UNIVERSITY OF CALGARY

# Elevating Communication, Collaboration, and Shared Experiences between Peers in Mobile Video Communication using Drones

by

Brennan Jones

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN COMPUTER SCIENCE

CALGARY, ALBERTA

DECEMBER, 2016

# Abstract

People are increasingly using mobile video conferencing (e.g., Skype, FaceTime, Hangouts) to communicate, collaborate, and share experiences while on the go. Yet this presents challenges in adequately sharing camera views with remote users. In this thesis, I study the use of semi-autonomous drones for video conferencing, where an outdoor user (using a smartphone) is connected to a desktop user (e.g., who is at home or in an office) who can explore the environment from the drone's perspective. I describe findings from a study where pairs collaborated to complete shared navigation and search tasks. I illustrate the benefits of providing the desktop user with a view that is elevated, manipulable, and decoupled from the outdoor user. In addition, I articulate how participants overcame challenges in communicating environmental information and navigational cues, negotiated control of the view, and used the drone as a tool for sharing experiences. This provides a new way of thinking about mobile video conferencing, where cameras that are decoupled from both users play an integral role in communication, collaboration, and sharing experiences.

# Research Acknowledgements

Materials, ideas, and figures from this thesis have appeared previously in the following work:

Jones, B., Dillman, K., Tang, R., Tang, A., Sharlin, E., Oehlberg, L., Neustaedter, C., and Bateman, S. (2016). **Elevating Communication, Collaboration, and Shared Experiences in Mobile Video through Drones**. In *Proceedings of the ACM conference on Designing Interactive Systems 2016 (ACM DIS 2016)*, ACM, 1123-1135.

The work done for this thesis involves collaborations with other researchers, including fellow students Kody Dillman and Richard Tang who helped run the study and perform the data analysis, my supervisor Dr. Anthony Tang who gave me much-needed guidance through the completion of this work, as well as collaborators Dr. Ehud Sharlin, Dr. Lora Oehlberg, Dr. Carman Neustaedter, and Dr. Scott Bateman whose contributions helped make this work possible. While this is the case, this thesis serves as an account of my personal perspective of this collaborative work, and thus I am using the first-person singular pronoun (*I*, *my*, etc.) throughout this thesis in reference to the work completed for it.

# Acknowledgements

This thesis would not be possible without the help and support of a number of people.

First and foremost, I would like to thank my supervisor, Dr. Anthony Tang, for giving me much-needed guidance through my master's, for believing in me as a graduate student and researcher, and for helping me grow both professionally and personally. In addition, I would like to thank Dr. Ehud Sharlin, who provided me a drone in his graduate human-robot-interaction class and helped me pursue drone-supported video conferencing, which at the time was just a simple idea in my head. I would also like to thank my lab colleagues and friends, Kody Dillman and Richard Tang, who helped me test my ideas, run my study, and analyze my data. I would like to also thank Dr. Lora Oehlberg, Dr. Carman Neustaedter, and Dr. Scott Bateman, who I worked with to write the DIS paper, and with whom I've had various discussions that helped push my ideas forward. I would also like to thank Jennifer Payne and Dr. Jagoda Walny, who helped proofread my thesis and made it much more pleasant to read. Overall, I've had the privilege to work with a number of amazing people, including professors and other students, from both inside and outside the lab. Without their help, I certainly would not have been able to make it as far as I did.

I also want to thank my peers from the Interactions Lab: Kody Dillman, Richard Tang, Carrie Mah, Kevin Ta, Omid Fakourfar, Tiffany Wun, Jennifer Payne, Anna Witcraft, Sowmya Somanath, Teddy Seyed, Mandy Wong, Jiannan Li, Bon Adriel Aseniero, David Ledo, Lindsay MacDonald, Hesam Alizadeh, Dr. Xing-Dong Yang, and many many others. Your support and encouragement helped me throughout, and I am grateful that we had unique friendships that went beyond the lab—whether in the form of skating, hiking, cycling, potlucks, concerts, restaurant outings, hockey games, birthday parties, travelling, or something completely different. My favourite part about working in the iLab—the part that I always looked forward to every day—was the people who make up the lab.

Lastly, and above all, I would like to thank my family: my parents Cathy and Jeff, whose hard work, love, and great sacrifices helped me become successful and made me into the person I am today; and of course my siblings Becca and Cory, whom I was always able to talk to and who gave me much needed love and support during my studies.

# Dedication

*To my parents, Cathy and Jeff, and to my siblings, Becca and Cory.*

# Table of Contents

# List of Tables

# List of Figures and Illustrations

# List of Symbols, Abbreviations and Nomenclature

| Symbol | Definition |
|---|---|
| DP | Desktop Person, Desktop Participant, Desktop User |
| FP | Field Person, Field Participant, Field User |
| FOV | Field of view |
| UAV | Unmanned air vehicle |
| HMD | Head-mounted display |
| MRP | Mobile remote proxy |
| HCI | Human-Computer Interaction |
| HRI | Human-Robot Interaction |
| CSCW | Computer-Supported Cooperative Work |

# Chapter 1

# INTRODUCTION

## 1.1 Background and Motivation

Video communication (often referred to as *video conferencing*) is becoming increasingly popular with recent technological advancements (e.g., the Internet and webcams). There is an increasing desire to communicate, collaborate, and share experiences with distant others; and video communication is becoming the go-to medium to facilitate these kinds of interactions [1]. For instance, many people use video communication in the home and the workplace using tools such as Skype and Google Hangouts [1]. People use video communication to connect with friends and loved ones, collaborate with colleagues, and even share experiences (e.g., a baby's first steps) [1]. Video communication can help facilitate close interaction, collaboration, and shared experiences between two or more people separated by distance [1].

Video chatting while on the go (i.e. where parties are not confined to a desk or laptop) is also becoming more common, due to the popularity of portable devices like smartphones, and the growth of wireless networks. Such devices allow us to engage in on-the-go video

Figure 1.1: A mobile-video-conferencing scenario where one participant is out and about and the other is at a PC. This is the type of scenario that I focus on in this thesis. Image source: [21]

communication during activities such as travelling, shopping, hiking, etc. People are now engaging in mobile video chats while performing activities such as assisting with repair (e.g., helping someone change a flat tire) [1, 9, 10], giving a tour of a new place [1, 31, 35], giving navigational directions [21], going on a nature hike [69], sharing children's outdoor activities [19], shopping [21], and attending sporting events, concerts, and conventions.

Many of these activities involve 'bringing' a remote person into a local person's space. While the remote person is not being *physically* brought into the activity space, effective tools would allow them to actively engage in the activity or observe and experience the activity and the space.

A notable problem with today's commonly-used mobile-video-conferencing technologies (e.g., smartphones equipped with front and back cameras and basic video-conferencing software) is that they lack a means of providing the person outside of the activity space the ability to control his/her view into the space. The remote person's view of the local environment relies entirely on what the local person shows them and how they frame objects, landmarks, and people into view [21]. This can significantly hinder the remote person's ability to participate in the activity or feel like they are part of the activity [21]. Inability to directly control one's view into the remote environment can limit one's awareness of the spatial layout of the environment as well as the objects, landmarks, and people contained within it. This can, for example, hinder a remote person's ability to assist a local person with a searching task that is taking place in the space. Because the remote person's perspective is effectively controlled by the local person, view changes need to be communicated explicitly and verbally, putting undue burden on the local person and easily frustrating the remote participant.

In this thesis, I propose drone-supported video conferencing as a possible solution to this problem (illustrated in Figure 1.2). I present an early design space outlining a set of possible control strategies for drones (also referred to as unmanned air vehicles, or UAVs) in remote-

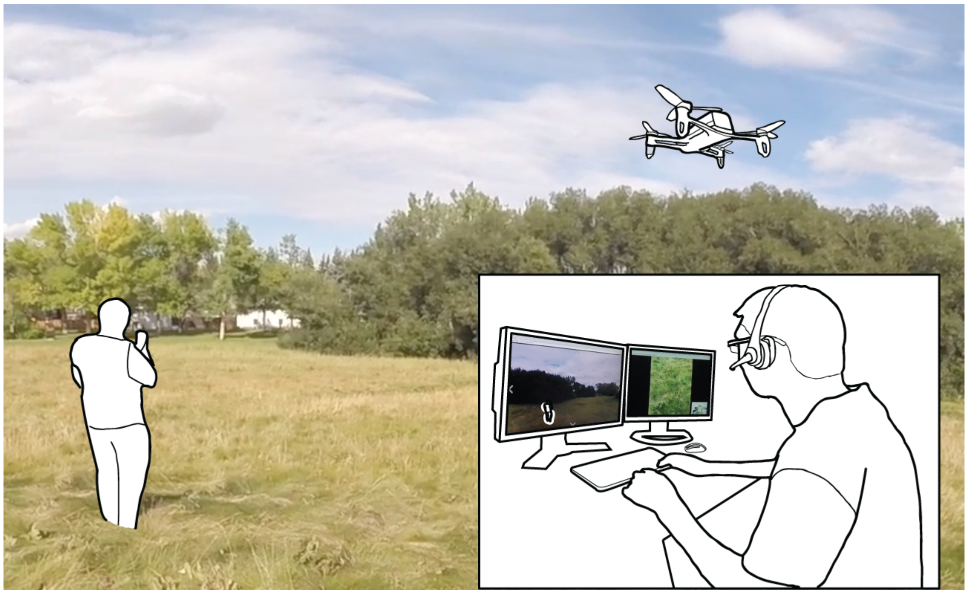

Figure 1.2: Two people collaborating remotely while communicating through a video-conferencing configuration involving a drone.

video-conferencing scenarios, and I present a prototype drone-video-conferencing system that adheres to one of the control strategies. I evaluate this approach by running a study to see how pairs of people communicate and collaborate with each other remotely through the drone-video-conferencing prototype system. The goals of this study are to explore and investigate the opportunities and challenges that drone-supported video conferencing brings. Based on this work, I outline several guidelines that could inform the design of future UAV-based video-conferencing systems.

## 1.2 Research Scope

The study of how individuals and groups of people communicate and interact through video chats falls within the field of Computer-Supported Cooperative Work (CSCW), which fall within the broader field of Human-Computer Interaction (HCI). HCI is a field that is concerned with how people use and interact with technology and with each other through the use of technology. HCI seeks to understand these things and to gain insight on how technology can be better designed to support people. CSCW is a field that is concerned with

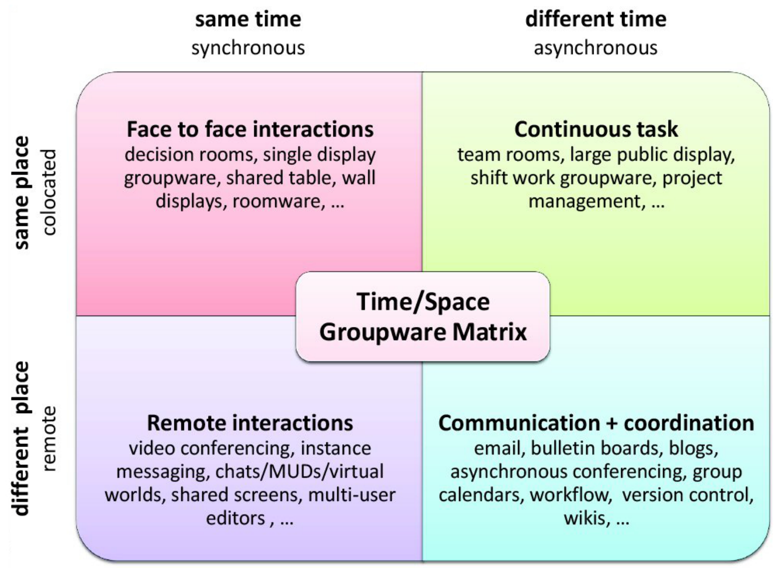


Figure 1.3: The CSCW Matrix.

how groups of people use technology to interact with each other and collaborate on work, social, and recreational activities. Johansen [20] introduced the *CSCW Matrix* (Figure 1.3), which conceptualizes contexts in which CSCW systems may be used. These contexts describe whether collaborators are collocated or not collocated, and whether they're working synchronously (i.e., at the same time) or asynchronously (not at the same time). Live video communication falls within the *Same Time/Different Place* collaboration context.

Live video communication has many different usage contexts, several of which have been explored in previous work and are practiced today. I will touch on these usage contexts in greater detail in Chapter 2. The most common usage context is two people–one on each end—communicating through live video, with each person in a static location (e.g., a desk or a couch). Another usage context, which is becoming more common, involves two people communicating through live video, with one person in a static location, and the other moving around some large environment (e.g., a field, a mall, a plaza, etc.). For this thesis, I focus solely on the latter usage context.

There are several video-conferencing activities that could fall into such scenarios. Many of these I cover in Chapter 2, and they include (but are not limited to): search and rescue;

site inspection; event set-up; guiding and navigating in large unknown environments; giving a tour; observing events such as sports, birthday parties, and festivals; and sharing experiences (such as nature hikes, kayaking, cycling, etc.).

## 1.3 Drone-Supported Mobile Video Conferencing

One of the primary issues with communicating through contemporary mobile-video-conferencing technologies while one communicator on the go is that the desk-bound person (DP) has very little control of their view into the activity space. Because the field person (FP) is constantly moving around his/her environment, the camera view that the FP is providing the DP is constantly changing. The camera is constantly moving—yet the DP has very little control of how it is moving. Information in the video frame is constantly changing at a rate that is too high for the DP to be able to effectively comprehend it [21]. The field of view (FOV) of the video frame is too small, and the movements of the camera are often so quick that the objects in the scene are blurred and the degree of movement is often not understood. This causes problems in terms of the DP being able to effectively understand the spatial layout of the activity space, as well as the objects, people, and landmarks contained within it. Activities such as searching, navigating (e.g., giving or taking navigational instructions), and touring (e.g., giving a tour or being taken on a tour) require all participants to have a strong sense of both the spatial layout of the activity environment as well as what is contained in it. Thus, it can be difficult for remote collaborators, such as the DP in our scenario, to assist with or take part in such activities through contemporary mobile-video-conferencing technologies.

Thus the question becomes: what if the remote person (in our case, the DP) has more control of their view into the activity space? There are many ways that this could be done—some of which have been explored in previous works (e.g., [33, 32, 57, 25, 24, 9, 10, 26]). For example, the remote person could control a camera attached to the local person [33, 32, 57], they could control their view through an omnidirectional camera attached to the local person [25, 24], they could control their view of a reconstructed representation of the activity space [9, 10, 26], or they could be given a better (perhaps more intuitive) means to request to the local person what direction to point the camera in [26]. Each of these approaches has their own benefits and drawbacks.

Drones, or unmanned air vehicles (UAVs) could also be used to provide the remote person with more control of their view into the activity space. This is the approach that I take in this thesis. Drones could potentially provide many benefits over the other proposed approaches. For example, they can offer views into the activity environment that are otherwise difficult or impossible to achieve from a ground perspective (e.g., bird's-eye view). They can also move around the space relatively easily and with many degrees of freedom—a quadcopter drone, which is the most common type of consumer-grade drone on the market today, can move freely along three different axes (up and down, left and right, and forward and backward), and in addition can rotate (yaw) in any direction. In addition to this, many drones often carry multiple cameras—one facing forward, and one facing directly down at the ground—as well as cameras with wide-angle lenses and cameras that can be rotated and zoomed both mechanically and virtually (e.g., virtual pan-tilt-zoom). Almost any angle that a camera operator would want to achieve in an activity space can be achieved by a drone. For this reason, drones are already commonly used by both professionals and hobbyists to capture events, activities, and sceneries for watching and sharing at a later time. Drone cameras can capture action from high above with a wider FOV; providing a better sense of the overall context of what is going on in the scene. Drones are commonly used today for purposes such as filmmaking, sports broadcasting, and journalism.

Because of their abilities to provide rich camera views containing a great deal of contextual information, their relatively low cost, and their versatility, drones have the potential to be a useful tool for on-the-go video conferencing. It is anticipated that drones could soon become a popular and viable alternative for experiencing live shared experiences with others. However, we currently know little about how people would make use of them and what challenges they might experience. This is needed in order to design the best possible user experience involving drones as resources for mobile video conferencing.

## 1.4 Research Goals

The overarching research question I address in this thesis is:

> **How can we design mobile-video-conferencing technologies that better support active back-and-forth collaboration between peers?**

The purpose of mobile-video-conferencing technologies is to allow people to communicate through video while on the go. Being on the move provides a whole new set of challenges that are not seen in traditional static (office-bound, home-bound, etc.) video-conferencing scenarios. As was mentioned earlier, the remote person often desires to be aware of what is going on in the activity environment, and thus desires to have a strong knowledge of the spatial layout of the environment as well as the things contained within it. It can be difficult for the remote person to gain this understanding if he/she does not have good control of their view into the activity space. The remote person could continuously send requests to the local person to direct the camera one way or another; but constantly having to position the camera to satisfy the needs of the remote person can be both mentally and physically cumbersome, and it can hinder both people's abilities to focus on the task at hand [21]. Giving the remote person greater control of their view into the activity space, without having to rely too much on the person who is in the activity space, is an obvious solution. The remote person could look around more freely and be able to take in a greater amount of visual information about the environment and the activity, giving him/her greater ability to contribute to the activity at hand—at the same time, the local person can focus less on trying to provide the remote person a good view of the scene and more on actually participating in the activity at hand.

Exploring this question provides the basis for the three thesis questions that I will answer later in my thesis:

**Thesis Question 1:** *How can we design mobile-video-conferencing technologies that provide a remote individual with fluid and flexible control over his/her view of the environment?* As described earlier, one of the principal challenges of mobile video conferencing with today's tools is that the remote viewer has little control over the view. This can be frustrating, as lack of view control makes it difficult for the remote communicator to actively contribute to the activity. This is especially true for activities that involve searching, navigating, and exploring large environments.

**Thesis Question 2:** *What are some of the opportunities presented by supporting mobile video conferencing using drones?* As mentioned earlier, drones are relatively inexpensive, versatile, and capable of providing rich camera views containing a great deal of overview and contextual information. They already provide numerous benefits with regards to video recording, and for this reason are commonly used by professionals and hobbyists to capture and film scenes, people, and activities for viewing at a later time. It is quite easy to see that they also have the potential to provide numerous benefits in the realm of live video streaming and video communication. In addition, a drone as a physical object in the activity space could also potentially act as a physical embodiment of the remote person in the activity space.

**Thesis Question 3:** *What are some of the interaction challenges presented by drone-supported video conferencing?* While drones have the potential to provide rich camera views and act as a physical embodiment of the remote person in the activity space, they can also be difficult to control. Total control could be given to either communicator, or the drone could act autonomously. If the drone acts autonomously, then it must be decided how the drone acts so that it benefits both communicators in the best way possible. This is not an easy problem to solve, and a solution that might be most beneficial for one pair of communicators working on one activity might not be that beneficial for another pair of communicators working on another activity. In this thesis, I explore several control strategies for drones in remote video communication, discuss benefits and drawbacks of each strategy, and propose an early design space for drone-based video-conferencing systems that outlines these strategies.

## 1.5 Contributions

This thesis provides the following contributions:

1. The introduction of an early design space for drone-based video-conferencing systems that outlines several control strategies for drones in out-and-about mobile-video-conferencing scenarios.

2. The design and implementation of a drone-video-conferencing prototype system that follows a particular control scheme in my early design space.

3. The first study of drone-supported video conferencing, where I outline new communication and interaction challenges that augment existing human-robot interaction (HRI) awareness and human-UAV awareness frameworks.

4. A public codebase upon which others can build drone-based video-conferencing setups.

## 1.6 Overview

This thesis is structured as follows:

Chapter 2 provides background on "on-the-go" mobile video conferencing, robotic telepresence, and interaction with drones.

Chapter 3 discusses the design of my drone-supported video conferencing prototype. In this chapter, I outline a set of design guidelines for drone-based video conferencing systems based on common challenges discussed in related work. Using this, I describe an early design space for position and orientation control in such a system. Finally, I describe the design and implementation of my prototype that I built to study the use of such a system.

Chapter 4 describes the observational study I performed with the drone-based video conferencing system that I built. Here, I discuss the study design, the study goals in relation to the research goals of this thesis, and the study findings.

Chapter 5 contains some discussion about the study results, the prototype system, and the design space. Here, I discuss the opportunities and challenges of drone-based mobile video conferencing, as well as the study limitations, scope, and future work.

Chapter 6 concludes this thesis, and discusses the overall contributions and implications of my work.

# Chapter 2

# BACKGROUND

In this chapter, I provide an overview of related work and discuss its relevance to my work. I frame my research from three perspectives: mobile video conferencing, robotic telepresence, and interaction with drones.

First, I discuss the body of work related to mobile video conferencing. Here, I provide an overview of different mobile-video-conferencing usage contexts. I also discuss the role of mobile video conferencing in supporting communication, collaboration, and shared experience activities; as well as several examples from the related work of such activities and technologies used to support them.

Next, I outline the robotic telepresence literature. This helps inform our understanding of people's experiences of embodiment when it is mediated through a robotic entity in a remote environment.

Finally, I discuss some of the current efforts on remote and collocated interaction with drones, and the existing frameworks for human-robot interaction (HRI) awareness and human-UAV awareness.

## 2.1 Mobile Video Conferencing

With the rapid evolution and adoption of mobile devices and cellular networks, and the increasing availability of video-conferencing software (e.g., Skype, Hangouts, FaceTime), mobile video conferencing is becoming increasingly commonplace. Compared to conventional desktop video conferencing, mobile video conferencing enables entirely new scenarios of use where one or more participants are 'out and about' physically moving around 'in the real world'. People use mobile video conferencing to communicate, collaborate, and share experiences with distant others while in the home or the workplace [1], while shopping [21, 27], while travelling [69], and even during major life events such as weddings and funerals [38]. In addition, mobile video chats can occur in a variety of contexts involving multiple devices and/or multiple users, and depending on the situation at hand, those involved in a mobile video call might be interested in seeing one thing or another.

#### 2.1.1 Usage Contexts

Mobile video conferencing can support a variety of different usage contexts. One of the most common contexts seen in day-to-day life is one-to-one video chatting—i.e., two people in separate locations communicating via a video call. Software applications such as Skype, Hangouts, and FaceTime support this. Another common context involves multiple (more than two) people communicating as part of a single video call. Often, this is done for work-related purposes (e.g., colleagues in separate offices taking part in a video conference call), but is sometimes done for non-work purposes (e.g., a camera person using Google Hangouts to take a small group of people on a tour [69]).

The context of multiple people communicating through a video call can also be divided into several subcontexts. For example, is each participant in a separate location (and thus using a separate device and in a separate view)? Are there multiple participants in the same location/view? Are all participants being captured on video; or are there some connected only via text or audio? Is there a 'main' view/location that everyone is interested in or focused on (e.g., the tour environment during a virtual photo walk activity [69]), or are there things of interest in more than one view/location (e.g., two people geocaching in separate locations while video chatting, where both are interested in what is in the other's environment [55, 54])?

There are other important factors to consider in both one-to-one and many-to-many video conferencing. These factors are largely dependent on both the context and activity at hand. For example, is one or more participant moving around some large environment, or are all participants stationary? What is the nature of the activity that is taking place (e.g., is it work-related, play-related, or something else)? What are participants doing (e.g., are they communicating, collaborating, or sharing experiences)? Are all participants actively engaged in the activity, or are some just observing? If there is a workspace or activity space (i.e., a space containing things of interest), what is the nature of that space (e.g., is it large or small, is it physical or virtual, is it in one location or spread out across many locations)? What constitutes a 'good view' into the activity space? What are participants interested in seeing? Aside from video, what other information do participants want to see (e.g., location data, a screen capture, the status or score of a game)?

