5-2016

Stabilized annotations for mobile remote assistance

Omid FAKOURFAR

Kevin TA

Richard TANG

Scott BATEMAN

Anthony TANG
Singapore Management University, tonyt@smu.edu.sg

Citation

FAKOURFAR, Omid; TA, Kevin; TANG, Richard; BATEMAN, Scott; and TANG, Anthony. Stabilized annotations for mobile remote assistance. (2016). *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, United States, May 7-12*. 1548-1560.
Available at: https://ink.library.smu.edu.sg/sis_research/7894

Stabilized Annotations for Mobile Remote Assistance

Omid Fakourfar[1] Kevin Ta[1] Richard Tang[1] Scott Bateman[2] Anthony Tang[1]

[1]Department of Computer Science, University of Calgary, Canada
{omid.fakourfar, kta, richard.tang, tonyt}@ucalgary.ca
[2]Faculty of Computer Science, University of New Brunswick, Canada
scottb@unb.ca

ABSTRACT

Recent mobile technology has provided new opportunities for creating remote assistance systems. However, mobile support systems present a particular challenge: both the camera and display are held by the user, leading to shaky video. When pointing or drawing annotations, this means that the desired target often moves, causing the gesture to lose its intended meaning. To address this problem, we investigate annotation stabilization techniques, which allow annotations to stick to their intended location. We studied two annotation systems, using three different forms of annotations, with both tablets and head-mounted displays. Our analysis suggests that stabilized annotations and head-mounted displays are only beneficial in certain situations. However, the simplest approach of automatically freezing video while drawing annotations was surprisingly effective in facilitating the completion of remote assistance tasks.

Author Keywords

Mobile video conferencing; annotation systems; augmented reality; head-mounted displays; remote assistance.

ACM Classification Keywords

H.5.3. Information Interfaces and presentation: Group and Organization Interfaces – CSCW

INTRODUCTION

Mobile devices such as smartphones, tablets, and head-mounted displays have created new opportunities for expert guidance and remote assistance for ad hoc, unplanned situations (e.g., [1,11,13,21]). For example, an expert mechanic could help diagnose and guide an apprentice through the repair of an engine from a distance, where the worker shows the expert helper the problem using video from their smartphone (e.g., [5]). A growing body of design work is focused on building mobile systems to address such scenarios (e.g., [1,11,21]). One of the classic challenges in supporting remote collaboration is the need to reference objects or locations through video efficiently [6,9,14,26,27,34,40]. Mobile scenarios—where people hold the camera and can move it freely—present new challenges, including: providing the right camera view for the remote expert [11,21,22]; and, difficulties in manipulating and positioning cameras and devices while conducting physical tasks [11,22]. Collaborators often work around these challenges by adopting complex verbal negotiation, allowing them to get the information they need and to communicate their intended directions [14,22]. Many systems address this problem by providing an annotation subsystem, allowing a remote helper to annotate a scene as it is captured by a worker's camera.

CHI'16, May 07-12, 2016, San Jose, CA, USA

DOI: http://dx.doi.org/10.1145/2858036.2858171

In this paper, we focus on the design of this annotation subsystem, exploring several variations: stabilized annotations (annotations that stay affixed to the objects they were drawn upon) (like [11,12]), freeze-framed annotations (the video temporarily freezes when annotations are being drawn) (like [12]), and annotations atop live video. We also explore two device variations: a handheld tablet and a head-mounted display and camera. Although prior work has explored similar setups, previous evaluations have not made clear the benefits of stabilized annotations (if any) or head-mounted display and camera. Thus, we still do not have a good understanding whether or not such features are successful for improving collaborative support scenarios.

To address this question, we designed two studies to explore the trade-offs between different annotation techniques over a range of collaborative support tasks. In the first study, we examine the use of stabilized freehand annotations and freehand annotations atop video, contrasting their use with both tablet devices and head-mounted video displays. Our second study focuses on freeze-frame annotations (where the video freezes as soon as drawing annotation begins), and compared this technique with freehand annotations atop live video.

We found that stabilized annotations provided only marginal benefit to teams over normal annotations atop live video [11,13]. This is a consequence of how annotations were used to support interaction—rather than being revisited, annotations were generally used in an *ephemeral* way—once the message is clear, the annotations are generally no longer needed. We also found that head-mounted displays offered no meaningful benefit in terms of making or using annotations, but consistent with prior work, facilitated feedback on the current actions being carried out [21]. Nevertheless, participants generally preferred stabilized annotations with head-mounted displays over other conditions. Helpers had difficulty drawing annotations on an unsteady, shaky video stream; to this end, the freeze-frame annotation technique afforded parallel, independent work while addressing the issue of annotating on unsteady video. However, the transition from frozen to live video was confusing for the helper, because of a loss of awareness of what their collaborator was doing.

Our findings help to unpack the benefits of recent technological innovations for collaborative video support. We make two contributions in this work. First, we provide findings from two studies that show how both task and design factors can affect the utility of stabilized annotations and head-mounted displays. Second, we describe how annotations are used under different task factors, providing an explanatory rationale for when and why stabilized annotations and different device are likely to be of value. Together, we provide evidence that, contrary to conventional wisdom, the cheap-and-easy approach of freezing video while annotations are being created may be sufficient for supporting remote work in many cases (as opposed to more complex AR approaches).

RELATED WORK

Today, many people turn to online resources such as YouTube videos to receive support and instruction on completing everyday tasks (e.g. [30]). While such videos are useful for simple tasks, they are insufficient for more complex scenarios, where: (a) the solution is unclear, (b) an expert's guidance is required (e.g. repair of specialized motor), or (c) feedback is required following actions. To address these more complex scenarios, researchers have built systems designed for mobile remote support.

To set the stage for our work, we first describe the role of deixis, and how it is supported in many systems using an annotation subsystem. We then outline studies evaluating these systems, highlighting findings, and identifying gaps in the literature. Finally, we summarize by discussing the limitations in experimental task design as one potential reason for the mixed results of these studies.

Systems for Remote Support and Collaboration

Supporting remote assistance and collaboration through video has long been an interest of CSCW research. Early work generally focused on fixed-perspective video to support remote collaborative work (e.g., [18,40]), where the focus was on connecting remote collaborators with one another [4,18,27,34,40]. Single camera systems provide a constrained view of the workspace; meaning there is no ability to frame a remote scene, objects of interest, or people in the environment [10]. Newer designs explored multiple views of the workspace [10,30,33,35], or cameras that can be repositioned [15,29]. Several designs used multiple camera views of a remote workspace and displayed them across walls to help simulate a seamless connection between physical spaces [30,35]. Other designs provided overview+detailed views of a remote workspace, allowing remote workers to see both fine details of work and a contextual overview of the remote space [10,33,35].

Mobile Remote Support. The prototypes described above tend to rely on specialized spaces or equipment, meaning they would be difficult to use in ad hoc, mobile scenarios. To this end, recent work has explored how remote assistance can be augmented through mobile technologies (e.g., [8,13,38]). Such systems allow for both the camera (used to send a view of the local scene to a remote collaborator) and the view (the screen or other display used to show the remote scene) to be freely moved and repositioned by collaborators. These systems have employed mobile phones [12], tablets [8,11,13] and head-mounted cameras and displays [1,17].

