ASIAGRAPH 2009 PROCEEDINGS

ビデオコミュニケーションシステムにおいて遠隔ジェスチャ認識を支援する映像効果

A Method for Supporting Recognition of Remote Gestures in a Video Communication System

梶克彦/NTT コミュニケーション科学基礎研究所，山下直美/NTT コミュニケーション科学基礎研究所，
平田圭二/NTT コミュニケーション科学基礎研究所

Katsuhiko Kaji[1]/NTT Communication Science Laboratories, Naomi Yamashita[2]/NTT Communication Science Laboratories
Keiji Hirata[3]/NTT Communication Science Laboratories
*[1]kaji@cslab.kecl.ntt.co.jp, *[2]naomiy@acm.org, *[3]hirata@brl.ntt.co.jp

Abstract: The 3D structure of people's body movements and gestures become distorted when a sequence of those movements and gestures is captured by multiple cameras and displayed on multiple 2D screens at a remote site. This increases the chances of remote people losing sight of those gestures, possibly making it difficult to understand the meaning intended by the gestures. To alleviate such problems, we propose a visual augmentation technique, called "remote lag," that applies the concept of telepointer traces to the bodily gestures performed in a videoconferencing system. Remote lag is a visual effect that overlays a user's past motion image onto his/her present image. In our experiment, we compared several types of remote lag and examined their effects on how accurately a user can perceive a remote user's pointing location. Our overall results show that when a user misses gestures, remote lag contributes to the recovery of a gesture context. In particular, when a remote user stands right beside a user, a particular type of remote lag is helpful.

Keywords: Remote lag, distortion, videoconferencing system, lagged image, optical flow

1. Introduction

Video conveys a limited amount of information on the three-dimensional structure of remote scenes, and thus limits exploration, inspection, and peripheral awareness [2].

Gesturing in videoconferencing systems (VCSs) is often difficult because various levels of invisibility occur in remote gestures. For example, when user P glances at remote user Q performing a series of gestures across the gap spanning split screens, P may miss Q's gesture in a video-mediated space, preventing P from predicting Q's behavior with consistency. Since the gestures are inherently situated in the context of coordination, P has difficulty in recovering the missing context of coordination. As another example, when P does not face the front of the Q's image but gives Q's image a sidelong look, P inevitably sees a distorted image of Q. In this situation, P may misunderstand the direction to which Q is oriented and the object to which Q points.

As one of the earliest work, Gaver et al. reported that a multiple target video (MTV) system causes several invisibility problems in the understanding of remote gestures, some of which originate from the inadequacy of views and the different views provided by different cameras [3]. Kuzuoka et al. pointed out that an extravagant speaker's gestures displayed on a flat and small video monitor invokes disembodiment [7]. Subsequently, Agora was developed to achieve a roundtable meeting metaphor with two 60-inch screens set along two sides of each desk, which ensured a comprehensive perspective.

Along the lines of the previous work, we are developing a videoconferencing system, called t-Room [6], which is an attempt to alleviate the invisibility problems. In the t-Room, the space shared by remote and local users is created by

