MulDiRoH: An Evaluation of Facial Direction Expression in Teleconferencing on a Multi-view Display System

Shiro Ozawa, Satoshi Mieda, Munekazu Date, Hideaki Takada, and Akira Kojima

NTT Media Intelligence Laboratories, NTT Corporation, Yokosuka, Kanagawa, Japan
ozawa.shiro@lab.ntt.co.jp

Abstract. We have developed a teleconference system called MulDiRoH (Multi-Directional Representation of Humans). It features the use of a QDA screen, one of the newest multi-view display techniques. A principal benefit of multi-view displays is they can show views of a remote participant from the direction in which the participant's face is pointing. This enables other participants to directly see the face of a remote participant who is actually looking away from them. However, all multi-view display systems share a common problem in that users who stand outside of the center area cannot observe geometrically correct images. To addressthis problem, we propose the use of the perspective transform method. We also evaluate the conveying of a person's facial direction by a communication game for multiple users.

Keywords: Teleconferencing, Multi-view Display, QDA Screen.

1 Introduction

Recent years have seen teleconference system displays that show people in remote places become increasingly important. In connection with such displays, multi-view viewing display systems that provide multiple images in different viewing regions are being developed and are expected to give rise to new systems. Such a system is our goal of our research; Fig. 1 shows an image of it. Since a multi-view display can show the field of view of remote participants from the direction their face is facing, it is especially useful as a display for showing multiple facial directions in a video conferencing system. This allows participants to look at a remote participant's face directly even though the participant is actually looking away from them.

In order to develop a video conferencing system using a multi-view display, Jones et al. proposed a system using a high-speed rotating mirror and a high-speed projector [1]. This system uses a computer vision technique to acquire a three-dimensional model of the remote user's face. In particular, it applies a pre-determined light pattern to the subject and a camera takes a picture to obtain the pattern. It then puts a texture on the obtained shape and produces an image of the subject. This system has been evaluated quite highly; however, there are some problems with it. First, arranging the

S. Yamamoto (Ed.): HIMI 2014, Part I, LNCS 8521, pp. 525–535, 2014.

Fig. 1. Concept image of our goal

mirror and projector may require considerable work. Second, conversations may be hindered by noise generated from the mirror. Third, using a mirror means it is quite possible the image resolution will decrease over time.

Feldmann et al. proposed a two-on-two-on-two system with a nine-aspect auto- stereoscopic display using a lenticular lens [2]. In this system, six participants sit around a table and each of them is allocated one auto-stereoscopic display from each remote participant. Since the participants' freedom of movement is restricted, however, they cannot get more than four of the nine auto-stereoscopic display views. A computer vision technique is used to acquire a three-dimensional model of the subject, which is then converted into a depth map for the auto-stereoscopic display. This is a good example of combining market products well to achieve teleconferencing. However, the six system users have little freedom of movement because of the restrictions on where they can sit. Another problem is the two-on-two-on-two limitation.

Many depth cameras such as Kinect [3] have been developed for teleconference systems, and are able to capture the human body as a three-dimensional model. With this technique, however, there is a domain where acquisition of form or depth is difficult theoretically. Therefore, more deterioration will be seen in the reconstructed image than for displayed images taken with common cameras. Moreover, a person's expression will be spoiled.

On the other hand, Otsuka et al. have proposed a system in which a camera is installed for every participant and takes a photograph of that participant, the background is removed so that only the person's image is extracted, and the image is displayed on a back projection type display [4]. A rotary motor is built into a projection screen and it is assumed that rotating the screen dynamically makes it possible to reinforce transfer of the facial direction of a photographic subject according to the direction acquired from the picture. Although image communication systems are currently used to express actual conversation situations, this system creates a new possibility of calling it reinforcing the expression of a person's facial direction by rotating a physical screen. In the system, no image processing other than background removal is added to the pictured person's image, but the displayed images have the image quality of a camera, and the images are not degraded by the rotation of the screen.

