Collaboration in 360° Videochat: Challenges and Opportunities

Anthony Tang[1], Omid Fakourfar[1], Carman Neustaedter[2], and Scott Bateman[3]

[1]University of Calgary   [2]Simon Fraser University   [3]University of New Brunswick

{tonyt, omid.fakourfar}@ucalgary.ca   carman@sfu.ca   scottb@unb.ca

ABSTRACT

We designed a videochat experience where one participant can experience a remote environment from a 360° camera. This allows the remote user to view and explore the environment without necessitating interaction from the local participant. We designed and conducted an observational study to understand the experience, and the challenges that people might encounter. In a study with 32 participants (16 pairs), we found that remote participants could actively participate in the experience with the environment in ways that are not possible with current mobile video chat. However, we also found that participants had challenges in communicating location and orientation information because many of common communication resources we rely on in collocated chat are not available. Based on these findings, we discuss how future mobile video chat systems need to balance immersion with interaction ease.

Author Keywords

360 video, video chat, video conferencing

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

Mobile devices such as smartphones, tablets and head-mounted displays present new opportunities for sharing experiences and communicating with others through mobile video chat (e.g. [5,38,24,16]). They are now being used to share experiences such as treasure hunts [38], activities such as walks [46] or bike rides [35], and as tools for supporting ad hoc assistance such as repair (e.g. [12,18,9]). Researchers have also explored novel camera and devices for sharing activities with others [24,28,16,21,22].

The problem with conventional mobile video chat tools is that the remote participant has little autonomy over the video

Figure 1. The 360° camera (atop a monopod affixed to a backpack) streams 360° video (left) to a tablet (right).

scene—their view is strictly controlled by the other participant in the chat. This creates asymmetries in how the two participants can view and explore the environment: namely, the remote participant's access to the environment is mediated by the local participant (i.e. they can only see what the local participant points the camera at) [20]. While the remote participant is at a noted disadvantage, in addition to participating in the video chat, the local participant is now also responsible with performing camerawork: framing subjects or objects in the video stream, providing effective overviews, and steady camerawork [20,30].

Researchers have tried to address this asymmetry through novel hardware and software. For example, several researchers have explored novel hardware configurations of cameras such as head-mounted (or similar) approaches (e.g. [18,24]), telepresence robots [39], and even drones [19]. Other researchers rely on different types of form of scene reconstruction (e.g. using depth cameras), thereby allowing the remote participant to independently control his/her view into cached portions of the scene (e.g. [12,21]).

In this paper, we study the use of a fixed, streaming 360° camera on a monopod (affixed to a backpack carried by the local participation), where the video is streamed to the remote participant. This prototype gives the remote participant the freedom to visually explore the environment independently of the local user, while giving the remote participant a view of the local user's hands and head orientation. We were focused on three research questions:

- What new opportunities does a 360° video chat present over a standard mobile video chat?

- How do remote participants explore the environment, and contribute to collaboration with a 360° video chat?
- How do collaboration challenges manifest themselves with a 360° video chat?

To address these questions, we designed a study to simulate a guided tour (e.g. a remote tour of a factory), where participants would explore a part of our university campus together. We see this scenario as a common future scenario, where remote participants will need to rely on the local participants to guide them around or use 360° video panning feature to explore the environment (e.g. remote assistance). 16 pairs of participants used our prototype to perform a photo walk tour, where the remote participant directed the local participant to take photos of certain landmarks at specific angles. Our findings show that the prototype gave remote participants agency in the interaction: they could comfortably and confidently explore the environment independently, enough so that they could assist the other participant in visual search tasks. On the other hand, because participants could not easily understand what the other could see in the environment, we observed small communication break-downs between pairs. Thus, some of the benefits of having independent views for collaboration were offset by the absence of gaze awareness between collaborators.

We make four contributions in this paper: first, we contribute the one of the first studies exploring the impact of view independence in 360° video chat; second, we describe findings from our study that outline opportunities for distributing collaboration with this approach; third, we identify challenges that arise from the view independence afforded by this configuration, and finally, we provide concrete design guidance by outlining how opportunities to address these challenges in future 360° video chat designs.

RELATED WORK
Video media space research has long focused on bridging the gap between fixed, remote spaces (e.g. [13]). More recent efforts explore how video chat from mobile devices can connect people with others in remote environments, rather than fixed spaces (e.g. [30,20,16,24]). As it turns out, many communicative challenges that researchers are encountering with mobile video chat prototypes are reminiscent of challenges encountered decades ago: the problem of joint reference and fractured ecologies. We outline these problems to contextualize our work.

Mobile Video Chat
Mobile video chat relies on sophisticated devices with cameras and a sufficiently sized screen running atop a high speed network. Smartphones and tablets now provide these capabilities, and it is these platforms that have given rise to an entirely new class of video chat behavior [5]. Prior to this, video chat activities were restricted to where computers or laptops could be plugged in and used, but mobility has enabled behavior that includes: giving video tours [5,30,20,8], asking for help and assistance (e.g. with cooking

or repair) [5,20,18,9], and sharing experiences such as sporting events or outdoor activities [16,46,24,38].

These kinds of experiences are fundamentally different from the fixed environments that were common in traditional video media spaces, because they involve camera movement—in part by the person carrying the camera who is moving in the environment, but also camera movement to show some aspect of the scene. Many explorations have explored how commodity cameras can be used to support video chat (e.g. [38,20,30]), and a common theme arising from these explorations is the challenge of camera work—the mechanics of actually capturing and framing subjects and objects of interest to be shared in the video chat [20,30].

Asymmetries of control and participation. Conventional mobile video chat creates asymmetries of control, and consequently participation (e.g. [20,38]). The user with the camera controls the remote user's view of the environment, meaning that the remote user cannot independently explore and understand the environment. In at least one study [20], we have seen that it impacts collaboration by stifling the remote user's participation: constantly needing to ask the camera to be turned is socially awkward. Consequently, remote users have limited ability to control the flow of conversation or what is seen in the view.

Several prototypes have tried to address this problem by obviating camera control altogether—either by fixing the camera to a static location (e.g. [28,16]) or to the person in the scene (e.g. [28,24]), or by giving some control to the remote participant ([19,21]). More recent approaches have been designed to explicitly allow the remote participant to engage and experience the scene independently of the local participant. For example, PanoVC captures a remote scene as a cylindrical capture from a mobile phone, and allows a remote participant to explore the scene independently of the local participant [33]. However, the remote user still has to rely on the local user to start capturing a panoramic view of the scene and keeping it updated.