These different usage contexts often require different types of tools and features to support them. To illustrate a simple example: if the activity involves a lot of movement around the space, a moveable camera is needed. Another example: workspace awareness is crucial to the success of a task [12]; so if the work/activity space is large, then it is beneficial to give participants the ability to look around and see things in the space as much as possible. The amount (and type) of information that the remote viewer sees as well as his/her abilities to control the view and interact with people (and perhaps objects) in the activity space can determine whether or not the remote viewer is just simply viewing or is also able to actively participate in the ongoing activity. For example, remote participants who are given a limited-FOV and uncontrollable view into the activity space (akin to conventional mobile video conferencing, with a smartphone camera operated by the local user) often resort to just passively observing the ongoing activity rather than contributing to it [21]. On the other hand, when remote participants are given much richer views into the activity space (e.g., a 360-degree live panoramic video, as in [25]) and they make an effort to actively look around the space, they can finish collaborative tasks much more quickly and efficiently [25]. Different requirements for 'good views' into the activity space require different tool requirements for providing such views. In addition, depending on the activity at hand, participants might find it useful to see the status of the activity (e.g., [27]). Activity statuses can be represented to the user in various ways depending on what type of information it is For example, if the group is searching or inspecting a large space, perhaps seeing a map of the space showing which areas were covered would be beneficial. If the viewer is watching a game, perhaps seeing the score would be beneficial. Activity statuses are often presented to players of video games in the forms of player scores, task progress (e.g., how many enemies are left), and maps (e.g., a map showing the player's location and the locations of allies, enemies, and items in the game room). This information is presented for good reason: it provides the player with important situational awareness about the activity at hand.

Another activity that is similar to (but not the same as) video conferencing is video streaming/broadcasting. This involves a single person (or group of people) streaming video to a pubic audience. In some cases the audience is small (just a few people), but in other cases the audience could be large (hundreds, thousands, or even millions of viewers). Live video broadcasting has traditionally taken place on TV, and events such as sports, news broadcasts, concerts, and celebrations are typically broadcasted on TV or streamed on the Internet. In recent years, non-professionals have taken up video streaming/broadcasting. People use services such as YouTube [71], Ustream [67], and Twitch [66] to broadcast live video streams from the home, the workplace, and sometimes while on the go. Tools such as Periscope [53] and Meerkat [40] make it easy for people to live-stream video to public audiences from a mobile device while on the go. Many of the challenges, opportunities, and techniques that exist in relation to video streaming also apply to video conferencing.

The focus of this thesis is on one-to-one video conferencing between two people in separate locations, where one person is 'on the go'. For this thesis, I present the design of a mobile-video-conferencing system that supports such activities, with the on-the-go person moving around a large outdoor space and the other person at a desk. For the remainder of this chapter, I discuss background work related to this usage context.

### 2.1.2 Activities

O'Hara et al. [47] ran a diary study investigating why people make mobile video calls and what challenges they face. They found that, of the video calls collected from their sample, there were three primary purposes for video calling: (1) to show things to talk about (i.e., communication), (2) to achieve a particular goal (i.e., collaboration), and (3) to keep in touch with "small talk" (i.e., sharing experiences). In the context of my work and for simplicity, I treat these as three separate activities; but one of these activities might support another—for example, communication can help support collaboration and shared experiences, and collaboration toward a shared goal could also play a role in supporting shared experiences.

#### Communication

People video chat often to show things to someone [47]. Such activities include showing things in the home or workplace [47], showing important life events (e.g., a wedding [38], a kid's first steps [1]), taking someone on a tour [69], and showing items at a store or restaurant [21, 27]. In addition, the person not in the activity space (the space containing the items/landmarks of interest) may want to communicate certain things (intents, details, object references, etc.) to the person in the activity space—however, with conventional mobile video conferencing tools, this can generally only be done verbally, and oftentimes the remote person's attempts are unsuccessful [21]. Researchers have studied various ways of providing the remote person with a means of referencing things and conveying intention—for example, through on-screen annotations [6, 8, 9, 10, 13, 30] (e.g., Figure 2.1), showing hand gestures [29], and telepointing [26].

Fussell et al. [8] ran a study exploring the benefits of remote gesturing through drawing in supporting collaboration on physical tasks. Study participants worked in pairs, with one participant playing the role of the 'worker' and the other playing the role of the 'helper'. The worker had to perform three robot-assembly tasks, while the helper had to guide the worker through the instructions in completing the task. The worker was located in the room

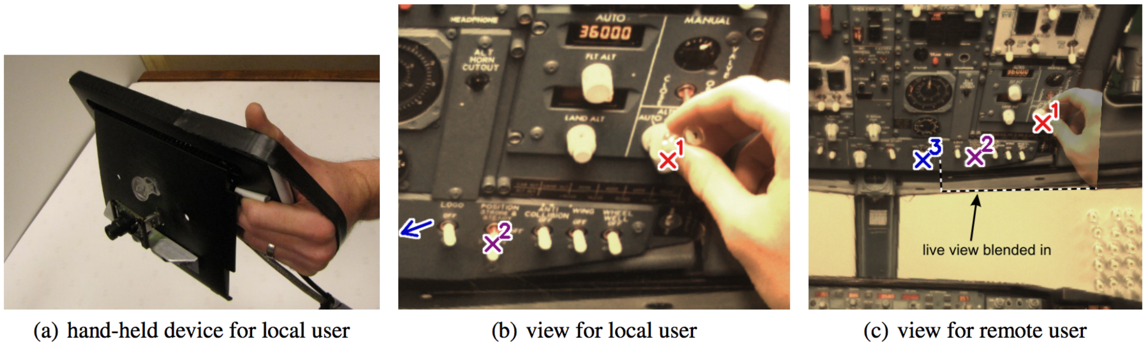


Figure 2.1: A remote user places world-stabilized annotations over objects on the video feed. These annotations are visible through the local user's device. Image source: Gauglitz et al. [9]

of the workspace while the helper was located in another room. They communicated to each other through a video-conferencing system with a camera fixed pointing down on the workspace. Pairs worked under three conditions: video-only, video with drawing, and video with drawing and manual erase. The researchers found that pairs performed better in both of the drawing conditions than they did in the video-only condition. The researchers also made interesting observations with regards to the types of drawings observed.

Kirk et al. [29] explored remote gesturing through the projection of each collaborator's hands onto the other's desk space (Figure 2.2). The purpose of this study was mainly to observe what types of hand gestures collaborators made to each other, and how those hand gestures supported collaboration on the given task. The researchers observed a series of common, yet very expressive hand gestures made by collaborators over this video conferencing system.

Additional challenges complicate remote referencing of objects in the context of a mobile video chat as opposed to desktop video conferencing, where the camera view is static. For one, a moving camera leads to a less consistent view, and objects that were in the frame at one moment might no longer be within view the next moment. Movement can be quite quick too. Because of quick camera and object movements, video images can also be quite blurry, and this can cause difficulty and frustration for the remote participant [21]. Environments

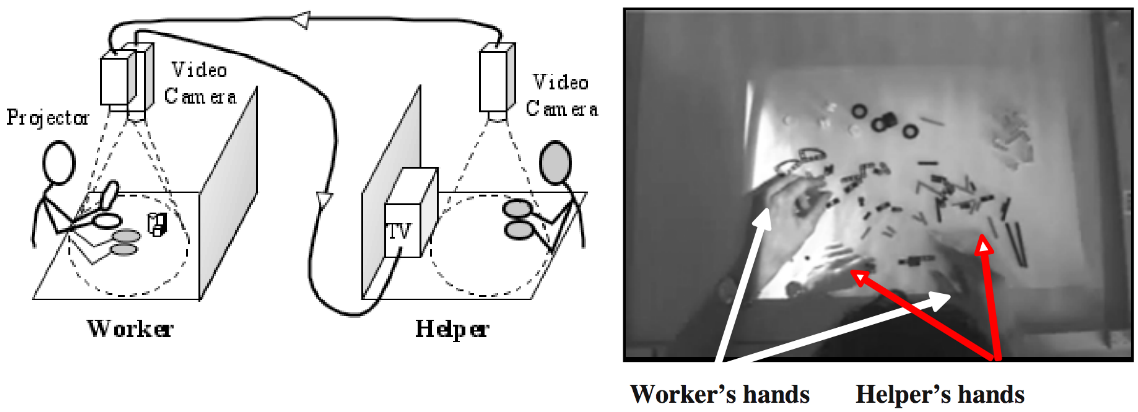


Figure 2.2: A prototype video conferencing system that projects a remote person's hands over the local person's workspace. Source: Kirk et al. [29]

can also be quite dynamic, with a lot going on, making it difficult to keep track of objects to draw attention to [21]. In addition, participants operating the mobile device may be hesitant to angle the camera certain ways in public spaces due to embarrassment or concern about drawing unwanted attention [21].

Researchers have explored means to tackle some of the challenges of referencing objects in an activity space where the camera view and objects in the space are consistently moving. Gauglitz et al. [9, 10] and Fakourfar et al. [6] explore world-stabilized annotations—annotations that remain in the same position relative to the workspace (e.g., an annotation that is placed on a static object stays on that object regardless of how the camera moves). Kasahara and Rekimoto [26] and Fakourfar et al. [6] explore displaying annotations through transparent head-mounted displays (HMDs). Kim et al. [28] and Fakourfar et al. [6] explore temporarily freezing the video frame while an annotation is being drawn, in order to keep the object(s) of interest in the picture.

To summarize, the literature has discussed the following common problems with regards to communication in mobile video conferencing:

- It can be difficult and frustrating for the remote user to reference objects in the activity space [21]. Proposed solutions include on-screen annotations [6, 8, 9, 10, 13, 30], projections of remote collaborators' hands [29], and telepointing [26].

- A moving camera or dynamic environment cause additional difficulty to the remote user in terms of understanding the environment and referencing things in the space [21]. Movement can cause a blurry video; and if objects pass through the frame too quickly, the remote user will not have sufficient time to understand what is going on, to analyze certain things in the scene, or to direct attention towards things in the scene. Proposed solutions include world-stabilized annotations [6, 9, 10] and freezing the video frame while focusing or directing focus on something in the scene [6, 28].

- Participants communicating via mobile video in public spaces may avoid angling the camera in certain ways due to embarrassment or concern about drawing unwanted attention [21]. Proposed solutions include the use of transparent HMDs [6, 26] and other wearables (e.g., [25]).

Collaboration

People also often use mobile video conferencing to work on collaborative tasks [47]. Collaboration simply involves two or more people working together to achieve a shared goal. Collaborative activities are often work-related, but can also be play-related. Example scenarios include assisting someone with a physical task (such as maintenance, construction, or repair) [9, 10, 26, 6, 21], searching for something together (e.g., in a retail outlet, on a diagram, or in a large outdoor space) [21], and making decisions together (e.g., shopping together and deciding what to buy) [21].

Communication helps support collaboration, and so providing good means of communication is beneficial to pairs working on collaborative tasks. Many of the solutions discussed previously (e.g., remote gesturing through annotations or hand gestures) have been proven to help collaborators convey intention, and thus improve awareness (for both collaborators) of what each person is focusing on, what everyone should be focusing on, and what the next step in the task is.

Figure 2.3: A camera, worn on the local user's shoulder, that can be mechanically controlled by the remote user. Source: Kratz et al. [32]

Aside from having sufficient means to communicate, collaborators should also have sufficient means to understand the work/activity space. If the remote collaborator does not have a sufficient view into the activity space (which is often the case, especially for conventional tools that provide a single view with limited FOV), he/she will not get a full understanding of the space and the status of the activity; and thus they may lose out on opportunities to contribute to the activity at hand [21]. Providing the remote user with greater workspace awareness via a richer view and more freedom to control their view can be beneficial in collaborative tasks [12]. If the remote user can see and look around more, they will be more easily able to find things that the local collaborator might not see. This would give the remote collaborator more opportunities to interject. In addition, it could allow the collaborators to complete the task much more quickly and efficiently. Researchers have attempted to address this by, for example, providing the remote person with a higher-FOV view into the activity space [25, 24], by providing the local user with a shoulder-worn mechanical camera that can be operated by the remote user [32, 33] (e.g., Figure 2.3), by providing the remote user with a large composite view into the activity space constructed by stitching sequences of images of the space taken over time [26], and by providing the remote user with a continuously-updating 3D reconstruction of the activity space [10].

Gauglitz et al. [10] developed a mobile video conferencing tool (Figure 2.4) that creates a

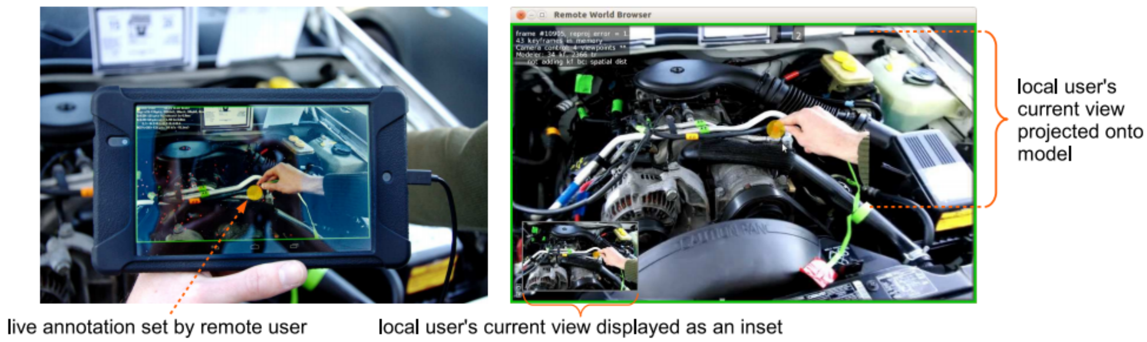


Figure 2.4: A mobile-video-conferencing system that creates a 3D model of the activity space and allows the remote user to place world-stabilized annotations, which are then viewable through the local user's device, overtop of the 3D model. Source: Gauglitz et al. [10]

3D reconstruction of the local task space in real-time and allows a remote user to add world-stabilized annotations to the 3D scene. The 3D reconstruction is created automatically with the image frames captured by the local tablet through computer vision techniques. The annotations are then displayed on the local user's tablet screen through augmented reality as the camera passes over the points where they are placed.

Kasahara and Rekimoto [26] designed a mobile video conferencing system, called "JackIn" (Figure 2.5), that provides the remote person (called the "Ghost") with the first-person camera view from the person in the activity space (called the "Body"). The Ghost is provided with a composite of images stitched together with video frames from the camera worn by the Body. The Ghost sees this composite of images on a large wall-sized display This composite view is updated live, and it provides a way for the Ghost to look around the space and view what was seen earlier (via the "out-of-body" view) while also knowing where the Body is currently looking (via the live video frame). The idea is to provide a "shared" common view, while also allowing the Ghost to look around independently (to some extent) in the Body's environment. In addition, the Ghost can gesture naturally (i.e., point) at the wall display while the system recognizes those gestures and send them to the Body user. These gestures can be viewed as annotations through a transparent HMD that the Body wears. The purpose of this system is to support assistance and guidance activities. One particularly interesting

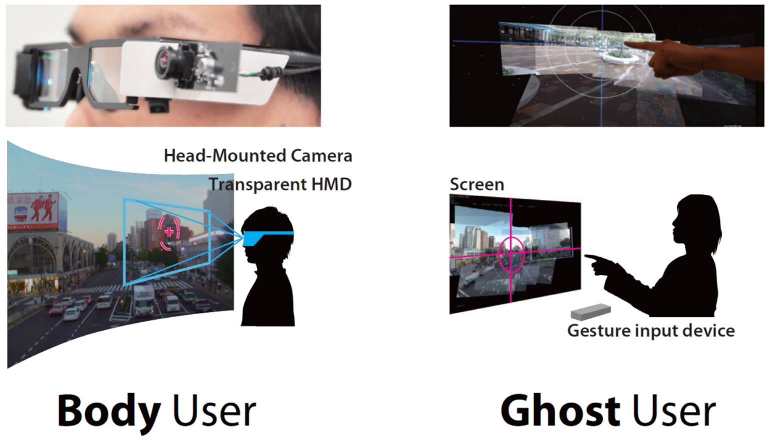


Figure 2.5: The "JackIn" mobile-video-conferencing system, which involves the use of a first-person camera and HMD by the outdoor user, and a wall-sized display and telepointing device by the indoor user. Source: Kasahara and Rekimoto [26]

finding from a study with this system was that Ghost participants spent a great deal of time looking at the different views of the composite image that were available to them, rather than just strictly looking at the live video feed coming from the Body's camera. In other words, participants took great advantage of the increased freedom they had to 'look around' the space.

Kasahara and Rekimoto [25, 24] designed a mobile-video-conferencing system (Figure 2.6) that provides the remote user with a 360-degree stabilized view into the activity space. The 360-degree video is captured from an omnidirectional wearable camera worn by the user in the activity space, and can be viewed by the remote person through a virtual-reality (VR) display. The local user's head direction is also indicated via an annotation on the remote user's interface. A user study [25] was conducted where remote users assisted local users in cleaning up a lab room. The remote users has to instruct their partners in where to put objects so that they would be in their proper places. All participants completed the task successfully, and interesting behaviours were observed. For example, remote participants matched their head movements and attentions (i.e., where they were looking) with the head movements of their partners whenever they wanted an understanding of what their partners were doing and to "establish common recognition", but they looked around the scene freely

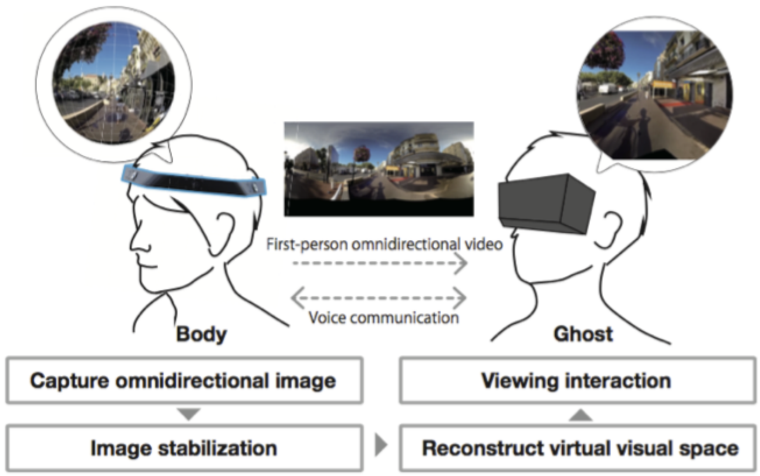


Figure 2.6: The "JackIn Head" mobile video conferencing system, which allows the remote user to view a 360-degree first-person live video view of the local user's space. Source: Kasahara and Rekimoto [25]

(independent from their partners' head movements) whenever they wanted to find or double check things on their own. Remote participants who moved their heads around on their own to explore the scene were able to perform more efficiently and with less back-and-forth communication and coordination with the local user.

To summarize, aside from having sufficient means to communicate and understand each other's intentions and focuses, collaborators should also have sufficient awareness of the work/activity space. Conventional tools often do not fulfill this requirement—a poor view into the activity space (with low resolution and FOV) and a lack of control over the view often hinder the remote collaborator's abilities to understand what is going on and contribute to the task [21]. Proposed solutions include increasing the FOV of the remove user's view [25, 24], making use of mechanical cameras operated by the remote user [32, 33], and providing reconstructed views showing what was seen before [10, 26].

Sharing Experiences

In addition to communicating and collaborating, people also use video conferencing for sharing experiences [1, 22]. This simply involves a person who is engaged in some activity or

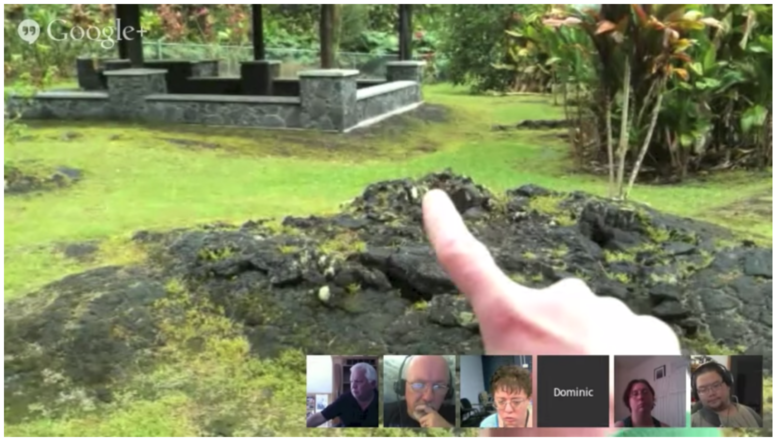


Figure 2.7: A 'Virtual Photo Walks' tour. Source: [69]

experience (e.g., sports, travelling, life events) using video conferencing to attempt to show or share that experience live with someone else, and in some cases make them feel (to some extent) like they are there and part of the experience. This happens a lot, and not just in the context of video conferencing—for example, people take photos and videos to share with others in attempt to get them to understand the experience, at least to a some extent. In addition, live-video-streaming services such as Periscope [53] and Meerkat [40] allow users to broadcast activities to a global audience and share experiences with them. In the context of video conferencing, people share experiences by, for example, sharing life events [38], taking someone on a tour of a new place [69], and showing kids' sporting events [19].

Virtual Photo Walks [69] (Figure 2.7) is a charity that allows isolated and disabled people to join Google Hangouts video-conference calls where a volunteer with a mobile device acts as a tour guide and gives a 'virtual tour' of a place (e.g., a city, a tourist attraction, or a forest). These calls are usually relatively intimate and only consist a small number of attendees (no more than ˜10 people). Attendees can see each other's faces, chat with one another, and even converse with the tour guide. Attendees can also direct the tour guide to specific spots and ask him/her to give specific camera shots. The main goal of the program is to allow people who are isolated and disable to participate in "immersive, real-time experience[s] of places they will likely never see first-hand."

Figure 2.8: The “Experiences2Go” mobile-video-conferencing technology probe. Source: Inkpen et al. [19]

Inkpen et al. [19] ran a study exploring the use of video conferencing technologies to share kids’ events occurring outside the home. Such events included soccer games, swimming lessons, and fishing. Two technology probes [18] were studied: an iPad, and a probe which consisted of a tablet, a high-quality video camera, and a tripod (Figure 2.8). The researchers found that providing multiple camera views helped in making the remote person feel like they were included in the shared activity. By doing so, the remote person was able to observe both the activity and the local parents’ reactions.

Massimi et al. [38] conducted an online survey to understand people’s experiences using mobile video conferencing in the context of major life events (i.e., events that are emotionally stimulating or life changing—such events include weddings, anniversaries, graduations, pregnancies, birth announcements, and death announcements). In general, while people do commonly use conventional mobile-video-conferencing tools (e.g., Skype, FaceTime, etc.) to connect with others during major life events, the technology falls short in providing an ideal experience. In some cases because of the less-than-ideal experience provided, some people opt to not share these types of experiences with a distant loved one, even when it is technologically feasible, because it would be deemed inappropriate for the circumstance. Some common challenges included (depending on the circumstances): the size of the space where the event was taking place (if it was too large, it was difficult and cumbersome to keep moving the camera around), the activeness of the space (if there was a lot of people or if the environment was very dynamic, it was difficult for the remote attendee to keep track of what was going on), and inability to communicate more meaningfully during emotionally-heightened moments (e.g., not being able to give a hug or kiss to a family member in a hospital).

