MoPaCo: Pseudo 3D Video Communication System

Ryo Ishii, Shiro Ozawa, Takafumi Mukouchi, and Norihiko Matsuura

NTT Cyber Space laboratories, NTT Corporation,
1-1, Hikari-no-oka, Yokosuka-Shi, Kanagawa 239-0847, Japan
{ishii.ryo,ozawa.shiro,mukouchi.takafumi,
matsuura.norihiko}@lab.ntt.co.jp

Abstract. We propose a pseudo 3D video communication system that imparts motion parallax which adjusts to the viewpoint position of a user and enables the user to view video pictures in which depth can be perceived with an ordinary equipment setup, namely a monocular camera and a 2D display. We have implemented the system and evaluation experiment results with it showed that its imparting of motion parallax allows it to represent distances that closely reflect actual face-to-face situations better than 2D video can. In addition, subjective evaluations confirmed that motion parallax gives users the feeling that the conversational partner is actually present and makes it easier for them to comprehend the positional relationship of the conversational partner in space.

Keywords: Video communication, motion parallax, depth perception, inter-personal distance.

1 Introduction

Our aim is to implement a video communication system which not only enables ordinary conversation, but which also provides a highly realistic feeling that enables users to work together in a natural manner while sharing a mutual space and observing each other, in a manner that is equivalent to actually being face-to-face. For this reason, it is important to give the user the feeling that the conversational partner is in front of him and have natural transmission of nonverbal information such as inter-personal distances, gaze, and pointing gestures, using a body that shares a mutual space with the conversational partner. An ordinary video communication system that uses a 2D display to show video footage which has been captured with a monocular camera lacks most of the depth information about the conversational partner and the space around him. In addition, the visibility of the conversational partner and background does not change to match the changing viewpoint when the user moves. For this reason, in addition to the problems of a feeling of spatial separation from the conversational partner [1, 2] and a lack of feeling of the presence of the conversational partner [3], there is a huge problem in that nonverbal information associated with depth (such as inter-personal distance and pointing gestures) cannot be transmitted correctly. In order to transmit such information, it is necessary to impart depth information within the video footage, and there are currently many research projects into tackling these challenges. It is known that

G. Salvendy, M.J. Smith (Eds.): Human Interface, Part II, HCII 2011, LNCS 6772, pp. 131–140, 2011.

motion parallax and binocular parallax are major cues that enable humans to sense depth at close distances [4]. Of these, binocular parallax can be used comparatively simply, so several methods have been proposed for using binocular parallax to represent depth by providing a stereoscopic view of the conversational partner [5]. To present correct depth information naturally in video communication, however, it is absolutely essential to present video that has motion parallax that corresponds to the observation position of the user.

In this study, we propose a method that implements video communication that is associated with motion parallax with a simple setup. Since this method is intended for systems that are generally available, it was implemented with a configuration similar to that of an existing video communication system, namely a monocular camera and a 2D display. In this paper, we also report on evaluation experiments using the thus-implemented system, by which we proved experimentally the perception of depth towards the conversational partner due to motion parallax, impressions relating to the conversation video, and the accuracy with which pointing actions that follow pointing gestures are transmitted, and thus report on the effectiveness of motion parallax.

2 Related Work

A video expression method that enables users to depth perspective using motion parallax alone has been proposed. Suenaga [6] proposed a 3-dimension perspective display with motion parallax. The method of display is to project video pictures corresponding to the user's viewpoint position tracked by stereo cameras onto a 2D display. However, the system provides only CG model contents. That is, the system cannot generate photographed video in real time. The objectives of this study are to propose a new method of implementing motion parallax by using only a monocular camera in real time, and also to verify the effect on the transmission of information about nonverbal behavior by motion parallax and impressions of the resultant video.

3 Proposed MoPaCo System

3.1 Overview of MoPaCo System

In this study, we propose a real-time video communication system called MoPaCo (which is a contraction of "Motion Parallax Communication") which implements motion parallax although using only a monocular camera. Fig. 1 shows images of motion parallax video representations by MoPaCo of a conversational partner which correspond to the viewpoint positions of different users. The display for a user who is at some distance from the conversational partner in the video can give the user the feeling they are linked as if through a window. It is thought that this motion parallax video representation will eliminate spatial separation, improve the feeling of presence

of the conversational partner, and enable the transmission of nonverbal information that is associated with depth by imparting depth information to video pictures.

