UNIVERSITY OF CALGARY

Stabilized Annotations for Mobile Remote Assistance

by

Omid Fakourfar

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE DEGREE IN COMPUTER SCIENCE

CALGARY, ALBERTA

DECEMBER, 2016

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, this thesis investigates an annotation stabilization technique, which allows annotations to stick to their intended location. I studied two different forms of annotation systems, with both tablets and head-mounted displays. To differentiate my work from the prior research, I considered a number of task factors that might influence system performance in remote assistance scenarios. My analysis suggests that stabilized annotations and head-mounted displays are only beneficial in certain situations. I conclude with reflections on system limitations and potential future work.

Publications

Some figures and material in this thesis have previously appeared in this prior work:

Fakourfar, O., Ta, K., Tang, R., Bateman, S., & Tang, A. (2016). Stabilized Annotations for Mobile Remote Assistance. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16 (pp. 1548–1560). New York, New York, USA: ACM Press.

Acknowledgements

The wonderful journey of my Masters' degree would not have been possible without the help and support I received from my family, colleagues and friends. Here I would like to thank a handful of those awesome people who have helped me academically and mentally.

First and foremost, I would like to thank my parents for all the support throughout my studies; from my first day at school up until now that I have been studying miles away from them.

I could never imagine an academic supervisor as passionate, dedicated and supportive as Dr. Anthony Tang. He familiarized me with my field of research and his support along with his constant feedback paved the way for our research idea to evolve into a CHI paper and my masters' thesis. Thanks Tony!

I also want to thank my CHI paper co-author Dr. Scott Bateman who helped me a lot in writing the paper as well as polishing my conference presentation.

Finally, I want to thank my peers in the Interactions Lab and RICELab. iLab has a very unique atmosphere with some of the brightest talents in HCI research who in an extremely friendly environment, support and help each other. It is not possible to list all the fantastic folks who have helped me along the way, but I would like to thank: Kevin Ta, Richard Tang, Brennan Jones, Hesam Alizadeh, Anna Witcraft, David Ledo, Teddy Seyed, Kody Dillman, Jennifer Payne, Fateme Rajabiyazdi, Bon Adriel Aseniero, Terrance Mok, Paul Lapides, Sowmya Samanath, Gabriele Kuzabavicute, Charles Perin, Jo Vermuelen, Lindsay MacDonald and our fantastic faculty members: Sheelagh Carpendale, Ehud Sharlin, Lora Oehlberg, Wesley Willett and Sonny Chan.

Table of Contents

List of Tables

List of Figures and Illustrations

Chapter One: Introduction

Mobile video chat is defined as the use of mobile devices such as smartphones, tablets and head-mounted displays to connect two (or more) remote users via video (and/or audio). (Mobile) video chat is becoming increasingly widespread across a number of scenarios, including but not limited to making video calls to communicate with remote loved ones (Judge & Neustaedter, 2010), giving tours of new places (Jones et. al., 2015), playing multiplayer games (Henrysson et. al., 2005) and remote assistance (Domova et. al., 2014, Fussell et. al., 2004, Gurevich et. al., 2012). Some scenarios where we would expect mobile video chat to be used include:

Figure 1.1. A sample task for remote assistance. Left: The owner uses the manual to fix the car. Middle: The owner uses a tablet to communicate with a remote helper. Right: The owner uses a head-mounted display to communicate with the helper.

- Using smartphones, tablets, and head-mounted displays for ad hoc, unplanned remote assistance (e.g., Bauer et. al., 1999, Gauglitz et. al., 2012, Johnson et. al., 2015). For example, Brubaker et. al. (2012) explored how 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.

- Fun, interactive activities or providing sightseeing tours. For example Procyk et. al. (2014) used mobile video chat in a geo-caching scenario, where two remote collaborators communicate to find hidden items in an environment.

An example of a remote assistance scenario could be a car repair/maintenance task. While vehicle owners might be able to fix their car while making use of the owner's manual, they can run into more complicated problems where they needs feedback or clarification from an expert. This can be addressed by setting up a mobile video chat session in which a remote expert can reach into the scene and help the car owner (Figure 1.1).

A growing body of design work is focused on building mobile systems to address such scenarios (e.g., Bauer et. al., 1999, Gauglitz et. al., 2012, Johnson et al., 2015). One of the classic

challenges in supporting remote collaboration is the need to reference objects or locations

efficiently through video (Buxton 2009, Fussel et. al., 2004, Gergle et. al., 2004, Kirk et. al.,

2006 & 2007, Ou et. al., 2003, Tang & Minneman, 1991). 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 (Gauglitz et. al., 2012, Johnson et. al., 2015, Jones et. al.,

2015); and, difficulties in manipulating and positioning cameras and devices while conducting

physical tasks (Gauglitz et. al., 2012, Jones et. al., 2015). 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 (Gergle et. al., 2004, Jones et. al., 2015).

Many systems address this problem by providing an annotation subsystem, allowing a remote

helper to annotate a scene (e.g. draw, write notes, etc.) as it is captured by a worker's camera.

The problem is that it is unclear how to design such an annotation subsystem for mobile video

chat mainly because prior studies have failed to identify specific task factors in which each type

of annotation subsystems might be useful.

In this thesis, I focus on the design and evaluation of such an annotation subsystem. I approached

this problem by developing a prototype system, exploring several design variations within this

prototype based on the history of research in this area: stabilized annotations (annotations that

stay affixed to the objects they were drawn upon – Figure 1.4), and annotations atop live video

(Figure 1.5). While considerable work has explored the space of annotations (e.g., Gauglitz et.

al., 2014 and Kim et. al., 2015), this work is distinct in that I also allow for freehand annotations

on predefined moveable annotation planes in the 3D environment. I also explore two device

variations: a handheld tablet and a head-mounted display and camera.

To evaluate my designs, I conducted a laboratory study where I recruited 16 pairs of participants to try different combinations of devices and annotation systems across 4 remote assistance tasks. The tasks were designed to mimic real-world remote assistance scenarios.

To foreshadow the results of this study, stabilized annotations provided only marginal benefit to teams over simple annotations atop live video (Gauglitz et. al., 2014). This is a consequence of how annotations were used to support interaction—rather than being used for information that needed to be revisited, annotations were generally used in an ephemeral fashion—i.e. in the moment. I 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 be carried out (Johnson et. al., 2015). Nevertheless, participants generally preferred stabilized annotations with head-mounted displays over other conditions – tablets and non-stabilized annotations.

1.1. Research Context

I situate my work at the intersection of Computer-Supported Cooperative Work (CSCW), which focuses on facilitating the way people collaboratively work in groups using technologies such as mobile video chat, and Augmented Reality (AR), which enables the users to have digital objects, drawings, etc. added on top of live video feeds. I approach this design problem from the perspective of a researcher in Human-Computer Interaction. As a Human-Computer Interaction (HCI) researcher, I am interested in variations on design, which allows me to understand trade-offs between different approaches to the same problem. Thus, I explore several variations on the technical implementation of the system, as well as form-factors (head-mounted display vs. tablet). Furthermore, the CSCW and HCI perspectives inform the design of my evaluation techniques, where I have tried to evaluate the efficiency of different design variations by observing the user behaviors while working with the systems through user studies. Although I have collected and reported some quantitative results from my studies, the main approach has been an observational and qualitative study of the user interactions with my systems.
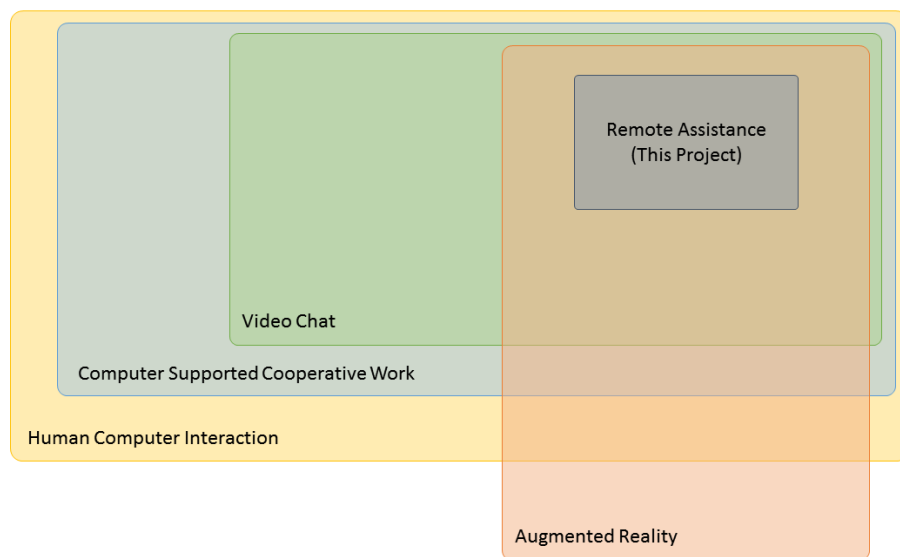
Figure 1.2. Thesis context.

1.2. Thesis Problem and Thesis Questions

The research problem this thesis addresses is:

How should we design effective annotation systems for mobile video chat?

In this thesis, I explore this problem from two different perspectives: first, by varying the physical form factor of the viewing system for annotations (manipulating the overall mobility of the system) and second by varying whether the annotations stay within the visual context of where they were drawn (manipulating the visual "situatedness" of annotations).

Thus, my thesis addresses the following thesis questions:

Thesis Question 1: What design aspects of a remote annotation system can best improve the interaction in a remote assistance scenario?

Although a considerable body of work has explored the use of annotation systems to support remote assistance, we still do not have a comprehensive understanding of how different features of remote assistance help the interaction. For instance, prior works have explored different types of annotation systems (stabilized, non-stabilized, temporally stabilized) and different device configurations (handheld, head-mounted displays, etc.). However, each study was run differently, on different tasks with inconsistent task factors. This means systems perform differently across tasks with varying factors. We need to have a systematic understanding of the relationship between different studies to understand what features are likely to produce the greatest benefit.

Thesis Question 2: How does changing different aspects of camera mobility affect interaction?
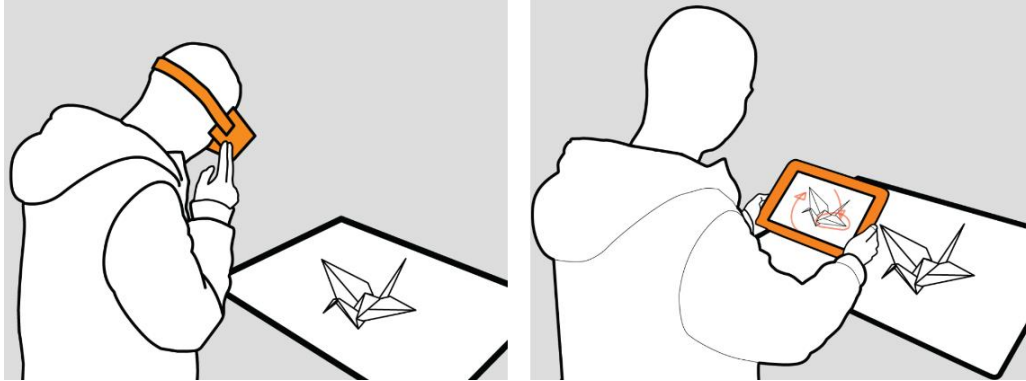
Figure 1.3. Left: A worker with a head-worn display, holding it with hands to make it steady vs. Right: with a hand-held device.

Others have explored how mobile devices such as smartphones and tablets are used to support remote assistance scenarios (Jones et. al., 2015). Yet, this presents several important challenges: the video may be shaky if it is being held by the user; the user is required to work with one hand (or forced to put the camera down), and finally, if a user does choose to put the camera/device down, this can cause the helper lose the context of what the worker is doing in the moment, which could lead to bigger breakdowns in action.

In this thesis, I explore how a wearable head-mounted display prototype that captures the environment via a camera compares to a typical handheld smartphone. In my prototype, the worker receives helper's annotations within the head-worn display, meaning the worker's hands are freed to do work.

Thesis Question 3: How does the "stickiness" of free-hand annotations (to the objects they describe) affect their utility within a remote assistance scenario?

With a typical free-hand annotation-atop-video approach, the problem is that as the video scene moves (e.g. if the camera moves), the annotations are no longer properly attached to the objects in the scene they are referring to. To address this problem, several researchers (e.g., Gauglitz et. al., 2012 & 2014) have designed annotation systems that allow helpers to pin markers onto parts of a 3D captured scene, meaning annotations can be properly attached to the objects, regardless of how the camera is subsequently positioned. Within the context of this thesis, I describe this latter style of annotation as "stabilized" (Figure 1.4), and the former, "non-stabilized" (Figure 1.5) annotations. Yet, this leads to two questions—how can we design a system that supports free-hand annotations in a tracked scene, and second, how does this support remote assistance?

1.3. Thesis Objectives

Although prior authors have developed similar systems in the past, previous evaluations have not made clear the benefits of stabilized annotations (if any) or head-mounted display and camera. This has been mainly because they have mostly relied on some common tasks (mainly construction tasks) with limited task factors. Thus, we still do not have a good understanding

Figure 1.4. Stabilized annotations, attached to the objects of the scene in spite of the camera movement. Left: annotations are visible through the camera. Right: worker has moved the phone to right. Annotations are still anchored to the objects despite not being visible in the camera.

8

whether or not such features are successful for improving collaborative support scenarios across different tasks. Thus, I outline the following objectives:

Thesis Objective 1. Conduct a meta-analysis of prior work involving annotation systems for remote assistance. To address Thesis Question 1, I will develop a meta-analysis of prior work where researchers have studied the use of remote assistance systems to understand which aspects of remote annotation systems are worth exploring for the purpose of design.

Thesis Objective 2. Design a system that enables stabilized free-hand annotations for remote assistance. Whereas prior work has allowed for "pointing" in a scene such that the annotations are properly stabilized, no prior work, to my knowledge, has allowed for free-hand annotations. Building this addresses part of Thesis Question 3.

Thesis Objective 3. Design and conduct a study that evaluates the use of an annotation system that contrasts the display/capture tool, and the style of annotations. I designed a study to explore the trade-offs between variations of annotation systems over a range of four different collaborative support tasks. In the study, I examine the use of stabilized freehand annotations and freehand annotations atop video, contrasting their use with tablet devices and head-mounted video see-through displays. The purpose of running this study is to observe how users communicate with one another using each of the four different configurations to identify when

Figure 1.5. Non-stabilized annotations, drawn on top of the video feed, moving with the camera movement.

9

each combination of annotation/device is likely to be useful. Results of this study address Thesis Questions 2 and 3.

1.4. Thesis Contributions

This thesis provides the following contributions:

Thesis Contribution 1. This thesis reviews and analyzes previous studies exploring annotations for remote support, where the meta-analysis provides guidance for designers of future systems for remote support.

Thesis Contribution 2. This thesis provides a first study of the idea of annotation planes as a technical contribution that allows freely drawn 2D annotations to be meaningfully made in 3D space.

Thesis Contribution 3. This thesis outlines a framework of annotation use that provides an understanding of when and why stabilized annotations are likely to be valuable.

Thesis Contribution 4. This thesis presents a study that provides new findings suggesting both task and design factors affecting the utility of stabilized annotations.

1.5. Roadmap

This thesis is structured as follows:

Chapter 2 provides background on remote support and collaboration systems and reviews prior work on different methods of providing assistance.

Chapter 3 discusses my system design process. In particular, I look at the remote assistance task requirements and shortcomings of the previous work, and then explain how I tried to address

those issues in my system design so that it would cover the widest possible range of scenarios. It also reviews the implementation of the tool used in my study.

Chapter 4 describes the laboratory study I performed on the prototype and analyzes data collected from participants and qualitative findings. This chapter particularly addresses the thesis question by providing an in-depth analysis of user behaviors while using different configurations of the system.

Chapter 5 provides a discussion of the implications of the findings of my study, and concludes the thesis. I will discuss my overall contributions and implications of my work for remote assistance task and future work.

Chapter Two: Related Work

Today, many people turn to online resources such as YouTube videos[1] (Figure 2.1) to receive

instructions for completing everyday tasks like fixing home utilities, cooking, etc. (e.g. Lee &

Lehto, 2013). While such videos are useful for simple tasks like cooking, cleaning, etc., 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. Examples of such tasks might be complex engineering tasks or non-trivial tasks which

require some level of expertise like repairing an engine or operating a complicated control board

in a power plant. To address these more complex scenarios, researchers have built systems

designed for mobile remote support. Specifically, these are systems where an expert can

---

[1] Example: How to tie a Windsor tie: https://youtu.be/KN_XhE2ma_E

Figure 2.1. A YouTube video demonstrating how to tie a windsor tie. Available at: https://youtu.be/KN_XhE2ma_E

remotely instruct his/her partner in the workspace. The worker operates a mobile device to show the current status of his/her surrounding to the expert (often through video chat).

To set the stage for my work, I describe several prior systems designed for mobile remote support. I then describe the role of deixis, and how this is addressed in many systems using an annotation subsystem. In the context of this work, I define deixis as words, phrases or pointers by the remote expert, such as "put 'this' thing 'there'", that cannot be fully understood unless more contextual information is provided to identify "this" and "there". This contextual information is often provided using an annotation subsystem for mobile remote support system. In an annotation subsystem, the expert is able to annotate the video stream for the worker using free-hand drawings, push-pins, etc. to refer to specific objects and/or locations in the workspace.

I then outline studies evaluating these systems, highlighting findings, and identifying gaps in the literature. Finally, I summarize by discussing the role of limitations in experimental task design as one potential reason for the relatively mixed, and sometimes conflicting, results of these studies. In this chapter, I have particularly addressed the following:

Figure 2.2. ClearBoard from Ishii et. al (1993), a shared workspace allowing remote coworkers to draw while maintaining eye contact and using natural gestures.

1. *Thesis Question 1*, by developing a meta-analysis of prior work where researchers have studied the use of remote assistance systems to understand which aspects of remote annotation systems are worth exploring for the purpose of design.

2. *Thesis Contribution 1*, by reviewing and analyzing previous studies exploring annotations for remote support.

2.1. Static Systems for Remote Support and Collaboration

Supporting remote assistance and collaboration through video has long been an interest of CSCW research. Early work by Ishii et. al. (1993) and Tang & Minneman (1991) generally focused on fixed-perspective video to support remote collaborative work, where the focus was on connecting remote collaborators with one another (Figure 2.2).

In terms of how the camera is set up in the workspace, prior work can be divided into the following categories:
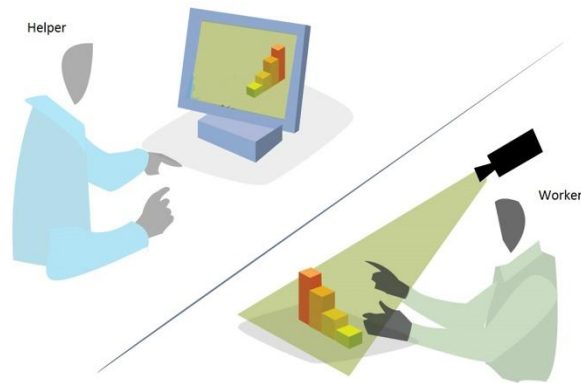
Figure 2.3. Helper and worker in a single-camera system similar to Fussell et al (2003). The helper uses a PC to view a video feed coming from the camera above the worker's table.

2.1.1: Single Camera Systems

These 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. As an example, Fussell et. al. (2003), introduced a fixed, scene-oriented camera providing a view of the work environment and assessed its value. In this evaluation, the scene-oriented camera proved to be useful; though it could not provide detailed information about the active work area because of its fixed perspective (Figure 2.3).

2.1.2. Multi Camera Systems

Newer designs by Fussell et. al. (2003) and Ranjan et. al. (2007) explored multiple views of the workspace. These designs typically provide overview+detail view of a remote workspace, allowing remote workers to see both fine details of work and a contextual overview of the remote space (Figure 2.4).

Fussell et. al. (2003) introduced a condition in which outputs from both a scene camera and a head-mounted camera were displayed on the helper's screen. The scene camera is meant to
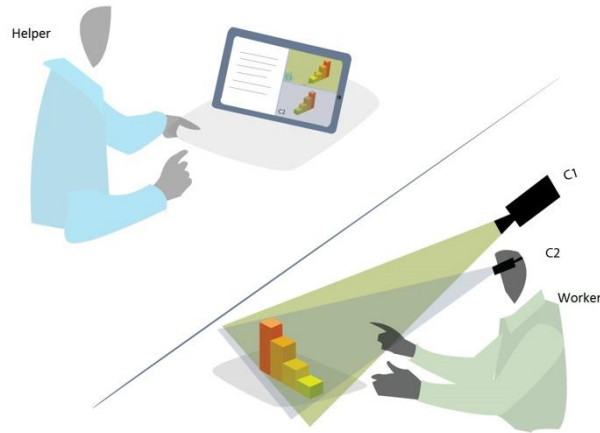
Figure 2.4. Helper and worker in a multi-camera system. The worker is wearing a head-mounted camera (C2). Both the static camera (C1) and C2 streams are shown to the helper. The helper also has access to the task manual on the left, similar to Fussell et al (2003).

provide the overview shot while the head-mounted camera could capture detailed views of the workspace (Figure 2.4). However, this configuration did not improve participants' effectiveness compared to the single scene camera condition. Ranjan et. al. (2007) integrated multiple automatic camera views that were guided partly by tracking worker's hands. The results of their study suggested that the automatic system provided performance benefits over the single static camera.

2.1.3. Movable Camera Systems

Some designs by Gurevich et. al. (2012) and Lanir et. al. (2013) also introduced cameras that could be repositioned.

Gurevich et. al. (2012) designed a system to support remote assistance named "TeleAdvisor" which consisted of a video camera and a projector mounted at the end of a tele-operated robotic arm. The worker can position the camera in their environment, directing the camera and projector to the point of need and carry on a voice conversation. The helper could also move the arm in 2 degrees of freedom to get a more detailed view, if needed. Helper's annotations were

then projected on to the workspace. This

prototype was evaluated through Lego

assembly and TV/DVD player assembly tasks.

They observed that TeleAdvisor was used

effectively to complete the tasks using

annotations and deictic references (Figure 2.5).
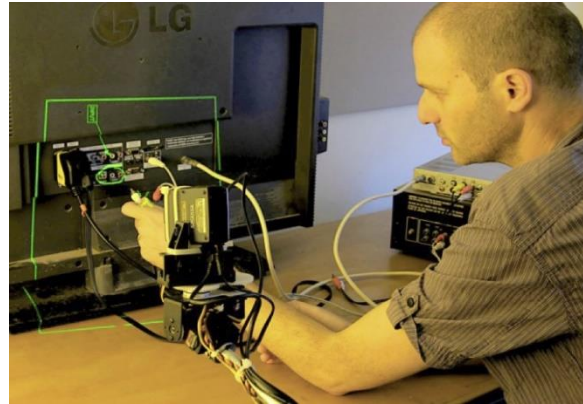
Figure 2.5. Gurevich et. al. (2012) prototype with a camera on a robotic arm.

2.2. Mobile Remote Support Systems

The prototypes described in section 2.1 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 by Domova

et. al. (2014), Gauglitz et. al. (2014) and Sodhi et. al. (2013) has explored how remote assistance

can be augmented through mobile technologies. 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 different device configurations:

2.2.1. Mobile Phones

Gauglitz et. al. (2014) introduced an annotation system for remote assistance in physical tasks

which is possible to be used either by a tablet or a smartphone. However, they only explored

hand-held devices in remote assistance tasks and did not contrast it with head-mounted displays.

In addition, they did not consider the various task factors in such scenarios and focused solely on

a car repair task. Also, Jones et. al. (2015) explored the mechanics of camera work in mobile

video chat by running a series of studies on users working with smartphones as the mobile

device. They stated the limited field of view and the lack of camera control as the main causes of

frustration in mobile video chat tasks (i.e., campus tours, shopping together, detail search and

17

collaborative physical task). They also observed the use of hand gestures to make deictic references, which suggests the use of annotation subsystems.

2.2.2. Tablets

Several prior works made use of tablets as the mobile device for the worker person. Gauglitz et. al. (2012) designed an interface to put augmented-reality annotations in shape of push-pins atop live video feed with a tablet as the device for local user. Domova et. al. (2014) also designed a remote video collaboration system for industrial settings where the worker could capture and stream video to a desktop application of the remote helper. The system also implemented synchronized snapshots and annotations between the two parties. However, it did not include a built-in voice communication mechanism which posed serious problems for participants. In both papers, they ran the study on a single task which does not resemble all the possible remote support tasks. In addition, both the smartphone and the tablet conditions can cause the problem of having to holding a device and framing the scene; thus disabling the user from working with both hands.