In an attempt to solve the problems of achieving sufficient image quality and communicating the facial directions of users, we have developed a system we call "MulDiRoH" (Multi-view Display system for Representation of Humans). It features a novel multi-view display we developed using a Quantized Diffusion Angle (QDA) screen [5].

In this paper we describe the system and our use of a communication game for multiple users to evaluate how well the system conveys a person's facial direction.

2 MulDiRoH

2.1 System Overview

The MulDiRoH system is basically classified as a rear projection display system and uses a QDA screen for the projection screen, three projectors, and three cameras for taking images of remote users. The QDA screen diffuses multiple, different images created by projectors at different directions to quantized diffusion angle regions, yielding multiple, different images in different viewing areas with no limit placed on the distance from the screen. Fig. 2 diagrams a MulDiRoH prototype system and Fig. 3 shows photographic views of the system.

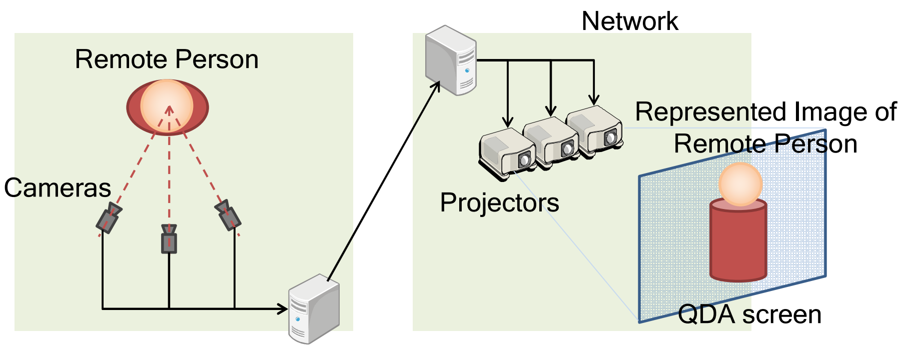

Fig. 2. Diagram of MulDiRoH

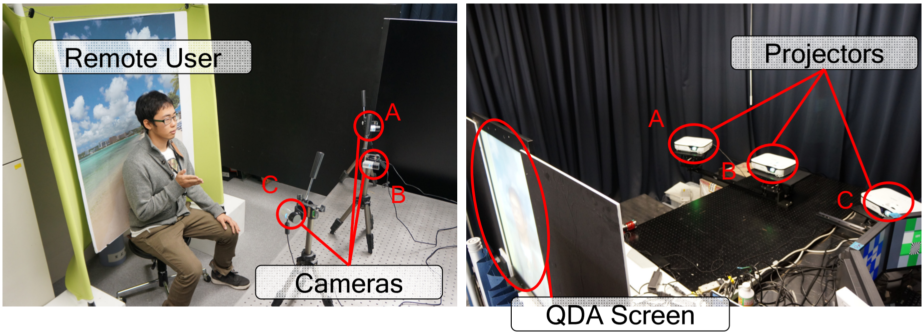

Fig. 3. Photographic Views of MulDiRoH

2.2 QDA Screen

A typical example of a multi-view display is a triple-view LCD using a liquid crystal TFT-panel and a parallax barrier as reported by Takaya [6]. With this method, however, increasing the number of viewing directions beyond four is quite difficult because of the attendant decrease in resolution and the requirement for high alignment accuracy in production. The method also has a very shallow viewing area because of the diamond shape of its individual viewing areas as shown by A, B, and C in Fig. 4. Outside of these diamond-shape areas, a mixture of individual images is likely to generate a cross-talk problem.