To address the problems of holding the device and framing the scene, work has made use of a head-mounted camera with a head-mounted display [1]—freeing the worker's hands to be used in the main collaborative task, and allowing the focus of work to be easily captured. Several recent prototypes decouple the remote collaborator's view from the worker's camera [24,38,39]. This is done by modeling the remote workspace, and allowing the helper to explore the modeled virtual environment on their own.

System Support for Deixis and Annotations

Remote support systems need to support deixis—the use of gestures toward objects in the context of speech (e.g., "move this one there")—to facilitate the basic mechanics of collaboration [16,22]. Deixis supports common ground [7], and reduces the number of speech acts needed to complete tasks (e.g., [9]). Remote support systems provide support for deixis in a variety of ways, including telepointers [9], push-pins [13], freely drawn annotations [12,18,40], and representations of collaborators' arms [19,41].

Deixis in Mobile Remote Support Systems. Gesturing into a video scene can be problematic, as elements in the scene may move while annotations are being drawn [25]. Similarly, if the camera position moves, annotations no longer point to the right object or location [11,13,25]. Recent systems have addressed this problem in two ways: by freezing the video, or by anchoring the annotations to elements in the scene. The simple approach of freezing video while annotations are drawn ensures that the annotations remain in place with the objects in the scene [13,25]. Stabilized annotations (that anchor to elements in the scene) have been explored by using fixed points of view [19], by tracking the point of view of the camera so that the annotations can be correctly positioned [12], or by dynamically modeling a remote environment [12,13].

Despite the effort that has gone into developing these systems, their benefits for collaboration have remained somewhat unclear. We next develop a meta-analysis of previous system studies, highlighting open questions.

SYSTEM STUDIES AND TASK BREADTH

Researchers frequently contribute studies of their systems, where the expected results follow conventional wisdom about the benefits of a given technology (e.g., if two hands are needed to work on task, a head-mounted camera could be used). Yet frequently, study results do not support such expectations. Below, we review study findings, organized by the expected benefits of particular technologies used for remote support, and the creation and use of annotations. We synthesize this work making a case for the use of a broader set of tasks when assessing new remote support systems.

Summary of Studies

The majority of previous studies use "collaborative physical tasks" (e.g., [9,21,26]). In general, such tasks have a "worker" perform physical tasks, such as building objects (e.g., with Lego) with the support of a remote "helper," who has a full set of instructions on how to complete the task. These tasks are designed to mimic scenarios where the expert has more knowledge than the worker about the task and often involve inspecting the workspace (for parts, or the current state of the object), selecting the correct pieces or tools, and then directing how they should be used.

Assumption: Stabilized annotations are better than non-stabilized annotations

Because the camera is mobile, several authors have attempted to provide stabilized annotations that stick to the objects in the scene. Several studies have examined the simple approach of using video pausing, and still frames to stabilize the scene and objects for annotation (e.g., [1,8,11,13,25]). In some implementations, both parties have control over when a still image is shown instead of live video [1,8], while others limit the control to one party or the other [11,13,25]. A final variation is to automatically freeze the scene while annotations are being made [25].

Annotations on a still frame do not seem to provide meaningful benefit in terms of task time [25]. Bauer et al. report more variance in how much the feature was used (some participants used it frequently; other participants rarely) [1]. Of note, is that people generally seem to prefer automatic freezing compared to manual freezing [25].

Another approach is to track and create a 3D model of the environment, allowing virtual annotations to adhere to physical objects even when the scene is changed (e.g., [11,13,24]). Yet again, surprisingly, two studies comparing the use of stabilized vs. non-stabilized annotations [11,13] have not found meaningful task performance differences between these interfaces. Nevertheless, the majority of participants in both studies preferred stabilized annotations.

While intuitively, stabilized annotations in mobile support scenarios make sense, it is not clear why they have not resulted in increased performance. More work is needed to understand the circumstances under which stabilized annotations provide benefit, and the lack of performance benefits likely relates to type and form of study tasks.

Assumption: Head-mounted cameras are useful

A useful property of the head-mounted camera is that it can be operated hands-free (i.e. compared to a tablet or mobile phone), giving the wearer operational use of both hands without the need to hold or position a camera. Furthermore, the view from a head-mounted camera tracks the worker's visual focus, working area, and area of interest.

Fussell et al. [10] found that the head-mounted camera did not provide the remote helper with a desirable view, and in fact, was barely an improvement over audio only. The argument was that the view was too limited, preventing the remote helper from understanding the entire space (as compared to the workspace camera). Similarly, Johnson et al. [21] found that head-mounted cameras did not result in reductions in task time when compared to a tablet-based camera. *However*, head-mounted cameras did change the effectiveness of the collaboration; remote helpers could anticipate trouble and proactively provide help.

Assumption: Head-mounted displays are better than tablets

In combination with head-mounted cameras, several prototypes have used head-mounted displays [1,17,24]. This has the benefit of freeing both the worker's hands for work. Furthermore, it allows information, such as annotations, to be displayed directly atop the scene [13,39]. In contrast, a handheld device requires a worker to position the screen to view annotations, and refer back to the workspace to take action, splitting their attention.

Researcher have also explored different head-mounted display technology. See-through video approaches obscure the view of the world completely but show a video feed of the workspace [3]. See-through transparent displays show information on top of the real world (e.g., [24,39], Epson Moverio). Finally, information has also been displayed via a small peripheral screen (e.g., [1], Google Glass).

In our review we did not find any work comparing head-mounted displays to handheld devices for remote collaboration. A peripheral work by Zheng et al. [42] compared head-mounted displays with tablets for static instructions in automobile repair (i.e., without collaboration), and found that head-mounted displays offer no improvement in completion time over tablets. In spite of this, head-mounted displays were preferred over tablets.

Factors of Concern in Remote Support

Many expectations about the benefits of technologies for remote support have not been borne out in studies. However, we noted that most studies have tended to rely on simple variations of a construction-style collaborative physical task (e.g., Lego assembly), where a "helper" is given step-by-step instructions to pass along to a "worker", and both helpers and workers had explicit roles. It may be that these tasks have not allowed the advantages of the technologies to surface. In this work, we aimed to explore a wider breadth of task styles for studying remote assistance systems. To this end, we identified several task factors that could change interactions during task execution:

Locus of knowledge. Most collaborative physical tasks place the onus on the helper to provide direction to the worker to work towards a known solution (i.e. one-way information transfer). We wanted to also explore scenarios where the solution was unclear at the outset, and where the worker has local knowledge to bring to bear on the solution.

Movement & Size of Workspace. Johnson et al. [21] explore the role of having to physically move about in the workspace—that is, environments that cannot be captured in a single camera frame. We are also interested in the movement required of objects in the workspace itself—for example objects that need to be inspected or operated on from different perspectives (e.g., [25]), rather than simply focusing on objects on a simple flat surface.