To present a motion parallax video of a conversational partner on a 2D display so as to correspond to the viewpoint positions of different users, the following processes are necessary:

(1) Measurement of each user's viewpoint position
(2) Construction of a 3D space having information on the dimensions and positional relationships of the people and the background, based on information obtained from a camera or other means
(3) Rendering of the 3D space constructed in step (2) on a 2D display, to correspond to each user's viewpoint position that was obtained in step (1).

We propose the technique to perform the steps (1) and (2) using only a monocular camera and implemented the MoPaCo system using a monocular camera. In this section, we describe the details for each process.

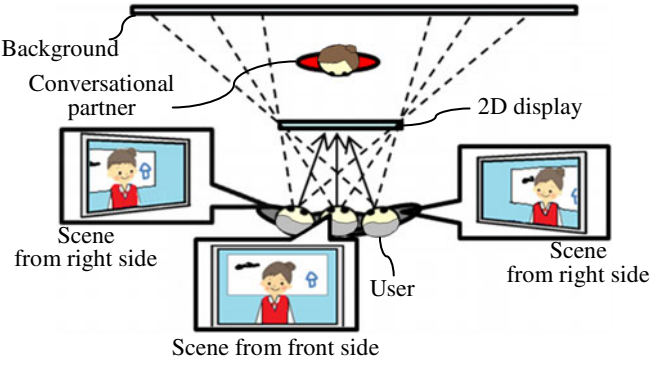

Fig. 1. Concept images of video representations caused by motion parallax

3.2 Measurement of User's Viewpoint

We propose the use of only a monocular camera to detect each user's viewpoint. Before calculating the 3D position from parts information (coordinate position) of each face in the 2D image, the system measures the eye separation distance of each user, as preprocessing. It then acquires the distance of each user from the camera, using the Depth from Focus function that is used for achieving focus in ordinary cameras. During this process, template matching is performed on the image captured from the camera to measure the positions of both eyes (2D coordinates within the image) and the orientation of the head. The system calculates the eye separation distance of each user from the distance from the user to the camera, information measured from the image, and the angle of view and resolution of the camera.

Based on this information, real-time capture is started and the system obtains the positions of both eyes (2D coordinates within the image) and the orientation of the head from the captured image, and calculates the viewpoint position z of that user from the camera from there at that time. Note that the x- and y-coordinates are calculated from the 2D coordinates within the image and the pixel pitch of the image. Based on this information, real-time capture is started and the system obtains the positions of both eyes (2D coordinates within the image) and the orientation of the head from the captured image, and calculates the viewpoint position z of that user from the camera from there at that time. Note that the x- and y-coordinates are calculated from the 2D coordinates within the image and the pixel pitch of the image.

3.3 Construction of a 3D Space

In this study we propose a method of constructing 3D information with respect to an image captured from a single camera. This is done by performing background difference processing using background information that was acquired beforehand, maintaining the 2D plane and dividing it into a personal area and a background area, and creating a multi-layer structure with those areas arranged as layers in accordance with their depth-wise positions. The use of 2D images ensures that a high-resolution display is possible. In addition, if only the background difference is subjected to image processing, the processing costs are low and thus real-time processing is possible.

Using the person and background images, the system then generates multiple layers having a distance relationship from the camera, at full size (in the rest of this paper, the layers that use the person image and the background image are called the person layer and the background layer). The method of calculating distance information measures for the background layer beforehand is by the depth-from-focus method provided in the autofocus function of the camera, when the background difference image is acquired. For the person layer, the user viewpoint position is used. These distances become distance information from the camera of the person layer and the background layer, respectively. Based on this distance information, the system uses Equation (1) to calculate the full size (width $wi \times$ height hi) of each layer i from the thus-acquired distance information di and the camera's angle of view (width θw, height θh). This procedure configures a 3D space having full size and position information.

$$w_i = 2*d_i*\tan\theta_w/2,\ h_i = 2*d_h*\tan\theta_h/2 \tag{1}$$

3.4 Rendering of 3D Space Based on User's Viewpoint

As shown in Fig. 2, the person layer and background layer generated by the 3D spatial information module are projected in perspective to match the viewpoint position of the user, using the 2D display as a projection surface. Thus motion parallax video is implemented.