In an attempt to solve these problems, we have developed the MulDiRoH system featuring the use of a QDA screen. This screen diffuses the different multiple images created by projectors at different directions to quantized diffusion angle regions,

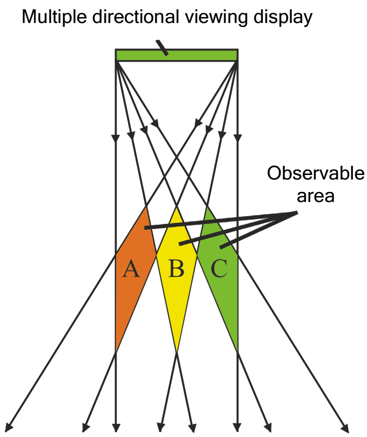

Fig. 4. Problem with diamond-shape areas [5]

which yields multiple (different) images in different viewing areas with no limit placed on the distance from the screen (Fig.5). This type of screen has three special features. First, the number of viewing areas can be increased by changing the lenticular lens parameters. Second, there is no need to install the projector within strict distance limits. This is because the screen's wide range of observable areas enables it to easily provide multi-view displays if the projector is set up within a range outside of that needed to quantize and radiate incident light. Third, its wide observable area enables images to be viewed further from given distances without restriction, though it is necessary for viewers to move somewhat away from the screen to see the images.

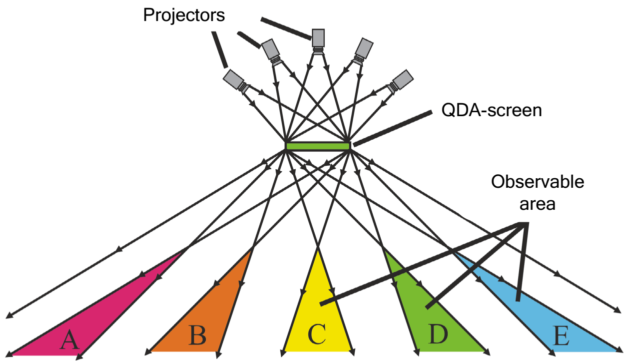

Fig. 5. QDA screen's observable areas [5]

2.3 Perspective Transform Method

All multi-view display systems share a common problem in that users who stand outside of the center area cannot observe geometrically correct images. Therefore, they cannot receive the correct facial direction of remote users. To solve this issue, we propose using the perspective transform method. Using this method, users can observe images of remote users that look as if the latter were standing right at the front of the display. Fig. 6 shows an example result obtained using the perspective transform method. The pictures displayed in a red lattice show the results for the right viewing area. The left side picture shows the original view and the right side one shows that obtained with our method. It can be seen that the pictures' horizontal sides are parallel and their displayed areas are rectangular.

Fig. 7 shows the results of representing a remote user from each of the three viewing areas. In each of the views the user's gaze is directed to the center viewing area. The results confirm that the system can accurately represent a remote user's facial direction. Since this enables the observer to feel that the remote user is looking directly at him/her, it allows the two persons to speak together in a natural manner.

