A Method for Supporting Recognition of Remote Bodily Gestures

Katsuhiko Kaji
NTT Communication Science Laboratories
kaji@cslab.kecl.ntt.co.jp

Naomi Yamashita
NTT Communication Science Laboratories
naomiy@acm.org

Keiji Hirata
NTT Communication Science Laboratories
hirata@brl.ntt.co.jp

ABSTRACT
The 3D structure of body movements and gestures becomes 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 made in a videoconferencing system. Remote lag is a visual effect that overlays a user’s past motion image onto his/her real-time image.

Author Keywords
Remote lag, videoconferencing system, lagged image, optical flow.

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation: User Interfaces: Prototyping.

General Terms
Design

INTRODUCTION
Video conveys a limited amount of information on the three-dimensional structure of remote scenes, and thus limits exploration, inspection, and peripheral awareness [1].

Gesturing in videoconferencing systems (VCSs) is often difficult because various levels of invisibility occur in remote gestures. As for a VCS that the screens are arranged discontinuously, 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 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.

One method for mitigating the invisibility problem is to augment screen images of the gestures, employing users’ past movements. This approach adopts well-established previous work (i.e., telepointer traces [3]) to apply it more generally. Telepointer traces are 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 render motion easier to interpret, and to provide context that helps people understand others’ behavior.

Remote Lag
To acquire a correct interpretation of a series of remote bodily gestures, we propose a visual augmentation technique, remote lag, that applies telepointer traces to bodily gestures such as pointing, moving, face/body orientating, and shape changes. Accordingly, remote lag is the visualization of the user’s past motion overlaid onto his/her real-time image. The several possible representations for telepointer traces proposed by Gutwin [2] 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 [3].

Remote lag consists of two parts: lagged image and motion flow (Fig. 1). The lagged image overlays the video projection while lagging behind the real-time image for a constant time. If user A misses user B’s gesture, the 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. Between the real-time image and lagged image, a set of broken line segments is drawn, and these have fading trails like motion blur. 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 the lagged image and motion flow as an afterimage. Accordingly, a user can correctly recognize the other’s movements, although he/she is viewing them in a constant time behind real-time.

Figure 1: Realistically colored image on the right-hand side is a real-time image of a remote user’s hand. Gray image on the left-hand side is lagged 1100 ms behind real time. Motion flow is displayed between real-time image and lagged image to show the trajectory of the remote object.
The screen images of a conventional VCS are clearly different from a telepointer image in terms of shape complexity and their time-varying nature. Since a telepointer is a cursor, its shape is simple and persistently 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 typically complicated and changing as communication proceeds. Therefore, it is difficult to straightforwardly apply the technique’s motion line and/or motion blur to a image of remote user’s hand and body. We examined several representations of remote lag in practical use of a VCS and, consequently, designed a representation that suited our purpose, with appropriate parameter settings for the time interval of the lagged image, the number of lagged images, the color, and so on.

Multiple images of a pointer can be projected along a trajectory in telepointer traces because the pointer image is small. On the other hand, the sizes of the images in VCS are typically larger than a pointer image. Therefore, to prevent a workspace from becoming cluttered with large lagged images, a single lagged image is overlaid. The lagged image is overlapped under the real-time image to hide the lagged image of motionless objects. Additionally, the lagged image is displayed in gray-scale to easily judge whether the image is the real-time one or the lagged one.

**Applying Remote Lag**

We applied these visual effects to our t-Room [4]. In the t-Room, the space shared by remote and local users is created by 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. Same t-Rooms are at the remote site, and the local images are projected at the correspondent remote screens. Local and remote users can freely move around the surrounding back screens. In addition, a central table is introduced to provide a shared work space. In the structure, recognition of remote gestures becomes difficult by a synergetic effect of the following two problems. First, the remote user’s image is projected behind the local user; second, invisibilities are caused by the gap between the surrounding back screen and the central table.

The following scenes illustrate the assistive capabilities of remote lag when local user P misses the pointing action performed by remote user Q (Fig. 2a) but manages to catch up by following the motion flow and the lagged image (Fig. 2b and 2c). In Figure 2a, Q is moving next to P and pointing at a green circle on the table while saying “this.” P mistakes Q’s table-side pointing as wall-side indication by looking at Q’s distorted image. Consequently, P pays attention to Q along the wall and thus misses Q’s pointing action projected on the table. P soon realizes that Q is pointing at a figure on the table, and thus P shifts his focus toward the table. At this time, Q has already finished pointing at the figures and pauses for a while, and at this point the motion flow is projected on the table (Fig. 2b). P predicts that the lagged image will appear through the motion flow. In Figure 2c, according to P’s prediction, the lagged image of a hand shape appears in the track of motion flow. Finally, P manages to answer about the pointed figure correctly. This scene is one of the most typical examples of how remote lag helps the recognition of a remote gesture.

**Figure 2:** Remote lag helped to ensure correct understanding of a remote gesture when local user missed it due to synergetic effect of two invisibilities: 1) gap between wall and table and 2) remote user being projected behind local user.

**REFERENCES**