2.2.3. Head-mounted Displays and Cameras

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 (Bauer et. al., 1999)—freeing the worker's hands to be used in the main collaborative task, and allowing the focus of work to be easily captured.

Huang et. al. (2013) presented another system named "HandsInAir" to support mobility of both the worker and the helper. In this system, both the helper and the worker use the same set of hardware consisting of a helmet mounted with a camera on top of, and a near-eye display connected to a wearable PC (Figure 2.6). Although they got positive feedback from the participants trying their system, they had not run a comparative study, they used a very low-fidelity prototype and did not integrate any annotation subsystem.

2.3. 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 (e.g. Gutwin et. al., 1996; Jones et. al., 2015). Deixis supports common ground, and reduces the number of speech acts needed to complete tasks (e.g., Fussell et. al,. 2004). According to Fussell et. al. (2003), one major drawback of audio-only systems (such as telephone calls) is the lack of visual information. As a result, helpers need to be far more explicit in describing object characteristics and current state of the task. While it is possible to provide instructions over phone, it is extremely hard to achieve a common ground and make deictic references with an audio-only setup. Remote support systems using video provide support for deixis in a variety of
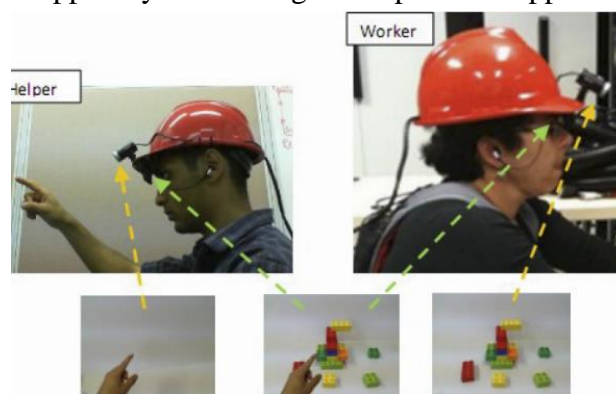
Figure 2.6. Head-mounted display and camera
by Huang et. al. (2013).

ways, including telepointers (Fussell et. al., 2004), push-pins (Figure 2.7 - Gauglitz et. al., 2012 and Seungwon et. al., 2013), freely drawn annotations (Figure 2.8 - Gauglitz et. al., 2014, Ishii et. al., 1993 and Tang & Minneman, 1991), and representations of collaborators' arms (Izadi et. al., 2007 and Tang et. al., 2006) (Figure 2.9).
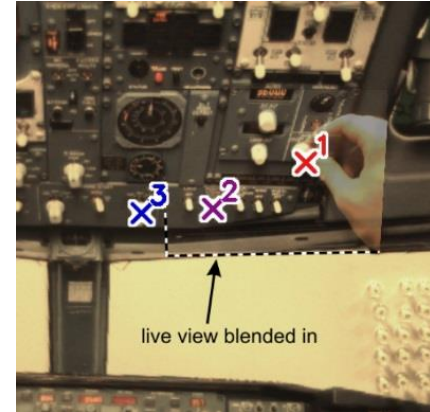
Figure 2.7. Push-pin annotations by Gauglitz et. al. (2012).

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 (Kim et. al., 2015). Similarly, if the camera position moves, annotations no longer point to the right object or location (Gauglitz et. al., 2012 & 2014, and Kim et. al., 2015). Recent systems have addressed this problem in two ways:

1. *Freezing the video while annotations are drawn:* This simple approach ensures that the annotations remain in place with the objects in the scene (Gauglitz et al., 2014 and Kim et al., 2015). As a result, the helper can concentrate on drawing the annotations without being distracted by camera movements.
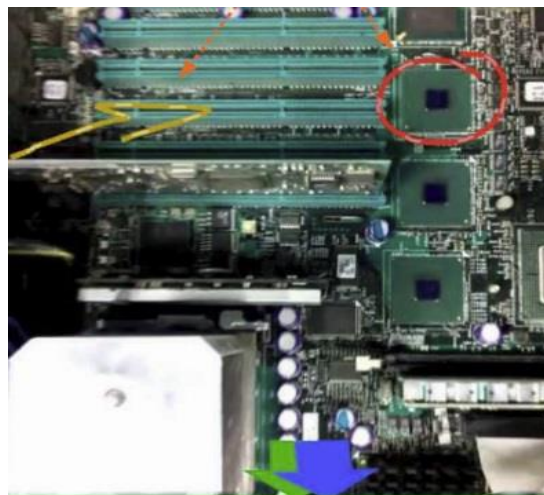
Figure 2.8. Freely drawn annotations by Gauglitz et. al. (2014).

20

2. *Anchoring the annotations to elements in the scene:* This approach, referred as stabilized annotations (Gauglitz et. al., 2014) has been explore by using fixed points of view (Izadi et. al., 2007), by tracking the point of view of the camera so that the annotations can be correctly positioned (Gauglitz et. al., 2014), and by dynamically modeling a remote environment (Gauglitz et. al., 2014).

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

2.4. System Studies and Task Breadth

Researchers 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, I review study findings, organized by the benefits, expected by a naïve designer, of particular technologies used for remote support, and the creation and use of annotations. I synthesize this work making a case for the use of a broader set of tasks when assessing new remote support systems.

2.4.1. Summary of Studies

The majority of previous studies use "collaborative physical tasks" (e.g., Fussell et. al., 2004, Johnson et. al., 2015 and Kirk et. al., 2007). In general, such tasks have a "worker" perform physical tasks, such as building
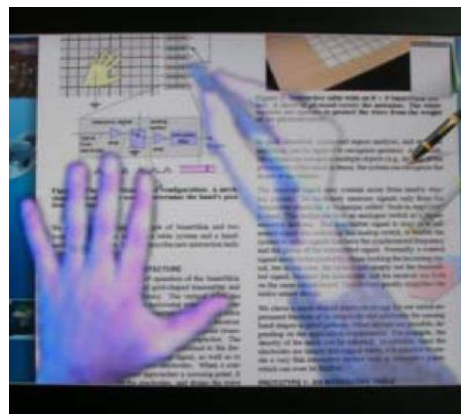
Figure 2.9. Representations of collaborators' arms and workspaces by Izadi et. al. (2007)

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 (Figure 2.4).

Assumption: Stabilized annotations are better than non-stabilized annotations

Because the camera is mobile, several authors have attempted to anchor annotations to elements

in the scene (often called stabilized annotations – Figure 2.10). Several studies by Bauer et. al.

(1999), Domova et. al. (2014), Gauglitz et. al, (2012 & 2014) and Kim et. al. (2015) have

examined the simple approach of using video pausing, and still frames to stabilize the scene and

objects for annotation. In some implementations, like Bauer et. al. (1999) and Domova et. al.

(2014) both parties have control over when a still image is shown instead of live video, while

others like Gauglitz et. al. (2012 & 2014) and Kim et. al. (2015) limit the control to one party or

the other. A final variation is to automatically freeze the scene while annotations are being made

(Kim et. al., 2015).

Figure 2.10. Stabilized annotations by Gauglitz et. al. (2014). In spite of the camera movements, the red annotation is still anchored to the object it was drawn upon.

Annotations on a still frame do not seem to provide meaningful benefit in terms of task time (Kim et. al., 2015). Bauer et. al. report more variance in how much the feature was used (some participants used it frequently; other participants rarely) (Bauer et. al., 1999). Of note, is that people generally seem to prefer automatic freezing compared to manual freezing (Kim et. al., 2015).

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., Gauglitz et. al., 2012 & 2014 and Kasahara et. al., 2014). Yet again, surprisingly, two studies comparing the use of stabilized vs. non-stabilized annotations (Gauglitz et. al., 2012 & 2014) 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 are beneficial, 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. (2003) 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. (2015) 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 by Bauer et. al. (1999), Huang et. al. (2013) and Kasahara et. al. (2014) have used head-mounted displays. 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 (Gauglitz et. al., 2014 and Tait & Billinghurst., 2015). 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.

Researchers 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. See-through transparent displays show information on top of the real world (e.g., Kasahara et. al. (2014), Epson Moverio). Finally, information has also been displayed via a small peripheral screen (e.g., Bauer et. al., 1999, Google Glass).

In this review I did not find any work comparing head-mounted displays to handheld devices for remote collaboration. A peripheral work by Zheng et. al. (2015) 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. To address this gap, I have set up my

user study in a way so that I would be able to observe and compare user behaviors while using a head-mounted display and a hand-held device. Furthermore, I have determined the advantages and disadvantages of each device configuration for remote assistance tasks.

2.5. Conclusions

In this chapter, I provided an overview of literature for remote assistance in physical tasks to particularly address Thesis Question 1 and Thesis Contribution 1. I started with the quick and easy approach of online video instructions and explained why they are not of interest for more complex scenarios. Then, I provided an overview of video chat systems with single or multiple views of the workspace to support collaborative task. This has led to the use of mobile video chat to support ad hoc, mobile scenarios.

Then I explained how and why supporting deixis is one of the main challenges in video chat systems, and how this issue is being addressed using different types of annotations. Particularly, my focus is on freeze-frame annotations and stabilized annotations as the two most common approaches to support deixis in mobile assistance systems. I then concluded this chapter by summarizing the previous user studies based on a few common assumptions.

In conclusion, although there has been extensive work studying different annotation systems for remote assistance, these studies have provided some mixed results and more importantly, have not provided the conditions and/or task factors in which each type of annotation system might prove to be useful. I address this issue in the next two chapters.

Chapter Three: System Design and Implementation

As described in previous chapter, system support for deixis through annotation systems is one of the main challenges that the researchers have faced in designing effective remote assistance systems. To address this issue, work has been done to design and evaluate different annotation and/or device configurations. However, remote assistance for physical tasks is a very general term and previous studies have failed to produce consistent results or to provide task factors under which each of the device and/or annotation configurations would be useful. This leads to two general problems in the research literature:

- First, we need a method of identifying and articulating the specific hardware/software factors that make different kinds of annotation tools/configurations useful.

- Second, we need to identify salient features of tasks that are representative of real-world tasks, and understand how/why an annotation subsystem would help.

In my work, I have tried to address these two problems by designing and implementing remote assistance systems and then evaluating them under different task factors to figure out which types of systems are useful for each task scenario. For each of the system configurations evaluated, I integrate an annotation subsystem into a video chat session, so the helper would be able to "point" towards objects in the workspace, show "processes", etc. I focus on four main configurations of interest (head-mounted display vs. handheld devices and stabilized vs. non-stabilized annotations). The goal is to explore user behaviors while using remote assistance systems under different task conditions. Thus, I implemented applications to allow helpers to draw on top of the video feed received from the worker.

Mobile remote assistance systems can be used in a wide variety of scenarios and for a variety of tasks. There is a broad range of possible tasks that their requirements might influence the system design. People might use remote assistance systems to give tours of new places (Jones et. al., 2014), playing games remotely (Procyk et. al., 2014) or to provide assistance in physical tasks such as home theater setup (Gurevich et. al., 2012), water treatment plant maintenance (Domova etl. al., 2014), robot construction (Fussell et. al., 2003) and vehicle repair (Gauglitz et. al., 2014) remotely. In this thesis, my focus is on designing systems to enable a remote expert assist a worker in the workplace during a physical task. Even here, the range of possible tasks varies from simple, everyday tasks - like cooking or assembling furniture (Rae et. al., 2014) -, to very sensitive tasks like repairing and maintaining a jet engine, or operating the control board in a power plant. Each of these tasks have their own requirements: sometimes the task can be performed by one hand or needs both hands, party with task specific knowledge can be on one or both sides, the task can have a 2D or 3D setup and different complexities, the instructions can be required only for a short amount of time or they need to be referred back later on, etc. I articulate

27

the salient features of this task space, selecting some of these features then creating a set of representative abstract tasks for studying how well an annotation subsystem supports the execution of the tasks. Before that, I decide on the configurations I want to be tested, the main focus of this chapter.

In particular, in this chapter I will describe my system design process; different devices and annotation systems that I evaluated by running a study. Then I will move forward and explain how we integrated two annotation subsystems, one with augmented-reality into a simple mobile video chat session.

Specifically, in this chapter I have addressed the following:

- Thesis Question 2 (how does changing different aspects of camera mobility affect interaction?), by designing a remote assistance system working under two different mobility configurations, head-mounted displays and tablets.
- Thesis Question 3 (how does "stickiness" of stabilized annotations affect their utility?) and Thesis Objective 2 (designing a system to enable stabilized free-hand annotations for remote assistance), by designing a remote assistance system working with two different annotation systems, stabilized and non-stabilized, to be evaluated by a user study.

3.1. System Design

As described earlier, I was interested in studying how different devices and annotation systems perform in remote assistance scenarios. In such scenarios, there would normally be two roles (Figure 3.1):

- Helper: This person has all or part of the knowledge needed to complete the task. They will use a device (e.g., a tablet) to: 1) receive video stream from the worker's workplace, 2) provide support (via audio or annotations) to the worker.

- Worker: This person is the one in need of help. They will be located in the workplace, working physically on the task, and communicating with the helper via a head-mounted display or a tablet.

The first step is choosing the configurations that I wanted to be tested; I chose two device settings with the following points in mind, based on a long line of research in this area:

- *Head-mounted camera and display*: There are multiple features that make HMDs unique. They can free the users' hands from the burden of holding a handheld device to be able to communicate their environment to the remote helper. As a result, the user is able to work on the tasks with both hands. Also, the head-mounted camera that comes with most HMDs aligns the camera view with the user's eye view. So the remote helper can see exactly what the worker is seeing in the workspace. Please note that normally head-mounted cameras are a *part* of head-mounted displays. However, there could be head-mounted cameras that do not have any displays attached to receive and show video, instead requiring an external display to show helper's annotations. Such cameras are not
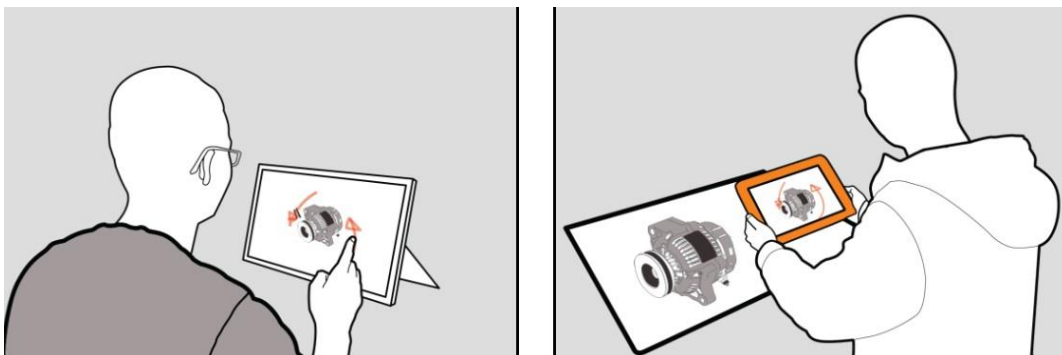
Figure 3.1. Left: Helper person working with a tablet, Right: Worker person working with either a tablet (can be substituted with a HMD).
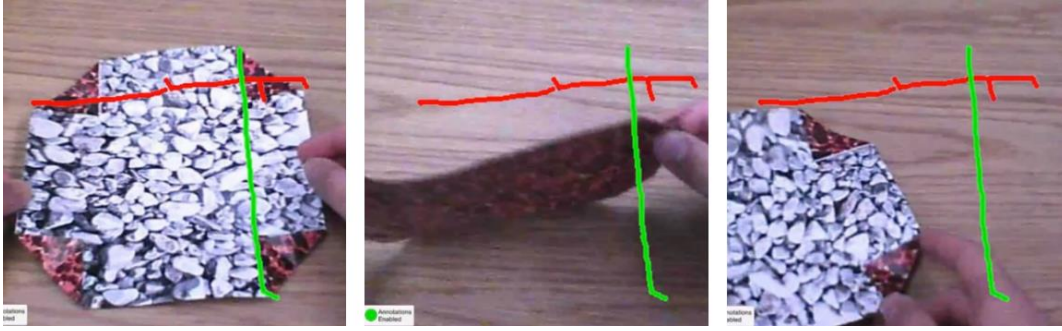
Figure 3.2. Non-stabilized annotations. Note that the annotations are not anchored to any object in the scene.

the focus of this work. In this thesis, I only focus on head-mounted devices that have both a camera and a display built in, and I will refer to them as HMDs.

- *Tablet*: Tablets and other handheld devices (e.g., smartphones) are generally cheaper and already available to the public. As a result, our system might be quicker and easier to be accessed through these devices. They also provide more mobility compared to head-mounted display, because the user can position the device in different angles. However, the user needs to hold the tablet with at least one hand to keep communication running. So they are not able to work on the tasks with both hands, something that might be required for some scenarios.

I also chose two annotation variations to be integrated into my mobile video chat systems:

- *Non-stabilized*: Non-stabilized annotations are a quick and easy way of integrating annotations in the video feed. While this type of annotation enables the helper to draw on top of the video feed, the drawings are not anchored to any objects in the scene and would move with the camera (Figure 3.2). These types of annotations are very easy to implement and can be integrated into common video chat platforms (e.g., Skype, Google Hangouts, etc.).
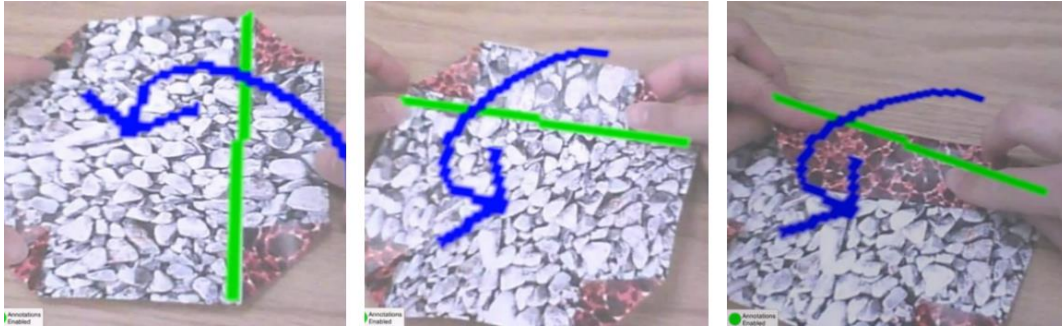
Figure 3.3. Stabilized annotations. Note that the annotations are anchored to
the object in the scene, even if it is moved.

- *Stabilized*: I also designed stabilized annotations which will be anchored to the objects

  they were drawn upon. This is accomplished by defining unique image targets in the

  scene and annotation planes which can be drawn on. Annotation planes are flat, semi-

  transparent virtual surfaces visible only through the systems interface. For the purpose of

  my work, I 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 (Figure 3.3). Stabilized annotations have already been introduced by Gauglitz et.

  al. (2014) but their benefits (and possible drawbacks) in different task configurations

  were not precisely identified. Stabilized annotations are harder to be implemented

  compared to non-stabilized annotations since they need augmented reality technologies to

  attach annotations on objects in the scene.

Note that for both types of annotations, I enabled freehand drawings to be made, not limiting the

users to any pre-defined shapes (e.g., push-pins, etc.).

In summary, my 2×2 device/annotation configurations are depicted in the matrix in Table 3.1:

	Head-mounted Display	Tablet
Non-stabilized Annotations		
Stabilized Annotations		

Table 3.1. Study device/annotation configurations.

3.2. System Implementation

The system was developed by a fellow undergraduate student of computer science with constant collaboration in all the stages of the design and implementation. It was developed on the Unity game engine along with Qualcomm's Vuforia for AR marker tracking. The Vuforia toolkit allowed us to paste a drawing plane to an image target placed in the real world. This gave the illusion that the drawings themselves are aligned to the target surface when users move the camera around. The planes are placed 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
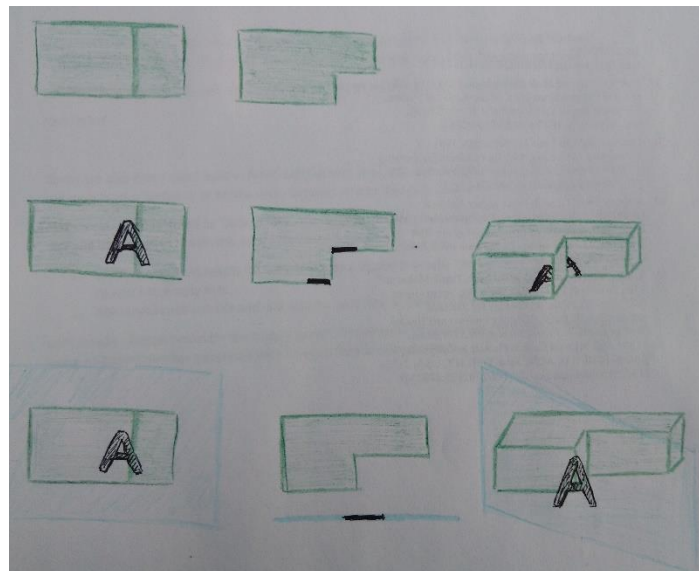
Figure 3.4. Idea of annotation planes – the "A" annotation is drawn on an annotation plane in the bottom row

32

are visibly detached from (but clearly associated with) the surface (Figure 3.4). For some tasks, we affixed this plane to the dominant surface. For others, where two sides of an object (for example Origami paper) were needed, we created two planes, one for each side. And for 3D tasks where the final structure has an inherited final structure (like a box in a Lego task where the structure looks like a cube – Figure 3.6) we created four such invisible planes around the box. In addition, Vuforia's extended tracking feature allows users look around more freely even when the image target itself is not in view of the camera.

Video streaming was accomplished through Unity's legacy Network View components as these were able to transmit the large byte data required for the video stream. The stream itself ran at 12-15 frames per second with JPG compression at a quality of 25% on Wifi. The quality between the camera feeds was lowered in this manner to keep the tablet and HMD video quality consistent on the helper's end.

With AR image targets (Figure 3.5) we are able to design tasks on any surface. The targets can be printed in a reasonably large form factor (5×7") and be placed arbitrarily on flat surfaces. With Unity, we can define any arbitrary model to act as a drawing plane and calibrate its actual size to the real world. We chose to implement a 3D box (used in the Lego task (Figure 3.6)) and

Figure 3.5. Image targets

Figure 3.6. Epson Moverio HMD

a flat plane (used in the tangram, origami and graph task). In the case of the origami task, the

paper acted as the plane. Because it could be lifted and flipped, this allowed us to display

annotation directly, and separate annotations for either side of the paper. More details about the

aforementioned tasks and their characteristics are introduced in the next chapter.

3.2.1. Hardware

This prototype uses a number of consumer oriented devices. The worker used either the Moverio

BT-200 (Figure 3.6) 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, for the helper a Microsoft Surface tablet was used as a drawing

tablet with a capacitive touch pen. All devices were connected on a Linksys Dual band router at

2.4 Ghz.

3.2.2. User Interface

T The user interface was designed in a simple and easy-to-use way so that all the tools would be easily accessible for the helper person (Figure 3.7). The helper had a number of tools at their disposal:

- Freehand drawing tool

- Line drawing tool

- "Clear all" button

- Eraser tool

- Changing the brush thickness

- Changing the brush color

- Turning the annotations on/off

They could either use a capacitive touch pen or their finger to interact with the system.
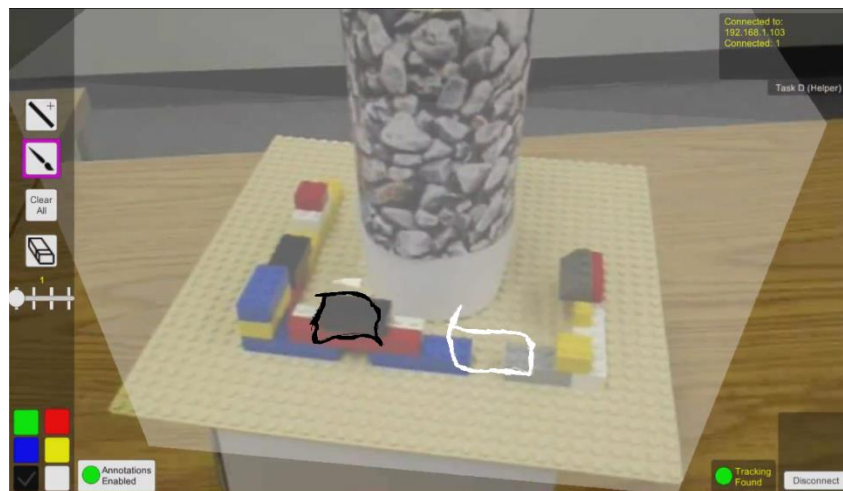
Figure 3.7. Box shaped transparent annotation planes and
the user interface of the system

3.2.3. Limitations

In addition, to setup the stabilized annotations, we needed mechanisms to track the position on which the annotations were drawn. This is usually achieved by using image targets, letter-sized papers with some unique patterns printed on them. Image targets do have their drawback, as users must maintain their camera within view to enable annotation drawing. These targets themselves may obstruct or be obscured by other objects as the participants progressed through the tasks. In addition, they might not be available for users in real-world scenarios.

