, we proposed a light projector technique involving projectors which illuminated the surface of real objects using directed illumination which could create shadows occluded by the virtual objects. Notably, this method represented the correct occlusion effect, not only with static objects, but also with moving objects. In this study, we implemented this technique in the ARView system and conducted an experiment to evaluate the effectiveness of this method.

1. Introduction

Recently, augmented reality (AR) technology has been applied to various fields such as entertainment [1] and medicine [2]. Currently, optical see-through or video see-through head mounted displays (HMD) have mainly been used for the augmented reality displays [3]. However, see-through HMDs have several problems or limitations. For example, the user must wear a heavy device, and when he turns his head quickly, he often observes a scene with positional gaps between the real and virtual objects due to the time delay in head tracking [4].

On the other hand, spatial augmented reality systems, such as the HoloPro Screen [5] or the Extended Virtual Table [6], have recently been proposed, and the authors have developed an immersive augmented reality display system named 'ARView' [7]. In ARView, immersive projection technology (IPT) is used, in place of the HMD technology, to represent a high presence augmented reality image.

ARView is a display system that optically combines a virtual scene, projected by projectors, with the real world by using a large, highly transparent semi-mirrored film. By using a large semi-transparent mirror, a user's wide field of view can be covered by the projected image. Moreover, in this system, the virtual world and the real world can be integrated three dimensionally, since the stereoscopic image of the virtual objects can be displayed in three-dimensional space. Therefore, the user can experience a high presence immersive augmented reality environment.

However, it is generally difficult for optical see-through augmented reality display systems using semi-transparent mirror to create the correct occlusion effect when the virtual object is displayed overlapped by a real object, as shown in figure 1.

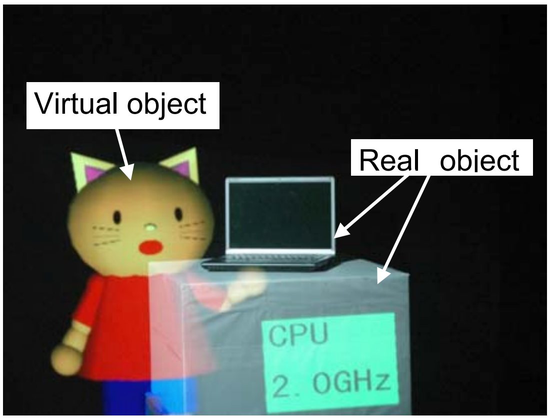


Figure 1: The example of occlusion between virtual objects and real objects.

In this paper, we introduce the technique that solves the occlusion problem in a spatial augmented reality display that employs a semi-transparent mirror. In particular, the method is discussed for controlling the light projector in an immersive augmented reality environment. In addition, we also present evaluation experiments to verify the availability of this technique.

2. Related Study

Several studies have been made regarding the techniques to create the correct occlusion effect in semi-transparent mirror based augmented reality display systems.

For example, ZARD [8], developed by Noda et al., created the correct occlusion effect between real and virtual objects by illuminating the real environment using a projector. The illumination pattern projected by the projector is created, based on the depth information from the real environment, while measuring the three dimensional shape of the real objects in real time. However, in this system, the user's viewpoint is restricted to a fixed position and stereoscopic images cannot be displayed.

Bimber et al. [9] presented illumination techniques to create the correct occlusion effects in optical see-through setups and implemented them in the desktop display known as Virtual Showcase [10]. Their techniques used the projectors to illuminate the real objects and represented the augmented reality scene according to the movement of the user's viewpoint. However, in this system, it was assumed that the depth information of the real environment was known and the positions and shapes of the real objects would not vary.

Although our technique also uses the light projectors, it has some features which differ from the above-mentioned related works. Firstly, our technique dynamically creates the correct occlusion effect according to the movement of the real objects, as well as the user's viewpoint, in a large region of the immersive augmented reality environment. Moreover, the virtual objects are represented using stereoscopic images, so that the virtual image and the real scene are integrated three-dimensionally in the immersive augmented reality environment.

