Can Eye Help You?: Effects of Visualizing Eye Fixations on Remote Collaboration Scenarios for Physical Tasks

Keita Higuch[1], **Ryo Yonetani**[1], and **Yoichi Sato**[1]

[1]The University of Tokyo

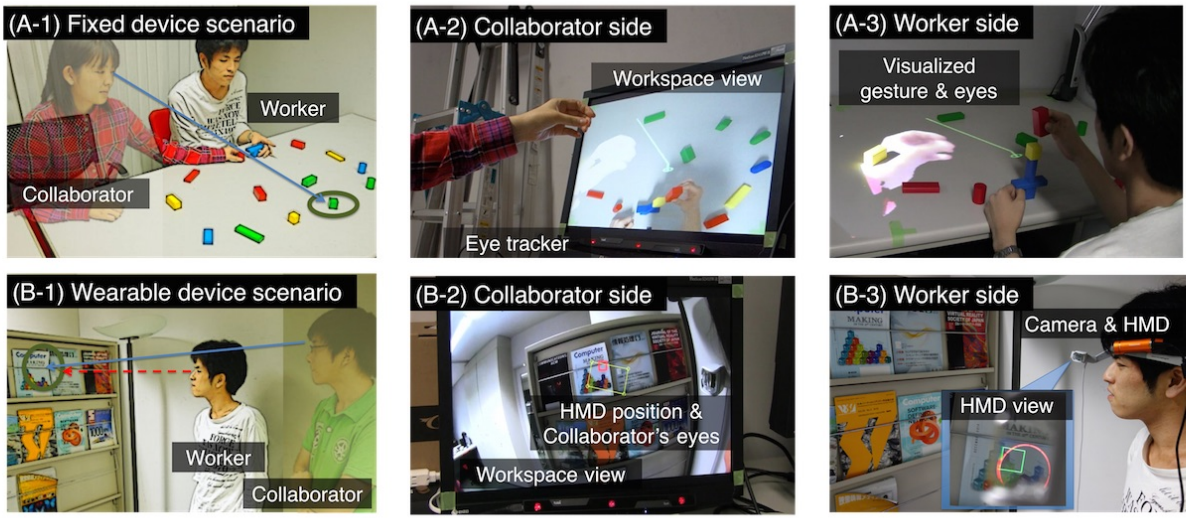

Figure 1: Conceptual figures and implementations of our remote collaboration framework under two different scenarios: (A-1) **fixed device** scenario offers remote collaboration experience for physical tasks on a small workspace. (A-2) a remote collaborator monitors a workspace through a fixed point-of-view camera and provides instructions to a worker using his/her eye fixations and hand gestures. (A-3) the system then visualizes a history of collaborator's point of gaze (green line) as well as hands directly in front of the worker. (B-1) **wearable device** scenario addresses remote collaboration in larger workspaces. (B-2) the system provides the view of a wearable camera worn by the worker. (B-3) the worker was allowed to see the relative direction and position of collaborator's eye fixations through an optical see-through head-mounted display.

ABSTRACT

In this work, we investigate how remote collaboration between a local worker and a remote collaborator will change if eye fixations of the collaborator are presented to the worker. We track the collaborator's points of gaze on a monitor screen displaying a physical workspace and visualize them onto the space by a projector or through an optical see-through head-mounted display. Through a series of user studies, we have found the followings: 1) Eye fixations can serve as a fast and precise pointer to objects of the collaborator's interest. 2) Eyes and other modalities, such as hand gestures and speech, are used differently for object identification and manipulation. 3) Eyes are used for explicit instructions only when they are combined with speech. 4) The worker can predict some intentions of the collaborator such as his/her current interest and next instruction.

Author Keywords

Remote Collaboration; Gaze

ACM Classification Keywords

H.5.3. Group and Organization Interfaces: Computer-supported cooperative work

INTRODUCTION

Remote collaboration is a form of interaction among two or more people at remote locations, who interact with each other to achieve a goal. Supporting remote collaboration by systems (*i.e.*, computer-supported cooperative work; CSCW) is a long-standing topic in HCI and is beneficial in various fields. For example, novice workers at a factory production line can gain the abilities of experienced people at another factory, making craftsmanship potentially ubiquitous. Remote collaboration at home enables people to take online classes while sharing their work in progress with a remote teacher. People may even casually consult on remote partners in their daily life as elders ask their children about the use of smartphones.

In particular, we are interested in a specific scenario where a worker involved in a physical task is supported by another remote collaborator who sees the worker's work in progress via video. One important technique in this scenario is the transmission of the collaborator's intentions and instructions to the worker such as what objects the collaborator is currently focusing on and how he/she wants the worker to manipulate these objects. Prior work does this by using different modalities including speech, annotations [10, 12, 17, 20, 36], and hand gestures [16, 30, 34] displayed to a physical work space to show the explicit instructions given by the collaborator.

In this study, we investigate how remote collaboration by a worker and a remote collaborator will change if the collaborator's eye fixations on a physical workspace are presented to the worker. We track the collaborator's points of gaze on a screen monitoring a workspace (Figure 1 (A-2), (B-2)) and project them on the space by a projector (A-3) or through an optical see-through head-mounted display (HMD) (B-3). Prior work has revealed the effectiveness of using eyes in both physical tasks [24, 32] and virtual tasks [41, 49]. We expect that people will also effectively use their eyes to accomplish a task in the context of remote collaboration. We will also see how the role of existing modalities, such as speech and hand gestures, will change when they are used in combination with eye fixations.

The proposed systems facilitate remote collaboration as follows: A physical workspace is first captured by a camera and transferred to a display screen monitored by a remote collaborator. The collaborator's points of gaze on the screen are then captured by an eye tracker and visualized directly onto the workspace. We design several ways of visualization to address two major scenarios of remote collaboration, which introduce 1) a pair of fixed point-of-view (POV) camera and a projector for monitoring relatively small workspaces, and 2) a wearable head-mounted camera and an optical see-through HMD to cope with larger workspaces.

The contribution of this work can be summarized by the following four main findings obtained from three user studies:

- Eye fixations visualized on a physical workspace serve as a fast and precise pointer, allowing collaborators to reduce the number of incorrect directions or to decrease the task completion time. This feature further leads to joint pointing to multiple objects in a static scene captured by a fixed POV camera or stable pointing in a dynamic scene with large motion caused by a moving wearable camera.
- We observe the different roles of eyes and other modalities. Instructions given by eye fixations are mainly used for identifying objects of the collaborator's interest. On the other hand, the collaborator uses hand gestures for describing object manipulation such as rotation and attachment.
- Eye fixations are used for providing explicit instructions only when these fixations are combined with speech. The worker can therefore distinguish such instructions easily even if the eye fixations are visualized throughout a task in a non-salient manner.
- Eye fixations indicate the collaborator's implicit intentions. In particular, the worker can predict what the collaborator is interested in and what his/her next instruction will be.