3.3. Summary

In this chapter, I provided a detailed overview of my design process. Since the main goal of the project is to study different annotation systems for mobile remote assistance, I first introduced the context in which such systems might be used in real life, then selected the various configurations of interest. Thereafter, I moved on to the prototype design and implementation and how these prototypes could then be involved in studies across different conditions. Particularly, I addressed Thesis Questions 2 & 3 and Thesis Objective 2 by designing a remote assistance system which incorporates different annotation systems and device configurations.

The next chapter describes the studies in more detail, and introduces findings.

Chapter Four: User Study

In the previous chapters, I reviewed the main challenges that the researchers have faced in designing effective remote assistance systems. What we have seen is that various kinds of remote assistance features can aid collaboration; however, it is likely that different task factors affect the utility of different features. But, we do not have a systematic understanding of what this relationship between task factors and features is. To address this issue, I designed a comprehensive study to observe user behaviors while using such systems. In particular, I took as many task factors as possible into account to design study tasks that could mimic a broader range of real-world scenarios.

This study explores the trade-offs between variations of annotation systems over a range of four different collaborative support tasks. In the study, I examine the use of stabilized freehand annotations and freehand annotations atop video (non-stabilized annotations), contrasting their

use with tablet devices and head-mounted video see-through displays. The purpose of running this study is to observe how users communicate with one another using each of the four different configurations to identify when each combination of annotation/device is likely to be useful.

Specifically, in this chapter I address the following:

- Thesis Question 2 (How can changing different aspects of mobility improve interaction?) and Thesis Question 3 (How does "stickiness" of stabilized annotations affect their utility?) by conducting a comprehensive user study exploring the usage of different devices and annotation systems for remote assistance tasks.

- Thesis Objective 3, which is to design and conduct a study that evaluates the use of an annotation system that contrasts the display/capture tool, and the style of annotations.

- Thesis Contribution 3 to outline a framework of annotation use that provides an understanding of when and why stabilized annotations are likely to be valuable, and Thesis Contribution 4 by presenting a study that provides new findings suggesting about both task and design factors that affect the utility of stabilized annotations.

To foreshadow the results of the study, I found that:

1. Most of the annotations made by the helpers were quick and ephemeral in utility and need. They were meant to be used to convey instructions for a very short amount of time.

2. The helpers found it very difficult and annoying to draw annotations when the camera or objects were moving. Thus, they repeatedly asked the workers to "stay still" and not to move the camera or objects.

3. People do prefer stabilized annotations and head-mounted displays for all the tasks. However, I did not observe any meaningful performance benefit for these technologies, so the preference might have been only because of an effect of novelty.

In this chapter, I will introduce my study, describing the study design process, my participants' demographics, setup and procedure and the findings.

4.1. Study Design

Mobile remote assistance systems could be used for a wide variety of scenarios. These scenarios are usually carried out by two roles (described in section 3.1): (1) A *Worker* who is the person in need of help to complete the task, and (2) a *Helper,* who has all or part of the knowledge to complete the task, but is not physically co-present with the Worker.

There is a broad range of possible tasks and their requirements might influence system design. In this thesis, my focus is on designing systems to enable a remote expert assist a worker in the workplace during a physical task. Even here though, the range of possible tasks varies extensively with each task having its own requirements and task factors. As we saw, the problems with interpreting the results from the prior work has been their inability to articulate the impact of different task factors in designing the studies. Each study was run differently, considering only a small number of task factors and device configurations, which might have influenced the findings. In addition, most of the previous study tasks were simply step-by-step construction tasks which do not represent the variety of possible tasks in remote support scenarios. In this work, I articulate the salient features of this task space, select and then create a set of representative abstract tasks for studying how well an annotation subsystem supports the execution of tasks.

These task factors can have a huge impact on the performance of remote support systems, leading one system to perform differently across different tasks. To address this, I explored different types of remote assistance task from a few different perspectives. I 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, I aimed to explore a wider breadth of task styles for studying remote assistance systems. To this end, I 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). I also wanted to 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 (2015) explore the role of having to physically move about in the workspace—that is, environments that cannot be captured in a single camera frame. I am 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., Kim et. al., 2015), rather than simply focusing on objects on a simple flat surface.

Complexity of physical task. Finally, I was 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.

Thus, I designed tasks for my study that systematically varied each of the mentioned factors, as I expected different combinations to produce different kinds of interaction needs; I expected this to help tease out the benefits of the different technology designs for remote support. To arrive at my final task selections, I brainstormed 20 potential tasks which could vary different aspects of the aforementioned factors. Then I grouped them into sub-categories of similar tasks and finally chose the tasks that I thought annotations could be most useful. This was done in consultation with the team of researchers that I worked with.

4.1.1. Study Tasks

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 (e.g. the workspace could be contained in 2' × 2'), and while challenging, is not a complex task (it can be completed with a

Figure 4.1. Tangram task

single hand) (Figure 4.1). This is similar to a furniture assembly task in which the end state is obvious for the helper, but it might be hard to detect and point towards different parts in the scene (mainly because of video scaling issues – size of the objects in the scene are not exactly the same as their sizes in the video, and this is important for completing the task). In this task, helpers were not allowed (and didn't have access to any mechanisms) to pass the final shape to the worker. Tangram task instructions are shown in Appendix A.3.

Task B: Graph. Participants find a least-cost path between two nodes in a large (approximately 4' × 4') graph (comprised of nodes and edges), where each participant is only given information about half the edges. As shown in Appendix A.4, there were two variations of this task. In each variation, the worker and the helper were given a table of 5 or 6 separate entries. Each row indicates two nodes and the cost associated with the link connecting the two. The node symbols were already put up on a whiteboard in the worker's workspace. Since finding the least-cost path between two specific nodes (instructed by the study coordinator) required knowledge about link costs from both the worker and the helper tables, participants had to communicate their share of

Figure 4.2. Graph task
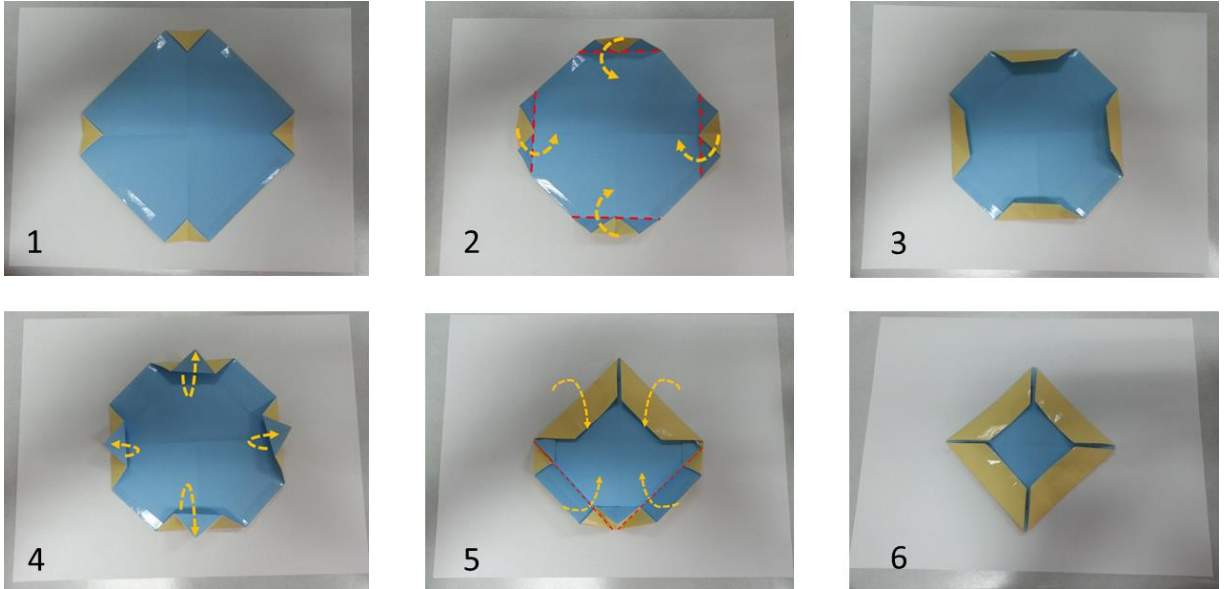
Figure 4.3. Origami task

nodes to their partner. A possible approach to accomplish this task is to first: trying to

communicate each person's share of edges so that they have a good understanding of the graph

layout and second: communicating verbally to identify the path with the least cost. The task here

is essentially a graphical, problem-solving task where knowledge is explicitly distributed, and the

physical complexity of the task is low (Figure 4.2). This task is similar to some tasks in industrial

settings (e.g. in power plants) where the area to work on is quite large, and requires contribution

from both sides to solve a problem.

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. An example

subset of the steps for this task is depicted in Figure 4.3, while the complete task instructions are

available in Appendix A.5. 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 requires two hands to make folds (Figure 4.3). This task is quite

similar to a task like assembling furniture, since it follows concrete, step-by-step instructions.

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 (Appendix A.6). 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 (Figure 4.4). This task could mimic a simple home theatre setup where pointing, detaching and re-attaching the objects happen.

In terms of external validity, my tasks are based on a long history of research in this area (Bauer et. al., 1999, Fussell et. al., 2004, Johnson et. al., 2015, Kirk et. al., 2006 & 2007, Ou et. al., 2003), and are also related to scenarios in everyday life. These specific tasks are unlikely to be encountered in a real-world remote support scenario. But I have tried to mimic and simulate the core sub-tasks in a diverse set of real-world tasks. 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.

Overall, major characteristics of the tasks could be summarized in Table 4.1:
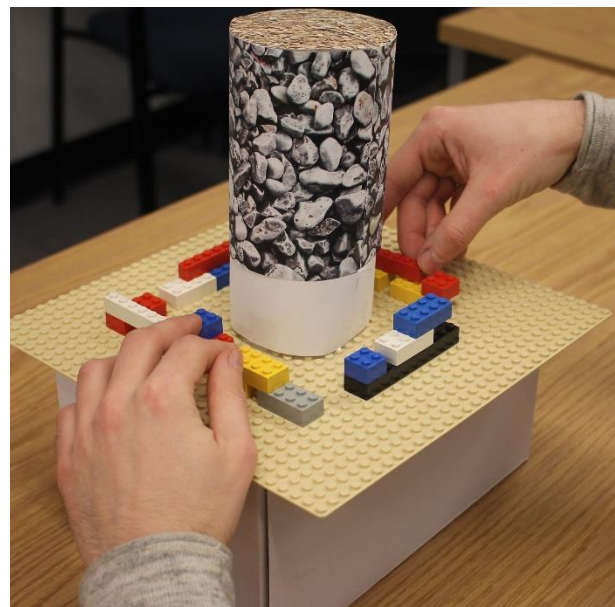
Figure 4.4. Lego task

Task	Knowledge	2D/3D	Use of hands
Tangram	Helper	2D	One
Graph	Shared	2D	One
Origami	Helper	3D	Both
Lego	Shared	3D	Both

Table 4.1. Summary of task factors

4.2. Study

As described in earlier and in section 3.1, I designed my observational study with two roles, worker and helper. The helper person always used a tablet to communicate instructions to the worker while the workers used either a tablet or a head-mounted display. The helpers were also able to draw annotations to be appeared on workers' screen (Figure 4.5).

As described in Chapter 3 (section 3.1), I considered four different conditions with 2 annotation systems (stabilized vs. non-stabilized) and 2 device configurations (tablet vs. head-mounted display) to be explored in the study. Then, I designed the four different tasks for my study (section 4.1.1) that systematically varied each of the mentioned task factors in section 4.1. Finally, the prototypes were built by a fellow student with my supervision in such a way that they would suit the study requirements. The prototype design and implementation were described in sections 3.1 and 3.2. In this section, I elaborate on my study specifications and the procedure, followed by the findings.

4.2.1. Participants

I recruited sixteen pairs of participants (32 total; 13 female) through posters displayed on a university campus. Participants were recruited in pairs and they all 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 chat as 2.9 (sd. 1.33, med. 3), and experience with augmented reality was low (avg. 1.71, sd. 0.87, med. 1).

4.2.2. Design

The tasks were set up as 2×2 within subjects design (stabilized vs. non-stabilized, and head-mounted display vs. tablet). I chose the within subjects design over the between subjects design because it allowed me to gather sufficient data from the study with a limited number of participants. It is also more generalizable than a between subjects design, since the individual variability and task assignment bias do not affect the study.

Ordering of the tasks was counter-balanced between the groups using the Latin square method to make sure factors like learning effect do not influence my observations. There were 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. I collected video and audio of the workspace, logged participants' interactions and took field notes.

4.2.3. Procedure

Participants were given a demographics questionnaire and taught how to: draw annotations with different tools, change the drawing color and thickness and turn the annotations on and off. 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. Basically, they had to collaboratively solve a math/geometry puzzle while each

having access to a subset of information. In this step, both the helper and the worker were introduced to both helper and worker systems to make it clear what their partner could or could not do.

Once the training task was completed, participants completed the four tasks (described in section 4.1.1); task presentation order was counterbalanced across groups.

I limited participants to seven minutes for each task, as I found from piloting the tasks that this allowed participants to comfortably complete or nearly complete each task, but allowed me to put an upper limit on the study length. After each task, I 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 each.

4.2.4. Tasks and Physical Setup

As described in section 4.1.1, I designed four tasks (Tangram, Graph, Origami, Lego) with different characteristics to mimic real-world remote assistance scenarios. All the tasks could be carried out indoor, in a small room. 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
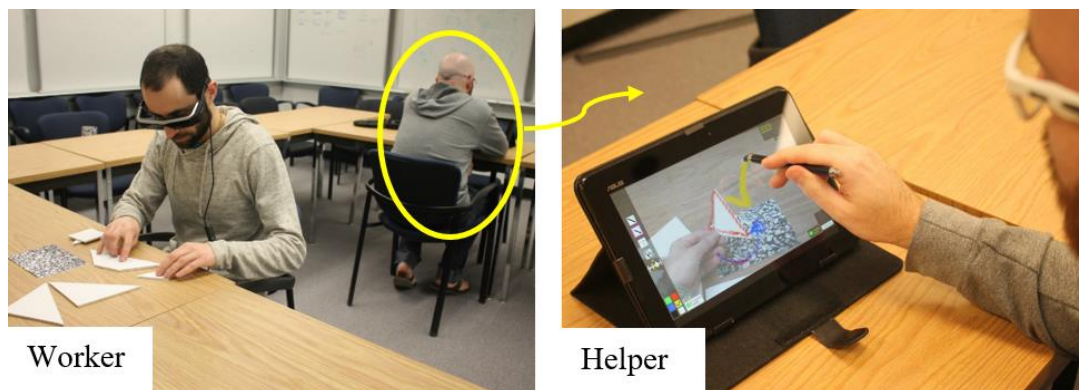
Figure 4.5. Study setup. The participants were actually sitting in the same room. The helper (right) guiding the worker (left) through the Tangram task.

47

hear one another (Figure 4.5). 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.

4.2.5. Data and Analysis

For each study, I took note of the observations right after the study as a content log. I 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, I used interaction analysis (Jordan et al., 1995) to analyze the annotations in context of the collaborative activities. I 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 the analysis, which referred to meaningful communicative acts throughout the tasks).

4.3. Findings

Overall, annotations played an important role in how participants completed the tasks. Of particular interest to me was how people might use annotations differently between task types. I also observed interesting tradeoffs between stabilized and non-stabilized annotations depending on the type of technology used. Some of the interesting findings were:

- I could classify helpers' annotations into three categories and count them across different tasks. This gave insight on which types of annotations might be useful for each task type.

- I observed other forms of communication, for example hand gestures or camera movements.

- Despite user preference, the more expensive technologies (head-mounted displays and stabilized annotations) did not necessarily outperform cheaper solutions in all scenarios.

Study findings are described in more detail below.

4.3.1. Role of Annotations

Although researchers (e.g., Ou et. al. (2003)) have introduced classifications of free-hand drawn annotations based on their shape (for example arrows, crosses, starts, circles, etc.), I could not find any schema based on the role of annotations and how they are used over time or in relation to other annotations. Thus, through analysis of annotations in my study, I developed a new three-category schema. The analysis was done by first watching all the video recordings and taking note of all the annotations made by the participants across groups. Then I grouped similar annotations based on: 1) frequent patterns of drawing annotations (whether they were accompanied by deictic references or by "action" verb instructions and 2) whether they were left on screen for a while or they were erased after a very short amount of time. Table 4.2 summarizes the three categories of annotations that I observed (reference, procedural, pointer; described below), along with the number of instances across each of the four tasks. More examples of each category are provided in Appendix A.11. Because I did not observe a noticeable difference in the distribution of these annotations across the hardware dimension or stabilized/non-stabilized dimension, I do not break out the distribution in this way.

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

Table 4.2. Annotation type counts across different tasks

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 I saw included "end-state" annotations, which showed how the current task components would appear once the task was complete, and "legend" annotations, which explained mappings between symbols and meaning. Figure 4.6 (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. In this example, the helper drew the outline of the final puzzle shape which was provided to him and waited for the worker to figure out how he should put the Tangram pieces together to solve the puzzle. Figure 4.6 (right) illustrates an example of a "legend" annotation observed in the Origami task. Here, the annotation indicates
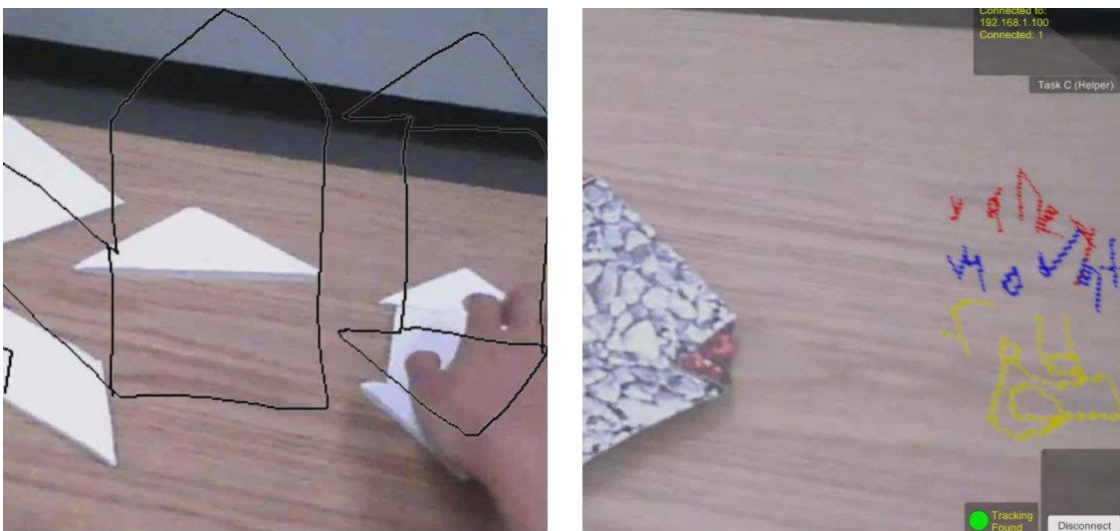
Figure 4.6. Left (group 11): An "end-state" annotation in the Tangram task. Right (group 2): A "legend" annotation in Origami task.

50

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. In this example, the helper has tried to convey the three types of fold with different colors: red lines will illustrate valley folds, blue lines will illustrate mountain folds and yellow lines will represent simple folds.

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 4.7 (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. Specifically, the helper asked the worker to fold the Origami paper along the green and red lines. Use of lines and arrows was also observed in this type (Figure 3.3). Figure 4.7 (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.
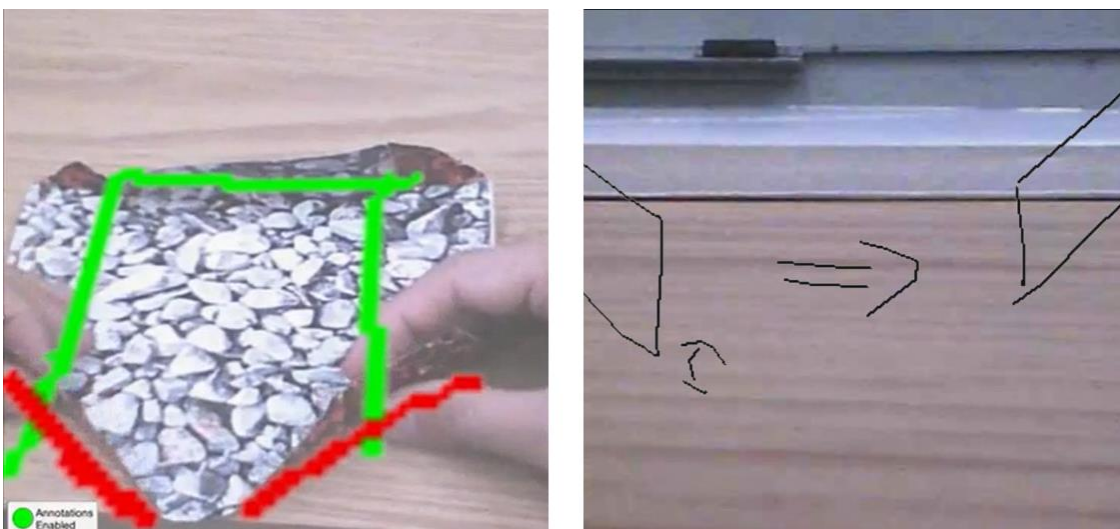
Figure 4.7. Left (group 2): Fold lines as a procedural annotation in Origami task. Right (group 2): A procedural annotation asking for the rotation of a Tangram

Pointer annotations. I also observed a variety of pointers to temporarily point at objects in the workspace. I saw instances of dots, arrows, circles or scribbles to point at objects (or targets) in the workspace. Figure 4.8 illustrate instances of these in the Lego and Origami tasks. In Figure 4.8 (left) the helper puts a check mark on the Lego blocks that are correctly positioned and in Figure 4.8 (right) the helper puts an "X" on one edge of the Origami paper that should not be folded. This type of annotations is particularly relevant to the role of deixis, since they were often accompanied by deictic words or phrases such as "this object" or "put it there".

4.3.2. Annotations in Task Completion

Participants used annotations as a means to support the completion of the tasks. Yet as illustrated in Table 4.2, the distribution of such annotations varied widely across tasks. I examine this result next.

Tangram Task: As helpers could not pass the outline of the goal shape as an image to the worker, most participants completed this task by first drawing (and then leaving) an outline of goal shape—an end-state reference annotation. Then, most pairs would operate in such a way that the
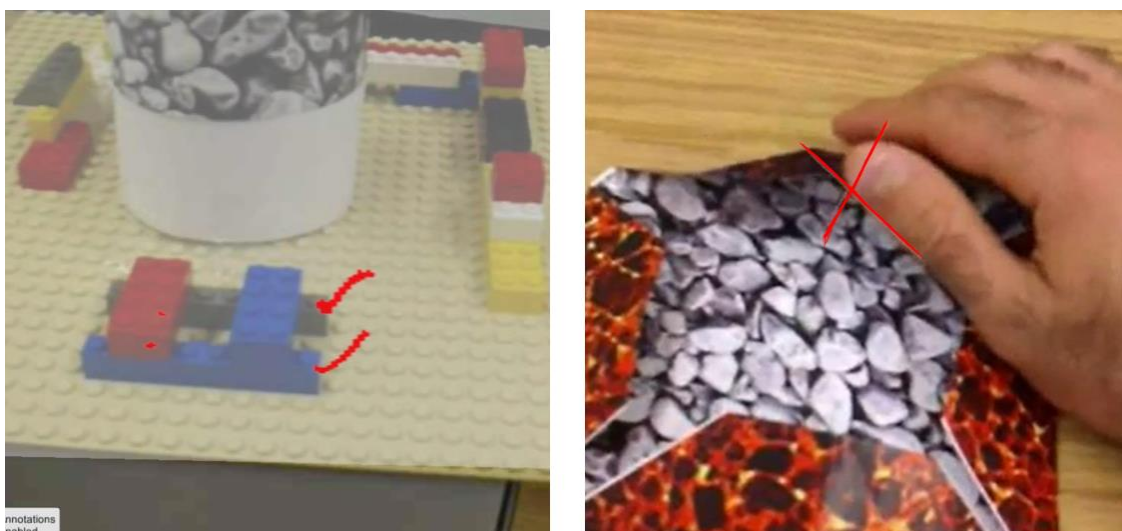
Figure 4.8. Left (group 6): Helper marks correctly placed Lego blocks. Right (group 16): Helper puts "X" on the edge that should not be folded.