Some of the challenges that were outlined above for using video conferencing for communication and collaboration also apply to its use for sharing experiences. For example, in many 'shared-experience-type' scenarios (e.g., sports, life events), a poor view of the action renders the connection unviewable and frustrating, resulting in a poor experience for the remote person. In addition, people observing an activity remotely might want to control their view into the environment [19], interact with the people who are part of the activity [19, 38], or perhaps even participate in the activity on their own in some way [19]. Conventional mobile video conferencing tools make this difficult; and consequently, remote users often disengage in the shared activity [21]. While not necessarily video conferencing, Periscope [53] provides viewers with a basic means of 'participating' in an activity and interacting with the people involved by sending text-based messages and 'hearts' to the user operating the camera. While technology and research have come a long way in providing improved means for the remote person to observe an on-the-go activity, there is still much work that can be done in terms of allowing remote users to more meaningfully participate in these types of activities.

## 2.2 Robotic Telepresence

Telepresence researchers have long tackled a similar problem space, where the principal concern is to allow a remote participant to engage effectively in the local space. Explorations frequently focus on different ways of embodying the remote participant, and the social con-

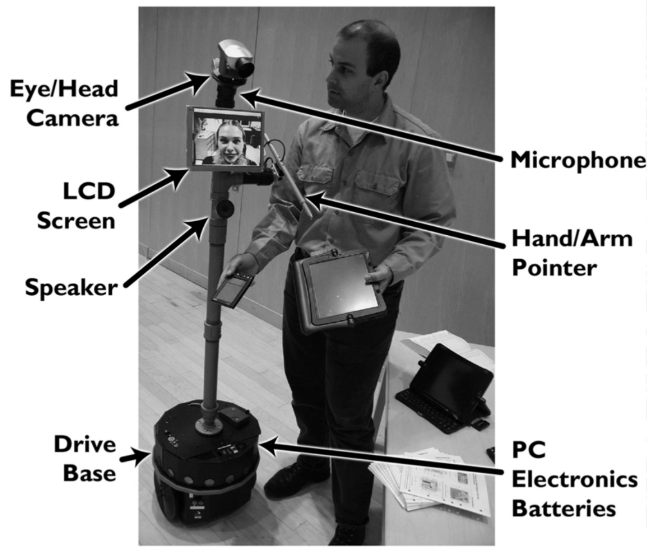


Figure 2.9: A PRoP surface cruiser telepresence robot. Source: Paulos and Canny [52]

sequences of these styles of embodiment. Paulos and Canny [51, 52] outline several of these challenges in their explorations of various form factors, including mobile blimps and surface cruisers (e.g., Figure 2.9) that anticipated many of the current commercially-available devices.

Several authors have explored the use of such mobile remote proxies (MRPs) in office environments (e.g., [34, 68]) to understand their use, and the new kinds of social norms that evolve. MRPs provide a richer experience compared to standard video conferencing, and the physical embodiment of the MRP gives remote participants the ability to engage in many activities that would otherwise be impossible (e.g., casual conversations with colleagues as they walk along a hallway [34], or casual "pop in" conversations by co-workers [68]). They provide a sense of presence for the person using the MRP, and the co-workers that engaging with the MRP, because the MRP is autonomous and mobile [56].

These technologies still suffer from significant drawbacks. Some authors argue that the video representation of the remote user is insufficient, resulting in prototypes where the video representation is larger than life (e.g., [65]). A challenge anticipated by Paulos and Canny [52] frequently reported in the literature (e.g., [34, 56, 68]) is that piloting MRPs is cognitively demanding: in addition to participating in the activities in the remote space, the MRP still needs to be piloted. In some cases, the efforts were so taxing that MRPs are sometimes left in awkward locations ([34]) rather than being properly piloted "home" (i.e., charging stations). Finally, referencing objects in collaborative physical tasks can be a challenge, because the MRP's mobility impairs people's ability establish a coherent spatial reference [37, 45, 56].

I attempt to address the piloting challenge by separating the positional aspects of piloting from the viewing aspects, and relegating control of positional piloting to a semi-autonomous algorithm. This should reduce the cognitive burden of participating in the remote activity space.

## 2.3 Interacting with Drones

Drone interaction has long been, and continues to be of interest in the HRI community. For example, thousands of military, reconnaissance, and search-and-rescue drones are being used around the globe in remote teleoperation settings, and in highly critical tasks [3, 11]. Collocated interaction with drones is a relatively new design space, which can be attributed to the recent profusion of inexpensive yet capable drones (e.g., the Parrot Bebop drone [50]; Figure 2.10). Work on collocated interaction with drones was inspired by falconeering and human-pet interaction metaphors [2, 46], employing human-drone interaction techniques based on gestures or face pose [43]. Higuchi et al. [16] map a person's head position to flight control, allowing for fairly intuitive position and view control. Hayakawa et al. [14] extend this work to physical movements (i.e., walking around in space). Higuchi et al. [15] explore the use of an HMD to provide athletes with a third-person perspective from a collocated drone as they navigate a space. Drone following algorithms typically use on-board GPS (e.g., [36]), though some researchers have explored the use of on-board computer vision (e.g., [15, 44]).

Figure 2.10: The Parrot Bebop drone [50].

Mueller and Muirhead [42] built a jogging companion from a drone, and ran a qualitative study investigating its use with 13 individual joggers. In general, people enjoyed having the drone as a jogging companion. Some participants were concerned about safety—for example, the possibility of the drone falling out of the sky and landing on the jogger. Some participants felt that the drone provided a good distraction from the "discomfort of exercise." Some people interpreted the slight movements of the drone, caused by the wind and sensor inaccuracies, as the drone trying to communicate with them. It was shown that the drone had a lot of potential to be seen as a social companion, and perhaps as somewhat of a coach or mentor.

Interesting challenges of interacting with collocated drones emerge from people's tendency to perceive and infer affect and intent from the flight paths of collocated drones [58, 60]. Mueller and Muirhead's [42] exploration of drones as running companions provides a practical example: runners were frequently uncertain of how to control the collocated drone, and/or whether the drone understood what they were doing as runners. As a result, researchers have developed simple ways of expressing and communicating drone intent expression. For example, Szafir et al. has used an LED ring under the drone [61], and flight primitives based on natural motion principles to indicate movement intent [60].

Yanco, Drury, and Scholtcz analyze users' understanding of the robots they interact with, and robots' understanding of the humans that interact with them, deriving a framework for HRI awareness [5, 70]. Based on this work, a human-UAV awareness framework [4, 3] describes how groups of operators maintain awareness of one another through the drones they are individually piloting. I extend this work by considering how awareness is maintained when each person has a different relationship with the drone. In particular, the drone mediates interaction between a field person, who is collocated with the drone, and a desktop person, who nominally "operates" the UAV. What existing frameworks do not make clear is how the field person views the drone: is the drone a proxy for the desktop person, or is the drone a shared resource to be used by both? My work enables us to explore these questions within the context of shared collaborative tasks.

## 2.4 Summary

In this chapter, I provided an overview of related work on mobile video conferencing, robotic telepresence, and interaction with drones. I discussed different video-conferencing usage contexts and scenarios, as well as what users are interested in seeing, knowing, and doing in each context and scenario. I also discussed three primary activities in mobile video conferencing: communication, collaboration, and sharing experiences. For each of these three activities, I touched on common challenges, related sample scenarios and past research, as well as technologies and solutions from both research and industry that aim to support these activities in one way or another. My own approach for supporting these three activities, which involves the use of a drone, will be described in the next chapter. I also touched on past research on robotic telepresence, where a remote user views into an environment through a mobile physical proxy and controls the position and/or orientation of the proxy. While these technologies have their advantages—some of which tackle problems in mobile video conferencing—they still suffer from drawbacks. My approach attempts to address some of these drawbacks. Finally, I touched on remote and collocated interaction with drones. Careful consideration of both the remote and collocated users is important when designing a drone-based tool to support shared communication and collaboration between the users; and it also informs how such a system could support use of the drone as a shared resource.

These works provided a basis on where to begin my own work. They provided me with a basic understanding of shared video-conferencing activities, techniques used to support such activities, and existing challenges in mobile video conferencing. My approach involves designing a system that utilizes a drone. Background in robotic telepresence, teleoperation, and human-UAV interaction helped inform the design of my system. In the next chapter, I discuss my design, introduce a potential design space for drone-based video conferencing systems, and discuss which mobile-video-conferencing challenges it attempts to address.

# Chapter 3

# DESIGN AND IMPLEMENTATION OF DRONE-SUPPORTED VIDEO CONFERENCING

In this chapter, I address Thesis Question 1: "How can we design mobile-video-conferencing technologies that provide a remote individual with fluid and flexible control over his/her view of the environment?" To do this, I discuss the design of my drone-supported video-conferencing prototype. I summarize some of the common challenges in mobile video conferencing discussed in the previous chapter, outline a set of design guidelines for drone-supported video conferencing that aim to address those challenges, introduce an early design space for position and orientation control of the drone, and describe the design and implementation of my prototype.

As discussed in the previous chapter, video conferencing while on the go introduces a whole new set of challenges. Many of these challenges depend on the activity and the space. As I mentioned earlier, the focus of my thesis is on mobile-video-conferencing activities between two people, where one (the FP) is moving around outdoors and the other (the DP) is stationary at a workstation. Examples of such activities could include (but are not limited to): giving a tour, searching an environment, inspecting an environment, laying out objects around an outdoor space (e.g., for event/venue setup), and accompanying someone during a recreational activity (e.g., hiking, cycling, running). My goal is to design a system that supports these types of activities.

## 3.1 Design Challenges and Considerations

As discussed in the previous chapter, there are numerous challenges that arise in these types of mobile-video-conferencing scenarios. Some of these challenges I address in my design approach, while others I leave for future work.

**Design Challenge 1: Camera and view limitations.**

In many cases, the phone camera can be cumbersome for the FP to operate [21]. This is especially true when the pair is engaged in a collaborative task that requires a lot of focus; and also in cases where the FP has to physically handle other objects (e.g., while building, repairing, or moving things around). The FP has to divide his/her attention between trying to provide a 'good' camera shot for the DP and actually participating in the activity at hand. 'Good' camera shots can often be very difficult to attain, especially with a camera that provides limited freedom (in terms of where it can go and how far it can zoom) and low fidelity. In addition, the fact that the camera is coupled with the FP limits the types of views that can be attained (for example, the DP cannot see a third-person view of the FP with this kind of setup). If the DP wants to see something or the FP wants to show something, the FP must move to the particular location in the environment containing the object of interest. A view that is coupled with the FP means that the DP's view into the space cannot be separate from where the FP is—the FP and DP cannot be looking in separate locations. This adds another level of inconvenience and inefficiency in task completion, especially for activities that could benefit from two people looking or being in separate spots at the same time (e.g., searching, inspecting).

***My Approach: Provide a view that is decoupled from the FP.*** Decoupling the DP's view into the space from the FP can help address many of these issues. Allowing the DP to control the *spatial* position of the camera (alongside the orientation and zoom) gives the DP greater freedom, more potential to see things that the FP might not see, and could potentially make task completion simpler and more efficient in many collaborative activities (e.g., search, inspection, and setup activities).

**Design Challenge 2: Asymmetry of view control.**

With conventional tools, the FP has full control of the DP's view into the space, whereas the DP can only *suggest* where the camera be pointed. This creates an asymmetry of view control, which can be an issue [21]. The FP and the DP might not always agree on what constitutes a 'good view'. For example, the FP might be providing what he/she thinks is a good view to the DP, while the DP wants to see more (or something else) in the view. In addition, there is also a question as to whom the camera view should serve—the FP, or the DP? The FP might want to use the camera to *show* something, while the DP might want to use the view to *see* something. There may be cases where what the FP wants to show and what the DP wants to see are not the same thing. These kinds of cases could create a tension between the FP's needs and the DP's needs. In standard mobile video conferencing, because the FP is in direct control of the camera, the DP's needs often become unsatisfied; either because the FP chooses to satisfy his/her own needs first (because it is easier), or because it is difficult for the DP to communicate his/her needs to the FP. This raises an interesting question: what happens if the DP has more control?

***My Approach: Balance view control between the FP and DP.*** Providing the DP with a way to directly control the camera view can help in situations where the DP may have difficulties conveying or communicating his/her intentions to the FP. In addition, if the DP can directly control the view on his/her own, then he/she may be more motivated to look around the environment knowing that his/her actions are not causing a burden on the FP.

While it is beneficial to give view control to the DP, since the DP uses the view to his/her advantage to see things in the activity space, the FP might also want to show things to the DP; and thus utilize the view for that purpose. Therefore, the view should be thought of as a *shared resource* used by both communicators rather than as tools owned and used only by one communicator. The question then becomes: how can we design the view in such a way? In conventional mobile video conferencing, the single camera view often cannot serve both the FP's needs and the DP's needs at the same time. Sometimes the group has to choose whose needs to satisfy. A solution could involve balancing control of a single view between communicators (e.g., through control switching, negotiation, or shared control) and/or providing separate camera views that each belong to a single communicator.

**Design Challenge 3: Limited awareness of the activity space.**

Conventional mobile-video-conferencing tools often provide the DP with insufficient awareness of the activity space [21]. This is caused primarily by two problems: (1) the DP's view into the activity space is insufficient (low FOV, low image quality, low frame rate, etc.), and (2) the DP lacks control over his/her view. Video conferencing while on the go adds to this challenge—mobile video chats often take place in dynamic environments with a lot of activity and movement, including movement of the objects and people in the space as well as the movement of the camera within the space.

***My Approach: Provide the DP with greater awareness of the activity space.*** Workspace awareness is important in collaborative activities [12]—if the DP is more aware of the environment and what is contained within it, then he/she will have more opportunities to interject and contribute to the activity at hand. This is especially true for activities that involve searching, inspecting, navigating, and laying out objects in a large environment. There are means of doing this that can be achieved through design—for example, many smartphone cameras lack the ability to optically zoom, lack image-stabilization, and provide a low-FOV video feed. Providing these features would be a simple step forward. Other potential solutions include providing multiple camera views, views that show a lot of information about to the workspace and the activity, and providing contextual information (e.g., maps, activity statuses, etc.) [27]. Increasing the video quality would also constitute a major improvement—but if this cannot be achieved due to bandwidth limitations, then another solution could involve providing the ability to temporarily raise the quality at the expense of frame rate, or allowing either user to snap a high-resolution photo with the phone camera at any moment for the DP to view.

## 3.2 Motivation for Drone-Supported Video Conferencing

I propose the use of a drone to support mobile-video-conferencing. A drone, or unmanned air vehicle (UAV), can essentially act as a camera in the sky, providing a rich view of the activity space that is decoupled from the FP and easily controllable (addressing all three design challenges). Drones are already used extensively for video recording and broadcasting for purposes such as filmmaking and journalism. They can move around a space with numerous degrees of freedom—up and down, left and right, forward and backward, and rotation (yaw)—and provide varying types of shots of the space, of objects and people within the space, and of course of the FP, showing his/her actions and location relative to the space. Shots that would otherwise be difficult to attain with a smartphone or other coupled camera can be easily attained via a drone. In addition, drones can be remotely controlled, introducing the possibility for direct control by the DP.

The view from a drone can either be used to completely replace or supplement the view from the FP's phone. Communicators may wish to use it either way depending on the types of shots they wish to attain and the information they wish to communicate. For example, it is possible that communicators could use the smartphone camera to provide up-close detail shots from on the ground while using the drone camera for providing overview or contextual shots (e.g., showing the object of interest captured on the phone camera in relation to the larger environment). Depending on how much control the DP is given over the drone, it is also possible that the FP could use the phone camera as his/her view to *show* things to the DP while the DP uses the drone view as his/her view to *see* or look for things in the space. If the FP is given the ability to view the drone view while the DP is given the ability to control it (while of course also being able to view it), the DP could also use the drone view to *show* things to the FP that the FP might not initially notice. These are only some of the possible ways that communicators might utilize the drone view.

In addition to the camera view, the drone could also potentially give the DP a physical embodiment in the activity space. Aside from just 'seeing through' the drone, it is possible that the FP could see the DP as being 'present' in the space through the drone as a proxy—similar to how a telepresence robot acts as a proxy of a remote operator. The drone as a proxy would essentially give the DP a 'location' within the space. In addition (depending on how much control the DP has), it would also give the DP a means of changing his/her location in the space and moving around the space to show or guide the FP.

## 3.3 Drone and View Control

One of the challenges we need to address (Design Challenge 2) is asymmetry of view control. Balancing view control between the FP and the DP is something that should be considered. Control of the drone can be given to either the FP or DP; or both communicators could share control of it to some degree. There are many ways in which a drone can be controlled. One of the most common ways is through a joystick-like interface, such as those provided by many out-of-the-box consumer drones (e.g., [50]). Researchers and designers have also explored other means, including defining flight plans on a map (e.g., [7]), control with head movements [16], control by walking around in space [14], control with a Sphero robotic ball [59, 62], and 'semi-autonomous' control by following a person or target [36, 15, 44].

One important thing to consider is that having too much control over the drone can be cognitively demanding, especially if the operator lacks training and has to also focus on another ongoing activity at the same time. The operator has to constantly keep track of the position of the drone and manoeuvre it so that it always stays where it should. In addition, controlling with a joystick interface or any other interface that requires the use of one's hands leaves the operator's hands unavailable for other tasks, such as picking up and carrying objects. In HRI, the concept of *shared control* [49, 23, 11] refers to the idea of designing interfaces that balance control between the human operator and the machine itself. Many human-robot interfaces, including drone interfaces (e.g., [41]), are designed with this idea, usually to allow the machine to carry out low-level operations (e.g., adjusting yaw, pitch, roll; changing the speed of certain propellers) while the human operator defines higher-level goals (e.g., keep following *this* target, fly to *this* location, etc.). There is also a notion of shared control between two or more human operators (e.g., [3, 39]), where (for example) each person controls one aspect of the operation of the machine, or each person passes control of the machine from one to another.

In my design, I apply the concept of shared human-machine control by designing the drone to keep track of the FP in the activity space and always stay relatively close to him/her (i.e., within an imaginary bubble around the FP). In addition, I apply (to some extent) the concept of shared human-human control in the following ways: (1) by giving the FP implicit (indirect) control of the drone's position relative to the activity space (by moving around the space and having the drone follow him/her to stay within the bubble around the FP), (2) by giving the DP control of the drone's position relative to the FP (i.e., its position *within* the bubble around the FP), and (3) by giving the DP control of the camera pan and tilt. In addition, beyond control of the drone, the FP has control of the camera view from the smartphone just as he/she would using a conventional mobile-video-conferencing system.

### 3.3.1 Early Design Space for Control

I present a simple two-dimensional spectrum (Figure 3.1) of levels of control. The major axes of this spectrum are control of the direction of the drone's camera view (i.e., the pan/tilt/zoom of the front-facing camera, as well as the yaw/orientation of the drone), and control of the drone's X-Y-Z position. Designs further along each axis give the DP more control, while designs less far along give the DP less control and the drone more autonomy.

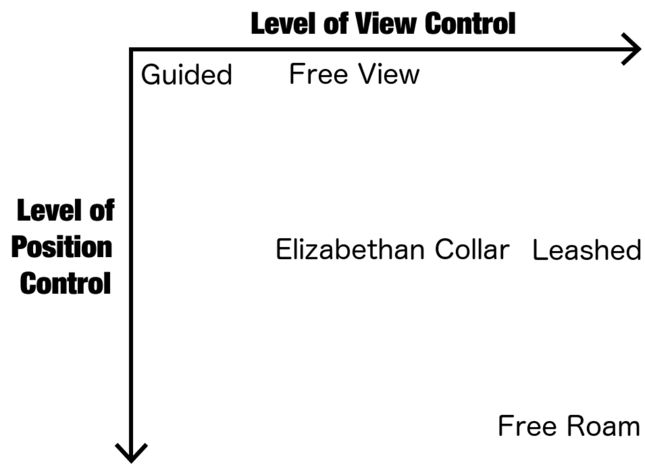


Figure 3.1: A two-dimensional design spectrum of drone-control strategies. Designs further along each axis give the DP more control, while designs less far along give the drone more autonomy.

| Style | View Control | Position Control |
|---|---|---|
| Free Roam | Yes | Yes |
| Leashed | Yes | Partial |
| Elizabethan Collar | Partial | Partial |
| Free View | Partial | None |
| Guided | None | None |

Table 3.1: A list of sample drone-control strategies, showing what each provides to the DP in terms of view and position control.

Designs less far along each axis also implicitly give the FP more control; although the FP's control is more indirect (i.e., because the FP would have to physically move around the space to control the drone, he/she still would not have 'direct' control). Next, I present a sample of control strategies (Table 3.1) along this spectrum. Each control strategy is illustrated in Figure 3.2.

Control Strategies

While the following not an exhaustive list of control strategies, it articulates meaningful strategies that could be explored based on our existing experiences and expectations from

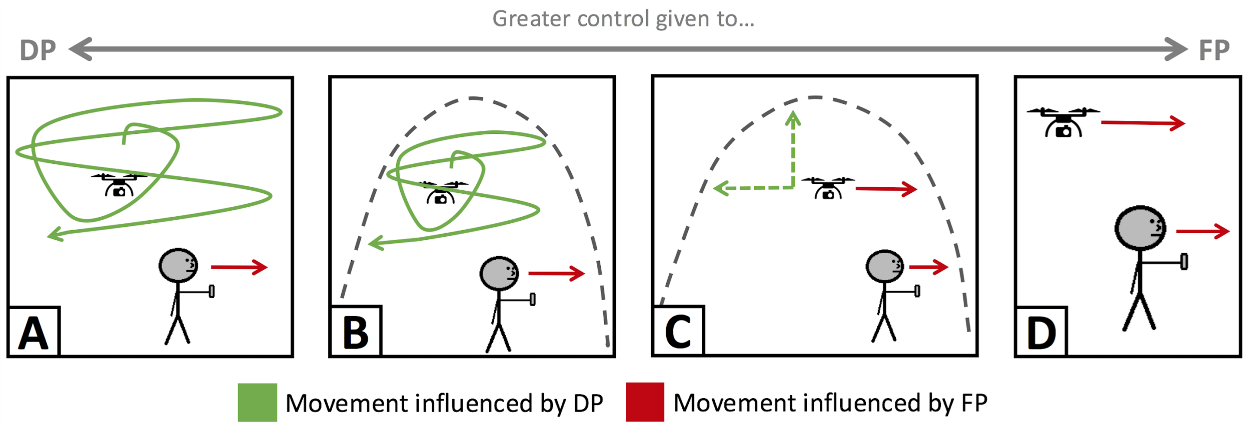


Figure 3.2: Illustrations of sample drone-control strategies. From left to right: (A) Free Roam; (B) Leashed; (C) Elizabethan Collar; and (D) Free View and Guided.

everyday life.

**Free Roam:** In this strategy (illustrated by Figure 3.2a), the DP has full control of the movement of the drone and its attached camera. The DP can fly the drone to any X-Y-Z position, orient the drone in any direction, and adjust the pan/tilt/zoom of the front-facing camera in any way. This is akin to many fully-featured telepresence robots (e.g., [51, 52]), and offers the most flexibility. However, this requires the most cognitive effort to control.

**Leashed:** In this strategy (illustrated by Figure 3.2b), the DP has full camera pan-tilt-zoom control and full yaw control, but only partial control of the drone's position. The DP may fly the drone anywhere within an imaginary sphere of a pre-defined radius around the FP (i.e., the extent of the leash). This imaginary sphere moves with the FP, thus causing the drone to stay 'tethered' or 'leashed' to the FP. Within this sphere, the DP retains full control of the view. Here, the drone's orientation and camera pan/tilt/zoom are fully controlled by the DP, while the drone's X-Y-Z position is influenced by both the FP and the DP.