Similar to the present work, the JackIn projects [21,22,23,34] explore scenarios that push the boundaries of this, where the local participant wears a 360° camera (one that captures all the way around oneself), and the remote participant can explore the scene as it is captured from the local participant. In these prototypes, the remote participant wears a head mounted display to gain fully immersive experience (i.e. they can look around, but in the remote space). We build these findings in three ways: first, we explore a larger interaction space (i.e. several buildings) to mimic a remote tour; second, we carefully consider the sub-task of conveying perspective (i.e. "Look at this,"), which is the point of a remote tour, and finally, we use a tablet for viewing the remote space (rather than an HMD).

Yet, even with these prototypes, a central issue remains: the remote user is not *actually* there—this is a one-way illusion. What new challenges arise because of this asymmetry?

Figure 2. The left illustrates the entirety of the 360° capture. The remote user can see a small portion of the view (that can be panned). From where the camera was affixed, it is possible to see which direction the local user is looking, and their hands.

Fractured Ecologies & Challenges of Joint Reference

Video media spaces bridge spaces with audio-video connections, but because people are not physical present in the remote space, communication about objects across the video boundary is challenging. Luff et al. [31] call this the problem of "fractured ecologies", where the ecology of resources available to facilitate communication (including people's physical bodies—their orientation, their heads, their eye gaze, their limbs, and the environment—containing objects), which support people's speech acts, are inconsistent between participants. For example, deictic references—the ability to point at an object and say, "Look at this,"—rely on several aspects of this ecology: the partner must be able to see one's body, one's head orientation (likely pointed at the object), one's finger and gesture, and finally at the object itself. Being able to understand the ecology and the environment allows people to make these joint references, and in so doing, to help create common ground for conversation. This common ground is crucial if people are to interact with one another. In a conventional video chat system, several communication resources are not present: when a remote partner looks away from the camera, it is difficult to determine what s/he is looking at or why; similarly, when a s/he points at the laptop screen through a video chat, it is difficult to know what is being referenced. For example, knowing what others are looking at helps to resolve shared tasks [6]. Although people generate workarounds by referring specifically to objects rather than using deictic references [7], these are cumbersome and can distract from the collaboration, and many efforts have gone into designing tools to address these challenges.

Some solutions rely on providing good overviews of the work surface [31,33,10], or fusing the work surface with people's bodies in such a way to facilitate gestures (e.g. [44,17,27,44]). Another class of prototypes address this problem by providing means to annotate the video stream [11,9,25,26], or a model of the scene [12,21,40]. While many of these approaches are one-way (i.e. one person draws for another), some systems provide a two-way mechanism to convey gestural and annotation information (e.g. [26,21,3]).

Our work builds on this rich lineage of research by exploring how these types of problems occur with a 360° camera. Because the remote participant has a (very) wide-angle camera access to the environment, we were uncertain how these challenges would manifest.

360° VIDEO CAMERA PROTOTYPE

We designed a 360° streaming prototype that gives the local participant complete freedom with his/her arms. As illustrated in Figure 1, we affixed a 360° camera to the top of a monopod, and secured it to a backpack. The view from the camera is thus a "third person perspective" of the local user. The remote user views the 360° feed through a tablet application. Figure 2 illustrates what is visible in a captured scene, and the application allows a user to drag the view to see different parts of the scene. The video streams using a proprietary UDP-based prototcol. While participants experienced variable video latency (1~5s), this issue could be resolved by explicitly resynchronizing the video (through the tablet UI). We used a phone connection for audio, with only the negligible latency of the cell network.

We arrived at this "above the backpack" view based on earlier prototypes where the monopods were held, but found the weight of the camera too heavy. In contrast, the backpack frees the local user's hands for activities and for gesturing. The view is also somewhat different from the approach of [34], where the view comes around eye-level. A nice consequence of our prototype is that remote users can see their partner's heads and hands, revealing not only which direction the local participant's head was facing, but also what they were doing with their hands.

STUDY

We designed a study where pairs of participants used our 360° mobile video chat prototype: the local participant wearing our 360° prototype, and the remote participant viewing the video stream through a tablet and connected via a phone call. Together, they completed a photo walk tour, where the remote participant directed the local participant to landmarks to take camera photos from specific angles. The remote participant was also given ad hoc tasks, where s/he was to report on characteristics of the environment.

Figure 3. Five Landmarks (across) each had different characteristics. Across the top are the reference images. The bottom row are photos taken by Group 8.

Our study simulated the experience of going on a "remote tour" (e.g. touring a remote factory, or taking a tour of campus) [5,46,8]. Here we would expect the remote user to occasionally ask about or talk about something in the environment that they can see. To do this, the remote participant would direct the local user's attention to specific things in the environment (perhaps only visible from certain angles). We mimicked this type of interaction using the photo walk tour, where the remote user directs the local user to take photos of landmarks from specific perspectives. We were interested in the interaction between pairs as they completed these tasks.

We similarly expect that in remote tour scenarios, remote users will occasionally spontaneously see something in the environment (e.g. [24]). We mimicked these situations using "ad hoc tasks," where remote participants had to answer questions about aspects of the immediate visual environment. These questions were asked at contextually appropriate times, but participants did not know about them *a priori*. We were interested in whether remote participants would complete these visual tasks independently (i.e. without support from the local participants).

Photo Walk Tour
The local participant takes photos of landmarks using a phone camera, directed by the remote participant (who views the environment via a streaming 360° camera connection). The remote participant is given reference photos for each landmark (Figure 3, top row), and a map of the environment (~5 minute walk end to end) that outlined where each photo should be taken.

Landmarks and Reference Images. We chose five different landmarks for the photo walk task (Figure 3). At each of these, we created a target image where the participants were to copy the image—not only of the landmark, but also of the perspective of the photo.

We designed the landmarks and target photos to vary on several dimensions (landmark characteristics, context characteristics, and target photo characteristics). Examples of the characteristics we varied include: landmark characteristics (how easy is it to describe the landmark: abstract *vs* concrete; 2D *vs* 3D; above eye-level *vs* eye-level); context characteristics (visibility along walking path—e.g. in front vs. to the left/right; visual milieu—i.e. is the object alone or amongst a group), and target photo characteristics (standing height *vs* squatting height; orientation—e.g. straight ahead *vs* pointing upwards; whole *vs* part; portrait *vs* landscape). For example: Landmark 1 is a distinctive wood sculpture at the corner of a hallway, plainly visible from eye-level. Landmark 2 is also visible from the hallway, though above eye-level; the photo is taken at an oblique angle and captures only a subset of the photo frames. In contrast, both Landmarks 4 and 5 are landmarks that are not visible from the main walking path (they require taking a "button-hook" turn), and whereas Landmark 3 is abstract and difficult to describe, Landmark 5 is easier, but among a series of similar paintings.