RELATED WORK

Remote Collaboration

Prior work on remote collaboration, as typified by [34], has sought a method to enhance the abilities of the worker with respect to physical tasks, *e.g.*, construction, navigation, and decision making, with support from the collaborator. Systems to support the remote collaboration are characterized by a camera device used for monitoring a physical workspace such as fixed POV camera systems [1, 2, 20, 30, 43, 50] and mobile camera systems [10, 12, 17, 26, 33, 36, 47]. In particular, Fussell *et al.* studied the effect of camera settings on a remote collaboration scenario [14, 15]; they observed that fixed POV cameras providing a static view of workspaces offered valuable visual information to a collaborator. Fussell also suggested that wearable cameras could be used to capture wider workspaces than the fixed cameras. These findings have inspired our use of the collaborator's eye fixations for providing instructions as eyes can be used for quickly scanning the workspaces.

Gaze-based Interfaces

After the pioneering work by Jacob on eye movement-based interaction techniques [22], eye movements have been adopted in a wide range of applications such as GUI manipulation [38, 25], daily life [8, 21], emergency guidance [40], and robot operation [35], as high-accuracy eye tracking has become available at a low cost. A considerable amount of recent work has focused on the enhancement of touch/gesture interaction with gaze information [42, 48, 51], demonstrating that users can use their eyes to quickly point at target objects and enhance the precision and speed of touch/gesture interactions. We expect that this ability of the eyes can also effectively work in object identification tasks in the context of remote collaboration.

Eye-tracking for collaborative tasks

Eye-tracking technologies have already been used in the field of CSCW, but in a different configuration. Dual eye-tracking (DUET) studies use an eye tracker for two workers to visualize their points of gaze on a shared screen [9, 11, 23, 39, 41, 44, 49]. Another method of using eye tracking is to present the eye movements of a worker involved in a physical task to a collaborator [31]; this has revealed a close relationship between the worker's gaze and the collaborator's verbal instructions. We will use the eye fixations of a remote collaborator for remote collaboration on physical tasks, standing for a complementary work to [31]. Our work also attempts to combine eye fixations and hand gestures used along with speech [34, 10, 16, 30].

PROPOSED REMOTE COLLABORATION FRAMEWORK

Assume that two persons at remote locations are trying to collaborate with each other to perform a physical task, including fine operations on physical objects such as cooperatively assembling implements at a factory production line or building a showcase of items at a store, *etc.*In this setting, we consider the following two roles to establish remote collaboration: 1) a *worker* who faces a workspace in physical spaces (*e.g.*, a workbench or an entire factory) to perform a task, and 2) a *collaborator* who monitors the workspace through a camera and instructs the worker.

The key question of this work is how the interactions between a collaborator and a worker working together will change if the collaborator's eye fixations on a monitor screen are physically presented onto the workspace. As the collaborator and the worker are physically separated, it is important through interaction to share objects of interest to manipulate them toward the goal of the task. Therefore, many user studies with a remote-collaboration setting have consisted of object identification (*e.g.*, [26, 31]) and manipulation (*e.g.*, [14, 34]).

In particular, we are interested in how systems can facilitate remote collaboration when remote collaborators wish to manifest their intentions or instructions, such as what objects they are currently focusing on and how they want a worker to manipulate these objects. In previous studies, remote collaborators did this explicitly using hand gestures [10, 26, 30, 34], annotations [12, 17, 20, 36], and so on. While these modalities are capable of describing detailed instructions, such as object manipulation or the overall context of a task, they have also been used only for identifying the objects of focus.

Our proposal to this end is to allow collaborators to use their eyes together with their hands and speech for describing their intentions and instructions. We hypothesize that if the eye fixations and the hand gestures of the collaborators are visualized in front of a worker, the collaborators will mainly use their eyes to identify the objects of interest, and use their hands and speech for describing other detailed instructions such as object manipulation. In fact, we see objects implicitly before manipulating them by hand [24], and such implicit gaze behavior alone is often critical for identifying which objects should be of the collaborator's interest [3, 7, 19]. To help workers infer the collaborator's intentions from such implicit behavior, we visualize eye fixations on the workspaces throughout the tasks, while keeping them less salient so as to not distract the workers.

Another advantage of using eye fixations is to enable fast pointing to objects, as we have already reviewed in the related work section. Even if the eye fixations are visualized in a less-salient way, they can be used for supporting the other modalities (*i.e.*, speech and hands) in the form of specifying the objects being referred to.

In the remainder, we will further discuss how eye fixations potentially work on more specific scenarios such as when using fixed POV cameras and wide field-of-view projectors, or when using wearable head-mounted cameras and relatively narrow field-of-view optical see-through displays. We will also present how to visualize eye fixations tailored to each of the scenarios.

Visualizing Eye Fixations with Fixed POV Devices

We first consider a specific scenario where a fixed POV camera and a projector are installed to deal with relatively small workspaces. In this scenario, a sequence of the remote collaborator's eye fixations can be visualized on the workspaces, enabling the collaborator to instruct a worker on more than a single object of current interest.

Some psychological studies have revealed that eye movements were indicative of their background factors such as critical points in the manipulation tasks [24], task-relevant locations in a scene [18, 52], next set of manipulations [6], types of tasks [5, 45, 52], and conversational cues [4, 28]. We can therefore expect that workers will be able to infer some of the intentions from the collaborator's eye movements. For example, eye fixations that traverse an entire workspace region implicitly indicate that the collaborator is not sure of which objects to process at that moment. In contrast, sequential fixations on several objects could allow workers to predict the next set of objects to focus on.

Another possible behavior enabled by a sequence of eye fixations is a joint reference of two or more objects. This behavior can help collaborators to describe the relationships between objects (*e.g.*, screws to attach and boards to be attached).

Visualizing Eye Fixations with Wearable Devices

When workspaces are larger than what a worker can see at once, we choose to use a pair of wearable head-mounted camera and an optical see-through HMD for the worker to interact with his/her collaborator. Unlike other camera devices, such as shoulder-mounted cameras [33] and portable computers [17], the head-mounted cameras allow the worker to use his/her hands for physical operations. The optical see-through HMD also has an advantage in physical tasks over display tools, such as immersive video see-through displays, particularly when the worker is involved in fine operations on physical objects.