3. Immersive AR display system: ARView

We now consider the system architecture of the immersive augmented reality display 'ARView', developed and used in this research.

Figure 2 shows the system appearance and figure 3 shows the structure of the ARView.

In this system, two stereo projectors are placed, side by side, at ceiling height to display a large image of the virtual world. The images of the virtual objects are projected onto the floor screen, partly overlapped and blended with each other to represent an image with large field of view. A semi-mirrored film is placed at an angle of 45 degrees to the floor. Thus, the users can see the virtual image, reflected by the mirror film in front of them. In this system, since the transparency of the semi-mirrored film is high (visible light transmission is 87.8%), the users can observe the real world behind the mirror as well as the real world in front of it, together with the images of the virtual objects.

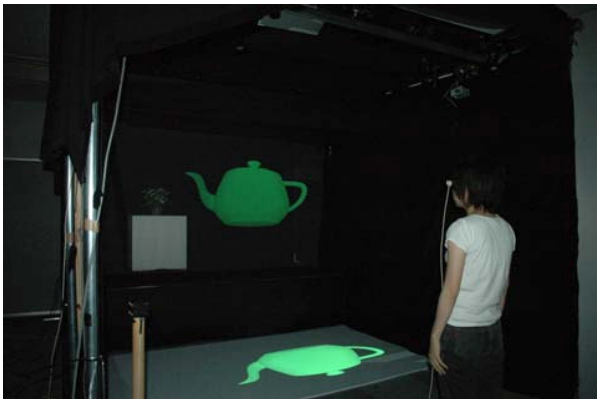


Figure 2: Photograph of 'AR View' equipment.

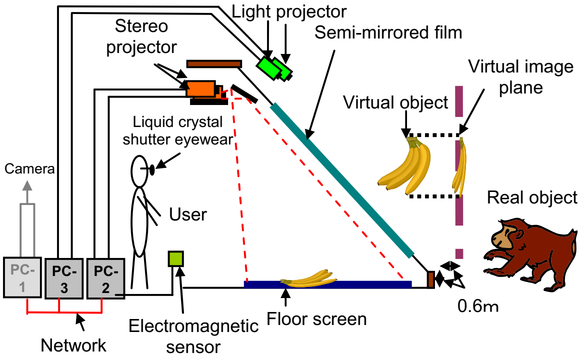


Figure 3: Structure of 'AR View' system.

Regarding the stereo projectors, InFocus DepthQ DLP projectors are used to project the active stereo images of the virtual objects. The left eye and right eye images are projected, alternately, at a frequency of 100Hz, so that the user can see the stereo image of the virtual objects while wearing the StereoGraphics CrystalEYEs3 liquid crystal shutter glasses that are synchronized with the refresh rate of the graphics output. In addition, since an electromagnetic sensor, Polhemus FASTRAK, is attached to the shutter glasses to track the user's viewpoint, the stereoscopic images, seen from the user's viewpoint, can be generated in real time.

The size of the semi-mirrored film is 2.60m wide and 1.95m high. The shape of the displayed virtual image is an isosceles trapezium. The height of the image is 1.40m, and the width of the upper and the lower sides are 2.22m and 2.54m, respectively. The keystone distortion of the displayed image is corrected by the software.

In addition, light projectors that are used to represent the correct occlusion effect between the real objects and the virtual objects (cf. section 4) are placed at the top of the ARView system. Two light projectors are used for illuminating a wide area. These projectors illuminate the real environment behind the semi-mirrored film using directed illumination.

Three PCs are used to control the ARView system. The PC-1 (CPU: Intel(R) Pentium(R) 4 CPU 3.20GHz) is used to capture camera images when the real objects move behind the semi-mirrored film (cf. section 4.3 and 5.2). The PC-2 (CPU: Intel(R) Pentium(R) 4 CPU 3.80GHz, video card: nVidia Quadro FX 1400) is used to receive the positional data from the electromagnetic sensor and to generate the stereo images for the two stereo projectors, while the PC-3 (CPU: Intel(R) Pentium(R) 4 CPU 3.80GHz, video card: nVidia Quadro FX 1400) is used to control the light projectors. To synchronize the images created by the stereo projectors and the light projectors, the positional data of the moving objects and user's viewpoint are sent from the PC-2 to the PC-3 using the UDP protocol.