**Elizabethan Collar:** In this strategy (illustrated by Figure 3.2c), the DP has camera control within limited viewing angles, and partial positional control. Here, the drone autonomously follows the FP and always orients itself toward the FP. The DP can adjust the altitude and following-distance at which the drone follows the FP, and in addition can still adjust the pan/tilt/zoom of the front-facing camera. Here, the drone's orientation is fully influenced by the FP, the camera pan/tilt/zoom is fully controlled by the DP, and the drone's X-Y-Z position is influenced by both the FP and the DP. This control strategy is similar to those found in third-person perspective video games (e.g., Tomb Raider), where the view is limited because the avatar's head can only turn within a limited angular range.

**Free View:** This strategy (illustrated by Figure 3.2d) is exactly like the Elizabethan-Collar strategy, except here the DP has no position control. In this case, the drone autonomously follows the FP, and the DP only has control of the pan/tilt/zoom of the front-facing camera. Here, the position and orientation of the drone are entirely dependent on the FP's movements within the space.

**Guided:** This strategy is exactly like the Free-View strategy (Figure 3.2d), except here the DP has no camera control. In this case, the drone always autonomously follows the FP and automatically adjusts its camera so that it always points directly at the FP. This is somewhat akin to the complete control that the FP has in many mobile video conferencing scenarios (e.g., [21, 27]), where there is only the smartphone view and the FP has complete control of it. Here, the drone's position, orientation, and camera pan/tilt/zoom are entirely dependent on the FP's movements within the space.

## 3.4 System Design

I designed a mobile-video-conferencing system (Figure 1.2) that utilizes a drone adhering to the Elizabethan-Collar control strategy. The drone autonomously follows the FP, showing the DP a third-person bird's-eye view of the FP and his or her surrounding environment. The purpose of this view is to provide an overview of the FP's environment to the DP and to show the FP in context of his/her larger environment. While the DP has no direct control of the drone's movements (since the drone always autonomously follows the FP), the DP can control the (virtual) pan and tilt of the camera attached to the drone (within a ˜180-degree

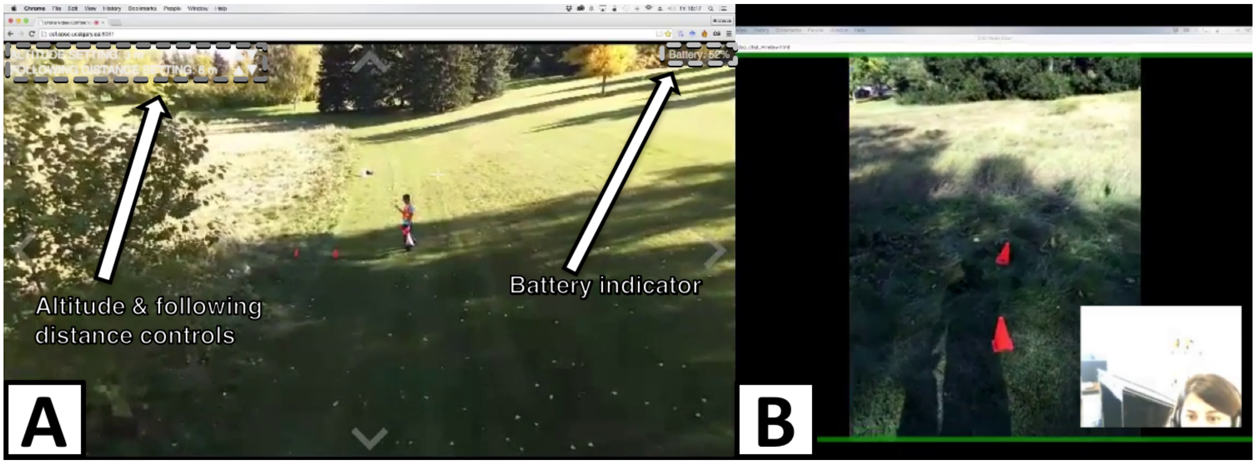


Figure 3.3: The desktop interface—the left monitor (A) contains the drone-camera view and a number of indicators and camera controls, while the right monitor (B) contains the Skype-like video-chat interface.

range for both pan and tilt), and adjust the altitude above the ground and the distance from the FP. The FP has a mobile phone that he/she can use to show detail views of the environment if necessary.

### 3.4.1 Interface

Figure 3.3 illustrates the interface that the DP uses: it provides two separate views—the view from the phone held by the FP, which resembles a traditional video-chat interface (e.g., Skype), and a view from the drone. The “drone view” has simple arrows (mapped to keyboard arrow keys) that the DP can use to manipulate the virtual pan and tilt of the drone camera. This view also has indicators and buttons allowing the DP to view and adjust the altitude and following distance parameters that the drone is trying to maintain. Finally, it provides a simple green outline indicator to show if the FP’s phone interface is displaying the drone view or the video-chat view.

The FP’s phone has a very simple interface (Figure 3.4), showing either the drone or the regular video-chat view, where the FP can toggle between each. The FP can also toggle between the front- and back-camera views from his/her phone. This allows the FP to have a

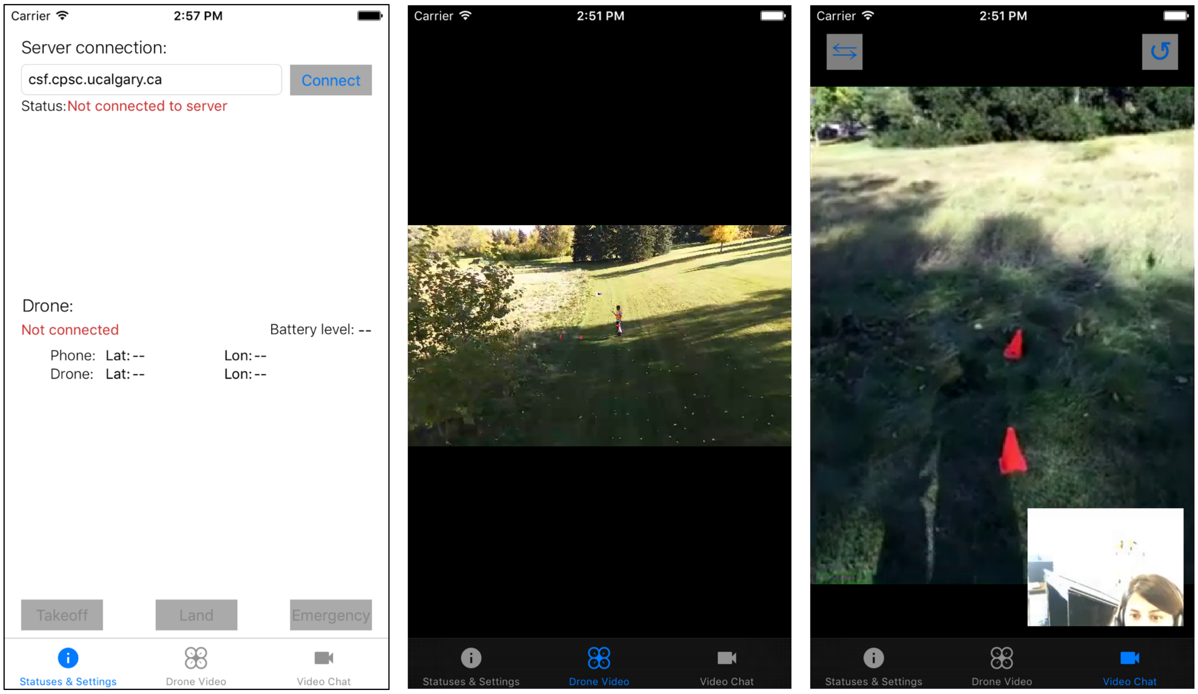


Figure 3.4: The mobile interface—the three tabs along the bottom allow the user to switch between three screens: (1) the statuses-and-settings screen (left), to change settings and view statuses; (2) the drone-video screen (centre), to view the drone-video view; (3) and the video-chat screen (right), to view a Skype-like interface showing the phone-camera view and the view from the DP's side. A two-way audio/video link is maintained regardless of which screen is being viewed.

conversation as 'talking heads', or to show the DP something in the local space. A two-way audio/video link is maintained regardless of which view is being used.

## 3.5 System Implementation

The system as a whole utilizes consumer-grade hardware and was implemented in C, Objective-C, Swift, JavaScript, HTML5, and CSS.

### 3.5.1 Components

Figure 3.5 illustrates the four components of the system: a Parrot Bebop drone, an iOS smartphone, a Node.js server, and a Mac desktop machine running a web-based desktop client in Chrome. The drone has a virtual pan-tilt camera providing a stabilized video feed

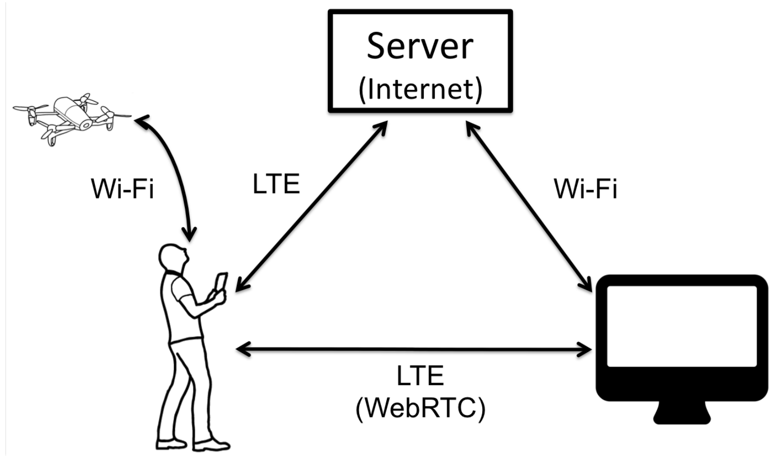


Figure 3.5: Block diagram of the system implementation.

with a ˜180-degree FOV (i.e., a fisheye lens), and is controlled by a custom iOS smartphone app connected via an internal Wi-Fi connection. The iOS app (developed in Objective-C, Swift, and C) runs on the FP's smartphone (iPhone 6 Plus), where the smartphone acts as the gateway (via LTE) to the other devices. A Node.js server (developed in JavaScript) relays information and commands between the smartphone and desktop, and helps establish a WebRTC video/audio call between the phone and desktop web application. The desktop web application (developed in HTML5, JavaScript, and CSS) allows the DP to adjust parameters (following distance and altitude) and control the pan and tilt of the drone camera.

### 3.5.2 Follow Algorithm

The drone uses a simple control loop running at 40 Hz to follow the FP, where both the drone and FP are tracked via GPS (the FP via the smartphone's GPS):

```
while (droneIsFlying) { // this loop iterates once every 25 ms

    // lowerBound and upperBound defined based on altitude settings set by the
        DP. The difference between the two bounds is always 2 metres.

    if (droneAltitude < lowerBound) {
        drone.pitchStill()
        drone.climbUp()
    }

    else if (droneAltitude > upperBound) {
        drone.pitchStill()
        drone.climbDown()
    }
```

```
    else {  // drone is at the proper altitude

        if (droneIsProperlyOriented) {

            // innerBound and outerBound defined based on following distance
                settings set by the DP. The difference between the two bounds
                is always 2 metres.

            if (distanceBetweenDroneAndPhone < innerBound)
                drone.pitchBackward()
            else if (distanceBetweenDroneAndPhone > outerBound)
                drone.pitchForward()
            else
                drone.pitchStill()
        }

        else {
            drone.pitchStill()
            drone.orientTowardPhone()   // adjust yaw to face the phone
        }
    }
}
```

At each step, the drone checks three things: (1) its bearing (i.e., pointing direction) to ensure the FP is within an appropriate range (20 degrees), (2) the distance from the FP (to ensure the FP is within range), and (3) the altitude (to ensure wind has not blown it off course). Once this check is complete, the drone makes micro-movements to correct for each of these factors (i.e., turn if necessary, move toward the FP if necessary, and fly up/down as necessary). This happens without explicit action on the part of the DP.

### 3.5.3 Safety

Not shown in Figure 3.8: an additional component of the system is a mobile app that runs on the field investigator's phone. The purpose of this is to provide the field investigator with manual-override control of the drone in case something goes wrong (for example, the drone gets close to colliding with a tree or with a person, the drone tries to leave the bounds of the activity space due to GPS or compass inaccuracies, etc.).

### 3.5.4 Source Code

The full source code of the prototype system can be found here: `https://github.com/BrennanJones/DroneVideoConf`.

## 3.6 Conclusion

In this chapter, I discussed the design and implementation of drone-supported video-conferencing. I reviewed a set of common challenges in one-to-one mobile video conferencing between a DP and an FP, and based on those derived a set of design considerations for systems that support such activities: (1) *provide a view that is decoupled from the FP*, (2) *balance view control between the FP and DP*, (3) *provide the DP with greater awareness of the activity space*. I explained my motivation for exploring mobile-video-conferencing systems that make use of a drone—drones are relatively inexpensive, they are versatile, and they are very capable of moving freely around the space and providing varying types of camera shots. They can easily provide camera views that are decoupled from the FP (addressing Design Challenge 1), they can be controlled by an operator who is collocated or remote (addressing Design Challenge 2), and they can provide up-in-the-air overview shots that show a lot of information (addressing Design Challenge 3). I discussed the concepts of shared human-machine and human-human control in relation to drone-supported video conferencing, and I introduced an early design space for shared-control strategies as well as sample strategies and their places within the design space. These sample control strategies were: *Free Roam*, *Leashed*, *Elizabethan Collar*, *Free View*, and *Guided*. I also discussed the design of my prototype system, adhering to the Elizabethan-Collar control strategy, which makes use of a consumer-level drone, a smartphone app, and a desktop app. Finally, I discussed the implementation of my prototype using a Parrot Bebop drone; an iOS mobile-client app developed in C, Objective-C, and Swift; a Node.js server developed in JavaScript; and a web-based desktop-client app developed in JavaScript, HTML5, and CSS.

My prototype system was designed for the purpose of understanding how pairs of people communicate, collaborate, and share experiences with one another through the use of a drone-supported video-conferencing system. While I did not implement every level of control available in my design space, I built the system to utilize a control strategy that roughly balances control between the FP and the DP. The goal is to see how communicators make use of this balanced control and to get an early understanding of how the drone and the camera views are utilized as shared resources to accomplish each communicator's goals as well as the collective goals of the group.

This chapter addresses Thesis Question 1: "How can we design mobile-video-conferencing technologies that provide a remote individual with fluid and flexible control over his/her view of the environment?" It also provides the following contributions: (1) the introduction of an early design space for drone-based video conferencing, and (2) the design and implementation of a drone-based video-conferencing prototype system following a particular control scheme in my early design space.

In the next chapter, I go over the design, the objectives, and the findings of an observational study that I conducted with my prototype system.

# Chapter 4

# OBSERVATIONAL STUDY OF DRONE-SUPPORTED VIDEO CONFERENCING

In this chapter, I describe the observational study that was conducted in order to understand how people would make use of a drone-based video-conferencing system to communicate and work on collaborative activities. This study and its results address Thesis Question 2: "What are some of the opportunities presented by supporting mobile video conferencing using drones?"; as well as Thesis Question 3: "What are some of the interaction challenges presented by drone-supported video conferencing?" In addition, the study and its results address the overarching research question I presented at the beginning of this thesis: "How can we design mobile-video-conferencing technologies that better support active back-and-forth collaboration between peers?"

This study was done in collaboration with two colleagues from my lab: a fellow MSc student and a research assistant. Because the study involved the interaction between participants and investigators in two separate locations, this collaboration was necessary.

The study was framed around three central questions:

- How do people make use of the drone view and the view from FP's phone to coordinate activity?
- How do people manage and understand both views of the environment, if at all?
- How do people use the control scheme, and is it sufficient?

## 4.1 Participants

We recruited eight pairs of participants through poster ads placed around our university campus and through online ads. One participant's (G3-FP) demographic data was lost. There was initially a ninth pair, but we dropped their data from the full study due unacceptable network latencies (i.e., greater than 10 seconds). The remaining 15 participants (eight female, seven male) ranged in age from 19 to 32 years old, with a mean age of 23.7 years and a standard deviation of 3.6 years. All but one pair were current university students, and the remaining pair had university education. All participants knew their partner prior to the study, with a wide range of existing relationships, including: friends (two groups), colleagues (two groups), spouses/dating (two groups), roommates (one group), and parent/daughter (one group). All participants had prior experience with video conferencing, with eight responding that they used it "very often", five saying they used it "sometimes", two using it "rarely", and zero using it "never". Eight participants reported previous experience with video chatting either while on the move (i.e., moving around) or with a mobile device. We recruited the participants as pairs to ensure they had pre-established rapport, asking only that one member of the pair was comfortable with walking around outdoors in a large field containing tall grass.

## 4.2 Method

Participants completed a warm-up task and two study tasks together. The DP worked from a desk in an office environment, while the FP worked from an outdoor field location. We asked the FP to wear a bright retro-reflective vest so that he/she was more easily visible from the drone-camera view. The overall duration of the study ranged from 60 to 90 minutes.

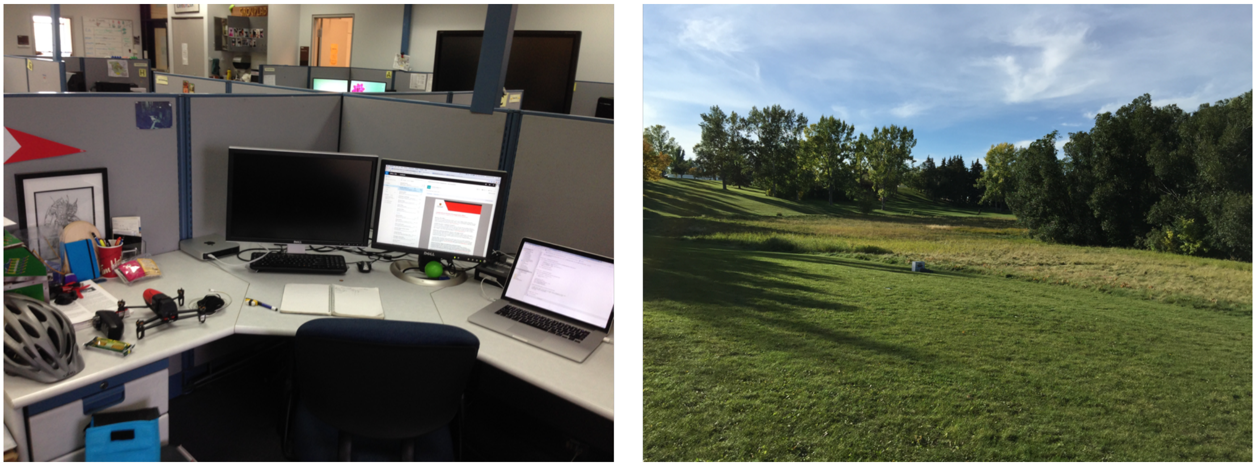

Figure 4.1: The study locations: a desk in our lab (left), and a park near our university (right).

### 4.2.1 Arrival, Locations, and Warm-Up Task

We asked participants in each pair to arrive at two different locations (Figure 4.1): one at our research lab, and the other at a nearby park. The park contained a field measuring roughly 70m x 140m that had a large, circular section of tall grass surrounded by trimmed grass. The field was also surrounded by irregular patches of bushes and trees. Upon arrival, each participant began the consent process, was briefed on the overall purpose of the study, and was introduced to their respective software interface. The participants then completed a warm-up task, giving them the opportunity to familiarize themselves with the system. For the warm-up activity, we asked participants to chat while the FP walked around the park with the drone following him/her around, and the DP practiced adjusting the controls and moving the camera on the desktop interface. The warm-up activity continued until either the pair said they were ready to continue or the drone battery depleted.

### 4.2.2 Study Tasks

Participants then completed each study task in turn. After each task, each participant provided a quick, independent verbal reaction to capture his/her immediate thoughts on the experience. Between the tasks, participants waited briefly (~5 minutes) so that the study investigator could switch batteries and restart the drone. Upon completing the second task, both participants completed a post-study questionnaire and demographic survey. Sample questionnaire questions included: “To what extent did you enjoy video conferencing with your partner?” and “How immersed did you feel in the activity?” Finally, we interviewed participants about their experiences as a pair through the video chat in order to understand their experiences as a group. Sample interview questions included: “What kinds of problems did you encounter as a group?” and “How do you think the experience would have compared to being physically together?”

We were interested in recreating scenarios where there was plenty of back-and-forth interaction between the FP and DP, in order to understand the role of our system in supporting communication and collaboration towards a shared goal. While this goes beyond activities that mostly involve passive observation from the remote user (e.g., watching a soccer game [19], viewing a tour [69]), these activities could still involve some active control and participation from remote users—indeed, one of the main goals of our design is to support and encourage active participation and engagement from the remote user. We designed the tasks based on anticipated styles of use: searching for things together (e.g., [54, 55]), and exploring environments together (e.g., [21]). In doing so, we aimed to include subtasks that comprise these broader activities, including: navigation style tasks (where one participant directs the other to a location), examining and exploring physical objects, and differential knowledge/knowledge sharing tasks. In designing these tasks, we were not so much interested in whether participants completed the tasks quickly; but rather in the ways in which they approached the tasks and communicated with one another.

**Study Task 1: Setting up the field for an event.** The DP is given a set of instructions for how the field ought to be set up with physical props (three pylons and two bags of coloured balls). These instructions include an overhead map (Figure 4.2) of the field with


Figure 4.2: An overhead map of the field location (included as part of the DP's set of instructions for Task 1) showing where certain props should be placed.

written directions about how and where the props should be placed. This task was designed to primarily be a navigation task, where the DP directs the FP to the correct location for each prop. The FP's role is mainly to move about the space, provide an 'on the ground' perspective from his/her mobile phone, and place objects according to the DP's directions.

**Study Task 2: Scan and search.** In this task, the DP and FP navigate through several locations within tall grass in the field together. Each location (20 in total) is marked by a hula-hoop, where each hula-hoop is of one of five colours (red, pink, yellow, green, or purple), and has one of four symbols (stripe, plus sign, hat symbol, or equal sign) through the centre of the hoop. The DP's instructions include the sequence of hula-hoop symbols to navigate through, but not the colours of the hoops. Meanwhile, the FP is given the colour of the first hula-hoop and each visited hoop provides the colour of the next to visit, but not the symbol. The task is to pass through as many locations as possible in the correct sequence within the drone's battery-life timeframe. As illustrated in Figure 4.3, the colour of a hoop

Figure 4.3: Hula-hoop colours are often visible from the FP's perspective (right), but not from the drone view (left).

is only visible from the ground, while the patterns in nearby hula-hoops are visible from the sky. In addition, because the hoops are place within tall grass, the patterns of hoops cannot be seen as easily from the ground as from the sky unless the FP walks closer to a hoop. This task is a dual searching task where both the DP and FP need to inspect the environment (at different scales), navigate (where the DP will direct the FP, and the FP informs the DP where he/she is), and exchange knowledge as the collaborators problem solve together.

### 4.2.3 Data Collection and Analysis

We collected position and bearing data of both the drone and the smartphone on a 0.5-Hz interval. We also collected and logged information about FPs' interactions with the interface (e.g., when does the FP have the drone view open, when is the FP using the front/back camera, etc.) We captured the video from the drone and smartphone, and the conversation across the video-conferencing link. We augmented this with video captured from a head-mounted camera worn by an investigator that we used to follow the FP around. Finally, we collected field notes from direct observations of the FP and the DP.

We conducted a thematic analysis based on field notes and interview data. Using this an initial set of themes, we iteratively reviewed and selectively transcribed the collected video data. This process resulted in an understanding of the general flow of participants' actions, as well as a set of critical incidents that illustrate or highlight challenges that participants experienced.

### 4.2.4 Additional Details in Appendices

Additional information about the study, including previous iterations of the study tasks and design, as well as supplementary study materials such as the consent documents, task information sheets, and other protocol documents, can be found in Appendices A and B respectively.

