Design and Evaluation of Mirror Interface MIOSS to Overlay Remote 3D Spaces

Ryo Ishii[1(✉)], Shiro Ozawa[2], Akira Kojima[2], Kazuhiro Otsuka[1], Yuki Hayashi[3], and Yukiko I. Nakano[4]

[1] NTT Communication Science Laboratories, NTT Corporation, Atsugi, Kanagawa, Japan
{ishii.ryo,otsuka.kazuhiro}@lab.ntt.co.jp
[2] NTT Media Intelligence Laboratories, NTT Corporation, Yokosuka, Kanagawa, Japan
{ozawa.shiro,kojima.akira}@lab.ntt.co.jp
[3] Colleage of Sustainable System Sciences, Osaka Prefecture University, Habikino, Osaka, Japan
hayashi@kis.osakafu-u.ac.jp
[4] Faculty of Science and Technology, Seikei University, Musashino, Tokyo, Japan
y.nakano@st.seikei.ac.jp

Abstract. The MIOSS mirror interface can overlay two remote spaces, enabling users to feel as if they are in the same room and thereby to share 3D objects in the spaces. MIOSS imparts motion parallax through a mirror that adjusts to the viewpoint of the user, in addition to providing geometrical consistency in the occlusion, size, and positional relationships in the two remote spaces. Experimental evaluations of an implemented MIOSS system show that users can recognize the exact positions of shared objects in the partner's space via the mirror video.

Keywords: Mirror interface · Motion parallax · 3D modeling · Overlaid space

1 Introduction

One of the big challenges in creating media spaces is how to achieve the sharing of remote spaces containing people and objects. If this can be achieved, we will be able to work closely together while sharing our respective spaces, to discuss things, and to smoothly perform collaborative work with the shared objects. Several studies have made attempts to create systems to share two remote spaces as one shared space [1–3]. These systems make it possible to share objects in a narrow area but do not permit complete sharing of the whole space. We aim to achieve an advanced media space that provides a seamless overlay between two remote spaces containing users and objects. This will enable users to share the objects in the two spaces and work closely together while sharing their respective spaces, to discuss things, such as furniture layouts, and to smoothly perform collaborative work with the shared objects. The enormous challenge in realizing such a media space is how to share two remote spaces with real objects and display the video to users naturally.

J. Abascal et al. (Eds.): INTERACT 2015, Part IV, LNCS 9299, pp. 319–326, 2015.
DOI: 10.1007/978-3-319-22723-8_25

To meet this challenge, we have developed a method, called MIOSS, for overlaying two remote spaces through a mirror video. The MIOSS enables users to feel as if they are in the same room and to share objects in their respective spaces. Figure 1 shows images of perspectives with a real mirror and MIOSS. The real mirror reflects the spaces of both user A and user B. With MIOSS, the display reflects the video of one space overlaid on the other space, and the users feel as if their space is overlaid on their partner's space. As an example of motion parallax, the yellow columnar object in user A's space is located behind user B on the display. The blue triangular object is located in front of user A on the display.

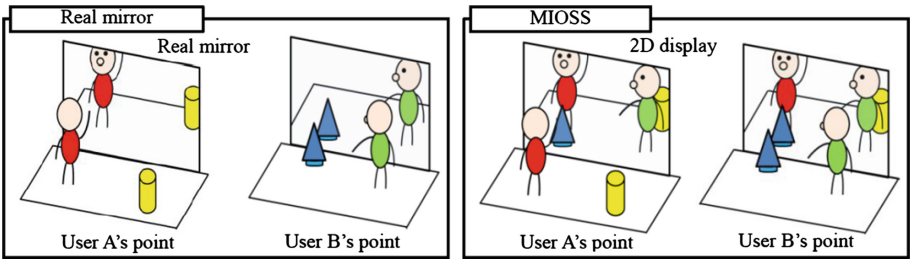

Fig. 1. Illustrations of real mirror and MIOSS.