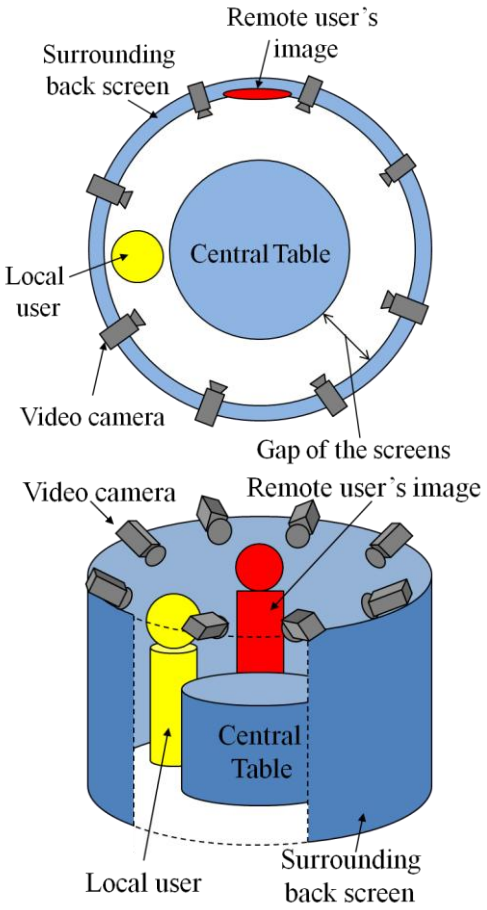

Figure 1: Arrangement of 2D screens, cameras, and users in a t-Room

installing multiple screens and video cameras so that the screens facing inside are arranged surrounding the space and the video cameras capture users standing in front of the opposite screens (Fig. 1). By seamlessly arranging the screens, we can construct the equivalent of a single continuous surrounding back screen, which ensures that users share a truly whole view angle, since whatever image object user P looks at in the local t-Room can be seen by all other users in the local and remote t-Rooms. Accordingly, in a sense, there is no invisibility in t-Room.

2. Problem

The fact that wherever users stand in the t-Room, they can share the whole perspective means that users can freely move within the t-Room. Consequently, another level of invisibility arises. If the image of remote user Q walks toward local user P and stops close to P, P may miss Q's image conveying his/her gestures displayed behind or around P. That is, while remote and local users standing opposite to each other in the t-Rooms can share the proper perspectives, the nearer they are to each other, the bigger the invisible, inconceivable areas become. Moreover, while the remote and local users standing opposite each other can also share the proper directionality, the nearer they are, the bigger the deviation of directionality between them becomes.

Both to improve users' convenience and naturally restrict users' standing positions within appropriate areas, we may introduce a central table providing a shared work space. In this case, unfortunately, a new level of invisibility arises due to the gap between the shared workspace on the central table and the surrounding back screens (Fig. 1). However, it is difficult to eliminate this gap, since users need to stand and move within the t-Room. That is, if the central table and the surrounding screens were placed closely side-by-side to avoid the gap, there would be no space for users to stand and move within the t-Room.

Recognition of remote gestures becomes difficult by a synergetic effect of the following two problems. One is the fact that remote user Q is behind local user P. Another is the invisibilities caused by the gap in screens. When remote user Q is close to local user P and Q's gesture is across the central table and the surrounding back screens, P may frequently miss the gestures. Such situation is possible to occur when screens of the VCS are arranged in three dimensions to display the remote images and the VCS has high flexibility of physical relationship between the user and the screens.

How can the gesture invisibility problem be mitigated? The first approach is to rearrange the screens and table as close as possible even though they are not completely not surrounding the virtual shared space or placed side-by-side, taking into account the requirements, tasks and applications to be carried out in a usable system. The second approach is to augment screen images of the gestures, employing users' past movements. These two approaches are actually complementary to each other. The second one is from solid previous work (i.e., telepointer traces [5], phosphor [1]) and could be more generally applied. Telepointer traces are

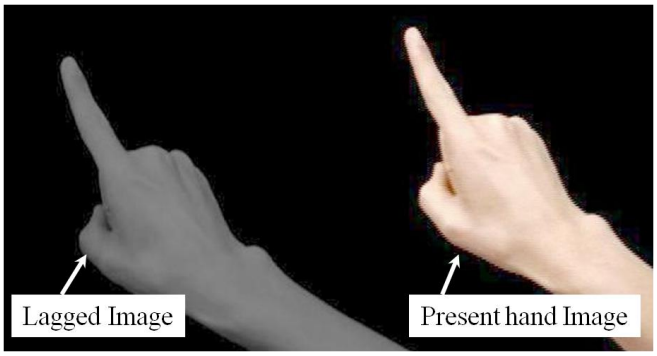

Figure 2: The realistically colored image on the right-hand side is an image of a present remote user's hand. The gray image on the left-hand side is lagged image of 1100 ms behind.