4. Technique for Correct Occlusion Representation

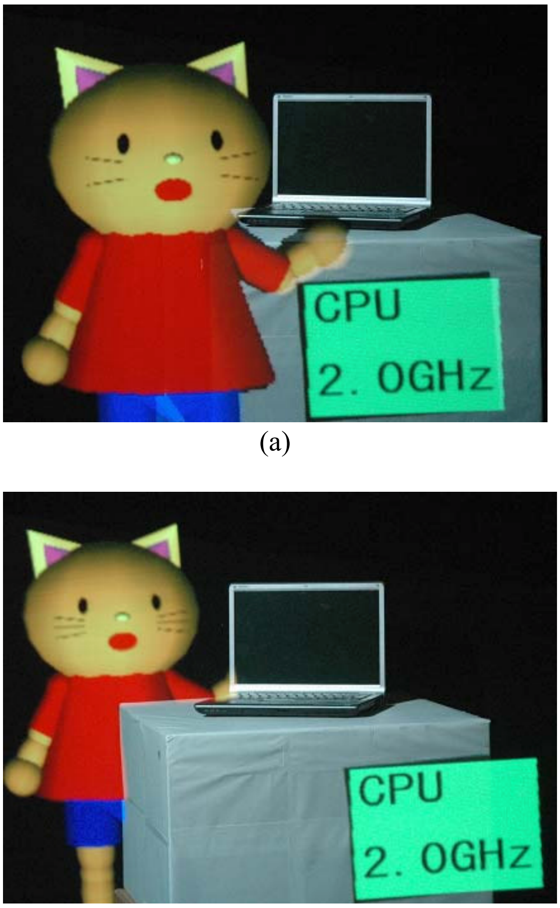
4.1. Occlusion Problem

The problem of incorrect occlusion in an augmented reality display using a semi-transparent mirror is caused by the mechanism whereby the users see the virtual image reflected by semi-transparent mirrors. The virtual images reflected by semi-transparent mirror cannot occlude the scene of the real objects behind the semi-transparent mirror. Therefore, the users perceive both the real object and the virtual object whenever the difference in luminance between the virtual image and the real object is not very great.

In this research, we introduced the light projector function to solve the problem of the incorrect occlusion in the immersive augmented reality display.

The light projectors are used, instead of the standard light bulbs, to project light directly onto the real environment. In other words, these projectors are used to provide computer-controlled light intensity. Since the real object is illuminated by the light projector using white light, an occlusion shadow that is overlapped by the virtual object's image can be created on the real objects' surface without illuminating it. Then, in the area of the occlusion shadow, the luminance of the virtual image would be much brighter than the real object. Thus, the user can see the virtual object situated in front of the real object, as shown in figure 4 (a). On the other hand, when the virtual object is located behind the real object, that part of virtual object's image, which is overlapped by the real object, is blackened to create the occlusion effect, as shown in figure 4 (b).

In this research, LED projectors, TOSHIBA TDPFF1A, were placed in the upper region of the ARView system and they were used as light projectors. Since the brightness of the LED projector was low (15 ANSI lumens), it could be used to illuminate a human who was standing behind the semi-mirrored film, together with the real objects.



(a)

(b)

Figure 4: Use of the light projector function.

4.2. Process of Creating Correct Occlusion Effect

In the augmented reality environment, the representations of correct occlusion effects in the cases such as when a virtual object is located behind a real object and when the real object is located behind a virtual object are necessary, as mentioned in section 4.1.