Complexity of physical task. Finally, we were interested in varying the physical complexity of the tasks such that some tasks can only be accomplished with two hands (with one hand holding something, and another affixing or manipulating something else), to vary from many tasks that could be completed with a single hand.

Tasks for Studying Remote Support

We designed tasks for our studies that systematically varied each factor, as we expect different combinations to produce different kinds of interaction needs; we expected this to help tease out the benefits of the different technology designs for remote support.

Task A: Tangram. Participants solve a tangram puzzle, where the helper is given a silhouette of the target shape, and the worker is given multiple smaller shapes to construct the target. This is a physical problem solving task frequently relying on trial and error. Knowledge is shared (helper has access to completed figure – worker has access to pieces available and relative sizes), the size of the workspace (comprised of all the pieces) is medium sized, and while challenging, is not a complex task (it can be completed with a single hand).

Task B: Graph. Participants find a least-cost path between two nodes in a large graph (comprised of nodes and edges), where each participant is only given information about half the edges. The task here is essentially a graphical, problem-solving task where knowledge is explicitly distributed, and the physical complexity of the task is low.

Task C: Origami. The helper provides step-by-step origami instructions to the worker to fold a piece of paper, where the instructions require flipping and turning the object around. Here, the helper has all the knowledge, but there is substantial movement in the workspace (i.e. flipping the paper around on different axes), and the task has higher physical complexity, as it require two hands to make folds.

Task D: Lego Repair. The helper is given pictures of a 3D Lego structure (about the size of a basketball), and needs to direct the worker to repair an existing 3D Lego structure to match the one depicted in images. This task mirrors many existing tasks from prior work, but removes explicit step-by-step instructions and uses a moderately sized workspace that requires viewing from multiple perspectives.

In terms of external validity, our tasks are based on a long history of research in this area [1,9,21,26,27,34,35,42], and are also related to scenarios in everyday life. For example, the Lego Repair and the Graph contain many of the basic elements of home theatre setup where pointing and/or moving behaviors occur frequently; the Origami task mimics systematic tasks like assembling furniture.

STUDY 1

We designed an observational study to explore the differences between stabilized and non-stabilized augmented reality annotations in mobile video conferencing scenarios with two different device configurations (tablet vs. head-mounted display). The study focused on how annotations are used to support the communicative acts between collaborators, and how hardware configurations might affect these processes.

System Overview

Our system was designed to optionally provide world stabilized virtual annotations that would seemingly stay affixed to real world objects viewed through the system's interface. Our system was developed on the Unity game engine along with Qualcomm's Vuforia for AR marker tracking. As shown in Figure 1, the system provides several simple annotation tools: line tool, freehand drawing, eraser, change of thickness, six colors and a 'clear all' tool.

In our system, stabilized annotations are drawn on *annotation planes*—flat, semi-transparent virtual surfaces visible only through the systems interface. While other approaches rely on sophisticated techniques to place these planes based on reconstructed scenes [12], for the purpose of our work, we place them atop surfaces that are likely to be annotated. Thus, annotations (made on the annotation planes) remain anchored to the surface upon which they are drawn. This plane is semi-translucent (like frosted glass), meaning that annotations on the plane are visibly detached from (but clearly associated with) the surface. With the Tangram and Graph tasks, we affixed this plane to the dominant surface. With the Origami task, we created two

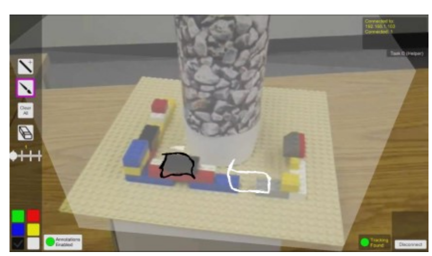

Figure 1. Viewing the Lego task through the system interface. Annotations made on an annotation plane (transparent white box) are visible.

planes (one for each side of the origami paper). In the Lego task, the final structure has an inherent 'box' shape, and so we created four such invisible planes around the box.

The worker used either the Movario BT-200 as the head-mounted display or the Asus MemoPad 10" Android tablet. The HMD display had shades, which meant that the wearer could only see the screen (and not see beyond the screen). Finally, the helper used a Microsoft Surface as a drawing tablet with a capacitive touch pen. All devices were connected on a Linksys Dual band router at 2.4Ghz.

Participants
We recruited sixteen pairs of participants (32 total; 13 female) through posters displayed on a university campus. All participants knew their partner prior to the study. Participants ranged from 18 to 32 years of age with an average of 23.4 years. On a five-point scale (5 being very experienced), participants self-rated experience with video conferencing as 2.9 (sd. 1.33, med. 3), and experience with augmented reality was low (avg. 1.71, sd. 0.87, med. 1).

Procedure
Participants were given a demographics questionnaire and taught how to use several aspects of the system: how to draw annotations with different tools, change the drawing color and thickness and turn the annotations on and off (Figure 1). Participants were then able to try all combinations annotation-type (stabilized or non-stabilized annotations), with two device types (a tablet or head-mounted display). Participants were assigned to be either the 'helper' or the 'worker'. The worker would have the objects being manipulated in front of them. Participants then completed a simple training task using the tablet version of the system, which provided participants the opportunity to get used to the type of tasks in the study. Once the training task was completed, participants completed the four tasks (described above); task presentation was counterbalanced across groups.

We limited participants to seven minutes for each task, as we found from piloting the tasks that this allowed participants to comfortably complete or nearly complete each task, but allowed us to put an upper limit on the study length. After each task, we conducted short interviews to elicit participant experiences with the particular task and technology combination they completed. Finally, a questionnaire was administered that asked participants to reflect on their overall experience. The study required around 75 minutes, and participants were paid $20.

Setup
Each participant pair sat in the same room, back-to-back, allowing us to mimic a remote scenario in that they could not see one another, but could easily hear one another. The helper used a tablet to receive video from the worker's camera and to communicate annotations. The worker used an Android tablet with rear-facing camera or a head-mounted display to share their environment with the helper and to receive annotations. The helper was the only person who was able to draw the annotations.

Data and Analysis
The tasks were setup as 2×2 within subjects design (stabilized vs. non-stabilized, and head-mounted display vs. tablet). There were also 16 (4×4) unique Task×Device configurations. Each participant pair went through the four tasks using only one of the combinations, for a total of 64 trials. We collected video and audio of the workspace, logged participants' interactions and took field notes. For each study, we took note of our observations right after the study as a content log. We then transcribed the video recordings and took note of three things: (a) the form of the annotation (i.e. how it actually appeared), (b) its relation to other nearby annotations (temporal and spatial), and finally (c) its role in the interaction between partners.

To analyze the data, we used interaction analysis [23] to analyze the annotations in context of the collaborative activities. We transcribed 226 total annotations that related to critical incidents across the experimental 64 trials (i.e., there were a larger number of annotations recorded, but 226 critical incidents were used in our analysis, which referred to meaningful communicative acts throughout the tasks).

STUDY 1: FINDINGS
Overall, annotations played an important role in how participants completed the tasks. Of particular interest to us was how people might use annotations differently between task types. We also observed interesting tradeoffs between stabilized and non-stabilized annotations depending on the type of technology used, which we described below.