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 (Gauglitz et al., 2012) 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, I did not observe meaningful performance gains (e.g., completion time) even though 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 were 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.

The distribution of different annotation types across the four tasks is depicted in Figure 4.9. As shown here, the majority of annotations (procedural and pointers) are used in ephemeral fashion (61 percent of annotations) especially in Origami and Lego tasks. This implies that although stabilized annotations are beneficial for certain scenarios in which annotations might be used after a while, cheaper non-stabilized annotations can be enough to make quick, ephemeral annotations.
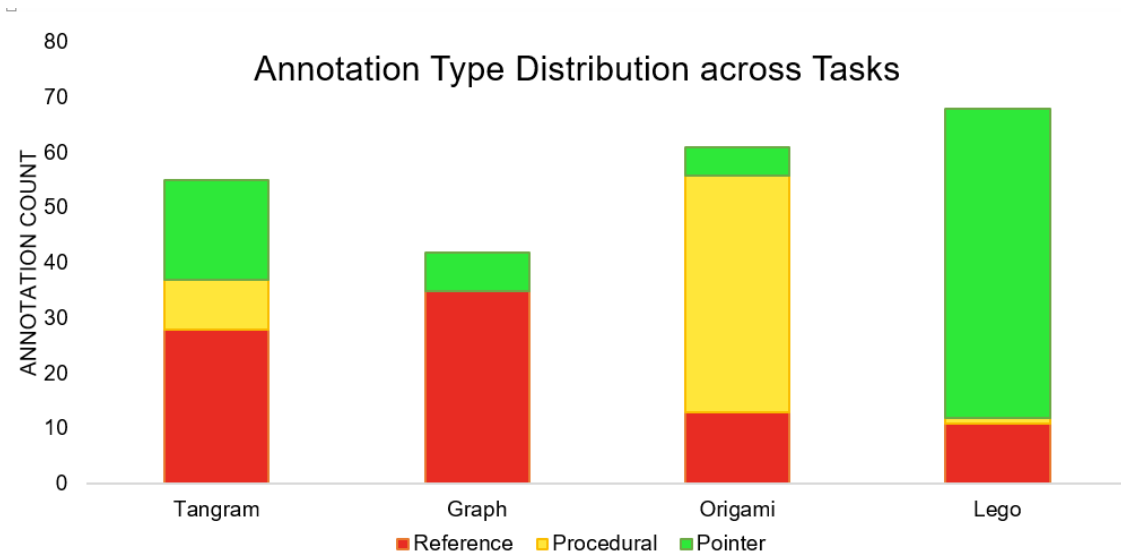
Figure 4.9. Annotation type distribution across tasks

4.3.3. 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) I 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.

"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 4.6 (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 (Bier et al., 1993) 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.

4.3.4. Other forms of Communication

Camera Mobility: In four groups, the tablet's mobility was used as a conversational resource. As

illustrated in Figure 4.10, 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. I 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. I observed this behavior in 5 of the 16 groups.

4.3.5. 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-

Figure 4.10. The worker (group 10) "zooms in" to provide a more detailed
shot of the graph task for the helper.

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 seen as having low-resolution, giving a poor field-of-view at an awkward angle, and heavy. It was also a bit hard for users wearing glasses. The following vignette from Group 15 demonstrates this issue:

> *Worker:* They (HMD) were heavy, plus I wear glasses and I had to take them off to wear those, so I had trouble seeing the symbols and they weren't clear."

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

58
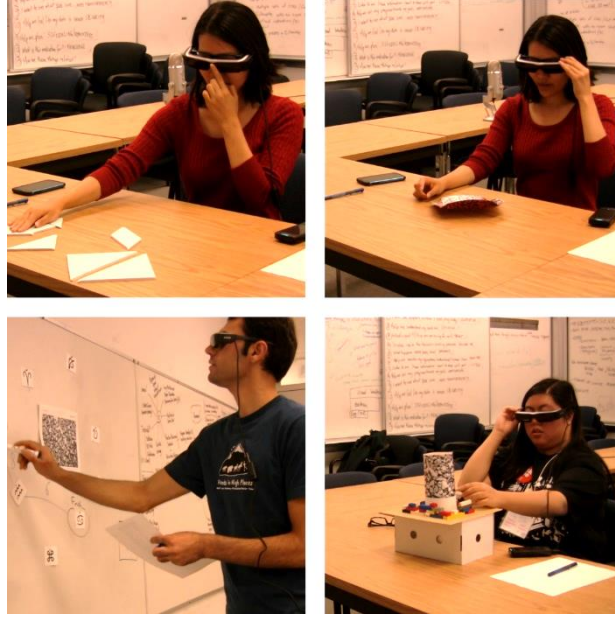
Figure 4.11. Workers holding the HMD with one hand (top, bottom-right) or
looking below the goggles (bottom-left).

workspace through the goggles). The following vignette from Group 15, together with Figure

4.11 illustrate this point:

> *Worker:* Yeah it (HMD) 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."

4.3.6. Preference by Task

Table 4.3 shows worker device and annotation-type preference by task condition (one group

failed to complete the questionnaire). I saw general preference for the head-mounted display over

the tablet, and participants preferred stabilized annotations over non-stabilized annotations.

	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 4.3. Workers' device and annotation preferences

59

4.3.7. Task Completion Time

The average completion time of tasks across different configurations is shown in Table 4.4 (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. Also I conducted a t-test to compare the two variations of annotations. Unlike what I expected, here was no significant difference in the completion times for stabilized annotations (M=351.12, SD=80.62) and non-stabilized annotations (M=334.06, SD=96.67); $t(31)=0.71679$, p < 0.05. These results suggest that annotation type really does not have an effect on task completion time.

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

4.4. Study Limitations

The recruited participants were local university students with a limited age range and experience level rather than task experts. I tried to mimic many real-world remote assistance scenarios by identifying potential task factors and choosing my study tasks based on those factors. While this is a step forward compared to the previous work, there is a possibility to run studies with more specific tasks (i.e. car repair) and field experts (i.e. mechanics). I selected this participant pool due to the early state of the prototypes and difficulties recruiting and working with more

specialized populations. Specifically, with the more "expert" population, I was concerned about time and transportation issues that would affect their availability. I took these problems into consideration before running the study and chose to use a more readily available student population to gather more results for future iterations.

In addition, although the worker role can be assigned to anyone who has little or no knowledge about the task which can be applied to the students taking on this role, the students taking on the helper role were not necessarily the intended user base; since they had no previous knowledge about the tasks. Overall, while participant pool is a tricky limitation to many user studies and I have tried to address this issue by designing multiple tasks with different task factors, it is still not clear how my findings will transfer to "real-world" scenarios.

Although this was the first study being conducted with such a diverse set of task factors, there are still other factors that could be explored in future work. Some of the extra factors that I have identified as potential additions are: 1) whether the task is inherently physical (like the Origami task) or visual (like the graph task), 2) whether the task is carried out on a vertical surface (like the Graph task) or on a horizontal surface (like the Tangram task) or potentially in-the-air (like knitting), 3) what kind of knowledge matters the most to complete the task (end state or step-by-step process), etc.

Finally, there is a missing "in-between" condition in the head-mounted device configuration for which I did not explore in this study. While most consumer-grade head-mounted devices have integrated both a camera and a display and I narrowed down my focus on such devices in this thesis, future work can explore the two other possible variations:

1) Head-mounted cameras without any display screen: In this condition, the helpers' annotations can be shown on an external device (e.g., a tablet).

2) Head-mounted displays without any camera: In this variation, the helpers' instructions can be shown on the head-mounted device but the worker's camera can be a static or mobile device (e.g., a webcam or a smartphone).

4.5. Summary

In this chapter, I introduced my observational study to explore how users interact with remote assistance systems with different devices and annotation system. Of particular interest was to determine the task factors which affect the usage of our system, and what configuration can be useful for each type of task.

I started by introducing my study design process and then moved on to describe the study setup, procedure and findings. Based on the observations, I classified the annotations into three categories: reference, procedural and pointer annotations. I noticed that some types of annotations are drawn more frequently in certain tasks based on task characteristics and requirements. Correspondingly, I observed participants employing different strategies for different tasks.

I 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 by freeing up both the workers' hands, it still has limitations; it is heavy and it can be difficult to see through.

Finally, I 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.

In particular, I addressed Thesis Questions 2 & 3, Thesis Objective 3 and Thesis Contributions 3 & 4 by designing and running an observational user study to explore user behaviors while using different annotation systems and device configurations under different task specifications.

The next chapter summarizes the findings for this thesis and discusses current limitations, as well as suggestions for future work.

Chapter Five: Conclusions

In this chapter, I conclude my work with the remote assistance systems. Through running my user study, I have answered the thesis questions originally raised in the first chapter. I have also addressed the thesis contributions and thesis objectives. I now reflect on my work to discuss potential areas for future work and my contribution.

The main research question for this thesis was to develop an understanding of how *to design effective annotation systems for mobile video chat*. I address this general research question by dividing it into three thesis questions: the study of 1) the aspects of mobile remote support systems that needed to be improved, 2) mobility of different devices to support such tasks and 3) utility of stabilizing the free-hand annotations.

Throughout this thesis, I described the process I took to design and implement prototypes for remote assistance scenarios. As the main contribution of my work, I conducted a series of user

studies to evaluate the effectiveness of those prototypes under different task factors and conditions. Particularly, I tested prototypes with two different device configurations (tablets and head-mounted displays) and two different annotation systems (stabilized and non-stabilized annotations). Overall, the results of the studies suggested that although newer technologies (head-mounted displays and stabilized annotations) are generally preferred by the users, they do not necessarily outperform the cheaper and easier approaches in all task scenarios. As a result, when designing remote support systems, we have to take a step back and consider the task factors and conditions associated with our task to be able to design effective remote support systems.

That being said, there could also be some immediate improvements on current prototypes including the introduction of temporally stabilized annotations which sit in between non-stabilized and stabilized annotations in terms of implementation cost and effort but they might still be enough for many tasks, as well as the integration of hand gesture visualization mechanisms as we saw many helpers trying to convey instructions with their hand gestures.

In this final chapter, I first discuss major findings observed in the study. Following that, I raise topics for future work to guide work on similar systems that may build on the work in this research. I then conclude with my contributions and final remarks.

5.1. Discussion

5.1.1. Utility of Stabilized Annotations

Overall, and consistent with previous work by Gauglitz et al (2014), I did not find stabilized annotations to be a clear winner. However, my 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 annotations 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, I still saw the use of arrows, curved lines, etc., suggesting that the free-drawing annotation tool was still valuable.

As Table 4.3 (Chapter 4) 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 the 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 (Gauglitz et al., 2014). I avoided this problem by predefining drawing planes. I 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, my 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 my study design, which employed a range of different tasks varied over my 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 the Lego task). My results show that having a range of tasks is extremely important in the assessment of new technology and techniques. My 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.

5.1.2. 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).

5.2. Future Work

Based on the current state of my prototype, I have identified several areas for future work, including immediate system improvements and additional implementation beyond it. I have

described these to provide context for where my remote support systems stand at the time of writing this thesis, and to provide readers with possible areas for later work in this field.

5.2.1. Immediate Improvements

In terms of the remote assistance prototypes, some of the present limitations may be resolved by immediate technical improvements. For example, there were some problems regarding the head-mounted displays being bulky that could be resolved by using other, lightweight devices. We can also consider deploying a better network connection to get rid of minor video glitches.

5.2.2. Temporal Stabilization

With my study, I found that stabilized annotations were potentially useful in many situations. However, this stabilized annotation system required a sophisticated library and trackers to be incorporated into the workspace and objects. Further, I 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. In fact, stabilized annotations only provided stabilization *after* the annotation was drawn. However, we observed most problems arose from the lack of stabilization *during* the annotation creation. To address these issues and as an immediate upgrade, we can design another prototype to freeze the current video frame while an annotation is being drawn. This allows users to draw *temporal 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). This upgrade is also easy to implement in traditional video chat systems such as Skype, Hangouts, etc. To explore this idea, we implemented and studied a temporally stabilized annotation system and reported the findings in the conference paper that resulted from this work. To briefly foreshadow the results, this simple design was extremely effective, making it easier for the helper to draw the 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.

Furthermore, because helpers had not observed what the worker had done in the meantime, they

were unable to provide "real time" feedback on the worker's activities. We can address this issue

easily by providing either manual control over when to return to live video, or by providing a

smaller live view (e.g., in the corner of the screen).

5.2.3. Gestures

As described in Chapter 4 (section 4.3.4), I observed that the helpers frequently made hand

gestures without actually drawing anything, and without the gestures being visible to the other

party, to communicate some actions that needed to be taken by the worker. These gestures

seemed mostly unintentional but can give a clue on a possible future work. The current prototype

can be upgraded by capturing, conveying and visualizing helpers' hand gestures for the worker

to provide a more intuitive user experience (Figure 5.1).
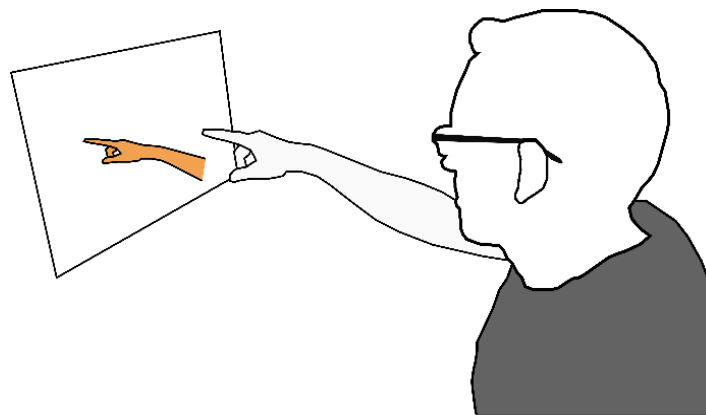
Figure 5.1. Capturing and visualizing helpers' hand gestures.

69

5.2.4. New Device Configurations

As discussed in Chapter 4 (section 4.3.4), some workers tried to provide more detailed views by "zooming in" the camera view with their tablet. This behavior makes sense especially in the tasks with a larger work area (e.g, Graph task). In such tasks, the user needs to have both overview and detail shots of the workspace. This issue can become more critical in outdoor scenarios where the workspace is much larger. To address this issue, we can integrate new devices such as drones (e.g., Jones et al., 2016) or 360-degree cameras to provide the overview shots and leaving the detail shots for handheld devices or head-mounted displays.

5.3. Design Implications

Remote support systems can incorporate a wide range of device configurations and annotation subsystems to enable effective communication between parties. However, I figured out that there is not a single solution for all possible task scenarios. When designing remote support systems, it is crucial to take as many task factors as possible into account and build a system which suits the task requirements. To summarize my findings and guide the design of future mobile remote support systems, I outline a set of design implications based on the observations from my study:

1. Annotation systems are extremely useful to make deictic references for remote support tasks. However, stabilized annotations are not clear winner for all task types. They are more useful for tasks where annotations are needed to be referred back over time. In other cases when only temporal annotations are required, non-stabilized or temporally stabilized annotations are sufficient for making ephemeral annotations.

2. Head-mounted displays are useful for tasks which require the worker to use both hands. They also provide continuous video stream of the workers' actions. However, this feature

is not critical in tasks which only one of the parties works at a time (e.g., the Graph task). In such tasks, a hand-held device is sufficient.

3. In the follow-up study to this thesis, I found that temporally stabilized annotations are a very cheap and effective way to provide stabilization. Users mostly made ephemeral annotations which can be addressed with this approach. Furthermore, workers do not need to hold the camera perfectly still while the annotations are being drawn.

5.4. Contributions

To state my contributions, I must first restate my research goals and thesis questions from the first chapter, I began this thesis with my research question:

How should we design effective annotation systems for mobile video chat?

To answer this question, I then posed three thesis questions:

Thesis Question 1: What design aspects of a remote annotation system can best improve the interaction in a remote assistance scenario?

Thesis Question 2: How does changing different aspects of camera mobility affect interaction?

Thesis Question 3: How does the "stickiness" of free-hand annotations (to the objects they describe) affect their utility within a remote assistance scenario?

To address Thesis Question 1, I studied the prior work in Chapter 2 and outlined the limitations of the prior studies to design our own prototype and user study in Chapters 3 and 4. By studying the prior work, I figured out that although there have been valuable work to study remote support systems, they have failed to identify system requirements in a per-task basis. As a result, I tried to address this issue in my thesis.

71

I answered Thesis Questions 2 and 3 by running a comprehensive user study on mobile support systems, considering several task factors that have been neglected in prior work, and comparing different types of annotations and different device configurations. In conclusion, it turned out that although there are commodity technologies (such as head-mounted displays or augmented-reality annotations) that could support remote assistance scenarios (and they might even be preferred by users), they are not necessarily the ultimate solution for all tasks. Instead, the cheaper and easier solutions (such as handheld devices or non-stabilized annotations) are sufficient for many scenarios. Particularly, I was able to classify the annotations made by the helpers and identify the ephemerality of most of them which could suggest that stabilized annotations actually overkill in many circumstances (compared to non-stabilized or temporally stabilized annotations).

Reflecting again on my research question, I have addressed the question and make a contribution by outlining the task factors involved in remote support scenarios and their influence on designing remote support systems. However, this thesis does not (and was not meant to) present a technical contribution, since our prototype is similar to some others in prior work. As a result, now that I have clearly outlined the task factors and their requirements, one can build on top of these and build more novel prototypes to support remote assistance scenarios (i.e., making use of helpers' gestures or new augmented/virtual reality technologies and devices).

5.5. Final Conclusions

Current designs of mobile video chat 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 thesis, I designed and evaluated a system to support remote assistance. The system

allowed users to use AR stabilized annotations on live video and incorporated a head-mounted display to free workers' hands. I saw that although stabilized annotations can improve performance in specific tasks, they do not necessarily outperform non-stabilized annotations in all tasks. Also head-mounted displays were valuable for freeing up the workers' hands.

My study has provided new insights on the ways annotations and different device configurations support communication in remote assistance tasks. Based on these findings, I have outlined several implications that will direct the design of future mobile video tools for remote support.

References

Bauer, M., Kortuem, G., & Segall, Z. (1999). Where Are You Pointing At? A Study of Remote Collaboration in a Wearable Videoconference System, 151. Retrieved from http://dl.acm.org/citation.cfm?id=519309.856501

Bier, E. A., Stone, M. C., Pier, K., Buxton, W., & DeRose, T. D. (1993). Toolglass and magic lenses: the see-through interface. *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, (Mmm), 73–80. http://doi.org/10.1145/166117.166126

Bly, S. a., Harrison, S. R., & Irwin, S. (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

Brubaker, J. R., Venolia, G., & Tang, J. C. (2012). Focusing on Shared Experiences : Moving Beyond the Camera in Video Communication. *DIS '12: Proceedings of the Designing Interactive Systems Conference*, 96–105. http://doi.org/10.1145/2317956.2317973

Buxton, B. (2009). Chapter 13 Mediaspace – Meaningspace – Meetingspace. *Media Space: 20+ Years of Mediated Life*, (2009), 1–13. http://doi.org/10.1007/978-1-84882-483-6

Collins, A., Apted, T., & Kay, J. (2007). Tabletop file system access: Associative and hierarchical approaches. *Tabletop 2007 - 2nd Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems*, 113–120. http://doi.org/10.1109/TABLETOP.2007.34

Domova, V., Vartiainen, E., & Englund, M. (2014). Designing a Remote Video Collaboration System for Industrial Settings. *ITS - Interactive Tabletops and Surfaces*, 229–238. http://doi.org/10.1145/2669485.2669517

Fussell, S. R., Setlock, L. D., & Kraut, R. E. (2003). Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. *Proceedings of the Conference on*

Human Factors in Computing Systems - CHI '03, (5), 513.
http://doi.org/10.1145/642611.642701

Fussell, S., Setlock, L., Yang, J., Ou, J., Mauer, E., & Kramer, A. (2004). Gestures Over Video
Streams to Support Remote Collaboration on Physical Tasks. *Human-Computer Interaction*,
19(3), 273–309. http://doi.org/10.1207/s15327051hci1903_3

Gauglitz, S., Nuernberger, B., Turk, M., & Höllerer, T. (2014a). In touch with the remote world.
*Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology -
VRST '14*, 197–205. http://doi.org/10.1145/2671015.2671016

Gauglitz, S., Nuernberger, B., Turk, M., & Höllerer, T. (2014b). World-stabilized annotations
and virtual scene navigation for remote collaboration. *Proceedings of the 27th Annual ACM
Symposium on User Interface Software and Technology - UIST '14*, 449–459.
http://doi.org/10.1145/2642918.2647372

Gauglitz, S., & Tobias, H. (2012). Integrating the Physical Environment into Mobile Remote
Collaboration, 241–250.

Gergle, D., Kraut, R. E., & Fussell, S. R. (2004). Action as language in a shared visual space. In
*Proceedings of the 2004 ACM conference on Computer supported cooperative work -
CSCW '04* (p. 487). New York, New York, USA: ACM Press.
http://doi.org/10.1145/1031607.1031687

Gurevich, P., Lanir, J., Cohen, B., & Stone, R. (2012). TeleAdvisor: a versatile augmented
reality tool for remote assistance. *Proceedings of the SIGCHI Conference on Human
Factors in Computing Systems*, (Figure 1), 619–622.
http://doi.org/10.1145/2207676.2207763

Gutwin, C., Roseman, M., & Greenberg, S. (1996). A usability study of awareness widgets in a
shared workspace groupware system. In *Proceedings of the 1996 ACM conference on
Computer supported cooperative work - CSCW '96* (pp. 258–267). New York, New York,
USA: ACM Press. http://doi.org/10.1145/240080.240298

Henrysson, A., Billinghurst, M., & Ollila, M. (2005). Face to Face Collaborative AR on Movile Phones. *Mixed and Augmented Reality, 2005. Proceedings. Fourth IEEE and ACM International Symposium on*, *1*, 80–89. http://doi.org/10.1109/ISMAR.2005.32

Huang, W., & Alem, L. (2013). Handsinair: A Wearable System For Remote Collaboration On Physical Tasks. *Cscw*, 153–156. http://doi.org/10.1145/2441955.2441994

Ishii, H., Kobayashi, M., & Grudin, J. (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

Johnson, S., Gibson, M., & Mutlu, B. (2015). Handheld or Handsfree?: Remote Collaboration via Lightweight Head-Mounted Displays and Handheld Devices. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 1825–1836. http://doi.org/10.1145/2675133.2675176

Jones, B., Dillman, K., Tang, R., Tang, A., Sharlin, E., Oehlberg, L., … Bateman, S. (2016). Elevating Communication , Collaboration , and Shared Experiences in Mobile Video through Drones, 1123–1135.

