Watching 360° Videos Together

Anthony Tang **Omid Fakourfar**
University of Calgary
Calgary, Canada
{tonyt, omid.fakourfar}@ucalgary.ca

ABSTRACT
360° videos are made using omnidirectional cameras that capture a sphere around the camera. Viewers get an immersive experience by freely changing their field of view around the sphere. The problem is that current interfaces are designed for a single user, and we do not know what challenges groups of people will have when viewing these videos together. We report on the findings of a study where 16 pairs of participants watched 360° videos together in a "guided tour" scenario. Our findings indicate that while participants enjoyed the ability to view the scene independently, this caused challenges establishing joint references, leading to breakdowns in conversation. We conclude by discussing how gaze awareness widgets and gesturing mechanisms may support smoother collaborative interaction around collaborative viewing of 360° videos.

Author Keywords
360° videos; omnidirectional videos; shared experience

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

INTRODUCTION
People watch fun and exciting videos with others using smartphones and tablets because it can lead to interesting conversations [14,16]. Yet, these conversations rely on sharing a common frame of reference—knowing what others have seen in these videos grounds the conversation. While this is straightforward with traditional 2D media, this becomes more challenging with so-called "360° videos," which provide an immersive omnidirectional view: viewers can freely look around while these videos play out.

The common intention of 360° videos is to provide the viewer with an immersive experience, where they are encouraged to explore the scene themselves. Popular 360° videos feature exotic destinations (e.g. Mecca), extreme sports (e.g. mountain climbing, skydiving), wildlife tourism (e.g. dolphins, sharks) and music experiences (e.g. concerts), where the action happens around the viewer from all sides. 360° video players on mobiles use the gyroscope sensor, allowing people to turn the display around to view different parts of the scene. This freedom to control the view makes 360° videos exciting; however, such an interface is unlikely to work well in multi-person scenarios.

Our interest is understanding how people experience these 360° videos when physically collocated. We designed and conducted a study of 16 pairs watching 360° videos side-by-side. Together, they watched a moving video tour of our university campus, where the capture camera was ridden (on a bike) around near major landmarks around campus. In each pair, one participant provided an accompanying oral "tour" of campus to the other participant who was new to campus. We sought to understand how pairs would use virtual (i.e. in-video) and physical cues and frames of reference to orient one another during viewing.

Our findings suggest that participants enjoy the freedom of an independent view (like [19]). But, this independence creates challenges for properly orienting and understanding spatial references when watching 360° videos with others. Participants used visual cues such as the movement of the video and distinctive landmarks to overcome the challenge of disjoint perspective, and that they took advantage of being able to physically see one another to overcome these problems. Based on their experiences, participants suggested new features for such joint viewing experiences: gaze awareness cues (e.g. [15,11,2,3]), the ability to control playback speed, to smoothly move between disjoint views and shared (i.e. locked) views, and finally to tag moments in the video that might be of interest to later viewers.

This paper makes two contributions: first, we contribute the first study of collaborative 360° video viewing; second, based on the findings of the study, we contribute a set of design implications for 360° video players.

BACKGROUND
Watching Videos from Mobiles with Others. Increasingly, our social lives are mediated by and supported through mobile device interaction [20,17]. Porcheron et al. [18] describe how mobile phone use is interleaved into everyday collocated interaction with others—for instance, to support conversation (e.g. to get more information about a topic through search). Several authors have documented how people share video with one another to enjoy a shared experience [16], or to support conversation [14].

CHI 2017, May 06 - 11, 2017, Denver, CO, USA

ACM 978-1-4503-4655-9/17/05…$15.00
DOI: http://dx.doi.org/10.1145/3025453.3025519

Sharing videos scaffolds conversation and interaction: watching videos together makes them a shared, common frame of reference, knowledge of which allows people to engage in a shared social experience. Based on this, several design explorations put media sharing on mobile devices at the centre of a collocated social experiences (e.g. [1,4,22,13]). Yet, how to do this sharing effectively is unclear with 360° videos, where the "action" may not be in the centre or default field of view.

360° Videos. 360° videos put the viewer in the centre of a video sphere. As the video plays out in time dimension (*t*), the capture location of the original camera can move (*x,y,z*), and the viewer controls the pitch and yaw of their view. Viewing interfaces for these videos use two mechanisms for controlling pitch and yaw: *drag* and *gyroscope*. The drag control (activated using a mouse on a PC, or touch on tablets and phones) moves the spherical sector. The gyroscopic interface (available on tablets, handhelds and head-mounted displays) moves this spherical sector in relation to the movement of the device. On tablets and handhelds, both interfaces are available simultaneously.