As the field-of-view of HMDs is often considerably narrower than that of wearable head-mounted cameras, a collaborator's instructions on the object outside the display view cannot be directly transmitted to a worker. This constraint makes it difficult or the collaborator to use his/her hands for manifesting the instructions; when pointing out the objects of interest, the collaborator must change his/her hand poses frequently to describe the relative direction of the objects as the head-mounted cameras move according to head motions of the worker.

Therefore, we design an indicator for the direction and distance of an eye fixation from the current view of the worker as if the worker and the collaborator were focusing on the same objects, as depicted in Figure 1 (B-1). The collaborator can then specify the objects of interest just by seeing them, making it easier to give instructions by other modalities such as speech and hand gestures.

From the worker's perspective, the eye fixations of the collaborator are not always salient as the worker sometimes averts his/her eyes from the HMD. Fixations will explicitly work only when they are used for instructions in combination with the other modalities.

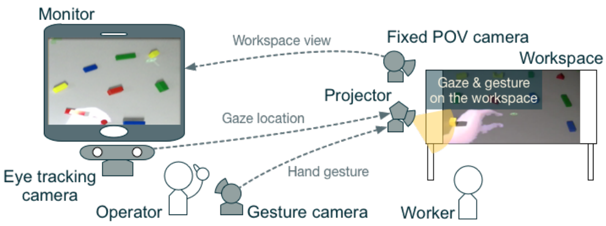

Figure 2: Overview of the system with fixed POV devices

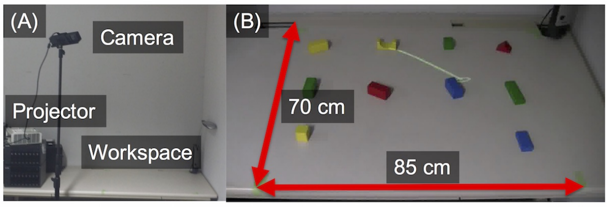

Figure 3: (A) Workspace captured by a fixed POV camera (B) Dimensions of the fixed device setting

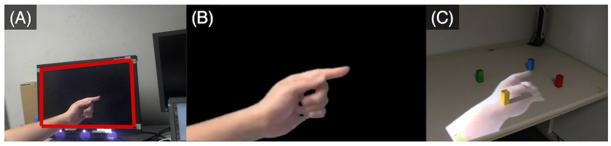

Figure 4: Visualizing Hand Gestures: (A) Video from web camera with a polarized filter (B) Cropped hand (C) Projected hand on the workspace

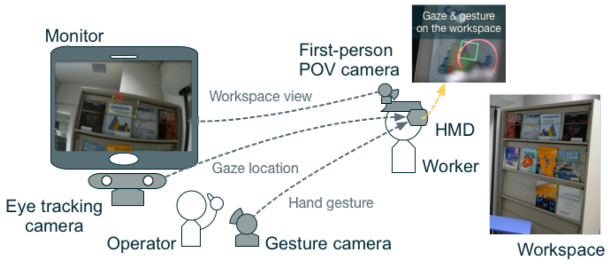

Figure 5: Overview of the system with wearable devices

IMPLEMENTATIONS

Implementation with Fixed POV Devices

We first introduce the implementation of our framework aimed for relatively small workspaces with a pair of a fixed POV camera and a projector (see Figure 2 for its overview).

Figure 3 shows the physical setup of the worker side; here, the system is based on a homographic projection mapping method. The dimensions of the setup are 85 (width) × 70 (depth) centimeter, but these dimensions can be extended by changing the projector and camera positions.

Projector-camera calibration

To map the collaborator's eye fixations on a display monitor onto the physical workspace of a worker, we computed the perspective projection between a projector and a camera by manually pointing where markers placed on the workspace appear in the view of the camera and the projector.

Visualizing eye fixations

The collaborator's eye and hand movements are captured using an off-the-shelf eye tracker (Tobii EyeX) and a webcam (Microsoft LifeCam), respectively. The eye tracker measures the collaborator's 2D eye position on the display at 30 fps. The current point of gaze is displayed as a green circle together with five preceding points of gaze shown with a polyline, which corresponds to approximately a 0.2-s history of eye movements (Figure 3 (B)).

Visualizing hand gestures

To visualize hand-based instructions, such as hand gestures, we referred to previous researches [13, 37]. Polarized light from an LCD monitor is blocked by a polarized filter attached to the camera, while an unpolarized light reflection from a hand is still visible. This makes it easy to find the hand regions irrespective of what is being displayed on the monitor (Figure 4). Then, the system visualizes the hand gestures onto the workspace.

Implementation with Wearable Devices

As shown in Figure 5, we adopt the same implementation for the collaborator's side for wearable-device scenarios tailored to larger workspaces. The main difference is that the worker wears a head-mounted camera (Panasonic A1H) and a light-weighted optical see-through HMD (Brother AirScouter WD-200S). The HMD is placed in front of the wearer's line-of-sight to achieve better task performance [53].

HMD-Camera Calibration

The system uses optical see-through HMD-based mobile augmented reality (AR) technologies that require a calibration process to map collaborator's eye fixations onto the HMD view from a worker. To do this, we followed [27] and asked an HMD wearer (worker) to align an AR marker at several locations specified in the HMD view by moving his/her head. We could then achieve the HMD-camera calibration by seeing the marker through the head-mounted camera.

Visualization in optical see-through displays

As the HMD has a considerably narrower field of view than the camera, the collaborator's eye position on the display showing the view from the wearable camera may be outside of the HMD's view. In such cases, the direction of and the distance to the collaborator's eye position is visualized on the HMD as shown in Figures 6 (A) and (B). The direction of eye positions from the center of the HMD's view is indicated by a circular edge, where the number of red edges increases as the eye positions get closer. When the collaborator sees inside the HMD's view, the eye positions are indicated by a green square (Figure 6 (C)). The worker is expected to see whether he/she and the collaborator can share the view or not.

The collaborator's hands are also shown on the HMD. The system crops hand images by using the same technology as the fixed device system (see Visualizing hand gestures). The system also directly visualizes the cropped hand images on the HMD (Figure 6 (D)).