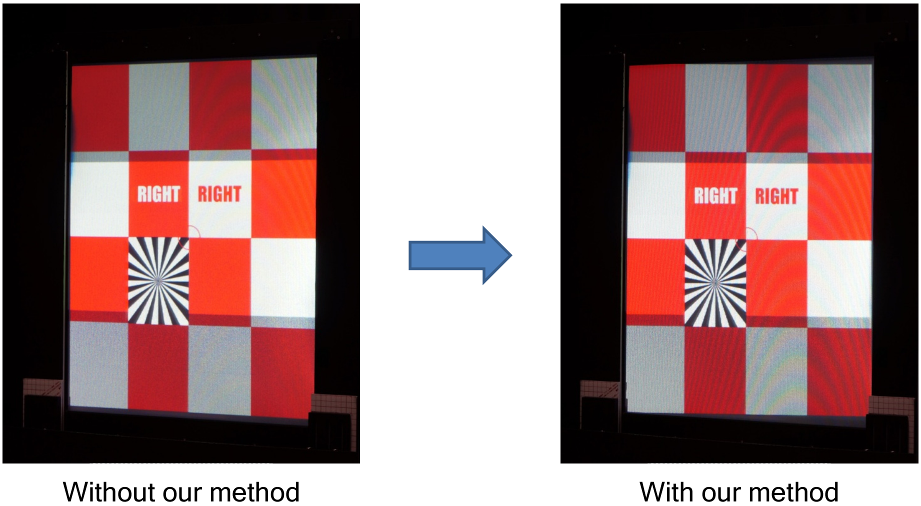

Fig. 6. Example result obtained using the perspective transform method

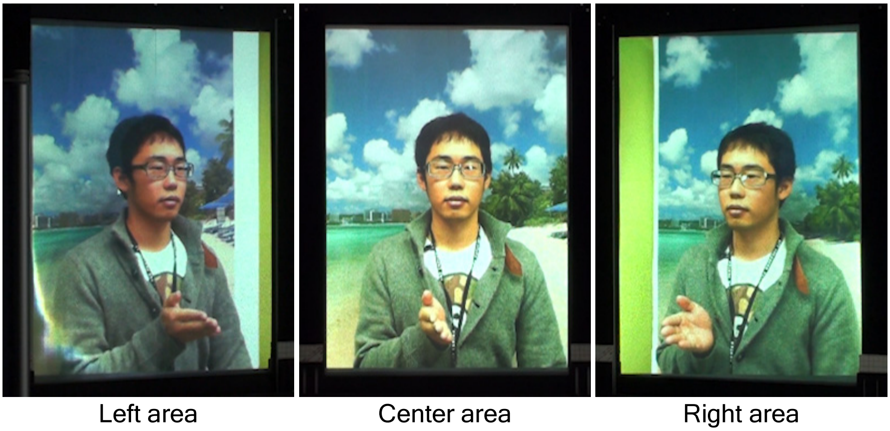

Fig. 7. Views produced by MulDiRoH

3 Experiment

3.1 Design and Experimental Conditions

To evaluate the MulDiRoH system's effectiveness in expressing facial directions, we designed a communication game for users and asked them to play it. In this game, all players have to pay attention to the direction of other player's action while playing. The procedure for the game is illustrated in Fig. 8. A game set/session starts with four users standing in a circle facing each other. One of the users (the “start” player) points to another, who becomes the “first” player (STEP 1). That player then points to another, who becomes the “second” player (STEP 2). Then the two persons flanking the second player make a prearranged pose to end the set (STEP 3). All moves must be made at rapid-fire speed. If one of the players makes a mistake or stops the flow of the game, the set/session is over and the next set/session starts with the player at fault becoming the “start” player. The idea of the game is that all the users have to pay close attention to the pointing gestures of the others. A set/session ends when one of the following occurs:

1. The first player does not point to any second player.
2. A player other than the first player points to a second player.
3. One or both of the players flanking the second player do not make the pose.
4. A player other than the players flanking the second player makes the pose.
5. One of the players is too slow in making a pointing gesture or a pose.

Fig.9 shows the flow of the game. One cycle from STEP1 to STEP3 is defined as a “set,” and a sequence of sets ending with a failure is defined as a “Session.”

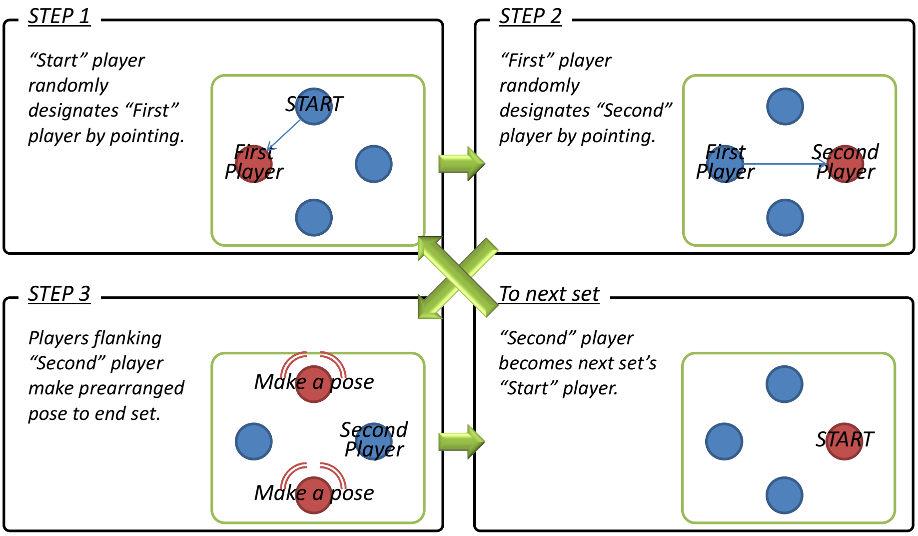

Fig. 8. Game procedure (example)