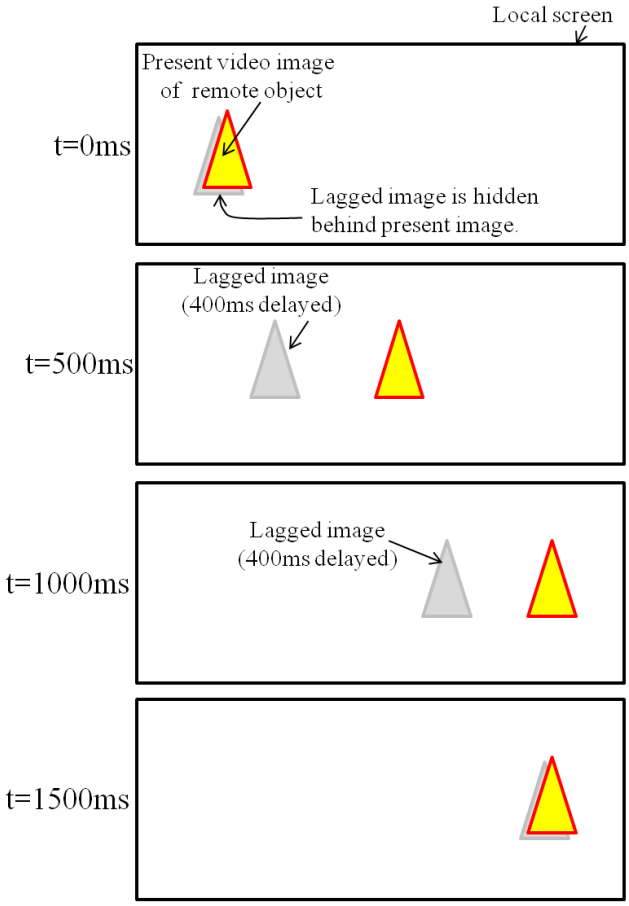

Figure 3: Lagged image starts to chase the remote object 400 ms after it moves from left to right,

visualizations of the previous motion and location of a remote mouse cursor. We share similar goals with Gutwin which are to make gestures easier to see, to make motion easier to interpret, and to provide context that helps people understand others' behavior.

To acquire a correct interpretation of a series of bodily gestures across the gap spanning split screens performed by a remote user standing close to a local user, this paper investigates a visual augmentation based on the telepointer trace technique. The organization of this paper is as follows.

First we present a new visual augmentation method in a VCS, called remote lag, that overlays past movements onto present images. Then we conduct an experiment using the t-Room system that observes a user creating gestures of pointing and bodily movement. The experimental results indicate that remote lag is effective in supporting the understanding of pointing gestures, especially in a situation where a remote user stands close to a local user. Finally, we discuss the applicability of remote lag for improving the usability of shared workspaces in a broader sense.

3. Remote lag

We propose a visual augmentation technique, remote lag, that applies telepointer traces to bodily gestures including pointing, moving, face/body orientating, and shape changes. Accordingly, remote lag is the visualization of the user's past motion to be overlaid onto his/her present image. The several possible representations for telepointer traces proposed by Gutwin [4] include motion line, motion blur, stutter blur, and their combinations. Gutwin suggested that we could design useful representations and choose the best one from them and their combinations [5].

3.1. Lagged image and Motion Flow

We introduce two representations of remote lag: lagged image and lagged image plus motion flow. Lagged image is overlaying the video image(s) lagging for constant time(s) on the present video image (Fig. 2). Figure 3 represents a sequence in which a remote object moves from left to right. At the start of the scene, the remote object remains stationary so that the lagged image is hidden behind the present image. The lagged image starts to chase the present image 400 ms after it moves from left to right. If user A misses user B's gesture, lagged image enables A to watch an instant playback of this gesture *in situ*, which facilitates recovering the missing context of coordination.

Motion flow means motion lines drawn by optical flow (Fig. 4). Between the present image and lagged image, a set of broken line segments is drawn, and these have fading trails like motion blur. In the figure, we use the function for calculating optical flow provided by OpenCV, which contains an algorithm for interest-point detection. Figure 5 represents a sequence in which a remote object moves like a sine curve. The local user can predict sine-curved movement of the lagged image's next moment by seeing rounded motion flow. Since motion flow expresses dynamic movements of the screen images of people and physical objects, if such screen images are missed by the user or disappear, a user can trace lagged image plus motion flow as an afterimage. Accordingly, a user can correctly recognize the other's movements, although he/she is in a constant time behind real time.