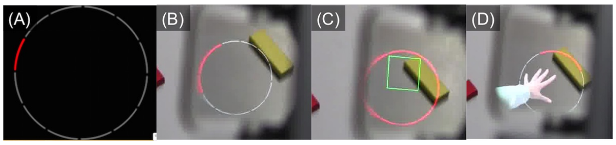

Figure 6: Visualizations through an optical see-through HMD: (A) and (B) circular-edge indicators to show the direction of the collaborator's points of gaze, (C) point of gaze inside the HMD view indicated by a green square, and (D) hand gestures

USER STUDY 1: ASSEMBLING TARGET OBJECTS UNDER FIXED DEVICE SCENARIO

In this section, we first report how eye fixations work on a fixed POV device scenario. Similar to [43, 46], participants were asked to perform remote collaboration to assemble a certain object (*e.g.*, a building) with the blocks scattered on a desk. We found the statistical evidence on fast pointing by eye fixations. We also obtained some feedback that implied the capability of eye fixations, enabling multiple pointing and conveying the collaborator's intentions.

Task and Procedure

In this user study, we compared the following two conditions to see how eye fixations were used for collaboration and how the use of the other modalities changed: a **Gesture** condition where a collaborator was allowed to use hand gestures and speech and a **Gesture + Eye** condition where eye fixations are also available for the collaborator. We recruited eight participants who were students or postdoctoral researchers at a graduate school of computer science; they belonged to the age group of 20 to 30 years and had limited experience on computer-supported remote collaboration.

Four pairs of participants were asked to complete an assembly task by using the following steps. First, one of the participants in each pair assigned the role of a remote collaborator saw the as-built drawings of a target object. Then, the collaborator instructed the other participant (*i.e.*, a worker) on how to assemble the target object by using blocks. Each participant took turns playing the collaborator and the worker for each condition, and the order of conditions was randomized to maintain the counterbalance (*i.e.*, each pair conducted four sessions in total). For each session, the participants were given an additional 5 min in advance for practice with an instruction to use our remote collaboration system by an experimenter. In the case of the **Gesture + Eye** condition, we calibrated the eye tracker before each session.

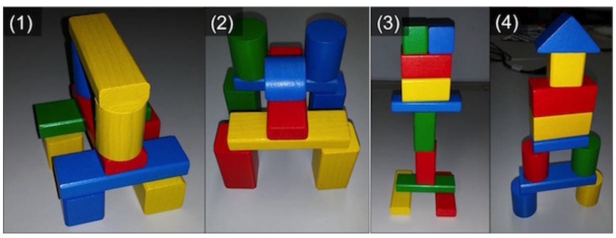

Figure 7: Four target objects used in User Study 1

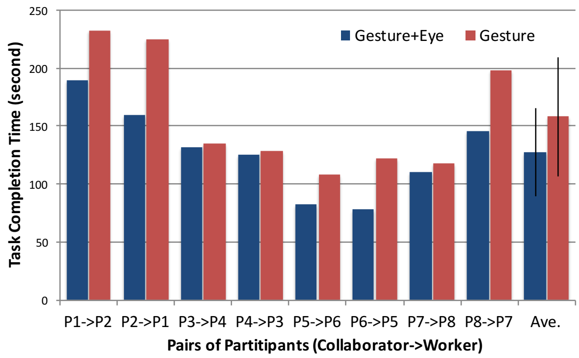

Figure 8: Task completion times in User Study 1. Bars show standard deviation.

The target objects were randomly chosen from the four depicted in Figure 7. They all consisted of 10 blocks from 25 candidate blocks with a variety of shapes and colors. The as-built drawings of the target objects were the photographs of these same objects taken from three different points of view so that the collaborators could easily figure out the component blocks.

Evaluations

The time taken to complete a task was measured for each session to observe the statistical difference between the conditions. We hypothesized that the task completion time would decrease in the **Gesture + Eye** condition than in the **Gesture** condition owing to the fast pointing by eye fixations. We validated this hypothesis by using a paired t-test as standardized test statistic.

After each session, the participants answered the list of questions presented in Figure 9 with a seven-point scale (disagree = 1, agree = 7). We also investigated using the Wilcoxon matched-pair signed-rank test as a non-parametric statistical hypothesis test whether there were significant differences in the participants' experience. Finally, we interviewed the participants regarding their experience with our remote collaboration system at the end of the experiment.

Results

Figure 8 shows the task completion time for each session. We confirmed the statistical significance in the task completion time ($p = 0.01$), indicating that the eye fixations were certainly used as a fast pointer. This finding was also supported by the statistical significance obtained in the case of the questionnaire shown in Figure 9 ($p \leq 0.05$). We also confirmed that the participants successfully shared the instructions and intentions, on the basis of the responses to questions.

Observation and Feedback

Under the **Gesture + Eye** condition, most participants serving as a collaborator used eye fixations together with speech to instruct a worker regarding a block position. They quickly

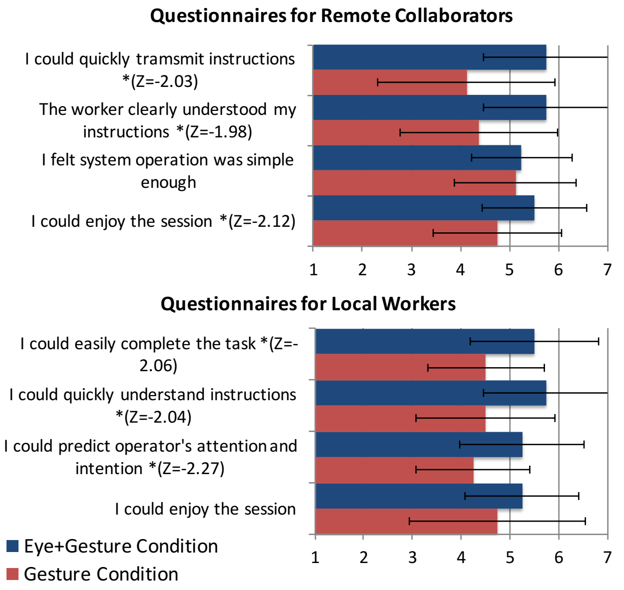

Figure 9: Questionnaires in User Study 1: Pairwise Wilcoxon signed-rank tests. "*" indicates the significance. Bars show standard deviation.

