

The Nomad and the Couch Potato: Enriching Mobile Shared Experiences with Contextual Information

Seungwon Kim^{1,2}, Sasa Junuzovic¹, Kori Inkpen¹

¹Microsoft Research
Redmond, WA 98052, USA
{sasajun | kori}@microsoft.com

²HITLab NZ, University of Canterbury
Christchurch, 8140 New Zealand
seungwon.kim@pg.canterbury.ac.nz

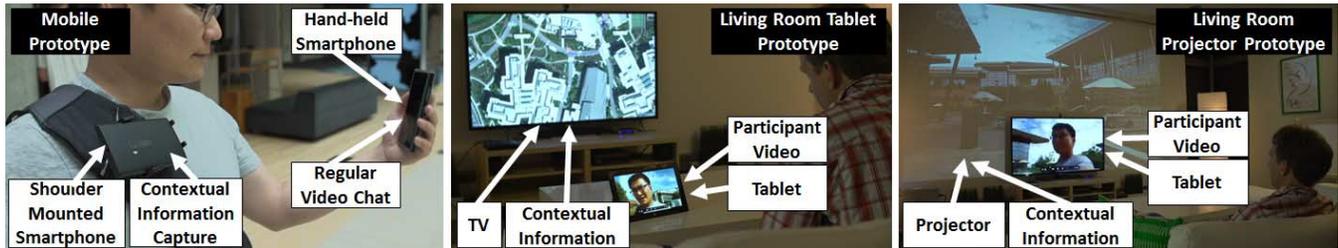


Figure 1. Mobile prototype (left); living room TV-Tablet prototype (center); living room Projector-TV prototype (right).

ABSTRACT

Mobile videoconferencing is increasingly being used to bring remote friends or family along to an activity happening outside the home, such as shopping or visiting a tourist attraction. We explored how including contextual information of the event, in addition to audio and video of the person at the event, impacts the shared experience. We studied three kinds of information: a map showing the position of the person at the activity, a second live video showing what was in front of that person, and periodic high quality images showing what was in front of the person. We carried out a field study with twelve pairs of participants, where one participant (the nomad) was at a self-selected activity while the other (the couch potato) joined the activity from our living room lab. The study results show that including contextual information significantly improved connectedness and the sense of presence for both participants. Each type of contextual information offered unique benefits. The map was used for orientation and to provide directions, the live video for “do you see this” moments and to maintain a sense of liveliness, and the periodic images for “did you see that” moments and to see greater detail. Together they led to smooth view negotiation, activity input from the couch potato, and high levels of engagement.

Categories and Subject Descriptors

H.4.3 [Information Systems Applications]: Communications Applications – computer conferencing, teleconferencing, and videoconferencing.

General Terms

Human Factors.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

GROUP '14, November 9–12, 2014, Sanibel Island, FL, USA.

Copyright 2014 ACM 978-1-4503-3043-5/14/11...\$15.00

<http://dx.doi.org/10.1145/2660398.2660409>

Keywords

Telepresence; consumer; mobile; wearable; video; periodic snapshots; map; field study; connectedness; sense of presence.

1. INTRODUCTION

People are increasingly using mobile devices to capture and share events with friends or family who could not be there in person. For example, they share shopping trips with friends to get advice on what to buy, kids’ soccer games with grandma so that she can attend from across the country, and visits to tourist attractions with loved ones back at home. They share events asynchronously through social media and text messages, and synchronously using mobile video chat apps. In this paper, we focus on synchronous shared mobile experiences.

Synchronous shared mobile experiences go beyond the traditional “talking head” conversations over video chat by enabling people to share rich experiences as they do something together. Two important shared experience factors are connectedness and the sense of presence. *Connectedness* can be defined as the degree to which participants relate to each other in terms of what they are experiencing and feeling. Meanwhile, the *sense of presence* can be defined as the degree to which participants feel as if they are physically side by side during the experience.

We were interested in exploring how to increase connectedness and the sense of presence in shared mobile experiences. Our approach was to provide the person joining the experience remotely with additional live views showing the context of the activity in combination with the live audio and video of the person at the activity. As people already frequently share videos and photos, and sometimes geo-tag them, we wanted to study the benefits of providing contextual views that convey these types of information. We focused on three views: a map showing the position of the person at the activity, a second live video showing what was in front of that person, and periodic high quality images showing what was in front of the person.

To better understand the impact of adding contextual information to mobile shared experiences, we built and field-tested a prototype system that shared contextual views of an activity, as well as, audio and video of the person at the activity. The field study had

twelve pairs of participants, where one participant was at a self-selected outside location while the other was in our living room lab. The field test results showed that including contextual information significantly improved connectedness and the sense of presence for both participants. Each kind of contextual information offered unique benefits. The map was used for orientation and to provide directions, the second live video for “do you see this” moments and to maintain a sense of liveliness, and the periodic images for “did you see that” moments and to see greater detail. Together the different sources of context led to smooth view negotiation, activity input from the couch potato, and high engagement levels.

The rest of this paper is organized as follows. First, we present prior work. Then we describe our prototype and the field study. We end with a discussion, conclusions, and directions for future work.

2. RELATED WORK

The use of mobile videoconferencing is on the rise, and as O’Hara et al. [13] discovered, many mobile video calls occur on the go, outside home and work settings. People are increasingly going beyond video calls that primarily focus on conversations (talking heads) to video calls that enable people to do things together or share rich experiences. Recently, shared experiences have attracted significant media attention, such as the iPad Bridesmaid [9] and a deployed soldier who watched the birth of his son on Skype [18].

Current mobile devices are well suited to sharing experiences because they can capture both video of the participant at the event (front-facing camera) and contextual video of the activity (rear-facing camera). They also support switching between the cameras during a call, which is important for collaboration scenarios. Prior research in the workspace supports this notion with Olson et al. [15] reporting that people like seeing each other during remote collaborations, and Gaver et al. [7] demonstrating that in some situations, they prefer seeing video of the activity rather than video of the people. Generalizing these results to mobile shared experiences suggests that would be useful for a remote attendee to have a view of the person at the event, additional contextual views of the event, and the ability to switch among these views.

Inkpen et al. [8] experimented with contextual information in the form of a second live video stream of an activity. They created and field tested a device that could stream front and rear-facing camera videos at the same time and found that this increased the remote attendee’s engagement and feelings of being together and at the activity. Recently, Procyk et al. [16] explored sharing live head-mounted video in addition to audio between two remote geocaching partners. They found that the video was useful for coarse-grained navigation but not for fine-grained search. The issues with using video for fine-grained tasks were low resolution (640x480) and difficulties with framing. The GestureCam by Kuzuoka et al. [12] and other prior work have addressed the video framing issue by allowing the remote person to direct the camera. The video quality issue remains, however, and is largely dependent on infrastructure.