3.2. Parameters of the Remote lag

The screen images of a conventional VCS are especially different from the telepointer image in terms of shape complexity and their time-varying nature. Since a telepointer is a cursor, its shape is simple and persistently

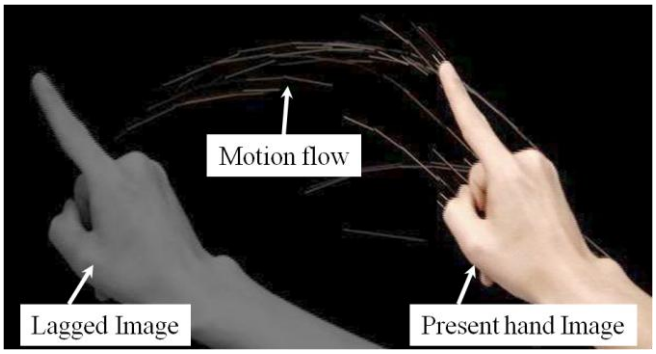

Figure 4: Motion flow is displayed between present image and lagged image to represent trajectory of the remote object.

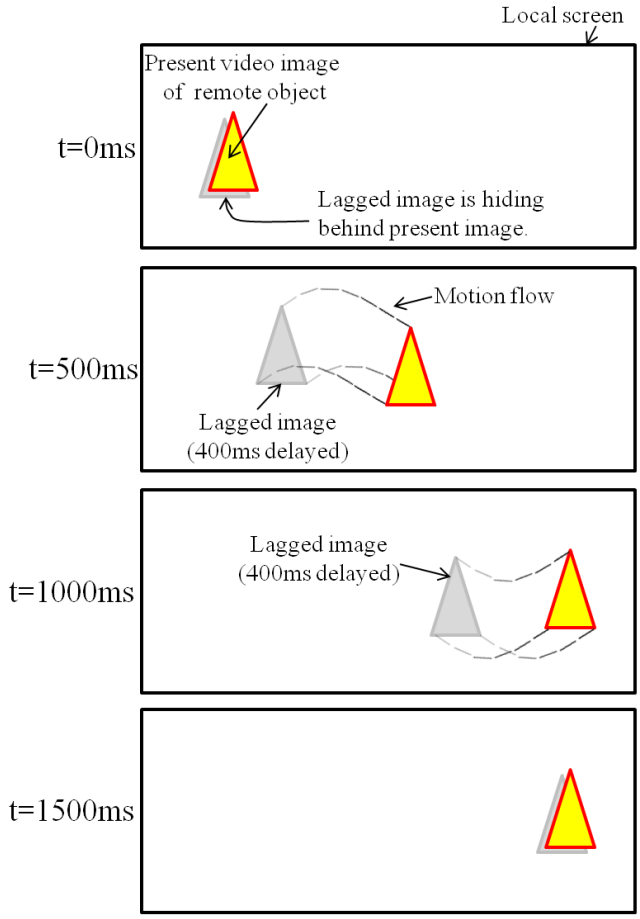

Figure 5: Motion flow connects the present image and the lagged image. When a user looks at the motion flow, he/she can predict the rounded movement of the lagged image.

unchanged. In contrast, since the screen images in a VCS are often 2D images of real people and physical objects, the shapes of these images are often complicated and changing as communication proceeds. Therefore, it is difficult to straightforwardly apply the techniques of motion line and/or motion blur to a VCS image. Prior to the experiment presented in this paper, we examined several representations of remote lag in a practical use of the t-Room and,

consequently, designed two representations for our purpose, with appropriate parameter settings for time interval of lagged image, the number of lagged images, the color and contrast and so on.

Multiple images of a pointer can be displayed along a trajectory in telepointer traces because the pointer image is small. On the other hand, the size of the images in VCS is larger than a pointer image in many cases. Therefore, to prevent a workspace from becoming cluttered with large lagged images, in the experiment described later, a single lagged image is overlaid. The lagged image is overlapped under the present image to hide the lagged image at motionless parts. Additionally, the lagged image is displayed in gray-scale to more easily judge whether the image is the present one or the lagged one. Through a preliminary experiment, we consider 400 ms and 1100 ms appropriate as the time interval between present image and lagged image.

4. Experiment

4. 1 Experimental Design

To examine the effect of remote lag on participants' ability to recover the missing context of coordination, we employ a task in which a participant perceives a remote director's pointing locations. We compared five conditions of remote lag: (a) without remote lag; (b) with lagged image of 400 ms behind the current position; (c) with lagged image of 400 ms behind the current position and motion flow; (d) with lagged image of 1100 ms behind the current position; and (e) with lagged image of 1100 ms behind the current position and motion flow.