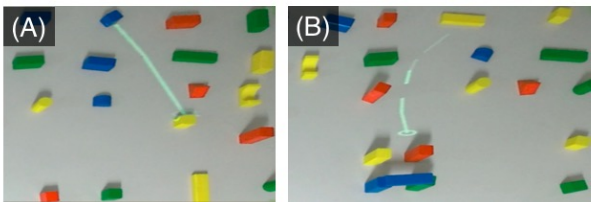

Figure 10: Multiple pointing in User Study 1 (A) Two blocks (B) Block and its destination

looked at a target block and gave simple instructions: *"I felt the worker knew my gaze position, so I just said* ***take this*** *to identify an object* [1]*."*

We also observed a particular role of hand gestures. The collaborators mainly used hand gestures to provide detailed instructions of a manipulation of blocks such as rotations and attachments: *"pointing by eye fixation is very easy and fast, but it can only describe locational information. So, I used gestures for describing the posture of blocks."*

Further, we sometimes observed multiple pointing by using eye fixations; the collaborators looked at several positions at a time. *"I could easily point out two blocks with my eyes"* and *"I pointed the block and destination positions together."* Multiple pointing is a characteristic capability of visualizing eye fixations in a fixed device setting that might support fast instructions for shorter task completion times (Figure 10).

The workers almost correctly understood the collaborator's instructions in the **Gesture + Eye** condition: *"I understood the blocks looked at by the collaborator. I just took these blocks in time with the collaborator's speech."*

The workers could also distinguish the collaborator's instructions from other eye movements: *"I followed the collaborator when he/she spoke some instructions,"* and *"I was able to predict the next instruction from a history of collaborator's eye positions."*

We received negative feedback mostly on projection problems because of the incorrect projection size and the 2D projection: *"The size of the collaborator's hands was unnatural to me; it was too large."* In addition, blocks rarely occluded instructions provided by visualized eyes from the workers: *"Sometimes, I lost the collaborator's eyes by occlusion"* and *"When the worker lost my eyes, I needed to keep looking at the target block to describe its position."*

USER STUDY 2: OBJECT IDENTIFICATION UNDER WEARABLE DEVICE SCENARIO

Next, we evaluated our remote collaboration system in a wearable device setting. We aimed to see how fast and precisely the collaborators could use gaze to specify the target objects captured with a head-mounted camera of a worker. We designed a simple task since it was unclear whether people could point objects reliably when the view of a camera changed dynamically with a worker's movement.

Task and Procedure

Two conditions were compared in this experiment, where a collaborator could use eye fixations (**Eye** condition) or hand gestures (**Gesture** condition). The eight participants had the same background as those in User Study 1. This time, the participants all played the role of the collaborator, while one experimenter served as a worker.

We performed the following simplified task where only a collaborator pointed to objects. As shown in Figure 11 (A), the worker was asked to stand in front of the 18 blocks placed on a desk. When a session began, the image shown in Figure 11 (B), which indicated one of the 18 blocks, was first presented to a collaborator. The collaborator then navigated the worker to the indicated block, by seeing it or describing its location by hand. The worker moved his/her head to place an object inside the view of an HMD and confirmed whether it was the correct target. This pair of navigation and confirmation was repeated until the collaborator reached the correct target. The collaborators were allowed to say "yes" or "no" to tell whether the worker correctly found the target. The number of mistakes were measured by counting the number of "no" responses in nine sessions. Please note that multiple mistakes could occur in a single session.

For each participant, we conducted the aforementioned procedure nine times for each condition preceded by three additional trials for practice. We calibrated an eye tracker once before the **Eye** condition. The order of conditions was randomized to maintain the counterbalance. We also performed experimental sessions to use wearable devices for the experimenter in advance in order to reduce the learning effects.

[1]In this paper, italic fonts in double quotations denote translated speech from Japanese.

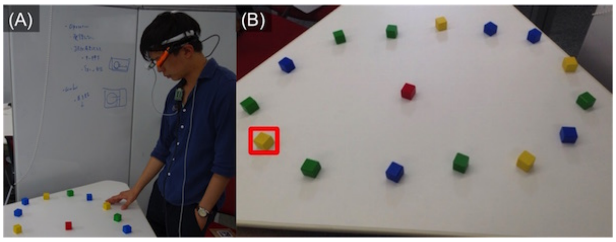

Figure 11: (A) Experimental setting for User Study 2 (B) example of a target block picture presented to participants

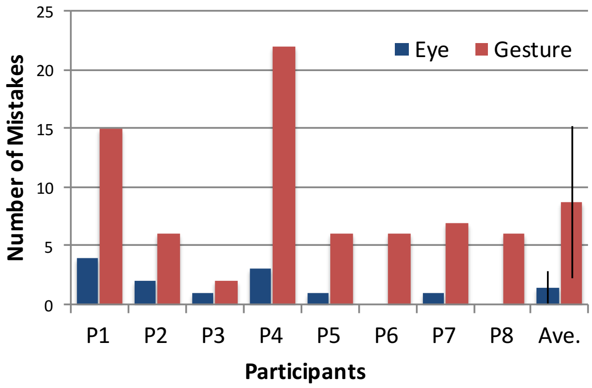

Figure 12: The number of mistakes in User Study 2. Bars show standard deviation.

Evaluations and Results

As an evaluation measure, we counted the number of times where the worker failed to specify the object navigated by a collaborator, accumulated over nine trials. We also measured the total task completion time in the nine sessions. We expected the number of mistakes and the total time to decrease in the **Eye** condition and tested them by the paired t-test.

Figure 12 shows the total number of mistakes for each participant. Overall, all the participants precisely specified the target objects with eye fixations, which was supported by the paired t-test at $p = 0.01$.

Figure 13 also shows the total task completion times for each participant. Some participants quickly specified the target objects with eye fixations, which was supported by the paired t-test at $p = 0.05$.

USER STUDY 3: ARRANGING SHOWCASES OF BROCHURES UNDER WEARABLE DEVICE SCENARIO

We further evaluated our system with wearable devices qualitatively to see how eye fixations affected the remote collaboration experience in a practical task. We again recruited eight participants who had the same background as those in User Studies 1 and 2. Four pairs of participants were asked to arrange a showcase of brochures for travel fairs, as shown in Figure 15, and to answer the interview questions on their experience of the task.

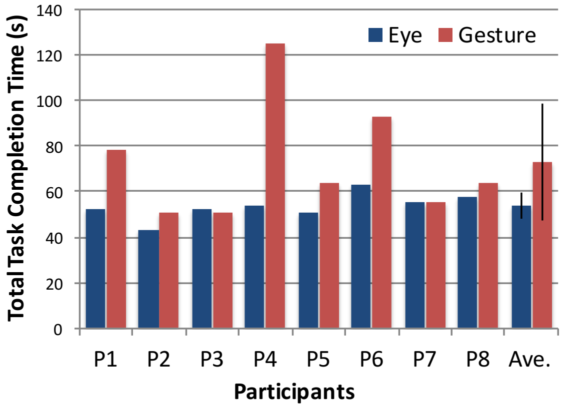

Figure 13: The task completion times in User Study 2. Bars show standard deviation.