The MIOSS system constructs 3D models of two remote spaces and displays the mirror video of a 3D model generated by overlaying the 3D models of the two spaces on a 2D display. The video provides geometrical consistency in occlusion, size, and positional relationships in the two remote spaces. Moreover, MIOSS imparts motion parallax through a mirror that adjusts to the viewpoint of the users. In this paper, we describe an implementation tool of MIOSS with a minimal setup as a first attempt to develop a MIOSS prototype. The setup comprises multiple Kinects and a 2D display in each user's space. We evaluated whether the implemented MIOSS enables users to recognize the exact positions of shared objects in the partner's space. The results show that the advanced functions of MIOSS— construction of 3D models of two remote spaces; reproduction of geometrical consistency in the position, size, and occlusion relationships among objects in the two spaces: and motion parallax to adjust to a user's viewpoint—enable users to recognize the exact positions of shared objects in the partner's space via the mirror video.

2 Related Work

Several studies have proposed systems for sharing two remote spaces as one shared space. Agora [1] provides the shared space on a desk, and users can share real objects in each space on the desk. In t-Room [2], each space has a 2D display. Video of users and objects in front of one of the displays is projected and displayed on the 2D display in the other remote space. Users and objects directly in front of the 2D displays can be shared, but those not directly in front of them cannot be. HyperMirror [3] overlays the

2D image of a space on the 2D image of another space. It doesn't construct a 3D model of the remote spaces and doesn't reproduce the geometrical consistency in the position, size, and shielding relationships of real objects in two remote spaces. In addition, it doesn't impart motion parallax to adjust to a user's viewpoint.

On the other hand, 3D modeling of users and objects with multiple depth sensors, such as Kinect, has been attracting attention lately [4–8]. In addition, media spaces that display a 3D model of a remote space on a 3D display have been developed [6–8]. These systems aimed to connect the two spaces via the display as a bonded surface and join the two remote spaces.

Holoflector [9], which has a half-silvered mirror three feet in front of a large LCD screen, can superimpose 3D modeling data on an image reflected on the mirror. This lets the system create some interesting interactive effects, such as turning the user into a pixelated mannequin generated by 3D modeling, displaying a floating "hologram" above the user's outstretched palm, or raining little bouncing balls all around you. However, this system cannot overlay two remote spaces reconstructed by 3D modeling.

MIOSS, the system proposed here, expands the functions of HyperMirror [3] to construct 3D models of two remote spaces and reproduce geometrical consistency in the position, size, and occlusion relationships among the objects in two spaces. In addition, it imparts motion parallax to adjust to a user's viewpoint. This research is the first attempt to create a system that can display a mirror video that reproduces geometrical consistency with motion parallax.

3 Mioss

3.1 System Summary

Figure 2 shows the system architecture of MIOSS for user A (the architecture for user B is the same). In each space, there are two Kinect cameras, which capture RGB and depth images, and a 2D display or projector screen. The processing steps for presenting a mirror video on user A's 2D display are as follows:

- Measure user A's viewpoint position: The 3D position of the center of the user's eyes in a world coordinate system is measured as the user's viewpoint by using robust face-tracking technology with a memory-based particle filter [10].
- Construct a 3D model of user A and user B: To construct the 3D model of the spaces of users A and B, the system uses the RGB and depth images from the Kinects and combines the 3D models of the spaces of users A and B in the same way as in previous research [4–8]. To construct a 3D model of a space, the system captures the RGB and depth images from the two Kinect sensors in each space. The system generates the two sets of 3D point cloud data from data captured from each Kinect using the of Point Cloud Library (PCL) function [11] for each space. The two sets of 3D point cloud data are combined with calibration data generated by a preprocessing for calibration by Zhang's method [12] between the two Kinects for each space. Finally, the two sets of 3D point cloud data for the spaces of users A and B are combined. At this time, the actual geometric consistency between the two

spaces is realized in consideration of the positional relationship between the position and mounting posture of the Kinects in the spaces.

- Generate the mirror image: With a 2D display used as a projection surface, the generated 3D model is projected in perspective to match the measured user's viewpoint. At this time, the world coordinate systems of the 3D position of the user's viewpoint and the 3D model are converted into the same coordinate system. Thus, mirror video on the display is implemented.