Role of Annotations
Through our analysis of annotations, we arrived at a three-category schema. Table 1 summarizes the three categories of annotations that we observed (reference, procedural, pointer; described below), along with the number of instances across each of the four tasks. Because we did not observe a noticeable difference in the distribution of these annotations across the hardware dimension or stabilized/non-stabilized dimension, we do not break out the distribution in this way.

Reference annotations. Reference annotations were intended to be used over time (or referred to at a later time). That is, while the intent of the communication was to convey information in the moment, there was also the intention for use in the future. Two very common examples of reference annotations we saw included "end-state" annotations, which showed how the current task components would appear once the task was complete, and

Annotation Type	Tangram	Graph	Origami	Lego
Reference	28	35	13	11
Procedural	9	0	43	1
Pointer	18	7	5	56

Table 1. Annotation type counts across different tasks

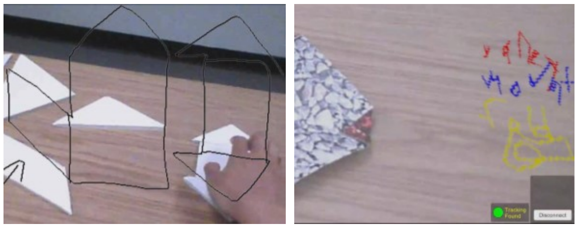

Figure 2. Left: An "end-state" annotation in the Tangram task. Right: A "legend" annotation in Origami task.

"legend" annotations, which explained mappings between symbols and meaning. Figure 2 (left) illustrates an example of an "end-state" annotation, which we frequently saw in the Tangram task– helpers would provide the workers with an outline of the desired object. Figure 2 (right) illustrates an example of a "legend" annotation observed in the Origami task. Here, the annotation indicates what fold each color will represent, and this group anchored the legend to the corner of the paper so the legend would always appear regardless of the state of the task.

Procedural annotations. Procedural annotations depicted actions that were needed on or close to the location of task objects. These annotations were intended to convey a verb or an action. Figure 3 (left) illustrates an example of this type of annotation in the Origami task, where the helper has indicated where to fold, and how to do the fold. Figure 3 (right) shows the helper communicating how to rotate a tangram piece using side-by-side shapes connected with an arrow. Note this second example differs from an "end-state (reference) annotation", where after the communicative step is complete, the annotation is no longer needed; in an end-state annotation, it would be left for later use.

Pointer annotations. We also observed a variety of pointers to temporarily point at objects in the workspace. We saw instances of dots, arrows, circles or scribbles to point at objects (or targets) in the workspace. Figure 4 illustrate instances of these in the Lego and Origami tasks.

Annotations in Task Completion

Participants used annotations as a means to support the completion of the tasks. Yet as illustrated in Table 1, the distribution of such annotations varied widely across tasks. We examine this result next.

Tangram Task: Most participants completed this task by first drawing (and then leaving) an outline of goal shape—

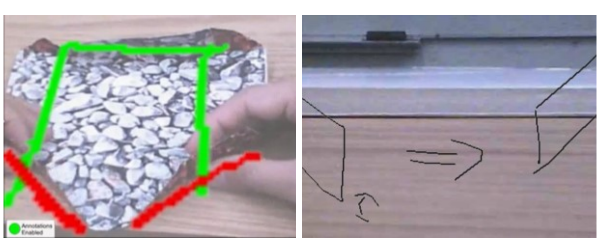

Figure 3. Left: Fold lines as a procedural annotation in Origami task. Right: A procedural annotation asking for the rotation of a Tangram piece.

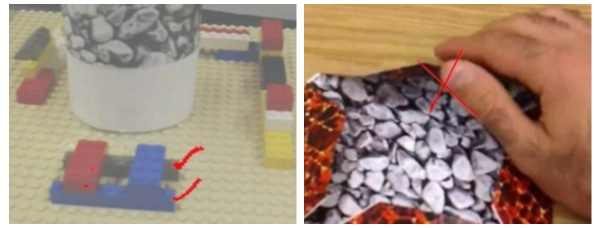

Figure 4. Left: Helper marks correctly placed Lego blocks. Right: Helper puts "X" on the edge that should not be folded.

an end-state reference annotation. Then, most pairs would operate in such a way that the helper would direct the worker to pick up certain pieces via pointer annotations, and indicate how to put them together. This kind of behavior was unusual to us, but later interviews provided some insight. Because the Tangram outline was a free-hand drawing, it was difficult for the helper to interpret the relative sizes of the target object (and its components) in relation to the shapes in the physical space. Given that our Tangram task provided several sizes of each shape, it became problematic for the worker to match the size of the drawn shape with its actual size and to meaningfully operate on their own without helper guidance.

Graph Task: In this task, pairs tended to use mainly reference annotations, putting together a consolidated drawing of the graph by drawing edges and annotating them with weights. Notably, participants employed a wide variety of strategies to complete the task.

As noted earlier, the task was designed to encompass two competing needs. First, there was a need to provide an overview of the space (i.e., the entire whiteboard needed to be seen, requiring the worker to back up so it could be in frame for the helper). Second, the need to work up close (i.e., workers need to approach the board to reach the board and make a change). Five pairs addressed this problem by recreating the entire graph as a set of annotations– essentially creating a radar view [11] to provide an overview of the space. Seven groups used only annotations on the physical whiteboard; the helper referred to nodes using pseudonyms, and rejected or confirmed workers' pointer annotations drawn on the board. A further three groups used a mixed set of annotations (some as digital annotations, some as markings on the whiteboard), while one group refrained from using annotations altogether. Both digital and physical annotations allowed groups to think aloud without the additional overhead of the worker struggling to frame the whiteboard correctly for the helper.

One pair used a mixed set of annotations, which allowed them to work independently. In this group, using non-stabilized annotations, the helper copied his node symbols from his tablet screen, drawing his share of edges. Meanwhile, the worker's tablet was placed on the table and the worker drew her share of the edges on the whiteboard. Once the helper finished copying the graph nodes, the two were able to draw edges independently. However, we did not observe meaningful performance gains even though

	HMD-S	HMD-NS	HH-S	HH-NS	Overall
Tangram	398.5	255	290.2	325.7	317.3
Graph	383.2	363.5	316.7	303.2	341.6
Origami	419.7	373.5	380	375	387.0
Lego	381.2	358.5	239.2	318	324.2

Table 3. Average completion time of tasks across conditions (in seconds)

work was divided and in parallel, because the stabilized annotations required them to work sequentially; approaching the board would not provide a useful (overall) view of the nodes for the helper.

Origami: This task was completed in a step-by-step fashion with the use of procedural and end-state annotations. A very common annotation was to draw a line to indicate the next fold's seam, along with an arrow to indicate the direction in which to make the fold. End-state annotations were occasionally used to depict what the piece should look like at the end of a step (i.e., as a reference).

Lego: The vast majority of annotations that we observed in the Lego task were pointer annotations. Such annotations indicated what piece to pick up, or a location for a piece. 14 groups used pointer annotations to complete the task.