In addition to live data, contextual information may be synthetic, such as websites relevant to the shared activity, as well as, mash-ups of live and synthetic data, such as a live video annotated with digital content. The Chili system by Jo and Hwang [10] supports a mash-up view in which both users can annotate a shared live video. In addition, Stafford et al. [19] created a system in which a non-mobile user can help give directions to a mobile user. The

non-mobile user places pins on locations in a digital map, and the mobile user sees that data superimposed on a mobile device when the device’s camera sees those locations. Prior work has also enriched the real world with participants’ videos. For instance, Billingham and Kato [1] show videos of remote participants overlaid on the real world when looking at the world through a head mounted display.

These previous works studied contextual information for shared activities. Contextual information has also been studied and applied differently in other research areas. For instance, in ubiquitous computing, devices compute decisions by sensing and responding to the environment around them, which forms the context for the computation [5]. Meanwhile, in studies of organizational processes and meetings, context has been used to describe relationships between people and entities, such as documents and devices, within the institution [4]. We focus on the notion of context for shared experiences, where context of a shared activity conveys additional awareness of the activity to remote attendees.

As a final note on prior work, while we focused on connectedness and the sense of presence, there are other factors that also impact remote shared experiences. At a low level, these factors include aspects such as the richness of the communication channels [2][6], mutual and directional gaze [14][17], and referential awareness [3]. These low-level factors drive high-level experience metrics such as task completion time [6], trust [2], and others, including the two we focused on, the sense of presence [14] and connectedness.

To summarize prior work, some systems have provided contextual activity information to a remote participant, but there have been few studies of this information, especially in the wild. Even fewer of these evaluations were carried out for activities that users themselves chose as something they would like to share remotely. Finally, prior work has not studied how users manage multiple types of contextual information when they are available simultaneously. Our work addresses these outstanding issues.

3. PREPARING FOR FIELD STUDY

To better understand how contextual activity information impacts connectedness and the sense of presence during a shared experience, we needed to observe it in the wild rather than in a lab. In a lab, it is not practical to replicate multiple real world shared events because of costs and scale. Thus, a lab could support only a few shared activities, which may not be reflective of the types of activities people actually want to share in their everyday life. As a result, we decided to run a field study in which the activity and the person at the activity were out in the wild while the remote attendee joined from our lab. While ideally the remote attendees should have joined from their own homes, having them in the lab made both administrative and data collection tasks more manageable.

By taking the shared activity out of the lab, the equipment and software that we could use for the study were limited. In particular, we could not rely on super-high resolution cameras, virtually unlimited bandwidth, and high-quality audio that exists in lab settings. Instead, we had to use mobile devices to capture both the person at the event and the contextual information of the event. We also had to rely on cellular networks for connecting the lab and the activity. The use of real-world devices and networks undeniably impacted the quality of the shared audio, video, and data. However, and more importantly, this quality was realistic.

The preparation complexity was further increased by the fact that no existing applications share contextual information of an activity together with the audio and video of the person at the activity. Thus, we built our own prototype to support this kind of sharing. To reduce the prototype build time, we leveraged a combination of existing commercial products and added custom components where required. The result was a prototype with a mobile and a living room end-point that connected our lab to an event in the wild.

3.1 Mobile Prototype

The mobile prototype had to perform three tasks: present the audio and video of the remote attendee; capture the audio and video of the person at the activity; and capture contextual information.

Standard video chat applications on a smartphone with a front-facing camera already accomplish the first two tasks. As a result, one part of the mobile prototype was a *handheld smartphone* (Figure 1 left), in our instance a Lumia 920 Windows Phone 8, running Skype mobile. As with most video chat applications and smartphones today, users could switch to the rear-facing camera if they wanted to use the phone to show a video of something in front of them. To reduce the impact of noise in outdoor settings, a headset was connected to the phone. The headset also made it possible to have audio when users placed the phone in a pocket or purse to free up their hands.

Standard video chat applications, however, do not currently capture contextual information in the form of live video, snapshot history, and user location together with the audio and video of the person using the phone. For this, we created a new system. Since a smartphone could capture photos and videos using one of its cameras and user location using the built in GPS sensor, we used a *second, shoulder-mounted, wearable smartphone* to capture the contextual information. We again used a Lumia 920. We mounted it in landscape orientation at shoulder level using a sash-like belt so that the rear camera was facing forward (Figure 1 left). Thus, users could not see or use the wearable phone's screen as it was pressed flush against their bodies.

The wearable smartphone captured three types of contextual information: **Map** – GPS location at one second intervals; **Video** – a video showing what was in front of the user wearing the phone; and **Images** – automated high-quality images taken every five seconds and capturing what was in front of the user. Because only one application at a time can use the camera on the Lumia 920, we used the rear camera to capture Video and screenshots of the camera preview window as Images. Thus, the resolution of the images matched the 800x480 resolution of the Lumia 920 screen.

To share the contextual information with the remote attendee, we used two separate channels. To share Video information, we setup a second Skype session with one-way muted video stream from the wearable phone. To share the Images and Map information, the wearable smartphone uploaded data to an Azure cloud service that could serve that data on demand.

3.2 Living Room Prototype

The living room prototype space needed to look like a living room so that users in the room could at least partially forget that they were in a lab. Thus, we created a space that had a couch, a 55" TV placed a comfortable distance from the couch, a coffee table, and some simple living room décor, such as rug, floor lamps, fake plants, and wall hangings (Figure 1 center and left).



Figure 2. The context view stretches across the TV and the projector when the hand-held phone is put away.

The prototype had to execute three tasks: capture the audio and video of the person in the room; present the audio and video of the person at the activity; present contextual information of the activity.

To capture the video of the person in the living room, we placed an HD webcam on top of the TV and digitally zoomed it in on the person. To capture the audio, we placed a ClearOne speakerphone on the coffee table. The speakerphone also played back the audio of the person at the activity.

Displaying content is more complex in the living room than in the mobile case. The mobile prototype had to display only the video of the remote attendee, while the room prototype had to show both contextual information and the video of the person at the activity.

An important question was whether to display contextual information and video of the person at the activity on a single display or multiple displays. With a single display, both could have been shown at the same time using a tiled or picture-in-picture view. However, this could have made some things difficult to see. Another possibility was to let users choose what to show on the display. However, switching between the views would have been burdensome to users. We instead chose to utilize two displays, so that the video of the person at the activity and the contextual information could be shown simultaneously on different displays. We used a TV and one additional display in the living room.