Ad hoc Tasks
During the photo walk tour, we occasionally asked the remote participant about the environment (e.g. "How many people are sitting in the study area?"). Participants were not told in advance when these would happen, and needed to respond in the moment by visually inspecting the scene. We gave two such tasks—one near Landmark 3 and one near Landmark 5.

Participants
We recruited 16 pairs of participants (32 participants total; 17 female) based on ads placed around our university. Participants' age ranged from 16 to 30 (median: 23.5). Within each pair, participants knew each other and only one participant was familiar with campus. However, our university is a public institution, so many of these

participants would have visited campus before but not be familiar with landmarks on campus. 29 participants reported having seen a 360° video before, and 11 participants had tour guide experience (e.g. as a volunteer).

Procedure

Participants were briefed on the tour-based theme of the videochat study. Participants then watched 360° videos from a tablet so they could understand what the experience would be like from the remote participants' perspective. Once they were familiar with the interface, one participant was outfitted with the 360° video camera prototype.

The participants were separated, and instructed to go for a short guided tour to familiarize themselves with the interface and experience. In all cases, this lasted no more than 5 minutes. At this point, the local participant was led to the "start point" for the Photo Walk. Once participants had completed the photo walk (along with the two ad hoc tasks), they were brought back to our lab to be debriefed. They each filled in a questionnaire about their experience separately, and were remunerated for their time.

Data and Analysis

Each participant was accompanied by an experimenter at all times. We collected field notes from both the perspective of the local and remote participants, and participants' responses from pre- and post-study questionnaires. The post-study questions were mostly qualitative and open-ended which gave us insight about participants' experience. For example, we asked about how participants would redesign the system and what aspects of the task they found difficult, and so forth. The remote participant was video recorded so we could hear both sides of the conversation.

We performed a thematic analysis on the data, where we iteratively identified salient aspects of the experience from both participants. Our goal was to understand the overall experience from each participant's perspective, and the challenges participants encountered when communicating with one another. These experiences were iteratively grouped together into provisional themes, where the themes changed through our data collection process.

RESULTS & FINDINGS

All participant pairs completed the main photo walk task within 20-40 minutes. The prototype failed temporarily (WiFi connectivity) for two participant groups; in both these cases, the study needed to be interrupted for about five minutes to restart the connection, after which the group continued on. Figure 3 illustrates a selected summary of the images taken by participants (the complete set of images taken by participants can be viewed at [43]).

Participants enjoyed the study, and the experience that the 360° videochat prototype provided. This was reflected both in the way in which pairs interacted with one another during the study (i.e. frequently laughing and joking around about what could be seen in the local environment), and by their post-study comments in the questionnaire:

It was pretty great guiding around my friend using a 360° camera so I could see what he was seeing while he walked around. [G2-remote]

It was quite easy to find my way when connected with the tour guide, and it felt more independent […] since I could go at my own pace and look at interesting things. It was like having a real life tour without the tour guide with me in person. [G2-local]

Locations and Photos

Participants found every landmark. As expected they had more difficulty navigating to Landmarks 3, 4, and 5 because these were not plainly visible from the walking path. When the landmark was more abstract (Landmark 4), remote participants had a harder time describing it. As illustrated in Figure 3, capturing the photo from the correct perspective/angle was challenging: for remote participants, it was difficult to articulate unambiguously the desired orientation of the photo. This difficulty was echoed in participants' reactions to this aspect of the task:

I wasn't sure how to orient the camera or where to focus the camera because sometimes descriptions were vague [G1-local]

It was difficult to know exactly the angle of the picture [G3-local]

Remote participants seemed frustrated at times when trying to describe these angles to their local partners:

Describing the angles was particularly challenging as even though my partner explained very carefully, there was still a lot of ambiguity as to how exactly the picture should've been taken. [G4-remote]

Nevertheless, participant pairs did complete these tasks, and we revisit both the communicative strategies they developed, and the challenges they encountered later in this paper.

Ad hoc Tasks

Remote participants could not anticipate when the ad hoc tasks would occur, nor what the content of these tasks would be (e.g. "Are there any vending machines nearby? If so, how many dispense only snacks?"). In a conventional mobile video chat scenario (e.g. Skype or FaceTime on a mobile phone where the local participant holds the mobile device [20]), a remote participant cannot answer such a question alone: s/he would not be able to see, and would need to either ask the local participant to turn the camera, or to answer the question.

In total, we captured 32 instances of these ad hoc tasks (two per pair). Of these, the remote participant completed 23 by themselves (i.e. without involving the local participant). This result suggests that participants felt comfortable and confident enough in the 360° camera view to complete the tasks themselves. In 5 instances, the remote participant relayed the task to the local participant to complete (i.e. much like in a conventional mobile video chat situation). Finally, in 4 cases, the remote participant completed the task with some assistance from the local participant—e.g. to confirm the result, or to move to a particular location with a better view. Kasahara & Rekimoto [21] describe this behavior as "Ghost" behavior, where the remote participants can see and

hear everything that is happening in the local environment, but cannot actually affect anything in the environment (aside from telling the local participant to move). Participants echoed how awkward it felt:

It is a strange feeling as if I am present [in] a certain area, but have no control/ability to engage with what I am seeing in front of me. [G10-remote]

Opportunities with 360° Chat
The 360° view provides a comparatively far larger field of view than a typical mobile phone camera [20,30]. Remote participants took advantage of the 360° view provided by the prototype during the photo walk task. We saw this reflected in a number of ways: first, remote participants would look around independently, *initiating* conversation about topics and objects in the scene (i.e. rather than respond to a view directed by the local participant); second, remote participants would actively help look for landmarks when the local participant was confused or could not locate them him/herself; third, remote participants would watch their partner's actions, providing assistive guidance where necessary, and fourth, remote participants would frequently describe the visual *context* of a target image, rather than strictly features of the target image.

Looking Around Freely
Based on prior research on mobile video chat, we know that the remote participant's view is strictly dictated by the camera work of the local participant [20,38]. In contrast, we saw that remote participants would freely manipulate this view independently. While remote participants generally stayed on task by strictly looking at what the local participant was looking at, there were also many instances when doing so was not important, for example, when the local participant was moving from one landmark to another. During these times, we observed remote participants looking around the environment—sometimes at objects of interest, and other times at people. This would sometimes lead to interactions that were initiated strictly by the remote participant about something that s/he had seen. Thus, the 360° camera combined with the independent view controls gives the remote participant the ability and agency to direct the attention of the local participant—an important role reversal from observations of mobile video chat [20,38].