Life of an Annotation

Creation: Annotations were typically quick, rough marks that were made within the context of conversation, aiding the ongoing dialogue. In a few limited cases (five groups) we observed participants taking the time to draw something very carefully (e.g., a realistic-looking end-state annotation of an origami piece). These were kept for a longer duration and were generally reference or end-state type annotations. Otherwise, all groups used quick, rough sketches and lines rather than spending time on more detailed drawings.

One of the core benefits of stabilized annotations was that when an annotated object returns into view, the annotation also returns into the view– that is, it remained "stuck" to the object. Yet, creating such annotations can be challenging. This is because annotations were drawn from the camera view provided by the worker, and if the worker moved the camera, even slightly, annotations were no longer properly aligned (regardless of the system used). Participants commonly noted this, and we frequently heard helpers yelling, "Don't move", "Stay still", and so forth. The vignette (Group 11) below illustrates a helpers' frustration:

Worker: (Holds tablet with one hand, Origami paper with the other)
Helper: "You have got to fold… I am going to draw this. Stay still!" [Begins drawing] "Fold like that…"
Worker: (Shifts the paper/camera slightly)
Helper: "Oh, you just moved it! Okay, don't move the paper!"

This also became frustrating for the workers. For example, when using the tablet, workers needed to hold very still, resulting in an awkward, unnatural position. With the HMD, workers would sometime hold the headset to keep the video as still as possible for the helper.

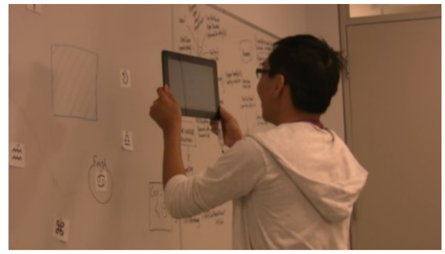

Figure 5. Worker moving the tablet close to the whiteboard to draw helper's attention to a symbol

"Yeah [the head-mounted display] was heavy. I had to hold it with one hand the whole time. And to see better, I had to keep holding it up my nose" – Group 15 Worker

Life and End: In the vast majority of cases, annotations were erased almost immediately after the corresponding action was completed—using either the erase tool, or "clear screen" tool. In contrast, some groups made use of longer-lived reference annotations as a strategy to complete their tasks. As illustrated in Figure 2 (right), Group 2 created a legend to illustrate a mapping between colors and "types of folds" for the Origami task. Similarly, some pairs would reuse these annotations as a magiclens-like overlay [2] either to confirm the correctness of a step, or to "compare notes". For example, when Group 9's worked on the Origami task, the helper drew the final state of the step, the worker completed the task with the tablet on the table, and then held up the tablet to see if he had done it correctly. The annotations were then cleared to make space for next steps.

Other forms of Communication

Camera Mobility: In four groups, the tablet's mobility was used as a conversational resource. As illustrated in Figure 5, a worker moves the tablet back and forth, right up close to the board (Graph task). The purpose of this "zooming" action was to draw the helper's attention to the particular symbol. We did not see such actions using the head-mounted display.

Hand Gestures: Despite the fact that we had no mechanism for the helpers to convey hand gestures to workers, (they would often use annotations to convey such intentions), helpers frequently made hand gestures without actually drawing anything, and without the gestures actually being visible to the other party. These actions seemed mostly unintentional: pointing towards certain objects on the video feed, or doing the origami folds with hands in the air and trying to link two nodes in the graph task. We observed this behavior in 5 of the 16 groups.

Head-Mounted Display and Camera Use

Using the head-mounted display and camera allowed workers to use both hands to complete the task, something that was extremely valuable in the Origami task. When workers wore the head-mounted display, helpers were able to provide timely feedback on the actions the workers were doing, correcting them if they were completing the task incorrectly. In contrast to this, workers in the tablet conditions would frequently put the tablet down (i.e., on the table) to complete tasks– particularly those that required two hands (i.e., Lego and Origami). Doing so prevented the helper from seeing what the worker was doing, because the camera was touching the table. This frequently allowed workers to go down the wrong path, and was observed in 12 of 16 pairs. The following vignette (Group 3) illustrates how the worker, in placing the tablet on the table, ends up doing something incorrectly because the helper cannot see the worker's actions until it is too late:

Worker: (Holding the tablet)
Helper: "Let me draw the line you need to fold. Put this triangle out."
Worker: (Puts tablet on table, folds paper, and picks up the tablet again to show the outcome) "Like this?"
Helper: "What did you do?! Oh no, no, no…"

Surprisingly, the HMD was not a complete solution. The HMD was sees as having low-resolution, giving a poor field-of-view at an awkward angle, and heavy. Our implementation's annotations were not translucent enough and often occluded the worker's view of the workspace. On top of this, the monocular camera only gave a 2D view, meaning that helpers lost considerable depth perception. Finally, in six groups, the helper would lift the goggles (or look over the top) to see the workspace when working (rather than look at the workspace through the goggles).

Preferences by Task

Table 2 shows worker device and annotation-type preference by task condition (one group failed to complete the questionnaire). We saw general preference for the head-mounted display over the tablet, and stabilized annotations were preferred over non-stabilized annotations.

Task Completion Times

The average completion time of tasks across different configurations is shown in Table 3 (recall that tasks were limited to 7 minutes, or 420 seconds). In terms of task difficulty, the participants seemed to struggle more with the Origami task (9 pairs failed to fully complete the Origami task). Interestingly, these results suggest that although the head-mounted display and stabilized annotations were popular among participants, they actually do not provide benefit regarding completion time. In fact, this condition has always yielded the longest completion time.

Summary of Study 1

Based on our observations, we were able to classify the annotations into three categories: reference, procedural and pointer annotations. We noticed that some types of annotations are drawn more frequently in certain tasks based on task characteristics and requirements.

	HMD	Tablet	Stabilized	Non-stabilized
Tangram	13	2	10	5
Graph	7	8	12	3
Origami	10	5	11	4
Lego	11	4	12	3

Table 2. Workers' device and annotation preferences

Correspondingly, we observed participants employing different strategies for different tasks.

We also examined the performance of the two device configurations employed by the worker in detail: tablet and head-mounted display. While head-mounted displays provides key benefits for some tasks mainly because it frees up both the workers' hands, it still has limitations; it is heavy and it can be difficult to see through.

Finally, we introduced and compared stabilized and non-stabilized annotations for different tasks. Although the stabilized annotations could make specific types of annotations possible and were more preferred by the workers, they did not seem to provide substantial performance benefit in the tasks.

STUDY 2

In Study 1, we found that stabilized annotations were potentially useful in many situations. However, our stabilized annotation systems required a sophisticated library and trackers to be incorporated into the workspace and objects. Further, we observed difficulties for participants when the helper intended to draw annotations while the worker's camera was shaking, leading to frustration and distraction from the main task.

To address these issues, we also designed a tablet-only video conferencing prototype called F2 ('freeze-frame'), which freezes the current video frame while an annotation is being drawn. F2 allows users to draw annotations on a stable, consistent image, avoiding some of the drawbacks of drawing annotations over live video (which were observed even when using an AR-stabilized system). We were motivated to explore the simple annotation feature, because it could be easily included in traditional video conferencing systems such as Skype, Hangouts, Facetime, etc. [25].