## 4.3 Findings

In total, our analysis covers the experiences of eight pairs. For three pairs, the drone crashed into trees or experienced a glitch in the altitude control, but these all occurred late during the task at hand, so we kept the data for analysis. Participants completed the first task without too much difficulty, though all pairs struggled to complete the second task in the allotted time.

We outline three main themes that emerged from our observations of participants' experiences: the DP's experiences of the drone's perspective, the FP's experiences being collocated with the drone, and utilization of the drone as a shared resource. Throughout our results, we refer to audio transcripts and participant quotes where we label them with G# for group number followed by DP or FP, depending on the participant's role.

### 4.3.1 Experiences with the Drone as a DP

Participants, particularly DPs who viewed the scene primarily through the drone view, enjoyed the experience. Most salient of these responses was from G7:

> **[G7]** DP: This is super cool... It was cool that I could almost be in the park with [FP]. I feel like I've been outside this morning.

Figure 4.4: An over-the-horizon view showing the park and the autumn-coloured trees. G8-DP adjusted the drone to attain this view, then paused and admired the view while talking about it to her partner.

Bird's-Eye View

While participants generally expressed excitement over the novelty of the bird's eye view, these positive feelings were also reflected in comments about the utility of the view in terms of the task:

> **[G7]** FP: It was helpful to have the bird's eye view. [DP] could see things I couldn't see, and I could see things she couldn't see.
>
> **[G4]** DP: The aerial view was a massive advantage [over using the mobile phone].
>
> **[G2]** DP: Better [using the drone] to see the world view and help him see what's available.

DPs lingered on views that provided them with great perspective over the scene (e.g., Figure 4.4) that would be otherwise impossible to have from the ground.

Dual View

Pairs used the smartphone cameras in a variety of ways where both the back camera (first-person view) and front camera (third-person view of the user's face) were valued. Based on

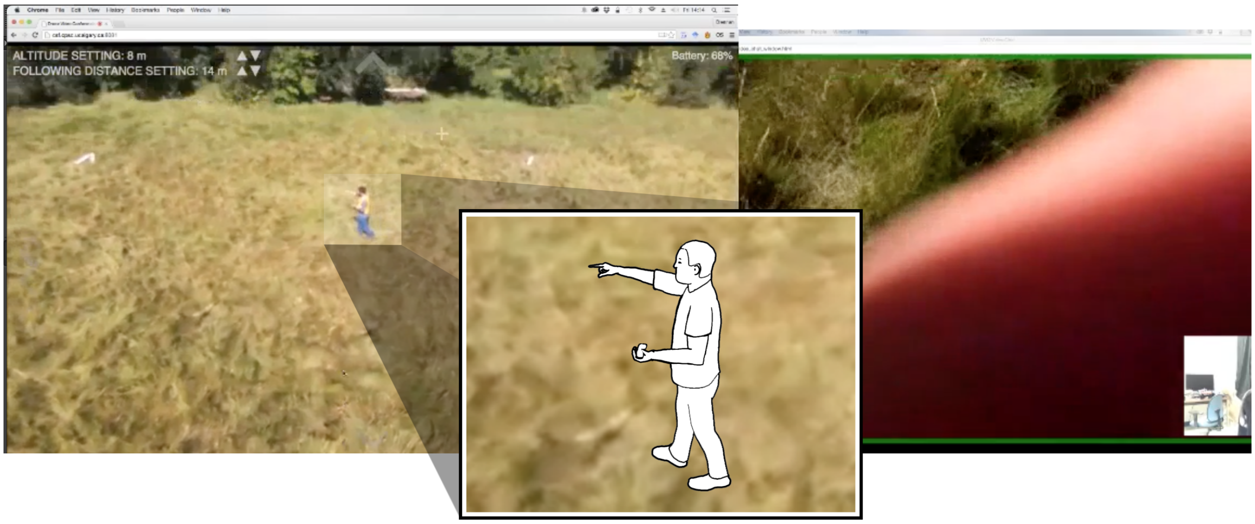


Figure 4.5: The FP (G5) points with the intention of his gesture being visible from the drone view (left). At the same time, he is ignoring what is visible in the video-chat view (right).

prior work, we generally expected FPs to use the back camera to provide DPs with detailed views and information (e.g., [21]) that could not be seen from the drone. Yet, FPs used these views in a number of unusual ways, for example: using the back camera to show scenery to establish common ground, showing a view of themselves, and other times (as can be seen Figure 4.5, where G5-FP covers the camera with his finger), they would forget the camera was there altogether. The use of the front camera to show the FP's face provided enjoyment to the DPs (similar to [19, 68]) allowing the DPs to see both the task environment and the FP's reactions to the task.

> **[G6]** DP: Sometimes while travelling, you would video conference with someone. He or she can watch your face, and if you want to show them something, you need to turn the camera. [With this, you don't have to, and] having the possibility of having both is nice.

> **[G7]** DP: It was kind of fun to watch. It is like doing something vicariously from the drone's eye view. Being able to see the face and the drone view—you feel like you're in pretty close contact.

Participants laughed and joked with one another, looking at each other's faces as they did so. This suggests that while participants were engaged in the task, they were still very much enjoying their time in relation to one another's experiences.

### Altitude-vs.-Resolution Tension

The altitude and following distance of the drone were the only sources of flight control for DPs. This brought with it an interesting tension—on the one hand, DPs wanted to see more of the field to gain visual context; on the other hand, being very high meant that DP had difficulty seeing visual details in the scene. All DPs mentioned having difficulties seeing certain details (such as the colours of the hula-hoops, as can be seen in Figure 4.3) in the drone view.

### Providing Navigational and Directional Cues

While the system provides a means for the DP to look around, DPs had a difficult time communicating where the FP should go. As the DP was tasked with helping the FP to navigate the environment in both tasks, we observed breakdowns in these communicative acts across all pairs. We consider two distinct challenges faced by our participants: first, that DPs were not always aware of where the FP was located relative to the drone; and second, that the DP and FP had completely different frames of reference.

**Location Awareness.** Providing directional cues for someone necessitates an understanding of their location and perspective on the world. Without this, it was difficult for DPs to provide FPs with directions. In spite of the drone's "follow" algorithm, the FP would sometimes move out of view of the drone. For example, in one instance (G5), the FP was out of view because of a tree that blocked the drone's view (Figure 4.6):

> **[G5]** FP: Can you see me?
>
> DP: *[Has been exploring the view on her own.] [Turns the view, and pans around.]* Wait, wait, wait... I can't find you! This is... Wait! Are you behind a tree?
>
> FP: No, I'm not. *[By this time, he is not behind the tree.]*
>
> DP: Oh, there you are! You were behind the tree.


Figure 4.6: In this example vignette, the FP walks by a tree (A) while the drone flies by, putting the FP out of frame (B). By the time the drone rotates to correct its orientation (C), the FP is behind the tree and the DP cannot find the FP. Eventually, the FP appears in frame again (D) as he continues walking.

In spite of the fact that the field setting was fairly sparse with few high obstacles, we see here the the FP disappeared from the DP's view. We expect that these types of occurrences would happen more frequently in general—for example, in urban environments that contain more tall obstacles (trees, buildings, etc.).

In the above vignette (G5), once the FP was within view again, the DP did not recognize his presence immediately, and still needed to carefully examine the drone video to find the FP. Without knowing where the FP was, the DP could not provide further instruction or navigational cues. In our study, the FP wore a bright retro-reflective vest, but in a more general scenario, it could be difficult to identify the FP in a crowd of people or in a busy scene.

**Reframing Directions for the FP.** Assuming the FP was in view, the next challenge was that the DP and FP had completely different frames of reference: the FP had an egocentric view from the ground, while the DP had a view from the air without necessarily being oriented in the same direction as the FP. While DPs understood this difference, it was sometimes difficult for them to remember it while deeply focused on completing a task. For instance, in G1, the DP occasionally told the FP to move straight up or down, but this was in reference to the screen space of the drone's view rather than asking the FP to move forward or backwards (or up into the air, for that matter).

A common workaround was for the DP to try to provide directions in the FP's frame of reference. In the following vignette during Task 1, we see that the DP continually provides directives to the FP. The DP is actively watching the direction that the FP is facing, and providing instructions based on that orientation:

> [**G5**] DP: There should be one... like to your... wait... turn around... wait... turn around slightly... oh yeah, there should be one in that vicinity.
>
> FP: Like, over here?
>
> DP: Like to your... your right
>
> FP: Like this direction?

The DP's directions are framed in the context of the FP. Each time the DP pauses, she carefully examines the visual scene for FP's orientation, and provides timely information about when to turn. Thus, directions were necessarily piecemeal, and depended on a continual stream of feedback. Here, the DP was unable to provide a coherent set of instructions ahead of time, and instead needed to monitor the FP's position and direction, and then provide just-in-time directions to the FP. DPs were thus constantly engaged in a monitoring sub-task rather than being able to do other things, because doing so might mean losing perspective of the FP's location and the FP's relationship with the destination.

Unsurprisingly, FPs would ask questions in reference to themselves, again requiring the DPs to watch and monitor the FP from drone view:

> [**G6**] DP: Is it this one? Can you see me here?

**Using the Environment as a Referential Resource.** Where there were useful landmarks, participants would find and use the environment to support navigation. The utility of this approach was that once the language and the frame of reference was established externally, it did not become a problem if the FP or target went out of view. Here, G7-DP used some distinct trees in the background and grass height as such features:

> [**G7**] DP: From where you are, when you look across the park, can you see those really bright gorgeous fall trees?
>
> *[FP looks around.]*
>
> FP: Yes?
>
> DP: Yep! [...] Those really bright gorgeous ones.
>
> FP: Yep.
>
> DP: On the edge of the long grass there, we're gonna put two pylons down at about the same distance [apart] as the two that you just passed over there [behind you] before you got there [where you are now].

These landmarks became references to which both participants could refer. But, this workaround breaks down when the two views are inconsistent. In the following vignette, the video capture of the drone is incapable of providing the DP with a sense that the FP is walking up a hill. Instead, because the grass here is recently mowed, it appears "flat":

> **[G5]** DP: Walk across the flat part. Walk across the flat part! The flat part! Turn 90 degrees and go left! No the other way! The flat part, the flat part!
>
> FP: Up? [...] You said flat part, this is the opposite of flat—this is a hill!
>
> DP: Stop, okay, stop. Okay, where are you right now? *[Drone has flown out of view.]* Walk down the flat part of [the hill].
>
> [...]
>
> DP: Walk towards the box. Wait, not that close! Not that close!

Later in the same sequence, the DP opportunistically uses a box in the field as an external reference. The FP moves quickly toward the box, as it is a concrete, grounded target, only to have the DP tell him to back off, as it is not actually the target—just an intermediary for the purpose of communication.

#### Limited Drone Control

DPs were motivated to "look around" from the drone for two reasons: first, DPs looked around to explore the space; and second, DPs looked around to try to locate the FP when the FP was no longer in view (as discussed above). The Elizabethan-Collar control strategy, which limits the drone's view perspective, proved frustrating to our participants. While participants understood the limited view cone of the drone view, we nevertheless observed DPs tapping on the keyboard furiously to try to see more of the scene. When trying to provide navigational cues, not being able to see the FP was frustrating for DPs—two of whom (G3 and G5) resorted to shouting, "where are you,"—especially when they knew that the FP was immediately under the drone itself.

### 4.3.2 Experiences with the Drone as an FP

The drone itself was noticeably noisy. While most FPs reported feeling comfortable around the drone, some FPs reported feeling somewhat disturbed by the noise:

> **[G5]** FP: When you hear it getting louder, you're like, 'Is it coming in for the kill?' But once you look back and realize it's not that close, it [feels] okay. It kept its distance.

Other FPs felt a certain sense of responsibility for the drone, knowing that their movements on the ground were what moved the drone around. In this sense, because the drone's position was influenced by the FP, the FP had an increased stake in staying focused on the task and his/her partner's needs as opposed to losing interest in the task or moving to more loosely-coupled forms of collaboration.

> **[G4]** FP: I felt a bit responsible for it, in terms of making sure that it flew properly. That is, I was worried it might fly where it shouldn't.

Some DPs even noticed this sense of concern in their partner. Others expressed concerns over the safety of the drone and would suggest certain movements to their partners to minimize potential damage to the drone.

> **[G4]** DP: It's going for the tree! Can you just walk [quickly] away from the tree?

In one case, the DP makes explicit mention of this concern in a comment to the FP:

> **[G7]** DP: I think the one question we should ask [to the experimenter] is how to safely land the drone. *[DP lowers the altitude of the drone to the initial setting, thinking that this will keep things safe.]*

FPs also became concerned when the drone flew too far away from them—this likely stems from the feeling that the FP lost the drone (and perhaps by extension, the DP) or that the drone is wandering off.

In our study, the investigator in the field had the ability to override the drone in case something went wrong. In a real-world scenario, this type of intervention likely would not exist—or if it did, it would likely not be in the hands of a third person. In our study, even though the investigator was fully responsible for the drone (and participants were aware of this), FPs still felt some sense of responsibility and concern for the drone. This sense of responsibility would likely increase in a real-world situation where the drone (semi-) autonomously follows the FP without any direct intervention from a third person.

Because the drone was originally designed for the DP, we had neglected to fully consider the experience of the drone from the FP's perspective, or even of it as a separate entity. We see now that the FPs are very much aware of the drone's presence, even leery of its presence, and that both FPs and DPs were concerned about the flight path of the drone.

### 4.3.3 The Drone as a Shared Resource

Rather than strictly using the drone as the DP's view into the activity space, pairs also utilized the drone as a shared resource. This is a reversal of the roles that have been reported in the literature, where the field person needs to work at the behest of the desktop person (e.g., [21, 27]).

#### Position, View, and Attention

Because the drone movement was autonomous (and out of the direct control of either participant), participants devised workarounds to "control" the drone. In this vignette (during the warm-up activity), the FP explains out loud what he is doing to the DP and continuously asks the DP if she has a good view:

> **[G3]** FP: Can you see me? [...] I [am making] slow movements, so that the drone has to follow me. [...] It's very close to me now, so I'm just going a little away from the drone. [...] Are you able to see the distance between me and the drone which follows me? [...] Is it okay now?

It was in the interests of both the DP and the FP to make sure that the drone was providing a good camera shot. Several FPs regularly checked during the tasks to make sure that their partners had a good view:

> **[G1]** FP (while looking at the drone): Do you see [the pylons]?

[**G3**] FP: I'm near the first pylon. Are you able to monitor?

[**G4**] FP: Do you see the drone view?

DP: Yeah.

FP: Okay, you should be able to see [the pylon] from the drone view.

FPs also moved around the environment to guide the drone's position and orientation:

[**G8**] FP: I'm behind the drone. *[Drone begins rotating.]* So now the drone rotates, and you can see me here. *[FP waves to the drone while walking forward (Figure 4.7).]*

FPs would also guide the DP's actions by suggesting that the DP adjust the drone's altitude, following distance, or camera position. While doing so, FPs would also often talk about what is (and what is not) visible in the video frame:

[**G4**] *[The FP watches the drone view as the DP adjusts the pan/tilt of the drone camera to find the FP. A few seconds later, the FP is in the drone video frame.]*

FP: That's me. *[Referring to himself in the video frame.]*

[**G8**] FP: Can you move the drone up to see more?

FPs also commonly performed other physical actions (e.g., waving, as seen in Figure 4.7; gesturing; and pointing, as seen in Figure 4.5) in front of the drone camera, sometimes while watching the drone video feed, to draw attention to themselves or to other objects in the video frame. This act requires the DP to orient the drone camera properly to see both the FP and the target, so much so that in some cases, the FP would move to actively ensure that the DP had a good view of him. The following vignette illustrates how FPs were cognizant of the drone's view:

[**G8**] FP: I think if you go lower you can have a better view of me... Go lower if you can.

[...]

FP: I think the drone might see me now. Can you turn it to the right? *[DP turns drone camera.]* A little bit down. I should be visible. Yeah, can you see me? I'm in the bottom left.

Figure 4.7: The FP waves back at the drone while walking forward, in order to draw attention to himself.

In some instances, DPs explicitly requested FPs to move around the environment to guide the drone's position and orientation and provide better camera shots:

> **[G4]** DP: Can you walk the opposite way... so [the drone] turns around?
>
> **[G7]** DP: Since the drone faces you, do a little bit of walking around in a circle, so I can figure out where you are on the map that I have.
>
> **[G7]** DP: I don't see pluses in my view... if you turn around [I can look].

In another instance with Group 1, the DP attempted to direct the FP to a pink hoop. The FP pointed towards a possible pink hoop just as the drone flew out of view. The DP thus needed to turn the camera down to try and keep the FP in view. From here, the DP instructed the FP to keep moving in that direction so the drone would eventually reorient on him.

FPs Making Use of the Drone's View

The setup of the system and the tasks were designed so that the DP would watch the drone-camera view and guide the FP based on that view. Even though we designed the setup with this in mind, we also gave the FP the ability to view the drone-camera view. In some instances, the FP made use of the drone view to decide what actions to take next. For example, G5-FP did this during Task 2:

> **[G5]** FP (after switching to the drone view): Oh damn! [...] Let me see where [the lined hoop] is. *[Looks at the drone view to see where to go, then walks toward where he thinks the correct hoop is.]*

Although this happened in some instances, in general FPs only had the drone-camera view open on the phone interface 17% of the time. FPs mainly did this at the beginning of the session to see themselves, but rarely used this view for orientation. One reason participants gave for not using the view was that, because studies were run on sunny days, the glare from the sun made it difficult to see the screen itself. The resolution of the phone's screen and the amount of small details in the drone view (due to it being up in the air capturing objects from afar) also meant that it was hard to see very much from the drone view on the phone screen.

Sharing Experiences through the Drone's View

The drone and its camera view were also utilized as a means of 'sharing experiences.' This was seen a lot particularly with Group 7—here, the FP and DP would often both look at the drone view together in order to take in the same scenery and experience the environment from the same perspective. For example (during the warm-up task):

> **[G7]** FP: Let's see what the drone sees. *[Switches to drone view.]* Oh, that's so cool! [...] Oh, this is so cool!
>
> *[DP adjusts the camera and altitude to keep FP in frame.]*
>
> DP: Can you see yourself?
>
> FP: Yeah!
>
> DP: On the bottom right.
>
> FP: I'm so little, it looks like I'm in a movie!
>
> *[DP moves the camera around to view the scenery while FP watches. They both converse about the scenery.]*

At one point for Group 7 during Task 1, the drone lost connection with the FP's phone and began hovering in place while the FP walked to the other end of the park. The investigator restarted the app on the FP's phone and reconnected the FP to the video chat. As soon as the FP reconnected, the drone began to fly back to the FP:

> [**G7**] DP: So cool, I'm coming to find you right now!
> FP: Oh, I wanna look at the view! *[Switches to drone view.]*
> FP: Ah! [...] It's like I'm the star of a movie!
> DP: You are the star of a movie!

#### 4.3.4 Comments about Potential Applications

In the questionnaires, participants were asked to imagine potential applications where a video-conferencing system like this could be useful. Participants came up with several specific ideas, which were then broken down into four broader categories: (1) *showing, seeing, and visiting*; (2) *instruction and guidance*; (3) *collaborative investigation, inspection, and search*; and (4) *sharing experiences.*

**Showing, seeing, and visiting.** This simply involves the use of drone-based video conferencing to show things to a remote person and to allow a remote person to explore and see things in the space, and in some sense to "visit" the space. Within this category, participants mentioned that drone-supported video conferencing could potentially support activities such as (but not limited to): an FP giving a DP a tour of an environment (e.g., as in [69]), a construction worker remotely conducting a site visit, and students remotely attending a field trip.

**Instruction and guidance.** This simply involves a remote helper guiding or instructing a field worker in completing some task. This is analogous to a lot of previous research in mobile video conferencing (e.g., [6, 8, 9, 10, 13, 29, 30]). In addition, Study Task 1 falls within this category. Within this category, participants mentioned that drone-supported video conferencing could potentially support applications such as (but not limited to): event management and setup, training of field employees, construction and repair of large objects and structures, and management and control of people (e.g., guiding or lining up people for a parade, formation, etc.).

**Collaborative investigation, inspection, and search.** This simply involves both the DP and the FP working collaboratively to investigate, inspect, and search for things within a space. Study Task 2 falls within this category. Within this category, participants mentioned that drone-supported video conferencing could potentially support applications such as (but not limited to): crime-scene investigation, disaster-scene inspection, and search and rescue.

**Sharing experiences.** This simply involves an FP, who is engaged in some activity (usually leisurely or recreational), using video conferencing to 'share' the experience of that activity with someone who is remote (often to try to get them involved in the activity in some way or to at least understand the experience). Past work has explored the use of video conferencing for sharing experiences with someone who is remote—for example, within the context of activities such as kids' sporting events (e.g., soccer games, swimming) [19], shopping [27], dog walking [27], major life events (e.g., weddings, graduations) [38], and travelling [69]. Within this category, participants mentioned that drone-supported video conferencing could potentially support applications such as (but not limited to): outdoor games, barbecues, get-togethers, hiking, golfing, camping, and sharing travel experiences.

## 4.4 Additional Quantitative Study Data

Additional quantitative data that was collected but not analyzed or discussed in great detail (including quantitative questionnaire data and mobile-phone action statistics) is included in Appendix C. In addition, a web link to some of the anonymized raw data from the system logs (user-interface logs and location/bearing logs) is also included in Appendix C.

## 4.5 Conclusion

In this chapter, I described the observational study that was conducted in order to understand how people would make use of drone-based video conferencing to communicate, collaborate, and share experiences across distances. The study was conducted using the prototype system described in Chapter 3, and was framed around three central questions: (1) "How do people make use of the drone view and the view from the FP's phone to coordinate activity?"; (2) "How do people manage and understand both views of the environment, if at all?"; and (3) "How do people use the control scheme, and is it sufficient?" We recruited eight pairs of participants (16 participants) to participate in three study activities using the prototype system; including a warm-up activity involving walking around and casual conversation, a setup activity involving the placement of objects within the park, and a detail-searching activity involving the search and inspection of a series of decorated hula-hoops placed around the park. The study findings indicate that, while the bird's-eye view of the drone is useful for getting an overall sense of the space and of the FP in relation to the space, participants encountered challenges in seeing details through the drone view and providing navigational and directional cues. Participants devised strategies to overcome these challenges, including making adjustments to the drone's altitude and following-distance settings, viewing up-close details through the phone view, giving directions in relation to the FP's frame of reference, and using the environment as a referential resource. FPs also felt a sense of responsibility for the drone; given that it followed them around and that they in a sense had influence on its position within the space. FPs made explicit comments about this sense of responsibility and were observed taking actions (such as moving to certain areas) to make sure that the drone would not collide with any obstacles. The drone was also utilized as a shared resource for communication, collaboration, and experience taking. FPs and DPs negotiated control of the view (e.g., by telling the other to move around a certain way or adjust the settings a certain way) in order to provide camera shots that satisfy both communicators' needs. Participants were also observed taking in shots of the scene together from the drone perspective and conversing casually about them as a means of experience sharing. Lastly, participants made comments about many potential applications that drone-based video conferencing could support, including some that fall within the contexts of *showing/seeing/visiting*, *collaborative investigation/inspection/search*, *instruction/guidance*, and *sharing experiences*.