Jones, B., Witcraft, A., Bateman, S., Neustaedter, C., & Tang, A. (2015). Mechanics of Camera Work in Mobile Video Collaboration. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 957–966. http://doi.org/10.1145/2702123.2702345

Jordan, B., & Henderson, A. (1995). Interaction Analysis: foundation and practice. The Journal of the Learning Sciences. http://doi.org/10.1207/s15327809jls0401_2

Judge, T., & Neustaedter, C. (2010). Sharing conversation and sharing life: video conferencing in the home. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems – CHI'10*, 655–658. http://doi.org/10.1145/1753326.1753422

Kasahara, S., & Rekimoto, J. (2014). JackIn. *Proceedings of the 5th Augmented Human International Conference on - AH '14*, 1–8. http://doi.org/10.1145/2582051.2582097

Kim, S., Lee, G. A., Ha, S., Sakata, N., & Billinghurst, M. (2015). Automatically Freezing Live Video for Annotation during Remote Collaboration. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '15* (pp. 1669–1674). New York, New York, USA: ACM Press. http://doi.org/10.1145/2702613.2732838

Kim, S., Lee, G. A., Sakata, N., D??nser, A., Vartiainen, E., & Billinghurst, M. (2013). Study of augmented gesture communication cues and view sharing in remote collaboration. *2013 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2013*, 261–262. http://doi.org/10.1109/ISMAR.2013.6671795

Kirk, D., Rodden, T., & Fraser, D. S. (2007). Turn it this way. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '07*, 1039–1048. http://doi.org/10.1145/1240624.1240782

Kirk, D., & Stanton Fraser, D. (2006). Comparing remote gesture technologies for supporting collaborative physical tasks. In *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06* (p. 1191). New York, New York, USA: ACM Press. http://doi.org/10.1145/1124772.1124951

Lanir, J., Stone, R., Cohen, B., & Gurevich, P. (2013). Ownership and Control of Point of View in Remote Assistance, 2243–2252. http://doi.org/10.1145/2470654.2481309

Lee, D. Y., & Lehto, M. R. (2013). User acceptance of YouTube for procedural learning: An extension of the Technology Acceptance Model. *Computers and Education*, *61*(1), 193–208. http://doi.org/10.1016/j.compedu.2012.10.001

Ou, J., Chen, X., Fussell, S. R., & Yang, J. (2003). DOVE : Drawing over Video Environment. *Proceedings of the 11th ACM International Conference on Multimedia – MULTIMEDIA'03*, 100-101. http://doi.org/10.1145/957013.957034

Ou, J., Fussell, S. R., Chen, X., Setlock, L. D., & Yang, J. (2003) Gestural communication over video stream: supporting multimodal interaction for remote collaborative physical tasks. *In*

Proceedings of the 5ᵗʰ international conference on Multimodal interfaces – ICMI'03, 242-249. http://dx.doi.org/10.1145/958432.958477

Procyk, J., Neustaedter, C., Pang, C., Tang, A., & Judge, T. K. (2014). Exploring Video Streaming in Public Settings: Shared Geocaching over Distance Using Mobile Video Chat. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2163–2172. http://doi.org/10.1145/2556288.2557198

Rae, I., Mutlu, B., & Takayama, L. (2014). Bodies in motion: Mobility, Presence, and Task Awareness in Telepresence. *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI '14*, 2153–2162. doi:10.1145/2556288.2557047

Ranjan, A., Birnholtz, J. P., & Balakrishnan, R. (2007). Dynamic shared visual spaces. *Proc. CHI '07, ACM Press*, *1*, 1177. http://doi.org/10.1145/1240624.1240802

Sodhi, R. S., Jones, B. R., Forsyth, D., Bailey, B. P., & Maciocci, G. (2013). BeThere: 3D Mobile Collaboration with Spatial Input. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, 179–188. http://doi.org/10.1145/2470654.2470679

Tait, M., & Billinghurst, M. (2015). The Effect of View Independence in a Collaborative AR System. *Computer Supported Cooperative Work (CSCW)*, *24*(6), 563–589. http://doi.org/10.1007/s10606-015-9231-8

Tang, A., Neustaedter, C., & Greenberg, S. (2007). VideoArms: Embodiments for Mixed Presence Groupware. *Proceedings of the International Conference on Human Computer Interaction (HCI'06): People and Computers XX - Engage*, 85–102. http://doi.org/10.1007/978-1-84628-664-3_8

Tang, J. C., & Minneman, S. L. (1991). Videodraw: a video interface for collaborative drawing. *ACM Transactions on Information Systems*, *9*(2), 170–184. http://doi.org/10.1145/123078.128729

Zheng, X. S., Foucault, C., Silva, P. M. da, Dasari, S., Yang, T., & Goose, S. (2015). Eye-Wearable Technology for Machine Maintenance: Effects of Display Position and Hands-free Operation. *Proceedings of the ACM CHI'15 Conference on Human Factors in Computing Systems*, *1*, 2125–2134. http://doi.org/10.1145/2702123.2702305

Appendix A: Study Materials

A.1. Consent Form

Name of Researcher, Faculty, Department, Telephone & Email:

Omid Fakourfar, Master's student – Department of computer science, University of Calgary
Hesam Alizadeh, Master's student – Department of computer science, University of Calgary
Kevin Ta, Undergraduate student – Department of computer science, University of Calgary
Supervisor:

Tony Tang, Assistant Professor - Department of computer science, University of Calgary
Title of Project:

Video Conferencing for Remote Collaboration Tasks

This consent form, a copy of which has been given to you, is only part of the process of informed consent. If you want more details about something mentioned here, or information not included here, you should feel free to ask. Please take the time to read this carefully and to understand any accompanying information.

The University of Calgary Conjoint Faculties Research Ethics Board has approved this research study.

Purpose of the Study

This study is designed to explore the role of augmented annotations in collaborative tasks via mobile video conferencing. In the context of a few short collaborative tasks which are done remotely, we will study the role of stabilized and non-stabilized annotations on video. In addition, we will compare the effectiveness of traditional hand-held devices for mobile video conferencing with a head-mounted display.

What Will I Be Asked To Do?

You will be asked to take part in a study which involves performing a series of pre-determined tasks collaboratively with a remote partner. The tasks might involve looking for visual clues, solving small puzzles, etc.
At the end of the study, you will be given a short questionnaire to share your experiences with the researchers.

Your participation is entirely voluntary. You may refuse to participate altogether, or may withdraw from the study at any time by stating your wish to withdraw to the researchers.

Each study, including going through all the tasks and responding to the questionnaire should take approximately 60 minutes. You will receive a $20 remuneration for your participation.

What Type of Personal Information Will Be Collected?

Should you agree to participate, you will be asked to provide your gender, educational level, eye condition and age during the interview. You may decline to answer this. If you do answer, however, this information may be published with information from your interview.

There are several options for you to consider if you decide to take part in this research. You can choose all, some, or none of them. Please review each of these options and choose Yes or No after carefully reviewing the information below.

Please note that the main purpose for collecting the video is analysis of the exploration session and the interview content. However, with your permission, we might want to use video clips or pictures in presentations or other electronic media, but this can only happen with your consent. Please, indicate below if you grant us permission to use video recordings or parts of it from this interview. Any video clips or photographs will **not** be associated with your name or contact information. If consent is given to present identifiable video clips and/or photographs (see table below), then no anonymity can be provided and you will be clearly recognizable as a participant in this study.

I agree to let whole or parts of recordings from the study to be used, for presentation of the research results: Yes: ___ No: ___

I agree to let recordings or parts of it from the study to be used, for data analysis only: Yes: ___ No: ___

I agree to let my conversation during the study be quoted and referred by a pseudonym, in presentation of the research results: Yes: ___ No: ___

I wish to remain anonymous, but you may refer to me by a pseudonym: Yes: ___ No: ___

Please note that once photographed or videotaped images are displayed in any public forum, the researchers will have no control over any future use by others who may copy these images and repost them in other formats or contexts, including possibly on the internet.

Are there Risks or Benefits if I Participate?

There are no known harms associated with your participation in this research. Feel free to ask questions about this study at any time.

What Happens to the Information I Provide

You are free to withdraw from this study at any point. If this occurs, we will immediately stop collecting data from you, ensuring that only data for which you have given consent is used.

All data received from this study will be kept for five years. The investigator indicated on this form will have access to the raw data, as will future investigators or research assistants on this project. While the exact composition of this team will change over time, the primary investigator will remain on the project.

In any reports created based on this study, you will be represented anonymously, using a pseudonym or participant number (e.g. Participant 4). With your permission (as indicated in the table above) we may use quotes from your interview or video clips/pictures of your session in our published results; these will not be associated with your name, contact information, pseudonym, or participant number. No personal or confidential information will be published. Please note that once videotaped images are displayed in any public forum, the researchers will have no control over any future use by others who may copy these images and repost them in other formats or contexts, including possibly on the internet.

I would like to be contacted for participating in the next phase of this study: Yes: ___ No: ___

Please also note that absolute anonymity cannot be guaranteed in a group setting, as the researchers will be unable to control what is said by individuals outside of the session.

Signatures

Your signature on this form indicates that 1) you understand to your satisfaction the information provided to you about your participation in this research project, and 2) you agree to participate in the research project.

In no way does this waive your legal rights nor release the investigators, sponsors, or involved institutions from their legal and professional responsibilities. You are free to withdraw from this research project at any time. You should feel free to ask for clarification or new information throughout your participation.

Participant's Name: (please print) _____

Participant's Signature: _____ Date: _____

Researcher's Name: (please print) _____

Researcher's Signature: _____ Date: _____

Questions/Concerns

If you have any further questions or want clarification regarding this research and/or your participation, please contact:

Dr. Tony Tang
Assistant Professor - Department of Computer Science
University of Calgary
(403)210-6912, tonyt@ucalgary.ca

Omid Fakourfar
Master's student - Department of Computer Science
University of Calgary
Omid.fakourfar@ucalgary.ca

Hesam Alizadeh
Master's student - Department of Computer Science
University of Calgary
hesam.alizadeh@ucalgary.ca

Kevin Ta
Undergraduate student - Department of Computer Science
University of Calgary
kta@ucalgary.ca

If you have any concerns about the way you've been treated as a participant, please contact an Ethics Resource Officer, Research Services Office, University of Calgary at (403) 210-9863; email cfreb@ucalgary.ca.

A copy of this consent form has been given to you to keep for your records and reference. The investigator has kept a copy of the consent form.

A.2. Task Configuration Assignment

Configuration Plan

Tasks:

Task A - Tangram

Task B - Graph

Task C - Origami

Task D – Lego

Configurations:

#	AR	HMD	Description
1	No	No	Non stabilized Tablet
2	No	Yes	Non stabilized HMD
3	Yes	No	Stabilized Tablet
4	Yes	Yes	Stabilized HMD

Configuration assignment

Group	T1	T2	T3	T4
1	A1	B2	C3	D4
2	A2	B1	C4	D3
3	A3	B4	C1	D2
4	A4	B3	C2	D1

A.3. Tangram Puzzle Task Instructions

- White silhouettes show the outline given to the helper, black shows the solution. There were two variations of the task.

Design A:
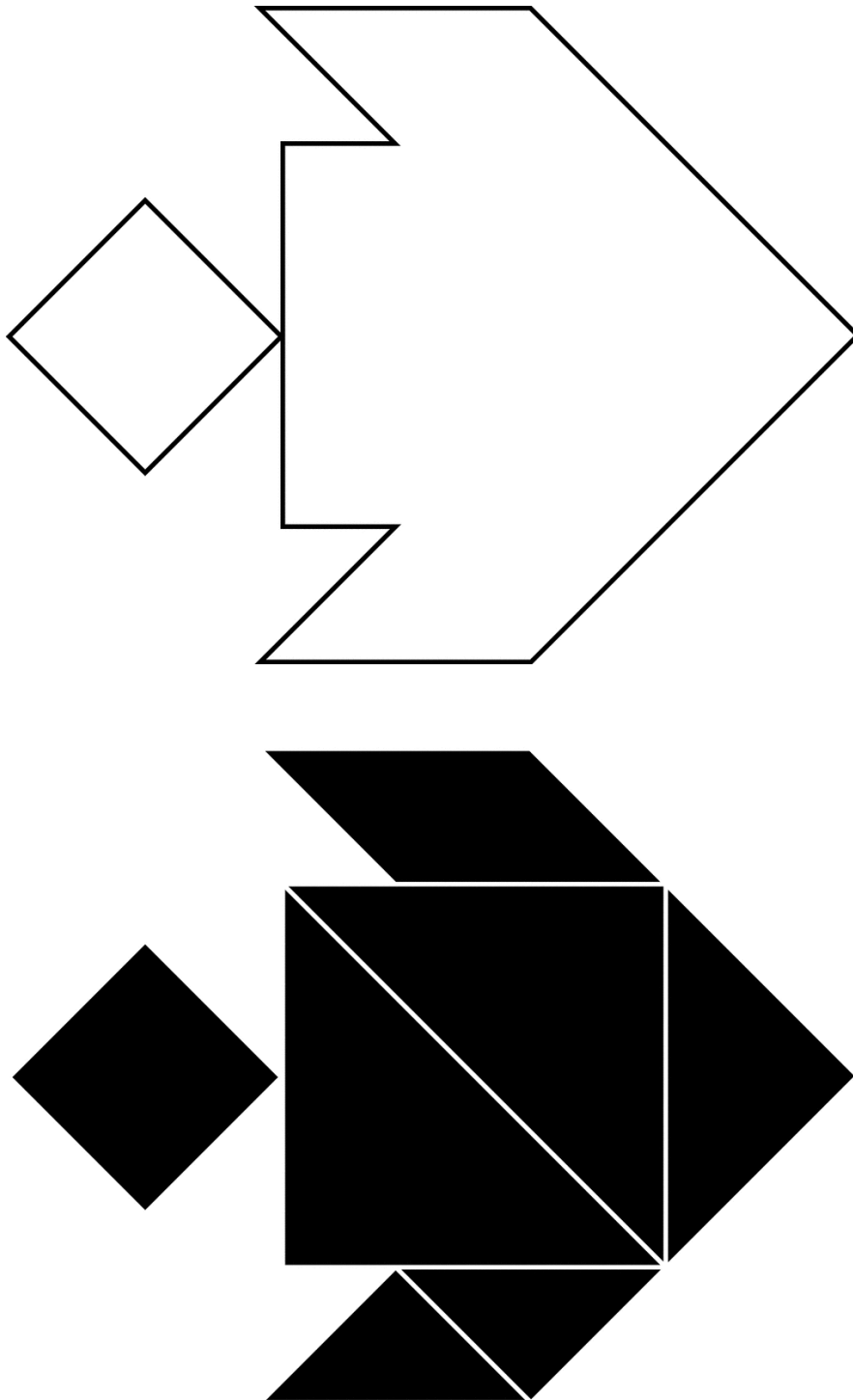
Design B:
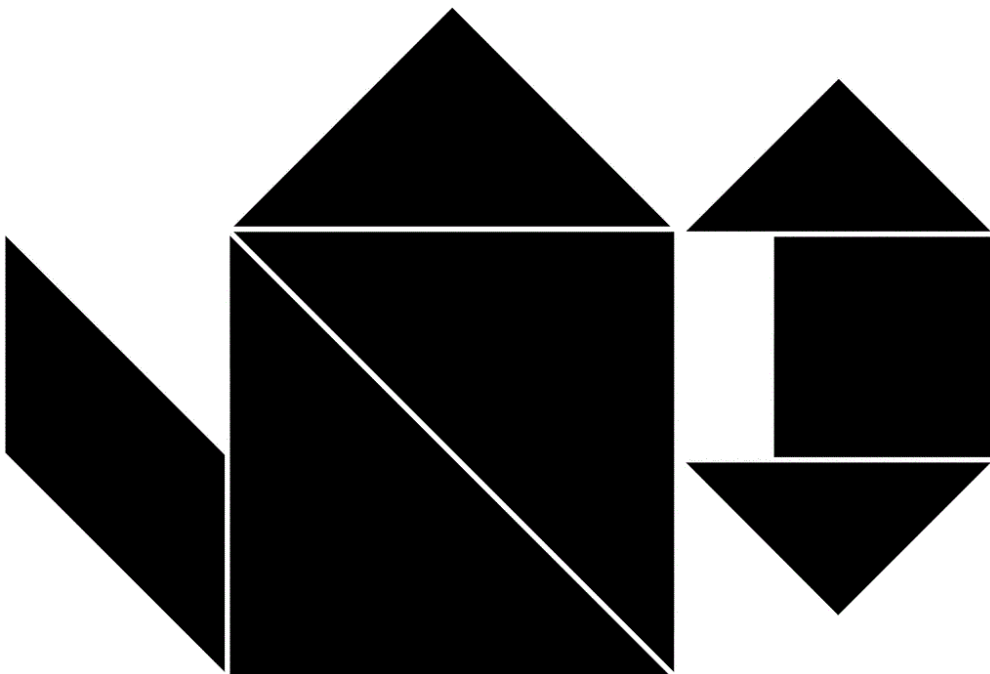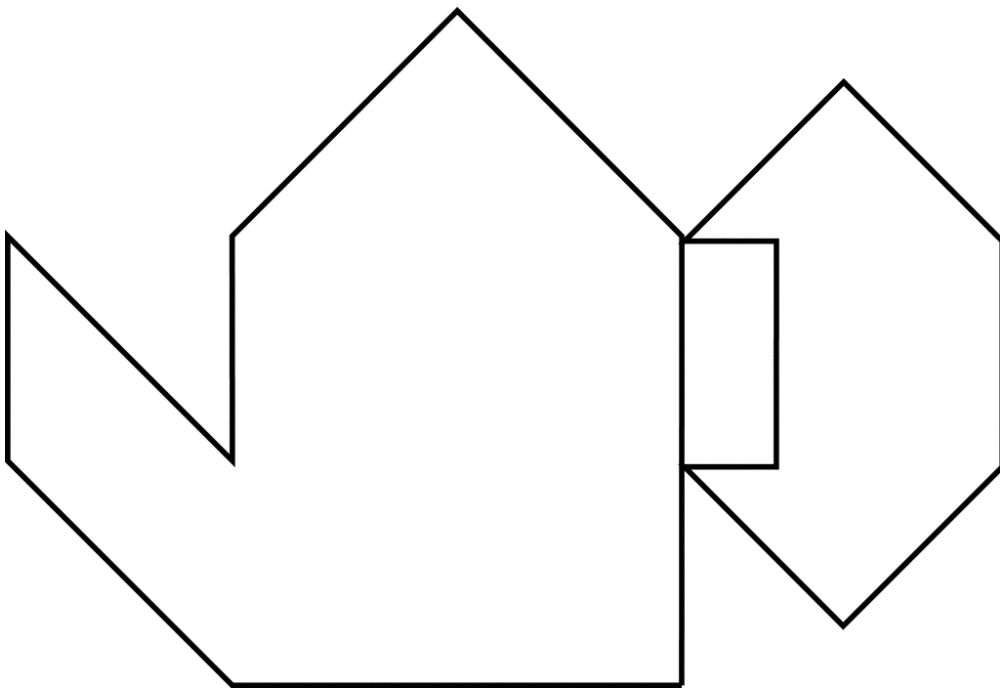
A.4. Graph Task Instructions

- The worker and the helper were given one table each, indicating a subset of link costs. There were two variations of this task.

Design A:

FROM	TO	COST
♉	♈	9
♉	↻	3
↻	♋	7
♎	♋	1
♎	⌘	6

FROM	TO	COST
♈	↻	7
♈	♋	4
♉	♎	2
♉	♒	9
♉	⌘	5
♒	⌘	7

Design B:

FROM	TO	COST
♉	♎	4
☊	♎	8
☊	♋	4
♎	♋	1
♎	⌘	1

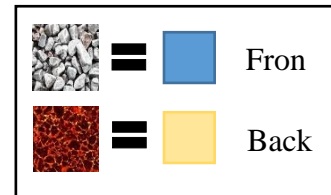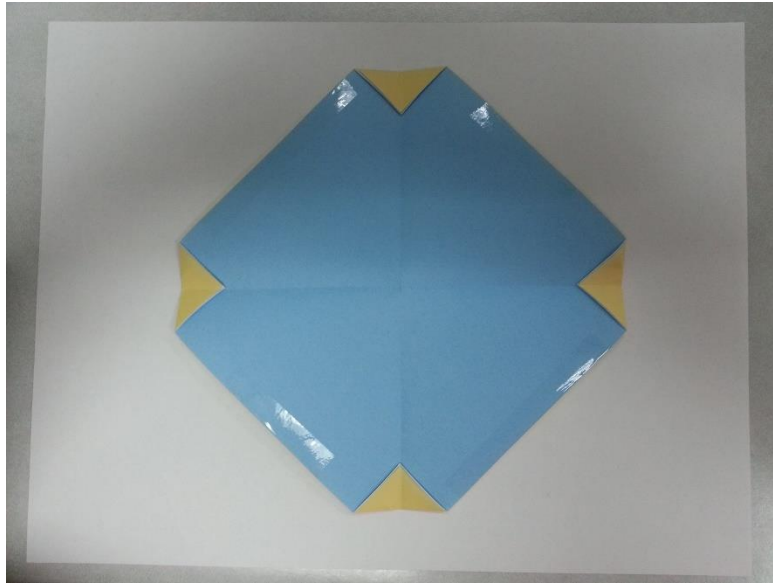| FROM	TO	COST
♈	♉	8
♈	♒	3
♈	♋	7
♉	♒	8
♉	⌘	9
♒	♋	6

Origami Design A

1.
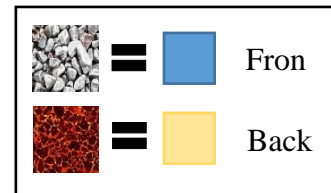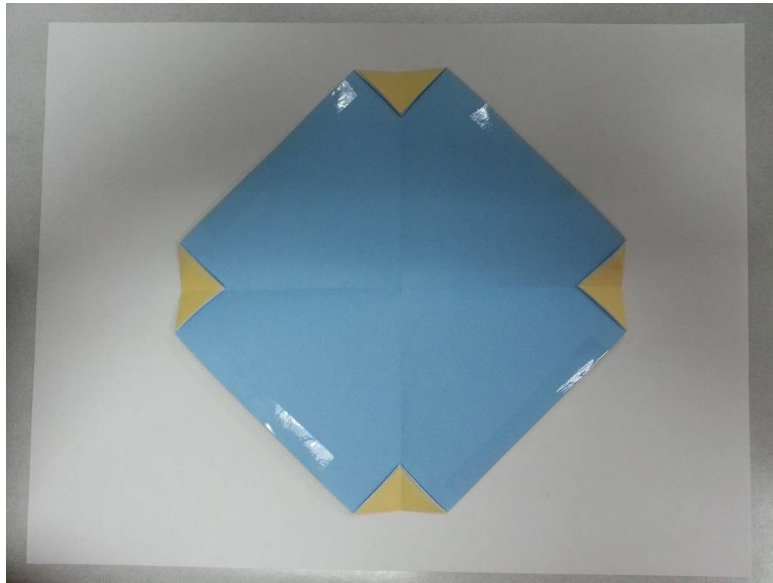
You are given the above.

These folds have been for you. Please proceed to the next step.

Legend

– – – – – –	Valley Fold
–·–·–·–·–	Mountain Fold
– – –►	Fold to
↻	Turn over

Example folds:

Mountain Fold   Valley Fold

🪨	= 🟦	Fron
🔴	= 🟨	Back
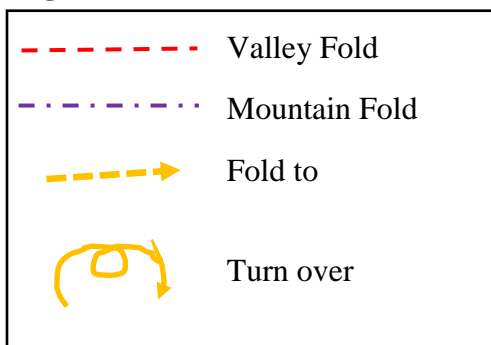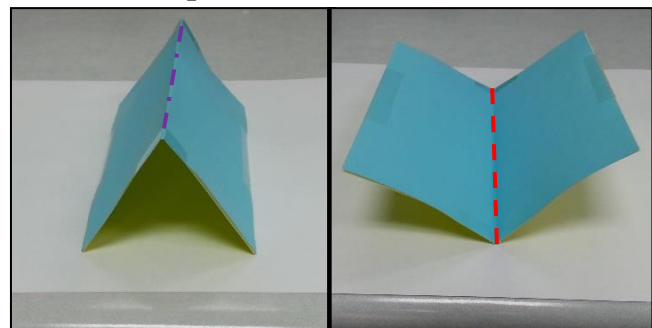
2.