Task and Procedure

This experiment was conducted with a larger workspace than that in the previous two user studies. As illustrated in Figure 14, the workspace consisted of a working desk, a shelf, and a whiteboard that could not all be captured at once by a wearable camera without moving the worker's head. A set of brochures was placed on the working desk. Various decorations and tools were stored on the shelf. The worker was able to attach them on the whiteboard to arrange a showcase.

The task proceeded as follows: First, we explained the use of our system and the goal of the task to a pair of participants. They then decided their role (the worker or the collaborator). We performed HMD-camera calibration for the worker and eye tracker calibration for the collaborator. During a session, the participants first checked all the travel brochures and discussed the concept of the showcase (*e.g.*, trip to Europe). Once the concept was fixed, they started to decorate the whiteboard by using brochures and tools on the showcase space. The task continued until participants were satisfied.

Interviews

At the end of the task, we interviewed the participants for 10 min. Our interviews were semi-structured to focus on the experience as a worker and a collaborator. First, we asked to the collaborator, "*Did you use your eyes for providing instructions, and if so, when?*" Next, we asked to the worker, "*Did you understand the collaborator's eye positions?*"and "*How did you find the visualization on the HMD?*" We then reviewed their showcase together and asked the participants to comment on their experience freely.

Observation and Feedback

All the pairs of participants were able to successfully realize their concept of travel fairs and satisfactorily arrange their showcases, as shown in Figure 15. The task completion times were between 22 min and 51 min. They enjoyed the remote collaboration experience: "*I really enjoyed this collaboration. I was able to realize our concept with my partner*" and "*We were satisfied with this showcase.*"

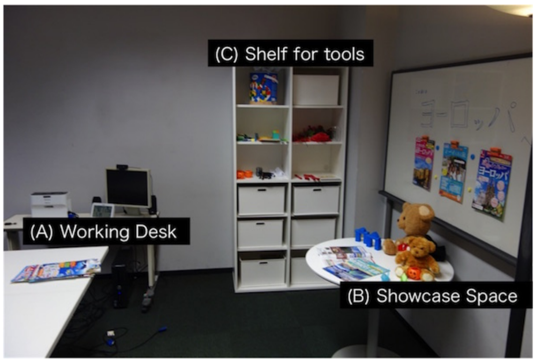

Figure 14: Experimental setting of User Study 3

Figure 15: Results of User Study 3: each picture shows a result of the showcase arranged by each pair.

Figure 16: Pointing instructions with eye fixations in User Study 3: (A) Object (B) Location. Blue circles show collaborator's eye positions.

The collaborator used eye fixations for pointing objects and some locations in the workspaces to facilitate joint attention with the worker (Figure 16). The worker could specify the objects mentioned by the collaborator before manipulating the objects: "*I felt the instructions provided by my eyes simple*" and "*I could look at the objects looked at by the collaborator*." In all the pairs, instructions with eye fixations were provided frequently when a worker moved around and the workspace visible to a collaborator completely changed. The collaborators were able to track the target objects/locations by their eyes under dynamic scenes.

Similar to the fixed device scenario evaluated in User Study 1, we found that the workers were able to distinguish explicit instructions from eye fixations continuously presented through an optical see-through HMD as such instructions were always accompanied with speech: "*I checked the HMD to receive instructions when my partner used speech*."

The visualization of eye fixations sometimes indicated the collaborator's interests: "*Sometimes, I saw the HMD to check the collaborator's interests*" and "*I could figure out where my partner was interested in.*"

Hand gestures were used for describing various object manipulations such as attaching brochures on a whiteboard, moving a toy, and cutting paper decorations. The workers felt that the hand gestures were more noticeable than eye fixations: "*I like seeing hand gestures. They are more attractive*" and "*I felt the eye movements non-salient compared to gestures*."

The participants who played the role of workers mostly liked our system setup and visualization: *"This setup is simple. I didn't feel a fatigue in this experiment."* We also received negative feedback from a few participants about the HMD-camera calibration process. The process needed trial and error for precise calibration that took up to 10 min: *"The calibration process was difficult for me; I wish it were easier."* On the other hand, the other participants were able to perform this calibration process quickly.

DISCUSSION

In this section, we discuss how remote collaborator's eye fixations visualized on a workspace can work for remote collaboration. Overall, we confirmed the effectiveness of visualizing eye fixations both in the user studies with fixed POV and wearable devices. In particular, we found the following features of the fixations in remote collaboration.

Finding 1: fast and precise pointing by fixations

Eye fixations visualized by both a projector (User Study 1) and a HMD (User Studies 2 and 3) show a fast and precise pointing capability over hand gestures. This feature of fast pointing further enables joint pointing to multiple objects at a time when the history of fixations is visualized in the fixed device scenario. Precise pointing effectively works in the wearable device scenario where workspaces monitored through a wearable camera often change dynamically.

Finding 2: different roles in fixations and hand gestures

Collaborators used their eyes and hands in different ways to instruct workers. Eye fixations were used mainly for identifying the objects of interest, while hand gestures were used for providing instructions regarding the manipulation of the identified objects. In User Studies 1 and 3, remote collaborators described the object locations by gaze, often followed by their hand gestures, to describe how they wanted the worker to manipulate the object of focus.

Finding 3: explicit instructions enabled by the combination of fixations and speeches

In User Studies 1 and 3, eye fixations were used for providing explicit instructions (*e.g.*, object identification) only when they were combined with speech. This feature allowed workers to easily distinguish instructions from the other unmeaningful eye movements even if the points of gaze were visualized throughout a task, in a non-salient manner. This feature was related to the finding in [31] that the collaborator's verbal instructions affected the worker's attention.

Finding 4: indicating collaborator's implicit intentions

Continuous visualization of the collaborator's points of gaze was also indicative of the implicit intentions of the collaborator such as his/her interest and future instructions. Indeed, some participants in User Study 1 reported that the history of points of gaze could be used for predicting the current state of the collaborator such as comparing multiple objects in the workspace and trying to provide the next instruction. In addition, the worker was able to see whether his/her interest aligned with that of a collaborator in User Study 3.

LIMITATION

In User Study 1, we observed the problems of projection mapping related to the size of the projected hands and the occlusions by blocks, because the fixed device system used a 2D projection mapping method (*i.e.*, homographic projection mapping). We believe that 3D projection mapping technologies will be able to address these problems.