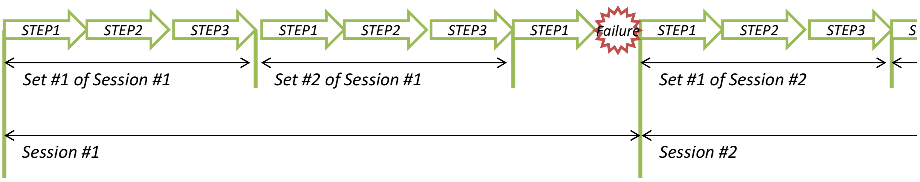

Fig. 9. Flow of the game

We conducted an experiment in which one of the users joined the game from a remote place by using MulDiRoH. Fig. 10 shows a view of users playing the game, with the remote user displayed on the screen. Four players at a time were given the task of playing the game for a three-minute period. They performed this task 12 times, six times using MulDiRoH and six times using a 2D display system for comparison. The remote player A was the "Start" player for each task. A total of 48 persons participated in the experiment.

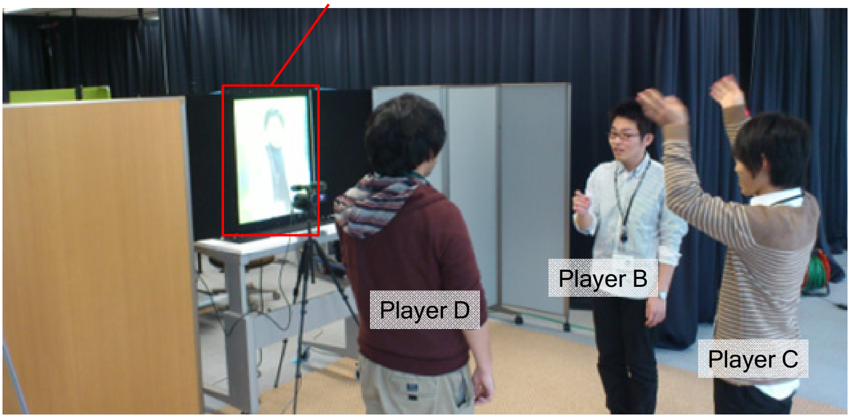

Fig. 10. View of experiment

3.2 Results

Fig. 11 shows quantitative results obtained in the experiment. T-test results for the average number of successful sets and unsuccessful sessions suggested that facial direction information is more easily communicated under MulDiRoH conditions than 2D display conditions.

After the experiment, we asked the subjects to subjectively assess the systems by answering questions about them. Fig. 12 shows the questions and Fig. 13 shows the mean opinion scores for them. The players in the B and D positions were particularly likely to be affected by the differences between the systems, and this was reflected in their scores: 3.9 for MulDiRoH and 2.5 for the 2D display system. This tends to demonstrate the superiority of the MulDiRoH system and it also suggest that the direction expression acts on an understanding in telecommunication. The players in the A and C positions were much less likely to be affected by the differences between the systems; therefore, their scores should not be considered significant for this evaluation.

We also performed an exploratory experiment to gauge reaction time under the 2D and MulDiRoH conditions as compared with that when no display was used. It was found that for the former case the reaction time was about 18% slower. This may be due to an inherent delay in using a display system or a kind of "mental block" on the part of subjects using such a system.