Group 12 has just completed the last photo walk photo, and are walking back to the lab. The remote participant in this group knows campus well, and as the local participant passes by a shuttered storefront, the remote participant remarks, "Wow." When the local participant asks what he's talking about, the remote participant directs the local participant to look at the shuttered storefront, talking about its history, and how sad that it is now closed.

Group 10 is walking between landmarks. While conversation has been ongoing about how to get to the landmark, the remote participant suddenly notices a passerby and remarks, "Oh! I think I know that girl." As the local participant continues moving on, the remote participant pans the 360° camera to keep watching the

passerby who is walking in the opposite direction of the local participant.

Looking to Help the Local Participant
Previous authors have noted that in mobile video chat, the onus is on the local participant to find objects/landmarks, and frame them for the remote participant [30,20,38]. In contrast, we observed remote participants taking on some responsibility for helping to find landmarks in the photo walk task—particularly when the local participant was lost or uncertain of where to go. This was particularly notable (and in many cases, necessary) for Landmarks 4 and 5, where both landmarks are not "forward visible" as the local participant walks. In almost every case, the local participant walked past the landmark, requiring intervention from the remote participant to help find the landmark. What we see from these instances is that the remote participant, given this power to explore the environment, now also shares some responsibility for finding landmarks—it is not the sole responsibility of the local participant.

Group 1 has walked past Landmark 4, and both participants are confused: the remote participant is sure the local participant has walked past the landmark based on the map. The remote participant directs the local participant to stop. At this point, the remote participant begins looking around herself, and in the distance, she spots what might be the landmark, "Wait, is that it?" She directs the local participant toward what ends up being the landmark.

Watching the Local Participant's Actions
Researchers have suggested that a major problem with mobile video chat is that it can be difficult to understand what the local participant is doing with his/her hands (without explicitly pointing the camera in this way) [30]. In contrast, because of the way our 360° prototype was positioned (atop a monopod affixed to the local participant's backpack), our prototype provided a third-person perspective on the partner. Remote participants made use of this as a resource when local participants were taking photos—watching how they were oriented to and in relation to the landmarks, and sometimes offering advice on how to reposition the camera based on what they could see in the camera screen. Using this view of the local participant, the remote participants could direct action on a moment-to-moment basis:

Group 4 is trying to take a picture of Landmark 4. The remote participant is having a difficult time explaining how the camera should be located and oriented in the scene. As the local participant attempts to decipher his instructions, she moves the camera slowly in space. The remote participant is watching her actions carefully, responding verbally to each of the movements she makes: "So [the] left side of the camera should be touching the white part of the wall, and then… Yeah… Yeah, kind of like… Only touch the left side of the wall."

Visual Context as a Resource
Remote participants with the 360° camera prototype frequently described the target landmark *within the visual context* of what they were seeing as the local participant got closer. This happened because the remote person could

actually see a wider range of visual context given the 360° view. This made the task easier for the local participant because as the remote participant could describe the context, where the pair built a shared understanding of what made up the visual context. Thus, specifying a particular target of interest becomes straightforward. In the following vignette, the remote participant identifies a feature of the target that distinguishes it from other similar targets:

Group 1 is attempting to take the photo for Landmark 5. Landmark 5 is made up of a set of 8 frames of pictures, arranged in an irregular grid. The remote participant, in describing which frame should be captured, describes the target: "It's the picture with the square in it." This is a distinguishing feature, as other pictures have other shapes.

This visual context provides a framing function as well—beyond simply describing what the target photo includes, the remote participant can use the description to articulate what aspects of the scene should be *outside* the photo.

Group 11 has arrived at Landmark 4, where the target image is of an abstract painting that is part of a set of three similar paintings. The remote participant asks the local participant to stop moving while she compares the target image to each of the three paintings. She directs the local participant to the middle painting, "Oh! It's the middle one with the bowling pin." She then further directs the local participant to not capture any of the other two other paintings in the photo.

Beyond using the framing technique to describe what was *outside* of the photo, remote participants also used this framing mechanic to describe what should be *in* the photo. Here, the participants would describe the target photos as frame, and describe what parts of the landmark should fit within the photo.

The remote participant in Group 6 is describing how to take a picture of Landmark 3, which is a display case. He bases his description strictly on the corners/interior of the frames on the target image: "The map will be in the picture. The top left is the map, and in the bottom right you'll have the text …"

The visual context also served as a common resource when the remote participant was navigating the local participant to the next landmark. Several times, remote participants would simply direct the local participant to go towards a sign, or to follow a passerby headed in the same direction. In these cases, rather than directing the local participant on a moment-to-moment basis, the remote participant is essentially deferring the navigation task to the local participant via an external (common) reference. By doing so, the remote participant frees him/herself of the obligation to navigate the local participant. The remote participant is able to do this here because s/he has independent access to a wide visual field of view. In contrast, with a conventional mobile video chat, a remote user would not be able to see what landmarks are available without asking the local user to move the camera.

Challenges with 360° Chat

We observed that remote participants had challenges communicating to the local participants *how* the camera should be oriented to take the photo of the landmark. We see four principal reasons for these difficulties: first, participants' views were decoupled—the remote participant's view was independent of the local participant, thus participants frequently needed to re-orient themselves during photo-taking; second, communicating unambiguously orientation was challenging as directional terms were overloaded, and finally, remote participants could not use gestures or rehearsal mechanics to communicate. These findings extend previous findings on mobile video chat [20,24,30,36] by explaining how they manifested in the 360° videochat scenario.

The Trouble with Decoupled Perspectives

With conventional mobile video chat, the local participant has a clear mental model of what the remote participant sees (since the local participant is directing the view and holding the camera in front of them with the viewfinder visible). Thus, when an object is in view, it is straightforward to understand what is being shown and discussed [20,30]. In contrast, because the 360° camera allows gives the remote participant the freedom to explore the environment, at any moment, participants could be looking in completely different directions. This decoupled perspective caused some problems, as one participant might reference something in the scene that was not visible to the other. Generally, remote participants understood what the local participant was looking at and could see: the local participant's hands and head were in view, and when the local participant was moving, the implicit motion provided a nominal, clear "forward" orientation. However, the local participant had no way of determining what the remote participant was looking at. Thus, the local participant might make reference to something in the immediate vicinity that required re-orientation by the remote participant:

Sometimes it was also difficult when [my partner] asked whether I could see "what was in front of him". Since I could move my video around whenever I wanted, I sometimes lost sense of which way was "front". [G10-remote]

Consequently, most remote participants kept the orientation of their view fairly closely aligned with the "forward" view, as they recognized not doing so would cause difficulty:

"It was important to see what she was looking at. I wanted to keep it the same as her view so as to not be incongruent with [the local participant's] view." [G1-remote]

Communicating Orientation is Ambiguous

To locate and orient the camera properly for the photo, participants needed to verbally coordinate the local participant's actions with instructions from the remote participant (and clarifying questions from the local participant). This was problematic, and we observed this problem most acutely on Landmark 4, where the target image was both abstract and taken from an extreme angle. Remote

participants simply lacked the resources (e.g. vocabulary, visual cues, etc.) to clarify what was needed.