Twenty-five participants took part in the experiment. As an experimental design, we used a within-subjects design where each participant completed five similar tasks, one in each of the five conditions stated above. Throughout the five tasks, each participant stood at a fixed location inside the t-Room located in Kyoto (Fig. 6). Two directors inside a remote t-Room located in Atsugi alternately pointed at one of the figures projected on the walls or table, and each participant was asked to identify the pointed figures on a set of answer sheets. Conditions were counterbalanced for order.

To eliminate the effects of network delay and the difference in directors' movements among the participants, we made use of t-Room's record & play function. Accordingly, we recorded the directors' pointing behavior in the Atsugi t-Room and played it in the Kyoto t-Room iteratively throughout the experiment so that all of the participants were subject to the same network delay and the same pointing behavior. Each task consisted of 32 questions, which were randomly assigned to the two directors. Each question issued by one of the two directors consisted of a sequence of 1 to 3 consecutive pointing(s).

4. 2. Categories of Standing Position

We further categorized the questions based on the director's standing position. This is because the standing position of a director changes the view of the participants, introducing different types and levels of distortions (i.e., the discontinuity of a 3D space, the deviation of directionality, magnification/reduction of images sizes and distances), which may lead to different effects of remote lag. The position where directors issued each question varied from (I) right beside the participant, (II) a few steps away from the participant, to (III) standing opposite side to the participant. In addition to these three categories, we added another category: (IV) walking toward the participant as the director points at the figure. In our earlier experience, people frequently faced difficulties in following the remote director's direction in such situations. To examine the relationships between the director's position and the effects of remote lag, we classified all 32 questions into four categories so that each category contained 8 questions.

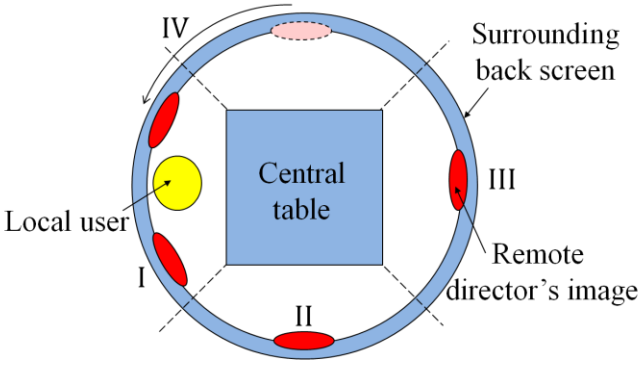

Figure 6: Variation of director's standing positions

Table 1. Overall performance in each condition

	(a) w/o RL	(b) w/ LI (400ms)	(c) w/ LI (400ms) + MF	(d) w/LI (1100ms)	(e) w/LI (1100ms) +MF
Overall Perfor-mance	71.1%	77.0%	80.6%	78.8%	81.0%

RL: remote lag, LI: lagged image, MF: motion flow

Table 2. Overall performance in each condition by standing position

	(a) w/o RL	(b) w/ LI (400ms)	(c) w/ LI (400ms) + MF	(d) w/LI (1100ms)	(e) w/LI (1100ms) +MF
(I) side-by-side	65.0%	76.0%	75.5%	75.0%	78.5%
(II) a few steps away	82.5%	77.5%	85.5%	79.5%	85.0%
(III) opposite side	81.0%	79.0%	80.0%	80.5%	85.5%
(IV) approach	70.0%	75.0%	71.5%	72.5%	79.0%

5. Results

We examined the effects of remote lag on participants' performance (i.e., scores of how accurately they perceived the pointed location from distant directors).

As shown in Table 1, participants' overall performances were highest (i.e., most accurate) in the (e) lagged image (1100 ms) + motion flow condition (M=81%, SD=11.3) and least accurate in the (a) without remote lag condition (M=71%, SD=15.2). A five (trial) by five (condition) repeated measures ANOVA on their scores indicated a significant main effect for trial ($F[4, 100] = 7.21$, $p < .0001$). Contrary to our expectations, however, remote lag did not significantly improve their overall performances.