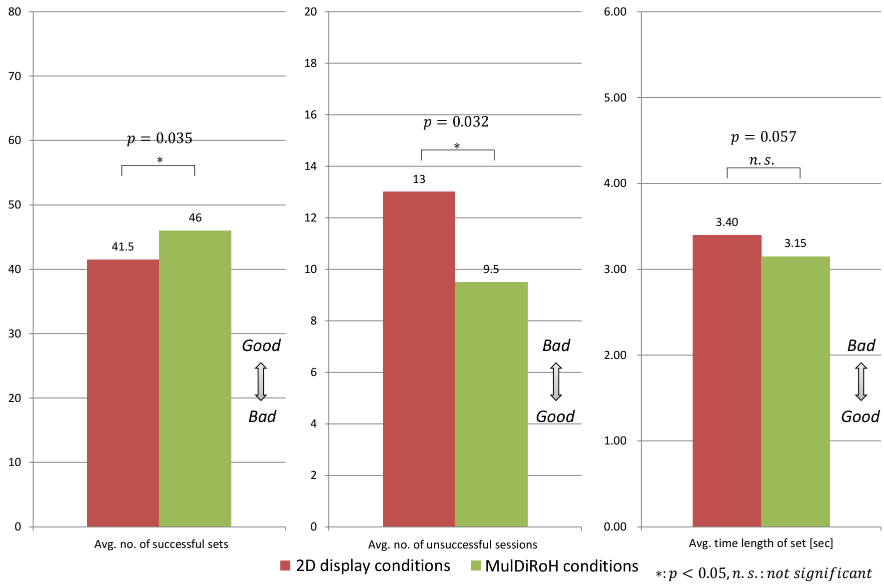

Fig. 11. Quantitative results of experiment

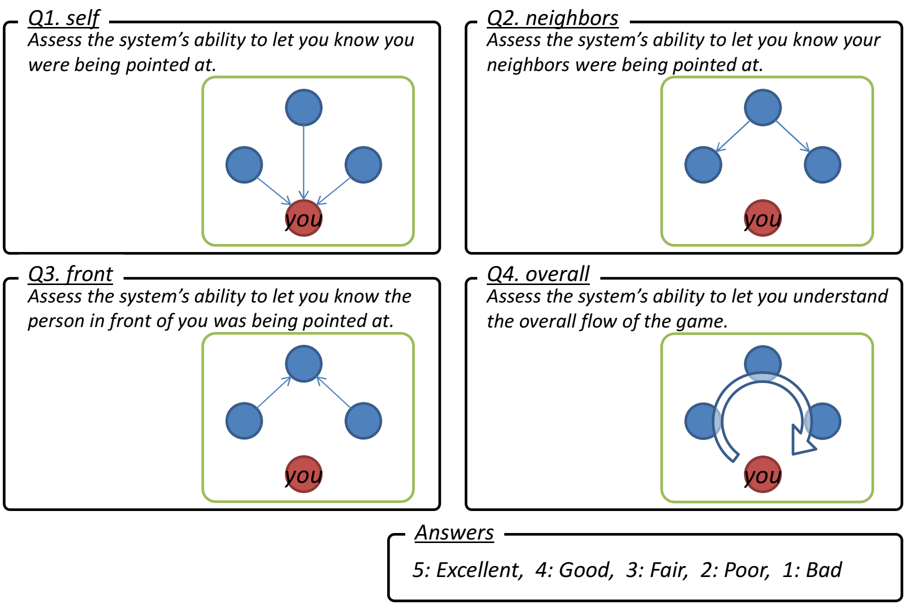

Fig. 12. Subjective assessment questions

We found that the reaction time for the subject in the B position tended to be slower than that for the subject in the D position. Since the conditions were the same for both, we had expected their results to be about the same. The difference may have been due to sunlight streaming into the room and producing an unfavorable field of view for the subject in position B.

The subjects were not informed in advance of the differences among the conditions. Under 2D conditions, however, we observed that some of the subjects were unsure about the manner in which pointing gestures should be made. This did not happen under the MulDiRoH conditions, which suggests that these conditions better enable directions to be transmitted naturally than 2D conditions.