The F2 System

F2 users are able to see the camera feed from their partner's device and draw annotations over them. This is done on a laptop or desktop with a mouse, while users on a smartphone or tablet can draw using a finger. Annotations drawn over the camera feed are visible to both users. When the user starts drawing, F2 freezes the camera on its current frame, and keeps it frozen as they continue drawing. After they release their mouse or fingertip, the annotations and current frame stay on-screen for approximately 2 seconds before fading away and returning to the live camera feed; frame-freezing can be turned on or off through a toggle. F2 is a web-based video chat system using HTML5 and Javascript for the interface, and a Node.js server. Touch support for drawn annotations is provided by Hammer.js. Our system is very similar to that of Kim et al. [37].

Study

We designed an observational lab study to evaluate and compare the use of F2's frame-freezing annotations to drawing annotations over live video ('live-video'), without freeze-framing. While this second formative study does not make use of our AR-stabilized annotation study, we made use of the same tasks and analysis as in Study 1; allowing us to make comparisons between studies. This study focused on addressing two questions:

- Does the simple freeze-frame approach for annotating provide benefit over annotating live video?
- Does F2 avoid the problems with in creating AR-stabilized annotations (observed during Study 1)?

This study followed the same design as Study 1. We recruited 6 pairs of participants (12 participants; 5 females) who already knew one another. Participants ranged from 24 to 32 years of age with an average of 27.5 years. As with Study 1, participants were asked to rate their previous experience with mobile video conferencing and augmented reality on a scale of 1 (none) to 5 (experienced). Mobile Video conferencing: range 1 to 5, mean 3.5, s.d. 1.45, med. 4. Augmented reality experience: range 1 to 4, mean 2.5, s.d. 1.17, med. 2.5 (6 rated 3 or greater).

For this study, we only focused on the tablet device configuration and two of the four tasks previously described: the Lego repair task and the Origami task. These tasks were selected because we found they consistently demonstrated the identified issues with the AR-stabilized system. The two tasks and the annotation style (frame-freezing and live video) were set up in a 2x2 within-subjects design, where the participant pairs used both annotation styles for both tasks. The tasks had two variations to avoid learning bias. Data collection and remuneration were administered as with Study 1.

Study 2: Findings

In Study 2, we found the same three types of annotations observed as in Study 1 (reference, procedural and pointing), which were executed in much the same style. Again, the annotations varied depending on the task. In the Lego task, helpers marked pieces with dashes or small circles to indicate a piece to be picked up, and drew arrows to convey where to move pieces (pointer). In the Origami task, helpers often drew lines to indicate where to fold and drew arrows to emphasize the type or direction of the fold (procedural). Helpers often drew an outline of what the paper should look like after the fold (reference, end-state). We also observed similar camera usage to Study 1; e.g., the worker often needed to set down the tablet in order to use both hands.

Advantages of Frame Freezing

As with Study 1, we observed workers frequently moving the camera before the helper finished drawing an annotation. In these cases, helpers often found frame-freezing easier to use when drawing annotations than with live video. When using live video, workers were unable to stay still, resulting in a shaky camera feed over which the helper could not easily draw. Echoing our previous observation, helpers often had to tell workers to stay still for a moment so they could draw.

In some instances, workers anticipated the instructions that helpers were giving while annotations were being drawn. In these situations, the worker would abruptly shift the tablet's camera, disorienting the helper who was still drawing. In these cases, the worker is able to parse and interpret the helper's intentions faster than they could draw. When frame-freezing was used this did not seem to be a problem. Because the helper was drawing an annotation over a still image, they could continue drawing uninterrupted, regardless of the worker's movements. Frame-freezing also benefited the worker, as holding the tablet absolutely still (to keep live video still) for the helper to draw upon was cumbersome and distracted from the collaboration.

Live video is still important

While frame freezing was useful in many situations, we also observed instances where it caused a loss of awareness for the helper. This occurred in situations where the worker started actions before the instructions were complete. During this time the helper continued drawing on the frozen video, while worker had been focusing on completing a task; both were unaware that the video was about to restart for the helper. The worker in the meantime had changed the camera angle while working on the task (sometimes putting the tablet down to use both hands). When live video would return, it resulted in some disorientation and confusion for the helper, not knowing what they were looking at or how they arrived at this view. This required the helper to reorient and reinterpret their scene before continuing. In this aspect, live video provided the benefit that any changes by the worker were immediately visible. Live video also allowed for faster and more immediate feedback. With live video, helpers were able to see their worker moving Lego pieces to the wrong spots sooner, and could immediately correct them. Frame freezing delayed such feedback until the view returned to live video.

DISCUSSION

Utility of Stabilized Annotations

Overall, and consistent with previous work [13], we did not find stabilized annotations to be a clear winner. However, our framework of the observed use of annotations does provide an indication of exactly when stabilized annotations were valuable. Stabilized annotations were greatly beneficial when the references made were *reference* type annotations, as opposed to procedural or pointing annotations. This makes sense, because with reference annotations the information that an annotation was meant to convey was needed over a longer period of time. Because stabilized annotations stick to a spot consistently, they do not lose their context and allow people to recall their intended meaning easily, even if many actions separated their creation and eventual (re)use.

In contrast, stabilized annotations provided little benefit when annotations were meant to be more ephemeral, which is the case for procedural or pointing type annotations. Once a helper had made these short, quick marks, there was really no need for their continued use. In these cases, simple telepointers might suffice, although, we still saw the use of arrows, curved lines, etc., suggesting that having our free-drawing annotation tool was still valuable.

As Table 2 illustrates, participants mostly prefer stabilized annotations to non-stabilized annotations. While this may be partially due to an effect of novelty, it does provide some evidence that our approach of using annotation planes was a usable and effective way to make annotations, and that stabilized annotations did not detract from the ability to complete tasks even if they were not really needed.

Previous work has used the term "annotations" broadly to refer to simple pushpins, marks and predefined shapes placed into the workspace. Rarely have freely drawn 2D annotations been used. This at least partially, is due to the challenge of meaningfully drawing 2D annotations into a 3D space [12]. We avoided this problem by predefining drawing planes. We identified these drawing planes by anticipating where and how people would draw annotations to support a particular task. While this additional design thought should be done carefully, our work shows that this approach is both feasible and can be a beneficial way that AR-stabilized annotations can support collaborative work.

Further, the additional clarification into when and how stabilized annotations are useful is a direct result of our study design, which employed a range of different tasks varied over our identified task factors. Previous evaluations have usually focused on physical tasks, where the helper has all of the information, and needs to direct the worker in a step-by-step process to solve the task (like our Lego task). Our results show that having a range of tasks is extremely important in the assessment of new technology and techniques. Our task factors can be employed and evolved to guide the design of experimental tasks to better represent the many forms that remote support can take.