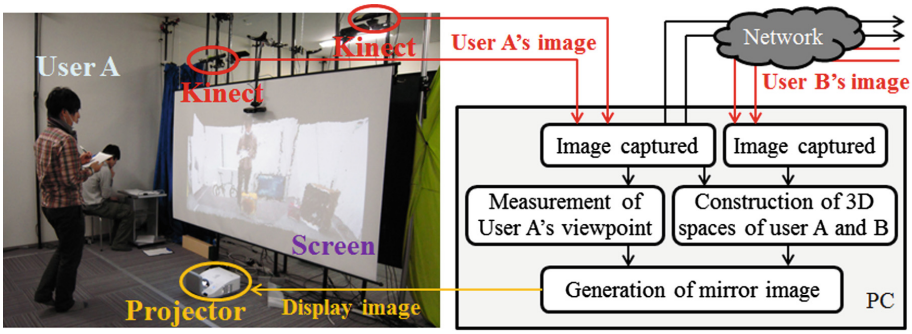

Fig. 2. System architecture of MIOSS.

3.2 Implementation

Using the above-described methods, we implemented a prototype of the MIOSS mirror interface. The development environment comprised two Kinects with VGA level resolution (640 × 480) of the RGB and depth images in each space, a computer with an Intel Core i7-3960X CPU and 16 GB of memory, and a NVIDIA GeForce GTX580 graphics board. The implementation results are summarized in Table 1, where "delay time of viewpoint movement" is the time from the user's viewpoint position's moving to the time the motion parallax appears in the video and "delay time of camera image" is the time until the captured video appears. In this regard, the prototype sends the RGB and depth image data from the Kinects to the partner's system directly via a non-computer network. The "delay time of camera image" has no network delay time for sending the data in this case.

Table 1. Performance of prototype of MIOSS

	Frame rate	Delay time
Camera image	About 15 fps	About 500 ms
Viewpoint movement	About 15 fps	About 500 ms

4 Evaluation of Recognition of Object's Position

4.1 Experimental Method

MIOSS expands HyperMirror [3] with two new functions. First, it constructs 3D models of two remote spaces and reproduces geometrical consistency in the position,

size, and occlusion relationships among the objects in two spaces. Second, it imparts motion parallax to adjust to a user's viewpoint. We conducted experiments to examine whether the new two functions in MIOSS contribute to the user's recognition of the precise positions of objects in the partner's space. We set the following three experimental conditions as within-subject factors.

- **2D condition:** Mirror video of overlay of the 2D image of a space on the 2D image of another room (same as HyperMirror [3]). The video doesn't reproduce the geometrical consistency and doesn't impart motion parallax to adjust to a user's viewpoint. The setting of a user's viewpoint is where the image is displayed when the user is in the center of the room. As a method to generate the video for the condition, only the user and the objects in the space are extracted from the RGB image using depth information. The extracted user and objects are overlaid on the partner's RGB image.
- **3D condition:** Motion parallax is excluded from MIOSS. The mirror video realizes the actual geometric consistency. The setting of the user's viewpoint position is where the image is displayed when the user is in the center of the room.
- **MIOSS condition:** MIOSS is used. The mirror video realizes actual geometric consistency and motion parallax.

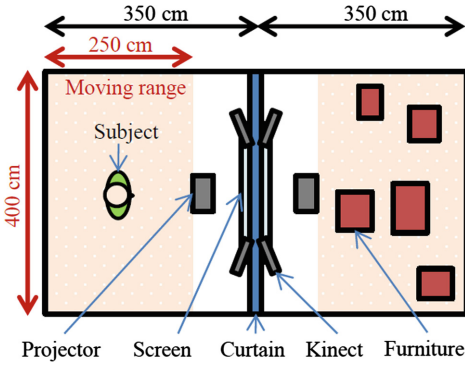

Fig. 3. Experimental equipment.

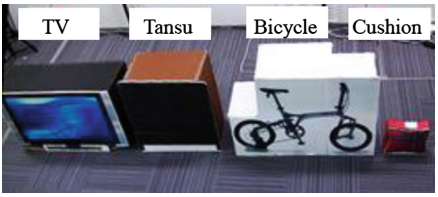

Fig. 4. Examples of objects used in the experiment.