.
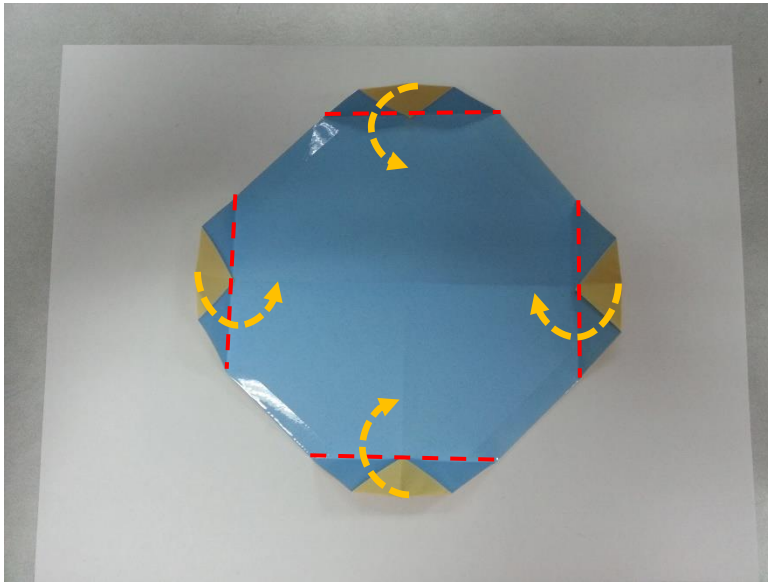
Fold along the lines, repeat on **all four** sides
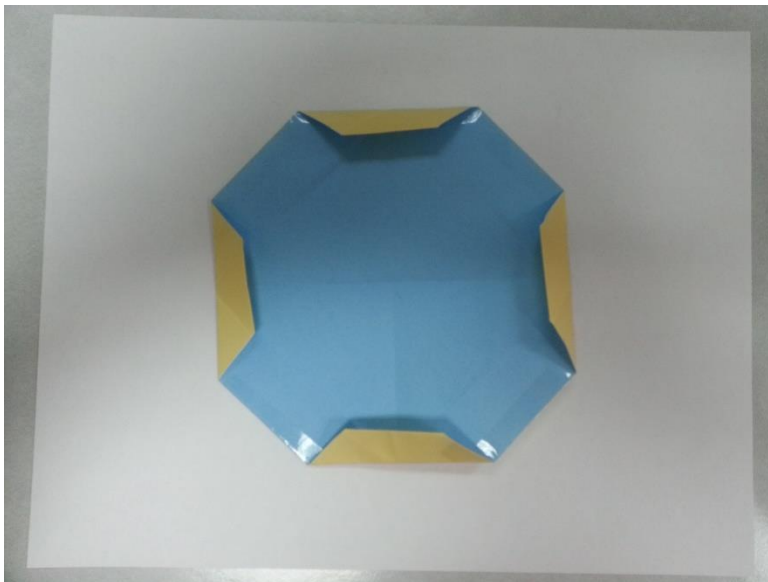
This is what you should have

3.
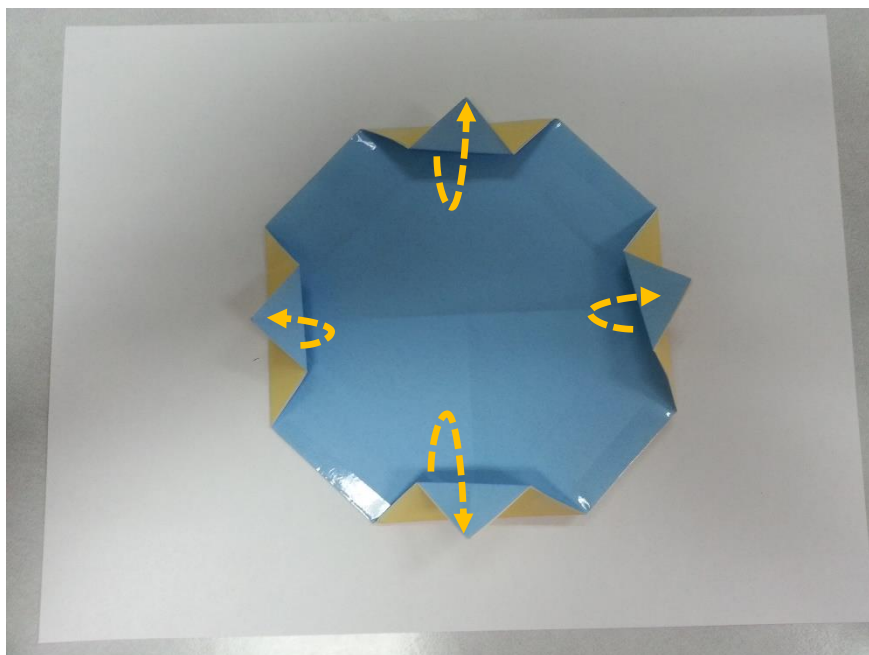
Pull out each smaller triangles and fold them up and out

Repeat on **all four** sides

4.
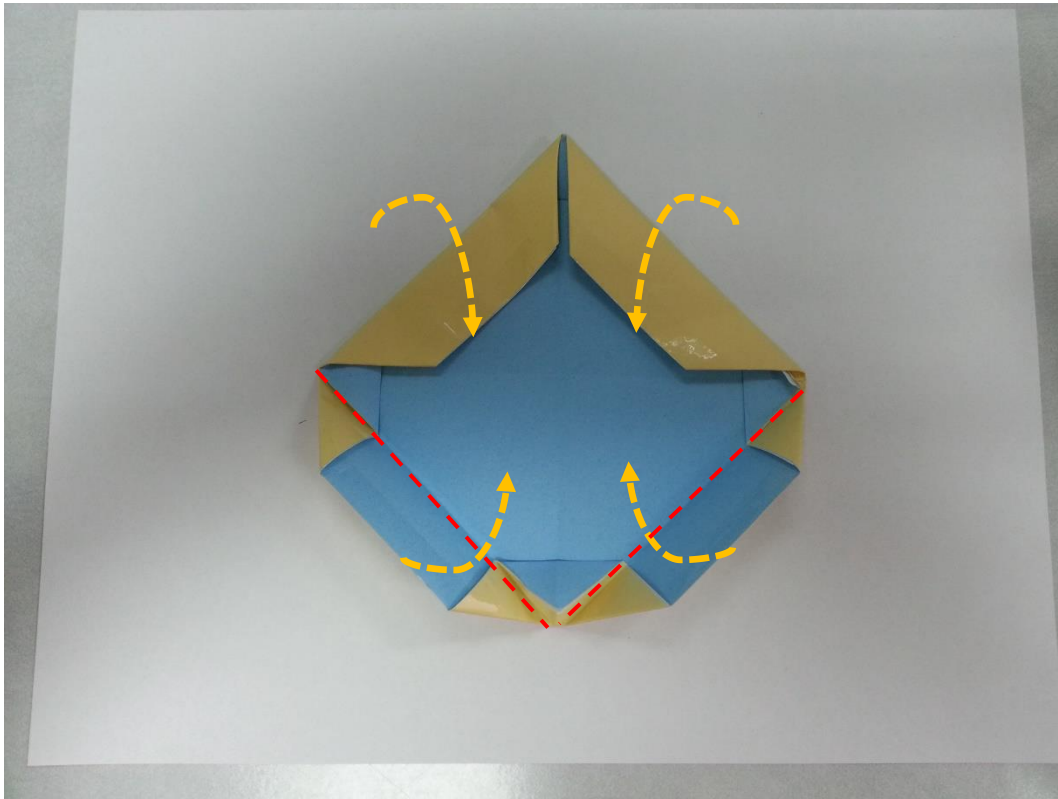
Fold along the lines, repeat on **all four** sides.

Notice that the folds are on the edges of the smaller triangles
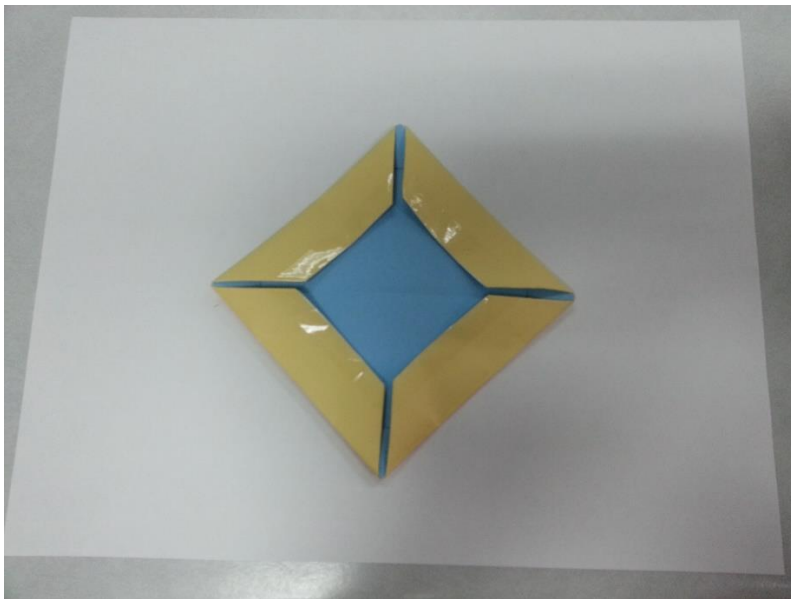
This is what it should look like after

5.
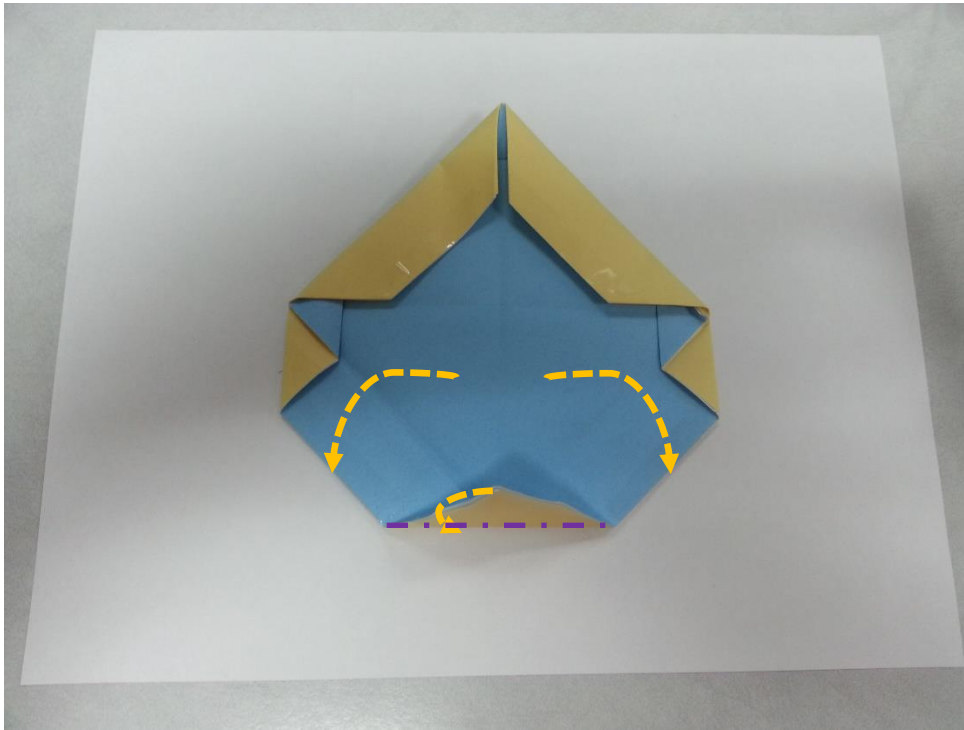
Now open any of the two edges and unfold the smaller triangle like so
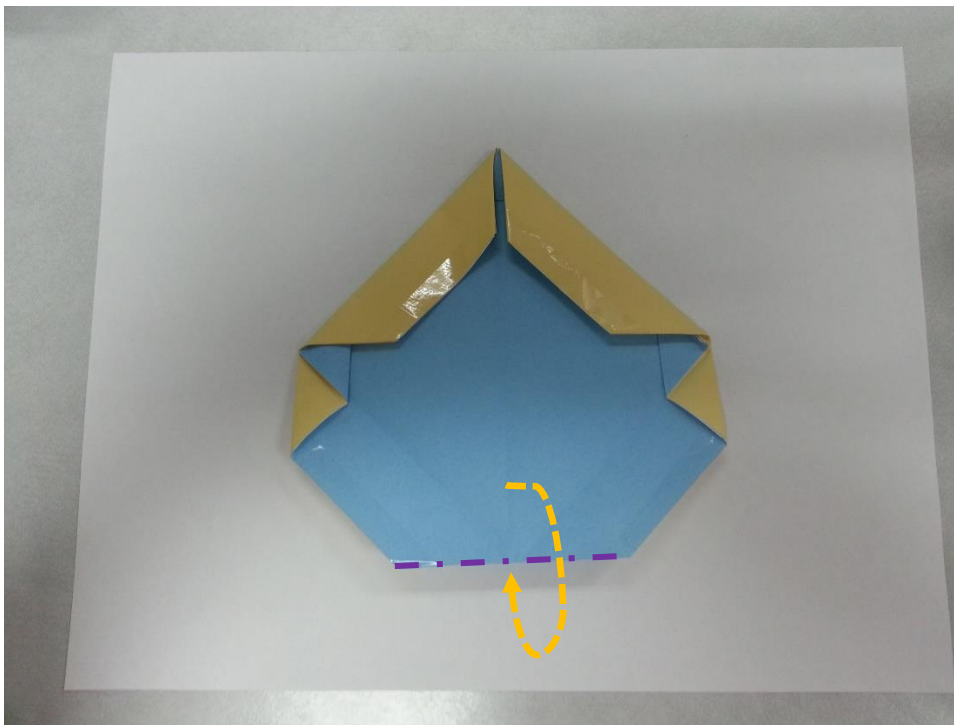
Then fold the flap under and behind the model

Notice that it is the longer edge being folded back

6.
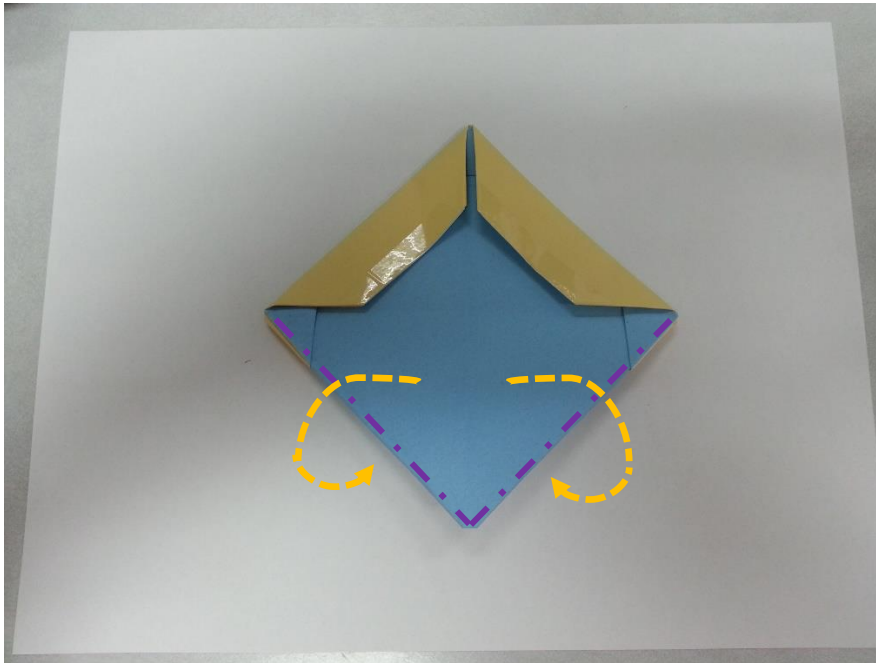
Now fold the two edges back behind the model

7.
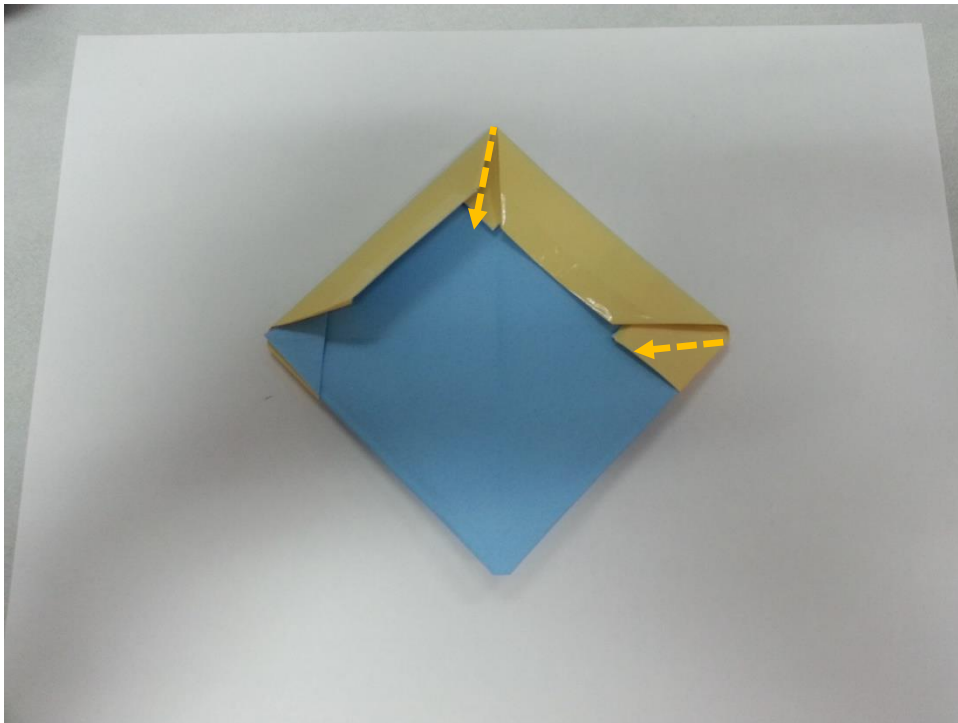
Pull out **only** the top and right triangles out

8.
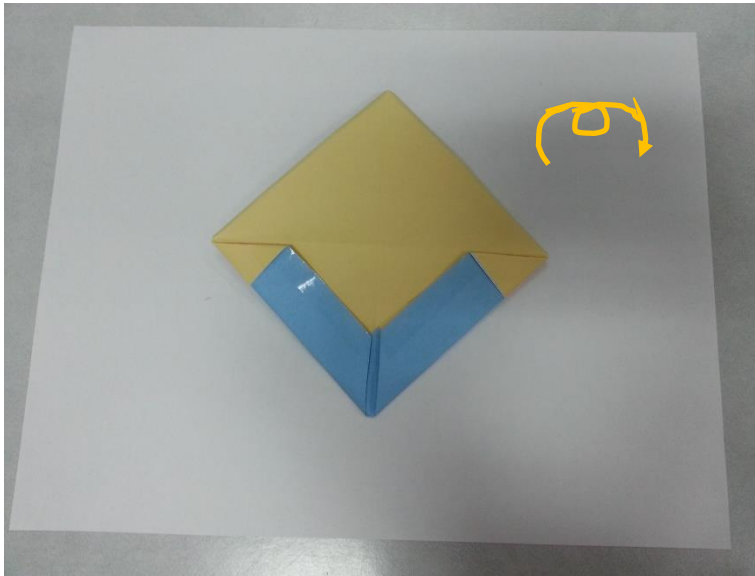
Turn over the model to see the back side

9.
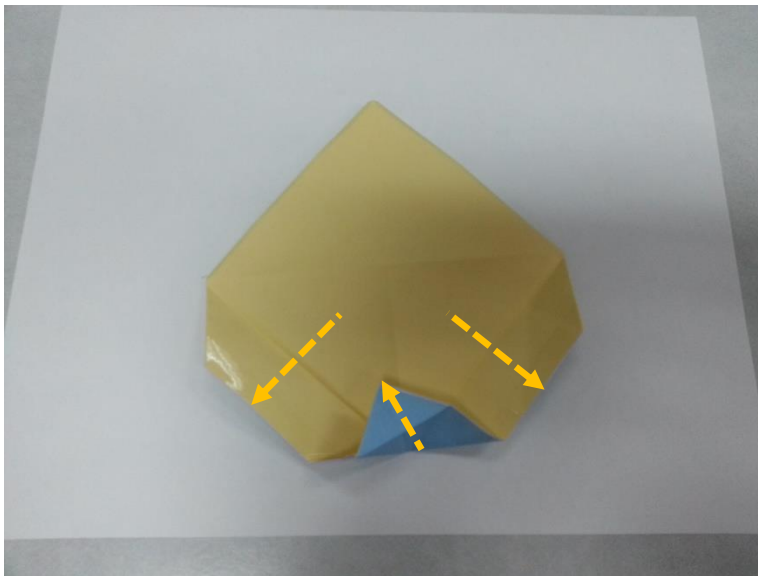
Unfold the two edges made earlier so that we can tuck the smaller triangle under
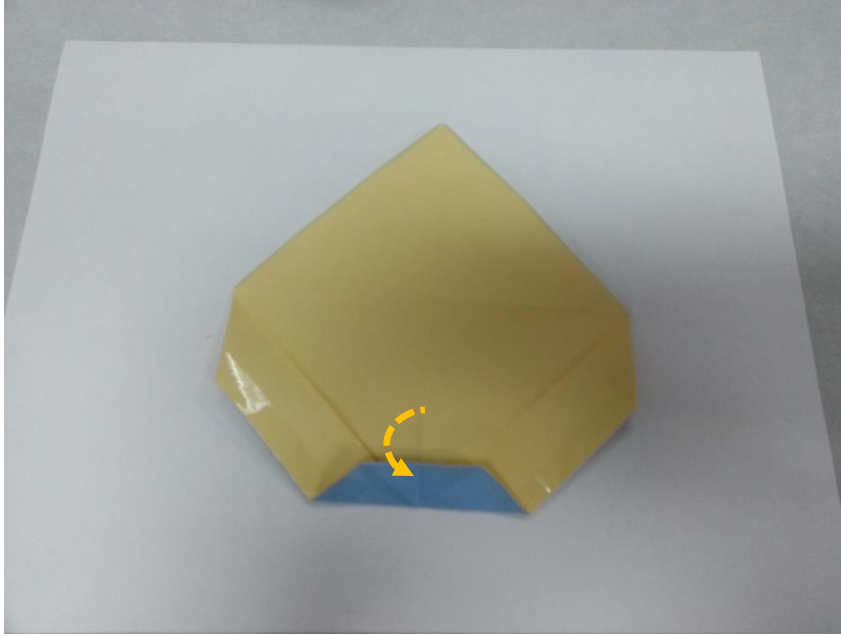
Tuck the small triangle under the flap like so
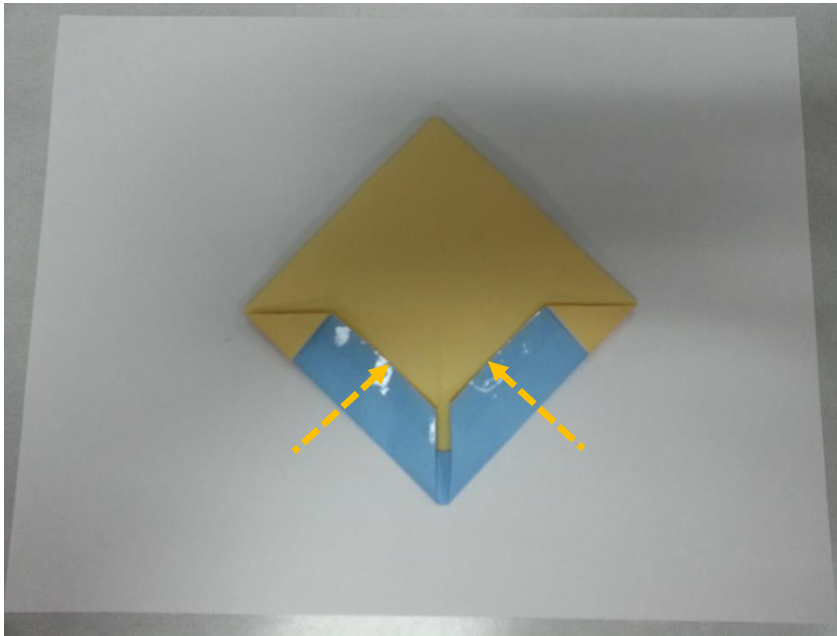
Fold the two edges back again

10.
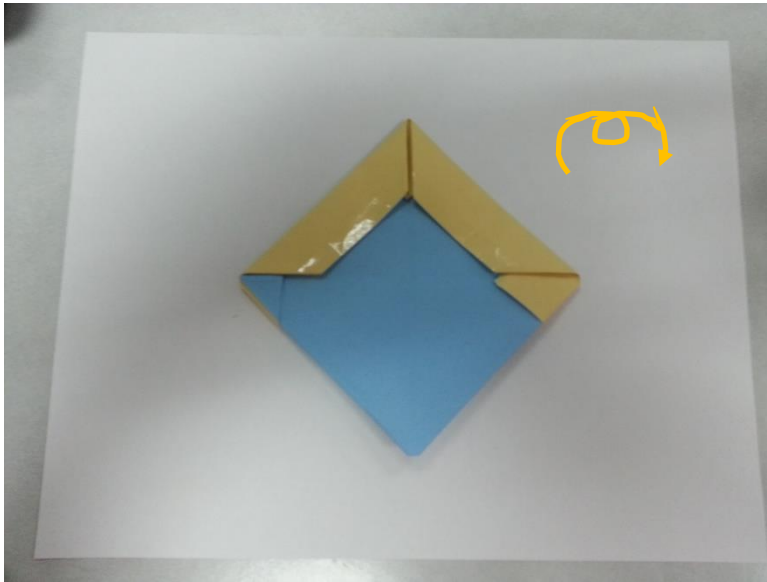
Turn the model over to see the front side

Complete!