Advantages/disadvantages of Temporal Stabilization

With our first study, we found that while stabilized annotations had some utility, they only provided stabilization *after* the annotation was drawn. However, we observed most problems arose from the lack of stabilization *during* the annotation creation. Our second study used frame freezing as a simple approach to provide a form of *temporal stabilization*. Temporal stabilization means that the entire scene stays consistent while annotations are created. This simple design was extremely effective, making it easier for the helper to draw annotations, and eliminating the need for a worker to hold the camera perfectly still while annotations were being drawn.

However, we found that returning to live video from a frozen frame was sometimes challenging for the helper—particularly when the perspective on the workspace had changed. Perspective change tended to be very disruptive, as it meant that helpers would need to reorient themselves to the scene. Furthermore, because they had not observed what the worker had done in the meantime, helpers were unable to provide "real time" feedback on the worker's activities. Similar systems should provide either manual control over when to return to live video, or provide a smaller live view (e.g., in the corner of the screen). This would allow annotations to be made over a stable image, while allowing the helper to monitor worker actions.

Head-Mounted Cameras and Displays for Collaboration

Consistent with much of the prior literature, a head-mounted camera does provide some utility. Workers are freed from the burden of holding another device, and able to use both hands to work. Further, helpers are provided with a continuous video stream of worker activity that they can monitor and provide feedback on. Participants also preferred the head-mounted for three of four tasks. It was not preferred for the Graph task, where the combination of the task design and technology meant that only one person could work at a time. When close to the board, only the worker could work (because the helper could not see enough information), and when far away, only the helper could work (because the worker could not reach the board).

CONCLUSIONS

Current designs of mobile video conferencing systems for remote assistance have placed little focus on the relationship between gestures, annotations and device type. Previous work suggests remote assistance scenarios need to be supported by such technologies in order to improve performance. In this paper, we designed and evaluated two systems to support remote assistance. The first system allowed users to use AR stabilized annotations on live video and incorporated a head-mounted display to free workers' hands. The second system, introduced a frame-freezing feature that lets the helper focus on drawing annotations without worrying about shaky cameras on the worker's side. We saw that although stabilized annotations can improve performance in specific tasks, they do not necessarily outperform non-stabilized annotations in all tasks. Head-mounted displays were valuable for freeing up the workers' hands. Meanwhile, frame-freezing shows promise for providing a simple mechanism to allow helpers to create stabilized annotations.

Our two studies have provided new insights on the ways annotations and different device configurations support communication in remote assistance tasks. Based on our findings from our studies, we have outlined several implications that will direct the design of future mobile video tools for remote support.

ACKNOWLEDGEMENTS

We thank NSERC and XMG Studios for funding and supporting this work. We thank Frederico Schardong and Nguyen Cong Van for developing F2. We also thank Freepik from www.flaticon.com for their UI icon design.

REFERENCES

1. Martin Bauer, Gerd Kortuem, and Zary Segall. 1999. "Where Are You Pointing At?" A Study of Remote

Collaboration in a Wearable Videoconference System. *In Proceedings of the 3rd IEEE International Symposium on Wearable Computers* (ISWC '99). IEEE Computer Society, Washington, DC, USA, 151-161.

2. Eric A. Bier, Maureen C. Stone, Ken Pier, William Buxton, and Tony D. DeRose. 1993. Toolglass and magic lenses. *Proceedings of the 20th annual conference on Computer graphics and interactive techniques - SIGGRAPH '93*, ACM Press, 73–80. http://doi.org/10.1145/166117.166126
3. Mark Billinghurst, Adrian Clark, and Gun Lee. 2015. A Survey of Augmented Reality. *Foundations and Trends® in Human–Computer Interaction* 8, 2-3: 73–272. http://doi.org/10.1561/1100000049
4. Sara A. Bly, Steve R. Harrison, and Susan Irwin. 1993. Media spaces: bringing people together in a video, audio, and computing environment. *Communications of the ACM* 36, 1: 28–46. http://doi.org/10.1145/151233.151235
5. Jed R. Brubaker, Gina Venolia, and John C. Tang. 2012. Focusing on shared experiences. *Proceedings of the Designing Interactive Systems Conference on - DIS '12*, ACM Press, 96. http://doi.org/10.1145/2317956.2317973
6. Bill Buxton. 2009. Mediaspace – Meaningspace – Meetingspace. In *Media Space: 20+ Years of Mediated Life*, S. Harrison (Ed.). Springer, London, UK, 217-231.
7. Herbert H. Clark. 1996. *Using Language*. Cambridge University Press.
8. Veronika Domova, Elina Vartiainen, and Marcus Englund. 2014. Designing a Remote Video Collaboration System for Industrial Settings. *In Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces* (ITS '14). ACM, New York, NY, USA, 229-238. http://doi.acm.org/10.1145/2669485.2669517
9. 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. Hum.-Comput. Interact. 19, 3 (September 2004), 273-309. http://dx.doi.org/10.1207/s15327051hci1903_3
10. Susan R. Fussell, Leslie D. Setlock, and Robert E. Kraut. 2003. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '03). ACM, New York, NY, USA, 513-520. http://doi.acm.org/10.1145/642611.642701
11. Steffen Gauglitz, Cha Lee, Matthew Turk, and Tobias Höllerer. 2012. Integrating the physical environment into mobile remote collaboration. *In Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services* (MobileHCI '12). ACM, New York, NY, USA, 241-250. http://doi.acm.org/10.1145/2371574.2371610
12. Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. In touch with the remote world: remote collaboration with augmented reality drawings and virtual navigation. In*Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology* (VRST '14). ACM, New York, NY, USA, 197-205. http://dx.doi.org/10.1145/2671015.2671016
13. Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. World-stabilized annotations and virtual scene navigation for remote collaboration. *In Proceedings of the 27th annual ACM symposium on User interface software and technology* (UIST '14). ACM, New York, NY, USA, 449-459. http://doi.acm.org/10.1145/2642918.2647372
14. Darren Gergle, Robert E. Kraut, and Susan R. Fussell. 2004. Action as language in a shared visual space. *Proceedings of the 2004 ACM conference on Computer supported cooperative work - CSCW '04*, ACM Press, 487. http://doi.org/10.1145/1031607.1031687
15. Pavel Gurevich, Joel Lanir, Benjamin Cohen, and Ran Stone. 2012. TeleAdvisor: a versatile augmented reality tool for remote assistance. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12). ACM, New York, NY, USA, 619-622. http://doi.acm.org/10.1145/2207676.2207763
16. Carl Gutwin, Mark Roseman, and Saul Greenberg. 1996. A usability study of awareness widgets in a shared workspace groupware system. *Proceedings of the 1996 ACM conference on Computer supported cooperative work - CSCW '96*, ACM Press, 258–267. http://doi.org/10.1145/240080.240298
17. Weidong Huang and Leila Alem. 2013. HandsinAir: a wearable system for remote collaboration on physical tasks. *In Proceedings of the 2013 conference on Computer supported cooperative work companion* (CSCW '13). ACM, New York, NY, USA, 153-156. http://doi.acm.org/10.1145/2441955.2441994
18. Hiroshi Ishii, Minoru Kobayashi, and Jonathan Grudin. 1993. Integration of interpersonal space and shared workspace: ClearBoard design and experiments. *ACM Transactions on Information Systems* 11, 4: 349–375. http://doi.org/10.1145/159764.159762
19. Shahram Izadi, Ankur Agarwal, Antonio Criminisi, John M. Winn, Andrew Blake and Andrew W. Fitzgibbon. 2007. C-Slate: A Multi-Touch and Object Recognition System for Remote Collaboration using Horizo. In *Second IEEE International Workshop on Horizontal Interactive Human-Computer Systems*