A related question was what size display should be used to show contextual information. Intuitively, a larger display should increase connectedness and the sense of presence because it brings the person in the living room closer to the action. To test this intuition, we explored two dual-display configurations with drastically different display sizes. In the *TV-Tablet condition*, a Surface Pro tablet was placed on the coffee table to serve as the second display (Figure 1 center). In the *Projector-TV condition*, we placed an In-Focus IN126ST short-throw projector under the coffee table and it projected a 14' diagonal image to the sides and above the TV (Figure 1 right). In both conditions, the video of the person at the activity was shown on the smaller display and the contextual views were shown on the larger display.

In order to display the participant video and contextual information on the two displays, the living room used two desktop computers. One computer joined the Skype session with the handheld phone in the mobile prototype and showed the participant video from that phone on the smaller display that was being used. The second computer joined the Skype session with the wearable phone and displayed the context video from that phone on the larger display that was being used. This desktop did not transmit any audio or video back to the wearable phone. In addition, it downloaded the Images and Map contextual information from the Azure cloud service. For the Map data, it

converted the GPS location to a pin on a Bing map and a line showing the path of the person at the activity.

The living room prototype allowed users to choose which context view to show and then interact with it. In the Map view, users could change the zoom level and clear the movement history. Map tilting, panning, and searching were disabled for simplicity. In the Images view, users could navigate through image history and jump to the latest image. The system did not provide any ways to interact with the Video view other than to watch it.

Finally, the living room prototype was mindful of the times when the people at the activity put their handheld phone away (e.g., into a pocket or purse) to free up their hands. When the handheld phone was put away, the video captured by the phone was simply black and not useful. In the Projector-TV condition, this meant that the TV occluded a part of the projected image for no reason. To improve this experience, when the system detected that the handheld phone was put away, it made things appear as if the TV and the projector were one large contiguous screen (Figure 2). The prototype accomplished this by determining the part of the projected image that was occluded by the TV and sending that part of the image to the TV. The prototype also enabled users to manually activate or deactivate this behavior.

4. FIELD STUDY

With the prototype system in place, we were ready to evaluate the impact of contextual information for activities people want to share. To identify these activities, we decided to allow the study participants to choose their own events to share, people to share them with, and their preferred time and location. While affording this freedom to the participants made our study more realistic, it also added some complexity.

4.1 Complexity

While the use of a field study was beneficial in terms of ecological validity, it also introduced additional complexity. First, there were cross-session differences since participants were allowed to choose their activity. This enabled us to observe different activities that took place in a variety of settings with different levels of ambient noise, lighting, and network coverage. However, it was difficult to compare results across the sessions. Second, our study design introduced a within-session difference because the participants out in the wild and those in the lab had asymmetric experiences. Thus, we had to analyze their feedback separately. Third, our use of two different display setups and multiple forms of contextual information in the living room added another layer of complexity. Therefore, we had to tease out the impact of each of these factors as we analyzed the study results. Finally, it is also important to recognize the overall quality and potential variability in the cellular network for connectivity. In particular, the quality of mobile video chat over 4G/LTE is still poor and will continue to be so in the near future. In the rest of this section, we describe our methodology.

4.2 Participants

For our field study, it was important to recruit participants who were already familiar with standard video chat features. Otherwise, their feedback could have been influenced by the novelty of using video chat rather than focusing on feedback about contextual information in mobile shared experiences. Therefore, we recruited twelve pairs of people who had used video chat at least once a month during the last six months. Our participant pairs were family members, couples, and good friends.

During recruitment, the recruits were told that one of them would attend an activity in person while the other would join them remotely. We provided some suggestions for activities (e.g., visiting a museum, attending a child's soccer game, etc.) but left it up to them to choose the event and who would attend it remotely.

4.3 Procedure

At the start of each session, one participant came to our lab while the other participant went to the event. We refer to the participants in the lab and at the event as *inside* and *outside participants*, respectively. One study administrator met with the outside participant while two administrators met the inside participant.

We started each session by explaining the purpose of the study. Then we demonstrated basic Skype features to the outside participant, such as switching between front and rear-facing cameras (on the handheld phone). We also explained that they can put the phone away in their pocket or purse any time they desired, and that if they did so, they could still talk to their partner through the headset. We then mounted the wearable smartphone on the participant and started the shared experience.

Each shared experience was divided into three sections: warm-up, familiarization, and free play.

Warm-up: Each shared experience started with a five-minute warm-up period of regular Skype video chat. Even though the outside user was wearing the second smartphone, we told them that it was not active. Meanwhile, the inside user could see video of the outside user on either the tablet or the TV, depending on which living room condition they were using. We used the warm up period to work out connection issues and get the participants talking. At this point, they were using a system that they were comfortable with since it was just basic Skype video chat.

Familiarization: Following the warm-up period, the participants spent three five-minute periods using each type of contextual information our system could provide. We explained each view to them as they started to use it. Since the outside participant could not see the living room, the administrator showed them what the current condition looked like using photographs. We used these five-minute periods to get the participants familiar with the contextual information. They could not choose what information they were seeing or otherwise interact with it. Of the twelve pairs in our study, six used the TV-Tablet condition in the living room and six used the Projector-TV condition. Within each condition, the order in which the contextual information views were used during the familiarization stage was counterbalanced.

Free-play: After the familiarization period, the participants entered a twenty-minute free-play period during which the inside user could control what contextual information was shown. We explained to the user how to 1) switch between Map, Video, and Images views; 2) zoom and clear the location history in the Map view; 3) navigate through image history in Image view; and 4) turn the "phone in pocket" feature on and off in any view. The free-play period was the period we were most interested in. We felt that it could help us better understand user preference for the contextual information, inform us of whether they would switch among the different views, and when a particular view is better than the others. Most of our data collection, analysis, and reported results are with respect to the twenty minute free-play period.

4.4 Data Collection

To assist in managing the study complexity, we collected multiple forms of data.

Questionnaires and interviews: We collected subjective data via two questionnaires and a debrief interview at the end of each session. The first questionnaire was completed at the start of each session and included demographic questions. At the end of the free-play phase, the participants completed the second questionnaire, which asked them to rate the usefulness, enjoyment, feeling of connectedness, feeling of being together, and the feeling of being a part of the activity for each contextual view. It also asked them to rank the views from best to worst and included two open ended questions about their likes and dislikes for each view. Once they completed the questionnaire, we conducted a semi-structured interview with the inside and outside participants separately.