Joint References in Collaborative Virtual Environments (CVEs). A related domain is prior work that explores how collaborators maintain awareness of one another's focus and intention in collaborative virtual environments (i.e. VR environments inhabited with others). Early work first identified this awareness problem within CVEs [8,9], and recent efforts have explored how to address these awareness through pointing gestures and expanded field-of-view [21]. As pointed out by Wong & Gutwin [21], however, little has been done to address this problem in the many CVEs that people use every day.

STUDY

We designed an observational lab study where pairs completed a "tour of campus" scenario, where one participant gave a verbal tour to accompany a time-synchronized 360° video of our university campus. We were interested in how pairs would experience immersive 360° videos together: specifically, how they would discuss things in the video, the kinds of challenges they would encounter, and how they would overcome these problems. The study task was chosen to ensure participants would consistently have something novel to see and discuss. By recruiting pairs where one participant was already familiar with campus, we ensured that at least one participant would have meaningful things to say about the video—i.e. rather than participants having only surface-level discussions.

Design. We used a two condition (interaction technique: gyroscope *vs*. touch) within-subjects design for the study. We expected that each would have different kinds of benefits to collaborative video viewing: that gyro would provide cues as to where someone was looking, but that drag would make it easier to see what a partner was actually looking at. Interaction technique presentation order was counter-balanced across groups.

Participants. We recruited 16 pairs of participants (32 total participants; 17 females; 16-30 age range, median: 23.5), where pairs knew each other beforehand, and consisted of one participant who knew campus well (all were students or ex-students) and the other not familiar with campus. 24 participants had viewed 360° videos in the past, 13 had viewed them from a phone or tablet, and 6 had experiences with head-mounted displays to view 360° videos.

Materials. Participants used two iPad Air 2 tablets (9.4"x6.6" physical dimensions; 2048x1536 pixel resolution) to display the 360° videos. We created a 14-minute 360° video tour of our campus, where an author rode a bicycle along the main outdoor thoroughfare of the university. This video was played back without audio.

Method. The participant familiar with campus was the "tour guide", while the other participant was asked to pretend s/he was a "new student" on campus. Prior to beginning the tour, the tour guide was given a map of where the video tour of campus would go, and asked to provide a tour to the new student, discussing interesting landmarks, personal anecdotes about buildings, daily routes, and so forth.

Pairs were seated on swivel chairs, and each given an iPad with the video. They completed the first half of the tour using one interface, and the second half of the tour using the second interface. The videos were started simultaneously, and oriented in the same direction to begin. Once the task was complete, participants completed a questionnaire about their experience and a verbal debrief.

Data. We collected a video recording of each session.

Findings

Viewing Preferences. Participants' preferences between the interfaces were evenly split (50% gyroscope; 50% drag). Participants' main concern over the gyroscopic interface was that it caused some dizziness due to the movement, and that it was somewhat inconvenient/unnatural. In contrast, several participants suggested that the drag interface was more comfortable, as it allowed them to explore the entirety of the video without needing to move around physically. Unless specifically noted below, these preferences did not seem to factor into people's experiences in the study.

Every participant watched the video independently from their own devices. The "new students" (who were not familiar with campus) particularly enjoyed the flexibility of being able to look around the environment independently of the tour guide. Thus, while the "tour guides" tended to stay focused on the path that the video was headed, where the bike ride revealed new parts of the environment to discuss, the new students tended to look around, exploring the scene independently. The associated video figure illustrates how much new students looked around.

We observed numerous instances when the tour guide would be speaking about a landmark or object of interest, but the new student was looking at something else in the

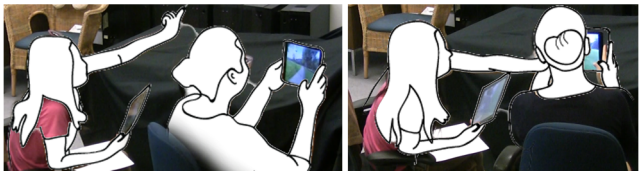

Figure 1. The tour guide points to a landmark (left), and then on the new student's screen (right).

Figure 2. The tour guide points with his foot to a landmark (left), and the new student re-orients herself (right).

Figure 3. The new student looks at the tour guide's view.

video. This freedom is much like how a real tour occurs, where rather than focusing strictly on what is being described by the guide, a tourist may only be partially listening, and looking around independently. Most poignant here was how the new students would initiate conversation about a landmark or building that was not necessarily in the direction that the video was headed. This is an important departure from a typical video tour, where the tour path prescribes exactly what is to be seen; here, the 360° view gives the new student not only the flexibility to "look around"—it also gives the new student agency to direct the shared experience.