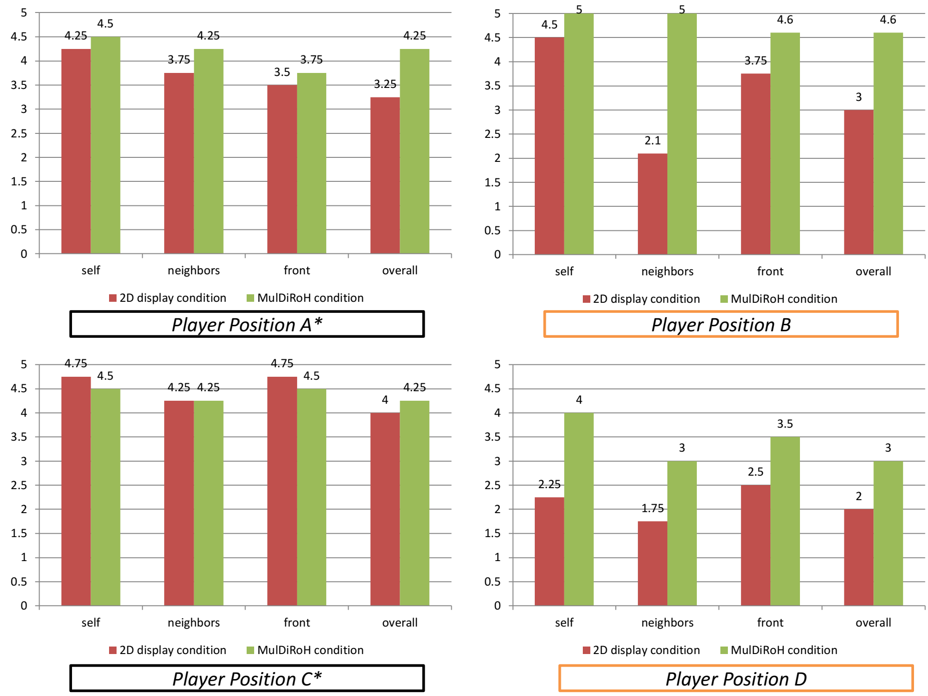

Fig. 13. Mean opinion scores for subjective assessment questions

4 Conclusions

We have developed a human representation system we call "MulDiRoH", which consists of a multi-view display that uses a QDA screen and multiple cameras for video conferencing. The system makes it possible to represent human beings in a remote space with a high degree of presence, because it provides multiple views of facial directions that enable observers to feel that remote users are looking directly at them.

In this paper, we reported an experiment using a communication game designed to evaluate the MulDiRoH system's effectiveness in expressing facial directions.

The results made it clear that giving directivity to images enables communication to progress more smoothly. They also suggested that the communication game progresses more naturally and smoothly under MulDiRoH conditions than the conditions of a 2D display system.

As a subject for future work, we plan to conduct an experiment in simultaneously providing multiple facial directions to a lot of people from persons in a remote place. We will evaluate the results and assess the impact they may have on computer supported cooperative work (CSCW) and on communications in general.

References

1. Jones, A., Lang, M., Fyffe, G., Yu, X., Busch, J., McDowall, I., Bolas, M., Debevec, P.: Achieving eye contact in a one-to-many 3D video teleconferencing system. ACM Trans. Graph. 28(3), Article 64 (2009)
2. Feldmann, I., Waizenegger, W., Atzpadin, N., Schreer, O.: Real-time depth estimation for immersive 3D videoconferencing. In: 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), pp. 1–4, 7–9 (2010)
3. Microsoft Kinect, `http://www.microsoft.com/en-us/kinectforwindows/`
4. Otsuka, K., Mucha, S.K., Kumano, S., Mikami, D., Matsuda, M., Yamato, J.: A System for Reconstructing Multiparty Conversation Field based on Augmented Head Motion by Dynamic Projection. In: Proceedings of ACM Multimedia 2011, pp. 763–764 (2011)
5. Kawakami, T., Katagiri, B., Ishinabe, T., Uchida, T.: Multiple Directional Viewing Projection Display Based on the Incident-Angle-Independent Diffusion-Angle-Quantizing Technology. In: Proc. IDW 2010, pp. 1479–1482 (2010)
6. Takaya, T.: A Dual Directional Viewing LCD and A Triple Directional Viewing LCD. Sharp Technical Journal 96, 21–23 (2007)