Logs of inside user activity: To complement the subjective data, we also logged the inside users' interactions with the contextual views. These logs included times at which the participants changed their contextual views, their interactions with the Map and Images views, and if the phone in pocket feature was on or off. From this data, we can obtain information such as the number and frequency of view changes and total time spent in each contextual view.

Observations, pictures, and screen and video recordings: To help us better understand the questionnaire and log data, we took notes and pictures during the sessions. Moreover, we video recorded the inside participant and screen captured their screens, while the outside admin captured the outside user's activity with a GoPro. From this data, we could get information that we could not get otherwise, such as whether the front or rear-facing camera was being used on the handheld phone.

4.5 Results

Despite our best efforts, the video quality during the study was fairly poor. As a result, Video quality was lower than Images quality. Unfortunately, our participants expected very good video because of two reasons they mentioned in their comments. First, they thought that mobile video on 4G/LTE would have similar quality as on their home Wi-Fi. Second, they believed that the quality of live mobile video should be similar to that of pre-recorded content, such as videos from a GoPro camera. These expectations influenced the feedback about the live video streams.

As predicted, our participants chose to share a variety of activities which resulted in many differences across the sessions. To help us understand the overall data, we first qualitatively analyzed what activities people shared and how our prototype system was used to share them. This initial analysis provided the scope for a deeper investigation into the usefulness of contextual information. Therefore, we start by presenting an in-depth description of what the participants did during the sessions.

4.6 Descriptions of Individual Sessions

The shared activities included two playtime sessions in a park, three shopping excursions, three tours, hanging out on the beach, walking the dog, visiting a farmers market, and fly fishing. We describe six unique activities in detail as we observed them to be representative of the remaining six, for which we provide only a brief summary. Three of the session we describe used the Projector-TV condition in the living room and three used the TV-Tablet condition. We focus on the twenty minute free-play periods in our descriptions.

Playtime in park (Group 2): In this session, a mom used the Projector-TV condition to join her husband as he took their young son to play in a park. The father and son walked around, climbed



Figure 3. (left) view of snack time from living room: front cam on TV and snapshots on projector; (right) outside view.



Figure 4. China Town from living room showing full screen map (left); walking around outside (right).



Figure 5. Shopping from living room: snapshot on TV with front cam (left); view from inside a store (right).

tables, visited a gazebo, and ate some snacks. Mom continuously interacted with them and suggested things to do in the park.

The mom spent almost the entire time in Map and Images views. She used Map mostly when the father and son were walking and often told them what was nearby: “at the end of that little curved road you’re going to be on, there’s going to be something, I don’t know what it is, but it’s an oval thing that’s green.” When the dad and son were stationary, she used mostly Images, which worked especially well when the father put the hand-held phone in front of their son (Figure 3). The son liked seeing his mom, while the mother enjoyed watching him stuff food into his mouth. She frequently navigated through the snapshot history until she found the photo she liked best. Her comments reflected her usage pattern: “I liked the map the best because I can zoom in and out and see where they were” and “I liked the snapshots because I could choose to go back and forth”. She did not feel the context video was useful because “that’s what Skype was already doing.” She also liked the ability to change views: “sitting and watching just the plain video is not as fun as when I have control.”

The dad enjoyed experience because “[the mom] could see around us, and what we were doing, and she could see us”; however, he found the system cumbersome to manage while also taking care of his son: “it was hard because there was gear strapped to me, and the headphones, and the other camera”.

China Town tour (Group 5): In this session, one participant gave a tour of Seattle’s China Town to one of his friends. The inside participant joined using the Projector-TV condition. They talked about nearby restaurants and the history and future of China Town.

The inside participant used all of the contextual views and frequently switched among them. Overall, he used Video and Images the most, and he used the Map for location references and to ask location specific questions. At times, he surprised his friend by asking questions about things he could see in the Images or Map views. For example, he asked “*Hey, are you walking towards the big arch thing?*” when he saw the China Town Gate on the map (Figure 4), or “*Hey what’s that Gossip restaurant on the corner there?*” when the sign appeared in a snapshot. In both cases, the friend in China Town paused in surprise at these questions because he did not expect his friend could see these things. However, even though the inside user was actively participating in the tour, when asked if the system helped him feel like he was there, he said “*not at all, it’s no substitute for being there.*” He added that “*I found myself gravitating toward the existing technology, like a basic Skype camera that you can switch between front and back views as being more familiar and comfortable.*” He liked that because “[the outside user] could control it and focus my attention on something.”

Interestingly, the perspective from the outside user was completely opposite. He felt that providing the additional contextual views automatically from the wearable phone made it feel like his partner was there: “*when I had regular Skype it didn’t feel like he was right next to me. I knew I was talking to him over Skype. No major difference over phone call. When I had the other camera I was more inclined to show him things, because I was able to gesture freely, because I had this one in my pocket. I knew he could see what I was seeing. He is getting the same experience, or a similar experience, to that if he were standing right next to me.*”

Shopping at a mall (Group 6): In this session, a boyfriend remotely joined his girlfriend for a shopping trip. He used the TV-Tablet living room condition. They shopped for clothes and vitamins.

The boyfriend used all contextual information, although he used Images the most and Video the least. He generally enjoyed the experience. At one point he said “*you know normally I hate shopping but this experience has been great sitting here on the couch.*” Once she held up a shirt for him to see, and he could see details such as the price tag because he was using the Images view: “*I can see a white shirt. For \$19*”. Later on, when shopping for vitamins, he used the Images view to find them himself. When she looked at the shelf, he said “*oh yes I can see something there*” (Figure 5). During the debrief interview, he said that “*overall, I was a bit surprised that I did have a feeling of being present.*” He mentioned that “*Skype by itself is not enough, but when you add more things that you can see at the same time, that gives a better feeling of the activity to the person.*”

The girlfriend wanted to see her boyfriend while shopping so she kept the front camera active on the handheld phone most of the time. She explained that she liked that she “*could see his reaction, if he liked it or not*”. One challenge this pair had was that she walked very fast, which made the video difficult for him to watch.

Beach (Group 7): In this session, two friends went to a farmers market and then walked along a beach. The inside participant joined using the Projector-TV condition.

The inside participant used all contextual information. Initially, she used Images and Map views interchangeably to get a sense of where her friend was. Later, she used Video and Images to see what her friend could see, such as a beautiful sunset (Figure 6).



Figure 6. Sunset at beach from living room: front cam on TV and snapshot (left) and (right) outside view.



Figure 7. Hanging out in the living room: no video on tablet and snapshots on TV (left); outside view (right).