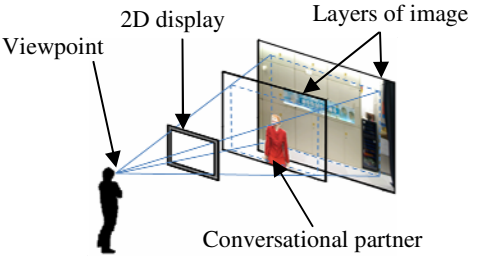

Fig. 2. The person layer and background layer are projected

3.5 Results of Implementation

Using the proposed method described above, we implemented the MoPaCo system which enables bidirectional viewing of motion parallax video in real time. The development environment was a computer with Intel Core i7 Extreme 980X as the CPU, 12 GB of memory, and a NVIDA GeForce GTX480 graphics board. The results of this implementation are shown in Table 1. The "Reflection of conversational partner's image" in this table is that the conversational partner's video appears in the user's display from the captured video, and "Reflection of motion parallax" motion parallax is the appearance in the video from a movement of the user's viewpoint position. The table also shows frame rate and response (time lag) in each part. The frame rate is sufficiently high. In contrast, response is somewhat late in arriving.

Table 1. Implementation results

Captured image size		1280×780 (HD)	1920×1080 (Full HD)
Reflection of conversational partner's image	Frame rate	30 fps	18 fps
	Response	260 ms	330 ms
Reflection of motion parallax	Frame rate	30 fps	
	Response	300 ms	

Fig. 3 shows a scene as observed in 2D video, motion parallax video with MoPaCo, and actual face-to-face (REAL) situations. In comparison with the face-to-face condition, there was no parallax in the video in the 2D condition, even if the user's head moved, so the human dimensions and positional relationships did not match. With the MoPaCo condition, in contrast, the dimensions and positional relationship between the person and background were reproduced in the video.

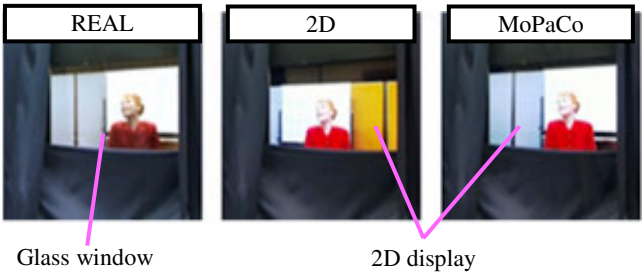

Fig. 3. Scenes for three observational conditions: actual face-to-face (REAL), 2D video, and motion parallax video with MoPaCo

4 Experiments

4.1 Experiment Procedure

Using the MoPaCo system, we conducted experiments with test subjects, with the objective of verifying the transmission by motion parallax video representation of inter-personal distances, which are particularly important in nonverbal information, and the effect of motion parallax on increasing the feeling of a partner's presence, the positional awareness of the partner's space and spatial comprehension, etc. In the experiments, the position of the conversational partner was varied in face-to-face conversational situations, conventional 2D video, and video with motion parallax. After each subject had observed a situation, they gave their evaluations. There were a total of 18 experimental conditions, consisting of combinations of three observation methods and six distances of the conversational partner, as shown in Table 2.

Table 2. Implementation results

Observation conditions (3 conditions)
• ***REAL condition***: Observing the conversational partner through a glass window
• ***2D condition***: Observing the conversational partner in an image displayed on a 2D display (in this case, the user's viewpoint position is where the image is displayed at a position when the user is sitting straight in the chair)
• ***Motion Parallax with MoPaCo (MP) condition***: Observing the conversational partner in an image having motion parallax, which is shown on a 2D display
Distance to the observer (6 conditions)
• The mannequin's position is set to *0, 20, 40, 80, 120, and 200 cm* from the partition (display plane), in other words, *150, 170, 190, 230, 270, and 350 cm* from the subject.

To measure the perceived distance to the conversational partner in the verification of the effect of inter-personal distance representation, we adopted a subjective evaluation by which subjects gave the distance from their chest to the chest of the conversational partner by a scale-less string length. Asking the subjects to answer

with the length of a piece of string is designed to eliminate any personal differences in the subjective criterion, in comparison with methods in which answers are numerical values. In this case, if the string length given as the answer under the REAL condition is the same as the length under the MP condition, we can say that motion parallax video makes it possible to perceive depth to the conversational partner in a similar manner to actual face-to-face conversation.

We obtained the subjects' impressions of motion parallax by asking them to evaluate the seven items listed in Table 3, such as the feeling that the conversational partner is present, using a six-step Likert scale (1-6 points, six being best) on a questionnaire sheet for the 2D and MP conditions.

Table 3. Subjective evaluation items

• ***Ease of viewing video***: Was the video of the conversational partner easy to see?
• ***Stereoscopic effect***: Did you feel that the conversational partner's space was in 3D?
• ***Intuition of distance comprehension***: Could you intuitively grasp the feeling of depth between you and the conversational partner?
• ***Presence of conversational partner***: Did you feel that the conversational partner was present?
• ***Face-to-face feeling***: Did you feel that you had actually met the conversational partner?
• ***Ease of spatial comprehension***: Was it easy to understand the positional relationship between the conversational partner and the background?
• ***Feeling through window***: Did you feel that you had met the conversational partner through a window via the display?