We evaluated the effect of the repetition of the geometric consistency in the occlusion, size, and positional relationships in the two remote spaces by comparing the 2D condition with the 3D condition. We evaluated the effect of the repetition of the motion parallax by comparing the 3D condition with the MIOSS condition. Moreover, we evaluated the multiple effects of the repetition of the geometric consistency and the motion parallax by comparing the 2D condition with the MIOSS condition.

The experimental setup is shown in Fig. 3. The participants entered adjacent rooms (400 cm × 350 cm) divided by a curtain. They were allowed to move freely within the area (400 cm × 250 cm) 100 cm away from the screen. There were 100-in. (125 cm × 221 cm) screens placed 60 cm above the floor in front of the curtain in each room. The mirror video, which is 80 in. (108 cm × 172 cm), was projected onto the

screens from the projector. Two Kinects were installed at the top of the screen on either side. Five objects were positioned randomly in the space where there was no participant. The objects were photographs of real furniture and a bicycle of actual scale pasted to cardboard boxes (Fig. 4). There were three different combinations of five pieces of furniture (the total number of the pieces of furniture was fifteen). They were used as the objects for each condition randomly. The experiment began with the subject standing in the center of the space. At a signal to begin, the mirror video was output. The subject was given three minutes to write down the positions and sizes of the five objects in the room plan. The room plan has lines that indicate intervals of 5 cm. To minimize order effects, the three experimental conditions were used randomly. Each pair of participants used a different set of objects in each condition. After executing the experiment in each condition, the participants filled in a questionnaire for subjective evaluation (six-point Likert Scale) of their impression of the ease of recognizing the objects' positions. Sixteen persons (12 males and four females in their 20 s) participated in the experiments.

4.2 Results of Recognition of Position

To evaluate how accurately the subjects were able to recognize the positions of the objects, we calculated the average error for all subjects between the center position reported by the subjects and the actual positions. We calculated the average error in the lengthwise direction, i.e., the direction perpendicular to the screen, and the crosswise direction parallel to the screen. The results are shown in Fig. 5. To determine whether experimental conditions made a difference in the position-recognition error of object's position of lengthwise and crosswise directions, we performed a one-way repeated factorial analysis of variance. The results showed a significant difference between experimental conditions ($F(1,45) = 3.55$, $p < .05$ for lengthwise direction; $F(1,45) = 2.51$, $p < .10$ for crosswise direction). Next, we performed multiple comparisons using the Tukey-Kramer method to identify differences between pairs of conditions. These tests showed that there are significant differences in the perception error of position in the lengthwise direction and that differences in the perception error of position in the crosswise direction trended to appear only between the 2D and MIOSS conditions ($p < .05$ for lengthwise direction; $p < .10$ for crosswise direction). The results demonstrate that position-recognition error was smaller for the MIOSS condition than for the 2D condition.

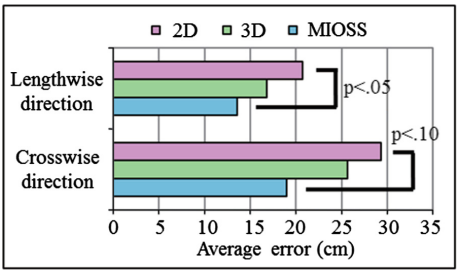

Fig. 5. Average perception error of position.

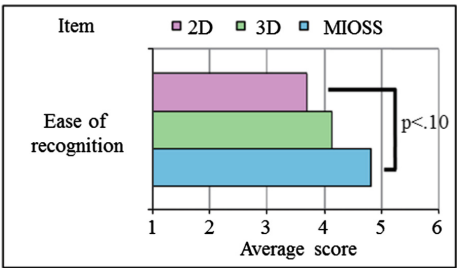

Fig. 6. Average score of subjective evaluations reported by subjects and the actual positions.