Figure 8. Farmers market from the living room: rear cam on TV and snapshot on projector (left); outside view (right).

She used Video when she wanted to see something that was currently happening, such as people walking by or paddle boarders in the water. According to her, what she liked best was “*switching between the videos and the images and Skype with her so that I could really see what she was talking about*”.

The outside participant also liked that “[my friend] could switch between what worked for her ... which was nice as well for me to not have to be turning the camera back and forth. That made it feel like she was here”.

Walking the dog (Group 8): In this session, a teenage girl walked her dog while a friend joined using the TV-Tablet living room condition. They generally just hung out during the session.

The teenager outside felt self-conscious about holding the phone out in front of her, so she kept it in her pocket the entire time. She explained that “*I like being able to see the person, but I don’t like when you walk around you have to hold it up*”.

The inside teenager used only Video and Images, so that she could see what her friend could see. At one point, the girl outside asked her friend “*can you see Seduce right now?*” to which she responded “*yeah, I saw it in the pictures*” and then she switched to Images (Figure 7). When asked how much the system helped her feel like they were together, she said “*a lot, I really liked the picture and video a lot just because I could see if there was a place there ... and it was really cool to talk to her about it while we look at it.*”

Farmers market (Group 11): In this session, two friends visited a farmers market. The one in the living room joined using the TV-

Tablet condition. They visited booths, tried samples, and generally enjoyed the market.

The inside friend briefly used the Map view a few times to see where her friend had been. For instance, one time she looked at the map, she commented “it looks like you’ve been around the entire place!” She spent the majority of time in Images view to see various fare and help her friend pick out things to taste test (Figure 8). At the end, she said “I felt pretty connected with her. It was neat to be able to click the map view to see where she was.” She added, “I really liked the pictures and being able to control what I see.”

The friend at the market was another participant who was self-conscious about wearing the prototype, and as a result, she was more negative about the setup. She also commented on the asymmetry of the experience: “my experience was probably really different than hers, because she is the one who is seeing everything and I’m just walking around really, I wasn’t really watching her as closely on the monitor because she wasn’t doing anything.”

Remaining groups: In Group 1, two friends in attended a **Viking festival** where they walked around booths and a Nordic museum, ate some food, and watched a mock sword fight. In Group 3, a boyfriend and girlfriend went **shopping**, bought cupcakes, window shopped, and looked at shirts for him and purses for his mom. In Group 4, a husband joined his wife at an **RC airfield** where they walked around and watched a person fly a plane. In Group 9, a husband and his wife went **fly fishing** and talked about the fishing area and strategies. In Group 10, two brothers went **shopping** at a mall. Finally, in Group 12, a husband joined his wife as she took their daughter to **play at a park**.

Summary of sessions: As these session observations illustrate, people used contextual information in a variety of ways. Each contextual view was useful at least some of the time, and some views were used in specific situations. The observations also show engagement by the participants in the living room, activity input from them, and smooth negotiations about what they looked at. Next, we unpack these findings and present higher level insights about the usefulness and impact of contextual information.

4.7 Contextual Information Evaluation

Based on the observations from the sessions, each of the three kinds of contextual information was useful. We were interested to see which of them had the highest impact on connectedness and the sense of presence, and if one was more useful than the others. To this end, we analyzed the rankings and the ratings from the participants’ questionnaire answers.

4.7.1 Connectedness and Sense of Presence

We found significant differences for the different types of contextual information on inside users’ ratings of *connectedness*, *feeling of being there*, and *feeling part of the activity* ($p < .01$, Table 1). The pairwise differences reveal that the participants rated having the Images view in addition to a pure Skype call

significantly higher than vanilla Skype (SkypeOnly) (*connectedness*: $Z = -2.85$ $p = .004$, *being there*: $Z = -2.84$ $p = .004$, and *part of the activity*: $Z = -3.20$ $p = .002$). Moreover, they rated the Images view significantly higher than the Map view for *feeling of being there* and *feeling part of the activity* ($Z = -2.81$ $p = .005$, $Z = -2.99$ $p = .003$). These results suggest that contextual information increased connectedness and the sense of presence for the inside participant. The Images view was the most effective, while the poor quality of the video worked against the Video view.

The debrief interviews corroborated the questionnaire results. Nine of the twelve participants reported that the system helped them feel like they were at the activity with their partner. Three stated that they felt like they were there with the Images but not with the Video view, and three reported that the system did not help them feel like they were at the activity, all giving poor video quality as the reason.

Meanwhile, the outside users experienced only two conditions, which were vanilla Skype (SkypeOnly) and contextual information with Skype (SkypeAndContext). In the SkypeAndContext condition, they knew the inside user had three types of contextual information available, but they did not see what the inside user saw. Nevertheless, we also found significant differences for having and not having contextual information on outside users’ ratings of *connectedness*, *feeling of partner being there*, and *feeling of partner being part of the activity* ($p < .05$, Table 1). They rated SkypeAndContext significantly higher than SkypeOnly on all three measures: (*connectedness*: $Z = -2.26$ $p = .026$, *partner being there*: $Z = -2.72$ $p = .006$, and *partner being part of the activity*: $Z = -2.54$ $p = .011$).

During the debrief interview, many of the participants commented on feeling more like their partner was with them in the SkypeAndContext condition compared to the SkypeOnly condition (see earlier comments from Groups 2 and 5). For instance, P_{7out} expressed that having Skype and contextual information was twice as good as regular Skype: “I had face to face with her, and then she could be looking where I was seeing as well, it was literally like she was here, she had every angle, pretty much, as if she was here. It’s like Skype but kind of twice as good that you can see both the person and the surroundings”.

4.7.2 Usefulness and Enjoyment

Like the inside users’ ratings for connectedness and sense of presence, we found significant differences for the different contextual views on inside users’ ratings of *usefulness* and *enjoyment* ($p < .01$, Table 1). Examining the pairwise differences, the inside participants rated having the Images contextual view in addition to a Skype call significantly higher than having only vanilla Skype (SkypeOnly) (Wilcoxon: *usefulness* $Z = -3.11$ $p = .002$, *enjoyment* $Z = -2.83$ $p = .005$). In addition, the Images view was rated significantly higher than the Map view for *enjoyment* ($Z = -2.69$ $p = .007$). These results suggest that including contextual information was generally beneficial for the remote attendee.