This study addresses Thesis Question 2: "What are some of the opportunities presented by supporting mobile video conferencing using drones?"; as well as Thesis Question 3: "What are some of the interaction challenges presented by drone-supported video conferencing?" By addressing these questions, the study also addresses the overarching question I presented at the beginning of the thesis: "How can we design mobile-video-conferencing technologies that better support active back-and-forth collaboration between peers?" It serves as an early investigation into the challenges, opportunities, and workarounds related to drone-based video conferencing; and it provides insight into one way a mobile-video-conferencing interface can be designed to support active back-and-forth collaboration between peers. While it does not explore all possible situations and challenges that might occur (e.g., potential long-term fatigue, challenges of operating in crowded or heavily-populated areas), it does provide a first look at the many potentials of this means of communication and the numerous challenges that need to be addressed. This study and its results have a number of implications in terms of technology design. These implications will be discussed in detail in the next chapter.

# Chapter 5

# DISCUSSION, CHALLENGES, AND OPPORTUNITIES

The goal of my research was to explore one possible solution for addressing a fundamental problem in conventional mobile-video-conferencing systems: that the desktop person does not have direct control over his/her view into the activity space. By using a drone, my design provides the desktop person with the ability to explore the remote environment (to a limited extent) from an elevated perspective. While my study participants thoroughly enjoyed the experience, my analysis reveals design implications, challenges, and opportunities that can inform future designs. These will be discussed in this chapter. This chapter will revisit all three thesis questions: (1) “How can we design mobile-video-conferencing technologies that provide a remote individual with fluid and flexible control over his/her view of the environment?”, (2) “What are some of the opportunities presented by supporting mobile video conferencing using drones?”, and (3) “What are some of the interaction challenges presented by drone-supported video conferencing?”

## 5.1 Challenges of Multiple Frames of Reference

A decoupled camera essentially guarantees that the FP and DP will have different frames of reference into the environment, and this introduces a number of challenges as was seen in my study and in related work (e.g., [48, 63, 64]). Designers should take into account these challenges while designing video-conferencing systems with decoupled cameras.

In particular, matching objects and people between two camera views is challenging when the two views are from different angle and scale perspectives (e.g., Figure 5.1). Some objects

Figure 5.1: Two synchronized views into the same activity space. If the two views are from different perspectives, matching information between the two views is not always trivial. The annotations indicate the same objects appearing in the two views.

might be visible in one view, but not the other. In addition, some objects might be visible in both views, but appear differently in the two views. It takes cognitive effort to figure out how the two views, and the information contained in them, relate to each other—or where one viewing perspective is in the environment in relation to the other. Within the context of on-the-go video conferencing, as was demonstrated in my study, this creates two notable difficulties: (1) knowing where the FP is; and (2) as the DP, giving directions that make sense to the FP.

There are a number of potential technology solutions for establishing the FP's position and orientation within the view and activity space—for example, placing a marker on the video feed showing where the FP is, giving the DP the ability to 'snap' the view orientation to the FP, displaying a map on the screen showing the positions and orientations of the FP and drone within the space, and showing the compass bearings of the camera views within each of the video feeds. Figure 5.2 illustrates some of these potential features in a static UI mockup.

In my study, participants made numerous references to established landmarks, such as a box placed in the field, pylons placed down by the participants, and specific trees within the park. While it often took some time for the FP and DP to establish and agree upon


Figure 5.2: An interface mockup illustrating some potential features that could address the challenges of establishing orientation and finding objects in the activity space. The features shown in the mockup are: an annotation highlighting the FP's location in the drone view, a button for snapping the drone view to show the FP, a map showing the locations and orientations of the drone and the FP, a compass indicator in the drone view indicating the orientation of the camera, and world-stabilized pins on the map and in the drone view indicating certain objects and points of interest.

these landmarks, they proved to be useful for the purpose of navigation. Future technology designs could help support this type of landmark establishment—for example, by giving communicators the ability to place virtual stabilized marker pins within the scene (e.g., on a map of the space or within one of the video views—illustrated in Figure 5.2), which could then be viewed by both communicators through (for example) augmented reality. This approach is very similar to one taken by designs for tools that support remote assistance on physical tasks (e.g., [6, 9, 10])—I expand the concept to larger activity spaces.

## 5.2 Communication and Information Retrieval

### 5.2.1 Phone View as a Complement to the Drone View

When a view from a drone camera and a view from a phone camera are both available to the DP, both views can complement one another. They can each be used to provide separate, yet important pieces of information for giving the DP sufficient situational and environmental awareness. These two pieces of information are: overview and detail (or focus and context). The phone camera can be utilized to show up-close details of something in relation to an up-in-the-air overview shot provided by the drone showing where the FP and object are in relation to the scene. Collaborators communicating through video conferencing while on the go strive to provide both overview and detail shots [21]. While phone cameras are typically good at showing details of something up close, phone-camera operators often have to move themselves and the phone around manually to provide good overview and contextual shots; and oftentimes these attempts fail to provide the remote person with anything meaningful [21]. In my study, when FPs used the phone camera, it was often to show close-up details of something that could not be seen from the drone view.

### 5.2.2 Viewing Details through the Drone

It would be beneficial to provide the DP with a greater ability to view details through the drone view. While the drone view proved to be useful for getting an overall sense of the activity space, most DPs expressed frustration and difficulty in seeing certain details (e.g., hula-hoop colours) that they had wished to see with the drone due to the low resolution of its video feed and its distance from objects of interest. While the phone camera can be used to show details, it is out of the control of the DP; and so in order for the DP to be able to see more details of something in the environment that grabs his/her attention, he/she must direct the FP to that object. As has been seen in previous work [21], this can be cumbersome for both participants, and attempts at giving these kinds of directions can often be unsuccessful. Easy interventions could include increasing the resolution and FOV of the drone view, allowing the DP to fly the drone up close to an object of interest to inspect it, or integrating a camera with the capacity for high-resolution still shots for inspection.

## 5.3 Drone Control

If the drone's positional movement is controlled algorithmically (as it was in my study), the viewing angle should be decoupled from the movement. In other words, adjustments in the drone-camera's viewing angle should not be dependent on the drone's positional movement. This decoupling could be accomplished by, for example, forcing the drone to continue to point in the direction it was pointing previously, even if it repositions itself (e.g., if the drone is initially facing away from the FP while the FP walks away from it, it follows the FP while still facing away from him/her by propelling itself backwards). In my study, the drone provided a view that was coupled with its movements—if the drone turned 180 degrees, the camera moved 180 degrees with it as well. This was disruptive to DPs as this occasionally caused drastic viewing-angle changes to occur.

My design highlights the fundamental dilemma of providing the desktop person with additional controls: while it adds more flexibility and power, it also adds considerable complexity. In its current state, participants seemed overwhelmed with the task of visually searching the field despite not having to explicitly pilot the drone. Yet, there were times where the DPs' ability to explore the environment was impaired by the limited controls (both in terms of movement and viewing). This is evidenced by how DPs asked the FPs to move in particular ways to indirectly control the drone's position or view. Much like commercial airplanes, I suggest that drone piloting in these scenarios be modal—allowing the DP to either pilot the drone directly or enter a semi-autonomous mode where the drone follows the FP.

## 5.4 The Drone-Camera View as a Shared Resource

Cameras that are decoupled from the FP may easily become valuable shared resources for use by both collaborators. While my expectation was that the drone would simply be the DP's view into the environment (i.e., the DP's to use), it was surprising how some FPs actively attempted to control the DP's view with the drone. In some cases, it was to help show the DP something; but in other cases, it was clear that the FP intended to use the view from the drone for him/herself. This raises several questions: should the drone view belong to the DP, the FP, or both; and, to what extent? How should collaborators negotiate the view? Conceivably, in scenarios where the FP is using the video chat to show the DP something in the FP's environment (e.g., a landmark), the FP may want more control over the drone (e.g., to show the DP around). Past research on mobile video conferencing has yet to show the value of such view decoupling since nearly all explorations have focused on cameras held or worn by the person in the activity space. In this work, I investigated what FPs and DPs do when they each share control of the same decoupled camera. As future work, it would be interesting to investigate (for example) what FPs and DPs would do if they were each given their own decoupled camera to control.

## 5.5 Elevated Roles for the DP

While the drone view has proven to be useful as a utilitarian tool for collaborating on the activity at hand, it has also been proven to be useful as a social tool for sharing experiences. Group 7, in particular, demonstrated this. Here, the DP did what she could to operate the drone camera to provide differing shots of the park and of her partner in the field. While doing so, she conversed with her partner casually while the two took in the shared view of the environment. This demonstrates a social role that DPs participating in activities like this could play—the role of the cinematographer. It is quite common at social gatherings for there to be one or a small handful of people with cameras taking photos and videos of the activity going on and sharing them with the rest of the group later on. It is not that difficult to imagine in the future a similar thing happening through video conferencing—for example: two people communicating remotely through a system like this, with one person out 'in the world' engaged in some recreational activity (e.g., hiking, cycling, kayaking), and the other person at a desk communicating socially with the field person and acting as the ‘camera person’, exploring the scene, taking shots of the activity at hand, and sharing those shots with the person in the field.

In addition to leisurely activities, participants mentioned that a drone-based system like this could easily support the DP in playing roles in collaborative work-related activities where having an elevated view may be beneficial. For example, a DP could play the role of a site overseer (e.g., for building construction or event setup) who watches over a job site from an elevated perspective and guides/instructs workers on where to go based on the information that he/she receives from his/her point of view. In addition, search and inspection tasks (e.g., search and rescue, disaster-scene inspection, etc.) could benefit from having one or more collaborators searching from different scales. While drones can be used for these purposes by collocated operators, they can also create potential opportunities for workers to play these types of meaningful collaborative roles without having to be physically in the space.

## 5.6 Being Collocated with the Drone

Balancing the size of the drone and its proximity to the FP are serious considerations to ensure the comfort of the FP. FPs were generally more concerned about the drone and its safety than for their own safety. This closely matches what has been seen in previous work investigating people’s interactions with drones [2]. As Mueller and Muirhead [42] point out: the larger the drone is, the more attention it draws. If the drone is too large, it causes a distraction. On the other hand, if it is too small, it becomes less noticeable. Similarly, if the drone flies too close to the FP, it becomes a distraction (and in addition, a safety concern). On the other hand, if it flies too far from the FP, then the FP becomes concerned about whether or not he/she has lost the drone (and by extension, the DP).

Awareness of safety warnings and protocols is also important. The system should communicate important safety messages, such as when the drone is taking off and landing, when its battery is running low, when it is too close to obstacles, and when it is malfunctioning. These are things that participants were evidently concerned about.

## 5.7 Limitations

While the drone's view provides a unique and useful view into the scene, current technologies limit the quality of the video feed. Current bandwidth limitations, along with the quality of the on-board camera (and limited zooming capabilities) mean that the quality is necessarily impaired. Many participants were surprised by the limited resolution and frame rate.

I also recognize that my tasks do not explore all possible situations and challenges that one might experience when using a drone for mobile video conferencing. For example, based on these tasks, I am unable to comment on DP fatigue from long-term navigation of the drone, or whether the novelty of video conferencing with a drone wears off for either participant. Beyond this, due to statutory limitations, I am limited in my ability to explore scenarios where the FP may be in heavily populated areas, or interacting with others. Finally, improvements to drone technologies in time will allow them to operate under less than ideal weather conditions (rain, high wind, etc.).

## 5.8 Future Work

Opportunities for future work fall into the following four categories: (1) evaluating different control strategies, (2) investigating different uses and setups of the drone, and (3) investigating use in specific scenarios.

### 5.8.1 Evaluating Different Control Strategies

A straightforward recommendation for future work would be to further investigate the balance of control of the drone and its camera view between the FP and DP. In Chapter 3, I presented five different drone-control strategies; but I only tested one of them. Participants' actions, behaviours, and strategies might change significantly depending on the control strategy being used. In addition, perception of the environment and ability to focus on the task might also change, as well as perceived presence, immersion, and how the FP perceives the drone (either as a tool for the FP, a tool for the DP, or an embodiment of the DP). Several research questions are open to investigation—for example: would FPs feel as concerned or responsible for the drone if they have less control of it? Would DPs get a better understanding of the space and of the relationships between the drone's and the FP's frames of reference if he/she has more control of the drone? A different control strategy could provide a solution to some of the challenges seen in the study, while also bringing about other challenges and insights. My work provides some vision and skepticism on what these challenges and insights might be; but has also only begun to scratch the surface.

Before programming other control strategies into the drone or testing them in the field, it would be completely reasonable to test in a game- or simulation-like environment (e.g., using a game engine, or programming camera movements inside of an actual game). Such a setup could involve two users—with one controlling a character in the game environment—collaborating on some in-game task, while the camera in the game is adhering to the control strategy being tested. This allows for risk-free and easy testing and comparison between multiple strategies without the challenges and limitations brought about in the real world—challenges such as potential GPS inaccuracies, winds, and weather conditions. In addition, the game environment can be built to mimic any environment and scenario, including those that cannot be tested or recreated in the real world (e.g., due to costs or inaccessibility). While this means of testing does take away much of the real-world experience, it is a safe and easy way to perform early testing and gain early insights on other control strategies and methods. Testing in a simulation before testing in the real world is already a very common approach taken in a variety of domains, including physics experimentation and aerospace engineering—it can also be done in testing drone-video-conferencing setups.

### 5.8.2 Investigating Different Uses and Setups of the Drone

In my study, there was one drone paired with a smartphone carried by the FP, and the FP talked to the DP through the phone speaker and listened to him/her through earbuds plugged into the phone. Other setups potentially worth investigating exist as well.

For example, some participants barely made use of the phone at all, other than for speaking and listening to their partners (i.e., no use of the screen or camera). It is easy to imagine providing a setup with just the drone as a main piece, where the FP speaks and listens to the DP through an earbud/mini-microphone piece, and where the phone is just a device in the FP's pocket acting as a relay for the audio channel and a GPS beacon for the drone to follow. Indeed, a phone may not even be required—if a drone with cellular-network capabilities is built, that drone can just communicate with a beacon carried by the FP that provides the audio relay and the FP's GPS position. If the drone is completely controlled by the DP, GPS might not even be required anymore. Attaching a speaker to the drone and allowing the FP to hear the DP's voice from the drone is another possibility—although, this would likely not work well in loud environments (e.g., with ambient noise or wind), might cause a disturbance to others in the space, and would likely ensure that any conversation between the FP and DP is no longer private.

Another possibility is to provide a setup with more than one drone. For example, rather than the FP and DP using a single drone as a shared resource, they could each be provided with their own drone to use as their own camera view. In addition, multiple autonomous drones could provide different camera angles—e.g., one up close and one far back, one from the front and one from the back, one focused on person A and another focused on a second person (person B) in the field, or one focused on the FP and another focused on an object that needs to be closely watched. Such a setup could also allow a DP to toggle between different viewpoints within the space (e.g., by clicking on points on a map) more quickly and with greater ease than if he/she only had a single drone and had to fly that drone from one viewpoint to another. Testing multi-drone setups is another thing that can be done in a game- or simulation-like environment before testing in the real world.

### 5.8.3 Investigating Use in Specific Scenarios

There are numerous potential applications where a drone-video-conferencing system could be useful. Some scenarios that come to mind, based on participants' thoughts and our explorations, include (but are not limited to): sharing experiences (e.g., nature hikes, sightseeing, touring, playing games), conducting worksite visits (e.g., of construction sites), assisting with instruction or guidance (e.g., for event setup, or for constructing or repairing large objects), and inspecting or searching a large site (e.g., crime scenes, disaster scenes). My study has only partially investigated a small subset of these potential applications. It would be useful to gain further insight on how drones would be used for video conferencing within these other contexts, and to understand how a drone system could be designed to support such scenarios. I therefore recommend, as future work, studying the use of drone-supported video conferencing within these specific real-world contexts and with real-world users (i.e., domain experts, or users already familiar with the activity at hand). Deploying such a system out in the field (e.g., to a company who would use the system for work duties, or to a family who would use the system for family communication) is an effective way to do this. The problem is, with the current average battery life of a drone being too low to justify prolonged use, and with there being various legal and statutory issues surrounding the use of drones in public spaces by untrained people, this type of deployment would not be feasible to run today. However, as technologies improve, laws and regulations evolve, and interest in this style of video conferencing potentially increases, it might become feasible to run such a deployment in the future.

## 5.9 Conclusion

In this chapter, I discussed the design implications, challenges, and opportunities that were revealed through my explorations of drone-supported video conferencing. I discussed challenges in controlling/guiding the drone view, using multiple frames of reference, maintaining awareness of one's partner, and maintaining awareness of the layout and contents of the activity space. In addition, I discussed design solutions that could potentially address these challenges. I also discussed the various ways in which participants used the drone, both as a tool for each of their own needs, as well as a shared resource between the two; and I discussed implications and raised further research questions and directions based on this discussion—e.g., how should collaborators negotiate or share control of the different camera views? How can we design to support efficient and meaningful view sharing between collaborators? I discussed the potential new roles that collaborators could play with the drone, mentioned the implications of having a user collocated with the drone (and the design considerations to take as a result), and the limitations of the technology and the study. In addition, I gave various recommendations for future work, such as investigating different control strategies, investigating different uses and setups of the drone, and investigating use in various specific (real-world) scenarios.

The results of this work provide insight on how people interact with each other remotely through video conferencing when provided with an elevated view. They also provide insight into the challenges that people face, the elevated abilities and opportunities that communicators are provided with, and ways in which we can better design this type of system (and indeed, any video-conferencing system for outdoor communication) to support users' communicative and collaborative needs. However, this work also opens up many other exciting avenues and research questions worth investigating—some of which would be difficult (or not feasible) to investigate now, but given time (with technological improvements and further evolution of drone laws and regulations) would become feasible and worthwhile to explore in the future. This chapter revisited all three thesis questions: (1) “How can we design mobile-video-conferencing technologies that provide a remote individual with fluid and flexible control over his/her view of the environment?”, (2) “What are some of the opportunities presented by supporting mobile video conferencing using drones?”, and (3) “What are some of the interaction challenges presented by drone-supported video conferencing?”

# Chapter 6

# CONCLUSION

In this thesis, I presented the concept of video conferencing using drones. This concept is based on the idea that drones could give remote users greater freedom to explore an outdoor environment alongside a friend, family member, colleague, or other person with whom he/she is communicating, collaborating, and/or sharing an experience with. I presented a design space of various strategies for controlling a drone in video-conferencing scenarios, I developed a prototype drone-video-conferencing system, and I ran an observational study outlining the various opportunities, challenges, and design implications of drone video conferencing.

In this chapter, I will revisit and reflect on the thesis questions and contributions raised in the first chapter.

## 6.1 Thesis Questions

**Thesis Question 1:** *How can we design mobile-video-conferencing technologies that provide a remote individual with fluid and flexible control over his/her view of the environment?* To address this question, in Chapter 2 I discussed various research outlining specific challenges that people have when communicating or seeking information through video conferencing. In Chapter 3, I presented a design solution, involving the use of a drone, that attempts to address those design challenges. I also presented a drone-control design spectrum, outlining various approaches to addressing the design challenges; with the idea that not all approaches are appropriate for all situations, but some approaches are better than others in certain situations, and that each approach provides its own benefits and drawbacks.

**Thesis Question 2:** *What are some of the opportunities presented by supporting mobile video conferencing using drones?* To address this question, I ran an observational study where pairs of participants communicated with each other and collaborated on setup and search/inspection tasks through a prototype drone-video-conferencing system. This study revealed an early look into the advantages of this type of communication—advantages which go beyond even what is available in physical co-presence [17]. The drone provides the remote user with elevated abilities (beyond those of traditional mobile video conferencing with just a phone) and an up-in-the-air overview perspective of the activity space which can prove to be beneficial in activities where being able to look into the space from different scales is important. Drone-video-conferencing has the potential to give remote users greater ability to more meaningfully contribute to activities that involve searching, inspecting, and instructing others in activities where being able to explore one's view into the space from different perspectives and scales is necessary. It also has the potential to contribute to shared experiences. As a shared resource, the drone can be used by the remote user to show things to the local user, and additionally as a tool for the local user to peek into the view to see what their partner is looking at from his/her perspective. It is possible that there are other potential benefits that have yet to be seen or explored—such as that of the drone being considered an embodiment for the remote person. These benefits could be explored and revealed with time, if different control strategies are explored and tested, and drone-supported video conferencing is tested long term and in various real-world situations. My work has only just begun to scratch the surface in the potentials of this style of communication. It will be exciting to see what can be done in the future.

**Thesis Question 3:** *What are some of the interaction challenges presented by drone-supported video conferencing?* My study also revealed a number of different challenges that drone-supported video conferencing brings, both to the user in the field as well as to the user at the desk. These challenges included: difficulties in getting the drone to do what the desktop user wants, challenges in matching between different views into the activity space, challenges in translating navigational instructions from the perspective of the desktop user to that of the field user, as well as safety concerns with regards to trying to keep the drone a safe distance from the field person and from obstacles. Various technical and statutory limitations, such as low battery life, occasionally low bandwidth, and restrictions on where drones can be operated in public, were also an issue, and may continue to be an issue for the time being. The design challenges presented also revealed a number of design solutions that could be implemented in drone video conferencing to better support the needs and intents of both communicators. In addition, the design challenges also revealed potential weakness areas of the Elizabethan-Collar control strategy which was tested in the study—areas which could potentially prove to be strengths of other control strategies, if they are tested in the future.

By answering the three thesis questions, my work also answered the overarching question I presented at the beginning of the thesis:

> **How can we design mobile-video-conferencing technologies that better support active back-and-forth collaboration between peers?**

Drone-supported video conferencing is simply one means of designing a mobile-video-conferencing setup that better supports active back-and-forth collaboration between a person on the go and a person at a stationary location. It promotes active participation, information seeking, instruction, and guidance from the DP, and it makes these actions much easier for the DP in comparison to (for example) a traditional setup where the only view into the activity space is that from the FP's phone. While the drone setup in this thesis does also introduce interaction challenges, those challenges provide additional insight on how a mobile-video-conferencing interfaces promoting active back-and-forth collaboration can be better designed to support users' goals.

## 6.2 Contributions

As discussed in Chapter 1, this thesis provides the following contributions:

**Contribution 1:** *The introduction of an early design space for drone-based video-conferencing systems that outlines several control strategies for drones in out-and-about mobile-video-conferencing scenarios.* This design spectrum, presented in Chapter 3, outlines a set of drone-control strategies along two axes: level of position control, and level of view control. Five strategies are presented—in order from least amount to most amount of control to the desktop user, they are: *Guided*, *Free View*, *Elizabethan Collar*, *Leashed*, and *Free Roam*.

**Contribution 2:** *The design and implementation of a drone-video-conferencing prototype system that follows a particular control scheme in my early design space.* This prototype, presented in Chapter 3, follows the Elizabethan-Collar control strategy, and was implemented in various pieces: a Parrot Bebop drone, a iOS smartphone app (developed in Objective-C, Swift, and C), a Node.js server, and an HTML5/JavaScript web desktop app.

**Contribution 3:** *The first study of drone-supported video conferencing, where I outline new communication and interaction challenges that augment existing human-robot interaction (HRI) awareness and human-UAV awareness frameworks.* This study and its findings were presented in Chapter 4. The implications of the study, including those for design, as well as the opportunities and challenges that this style of video conferencing present, were discussed in Chapter 4.

**Contribution 4:** *A public codebase upon which others can build drone-based video-conferencing setups.* My prototype was developed using open-source and readily-available software, hardware, and libraries. The full source code of my prototype can be found here: `https://github.com/BrennanJones/DroneVideoConf`.