The following vignette is representative of how many remote participants struggled with communicating orientation unambiguously. Here, we noticed the remote participant's instructions about how to position the camera conflated with how to angle the camera. These participants needed to "restart" the instructions several times before the photo was taken. Throughout the interaction, the remote participant continually dragging the tablet view between looking down at what the local participant is doing and back upwards at the painting.

Group 4 has reached Landmark 4. The remote participant tells the local participant to "Take the picture from the bottom left corner." As the local participant starts to take the picture, the remote participant realizes the photo will be taken straight-on, and tries to correct the angle of the photo by saying, "Take a picture up." The local participant points the camera straight up, rather than pitching the camera upward as the remote participant intended. They go back and forth three times as the remote participant tries to clarify how the photo should be taken. Both are frustrated.

Overloaded Terminology. One reason why communicating orientation was challenging was because participants lacked a precise vocabulary for articulating their ideas. Although a precise vocabulary exists (pitch, roll, yaw), these are not part of everyday vocabulary; instead, participants generally used familiar words and phrases drawn from our interaction with static 2D photos: up, down, left, right. Unfortunately, these carry multiple interpretations depending on whether one is thinking about location of oneself, location of the camera, or pitch. For instance, we observed many remote participants use the phrase, "Go back," and have the instruction misinterpreted. Based on our observations, we counted no fewer than four different meanings for what "back" referred to: "Turn around" (i.e. turn the camera view around), "Turn around and walk backwards from where you have just walked", "Take a step backwards", or "Go to your original location." Yet, the specific intention was frequently unclear to local participant; instead, it needed to be clarified with additional verbal exchanges.

Lost Resources: Gestures and Rehearsals
In a collocated setting, ambiguities around orientation and perspective could be easily resolved with a number of common approaches: one could physically position his/her partner in place, one could point to the target, or one could demonstrate how to achieve the angle, among others. In contrast, we noted that these common resources were not available to our participants.

Remote Gestures. Remote participants were active as they provided instructions to local participants, many gesturing in ways to emphasize their words or to point out objects in the world that the photos should be taken in relation to. Figure 4 (left) illustrates one such sequence: "Turn around, and take the photo diagonally in this direction [as he gestures]..."

Figure 4. (Left) Communicative gestures made by remote participants were lost. (Right) The remote participant mimics the target image (on the table) with her tablet as she tries to explain to her partner how to orient the photo.

Figure 5. Using the gyroscope control made it difficult to use external resources, where they would be left on the table (left), or balanced onto one's lap (right).

[G12-Remote] Here, not only is the hand gesture not communicated to the local participant (and therefore lost), the actual intention of the verbal instruction is lost, too.

While it was clear to all participants that these kinds of gestures would not be seen by the local participant, we nevertheless saw many instances of these kinds of gestures being produced by remote participants. The production of the gesture is closely timed with speech, and furthermore, some participants would hold the gestures in space—not because it would become clearer to the local participant—but rather to hold it as if to try to generate another verbal description of what they intended to say.

Rehearsal. In a collocated setting, a person would likely simply *demonstrate* the correct position and posture with a set of physical movements so the other person would be able to follow along. We observed that remote participants would do something similar—for instance, by re-orienting their own view (from the tablet) to mimic the reference image—however, it was not clear who this action was for:

Group 1 has arrived at Landmark 2. Unlike the first landmark, this photo needs to be taken at an angle to capture the length of the hallway. As the remote participant describes how the local participant should take the photo, the remote participant repositions her own view from the 360° camera so that the tablet view looks like the reference image (Figure 4, right). almost exactly—much like a demonstration or a rehearsal. From here, the remote participant spends some considerable time trying to describe how the picture should be taken based on the actions she took to manipulate her tablet. Throughout the interaction, she moves between looking at her own version of the target image,

and back downwards so she can see what the local participant is doing with respect to his location and the orientation of his camera.

We call this behavior *rehearsal* as it seems to be similar to a demonstration; however, the movement and posture is not for the benefit of the local participant. Rather, the primary purpose seems to be for the remote participant. The inability the local participant to see what the remote participant was looking at prevented this simple interactional resource from being used effectively.

Practicalities of 360° Viewing

At the outset of the study, we were interested in how participants would make use of the touch interface vs. the immersive gyroscope interface. Yet, only 2 of the 16 remote participants actively made use of the gyroscope interface; instead, most preferred the touch-based interface (though occasionally might make use of the gyroscope, though rarely for large movements). Participants reported preferring the Touch-based interface because it was either more comfortable or convenient. This is perhaps consistent with our general experience of using a tablet (where motion is not a normal part of interacting with a tablet).

We noted that participants using the gyroscope interface had difficulties using external resources, as it was unclear where to put these between looking around and providing instructions.

Participant 4, the Group 2 remote participant, is using the gyroscope interface. To match the local participant's "forward" view, he needs to be oriented away from the table (Figure 5, left). As the local participant reaches the Landmark, the remote participant realizes the instructions are on the coffee table. He spins his chair around to get the instructions, and then awkwardly holds them as he spins his chair back, re-orienting the tablet to see what the local participant is looking at.

Holding the instructions was no more comfortable. Participant 24 (the remote participant from group 12) is seen in Figure 5 (right) trying to balance these instructions on his lap and arm as he provides instructions to his partner.

Summary of Findings

Our study provides the first findings on the impact of 360° cameras on remote video collaboration. Our prototype system enables view independence, which provided three main benefits over conventional video chat:

1. Remote users can look and explore the environment independently, which better allows them to focus on tasks rather than coordination.
2. Remote participants can complete tasks without direct assistance of the local participant, and can better assist local participants with their tasks.
3. Because the remote user can better build an understanding of the environment, they can assist the local participant with navigation tasks.

However, this view independence also introduces three main challenges that can make collaborations awkward:

1. Without cues to indicate where the other collaborator is currently looking, participants had to use complex verbal negotiations to orient themselves.
2. Verbal communications could lose context when a collaborator did not share the same view.
3. Local participants missed gestures produced by the remote participants, which caused confusion.