Table 1. Mean (SD) ratings of the contextual views on a 10-point scale where 1 is low and 10 is high (* $p < .01$; ** $p < .05$)

	Inside Participant					Outside Participant		
	SkypeOnly	Map	Images	Video	Friedman Tests	SkypeOnly	SkypeAndContext	Wilcoxon Tests
Connectedness	6.08 (2.11)	6.75 (1.71)	7.75* (1.77)	7.00 (2.22)	$\chi^2_{12,3} = 12.34, p = .006$	6.67 (1.30)	7.50** (1.17)	$Z = -2.26 p = .026$
Being there	5.58 (2.35)	5.33 (2.27)	7.08* (2.15)	6.17 (2.73)	$\chi^2_{12,3} = 14.15, p = .003$	6.75 (2.67)	7.92** (1.24)	$Z = -2.72 p = .006$
Part of activity	5.42 (2.02)	5.25 (2.01)	7.33* (2.15)	6.33 (2.46)	$\chi^2_{12,3} = 22.21, p < .001$	6.58 (1.44)	8.08** (1.24)	$Z = -2.54 p = .011$
Usefulness	5.87 (1.95)	6.58 (1.56)	7.83* (1.40)	6.58 (2.28)	$\chi^2_{12,3} = 15.46, p = .001$	7.5 (1.73)	8.25 (1.14)	$Z = -1.27 p = .204$
Enjoyment	6.08 (2.47)	6.42 (2.11)	8.17* (1.53)	6.83 (2.69)	$\chi^2_{12,3} = 15.25, p = .002$	7.42 (1.62)	7.67 (1.97)	$Z = -0.73 p = .467$

They also show that a history of periodic high quality snapshots is the most useful and enjoyable form of contextual information our prototype provided.

The results for the Map view were activity dependent. In some cases, the map provided little additional information because the outside user was mostly stationary or the inside participant knew the area well. For instance, P_{12in} joined his wife and daughter for a walk that was within three blocks of their home. On the other hand, three participants ranked it as very useful for their activities.

The outside users, in general, rated their experience high on both *usefulness* and *enjoyment* for the SkypeOnly condition (median 7.0 and 7.5, respectively) and the SkypeAndContext condition (median 8.0 for both). While the ratings were higher for SkypeAndContext, the differences were not statistically significant (*usefulness*: $Z=-1.27$ $p=.204$, *enjoyment*: $Z=-0.73$ $p=.467$).

Overall, these results indicate that including contextual information did not burden the outside participants. Since they did not see any of the contextual views, it makes sense that they would not report these views personally useful or enjoyable.

4.7.3 Summary

Overall, our users found contextual information useful, enjoyable, and helpful with respect to connectedness and the sense of presence. Interestingly, the outside users reported increased connectedness and a higher sense of presence even though they could not see the contextual information. Overall, the Images view was found to be the best. The Video view suffered from poor video quality, while the Map view was useful only in certain situations. In the next section, we present the analysis of how the different views were used.

4.8 Contextual Information Use

Even though our participants rated the Images view as the best view on all measures, the Map and Video views were still used and were found to be useful. In this section, we present a closer analysis of how each of the views was used.

From the log data, we computed the number of times each view was utilized and the amount of time it was displayed (Map 15%, Images 47%, and Video 38%). Repeated measures ANOVAs with living room setup as a between subjects factor revealed a significant main effect for the amount of time each view was shown ($F_{2,20}=6.625$, $p=.006$), but no significant interaction effect ($F_{2,20}=3.418$, $p=.053$). The post-hoc pairwise comparisons revealed that the Images view was viewed significantly more than the Map view ($p<.01$). Usage of the Video view was in between Map and Images but not significantly different than either of them, which is consistent with the participants' ratings for the three views (Table 1).

4.8.1 Specific Uses of Contextual Views

During the sessions, we observed that there were specific instances when the inside participant looked at a particular context view.

The Map view was often used to get a sense of where the outside participant was. Sometimes these were quick glances to get a location reading. For instance, the living room participant getting a tour of China Town from his friend in Group 5 often glanced at the map to see where in China Town they were. Other times, the Map uses were longer to get a better sense of where the outside person is and where they have been. For instance, the inside

participant joining her friend at a farmers market in Group 11 used the map to check how much of the farmers market her friend had visited.

While the use of the Map view was mostly triggered by inside users' curiosity, the use of the Images view was often triggered by the outside user. Specifically, when the person outside asked the one inside "did you see that", the inside user switched to Images to review what occurred. An example of this was the teenager in Group 8 asking her friend in the living room if she saw some nearby store, which prompted the living room teenager to switch to the Images view. At times, the inside user decided to review what happened without a prompt from the outside participant. For example, the mom from Group 2 looked for the perfect shot of her son as he played with dad in the park. Finally, the Images view was used often to see details, which is not surprising given the poor quality of the video. For instance, in Group 6, the boyfriend in the living room used the Images view to see the details and price of a shirt his girlfriend held up for him at a store.

Video, on the other hand, was used mostly to see things happening live. For example, in Group 7 where two friends shared a walk along a beach, the outside participant saw paddle boarders and asked the inside participant if she could see them. The inside participant immediately switched to the Video view to see what was going on. The Video was often used in coordination with Images, with quick switches between them depending on whether the inside person was interested in things happening live or things that have already happened. Finally, we observed that the inside participant often wanted some video of the activity or of the outside partner to get some degree of live, real-time data. Hence, when the handheld phone was put away, use of the Video view increased. In the next section, we discuss this and other patterns of contextual view use that were driven by the state of the handheld phone.

4.8.2 Patterns of Contextual View Use

We discovered three patterns of contextual view use that were driven by what the outside user was doing with the handheld phone. To show these patterns, we created timelines of the free-play periods showing the contextual view that was used and the handheld phone state. One pattern mentioned already was that the participants always wanted some form of live, real-time video during the activity. For instance, participant who walked her dog in Group 8 put the phone away the entire time (Figure 9 top), which prompted the inside user to use the Video view heavily.

When the outside participant mostly used the front facing camera on the handheld phone, the living room participant hunted for views of the activity. For example, in Group 7's visit to a beach,

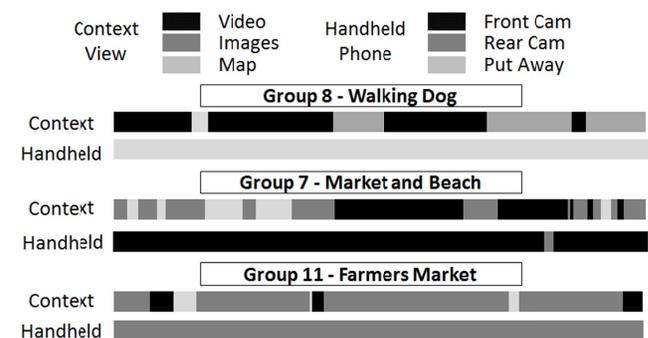


Figure 9. Timelines of context view use and handheld phone state for three different groups.

the outside user mostly used the front facing camera. As the session timeline shows (Figure 9 middle), the inside user changed views frequently as they tried to get an angle of the activity.