Tabletop 2007 October 10-12, 2007, Newport, Rhode Island, USA. pp. 3-10. http://doi.ieeecomputersociety.org/10.1109/TABLETOP.2007.5

20. Hyungeun Jo and Sungjae Hwang. 2013. Chili: viewpoint control and on-video drawing for mobile video calls. In CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13). ACM, New York, NY, USA, 1425-1430. http://doi.acm.org/10.1145/2468356.2468610
21. Steven Johnson, Madeleine Gibson, and Bilge Mutlu. 2015. Handheld or Handsfree?: Remote Collaboration via Lightweight Head-Mounted Displays and Handheld Devices. *In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '15). ACM, New York, NY, USA, 1825-1836. http://doi.acm.org/10.1145/2675133.2675176
22. Brennan Jones, Anna Witcraft, Scott Bateman, Carman Neustaedter, and Anthony Tang. 2015. Mechanics of Camera Work in Mobile Video Collaboration. *In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15). ACM, New York, NY, USA, 957-966. http://doi.acm.org/10.1145/2702123.2702345
23. Brigitte Jordan and Austin Henderson. 1995. Interaction Analysis: Foundations and Practice. *The Journal of the Learning Sciences*, 39-103.
24. Shunichi Kasahara and Jun Rekimoto. 2014. JackIn: integrating first-person view with out-of-body vision generation for human-human augmentation. *In Proceedings of the 5th Augmented Human International Conference* (AH '14). ACM, New York, NY, USA, Article 46, 8 pages. http://doi.acm.org/10.1145/2582051.2582097
25. Seungwon Kim, Gun A. Lee, Sangtae Ha, Nobuchika Sakata and Mark Billinghurst. 2015. Automatically Freezing Live Video for Annotation during Remote Collaboration. In CHI '15 Extended Abstracts on Human Factors in Computing Systems (CHI EA '15). ACM, New York, NY, USA, 1669-1674. http://dx.doi.org/10.1145/2702613.2732838
26. David Kirk, Tom Rodden, and Danaë Stanton Fraser. 2007. Turn it this way. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*, ACM Press, 1039. http://doi.org/10.1145/1240624.1240782
27. David Kirk and Danae Stanton Fraser. 2006. Comparing remote gesture technologies for supporting collaborative physical tasks. *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06*, ACM Press, 1191. http://doi.org/10.1145/1124772.1124951
28. Hideaki Kuzuoka, Shinya Oyama, Keiichi Yamazaki, Kenji Suzuki, and Mamoru Mitsuishi. 2000. GestureMan. *Proceedings of the 2000 ACM conference on Computer supported cooperative work - CSCW '00*, ACM Press, 155–162. http://doi.org/10.1145/358916.358986
29. Joel Lanir, Ran Stone, Benjamin Cohen, and Pavel Gurevich. 2013. Ownership and control of point of view in remote assistance. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13). ACM, New York, NY, USA, 2243-2252. http://doi.acm.org/10.1145/2470654.2481309
30. Doo Young Lee and Mark R. Lehto. 2013. User acceptance of youtube for procedural learning: an extension of the technology acceptance model. *Computers & Education 61*: 193-208.
31. Paul Luff, Christian Heath, Hideaki Kuzuoka, Jon Hindmarsh, Keiichi Yamazaki, and Shinya Oyama. 2003. Fractured Ecologies: Creating Environments for Collaboration. *Human-Computer Interaction* 18, 1: 51–84. http://doi.org/10.1207/S15327051HCI1812_3
32. Paul K. Luff, Naomi Yamashita, Hideaki Kuzuoka, and Christian Heath. 2015. Flexible Ecologies And Incongruent Locations. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, ACM Press, 877–886. http://doi.org/10.1145/2702123.2702286
33. James Norris, Holger M. Schnädelbach, and Paul K. Luff. 2013. Putting things in focus: establishing co-orientation through video in context. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13). ACM, New York, NY, USA, 1329-1338. http://doi.acm.org/10.1145/2470654.2466174
34. Jiazhi Ou, Xilin Chen, Susan R. Fussell, and Jie Yang. 2003. DOVE. *Proceedings of the eleventh ACM international conference on Multimedia - MULTIMEDIA '03*, ACM Press, 100. http://doi.org/10.1145/957013.957034
35. Abhishek Ranjan, Jeremy P. Birnholtz, and Ravin Balakrishnan. 2007. Dynamic shared visual spaces: experimenting with automatic camera control in a remote repair task. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07). ACM, New York, NY, USA, 1177-1186. http://doi.acm.org/10.1145/1240624.1240802
36. Kim Seungwon, Gun A. Lee, and Nobuchika Sakata. 2013. Comparing pointing and drawing for remote collaboration. *In 2013 IEEE International Symposium on Mixed and Augmented Reality* (ISMAR), pp.1-6, Oct. 2013. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6671833&isnumber=6671745

37. Kim Seungwon, Gun A. Lee, Nobuchika Sakata, Andreas Dunser, Elina Vartiainen, and Mark Billinghurst. 2013. Study of Augmented Gesture Communication Cues and View Sharing in Remote Collaboration. *In IEEE International Symposium on Mixed and Augmented Reality* (ISMAR), pp.261-262, Oct. 2013. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6671795&isnumber=6671745
38. Rajinder S. Sodhi, Brett R. Jones, David Forsyth, Brian P. Bailey, and Giuliano Maciocci. 2013. BeThere: 3D mobile collaboration with spatial input. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13). ACM, New York, NY, USA, 179-188. http://doi.acm.org/10.1145/2470654.2470679
39. Matthew Tait and Mark Billinghurst. 2015. The Effect of View Independence in a Collaborative AR System. *Computer Supported Cooperative Work (CSCW)*. http://doi.org/10.1007/s10606-015-9231-8
40. John C. Tang and Scott L. Minneman. 1991. Videodraw: a video interface for collaborative drawing. *ACM Transactions on Information Systems* 9, 2: 170–184. http://doi.org/10.1145/123078.128729
41. Anthony Tang, Carman Newstaedter and Saul Greenberg. 2006. VideoArms: Embodiments for Mixed Presence Groupware. *In Proceedings of the 20th British HCI Group Annual Conference (HCI 2006)*. London, UK, 85-102.
42. Xianjun Sam Zheng, Cedric Foucault, Patrik Matos da Silva, Siddharth Dasari, Tao Yang, and Stuart Goose. 2015. Eye-Wearable Technology for Machine Maintenance: Effects of Display Position and Hands-free Operation. *In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15). ACM, New York, NY, USA, 2125-2134. http://doi.acm.org/10.1145/2702123.2702305