## 6.3 Closing Remarks

Many outdoor activities can be made more enjoyable when we share them with others. In addition, some outdoor activities such as field setup, searching, and navigation may require collaboration with other people—some of whom might not be physically present. While conventional mobile-video-conferencing technologies allow us to communicate, collaborate, and share experiences with others while on the go to some extent, mobile users have to perform a large amount of camera work, which may not provide good views for the remote person. In this thesis, I introduced a style of mobile video conferencing provides the remote user with the ability to explore the environment from the elevated perspective of a drone. I then presented the results of a study that demonstrated that, while the experience still has far to go, drones provide remote users with a perspective that is enjoyable and useful. This, in turn, suggests new opportunities for mobile video conferencing where cameras can be decoupled from participants in the activity space to alleviate challenging camera work and give more freedom to the remote participant. While there is still much to be explored in this domain, and while the technology and regulations still have far to go before it can become feasible or justifiable to video chat in this way, this work shows that drones can potentially be a very useful tool in many outdoor video-conferencing activities, both work-related and leisurely. It has the potential to elevate a remote communicator's role, from simply being an observer with a lack of control, to being a contributor with a unique perspective and unique active roles to play. The sky is the limit.

# Bibliography


[1] Jed R. Brubaker, Gina Venolia, and John C. Tang. "Focusing on shared experiences: moving beyond the camera in video communication". In: *Proceedings of the Designing Interactive Systems Conference*. ACM, 2012, pp. 96–105.

[2] Jessica R. Cauchard et al. "Drone & Me: An Exploration into Natural Human-drone Interaction". In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. UbiComp '15. New York, NY, USA: ACM, 2015, pp. 361–365. ISBN: 978-1-4503-3574-4. DOI: 10.1145/2750858.2805823. URL: http://doi.acm.org/10.1145/2750858.2805823 (visited on 09/23/2015).

[3] Jill L. Drury, Laurel Riek, and Nathan Rackliffe. "A decomposition of UAV-related situation awareness". In: *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. ACM, 2006, pp. 88–94. URL: http://dl.acm.org/citation.cfm?id=1121258 (visited on 09/23/2015).

[4] Jill L. Drury and Stacey D. Scott. "Awareness in unmanned aerial vehicle operations". In: *The International C2 Journal* 2.1 (2008), pp. 1–10. URL: http://www.academia.edu/download/30260691/10.1.1.100.6194.pdf (visited on 09/23/2015).

[5] J.L. Drury, J. Scholtz, and H.A. Yanco. "Awareness in human-robot interactions". In: *IEEE International Conference on Systems, Man and Cybernetics, 2003*. Vol. 1. Oct. 2003, 912–918 vol.1. DOI: 10.1109/ICSMC.2003.1243931.

[6] Omid Fakourfar et al. "Stabilized Annotations for Mobile Remote Assistance". In: *Proceedings of the 34th Annual ACM Conference on Human Factors in Computing Systems*. CHI '16. ACM, 2016. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858171. URL: http://dx.doi.org/10.1145/2858036.2858171.

[7] *Flight plan update is now available for Bebop Drone!* Oct. 2015. URL: http://blog.parrot.com/2015/10/21/flight-plan-update-available/ (visited on 02/12/2016).

[8] Susan R. Fussell et al. "Gestures over Video Streams to Support Remote Collaboration on Physical Tasks". In: *Hum.-Comput. Interact.* 19.3 (Sept. 2004), pp. 273–309. ISSN: 0737-0024. DOI: 10.1207/s15327051hci1903_3. URL: http://dx.doi.org/10.1207/s15327051hci1903_3 (visited on 01/12/2016).

[9] Steffen Gauglitz et al. "Integrating the Physical Environment into Mobile Remote Collaboration". In: *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services*. MobileHCI '12. New York, NY, USA: ACM, 2012, pp. 241–250. ISBN: 978-1-4503-1105-2. DOI: 10.1145/2371574.2371610. URL: http://doi.acm.org/10.1145/2371574.2371610 (visited on 11/20/2015).

[10] Steffen Gauglitz et al. "World-stabilized Annotations and Virtual Scene Navigation for Remote Collaboration". In: *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*. UIST '14. New York, NY, USA: ACM, 2014, pp. 449–459. ISBN: 978-1-4503-3069-5. DOI: 10.1145/2642918.2647372. URL: http://doi.acm.org/10.1145/2642918.2647372 (visited on 11/20/2015).

[11] Michael A. Goodrich and Alan C. Schultz. "Human-robot Interaction: A Survey". In: *Found. Trends Hum.-Comput. Interact.* 1.3 (Jan. 2007), pp. 203–275. ISSN: 1551-3955. DOI: 10.1561/1100000005. URL: http://dx.doi.org/10.1561/1100000005 (visited on 09/23/2015).

[12] Carl Gutwin and Saul Greenberg. "A Descriptive Framework of Workspace Awareness for Real-Time Groupware". en. In: *Computer Supported Cooperative Work (CSCW)* 11.3-4 (Sept. 2002), pp. 411–446. ISSN: 0925-9724, 1573-7551. DOI: 10.1023/A:1021271517844. URL: http://link.springer.com/article/10.1023/A%3A1021271517844 (visited on 01/27/2016).

[13] Carl Gutwin and Reagan Penner. "Improving Interpretation of Remote Gestures with Telepointer Traces". In: *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work*. CSCW '02. New York, NY, USA: ACM, 2002, pp. 49–57. ISBN: 1-58113-560-2. DOI: 10.1145/587078.587086. URL: http://doi.acm.org/10.1145/587078.587086 (visited on 01/12/2016).

[14] Hirohiko Hayakawa et al. "Telexistence drone: design of a flight telexistence system for immersive aerial sports experience". In: *Proceedings of the 6th Augmented Human International Conference*. ACM, 2015, pp. 171–172. URL: http://dl.acm.org/citation.cfm?id=2735816 (visited on 09/23/2015).

[15] Keita Higuchi, Tetsuro Shimada, and Jun Rekimoto. "Flying sports assistant: external visual imagery representation for sports training". In: *Proceedings of the 2nd Augmented Human International Conference*. ACM, 2011, p. 7. URL: http://dl.acm.org/citation.cfm?id=1959833 (visited on 09/23/2015).

[16] Kenichi Higuchi, Kenichi Fujii, and Jun Rekimoto. "Flying head: A head-synchronization mechanism for flying telepresence". In: *Artificial Reality and Telexistence (ICAT), 2013 23rd International Conference on*. IEEE, 2013, pp. 28–34. URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6728902 (visited on 09/23/2015).

[17] Jim Hollan and Scott Stornetta. "Beyond Being There". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '92. New York, NY, USA: ACM, 1992, pp. 119–125. ISBN: 0-89791-513-5. DOI: 10.1145/142750.142769. URL: http://doi.acm.org/10.1145/142750.142769 (visited on 09/23/2015).

[18] Hilary Hutchinson et al. "Technology Probes: Inspiring Design for and with Families". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '03. New York, NY, USA: ACM, 2003, pp. 17–24. ISBN: 978-1-58113-630-2. DOI: 10.1145/642611.642616. URL: http://doi.acm.org/10.1145/642611.642616 (visited on 11/02/2016).

[19] Kori Inkpen et al. “Experiences2Go: sharing kids’ activities outside the home with remote family members”. In: *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 2013, pp. 1329–1340.

[20] Robert Johansen. *GroupWare: Computer Support for Business Teams*. New York, NY, USA: The Free Press, 1988. ISBN: 0-02-916491-5.

[21] Brennan Jones et al. “Mechanics of Camera Work in Mobile Video Collaboration”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI ’15. New York, NY, USA: ACM, 2015, pp. 957–966. ISBN: 978-1-4503-3145-6. DOI: 10.1145/2702123.2702345. URL: http://doi.acm.org/10.1145/2702123.2702345 (visited on 09/23/2015).

[22] Tejinder K. Judge and Carman Neustaedter. “Sharing Conversation and Sharing Life: Video Conferencing in the Home”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’10. New York, NY, USA: ACM, 2010, pp. 655–658. ISBN: 978-1-60558-929-9. DOI: 10.1145/1753326.1753422. URL: http://doi.acm.org/10.1145/1753326.1753422 (visited on 01/11/2016).

[23] David B. Kaber and Mica R. Endsley. “The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task”. In: *Theoretical Issues in Ergonomics Science* 5.2 (Mar. 2004), pp. 113–153. ISSN: 1463-922X. DOI: 10.1080/1463922021000054335. URL: http://dx.doi.org/10.1080/1463922021000054335 (visited on 02/12/2016).

[24] Shunichi Kasahara and Jun Rekimoto. “JackIn Head: An Immersive Human-human Telepresence System”. In: *SIGGRAPH Asia 2015 Emerging Technologies*. SA ’15. New York, NY, USA: ACM, 2015, 14:1–14:3. ISBN: 978-1-4503-3925-4. DOI: 10.1145/2818466.2818486. URL: http://doi.acm.org/10.1145/2818466.2818486 (visited on 11/20/2015).

[25] Shunichi Kasahara and Jun Rekimoto. "JackIn Head: Immersive Visual Telepresence System with Omnidirectional Wearable Camera for Remote Collaboration". In: *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology*. VRST '15. New York, NY, USA: ACM, 2015, pp. 217–225. ISBN: 978-1-4503-3990-2. DOI: 10.1145/2821592.2821608. URL: http://doi.acm.org/10.1145/2821592.2821608 (visited on 11/20/2015).

[26] Shunichi Kasahara and Jun Rekimoto. "JackIn: integrating first-person view with out-of-body vision generation for human-human augmentation". In: *Proceedings of the 5th Augmented Human International Conference*. ACM, 2014, p. 46. (Visited on 09/23/2015).

[27] Seungwon Kim, Sasa Junuzovic, and Kori Inkpen. "The Nomad and the Couch Potato: Enriching Mobile Shared Experiences with Contextual Information". In: *Proceedings of the 18th International Conference on Supporting Group Work*. ACM, 2014, pp. 167–177.

[28] Seungwon Kim et al. "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. New York, NY, USA: ACM, 2015, pp. 1669–1674. ISBN: 978-1-4503-3146-3. DOI: 10.1145/2702613.2732838. URL: http://doi.acm.org/10.1145/2702613.2732838 (visited on 01/19/2016).

[29] David Kirk, Andy Crabtree, and Tom Rodden. "Ways of the Hands". en. In: *ECSCW 2005*. Ed. by Hans Gellersen et al. DOI: 10.1007/1-4020-4023-7_1. Springer Netherlands, 2005, pp. 1–21. ISBN: 978-1-4020-4022-1 978-1-4020-4023-8. URL: http://link.springer.com/chapter/10.1007/1-4020-4023-7_1 (visited on 01/19/2016).

[30] David Kirk and Danae Stanton Fraser. "Comparing Remote Gesture Technologies for Supporting Collaborative Physical Tasks". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '06. New York, NY, USA: ACM, 2006,

pp. 1191–1200. ISBN: 1-59593-372-7. DOI: 10.1145/1124772.1124951. URL: http://doi.acm.org/10.1145/1124772.1124951 (visited on 01/12/2016).

[31] David S. Kirk, Abigail Sellen, and Xiang Cao. "Home video communication: mediating'closeness'". In: *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. ACM, 2010, pp. 135–144. URL: http://dl.acm.org/citation.cfm?id=1718945 (visited on 09/23/2015).

[32] Sven Kratz et al. "Polly: Being there through the parrot and a guide". In: *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*. ACM, 2014, pp. 625–630.

[33] Sven Kratz et al. "Polly Wanna Show You: Examining Viewpoint-Conveyance Techniques for a Shoulder-Worn Telepresence System". In: *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI 2014), Industrial Case Studies*. Copenhagen, Denmark: ACM, Aug. 2015.

[34] Min Kyung Lee and Leila Takayama. "Now, I have a body: Uses and social norms for mobile remote presence in the workplace". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011, pp. 33–42.

[35] Christian Licoppe and Julien Morel. "The collaborative work of producing meaningful shots in mobile video telephony". In: *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 2009, p. 35. URL: http://dl.acm.org/citation.cfm?id=1613903 (visited on 09/23/2015).

[36] *Lily - The Camera That Follows You*. URL: https://www.lily.camera/ (visited on 09/23/2015).

[37] Paul Luff et al. "Fractured ecologies: creating environments for collaboration". In: *Human-Computer Interaction* 18.1 (2003), pp. 51–84. URL: http://dl.acm.org/citation.cfm?id=1466533 (visited on 09/23/2015).

[38] Michael Massimi and Carman Neustaedter. "Moving from talking heads to newlyweds: exploring video chat use during major life events". In: *Proceedings of the 2014 conference on Designing interactive systems*. ACM, 2014, pp. 43–52.

[39] Jason S. McCarley and Christopher D. Wickens. *Human factors implications of UAVs in the national airspace*. University of Illinois at Urbana-Champaign, Aviation Human Factors Division, 2005. URL: http://www.tc.faa.gov/logistics/Grants/pdf/2004/04-G-032.pdf (visited on 02/12/2016).

[40] *Meerkat*. URL: http://meerkatapp.co/ (visited on 01/11/2016).

[41] P. M. Mitchell, M. L. Cummings, and T. B. Sheridan. "Management of multiple dynamic human supervisory control tasks". In: *10th International Command and Control Research and Technology Symposium*. 2005. URL: http://www.dodccrp.org/events/10th_ICCRTS/CD/papers/176.pdf (visited on 02/12/2016).

[42] Florian 'Floyd' Mueller and Matthew Muirhead. "Jogging with a Quadcopter". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. New York, NY, USA: ACM, 2015, pp. 2023–2032. ISBN: 978-1-4503-3145-6. DOI: 10.1145/2702123.2702472. URL: http://doi.acm.org/10.1145/2702123.2702472 (visited on 09/23/2015).

[43] Jawad Nagi et al. "Human Control of UAVs Using Face Pose Estimates and Hand Gestures". In: *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*. HRI '14. New York, NY, USA: ACM, 2014, pp. 252–253. ISBN: 978-1-4503-2658-2. DOI: 10.1145/2559636.2559833. URL: http://doi.acm.org/10.1145/2559636.2559833 (visited on 09/23/2015).

[44] Tayyab Naseer, Jurgen Sturm, and Daniel Cremers. "FollowMe: Person following and gesture recognition with a quadrocopter". In: *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE, 2013, pp. 624–630. URL: http:

//ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6696416 (visited on 09/23/2015).

[45] Carman Neustaedter et al. "To Beam or Not to Beam: A Study of Remote Telepresence Attendance at an Academic Conference". In: *Proceedings of the 2016 Conference on Computer Supported Cooperative Work*. 2016.

[46] Wai Shan Ng and E. Sharlin. "Collocated interaction with flying robots". In: *2011 IEEE RO-MAN*. July 2011, pp. 143–149. DOI: 10.1109/ROMAN.2011.6005280.

[47] Kenton O'Hara, Alison Black, and Matthew Lipson. "Everyday Practices with Mobile Video Telephony". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '06. New York, NY, USA: ACM, 2006, pp. 871–880. ISBN: 1-59593-372-7. DOI: 10.1145/1124772.1124900. URL: http://doi.acm.org/10.1145/1124772.1124900 (visited on 01/06/2016).

[48] Oscar Olmos, Christopher D. Wickens, and Andrew Chudy. "Tactical Displays for Combat Awareness: An Examination of Dimensionality and Frame of Reference Concepts and the Application of Cognitive Engineering". In: *The International Journal of Aviation Psychology* 10.3 (July 2000), pp. 247–271. ISSN: 1050-8414. DOI: 10.1207/S15327108IJAP1003_03. URL: http://dx.doi.org/10.1207/S15327108IJAP1003_03 (visited on 03/18/2016).

[49] R. Parasuraman, T.B. Sheridan, and Christopher D. Wickens. "A model for types and levels of human interaction with automation". In: *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* 30.3 (May 2000), pp. 286–297. ISSN: 1083-4427. DOI: 10.1109/3468.844354.

[50] *Parrot Bebop Drone. Lightweight yet robust quadricopter - 14 megapixel Full HD 1080p Fisheye Camera - Skycontroller - 3-axes image stabilization.* URL: http://www.parrot.com/products/bebop-drone/ (visited on 09/23/2015).

[51] Eric Paulos and John Canny. "PRoP: Personal Roving Presence". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '98. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1998, pp. 296–303. ISBN: 0-201-30987-4. DOI: 10.1145/274644.274686. URL: http://dx.doi.org/10.1145/274644.274686 (visited on 09/23/2015).

[52] Eric Paulos and John Canny. "Social Tele-Embodiment: Understanding Presence". en. In: *Autonomous Robots* 11.1 (July 2001), pp. 87–95. ISSN: 0929-5593, 1573-7527. DOI: 10.1023/A:1011264330469. URL: http://link.springer.com/article/10.1023/A%3A1011264330469 (visited on 09/23/2015).

[53] *Periscope*. URL: https://www.periscope.tv (visited on 01/11/2016).

[54] Jason Procyk et al. "Exploring Video Streaming in Public Settings: Shared Geocaching over Distance Using Mobile Video Chat". In: *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*. CHI '14. New York, NY, USA: ACM, 2014, pp. 2163–2172. ISBN: 978-1-4503-2473-1. DOI: 10.1145/2556288.2557198. URL: http://doi.acm.org/10.1145/2556288.2557198 (visited on 09/23/2015).

[55] Jason Procyk et al. "Shared Geocaching over Distance with Mobile Video Streaming". In: *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. CSCW Companion '14. New York, NY, USA: ACM, 2014, pp. 293–296. ISBN: 978-1-4503-2541-7. DOI: 10.1145/2556420.2557635. URL: http://doi.acm.org/10.1145/2556420.2557635 (visited on 01/18/2016).

[56] Irene Rae, Bilge Mutlu, and Leila Takayama. "Bodies in Motion: Mobility, Presence, and Task Awareness in Telepresence". In: *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*. CHI '14. New York, NY, USA: ACM, 2014, pp. 2153–2162. ISBN: 978-1-4503-2473-1. DOI: 10.1145/2556288.2557047. URL: http://doi.acm.org/10.1145/2556288.2557047 (visited on 09/23/2015).

[57] Nobuchika Sakata et al. “WACL: Supporting Telecommunications Using Wearable Active Camera with Laser Pointer”. In: *Proceedings of the 7th IEEE International Symposium on Wearable Computers*. ISWC ’03. Washington, DC, USA: IEEE Computer Society, 2003, pp. 53–. ISBN: 0-7695-2034-0. URL: http://dl.acm.org/citation.cfm?id=946249.946870 (visited on 11/20/2015).

[58] M. Sharma et al. “Communicating affect via flight path Exploring use of the Laban Effort System for designing affective locomotion paths”. In: *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Mar. 2013, pp. 293–300. DOI: 10.1109/HRI.2013.6483602.

[59] Sphero. *Sphero controls an AR Drone*. URL: https://www.youtube.com/watch?v=wGxIUXkneWE (visited on 02/12/2016).

[60] Daniel Szafir, Bilge Mutlu, and Terrence Fong. “Communication of Intent in Assistive Free Flyers”. In: *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*. HRI ’14. New York, NY, USA: ACM, 2014, pp. 358–365. ISBN: 978-1-4503-2658-2. DOI: 10.1145/2559636.2559672. URL: http://doi.acm.org/10.1145/2559636.2559672 (visited on 09/23/2015).

[61] Daniel Szafir, Bilge Mutlu, and Terry Fong. “Communicating Directionality in Flying Robots”. In: *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. HRI ’15. New York, NY, USA: ACM, 2015, pp. 19–26. ISBN: 978-1-4503-2883-8. DOI: 10.1145/2696454.2696475. URL: http://doi.acm.org/10.1145/2696454.2696475 (visited on 09/23/2015).

[62] *Thanavath/ARDrone-x-Sphero*. URL: https://github.com/Thanavath/ARDrone-x-Sphero (visited on 02/12/2016).

[63] Lisa C. Thomas and Christopher D. Wickens. *Effects of Display Frames of Reference on Spatial Judgments and Change Detection*. en. Tech. rep. Sept. 2000.

[64] Lisa C. Thomas and Christopher D. Wickens. "Visual Displays and Cognitive Tunneling: Frames of Reference Effects on Spatial Judgments and Change Detection". en. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 45.4 (Oct. 2001), pp. 336–340. ISSN: 1541-9312, DOI: 10.1177/154193120104500415. URL: http://pro.sagepub.com/content/45/4/336 (visited on 03/18/2016).

[65] Hiroaki Tobita and Takuya Kuzi. "Face-to-avatar: Augmented Face-to-face Communication with Aerotop Telepresence System". In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*. AVI '12. New York, NY, USA: ACM, 2012, pp. 262–265. ISBN: 978-1-4503-1287-5. DOI: 10.1145/2254556.2254605. URL: http://doi.acm.org/10.1145/2254556.2254605 (visited on 09/23/2015).

[66] *Twitch*. URL: http://www.twitch.tv/ (visited on 01/18/2016).

[67] *Ustream*. URL: http://www.ustream.tv/ (visited on 01/18/2016).

[68] Gina Venolia et al. "Embodied Social Proxy: Mediating Interpersonal Connection in Hub-and-satellite Teams". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '10. New York, NY, USA: ACM, 2010, pp. 1049–1058. ISBN: 978-1-60558-929-9. DOI: 10.1145/1753326.1753482. URL: http://doi.acm.org/10.1145/1753326.1753482 (visited on 09/23/2015).

[69] *Virtual Photo Walks, using Video conferencing tools to help people be interactive citizens again. - Home*. URL: http://www.virtualphotowalks.org/ (visited on 09/23/2015).

[70] Holly A. Yanco, Jill L. Drury, and others. "Classifying human-robot interaction: an updated taxonomy." In: *SMC (3)*. 2004, pp. 2841–2846. URL: http://www.researchgate.net/profile/Jill_Drury/publication/220755236_Classifying_human-robot_interaction_An_updated_taxonomy/links/0c9605229dde7b60e0000000.pdf (visited on 09/23/2015).

[71] *YouTube*. URL: https://www.youtube.com/ (visited on 01/18/2016).

# Appendix A

# STUDY-DESIGN ITERATION

Several versions of the study and the study tasks were iterated on before the final version was decided. Changes to the study tasks were made on several occassions, because they were either too easy, too difficult, or too time consuming. Before conducting pilot studies, we discussed several task ideas—some of which we could not run due to technological and logistical limitations. In this appendix, I report on the task ideas and pilot iterations that led to the final study design.

## A.1 Early Ideas

Before running pilot studies, we discussed and considered several ideas of potential study tasks. The following is a list of the main ideas we narrowed in on, based on possible use cases of drone-supported video conferencing and possible advantages of using a drone in mobile video conferencing.

**Geocaching:** In this task, the FP would search for three geocaches spread around a field. The FP would search for these geocaches with the assistance from the DP. The DP would have on hand a map of the field—showing the locations of certain key landmarks (e.g., an important tree, a flagged area, pathways, etc.)—and clues about where the first geocache is. Each geocache would contain clues about where the other geocaches are—these clues would use the names of landmarks marked on the DP's map.

**Laying out objects:** In this task, the FP would lay out objects (e.g., tent pads, tables, chairs, etc.) around a field based on a pre-defined plan. This task would be analogous to setting up for an event (e.g., festival, sporting event, etc.). The DP would have on hand the pre-defined plan that the layout has to adhere to, and the FP would have to follow the DP's instructions to lay out the objects in adherence to the pre-defined plan.

**Search and rescue:** In this task, the FP would look for a lost puppy (fake puppy) in a park, and would have to look through bushes and/or trees (no climbing of trees would be required). The DP would have on hand statements from the puppy's owner containing clues about the puppy's last-known whereabouts. Spread around the field would be clues (e.g., fake fur, fake footprints, etc.) about where the puppy visited.

**Disaster scene inspection:** In this task, the FP would inspect a fake disaster scene, looking for bodies, survivors, hazardous materials, and other important evidence.

**Tour:** In this task, the FP would take the DP on a tour of a park, showing the DP a set of key landmarks. The DP would be taking note of these key landmarks and drawing/filling out a map of the park.