Finally, when the outside participant mostly used the rear facing camera, the inside user tended to use the Images more. For instance, during Group 11's visit to the farmers' market, the outside user only used the rear facing camera, which lead the inside user to mostly use Images (Figure 9 bottom). Overall, for many participants, this was reported to be the favorite combination of views – rear-facing camera video combined with Images.

4.8.3 Impact of Contextual Views

The uses of the contextual views illustrate that the living room participants were engaged and trying to get the best view of the event. They also commented that they were more engaged because they had control of the context views. For example, the mom in Group 2 who was joining playtime at a park mentioned that having some control is more fun than just watching plain video.

In addition to following the activity with contextual views, the inside participants were also able to use the views to contribute to the activity. For instance, several of the inside participants used the Map view to give directions to the outside participant. Moreover, the details afforded by the Images view let them drive the conversation by asking specific questions that were not necessarily related to what the outside person was talking about at the moment.

Interestingly, although the participants in the living room often switched among the various contextual views, none of the outside participants were able to tell which view their partner was looking at unless they explicitly told them or commented on the view. None of the outside participants expressed concern over this. In fact, many commented that they liked their partner having the freedom of choosing whatever view they wanted: *"the fact that I don't have to constantly have to switch between the two for his convenience is a blessing"* P_{10out}. Some participants, like the one on the beach in Group 7, found that the inside participant's ability to switch among the context views on their own made it feel like they were there at the activity because they could see more. At times, the outside participants were surprised at all the things the inside participants could see. For instance, the person giving the tour of China Town in Group 5 was stunned a couple of times when his friend asked him about things nearby that he did not think his friend could see.

4.8.4 Summary

In summary, each contextual view contributed uniquely to the shared experience. Moreover, the ability of the inside person to switch among the context views, together with the ability of the outside person to choose how to use the hand-held phone, resulted in smooth negotiation of what view the inside participant was looking at. Finally, the availability of the various contextual views kept the inside participants engaged in the activity and enabled them to offer activity related input back to the outside people.

4.9 Other Results

In addition to analyzing the contextual views, we were interested in privacy concerns, the experience of wearing the mobile prototype, and the how display size impact the experience in the living room.

4.9.1 Privacy

Our participants did not express any concerns about privacy during the study. As our sessions involved family, loved ones, and good friends, we did not expect such concerns, except perhaps for times when they went to the bathroom (which never happened).

However, privacy concerns were raised by people around our participants. For instance, in both of the farmers market sessions, when our participant approached booths selling art pieces, the sellers asked that no pictures of the art were taken. Meanwhile, in one of the mall shopping events, a security guard asked us to stop filming. Even though our participants informed the offended party that they were just video chatting, the concerns remained.

4.9.2 Mobile Device Form Factor

The participants liked the general hands-free nature of the wearable phone. Two participants commented on the hands-free benefit: *"When I didn't want to hold up the phone with Skype she could still see where I was"* P_{8out} and *"I could put the phone in my pocket and he could still see"* P_{12out}. This was important for participants whose hands were occupied during the activity, such as holding a child, walking the dog, or carrying parcels, and for participants who felt self-conscious about having the phone out. As P_{12out} expressed: *"I liked that I could put the regular phone in my pocket if I had to, and so the other one was still taking pictures. Or if I'm dawdling, looking somewhere else, he is still getting something."*

The main issue with our wearable camera prototype was that it was awkward and cumbersome. Two participants felt extremely self-conscious about wearing it and explained they *"didn't like walking around with [it]"* P_{8out} and *"looked kind of funny [and felt] stupid"* P_{11out}. The wearable setup was in fact one of the things least liked about the whole experience. Four participants mentioned the bulkiness of the belt to which the phone was attached. When we probed the participants on what type of wearable form factor they would prefer, they generally desired something compact, discreet, sleek, waterproof, rugged, hands-free, and voice-activated.

4.9.3 Living Room Form Factor

We were also interested in understanding how the sizes of the displays in the living room affected the experience. During the debrief interview, we asked each inside participant to compare the two living room conditions. Since each participant experienced only condition one during the study, we verbally described the other condition to them. Those who tried the Projector-TV condition commented that the video was stretched out and pixelated, and that the same quality video on a smaller display may have felt better. This belief is supported by the questionnaire results. For both *connectedness* and *feeling of being there*, Video was ranked significantly higher in the TV-Tablet setup (median=1 for both measures) than in the Projector-TV setup (median=3.5 and 3) (*connectedness*: $Z=-2.19$, $p=.041$, *being there*: $Z=-2.18$, $p=.041$). However, as these results were affected by the poor quality of 4G/LTE video, they may not hold once video quality improves.

5. DISCUSSION

Expected vs. Actual Video Quality: One of the key observations during our study was the gap between the expected quality of live video when streamed over 4G/LTE and the actual quality. All of our living room participants complained about the low video quality. Two of their expectations were at the core of the issue. First, they assumed that mobile video quality is the same on Wi-Fi

as it is on 4G/LTE. Second, they thought that live streaming video should have the same quality as videos recorded on a GoPro camera. Unfortunately, because wireless bandwidth is being consumed as quickly as it becomes available, the expectations gap will impact all near future mobile shared experiences.

Types of Contextual Information: Even if the gap in the expected and actual streaming video quality on wireless networks were to be reduced, our study illustrates that video is not always the best form of contextual information. In particular, the results show that having a history of periodic images from the activity and a map of the user at the activity were also useful contextual views.

Although one reason for usefulness of Images view was their high quality, it was also useful because it enabled people to review what happened. While it is possible to review video streamed during a videoconference, Junuzovic et al. [11] found that it is difficult to do so without affecting the live conversation. This did not seem to be an issue in our study when the participants reviewed images.

The Map view helped the inside participants get a sense of bearing and enabled them to provide directions to their partners. Interestingly, Procyk et al. [16] found that having live video of what is in front of a remote partner helped with navigation tasks. Therefore, it may be useful create a system that begins to support very course-grained navigation using our Map view and then switches to a Video view as the target comes into visual range.