DISCUSSION AND OPPORTUNITIES FOR DESIGN

The 360° camera view opens up new opportunities for a user to engage with the remote scene: our study demonstrated that our remote participants actively explored the scene without mediation from the local participant. This changes the balance of control from conventional video chat scenarios, where the local user mediates the remote user's access to the scene via explicit camera control (e.g. [30,20,38]). Our approach here is similar to several prior works that affix a camera to a local user, but do not require explicit manual control (e.g. [28,21,24]). In following with the line of current research exploring mobile video chat, we focused specifically on collaboration issues as they relate to providing navigation and orientation cues to the local user (e.g. [20,38,22,23,9,24]). Aligned with this prior work, we noted that using the video link to help the local user to navigate the environment, or to provide direction and perspective information was extremely challenging. This is important in many scenarios when trying to remotely direct a person's visual attention—for instance in a remote tour of a building. While people use static visual objects as resources for communication (e.g. "Go closer to the blue object") [7], they cannot rely on these objects to be good visual resources when there is a lot of camera movement, or the objects are abstract. Thus, we argue that we need to restore some of communication resources that we use in everyday life, such as knowing where someone is looking, or conversational gestures. While these issues have been raised before in prior VMC literature, the way in which it manifests is somewhat unique in the 360° context, as local participant experiences the world in an unmediated way in contrast to the remote participant.

Design Opportunity: Bi-directional Gaze Awareness. While our prototype gave remote participants the ability to see where the local participant was looking, this was insufficient in two ways: first, the remote participant would only know what the local participant was seeing *if* s/he could see the local participant's head; second, the local participant had no way of knowing what the remote participant was looking at. The design opportunity here is to provide some mechanism to restore this gaze awareness in both directions. For the remote participant, for example, it might be useful to see an overview of the entirety of the captured space (e.g. as a radar view [13]), along with a clear indication of which way the local participant was facing. This would give awareness when the remote participant was not facing the same direction as the local participant. Similarly, the local participant needs to be provided with some way of understanding where the remote participant is looking. There

may be, for instance, ways of representing this gaze information through head-mounted AR. We have begun to see some early attempts to represent this information visually (e.g. [41,29]), and we need to consider how to evolve these for mobile platforms.

Design Opportunity: Representing Remote Gestures. Similarly, remote participants used hand gestures—even though these hand gestures were not visible to the other participant. For instance, we observed conversational gestures (e.g. Figure 4, left) that were lost, though more importantly, we observed "rehearsals" of physical movements that would be more useful as "demonstrations" of certain physical movements (e.g. twisting one's hand in a certain way to demonstrate a pitch vs. a yaw movement). Providing a mechanism to capture and view these gestures would be useful. We have seen steps towards this idea: [40] allows remote users to point (via a 3D telepointer) to objects in the environment; others have explored how to annotate the remote environment [12,25,9], and more recent approaches actually capture a remote user's hand gestures and visualize these (typically with a head mounted display) for the local user (e.g. [3,4]). Restoring these communicative tools will help make interaction in mobile video chat smoother.

Design Opportunity: Additional Visual Context in Mobile Video Chat. Our 360° prototype gave the remote participant the ability to see the local participant's body, and the most useful benefit of this design was that remote participants could *see* what the local participant was doing with his/her hands. This gave the remote participants the ability to correct actions mid-stream if it was appropriate. This is likely to be useful in a whole range of future applications of mobile video chat—for instance, in remote assistance and remote repair tasks (e.g. [12,40]). While this result is perhaps unsurprising given prior work (e.g. [10]), designing for this kind of visibility in a workspace is even more important in mobile video chat contexts, as the workspace is not fixed as in prior work (e.g. [13]).

The wide visual context provided by the 360° was useful for remote participants: it gave them the autonomy to explore the environment themselves, as well as to provide direction and help to the local participant. Even if 360° video is not used all the time, it seems clear that this additional visual context (i.e. beyond the field of view of a conventional video camera) is useful for mobile video chat. Designers of mobile video chat systems should consider how adding a wider lens or additional cameras can be used to augment mobile video chat.

360° Chat Creates New Asymmetries. Controlling a telepresence robot might afford the same kind of freedom to navigate and explore the local scene as we saw with the 360° prototype. Where the two experiences differ is that navigating a telepresence robot around obstacles is a non-trivial activity that demands attention and up-to-the-moment awareness of the environment [36,39]. In contrast, the 360° prototype was affixed to the local participant, thus obviating the need to navigate the environment. This is reminiscent of the approach in [24,28], where the remote participant is freed from the moment-to-moment challenges of navigating an environment.

Yet, this freedom was balanced by frustration of remote participants when something s/he was looking at went out of view (due to the local participant's movement). This is more likely now (than with a standard mobile video chat) because the remote user has freedom to explore and find things of interest him/her-self. We saw this play out with the ad hoc tasks, where generally remote users would complete these tasks on their own, but that in a rare minority of cases, the local user had kept moving, compromising the remote user's view of the scene.

Thus, our prototype changes a remote user's relationship with the environment compared to a conventional mobile video chat: with a standard mobile video chat, the issue is that remote user cannot *see* enough, whereas with this prototype, the remote user cannot *move* enough (i.e. they depend on the local user to move for them). The location of the camera still remains a monopoly of the local user.

Head-Tracked Displays for 360° Video Chat. Following Kasahara et al. [23,22], our early prototypes outfitted remote participants with a head-tracked display. The challenge with this approach was that the cost of exploring the environment (moving one's head) was very high; in contrast, the tablet interface allowed users to explore the scene quickly (via dragging). Further study is required to determine the trade-offs between head-tracked displays and tablets for 360° mobile video chat.

CONCLUSIONS
Many of the challenges we discuss in this paper have been identified by earlier video media space research (e.g. [32,13]). Yet, addressing these challenges in a mobile video chat scenario demands new types of solutions—in part because we cannot control the environment in the same way as in controlled video media spaces (e.g. [32,33]), but also because there may be opportunities to find solutions that can fit aboard commodity hardware and tools (e.g. mobile augmented reality). Our 360° video chat prototype addresses a recent interest in designing tools that allow us to immerse ourselves in places, environments and activities that are not within our physical reach. Based on the findings of our study, we outline how capturing and communicating gesture and gaze information can make collaborative experiences with mobile video chat more effective and enjoyable.