**Streaming/viewing sporting experiences:** Here, a mini soccer game (or other sporting event) would be streamed live to a participant acting as a spectator. The drone would take specific actions—e.g., following a key player, or following whichever player the participant chooses. A comparison between a standard mobile-video-conferencing setup (with just a smartphone) and the drone could be tested here, with the goal of seeing if the drone improves the experience compared to using just the phone. If no comparison is done here, another goal of this task could be to see how participants utilize the drone and take in the experience.

## A.2 Pilot Studies

We conducted four different pilot studies with four pairs of participants (eight participants) before arriving at the final study design. Based on the task ideas above, we initially narrowed down to two tasks: one based on the *laying-out-objects* task, and the other containing elements of the *geocaching* and *search-and-rescue* tasks. We narrowed down to these two tasks mainly based on the following practical reasons:

1. The drone technology we had on had only allowed for flight time of about thirteen

minutes before a battery swap is required, interrupting the activity or the participant's view into the activity. Thus, we needed tasks that could be completed within this time frame.

2. We needed tasks that required little to no specialized expertise, which could be done by most participants from the general public.

3. We were interested in how pairs would communicate back and forth between each other to coordinate activity and exchange information. We were also interested in how participants would actively inspect the environment at different scales. Thus, we wanted to include tasks that incorporate all of these actions.

### A.2.1 Pilot 1

The field location for this pilot was a park near the lab. For this pilot, Task 1 was an *event-setup* task. In this task, the DP is given instructions (in the form of a map) on how a field should be set up with physical props, and he/she must instruct the FP in placing the props around the field in accordance with the instructions. The physical props included pylons, tents, a sleeping bag, and a set of bocce balls. Task 2 was an escape-room-inspired *unlock-the-lock* task, where the DP was given clues, and the DP and FP had to inspect the environment to find digits to a lock combination. The digits were based on the placement of physical props in the space (e.g., the total number of red balls in the field, the total number of heavy balls in the field, etc.). Some features, such as the colours of props, could be observed by both the FP and the DP; while other features, such as the weight of a prop, could only be checked by the FP.

### A.2.2 Pilot 2

For this pilot, the field location was moved to a soccer field on the university campus. This was done mainly to make moving participants from the lab to the field location quicker and easier. The tasks were also modified a bit to make them easier to complete. The event-setup task was given a new layout for the new location, and the props were modified slightly—this time, the props included pylons, one tent, a bag of red balls, and a bag of blue balls. New and slightly simpler clues were used for the unlock-the-lock task. For this iteration of the task, all clues relied only on visual inspection of props, but inspection was required at different scales.

### A.2.3 Pilot 3

For the third pilot, the field location remained the same as the second pilot, but both of the tasks were changed a bit. For the event-setup task, rather than the DP's instructions on the layout of the field being in the form of a map, the DP was given six photographs showing where each pylon should be placed; but these photographs were taken from a ground perspective. The idea here is that the DP would have to inspect the environment through the visual information in the video feeds in order to figure out where pylons should be placed. The unlock-the-lock task was removed entirely, and replaced with another search-and-inspection task that we expected would make greater use of the two view perspectives together. For this task, hula-hoops of unique colours (red, pink, yellow, green) and symbols (stripe, plus, hat) through the centre were placed around the field, and participants had to visit as many hoops in the correct sequence as possible. The idea we aimed for here was that the colours of hoops would be visible only from the ground perspective (i.e., to the FP), while the symbols and locations of far-away hoops would be visible only from an up-in-the-air perspective (i.e., to the DP).

### A.2.4 Pilot 4

For the fourth pilot, we changed the field location back to the park in which we conducted the first pilot. We did this for three reasons: (1) we wanted to make use of larger field space, (2) we wanted to make use of a field space with a less obvious layout (most people have a mental model of what a soccer field looks like), and (3) we wanted to make use of a field space that contained tall grass, making it so that certain props placed on the ground would not be immediately visible from a ground perspective, but could be from an up-in-the-air perspective. We changed the study protocol so that participants would instead meet the investigators separately in the lab and field locations at the same time. For the event-setup task, we reverted back to giving the DP a map layout instead of photographs, to make the task easier to complete. The hula-hoop task did not change, except that the hula hoops were placed in tall grass, so that they would not be immediately visible to the FP but could be seen by the DP through the drone view.

## A.3 Final Iteration

After the fourth pilot, we were confident that we had narrowed down on a protocol and study tasks that would achieve our goals. The only changes from the fourth pilot to the final design were the number, colours, and patterns of the hoops laid out for the inspection task. Full details of the final study design, including descriptions of the tasks, the protocol, and our research questions and goals, are included in Chapter 4. The materials for the study, including task sheets and instructions given to participants, are included in Appendix A.

# Appendix B

# STUDY MATERIALS

## B.1 Consent Form

UNIVERSITY OF CALGARY

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

Brennan Jones, Master's Student – Department of Computer Science, 403-210-9499, bdgjones@ucalgary.ca

Dr. Anthony Tang, Assistant Professor – Department of Computer Science, 403-210-6912, tonyt@ucalgary.ca

**Title of Project:**

Drone-Supported Video Conferencing for On-the-Go Activities

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:**

The purpose of this study is to explore the use of a drone in on-the-go video conferencing. You are here because you heard about this study based on a poster, email, forum post, social media post, or word-of-mouth, and you volunteered to participate.

**What Will I Be Asked To Do?**

You will be asked to complete an activity with your partner. Your partner will be connected to you via a video conferencing system, meaning that you will have both an audio and video link to communicate with him/her. One of you will work from within our lab at a desk, and the other will work from out in the field, at a nearby public park. You will be provided with a task to complete with your partner. This task may involve speaking with one another, walking around a park, showing the your partner interesting things, collaborating to accomplish a shared goal, giving/following instructions, and learning something that your partner knows. What we are interested in is how you communicate with one another given certain technology conditions.

Note: we are *not* evaluating you. Instead, we are evaluating how different technology conditions work, and whether they work well. Thus, please relax and simply enjoy yourself while you complete these tasks.

You will also be asked to fill out two brief questionnaires and participate in a brief interview about your experiences. This study will be video recorded.

Your participation is entirely voluntary. You may refuse to participate altogether, or may withdraw from the study at any time without penalty by stating your wish to withdraw to the researchers.

This study should take approximately 1½ to 2 hours. You will receive remuneration in the form of $20 cash for your participation—you will receive this remuneration even if you choose to withdraw from the study.

**What Type of Personal Information Will Be Collected?**

We will be collecting video data. 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 clips or stills of the video 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 clips or still pictures from the exploration session and the interview. Any clips or stills of the video 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. 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

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 put a check mark on the corresponding line(s) that grants us your permission to:

| | | |
|---|---|---|
| I agree to let video to be collected from the study for data analysis. | YES __ | NO __ |
| I agree to let identifiable video clips or stills from the study to be used for presentation of the research results. | YES __ | NO __ |
| I agree to let my conversation during the study be directly quoted, anonymously, in presentation of the research results. | YES __ | NO __ |

### Are there Risks or Benefits if I Participate?

There is very minimal risk of harm associated with your participation in this research. While we expect no direct benefits to participants, this research will allow participants to experience working together using different mobile video conferencing technologies, and thus it might spark interest or curiosity. At the end of the session, participants will be able to ask questions about our research.

### 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 and destroy all data that was previously collected from you. You may withdraw your data from the study anytime until it is formally published or presented in a public forum.

All data received from this study will be kept indefinitely in a secure location. 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 stills 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.

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 (written consent)***

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

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:

Brennan Jones
Master's Student – Dept. of Computer Science
University of Calgary
Phone: 403-210-9499, bdgjones@ucalgary.ca

Anthony Tang
Professor – Dept. of Computer Science
University of Calgary
Phone: 403-210-6912, tonyt@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.

## B.2 Verbal-Consent Protocol

**Verbal Protocol**

This script is used prior to asking the participants to sign consent forms. This is part of the consent process.

*Hello. My name is [experimenter name], and I will guide you through this study. Thank you for your participation. You should feel free to ask me questions at any time.*

*Before we begin, you should understand that you have several rights as a participant:*

- *You may withdraw from this study at any time by indicating your wish to withdraw to me or my co-investigators.*
- *No data will be used without your explicit consent, as indicated on the consent form I am about to give you.*
- *Your participation in this experiment is confidential.*

*Now, please read this consent form carefully, which explains your right as a participant and the conditions of the study, and sign it if you agree with the terms. [Hand two copies of the form to each participant]*

*One of these copies is for you to keep for your own records.*

## B.3 Brief Study Protocol/Instructions

# Drone Video Conferencing Study Instructions

## Task Order

1. Getting Familiar
2. Setting Up for an Event
3. Visiting Hula-Hoops

## Checklist

### Before the study

- Node server running
- Drone batteries fully charged and ready to go
- Data logging cleared
- Screen recording set up
- Camera set up

### During the study

- Hand consent forms and pens to participants
- Briefly explain the study:
    - Thank you!
    - Evaluating a video conferencing prototype that uses a drone.
    - One of you will be in the field, and the other in the lab.
    - You will complete a series of tasks together while communicating through this video conferencing prototype.
    - The activities will take place in the field, but the person in the lab (who is, in this case 'remote') will be assisting the field person with the activities.
    - The video conferencing prototype works just like Skype, but with one additional feature…
    - The person in the field will have a drone follow them around, providing an extra camera view of them for the person in the lab to see.
    - Any questions?
- Explain the system to both participants:
    - Show the drone
    - Run the system
    - Walk through the interface with both participants
    - Explain safety precautions/guidelines
- Bring the field participant (hereby called the 'MC') to the field location.

## B.4 Task 1 Information Sheet for the DP

# Event Plan


Your task is to guide your partner in setting up the park for an event. The park must be set up as shown in the image above. Your partner has on hand a set of items. You must guide your partner in placing the items around the park in accordance to the plan.

Some additional notes:

- Pylons 1, 3, and 4 have already been placed down.
- A bag of red balls must be placed between pylons 1 and 2.
- A bag of blue balls must be placed between pylons 5 and 6.

## B.5 Task 1 Information Sheet for the FP

### Event Plan

Your task is to set up the park for an event. You have on hand a set of items that must be placed in certain spots around the park in a specific way. Your partner has instructions on how to place the items around the park in their correct places. You will follow your partner's instructions in placing the items.

## B.6 Task 2 Information Sheet for the DP

# Visiting Hula-Hoops

Your task is to help your partner visit the hula-hoops in the field in a specific sequence.

Each hula-hoop is unique, and has a specific colour and a pattern. A hula-hoop can be either red, pink, yellow, green, or purple; and can have one of the following patterns: |, +, Λ, or =

This is the sequence of patterns your partner must visit:

| = Λ Λ = | Λ + | + = Λ = + + + | | Λ =

Your partner knows the first colour he/she must visit. Each hoop has written on it the next colour to visit, as well as a number.

Please write the numbers you find written on the hoops in the order you find them:

__ __ __ __ __ __ __ __ __ __ __ __ __ __ __ __ __ __ __ __

Good luck! ☺

## B.7 Task 2 Information Sheet for the FP

### Visiting Hula-Hoops

Your task is to visit the hula-hoops in the field in a specific order, in order to help your partner figure out a sequence of digits.

Each hoop has a colour and a pattern. The first colour hoop you must visit is purple. Each hoop has written on it a number and the colour of the next hoop to visit.

Your partner has information on the sequence of patterns to visit. With that information, you two will decide which hoops to visit in which order.

## B.8 Task 2 Layout and Key

Visiting Hula-Hoops (Layout and Key)

The correct sequence of numbers is: **25631948825307474518**

## B.9 Post-Study Questionnaire for the DP

### POST-SESSION QUESTIONNAIRE

**For Desktop Participant #: _____**

Dear participant,

Firstly, we would like to thank you for taking the time to be involved in our study. There is no doubt that your comments and feedback would be valuable.

We would like to ask you a few questions about your experiences. Thank you for your cooperation!

*If there are any questions that you do not want to answer, feel free to leave them blank.*

**Demographic information:**

Age:

Gender:

Educational background:

Occupation:

**What is your relationship with your fellow participant (if any)?**

_______________________________________________

**To what extent did you enjoy video conferencing with your partner?**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Not at all | | | Somewhat | | | Very much |

**How efficient do you believe you were in completing the given tasks?**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Very inefficient | | Somewhat inefficient | | Somewhat efficient | | Very efficient |

**How immersed did you feel in the activity?**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Not immersed | | | Somewhat immersed | | | Very immersed |

**To what extent did you feel *present* in the activity space with your partner?**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Not at all | | | Somewhat | | | Very much |

**When could you imagine using a drone-supported system like this while communicating through video conferencing?**

____________________________________________________________

____________________________________________________________

____________________________________________________________

____________________________________________________________

____________________________________________________________

**How often do you use video conferencing tools (such as Skype, Google Hangouts, FaceTime, etc.)?**

Very often

Sometimes

Rarely

Never

**Have you used video conferencing *while on the go* (i.e., not in the home or workplace) using your phone or some other mobile device? If so, describe the last time you did this.**

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

**Do you have any additional comments?**

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

Again, thank you so much for participating in this study!

## B.10 Post-Study Questionnaire for the FP

### POST-SESSION QUESTIONNAIRE

**For Mobile Participant #: _____**

Dear participant,

Firstly, we would like to thank you for taking the time to be involved in our study. There is no doubt that your comments and feedback would be valuable.

We would like to ask you a few questions about your experiences. Thank you for your cooperation!

*If there are any questions that you do not want to answer, feel free to leave them blank.*

**Demographic information:**

Age:

Gender:

Educational background:

Occupation:

**What is your relationship with your fellow participant (if any)?**

______________________________________________________________

**To what extent did you enjoy video conferencing with your partner?**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Not at all | | | Somewhat | | | Very much |

**How efficient do you believe you were in completing the given tasks?**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Very inefficient | | Somewhat inefficient | | Somewhat efficient | | Very efficient |

**How immersed did you feel in the activity?**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Not immersed | | | Somewhat immersed | | | Very immersed |

**To what extent did you feel that your partner was present in the activity space?**

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Not at all | | | Somewhat | | | Very much |

**When could you imagine using a drone-supported system like this while communicating through video conferencing?**

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

**How often do you use video conferencing tools (such as Skype, Google Hangouts, FaceTime, etc.)?**

Very often

Sometimes

Rarely

Never

**Have you used video conferencing *while on the go* (i.e., not in the home or workplace) using your phone or some other mobile device? If so, describe the last time you did this.**

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

**Do you have any additional comments?**

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

Again, thank you so much for participating in this study!

## B.11 Interview Questions to both the DP and FP

**Interview Questions**

*What did you think?*

*Did you enjoy completing the activity together through this communication medium?*

*Was the activity easy to complete this way? Was it difficult?*

*How do you think the experience would have compared to being physically together?*

*How do you think the experience would have compared to having only a smartphone and communicating through video chat?*

*What kinds of problems or difficulties did you encounter while communicating and interacting this way?*

> *Did you develop any strategies to address these problems?*

*[Asking the desk-bound participant.] Is there anything you would have done differently if you were the person walking around the field?*

*[Asking the field participant.] Is there anything you would have done differently if you were the person at the desk?*

*[Ask to field participants.]:*

> *How did it feel having a drone follow you around during the activity?*
>
> *Did you feel comfortable having the drone fly near you?*

## B.12 Recruitment Poster

UNIVERSITY OF CALGARY

**LOOKING FOR PARTICIPANTS FOR A STUDY ON DRONE-SUPPORTED VIDEO CONFERENCING**

We are researchers from the Interactions Lab in the Department of Computer Science, University of Calgary. We are looking for pairs of people aged 18+ to participate in a study exploring the use of drones in supporting video conferencing while on the go. As part of this experience, your interactions will be video-recorded. Come participate to experience the future of video conferencing!

WHO: Pairs of people aged 18+

WHEN: August 26 to September 13, 2015

WHERE: University of Calgary

TIME: Approximately 1½ to 2 hours

REMUNERATION: $20/person

If you are interested in participating, or have any questions, please contact Brennan Jones (bdgjones@ucalgary.ca, 403-324-2790).

*The University of Calgary Conjoint Faculties Research Ethics Board (CFREB) has approved this research study.*

## B.13 Recruitment Notice

**SAMPLE Recruitment Notice**

This is a sample. While wording may change slightly, the intent of the language will stay the same. If there are substantial changes, we will report this to the CFREB (Conjoint Faculties Research Ethics Board)

---

**Title of Project:** Exploring the role of a drone to support collaboration and shared experiences in on-the-go video conferencing

---

We will recruit volunteer adults (aged 18 or over) who are willing to participate in our study using department mailing lists, in-person contact with colleagues, online forum posts, online social media posts, and advertisements in the form of posters. From the respondents, we will choose a subset of participants for the study, largely on a first-come first-served basis, but also with consideration of obtaining a good demographic spread.

These emails are not unsolicited bulk email (i.e., spam), but they will be posted to selected target communities by the most efficient methods currently available. The recruiting announcement will use language similar to the following:

*"We are researchers who are designing a system, utilizing a recreational drone, to support collaboration and shared experiences in video conferencing while on the go in large environments. If you agree, we would like to ask you and somebody you know to participate in a study to test our prototype. You will be asked to communicate with your partner through either a traditional video conferencing link (e.g., one that resembles Skype, FaceTime, etc.) or through our prototype of a video conferencing system utilizing a drone. Please be advised that we prefer recruiting participants in pairs, and one person in each pair should be able to comfortably walk around a large outdoor environment (e.g., a field). With your permission, we will video/audio record this study. All collected information will be anonymized. If you are interested in participating or have any questions, please reply to this email. We estimate a total time of approximately 90 to 120 minutes. In addition, if you have friends or colleagues who you believe might also be interested in participating in this research, we would be grateful if you could speak to them about this research opportunity and/or forward them this email with information about our study. Thanks so much for your help!*

*This study has been approved by the CFREB (Conjoint Faculties Research Ethics Board)."*

# Appendix C

# ADDITIONAL QUANTITATIVE STUDY DATA

In this appendix, I present some of the additional data that was collected during the study, but was not analyzed or discussed in greater detail in the body of the thesis. In addition, I also provide a web link to some of the anonymized raw data from the system logs (user-interface logs and location/bearing logs).

## C.1 Quantitative Questionnaire Data

The following Likert-scale questions (on a scale of 1-7) were asked to participants in the questionnaires:

1. *To what extent did you enjoy video conferencing with your partner?* [1 = 'not at all'; 7 = 'very much']

2. *How efficient do you believe you were in completing the tasks?* [1 = 'very inefficient'; 7 = 'very efficient']

3. *How immersed did you feel in the activity?* [1 = 'not immersed'; 7 = 'very immersed']

4. *To what extent did you feel that you were present in the activity space?* (in the case of the DP) or *To what extent did you feel that you were present in the activity space?* (in the case of the FP) [1 = 'not at all'; 7 = 'very much']

The bar chart in Figure C.1 shows the average response values (across all pairs) to these questions, separated by DP and FP. For more information on how the questions were worded, please refer to the post-study questionnaire documents in Appendix B (Section B.9 for the questionnaire given to DPs, and Section B.10 for the questionnaire given to FPs).

## C.2 Mobile-Interface Action Statistics

Tables C.1-C.4 indicate the proportion of time across the tasks that certain settings were in place on the FP's phone. The values in these tables are based on the data logged from the system.

## C.3 System-Log Data

Anonymized raw data from some of the system logs (user-interface logs and location/bearing logs) can be found here: `https://dataverse.scholarsportal.info/dataset.xhtml?persistentId=doi:10.5683/SP/VWCDKW`.

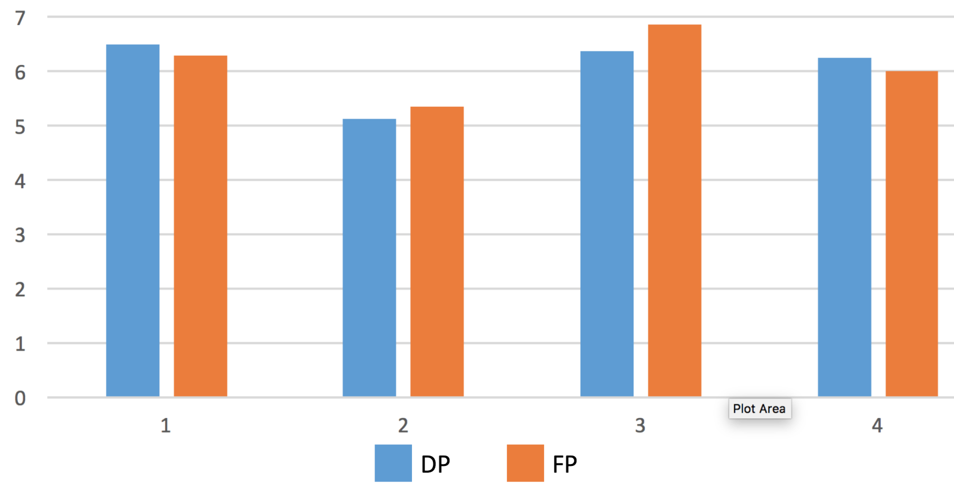


Figure C.1: Average Likert-scale response values (across all pairs) to the questionnaire questions, separated by DP and FP.

| | |
|---|---|
| Percentage of time FP using front camera | 15.1931796641 |
| Percentage of time FP using back camera | 84.8068203359 |
| Percentage of time phone in portrait orientation | 99.6581575278 |
| Percentage of time phone in landscape orientation | 0.341842472197 |
| Percentage of time FP has remote view enlarged | 10.082463173 |
| Percentage of time FP has local view enlarged | 89.917536827 |
| Percentage of time FP viewing the drone video view | 17.1685860781 |
| Percentage of time FP viewing the video chat view | 82.5873154871 |

Table C.1: Mobile-Interface action statistics across all tasks. These values are averages across all pairs.

| | |
|---|---|
| Percentage of time FP using front camera | 51.61205601 |
| Percentage of time FP using back camera | 48.38794399 |
| Percentage of time phone in portrait orientation | 100 |
| Percentage of time phone in landscape orientation | 0 |
| Percentage of time FP has remote view enlarged | 77.97958666 |
| Percentage of time FP has local view enlarged | 22.02041334 |
| Percentage of time FP viewing the drone video view | 15.4802659 |
| Percentage of time FP viewing the video chat view | 84.5197341 |

Table C.2: Mobile-Interface action statistics for the warm-up task. These values are averages across all pairs.

| | |
|---|---|
| Percentage of time FP using front camera | 17.5495100603 |
| Percentage of time FP using back camera | 82.4504899397 |
| Percentage of time phone in portrait orientation | 99.7456007873 |
| Percentage of time phone in landscape orientation | 0.254399212744 |
| Percentage of time FP has remote view enlarged | 0.184862393153 |
| Percentage of time FP has local view enlarged | 99.8151376068 |
| Percentage of time FP viewing the drone video view | 30.6287960041 |
| Percentage of time FP viewing the video chat view | 69.0320567843 |

Table C.3: Mobile-Interface action statistics for Task 1 (field setup). These values are averages across all pairs.

| | |
|---|---|
| Percentage of time FP using front camera | 4.0071416955 |
| Percentage of time FP using back camera | 95.9928583045 |
| Percentage of time phone in portrait orientation | 99.5039189235 |
| Percentage of time phone in landscape orientation | 0.496081076496 |
| Percentage of time FP has remote view enlarged | 0.0 |
| Percentage of time FP has local view enlarged | 100.0 |
| Percentage of time FP viewing the drone video view | 7.45864918527 |
| Percentage of time FP viewing the video chat view | 92.3058293036 |

Table C.4: Mobile-Interface action statistics for Task 2 (scan and search). These values are averages across all pairs.