The result of User Study 2 revealed fast and precise pointing of visualizing eye fixations in a wearable device setting. However, User Study 2 did not clarify the effectiveness of the collaborator's eye fixations from a worker's perspective, because the participants played only the role of the collaborator. Even so, the participants of User Study 3 used eye fixations for object and position identification in a practical task. We believe that this observation can support Finding 1. Additional studies will help to generalize the effectiveness of the proposed framework under different conditions (e.g., different person and target sizes).

CONCLUSIONS AND FUTURE WORK

In this study, we investigated how remote collaboration between a local worker and a remote collaborator changed if the eye fixations of the collaborator were presented to the worker. Through a series of user studies, we found some typical ways of using eyes for remote collaboration and how they effectively worked to accomplish physical tasks. Our findings can be enabled just by adding an off-the-shelf eye tracker on the collaborator's side. It covers not only the remote collaboration in small workspaces enabled by a pair of a fixed POV camera and a projector but also that for larger workspaces that require wearable devices.

Extending the visualization of eye fixations to the collaborator's side like [31] leads to a novel dual eye-tracking study. One promising direction is to install a pair of wearable head-mounted eye tracker and an HMD on both the people in the remote collaboration for two individual physical tasks. While this setting has the potential of enhancing their non-verbal interaction, it also raises a new question: How will people distribute their attention to their own workspace and that of their partner's visualized through the HMD.

Another direction is to apply our framework to considerably larger workspaces, such as an entire building or a city, where more than two workers will move around to perform a physical task. This application involves several important topics in CSCW, such as efficient monitoring and provision of instructions to multiple workers, long-term assistance with remote collaboration, and heterogeneous monitoring of workspaces by using smartphones [17, 29] as well as fixed POV and wearable cameras.

ACKNOWLEDGMENTS

This research was supported by CREST, JST. We thank Kei Nitta and Kana Misawa for their support.

REFERENCES

1. Matt Adcock, Stuart Anderson, and Bruce Thomas. RemoteFusion: Real Time Depth Camera Fusion for Remote Collaboration on Physical Tasks *(VRCAI '13)*.
2. Matt Adcock, Dulitha Ranatunga, Ross Smith, and Bruce H. Thomas. Object-based Touch Manipulation for Remote Guidance of Physical Tasks *(SUI '14)*.
3. Antti Ajanki, DavidR. Hardoon, Samuel Kaski, Kai Puolamki, and John Shawe-Taylor. 2009. Can eyes reveal interest? Implicit queries from gaze patterns. *User Modeling and User-Adapted Interaction* (2009).
4. Michael Argyle, Roger Ingham, Florisse Alkema, and Margaret McCallin. 1973. The Different Functions of Gaze. *Semiotica* 7 (1973).
5. Roman Bednarik, Shahram Eivazi, and Hana Vrzakova. 2013. A Computational Approach for Prediction of Problem-Solving Behavior Using Support Vector Machines and Eye-Tracking Data. In *Eye Gaze in Intelligent User Interfaces*. 111–134.
6. Roman Bednarik, Hana Vrzakova, and Michal Hradis. 2012. What Do You Want to Do Next: A Novel Approach for Intent Prediction in Gaze-based Interaction *(ETRA '12)*.
7. Boris Brandherm, Helmut Prendinger, and Mitsuru Ishizuka. 2007. Interest Estimation Based on Dynamic Bayesian Networks for Visual Attentive Presentation Agents *(ICMI '07)*.
8. Andreas Bulling, Daniel Roggen, and Gerhard Tröster. It's in Your Eyes: Towards Context-awareness and Mobile HCI Using Wearable EOG Goggles *(UbiComp '08)*.
9. Jean Carletta, Robin L Hill, Craig Nicol, Tim Taylor, Jan Peter De Ruiter, and Ellen Gurman Bard. 2010. Eyetracking for Two-Person Tasks with Manipulation of a Virtual World. *Behavior Research Methods* 42, 1 (2010).
10. Sicheng Chen, Miao Chen, Andreas Kunz, Asim Evren Yantaç, Mathias Bergmark, Anders Sundin, and Morten Fjeld. SEMarbeta: Mobile Sketch-gesture-video Remote Support for Car Drivers *(AH '13)*.

11. Mauro Cherubini, Marc-Antoine Nüssli, and Pierre Dillenbourg. Deixis and Gaze in Collaborative Work at a Distance (over a Shared Map): A Computational Model to Detect Misunderstandings *(ETRA '08)*.

12. Andrew J. Davison, Walterio W. Mayol, and David W. Murray. 2003. Real-Time Localisation and Mapping with Wearable Active Vision. In *Proceedings of the 2Nd IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR '03)*. 18–.

13. Kentaro Fukuchi, Toshiki Sato, Haruko Mamiya, and Hideki Koike. 2010. Pac-pac: Pinching Gesture Recognition for Tabletop Entertainment System *(AVI '08)*. ACM, 267–273.

14. Susan R. Fussell, Leslie D. Setlock, and Robert E. Kraut. Effects of Head-mounted and Scene-oriented Video Systems on Remote Collaboration on Physical Tasks *(CHI '03)*.

15. Susan R. Fussell, Leslie D. Setlock, and Elizabeth M. Parker. Where Do Helpers Look?: Gaze Targets During Collaborative Physical Tasks *(CHI EA '03)*.

16. Susan R. Fussell, Leslie D. Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer, and Adam D. I. Kramer. 2004. Gestures over Video Streams to Support Remote Collaboration on Physical Tasks. *Human–Computer Interaction* (2004).

17. Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. World-stabilized Annotations and Virtual Scene Navigation for Remote Collaboration *(UIST '14)*.

18. Mary Hayhoe and Dana Ballard. 2005. Eye Movements in Natural Behavior. *Trends Cogn. Sci. (Regul. Ed.)* 9, 4 (2005), 188–194.

19. Takatsugu Hirayama, Dodane Jean-Baptiste, Hiroaki Kawashima, and Takashi Matsuyama. 2010. Estimates of User Interest Using Timing Structures between Proactive Content-Display Updates and Eye Movements. *IEICE Trans. Inf. & Syst.* 93, 6 (2010).

20. Shinsaku Hiura, Kenji Tojo, and Seiji Inokuchi. 3DD Tele-direction Interface Using Video Projector.

21. Yoshio Ishiguro, Adiyan Mujibiya, Takashi Miyaki, and Jun Rekimoto. Aided Eyes: Eye Activity Sensing for Daily Life *(AH '10)*.