To see how and when remote lag affected participants' performance, we analyzed the data in further detail. Table 2 shows the average performance of how accurately participants perceived the pointed location from distant directors when directors stood at different standing positions.

Although remote lag was not sufficient to significantly improve participants' overall performances, we found that remote lag did help them to improve their performance when directors stood beside the participants (i.e., (I) side-by-side). A five (trial) by five (condition) repeated measures ANOVA on their scores indicated significant main effects for trial ($F[4, 100] = 4.47$, $p < .01$) and condition ($F[4, 100] = 2.78$, $p < .05$), but no interaction. *Post hoc* comparisons (Tukey's test) of conditions indicated that the scores were significantly higher in condition (e) lagged image (1100 ms) + motion flow than condition (a) without remote lag ($p < .05$).

The following scenes illustrate the capability of remote lag to be helpful when a participant misses the director's pointing action (Fig. 7a) but manages to catch up by following the motion flow and the lagged image (Fig. 7b and 7c). Figure 7b is 900 ms after Figure 7a, and Figure 7c is 100ms after Figure 7b. In Figure 7a, the director is moving next to the participant and pointing at a green circle on the table while saying "this." The participant misunderstood table-side pointing with wall-side by looking at the director's distorted image. Then the participant gives attention to the director along the wall and misses his pointing action projected on the table. The participant soon realizes that the director is pointing at a figure on the table, and shifts his focus toward the table. At this time, the director has already finished pointing at the figures and pauses for a while, and the motion flow is projected on the table (Fig. 7b). The participant predicts that the lagged image will appear through the motion flow. In Figure 7c, according to his prediction, the lagged image of a hand shape appears in the track of motion flow. Finally, the participant manages to answer about the pointed figure correctly. This scene is one of the most typical examples of remote lag overcoming the synergy of several problems that prevent the recognition of a remote gesture as described in Section 2.