When the virtual object is behind the real object, the 3D model of the virtual object is rendered with the black 3D model of the real object. Thus, the occluded area of the virtual object's image is blackened, so that the user can see the virtual object situated behind the real object, as shown in figure 4 (b).

When the real object is behind the virtual object, the light projectors project an illumination image onto the real object. The illumination image, projected by the light projectors, is generated using a shadow mapping algorithm [11] in the following process.

(1) Measuring 3D information of real object

When the real objects, illuminated by the light projectors, are static, the positions and the shapes of these objects are measured in advance. However, when the real objects are moving, the positions and the shapes are measured in real time, using the shape-from-silhouette method [12]. The detailed procedures for measuring 3D information from the moving objects will be explained in the next section.

(2) Creating depth texture of virtual object

The user's view position (the left eye position is used in this system) is measured by an electromagnetic sensor and the perspective projection image of the virtual objects, seen from the user's view position, is rendered. In this case, the depth values for each pixel are stored as depth texture.

(3)Projective depth texture mapping

A 3D model of the real objects in a white color is created using the position and shape information measured in (1), and the depth texture is mapped onto it using the projective mapping from the user's left eye position. In this case, the black color is mapped to the pixels whose depth value is greater than the corresponding value in the depth texture.

(4)Projecting illumination image by the light projectors

The 3D model of the real objects, on which the depth texture has been mapped, is then rendered, seen from the position of the light projector using the perspective projection, and it is projected directly onto the real environment by the light projectors as the illumination image.

Thus, the correct occlusion effect can be represented when the real objects are located behind the virtual objects, as shown in figure 4 (a), because only that part of real environment, which is not occluded by the virtual objects, is illuminated.

In addition, when the virtual object is located behind the real object as shown in figure 4 (b), the whole image of the real object with a white color is rendered and the image of the virtual object, that is not occluded by the real object, is projected by the stereo projectors.

4.3. Creating 3D models of moving real objects

In order to use the light projector function, it is necessary to measure the 3D positions and shapes of the real objects, as mentioned in section 4.2. In our technique, 3D information from the static real objects is measured in advance. However, when the light projector function is used for moving real objects, it is necessary to measure the 3D information in real time. In this study, the shape-from-silhouette method is used to measure the 3D information from the moving real objects.

This method is based on the principle that the shape of the 3D object exists within the volume obtained by the back-projection of the silhouette from the corresponding viewpoint. Based on this principle, a set of silhouette images specifies the boundary volume, called a visual hull, by intersecting the volumes created by each silhouette. Thus, the visual hull constructed by all silhouette images approximates to the shape of the 3D object [12]. Therefore, the 3D objects can be reconstructed from the 2D silhouette images. This method has been applied to construct the 3D shape and position of the moving real object.

The following is the procedure of creating the 3D models of the moving real objects for light projector function, using the shape-from-silhouette method.

(1) In this method, infrared floodlights are placed where they can illuminate the real objects, and two infrared cameras are placed at different locations to capture the images of the moving real objects from different viewpoints. An infrared filter is attached in front of the lens of each camera to permit penetration by only the infrared light.
(2) The extrinsic camera parameters and perspective projection matrix of each camera are estimated in the calibration process.
(3) Each camera takes an image of the real environment, where the moving real objects do not exist, beforehand, and these images are stored as the background images.
(4) When the moving real object exists in the display space, each camera takes an image of it. Then the binary silhouette images of moving real object are created, based on a background subtraction algorithm, using the background image stored in (3).
(5) The three dimensional space in which a moving real object can exist is divided into a grid and each grid-point is projected onto each silhouette image by the perspective projection matrix for each camera. If the projected grid-point is inside the corresponding silhouette, it is judged to be inside the silhouette cone (i.e. the volume obtained by

back-projecting the silhouette from the camera's viewpoint.) If the grid-point is judged to be inside all the silhouette cones, it is determined to be inside the real object and a voxel (small cube) is defined at the grid-point. This calculation is conducted for all grid-points. Thus, a 3D voxel model of the real object is created.