Front:
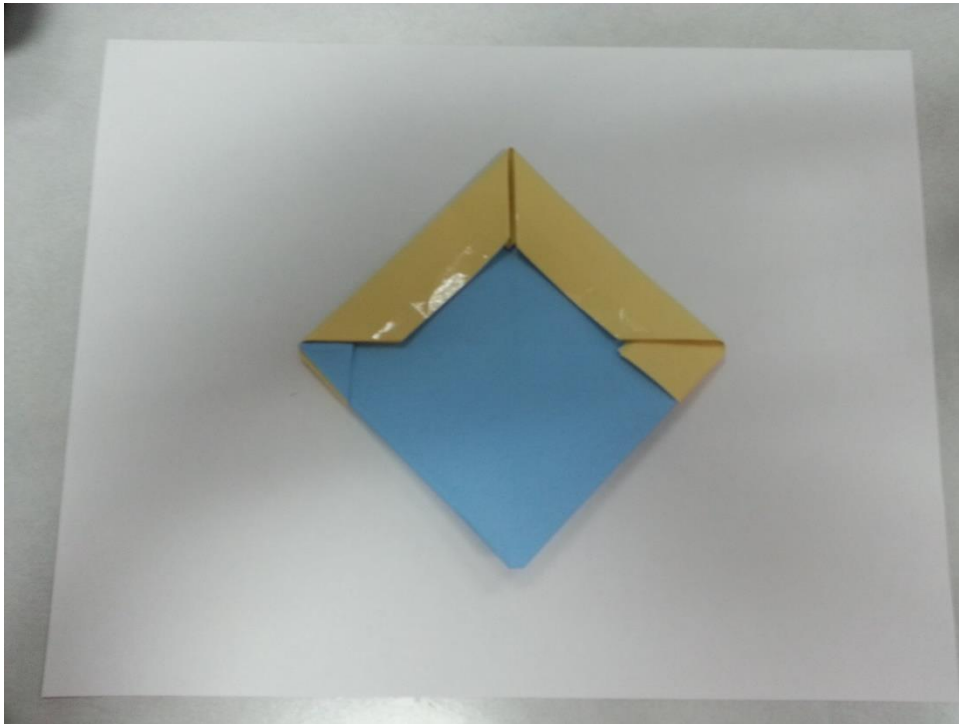
Back:
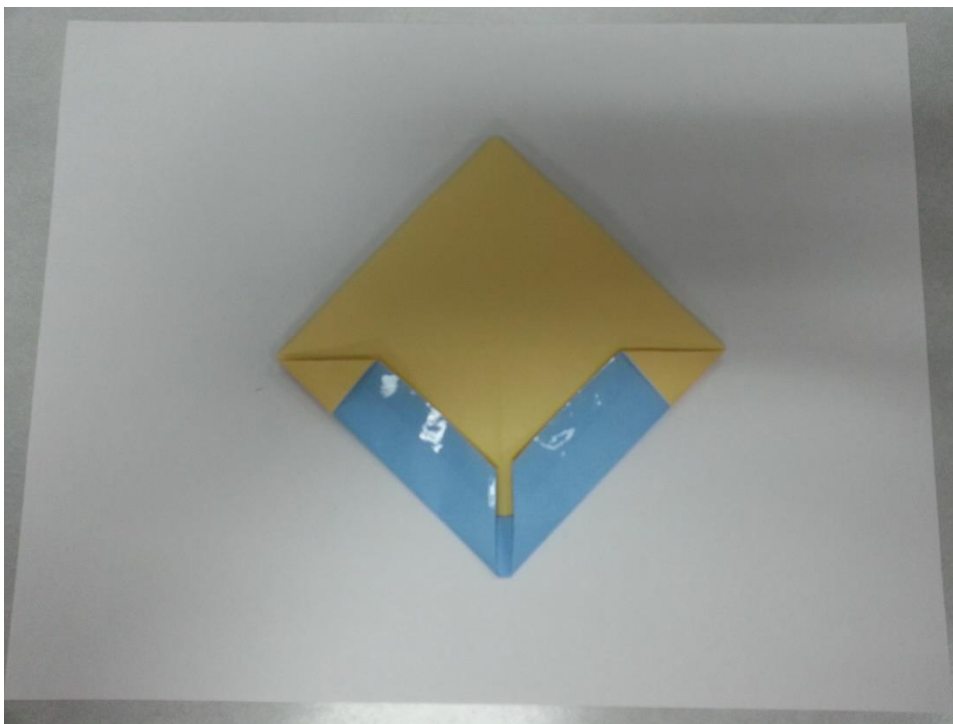
Origami Design B

1.

You are given the above.

These folds been <u>done</u> for you. Please proceed to the next step.

Example folds:

Mountain Fold   Valley Fold

2.
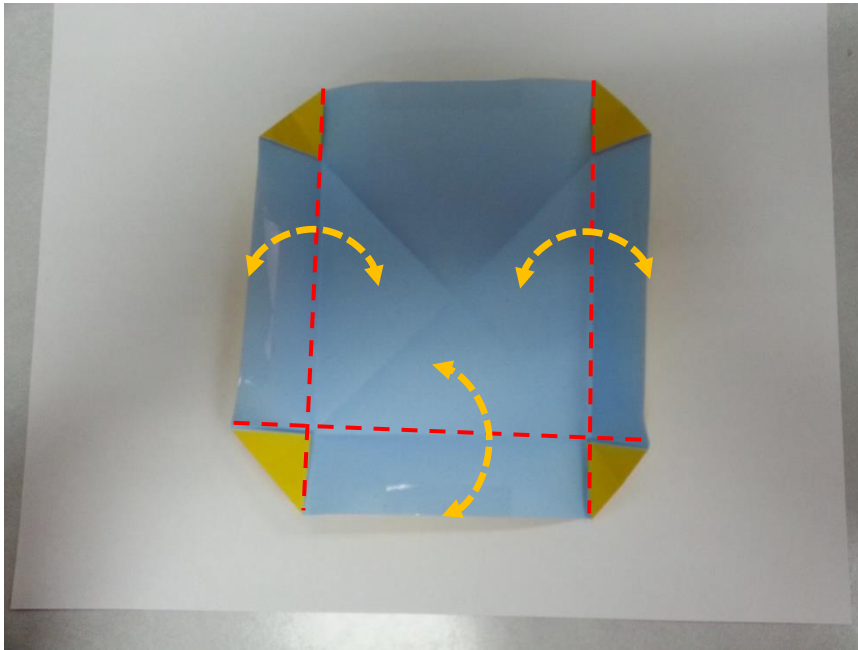
Turn the model so it faces like a square

Now fold **only** three of the edges as above, then **unfold** all three edges

Keep the **unfolded** edge at the top

3.
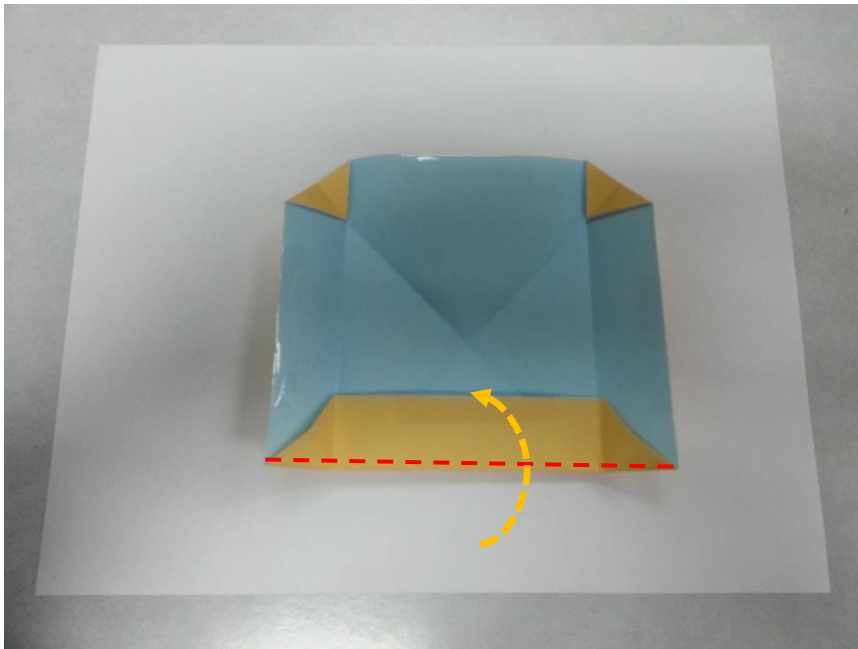
Fold up the bottom edge again on the same line

99

4.
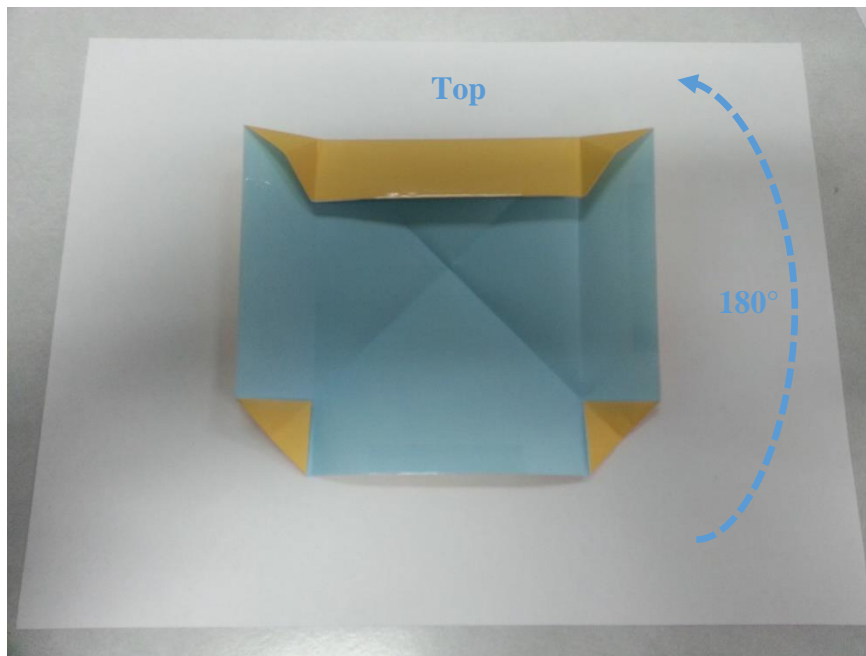
Turn the model around 180 degrees so that the **unfolded** side is on the bottom

5.
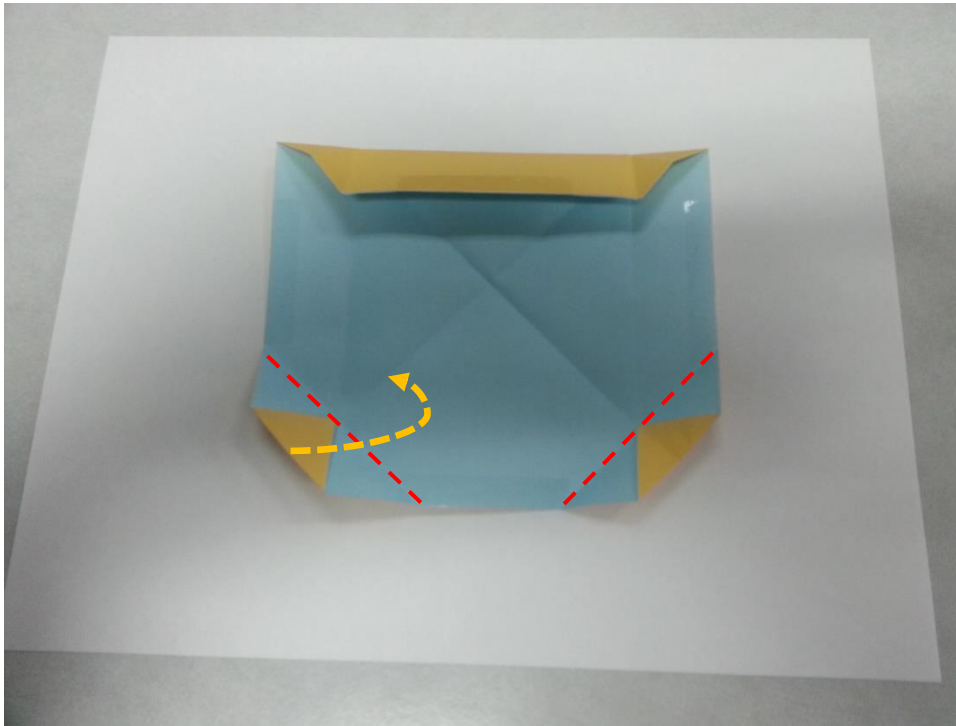
Fold the bottom two corners inwards
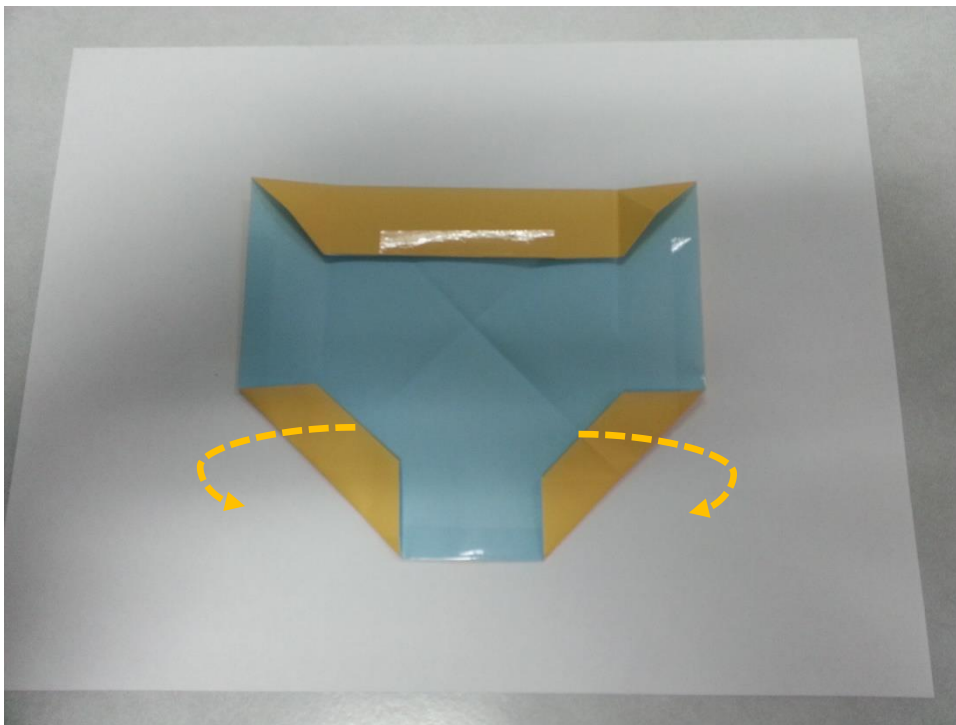
It should look like the above, **unfold** both folds
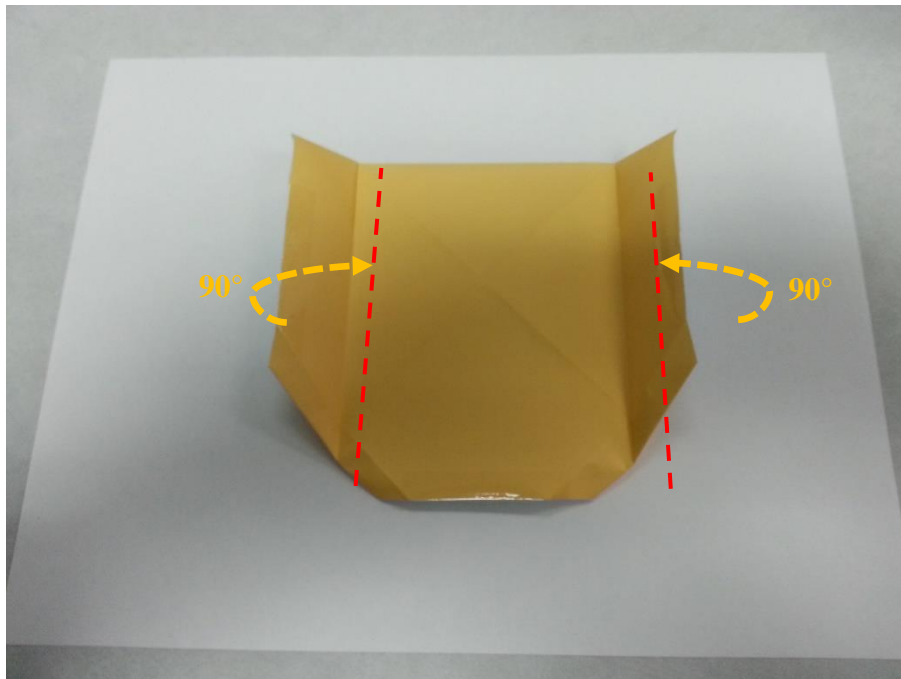
6.

Now fold the bottom corner halfway, meeting the line of the fold made earlier

7.

Turn the model over

8.

Fold the two sides up on the creases made in step 2

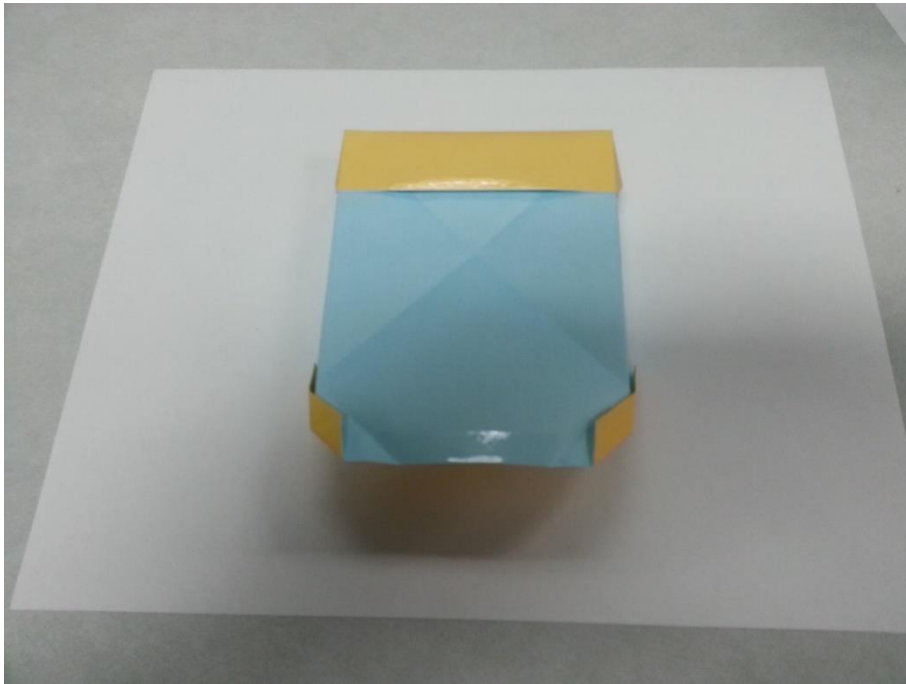These folds should be made to 90 degrees so that the model stands up when flipped over
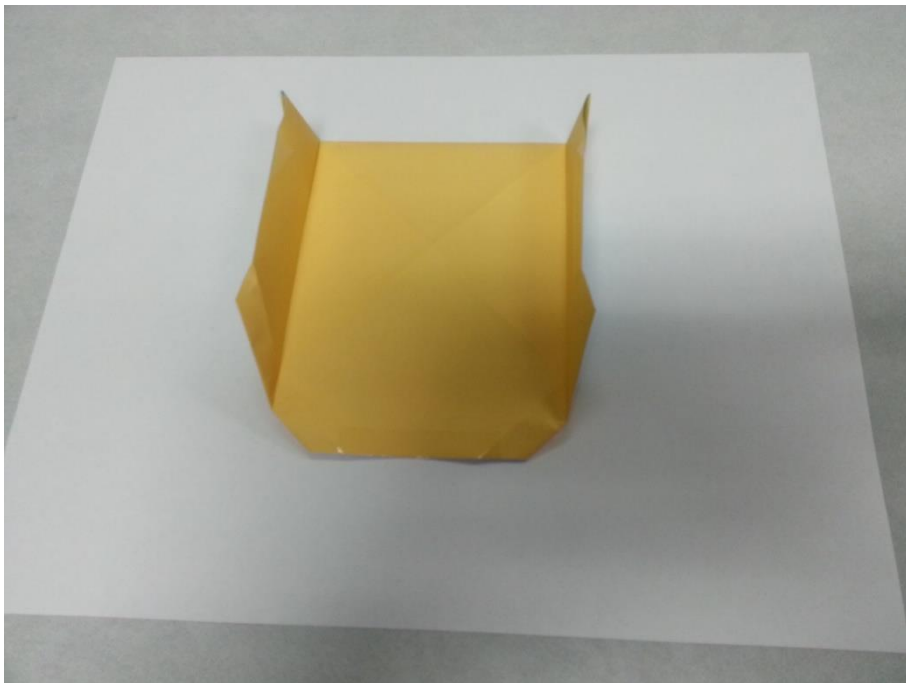
9.
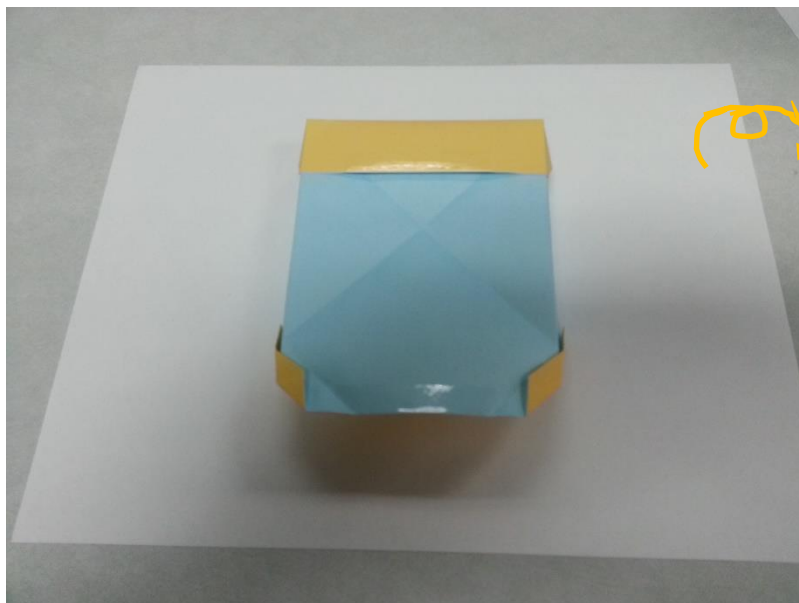
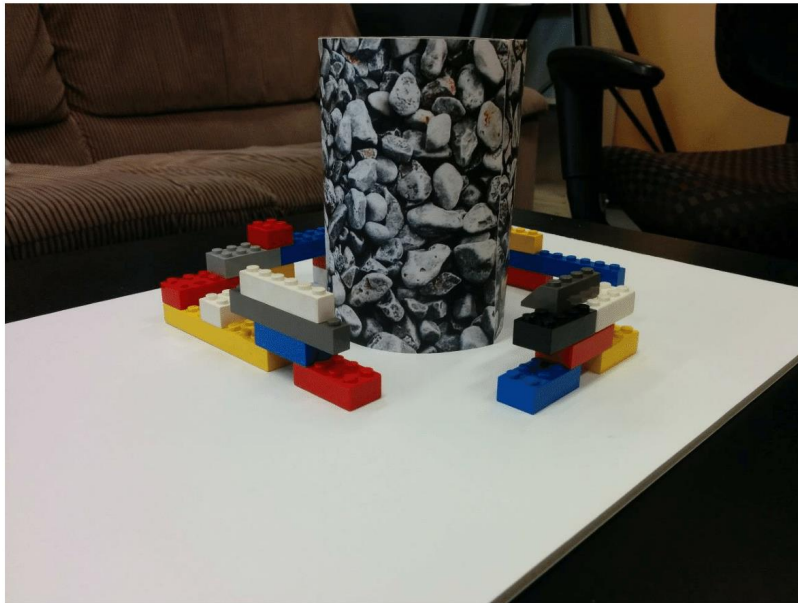Turn the model over, it should stand up
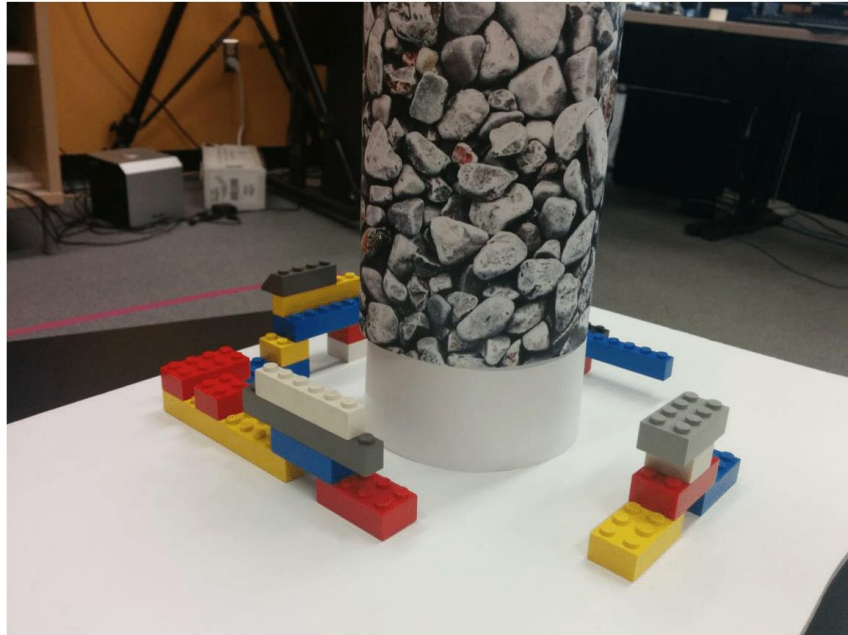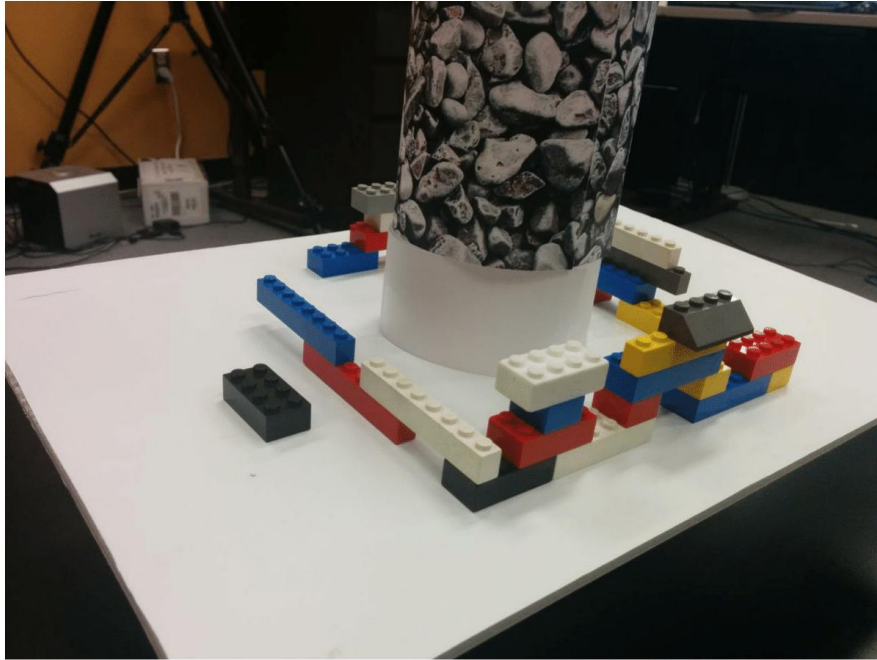
Complete!