22. Robert J. K. Jacob. 1990. What You Look at is What You Get: Eye Movement-based Interaction Techniques *(CHI '90)*.

23. Patrick Jermann, Darren Gergle, Roman Bednarik, and Susan Brennan. Duet 2012: Dual Eye Tracking in CSCW. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion*. 23–24.

24. Roland S. Johansson, Gran Westling, Anders Bckstrm, and J. Randall Flanagan. 2001. Eye-Hand Coordination in Object Manipulation. *JOURNAL OF NEUROSCIENCE* 21, 17 (2001), 6917–6932.

25. Yvonne Kammerer, Katharina Scheiter, and Wolfgang Beinhauer. Looking My Way Through the Menu: The Impact of Menu Design and Multimodal Input on Gaze-based Menu Selection *(ETRA '08)*.

26. Shunichi Kasahara and Jun Rekimoto. JackIn: Integrating First-person View with Out-of-body Vision Generation for Human-human Augmentation *(AH '14)*.

27. Hirokazu Kato and Mark Billinghurst. Marker Tracking and Hmd Calibration for a Video-based Augmented Reality Conferencing System *(IWAR '99)*.

28. Adam Kendon. 1967. Some Functions of Gaze-Direction in Social Interaction. *Acta Psychologica* 26 (1967).

29. Seungwon Kim, Gun Lee, Nobuyasu Sakata, and Mark Billinghurst. Improving Co-presence with Augmented Visual Communication Cues for Sharing Experience through Video Conference *(ISMAR'14)*.

30. David Kirk and Danae Stanton Fraser. Comparing Remote Gesture Technologies for Supporting Collaborative Physical Tasks *(CHI '06)*.

31. Nikolina Koleva, Sabrina Hoppe, Mohammed Mehdi Moniri, Maria Staudte, and Andreas Bulling. On the Interplay between Spontaneous Spoken Instructions and Human Visual Behaviour in an Indoor Guidance Task *(COGSCI '15)*.

32. Robert E. Kraut, Susan R. Fussell, and Jane Siegel. 2003. Visual Information As a Conversational Resource in Collaborative Physical Tasks. *Human–Computter Interaction* 18, 1 (2003).

33. Takeshi Kurata, Nobuchika Sakata, Masakatsu Kourogi, Hideaki Kuzuoka, and Mark Billinghurst. Remote Collaboration using a Shoulder-worn Active Camera/Laser *(ISWC '04)*.

34. Hideaki Kuzuoka. Spatial Workspace Collaboration: A SharedView Video Support System for Remote Collaboration Capability *(CHI '92)*.

35. Hemin Omer Latif, Nasser Sherkat, and Ahmad Lotfi. Teleoperation Through Eye Gaze (TeleGaze): A Multimodal Approach *(ROBIO'09)*.

36. Taehee Lee and Tobias Hollerer. Viewpoint Stabilization for Live Collaborative Video Augmentations *(ISMAR '06)*.

37. Kana Misawa and Jun Rekimoto. Wearing Another's Personality: A Human-surrogate System with a Telepresence Face *(ISWC '15)*.

38. A. Monden, K. Matsumoto, and M. Yamato. 2005. Evaluation of Gaze-Added Target Selection Methods Suitable for General GUIs. *Int. J. Comput. Appl. Technol.* 24, 1 (2005).

39. Romy Müller, Jens R Helmert, Sebastian Pannasch, and Boris M Velichkovsky. 2013. Gaze Transfer in Remote Cooperation: Is it always Helpful to See What Your Partner Is Attending to? *The Quarterly Journal of Experimental Psychology* 66, 7 (2013), 1302–1316.

40. Hideyuki Nakanishi, Satoshi Koizumi, Toru Ishida, and Hideaki Ito. Transcendent Communication: Location-based Guidance for Large-scale Public Spaces *(CHI '04)*.

41. Jiazhi Ou, Lui Min Oh, Jie Yang, and Susan R. Fussell. Effects of Task Properties, Partner Actions, and Message Content on Eye Gaze Patterns in a Collaborative Task *(CHI '05)*.

42. Ken Pfeuffer, Jason Alexander, Ming Ki Chong, and Hans Gellersen. Gaze-touch: Combining Gaze with Multi-touch for Interaction on the Same Surface *(UIST '14)*.

43. Abhishek Ranjan, Jeremy P. Birnholtz, and Ravin Balakrishnan. Dynamic Shared Visual Spaces: Experimenting with Automatic Camera Control in a Remote Repair Task *(CHI '07)*.

44. Kshitij Sharma, Patrick Jermann, Marc-Antoine Nüssli, and Pierre Dillenbourg. Gaze evidence for different activities in program understanding.

45. Jaana Simola, Jarkko Saloj"arvi, and Ilpo Kojo. 2008. Using Hidden Markov Model to Uncover Processing States from Eye Movements in Information Search Tasks. *Cognitive Systems Research* 9, 4 (2008), 237–251.

46. Rajinder S. Sodhi, Brett R. Jones, David Forsyth, Brian P. Bailey, and Giuliano Maciocci. BeThere: 3D Mobile Collaboration with Spatial Input *(CHI '13)*.

47. Aaron Stafford, Wayne Piekarski, and Bruce Thomas. Implementation of God-like Interaction Techniques for Supporting Collaboration Between Outdoor AR and Indoor Tabletop Users *(ISMAR '06)*.

48. Sophie Stellmach and Raimund Dachselt. Look & Touch: Gaze-supported Target Acquisition *(CHI '12)*.

49. Cara A. Stitzlein, Jane Li, and Alex Krumm-Heller. Gaze Analysis in a Remote Collaborative Setting *(OZCHI '06)*.

50. Chiew Seng Sean Tan, Johannes Schöning, Kris Luyten, and Karin Coninx. Investigating the Effects of Using Biofeedback As Visual Stress Indicator During Video-mediated Collaboration *(CHI '14)*.

51. Jayson Turner, Jason Alexander, Andreas Bulling, and Hans Gellersen. Gaze+RST: Integrating Gaze and Multitouch for Remote Rotate-Scale-Translate Tasks *(CHI '15)*.

52. Alfred Yarbus. 1967. Eye Movements and Vision. *Plenum* (1967).

53. Xianjun Sam Zheng, Cedric Foucault, Patrik Matos da Silva, Siddharth Dasari, Tao Yang, and Stuart Goose. Eye-Wearable Technology for Machine Maintenance: Effects of Display Position and Hands-free Operation *(CHI '15)*.