(6) By repeating the processes, (4) and (5) in real time, the 3D model of the moving real object can be created

In our approach, we have used the infrared cameras to capture the images of the real objects. If normal cameras were used, it would be difficult to create the correct silhouette images of the real objects, because the foreground object in an image cannot be correctly segmented from the background by the background subtraction algorithm. When a part of the real object is occluded by the virtual object, it cannot be captured by the normal camera because it is not illuminated by the light projectors. This is the reason that the infrared cameras illumination system has been employed. By using the infrared light that cannot be perceived by the user, the images of the real environment can be captured with a stable lighting condition and it is not influenced by the illumination from the light projectors.

Using the above mentioned procedure, the 3D models of the moving real objects are created and the light projector function can be used for the moving real objects.

5. Experiments

5.1. Evaluation of Light Projector Function

We conducted an experiment to evaluate the function of the light projector representing the occlusion effect in the ARView. The accuracy of the user's depth perception was compared, with and without the use of the light projector function.

5.1.1 Experimental Method

Before starting this experiment, a universal coordinate system, common to the virtual world and the real world, was defined, as shown in figure 5.

The x-axis was defined as parallel to the virtual image plane (i.e. the plane on which virtual image is displayed in front of the user) and the floor. The y-axis was vertical to the virtual image plane and the z-axis was vertical to the floor. The origin of x-coordinate was defined at the midpoint of the width of the floor screen, and the origin of y-coordinate was 1.0m in front of the virtual image plane, while that of z-coordinate was on the floor.

Firstly, while wearing the liquid crystal shutter glasses, the subject sat on the chair, placed at the position (0.0, -2.1, 0.0). The real object of a white box (0.31x0.09x0.26m), whose position and shape had been measured in advance, was placed behind the mirror film at the position (0.0, 1.14, 1.25). A stereoscopic image of the virtual cubic box was also displayed.

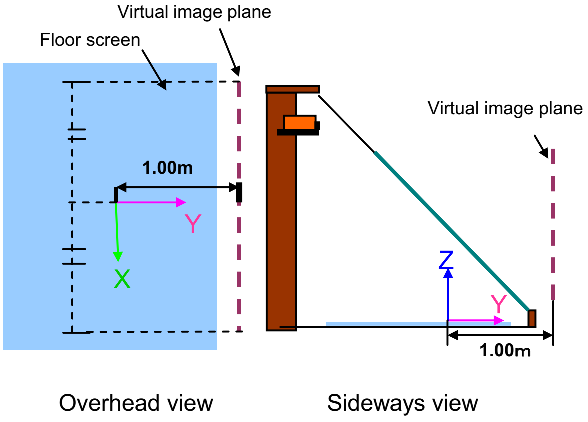


Figure 5: Co-ordinate system used.

In this experiment, the virtual cubic box was located at various depth positions, with y-coordinates, 1.8, 1.4, 1.0, 0.6, 0.2, -0.2, -0.6, and -0.8, so that part of the real box was overlapped by the virtual box when they were seen from the subject's viewpoint, as shown in figure 6. In this case, since the image of the virtual box was rendered using the perspective projection, the box located near the subject appeared larger than the box located farther from him on the virtual image plane.

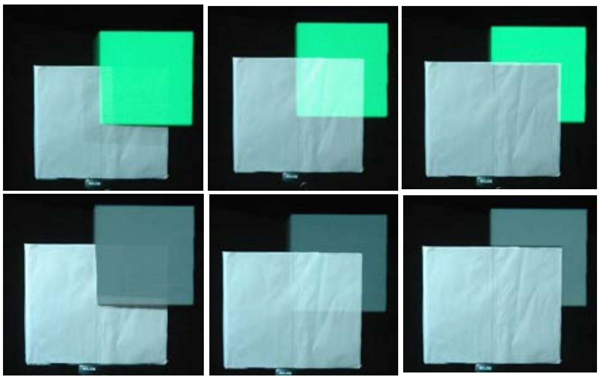


Figure 6: The examples of the displayed images.