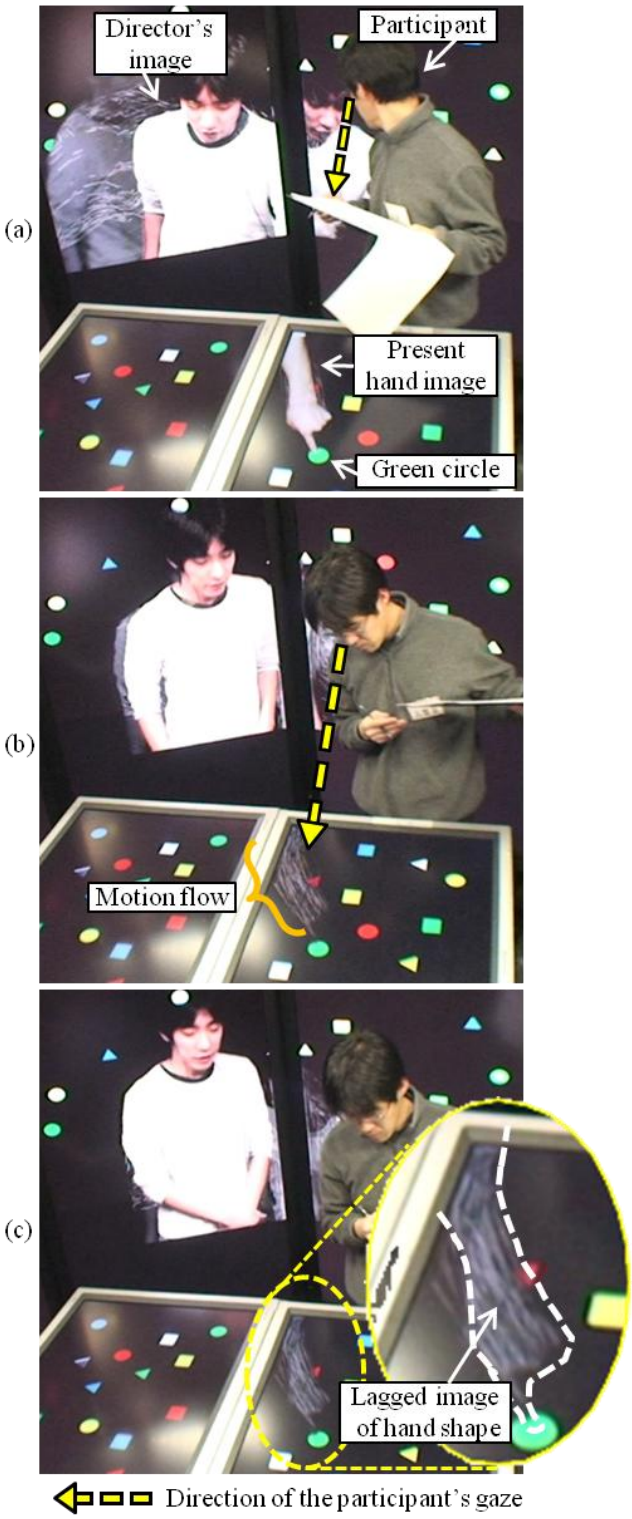

Figure 7: Remote lag contributed to correct understanding of the remote gesture when a participant missed it in the situation where there is a synergetic effect of the invisibilities due to the gap and the fact that remote user Q is behind local user P.

6. Discussion and Conclusion

When a sequence of human behaviors or actions (such as body movements and gestures) is captured by multiple cameras and displayed on multiple 2D screens at a remote site, the 3D structure of those behaviors usually gets distorted. This increases the chances of remote people losing sight of those gestures, possibly making it difficult to understand the gestures.

To alleviate such problems, we introduced the idea of

using remote lag (a technique of showing a lagged image). From our study, we found that remote lag, particularly with a lagged image of 1100 ms behind the current position plus motion flow, helped a participant to perceive the remote gestures more accurately when the director stood right beside the participant.

Although we found that remote lag improved performance when the director stood right beside the participant, it was still insufficient for improving performance when directors walked around the space. One reason may be that directors are typically targeted toward the figure when they walk while pointing at a figure; if a director focused on a particular position while walking toward the figure, the participants might guess approximately where the targeted figure is. Indeed, several participants mentioned in the post-experimental interview that the direction of the director's head helped them to guess approximately where the targeted figure was. When the director stood right beside the participant, maybe there was less time for the participant to make a guess, which led to effective use of remote lag.

Another design implication derived from this study is to show the motion flow between the present and the lagged image. As seen in Tables 1 and 2, the participants' performances were slightly better in those trials with the motion flow than in those without it. In our post-experiment interview, more than half of the participants mentioned that they preferred the trials with the motion flow, since it is easier to follow a lagged image of remote gestures with motion flow.

Our findings from the experiment are not so major, but our approach is the first solution to the gesture invisibility problem in which the recognition degree of remote gestures varies with the physical relationship of users and screens. Our next step is to make use of remote lag in the context of remote collaboration. In this study, we tested the effect of remote lag only in a simple one-way situation, where there is no interaction between remote sites. Moreover, the type of gesture that we tested was only pointing and the targets of pointing were static digital objects. We concentrated exclusively on how accurately participants could perceive remote gestures. However, it is still unclear whether remote lag works well in a more complicated situation such as remote collaborative work that requires various types of gestures.

References

1. Baudisch, P., Tan, D., Collomb, M., et al. Phosphor: Explaining Transitions in the User Interface Using Afterglow Effects. *Proceedings of UIST'06*, pages 169–178, 2006.
2. Gaver, W. The Affordances of Media Spaces for Collaboration. In *Proceedings of CSCW*, pages 17–24, 1992.
3. Gaver, W., Sellen, A., Heath, C., et al. One is not Enough: Multiple Views in a Media Space. *INTERCHI'93*, pages 335–341, 1993.
4. Gutwin, C. Traces: Visualizing the Immediate Past to Support Group Interaction. In *Proceedings of Graphics Interface'02*, pages 43–50, 2002.
5. Gutwin, C. and Penner, R. Improving interpretation of remote gestures with telepointer traces. In *Proceedings of CSCW'02*, pages 49–57, 2002.
6. Hirata, K., Harada, Y., Takada, T., et al. Basic Design of Video Communication System Enabling Users to Move Around in Shared Space, IEICE Transactions, Vol.E92-C, No.11, 2009. (To appear)
7. Kuzuoka, H., Yamashita, J., Yamazaki, K., et al. Agora: A Remote Collaboration System that Enables Mutual Monitoring. *Proceedings of CHI'99*, pages 190–191, 1999.