Front

Back

A.6. Lego Repair Task Instructions

- The "BEFORE" photos show the structure given to the worker at the beginning. The "AFTER" photos show the final desired outcome. There were two variations of this task.
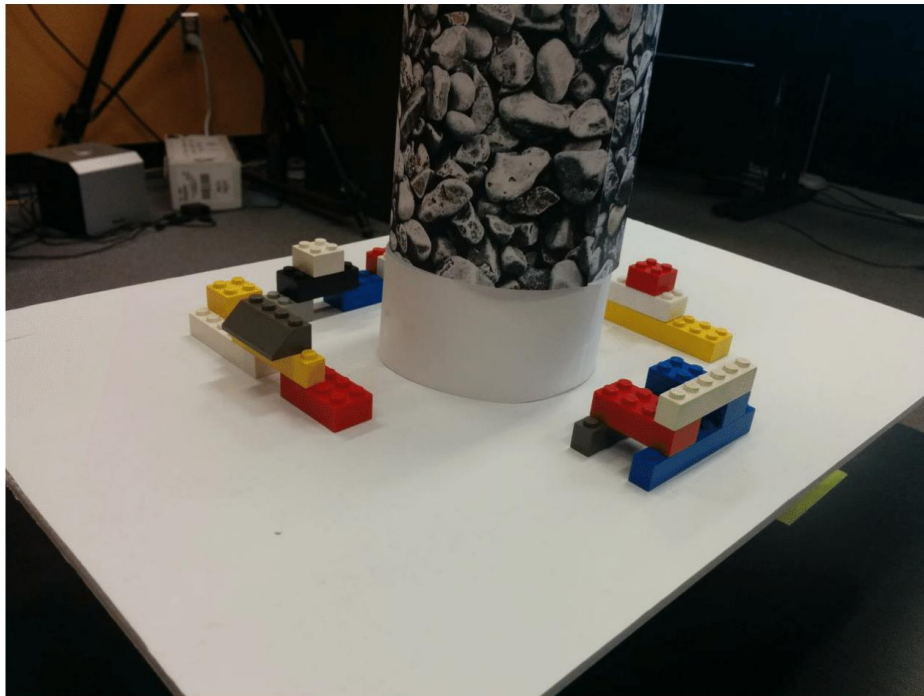
BEFORE 1

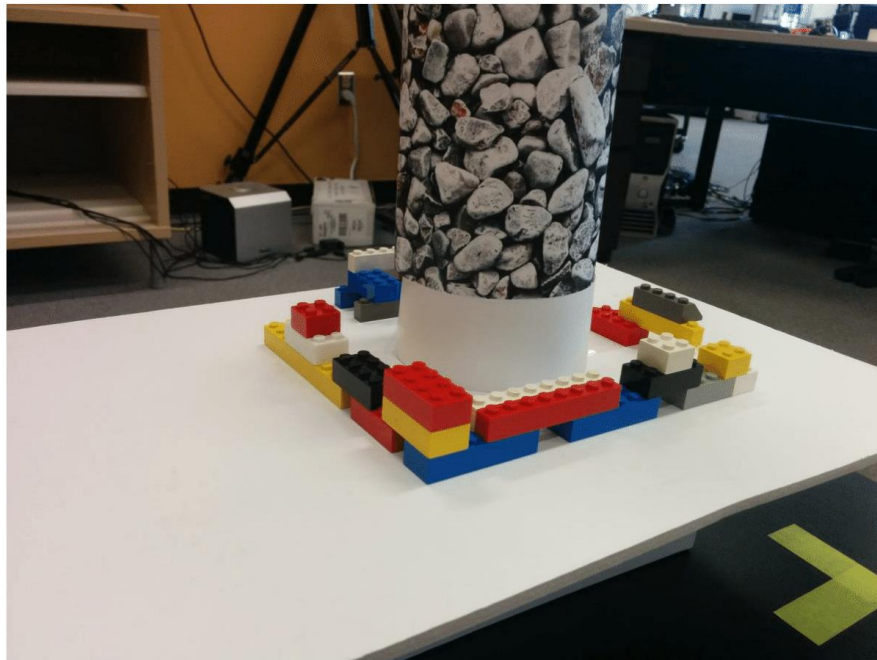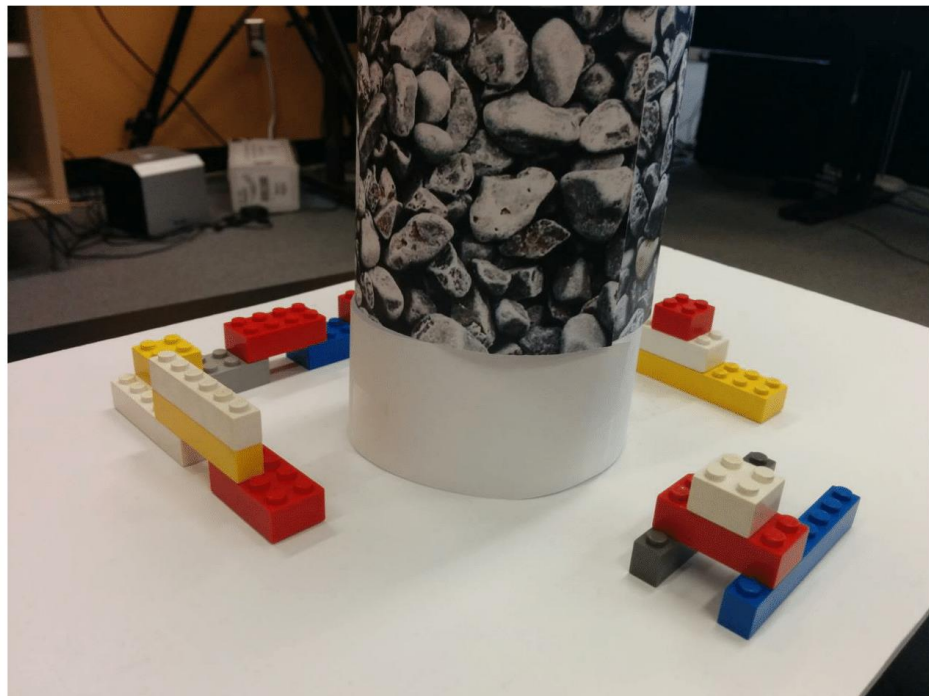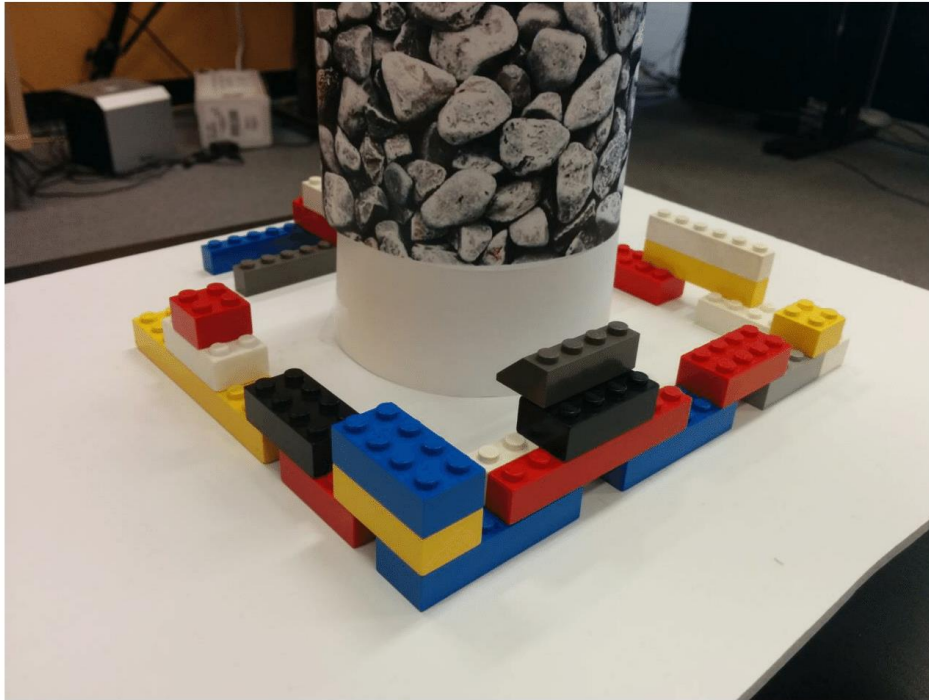## A.7. Post-task Interview Questions

Were the annotations helpful for this task? How could they be improved?

Did you run into any difficulties while sending/receiving assistance?

How was the device configuration? Did you feel comfortable with the HMD/tablet?

A.8. Post-study Questionnaire

POST-TEST QUESTIONNAIRE

Helper

Participant ID: _____

Dear participant,

We would like to thank you for taking the time to be involved in our study. Your comment and feedback will help inform our ongoing development of the systems you will be using.

Now, to conclude our study, we are going to ask you a few questions about your experience during the experiments. Please remember that we are evaluating these new technologies, not you personally.

Thanks for participating!

Personal Information

Age:

Gender:

Educational background:

Previous experience in working with mobile video conferencing systems:

None	Somewhat familiar	Familiar	Somewhat experienced	Experienced

Previous experience in working with augmented reality systems:

None	Somewhat familiar	Familiar	Somewhat experienced	Experienced

Do you have normal or corrected to normal vision? If not, please describe.

Do you have any color vision deficiencies (color blindness) that you know of? If yes, please describe.

Study-related Questions

1. How would you rate your overall experience in providing assistance with the wearable (head-mounted display) prototype? Please rate from 1 (bad) to 7 (excellent).

<div align="center">

1   2   3   4   5   6   7

</div>

2. How would you rate your overall experience in providing assistance using the mobile device (tablet)? Please rate from 1 (bad) to 7 (excellent).

<div align="center">

1   2   3   4   5   6   7

</div>

3. How would you rate your overall experience in providing assistance with stabilized annotations? Please rate from 1 (bad) to 7 (excellent).

<div align="center">

1   2   3   4   5   6   7

</div>

4. How would you rate your overall experience in providing assistance with non-stabilized annotations? Please rate from 1 (bad) to 7 (excellent).

<div align="center">

1   2   3   4   5   6   7

</div>

5. Based on your experience in using our technologies, which of the four conditions do you prefer for each task? Why?

Conditions
1 - Handheld device with stabilized annotations
2 - Handheld device with non-stabilized annotations
3 - Head-mounted display with stabilized annotations
4 - Head-mounted display with non-stabilized annotations

Graph task:

………
………
………
………
………
………………………………………………………………………………..

LEGO task:

………
………
………
………
………
………………………………………………………………………………..

Origami task:

……
……
……
……
……
………………………………………………………………………………..

Tangram task:

……
……
……
……
……
………………………………………………………………………………..

6. Please rank the efficiency of conditions to accomplish the tasks from 1 (most efficient) to 4 (least efficient).

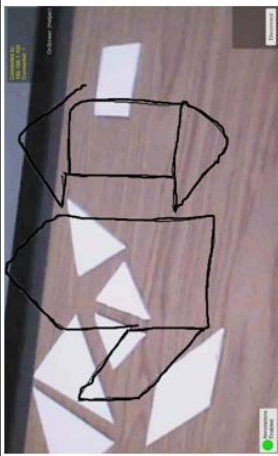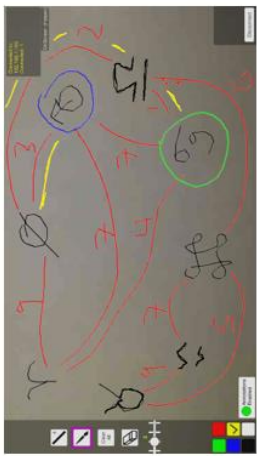| Condition	Rank
Handheld device with stabilized annotations	
Handheld device with non-stabilized annotations	
Head-mounted display with stabilized annotations	
Head-mounted display with non-stabilized annotations	

7. What other methods of giving/getting assistance do you suggest for future prototypes?

……
……
……
……
………………………………………

8. Any other comments:

……
……
……
……
…………………………………………………………..

Thank You!

A.9. Task Completion Times (in seconds)

Tangram

	HMD-S				HMD-NS				Tablet-S				Tablet-NS			
Pair No.	4	8	12	16	2	6	10	14	3	7	11	15	1	9	13	17
Time	343	420	420	411	133	318	150	419	330	420	130	281	392	188	420	303

Graph

	HMD-S				HMD-NS				Tablet-S				Tablet-NS			
Pair No.	3	7	11	15	1	9	13	17	4	8	12	16	2	6	10	14
Time	339	419	420	355	387	227	420	420	240	322	311	394	229	145	420	419

Origami

	HMD-S				HMD-NS				Tablet-S				Tablet-NS			
Pair No.	2	6	10	14	4	8	12	16	1	9	13	17	3	7	11	15
Time	420	419	420	420	234	420	420	420	420	316	419	365	420	420	319	341

LEGO

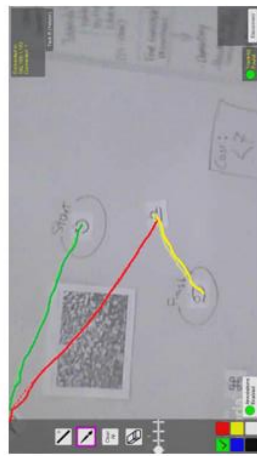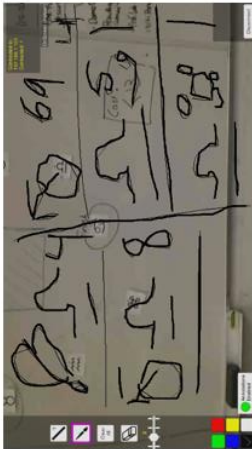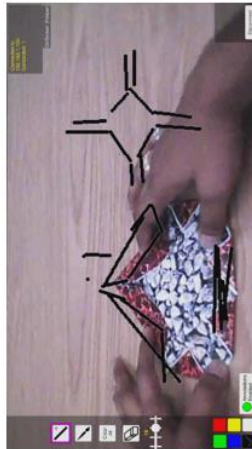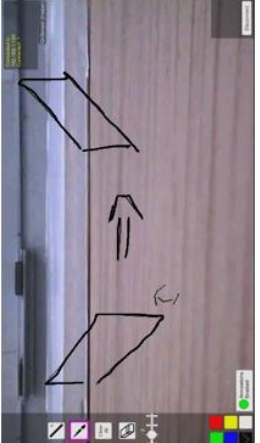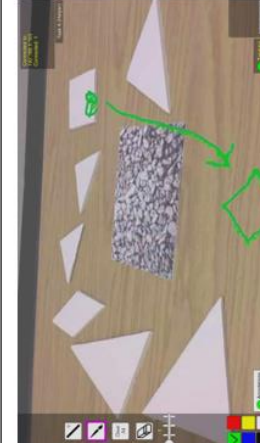|	HMD-S				HMD-NS				Tablet-S				Tablet-NS			
Pair No.	1	9	13	17	3	7	11	15	2	6	10	14	4	8	12	16
Time	420	265	420	420	420	420	293	301	256	229	209	263	229	323	300	420

HH: Handheld (tablet)

HMD: Head-mounted display

S: Stabilized annotations

NS: Non-stabilized annotations

Group #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Worker
Graph	HMD-S	HMD-S	HMD-S	HMD-S		HH-NS		HH-S	HMD-NS	HH-S	HH-S	HH-NS	HH-S	HH-S	HMD-S	HMD-S	HH-S	Graph
LEGO	HMD-S	HMD-S	HMD-NS	HMD-S		HMD-S		HH-S	HMD-S	HMD-S	HMD-S	HMD-NS	HMD-NS	HH-S	HH-S	HMD-S	HH-S	LEGO
Origami	HMD-S	HMD-S	HMD-NS	HMD-NS		HMD-NS		HH-S	HMD-S	HMD-S	HMD-NS	HMD-S	HH-S	HH-S	HH-S	HMD-S	HH-S	Origami
Tangram	HMD-S	HMD-S	HMD-NS	HMD-NS		HMD-S		HH-S	HMD-NS	HMD-NS	HMD-NS	HMD-S	HMD-S	HMD-S	HMD-S	HMD-S	HH-S	Tangram

Task-specific Preference

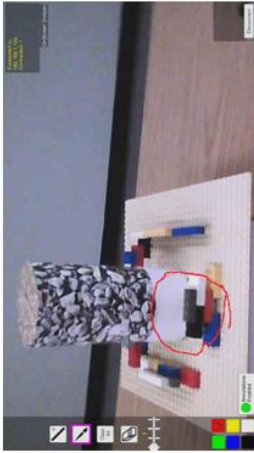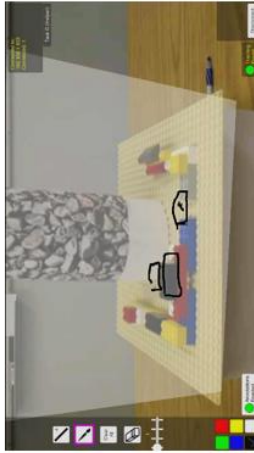## A.11. Examples of Different Annotation Types
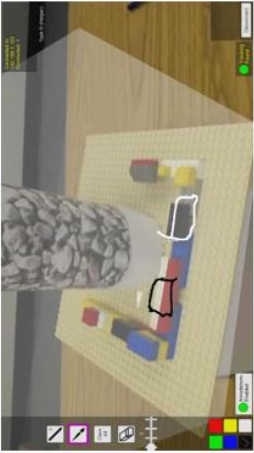
Reference Annotations	Description
	Tangram task Group 2 Non-stabilized annotations Head-mounted display Helper draws an outline of goal shape (an end-state reference annotation)
	Graph task Group 2 Non-stabilized annotations Tablet Helper copies his node symbols from his tablet screen, drawing his share of edges.
	Origami task Group 2 Stabilized annotations Head-mounted display Helper defines colors for different types of folds.
	Graph task Group 3 Stabilized annotations Tablet Helper draws her share of edges.

Reference Annotations	Description
	Origami task Group 3 Non-stabilized annotations Tablet Helper draws an end-state annotation to depict what the piece should look like at the end of a step.
	Graph task Group 10 Non-stabilized annotations Tablet Helper indicates his share of edges in a table-like structure.
	Origami task Group 12 Non-stabilized annotations Head-mounted display Helper draws an end-state annotation to depict what the piece should look like at the end of a step.
	Tangram task Group 15 Stabilized annotations Tablet Helper draws an outline of goal shape (an end-state reference annotation)

Procedural Annotations	Description
	Tangram task Group 2 Non-stabilized annotations Head-mounted display Helper indicates how to rotate a Tangram piece.
	Origami task Group 2 Stabilized annotations Head-mounted display Helper indicates how to fold the paper in the current step.
	Tangram task Group 3 Non-stabilized annotations Tablet Helper indicates where to move a Tangram piece.
	Origami task Group 3 Non-stabilized annotations Tablet Helper indicates how to fold the paper in the current step.

Procedural Annotations	Description
	Origami task Group 7 Non-stabilized annotations Tablet Helper indicates how to fold the paper in the current step.
	Origami task Group 10 Stabilized annotations Head-mounted display Helper indicates how to fold the paper in the current step.
	Origami task Group 15 Non-stabilized annotations Tablet Helper indicates how to fold the paper in the current step by drawing the fold lines.

118

Pointer Annotations	Description
	Lego task Group 2 Stabilized annotations Tablet Helper points towards where the Lego pieces should be placed.
	Lego task Group 3 Non-stabilized annotations Head-mounted display Helper points towards a group of Lego pieces.
	Lego task Group 10 Stabilized annotations Tablet Helper points towards where the Lego pieces should be placed.
	Lego task Group 14 Stabilized annotations Tablet Helper points towards the Lego pieces that should be taken off.

119