Therefore, the size of the virtual box was changed in proportion to the distance from the subject's view position so as to render the appearance of size of the virtual box constant. The virtual box was displayed in one of the eight positions, with the real box in two different conditions, where the function of the occlusion shadow was either used or not used as shown in figure 6. Namely, sixteen scenes were given randomly to several subjects, and they were asked to answer the question "Is the virtual box located in front of the real box or behind the real box?" with the options 'in front of the real box', 'behind the real box' and 'no idea'. The experiment was conducted using four different color conditions for the virtual box, including 'bright green', 'dark green', 'bright white' and 'dull white' as shown in figure 6. Five subjects each performed the experiment three times.

5.1.2 Results and Analysis

Figure 7 shows the results of this experiment. This graph indicates the correct answer rates in the cases where the light projectors were or were not used. The correct answer rates were calculated, ascribing 'no idea' to a wrong answer.

From the results, we can consider that the light projectors improved the users' depth perception in the immersive augmented reality environment where real and virtual objects were integrated.

We evaluated the results of this experiment by using three-way ANOVA; factor A is the method (light projector is used / light projector is not used), factor B is the color hue of the virtual objects (white / green), factor C is the brightness of the virtual objects (bright / dark). From the results, the use of the light projector was significant with a 1% level of significance, although the other factors were not significant.

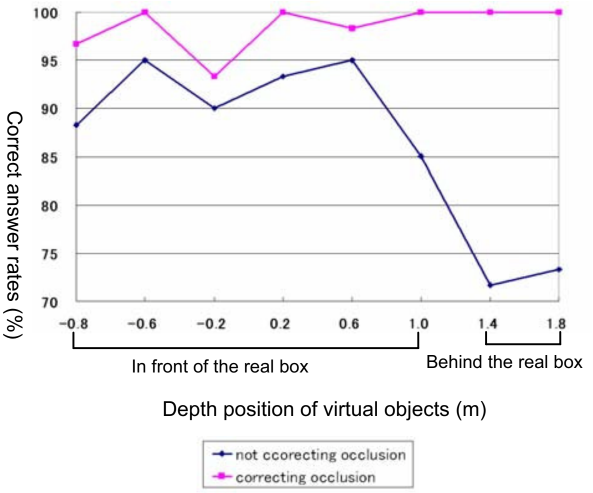


Figure 7: Comparison of results, with and without occlusion correction of images.

In the graph displaying the result when the light projectors were used, however, the correct answer rate was a little lower when the virtual object was located closer to the subject (i.e. the area from -0.2 to -0.8). The LED light projector used in this experiment could not output a stereo image synchronized with the virtual object image, because the refresh rate of the projected light was below 75Hz. For this reason, the light projector creates only the shadow image occluded by the virtual object seen from the use's left eye position. Consequently, the greater the difference in depth values between the virtual object and the real object, the larger was the positional error between the shadow that was projected onto the real object's surface and the virtual object's image seen from the right eye position, as shown in figure 8 (left). This seems to be the reason for the decrease in the correct answer rate when the virtual object was located close to the subject.

From the results of this experiment, we can conclude that the user can perceive the correct positional relationship between a real object and a virtual object when the distance between them is less than 1.0m (the direction error between the shadow and the virtual object's image for the right eye is less than 0.49 deg). In the experimental system, the low brightness LED projectors were used as the light projectors so that they could illuminate a person standing behind the semi-mirrored film. However, the above mentioned problem would be solved by using the light projectors that have a refresh rate of more than 100Hz and could be synchronized with the projected stereo image of the virtual objects.