ACKNOWLEDGEMENTS

REFERENCES
1. Deepak Akkil and Poika Isokoski. 2016. Gaze Augmentation in Egocentric Video Improves Awareness of Intention. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing*

Systems (CHI '16). ACM, New York, NY, USA, 1573-1584. DOI:
http://dx.doi.org/10.1145/2858036.2858127

2. Deepak Akkil, Jobin Mathew James, Poika Isokoski, and Jari Kangas. 2016. GazeTorch: Enabling Gaze Awareness in Collaborative Physical Tasks. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '16). ACM, New York, NY, USA, 1151-1158. DOI:
http://dx.doi.org/10.1145/2851581.2892459

3. Judith Amores, Xavier Benavides, and Pattie Maes. 2015. ShowMe: A Remote Collaboration System that Supports Immersive Gestural Communication. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '15). ACM, New York, NY, USA, 1343-1348.
DOI=http://dx.doi.org/10.1145/2702613.2732927

4. Xavier Benavides, Judith Amores, and Pattie Maes. 2015. Remot-IO: a System for Reaching into the Environment of a Remote Collaborator. In *Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (UIST '15 Adjunct). ACM, New York, NY, USA, 99-100. DOI:
http://dx.doi.org/10.1145/2815585.2815738

5. Jed R. Brubaker, Gina Venolia, and John C. Tang. 2012. Focusing on shared experiences: moving beyond the camera in video communication. In *Proceedings of the Designing Interactive Systems Conference* (DIS '12). ACM, New York, NY, USA, 96-105.
DOI=10.1145/2317956.2317973
http://doi.acm.org/10.1145/2317956.2317973

6. Sarah D'Angelo and Darren Gergle. 2016. Gazed and Confused: Understanding and Designing Shared Gaze for Remote Collaboration. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). ACM, New York, NY, USA, 2492-2496. DOI:
http://dx.doi.org/10.1145/2858036.2858499

7. Darren Gergle, Robert E. Kraut, and Susan R. Fussell. 2013. Using visual information for grounding and awareness in collaborative tasks. Human-Computer Interaction 28, 1: 1-39.

8. Lilian de Greef, Meredith Morris, and Kori Inkpen. 2016. TeleTourist: Immersive Telepresence Tourism for Mobility-Restricted Participants. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion* (CSCW '16 Companion). ACM, New York, NY, USA, 273-276.
DOI=http://dx.doi.org/10.1145/2818052.2869082

9. Omid Fakourfar, Kevin Ta, Richard Tang, Scott Bateman, and Anthony Tang. 2016. Stabilized Annotations for Mobile Remote Assistance.
In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). ACM, New York, NY, USA, 1548-1560. DOI:
http://dx.doi.org/10.1145/2858036.2858171

10. Susan R. Fussell, Robert E. Kraut, and Jane Siegel. 2000. Coordination of communication: effects of shared visual context on collaborative work. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work* (CSCW '00). ACM, New York, NY, USA, 21-30.
DOI=http://dx.doi.org/10.1145/358916.358947

11. Susan R. Fussell, Leslie D. Setlock , Jie Yang , Jiazhi Ou , Elizabeth Mauer , Adam D. I. Kramer, Gestures over video streams to support remote collaboration on physical tasks, *Human-Computer Interactio*n, v.19 n.3, p.273-309, September 2004

12. Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. World-stabilized annotations and virtual scene navigation for remote collaboration. In *Proceedings of the 27th annual ACM symposium on User interface software and technology* (UIST '14). ACM, New York, NY, USA, 449-459.
DOI=http://dx.doi.org/10.1145/2642918.2647372

13. Carl Gutwin and Saul Greenberg. 2002. A Descriptive Framework of Workspace Awareness for Real-Time Groupware. *Computer Supported Cooperative Work (CSCW) 11(3),* 411-446.

14. Steve Harrison (Ed). 2009. Media Space 20+ Years of Mediated Life. Springer-Verlag: London.

15. Keita Higuch, Ryo Yonetani, and Yoichi Sato. 2016. Can Eye Help You?: Effects of Visualizing Eye Fixations on Remote Collaboration Scenarios for Physical Tasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). ACM, New York, NY, USA, 5180-5190. DOI:
http://dx.doi.org/10.1145/2858036.2858438

16. Kori Inkpen, Brett Taylor, Sasa Junuzovic, John Tang, and Gina Venolia. 2013. Experiences2Go: sharing kids' activities outside the home with remote family members. In *Proceedings of the 2013 conference on Computer supported cooperative work* (CSCW '13). ACM, New York, NY, USA, 1329-1340.
DOI=http://dx.doi.org/10.1145/2441776.2441926

17. Hiroshi Ishii, Minoru Kobayashi, and Jonathan Grudin. 1993. Integration of interpersonal space and shared workspace: ClearBoard design and experiments. *ACM Trans. Inf. Syst.* 11, 4 (October 1993), 349-375.
DOI=http://dx.doi.org/10.1145/159764.159762

18. Steven Johnson, Madeleine Gibson, and Bilge Mutlu. 2015. Handheld or Handsfree?: Remote Collaboration via Lightweight Head-Mounted Displays and Handheld

Devices. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '15). ACM, New York, NY, USA, 1825-1836. DOI: http://dx.doi.org/10.1145/2675133.2675176

19. Brennan Jones, Kody Dillman, Richard Tang, Anthony Tang, Ehud Sharlin, Lora Oehlberg, Carman Neustaedter, and Scott Bateman. 2016. Elevating Communication, Collaboration, and Shared Experiences in Mobile Video through Drones. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (DIS '16). ACM, New York, NY, USA, 1123-1135. DOI: http://dx.doi.org/10.1145/2901790.2901847

20. Brennan Jones, Anna Witcraft, Scott Bateman, Carman Neustaedter, and Anthony Tang. 2015. Mechanics of Camera Work in Mobile Video Collaboration. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15). ACM, New York, NY, USA, 957-966. DOI: http://dx.doi.org/10.1145/2702123.2702345

21. Shunichi Kasahara and Jun Rekimoto. 2014. JackIn: integrating first-person view with out-of-body vision generation for human-human augmentation. In *Proceedings of the 5th Augmented Human International Conference* (AH '14). ACM, New York, NY, USA, , Article 46 , 8 pages. DOI=http://dx.doi.org/10.1145/2582051.2582097

22. Shunichi Kasahara, Shohei Nagai, and Jun Rekimoto. 2015. First Person Omnidirectional Video: System Design and Implications for Immersive Experience. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video* (TVX '15). ACM, New York, NY, USA, 33-42. DOI=http://dx.doi.org/10.1145/2745197.2745202

23. Shunichi Kasahara and Jun Rekimoto. 2015. 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), Stephen N. Spencer (Ed.). ACM, New York, NY, USA, 217-225. DOI=http://dx.doi.org/10.1145/2821592.2821608