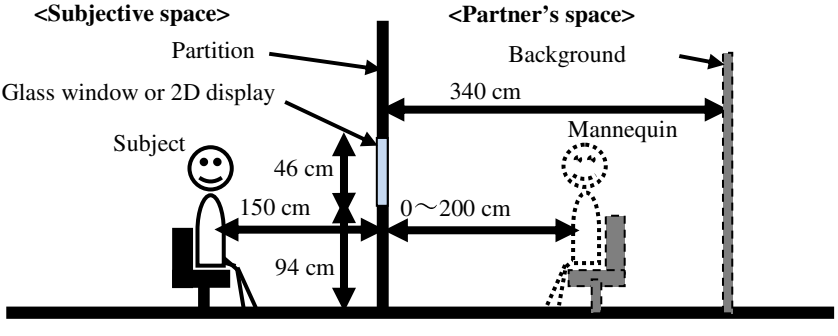

Fig. 4. Experimental equipment

The experimental setup was such that the subject was seated on a chair at a position 150 cm from a partition in which a glass window was installed, as shown in Fig. 4, and was able to observe the space of the conversational partner through the glass window and display. Note that with the 2D condition and the MP condition, the display was installed directly behind the glass window. The size of the glass window (46 cm high x 80 cm wide) was less than that of the display, so the edges of the display were not visible from the subjects. Note that we used a humanoid mannequin

in place of an actual person, in order to ensure that the visual stimulus of the conversational partner was uniform.

Before the experiment, each subject observed the actual mannequin and verified its size. Taking the effect of the order of trials into consideration, the sequence of trials was performed under randomized conditions. Subjects were asked to sit in the chair and raise their head upon hearing the start signal, then observe the mannequin while remaining seated in the chair but moving their head freely. No time limits were set, but all the subjects finished their observations within ten seconds and then made their evaluations. They made their perceived distance evaluations after undergoing a trial for each of the experimental conditions. In contrast, they made their video impression evaluations after going through all the experimental conditions.

4.2 Experiment Results for Perceived Distance

Experiments were conducted using ten subjects (nine males, one female, age range 20-60). We first show the results obtained for perspective distance in Fig. 5. The figure shows the average values of string lengths measured for all subjects. The actual distance from the subject to the conversational partner is plotted along the horizontal axis and the average value of the string length (perceived distance) is plotted along the vertical axis. The dotted line shows the values when the actual position of the conversational partner matches the string distance. It was clear that perceived distances given as string lengths were much smaller than the actual distances (the dotted line). Significant differences were seen overall, in that the string lengths became shorter in the sequence for the REAL → MP → 2D condition. In other words, it was confirmed that the MP condition is better able than the 2D condition to represent inter-personal distance in a manner close to that of the REAL condition. However, because a difference in depth perception exists between the MP and REAL conditions, we think that it is necessary to define the relationship between depth perception and motion parallax video, and to determine how distance in the MP condition can be represented accurately in the same way as it is in the REAL condition.