4.3 Results of Subjective Evaluation

The average score of participants' subjective evaluations for "ease of recognition" is shown in Fig. 6. We performed a one-way repeated factorial analysis of variance for the item to determine whether experimental conditions affected the values. Since the analysis showed a marginally significant effect of experimental conditions on the evaluation ($F(1,45) = 2.57$, $p < .10$), we performed multiple comparisons using the Tukey-Kramer method. A difference trend was found for the "ease of recognition" between the 2D and MIOSS conditions ($p < .10$). The average subjective score of the item of "ease of recognition" in the MIOSS condition is very high (4.8) and 1.2 score higher than in the 2D condition.

5 Discussion

In the results of the experiment, there was no difference in the average error of recognition of the object's position and subjective evaluation values between the 2D and 3D conditions. In contrast, the result showed that the average error of recognition of the object's position in the MIOSS condition is smaller than in 2D condition. The subjective evaluation of "ease of recognition" in the MIOSS condition was higher than in the 2D condition. The motion parallax, in addition to the reproduction of geometric consistency, is effective for enabling users to recognize an object's precise position. In the MIOSS condition, the average error of the recognition in the lengthwise direction was about 13.5 cm and that in crosswise direction was about 19 cm. The errors are very small. When a person observes an object that is directly in front of his/her eyes and writes the position on a sketch of the room, a little error is likely. In addition, the average subjective score of the item of "ease of recognition" in the MIOSS condition was very high (4.8). From the above, it is believed that users can recognize the position of an object accurately with MIOSS. Therefore, these results suggest that the reproduction of geometric consistency in video is, by itself, not sufficient for recognition of the precise position of objects in the partner's space. The motion parallax, in addition to the reproduction of geometric consistency implemented in MIOSS contributes to user's recognition of the precise position of objects in the partner's space. In this research, the prototype of MIOSS was implemented with a minimal setup with a 2D display (screen); it does not impart stereoscopic indications by means of monocular parallax with a 3D display. However, the results suggested that the motion parallax alone is quite sufficient for users to be able to recognize the position of objects in the mirror image.

6 Conclusion

We aim to create a media space that can overlay two remote spaces. We presented MIOSS, which enables users to feel as if they are in same room through the mirror and thereby to naturally share objects in two spaces. We developed a prototype of MIOSS that imparts motion parallax through a mirror that adjusts to the viewpoint of the user, in addition to providing geometrical consistency in the occlusion, size, and positional

relationships in the two remote spaces. Experimental evaluations with the implemented MIOSS showed that the video expression of the geometric consistency in the video and motion parallax enables users to recognize the exact positions of shared objects in the partner's space via the mirror video. In future work, we plan to evaluate the effect of MIOSS in terms of the smoothness of remote cooperative work in detail with a conversation analysis.

References

1. Kuzuoka, H., et al.: Agora: a remote collaboration system that enables mutual monitoring. In: CHI Extended Abstracts, pp. 190–191 (1999)
2. Hirata, K., et al.: t-Room: remote collaboration apparatus enhancing spatio-temporal experiences. In: Proceedings of CSCW (2008)
3. Morikawa, et al.: HyperMirror: toward pleasant-to-use video mediated communication system. In: Proceedings of CSCW, pp. 149–158 (1998)
4. Newcombe, E., et al.: KinectFusion: real-time dense surface mapping and tracking. In: Proceedings of ISMAR, pp. 127–136 (2011)
5. Kainz, B., et al.: OmniKinect: real-time dense volumetric data acquisition and applications. In: Proceedings of VRST, pp. 25–32 (2012)
6. Maimone, A., et al.: Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras. In: Proceedings of ISMAR (2011)
7. Maimone, A., et al.: A first look at a telepresence system with room-sized real-time 3D capture and large tracked display. In: Proceedings of ICAT, vol. 1 (2011)
8. Beck, S., et al.: Immersive group-to-group telepresence. IEEE Trans. Visual. Comput. Graphics **19**(4), 616–625 (2013)
9. Holoflector. http://research.microsoft.com/apps/video/default.aspx?id=159487&r=1
10. Mikami, D., Otsuka, K., Yamato, J.: Memory-based particle filter for tracking objects with large variation in pose and appearance. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part III. LNCS, vol. 6313, pp. 215–228. Springer, Heidelberg (2010)
11. PCL (The point cloud library). http://pointclouds.org/
12. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. **22**(11), 1330–1334 (2000)