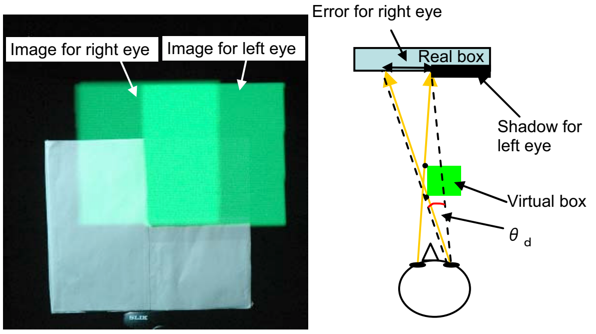


Figure 8: Stereo parallax image and shadow projected onto the surface of a real object.

5.2. Implementation for Moving Real object

In this research, we conducted an experiment in which the correct occlusion representation technique, described in section 4, was applied to a moving real object. The experimental conditions were as follows.

The display system used in this experiment was ARView. The stereoscopic image of the virtual cubic box was displayed while a person (i.e. a moving real object) was standing behind the semi-mirrored film.

Two infrared USB cameras (NETCOWBOY

DC-NCR131), equipped with the infrared filter (IR-78: Fuji film) in front of the lens, were used to capture the images of the person. One camera was placed behind the semi-mirrored film to capture the image seen from the side, while the other camera was placed in front of the semi-mirrored film. The extrinsic camera parameters were measured in the experimental environment.

These cameras were connected to PC-1. Additionally, PC-2 was used to create the images of the virtual object projected by the stereo projectors and to ascertain the user's viewpoint from the electromagnetic sensor. PC-3 was used to create the images for the light projectors. These PCs were connected through a Gigabit Ethernet. When the infrared images were sent to the PC-1 from two infrared cameras, the silhouette images were created by using the background subtraction algorithm in PC-1. In this case, the background noise was filtered out by a morphological opening operation [13][14] and the silhouette images were sent to the PC-2 and the PC-3. In the PC-2, the voxel model of the real object was created, based on the silhouette image using the shape-from-silhouette method, and the stereoscopic image of the virtual object, occluded by the real object, was rendered and then was projected by the stereo projectors. At the same time, positional data, measured by the electromagnetic sensor, was sent from PC-2 to PC-3. In PC-3, the voxel model was also created based on the silhouette images and the light image, occluded by the virtual object seen from the user's viewpoint, was projected by the light projectors.

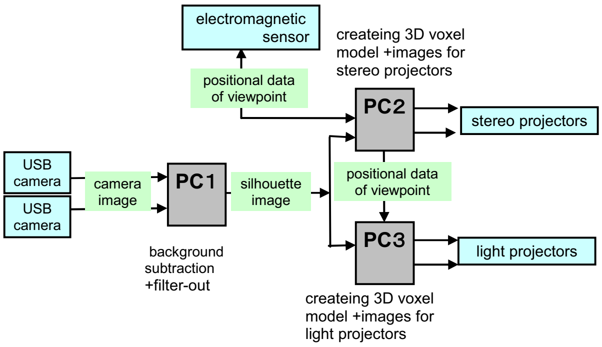


Figure 9: System structure and data flow

The size of the voxel used in the experiment was 2.5cm x 2.5cm x 2.5cm, and the resolution of the captured camera images and the silhouette images was 640 x 480.

In this condition, the correct occlusion effect between the moving real object and the virtual object was represented. The result is shown in figure 9. In figure 9, the person and the virtual object image, comprising a green cubic box, are represented in the immersive augmented reality environment.

In addition, table 1 shows the performances of processes in each PC when the virtual object is located in front of the real object and it is behind the real object. These values include the time for data transmission.

Though each camera can capture images at 30 fps, the PC obtained the images at less than 30 fps because two images captured by two cameras must be synchronized and the capture speed of the USB cameras depends on an amount of received light.

Figure 10: Use of light projector function with a person.

	PC-1	PC-2	PC-3
The speed of image creation (Hz)	16.05	15.93	16.05

Table 1: The speed of image creation on each PC.

From the results, the light projector function was able to realize the correct occlusion effect in real time for a moving real object, although the positional errors between the occlusion shadow, projected onto the real object's surface, and the virtual object's image were a little larger than the error in the case of static real objects.