24. Seungwon Kim, Sasa Junuzovic, and Kori Inkpen. 2014. The Nomad and the Couch Potato: Enriching Mobile Shared Experiences with Contextual Information. In *Proceedings of the 18th International Conference on Supporting Group Work* (GROUP '14). ACM, New York, NY, USA, 167-177. DOI=http://dx.doi.org/10.1145/2660398.2660409

25. Seungwon Kim, Gun A. Lee, Sangtae Ha, Nobuchika Sakata, and Mark Billinghurst. 2015. Automatically Freezing Live Video for Annotation during Remote Collaboration. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*(CHI EA '15). ACM, New York, NY, USA, 1669-1674. DOI=http://dx.doi.org/10.1145/2702613.2732838

26. Seungwon Kim, Gun A. Lee, Nobuchika Sakata and Mark Billinghurst. 2014. Improving co-presence with augmented visual communication cues for sharing experience through video conference. In *IEEE International Symposium on 2014 Mixed and Augmented Reality (ISMAR 2014),* 83-92.

27. David Kirk, Tom Rodden, and Danaë Stanton Fraser. 2007. Turn it this way: grounding collaborative action with remote gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07). ACM, New York, NY, USA, 1039-1048. DOI=http://dx.doi.org/10.1145/1240624.1240782

28. Sven Kratz, Don Kimber, Weiqing Su, Gwen Gordon, and Don Severns. 2014. Polly: "being there" through the parrot and a guide. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services* (MobileHCI '14). ACM, New York, NY, USA, 625-630. DOI: http://dx.doi.org/10.1145/2628363.2628430

29. Jerry Li, Mia Manavalan, Sarah D'Angelo, and Darren Gergle. 2016. Designing Shared Gaze Awareness for Remote Collaboration. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion* (CSCW '16 Companion). ACM, New York, NY, USA, 325-328. DOI=http://dx.doi.org/10.1145/2818052.2869097

30. Christian Licoppe and Julien Morel. 2009. 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* (MobileHCI '09). ACM, New York, NY, USA, , Article 35 , 10 pages. DOI=http://dx.doi.org/10.1145/1613858.1613903

31. Paul Luff, Christian Heath, Hideaki Kuzuoka, Jon Hindmarsh, Keiichi Yamazaki, and Shinya Oyama. 2003. Fractured Ecologies: Creating Environments for Collaboration. *Human–Computer Interaction* Vol. 18 , Iss. 1-2,2003.

32. Paul K. Luff, Naomi Yamashita, Hideaki Kuzuoka, and Christian Heath. 2015. Flexible Ecologies And Incongruent Locations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15). ACM, New York, NY, USA, 877-886. DOI: http://dx.doi.org/10.1145/2702123.2702286

33. Jýrg Mýller, Tobias Langlotz, and Holger Regenbrecht. 2016. PanoVC: Pervasive telepresence using mobile phones. In *Proceedings of 2016 IEEE International*

Conference on Pervasive Computing and Communications (PerCom 2016). IEEE, 1-10.

34. Shohei Nagai, Shunichi Kasahara, and Jun Rekimoto. 2015. LiveSphere: Sharing the Surrounding Visual Environment for Immersive Experience in Remote Collaboration. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction* (TEI '15). ACM, New York, NY, USA, 113-116. DOI=http://dx.doi.org/10.1145/2677199.2680549

35. Carman Neustaedter, Carolyn Pang, Azadeh Forghani, Erick Oduor, Serena Hillman, Tejinder K. Judge, Michael Massimi, and Saul Greenberg. 2015. Sharing Domestic Life through Long-Term Video Connections. *ACM Trans. Comput.-Hum. Interact.* 22, 1, Article 3 (February 2015), 29 pages. DOI=http://dx.doi.org/10.1145/2696869

36. Carman Neustaedter, Gina Venolia, Jason Procyk, and Daniel Hawkins. 2016. To Beam or Not to Beam: A Study of Remote Telepresence Attendance at an Academic Conference. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16). ACM, New York, NY, USA, 418-431.

37. James Norris, Holger M. Schnädelbach, and Paul K. Luff. 2013. Putting things in focus: establishing co-orientation through video in context. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13). ACM, New York, NY, USA, 1329-1338. DOI: http://dx.doi.org/10.1145/2470654.2466174

38. Jason Procyk, Carman Neustaedter, Carolyn Pang, Anthony Tang, and Tejinder K. Judge. 2014. Exploring video streaming in public settings: shared geocaching over distance using mobile video chat. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14). ACM, New York, NY, USA, 2163-2172. DOI: http://dx.doi.org/10.1145/2556288.2557198

39. Irene Rae, Bilge Mutlu, and Leila Takayama. 2014. Bodies in motion: mobility, presence, and task awareness in telepresence. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14). ACM, New York, NY, USA, 2153-2162. DOI: http://dx.doi.org/10.1145/2556288.2557047

40. Rajinder S. Sodhi, Brett R. Jones, David Forsyth, Brian P. Bailey, and Giuliano Maciocci. 2013. BeThere: 3D mobile collaboration with spatial input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13). ACM, New York, NY, USA, 179-188. DOI: http://dx.doi.org/10.1145/2470654.2470679

41. Mengu Sukan, Carmine Elvezio, Ohan Oda, Steven Feiner, and Barbara Tversky. 2014. ParaFrustum: visualization techniques for guiding a user to a constrained set of viewing positions and orientations. In *Proceedings of the 27th annual ACM symposium on User interface software and technology* (UIST '14). ACM, New York, NY, USA, 331-340. DOI=http://dx.doi.org/10.1145/2642918.2647417

42. Matthew Tait, and Mark Billinghurst. 2015. The Effect of View Independence in a Collaborative AR System. *Computer Supported Cooperative Work 24*: 563. doi:10.1007/s10606-015-9231-8

43. Anthony Tang, Omid Fakourfar, Carman Neustaedter and Scott Bateman. 2017. *Collaboration in 360° Videochat: Challenges and Opportunities*. Technical Report 2017-1094-01. Department of Computer Science, University of Calgary.

44. Anthony Tang, Carman Neustaedter, and Saul Greenberg. 2006. VideoArms: Embodiments for Mixed Presence Groupware. In People and Computers XX - Engage (Proceedings of BHCI 2006), Springer.

45. John C. Tang and Scott L. Minneman. 1991. Videodraw: a video interface for collaborative drawing. *ACM Trans. Inf. Syst.* 9, 2 (April 1991), 170-184. DOI=http://dx.doi.org/10.1145/123078.128729

46. Virtual Photo Walks. http://www.virtualphotowalks.org.