Cognitive Costs: Although our study shows the benefits of including the contextual information in a mobile video chat system, these benefits do not come for free. An important issue is the cognitive load forced onto the participants with additional channels of communication. In our study, neither the inside nor the outside participants complained about any stress from the contextual information being included. In fact, the cognitive load of the outside participants seemed to have been reduced as they mentioned that the activity context made them worry less about what their partners could see. Thus, it may be that cognitive load increases only for the inside participant. Since, overall, the contextual information benefited the inside participant more, perhaps this is only fair. Such fair asymmetry in the additional cognitive load is one of many [20] that future systems can leverage.

Implementation Costs: Another cost associated with contextual information is the cost of building the software and hardware needed to support it. The system infrastructure will need to be redesigned to make tradeoffs between the performances of the various views for whatever bandwidth is available. The UX of the applications will need to be redesigned in order to present the additional information in useful ways. Finally, additional sensors (e.g., wearable cameras, etc.) may be needed, and the additional hardware costs will impact both system builders and users.

Social Costs: The additional hardware also leads to social costs. Some of our participants reported that having just a smartphone out for a mobile video chat made them feel self-conscious, let alone the wearable phone. Ideally, the additional hardware should not increase the social stigma beyond that of regular mobile video chat. Therefore, its design will need to be as inconspicuous as possible.

Privacy: Related to social costs is the issue of privacy. Although our participants did not report any privacy issues, people at the activities had some concerns. These concerns did not seem to be

about the people themselves being in the shot (although we believe that this can also be an issue). Instead, the concerns seemed to be copyright related and policy driven. To the artists who asked for their pictures not to be photographed, the issue was that they did not want others to digitally reproduce their art. Meanwhile, the security guard at the mall who asked for filming to stop did so because of mall policies about video recordings. Interestingly, even when our users explained that they were not actually recording anything, the artists and the security guard were still concerned. As mobile video chat becomes more pervasive, we can hope that such issues will become a thing of the past. A more proactive way to address the issue is to make it easy for people to tell when a device is recording video and when it is in a video call.

Limitations: There are several limitations with our work that are important to mention. For one, the participants used the system for only twenty minutes, so the novelty effect could have influenced our findings. Also, the quality of 4G/LTE video clearly impacted the findings. A useful question to ask is how the results would have differed with perfect video. In addition, the wearable prototype was cumbersome and awkward, which may have impacted our findings. Finally, our study had an element of artificiality because the remote participants did not join from their own living rooms.

6. CONCLUSION AND FUTURE WORK

In this paper, we explored the impact of adding contextual information to mobile shared experiences in which one person is at an activity and another joins remotely. Through a field study of events happening in the wild, we found that contextual information increased connectedness and the sense of presence for both parties. The study results also showed that contextual information is not “one size fit all.” The three types of information we studied, Map, Video, and Images, all had unique positive impacts on the shared activity. Map was used for orientation and to provide directions, Video for “do you see this” moments and to maintain a sense of liveliness, and Images was used for “did you see that” moments and to see greater detail. Together the different sources of context provided additional benefits. They led to smooth view negotiation, activity input from the participant joining remotely, and high levels of engagement.

The benefits of including contextual information are not without costs. Cognitive load, social awkwardness, and privacy concerns may all increase. Thus, a careful comparison of costs and benefits is needed before adding such information to video chat systems.

In the future, we plan to build and evaluate systems that incorporate shared inking and augmented reality into mobile shared experiences. We also plan to study how stabilization of live video impacts the experience even if the video quality is poor like on current 4G/LTE networks. We will also design and evaluate new wearable form factors for these experiences.

7. REFERENCES

- [1] Billinghurst, M., Kato, H., Real World Teleconferencing. *CHI EA 1999*.
- [2] Bos, N., Olson, J., Gergle, D., Olson, G., Wright, Z. Effects of four computer-mediated communications channels on trust development. *CHI 2002*.
- [3] Buxton, W. Mediaspace – meaningspace – meetingspace. In Harrison, S. (Ed) *Media Space 20+ Years of Mediated Life*, Springer, 2009.

- [4] Dourish, P., Bellotti, V., Mackay, W., and Ma, C-Y. Information and context: lessons from a study of two shared information systems. *COOCS 1993*.
- [5] Dourish, P. What we talk about when we talk about context? *PUC 2004*.
- [6] Fussell, S.R., Setlock, L.D., Kraut, R.E. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. *CHI 2003*.
- [7] Gaver, W., Sellen, A., Heath, C., Luff, P. One is not enough: multiple views in a media space. *CHI 1993*.
- [8] Inkpen, K., Taylor, B., Junuzovic, S., Tang, J.C., Venolia, G. Experiences2Go: Sharing kids' activities outside the home with remote family members. *CSCW 2013*.
- [9] iPad bridesmaid attends wedding via FaceTime. http://news.cnet.com/8301-17938_105-20097249-1/ipad-bridesmaid-attends-wedding-via-facetime/.
- [10] Jo, H., Hwang, S. Chili: Viewpoint Control and On-Video Drawing for Mobile Video Calls. *CHI EA 2013*.
- [11] Junuzovic, S., Inkpen, K., Hegde, R., Zhang, Z., Tang, J., Brooks, C. What did I miss? In-meeting review using multimodal accelerated instant replay (AIR) conferencing. *CHI 2011*.
- [12] Kuzuoka, H., Kosuge, T., Tanaka, M. GestureCam: Video communication system for sympathetic remote collaboration. *CSCW 1994*.
- [13] O'Hara, K., Black, A., Lipson, M. Everyday Practices with Mobile Video Telephony. *CHI 2006*.
- [14] Okada, K., Maeda, F., Ichikawaa, Y., Matsushita, Y. Multiparty videoconferencing at virtual social distance: MAJIC design. *CSCW 1994*.
- [15] Olson, J.S., Olson, G.M., Meader, D.K. What mix of video and audio is useful for small groups doing remote real-time design work? *CHI 1995*.
- [16] Procyk, J., Neustaedter, C., Pang, C., Tang, A., and Judge, T.K. Exploring video streaming in public settings: shared geocaching over distance using mobile video chat. *CHI 2014*.
- [17] Sellen, A., Buxton, B., Arnott, J. Using spatial cues to improve videoconferencing. *CHI 1992*.
- [18] Soldier watches baby born over Skype. http://www.youtube.com/watch?v=Z_w8RzYVIM4.
- [19] Stafford, A., Piekarski, W., Thomas, B. Implementation of God-like interaction techniques for supporting collaboration between outdoorAR and indoor tabletop users. *ISMAR 2006*.
- [20] Volda, A., Volda, S., Greenberg, S., and He, H. A. Asymmetry in media spaces. *CSCW 2008*.