From the table 1, the images for the light projectors and the stereo projectors were rendered at more than 15 fps. Although, in this research, one person and one virtual object were integrated in an immersive augmented reality environment, these frame-rates depend on the number of polygons rendered in each PC.

6. Summary and Conclusions

In this paper, we have presented a technique to represent the correct occlusion effect between virtual objects and real objects in the immersive augmented reality display, using the light projector function. In particular, we realized the correct occlusion of the moving real object by measuring the 3D positions and shapes of the real objects in real time.

In the evaluation experiment, it was shown that the system of light projectors was useful for the perception of the positional relationship between a real object and a virtual object. In addition, the representation of the correct occlusion between a moving real object and a virtual object was realized in real time.

Future work will include improving the accuracy of integration between moving real objects and virtual objects.

Acknowledgements

This research was supported by NICT Tsukuba JGNII Research Center and G-COE (Center of Education and Research of Symbiotic, Safe and Secure system Design) program at Keio University.

References

[1] Bruce Thomas , Ben Close , John Donoghue , John Squires , Phillip de Bondi , Michael Morris , and Wayne Piekarski. ARQuake: An outdoor/indoor augmented reality first person application. In 4th Int'l Symposium on Wearable Computers, 139–146, 2000.

[2] Michael Bajura, Henry Fuchs, and Ryutarou Ohbuchi. Merging virtual objects with the real world: seeing ultrasound imagery within the patient. Computer Graphics (Proc. of SIG-. GRAPH'92), 26(2): 203–210, 1992.

[3] Ronald T. Azuma. A survey of augmented reality. PRESENCE: Teleoperators and Virtual Environments, 6(4): 355–385, 1997.

[4] Oliver Bimber, and Ramesh Raskar. Spatial Augmented Reality: Merging Real and Virtual Worlds. A K Peters Ltd, 74-76, 2005.

[5] Tetsuro Ogi, Toshio Yamada, Koji Yamamoto, and Michitaka Hirose. Invisible Interface for the Immersive Virtual World. IPT/EGVE2001, .237-246, 2001.

[6] Oliver Bimber, L. Miguel Encarnação, and Pedro Branco. The Extended Virtual Table: An Optical Extension for Table-Like Projection Systems. Presence, 10(6): 613-631, 2001.

[7] Kaori Murase, Tetsuro Ogi, Kota Saito, and Takahide Koyama. Immersive Augmented Reality Display System Using a Large Semi-transparent Mirror, IPT-EGVE Symposium (2007)

[8] Shinichi Noda, Yoshihiro Ban, Kosuke Sato, and Kunihiro Chihara. An Optical See-Through Mixed Reality Display with a Realtime Rangefinder and an Active Pattern Light Source. Transactions of the Virtual Reality Society of Japan, 4(4): 665-670, 1999.(in Japanese)

[9] Oliver Bimber, and Bernd Fröhlich. Occlusion Shadows: Using Projected Light to Generate Realistic Occlusion Effects for View-Dependent Optical See-Through Displays. Proc. International Symposium on Mixed and Augmented Reality (ISMAR '02), 186-195, 2002.

[10] Oliver Bimber, Bernd Fröhlich, Dieter Schmalstieg, and L. Miguel Encarnação. The Virtual Showcase. IEEE Computer Graphics and Applications, 21(6): 48-55, 2001.

[11] Mark Segal, Carl Korobkin, Rolf van Widenfelt, Jim Foran, and Paul Haeberli. Frobnication. Fast Shadows and Lighting Effects Using Texture Mapping. ACM SIGGRAPH Computer Graphics, 26(2): 249-252, 1992.

[12] Aldo Laurentini. How Far 3D Shapes Can Be Understood from 2D Silhouettes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(2): 188-195, 1995.

[13] Jean Serra. Image Analysis and Mathematical Morphology. Academic Press, Inc., 1982.

[14] Jean Serra. Image Analysis and Mathematical Morphology 2. Academic Press, Inc., 1988.