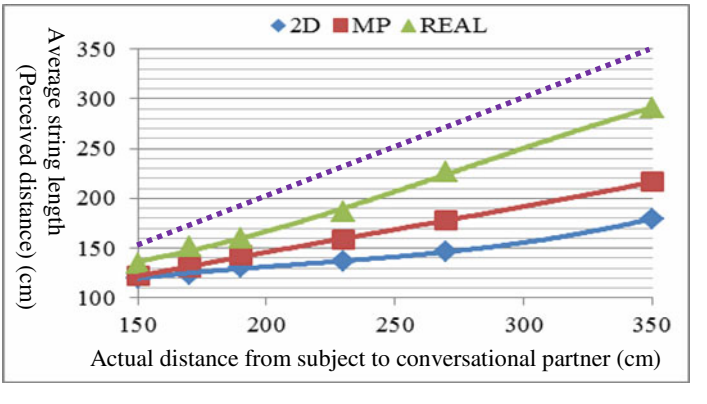

Fig. 5. Experiment results

4.3 Experiment Results for Video Impression

Fig. 6 shows the average values of the subjective evaluation scores for the questions that 10 subjects were asked to answer on the questionnaire. In this experiment, we performed paired t-testing for each evaluation item. These results are also shown in the graphs. The "Ease of viewing video" in the figure caused us some concern, in that since movements under the MP conditions result in large changes in the video, drunkenness or other abnormalities might not be noticed. However, the existence of such a problem could not be confirmed. We consider that the MoPaCo feature of imparting motion parallax to video pictures in accordance with viewpoint position will give humans a natural visual effect. This is supported by results showing that MP was evaluated significantly higher than 2D in two paired T-tests for the six evaluation factors other than "Ease of viewing video" (Stereoscopic effect : $t(9)=2.68$, $.10<p<.05$, Intuition of distance comprehension : $t(9)=2.03$, $p<.01$, Presence of conversational partner: $t(9)=5.51$, $p<.01$, Actual face-to-face feeling : $t(9)=4.00$, $p<.05$, Ease of spatial comprehension: $t(9)=7.39$, $p<.05$, Feeling through window: $t(9)=3.21$, $p<.01$). These results indicate that motion parallax video enables users to feel improved depth perception and enhances their feeling of distance between users more than 2D video does. It is also very interesting that motion parallax enhances face-to-face feeling and actual partner's presence. We believe this is the reason that motion parallax is better than 2D at giving users the same visual effects and depth cues. For the item "Ease of spatial comprehension" MoPaCo scored higher than 2D, indicating that motion parallax is effective in situations where users can observe their partner's space. Finally, in the "Feeling through window" item, the MPC score was more than five times higher than the 2D score, indicating that MoPaCo users may feel that they are seeing their partner and their partner's space through an actual window. This suggests the possibility that users can naturally feel the composition surface of each space and recognize the region in their own space that the partner is observing.

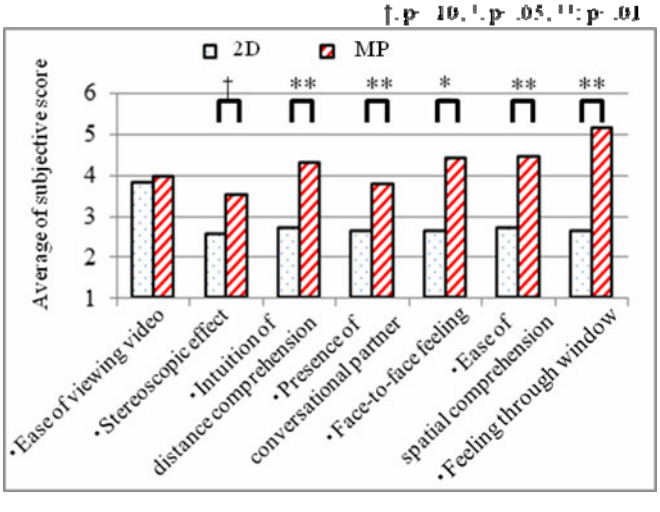

Fig. 6. Impression evaluation results for motion parallax video

5 Conclusion

We have proposed a pseudo 3D video communication system that imparts motion parallax which adjusts to the viewpoint position of a user and enables the user to view video pictures in which depth can be perceived with an ordinary equipment setup, namely a monocular camera and a 2D display. The system was implemented and evaluation experiment results with it showed that its imparting of motion parallax enables it to represent distances that closely reflect actual face-to-face situations better than 2D video can. In addition, it has been confirmed from subjective evaluations that motion parallax gives the user a feeling as if the conversational partner is actually present and makes it easier for him/her to comprehend the positional relationship of the conversational partner in space. In the future, we plan to conduct real-time conversations to further confirm the effectiveness of using motion parallax in video communication.

References

1. Williams, E.: Coalition formation over telecommunications media. European Journal of Social Psychology 5, 503–507 (1975)
2. Strickland, L.H., Guild, P.D., Barefoot, J., Paterson, S.A.: Teleconferencing and leadership emergence. Human Relation 31, 583–596 (1978)
3. Heath, C., et al.: Disembodied conduct: communication through video in a multimedia environment. In: Proc. CHI 1991, pp. 99–103. ACM Press, NY (1991)
4. Cutting, J., Vishton, P.: Perceiving Layout and Knowing Distances. In: Epstein, W., Rogers, S. (eds.) Perception of Space and Motion, pp. 69–117. Academic Press, New York (1995)
5. Towles, H., Chen, W.C., Yang, R., Kum, S.U., Fuchs, H., Kelshikar, N., Mulligan, J., Daniilidis, K., Holden, L., Zeleznik, B., Sadagic, A., Lanier, J.: 3DTele-Immersion Over Internet 2. In: ITP 2002 (2002)
6. Tsuyoshi, S., Yoshio, M., Tsukasa, O.: 3D Display Based on Motion Parallax Using Non-contact 3D Measurement of Head Position. In: Proceedings of OZCHI 2005 (2005)