Joint Reference. Tour guide and new student sometimes had momentary difficulty establishing joint references—i.e. ensuring that both understood what was being discussed (e.g. the tour guide talks about a landmark, or the new student asks about a landmark). This was not a problem when the landmark was particularly distinctive, though with subtle landmarks, or landmarks that were partially (or fully obstructed), this became problematic. Describing such landmarks would require describing its location *in relation* to another landmark, or with additional specificity: "[A challenge is] coordinating direction and pointing out specific places - the other person is also looking at the screen and sometimes will not be looking at you so you really have to be specific in the way you ask the question (ie: 'What is this blue building on the left,' rather than 'What's that')." [P1] This additional verbal coordination is needed because participants have independent views: "[With two iPads, we have] to describe where everything is. There was a small amount of confusion between which landmarks were being described because [we] were looking at different things." [P8]

Knowing where the other participant was looking helped support joint references. With the gyroscope interface, it was evident where someone was looking based on their body posture. This is illustrated well in the video figure, where a pair of participants engaged in "synchronized viewing", where they mimicked one another's body position. Generally, this gaze information was not available, but seeing it would be useful: "[I wish] I could have seen the tour guide's POV or some kind of indication of where the tour guide was looking at (ie: like a symbol on the screen)." [P1]

Pointing and Glancing. We observed participants pointing and gesturing with varying frames of reference. Figure 1 illustrates a sequence where a tour guide points to an off-screen landmark using a twisting hand gesture (left). This gesture uses the new student's view (that she can see): she expects the new student to turn her device to see the landmark is. Once the new student oriented properly to the landmark, she then points by tapping on the screen (right). In contrast, Figure 2 (left) shows a tour guide pointing using his foot to a landmark that is relative to his own point of view, and ultimately draws the new student's attention.

Figure 3 illustrates a common phenomenon where participants would glance at his/her partner's device. Sometimes, this glance was to understand the partner device's orientation to either correct oneself or one's partner (e.g. "No, turn like this so you can see over here…"). Other times, it was to understand what the partner was looking at. The physicality of the gyroscopic interface provided a shortcut to this information, as body orientation gave this cue without needing to look at the screen. In contrast, with the drag interface, participants could not intuit what was being looked at by glancing at the orientation of his/her partner's device. "[In the gyroscope condition, it] was easier to follow because I could see her body position and follow her, but in the [drag condition, it] was harder ... when I got lost. [P11]" Accordingly, we observed participants glancing at partner screens more often in the drag interface condition (Figure 3).

Not all participants arrived at these workarounds, which simply resulted in not being able to see the same things: "I wanted to point at things, but I couldn't exactly do that, so I basically just kind of had to let my friend find it for herself, or swipe [sic] on her screen so that she saw it." [P2]

Missing the Reference. We also saw numerous instances where new students missed seeing a landmark *because* they were looking at something else. In these instances, the landmark was no longer in view—for instance, the cyclist recording the video went around a building (e.g. "Oh, it's too late, you missed it."). This was not consequential in our task, but participants commented on this in relation to the

inability to control playback speed: "There were some points where I probably would've stopped and talked about some of the places, also I wanted the route to be different since I would want to go by more of the places I knew. [P3]"; "Sometimes if we went too fast, we would miss some of the buildings and then we would just have to move on to the next part of the tour. [P9]"

Direction. One strategy that most teams developed (15/16 pairs) was to use the direction of the cyclist in the source video as the "canonical" forward direction. Thus, the cyclist was always considered to be "straight ahead" or "forward", and directions could be described in relation to the implicit "movement" in the video. Directions would be articulated from the cyclist's perspective. For example: "If you see to the right, that building, with all the smaller buildings, that's the engineering building. Directly ahead [...], that's the new residence [building]. [P10; 0:10]" Here, we see that the video's movement provides not only a common anchor that participants used to resolve situations where each was looking in different directions or objects. Similarly, participants would refer to "left" or "right" in relation to this inherent movement in the video.

Using One Device. Although participants were free to use a single device, no groups chose to do this. Tour guides preferred sharing a single device (10/14), whereas most new students preferred the freedom of an independent view (8/14): "I really enjoyed my own ability to move the camera around. In real life, that's essentially what I would be doing - looking around as a tour guide gives the tour. [P21]"

DISCUSSION & DESIGN IMPLICATIONS

Participants regularly pointed to the challenge of building a shared understanding of what was being looked at and discussed. This challenge arises because people are given the freedom (and encouraged, in many ways) to look around in a 360° video. This presents challenges for people when they try to view a 360° video together. We synthesize the ideas that arise from our study, and suggest design possibilities for addressing these challenges.

Gaze Awareness and Gestures. The principal communicative problem in viewing 360° videos with others is understanding what they are looking at in relation to conversation. This challenge is somewhat reminiscent of the challenges people experience when using mobile phones to engage in video chat, as the framing of targets/objects of interest becomes the principle problem [5,12]. Providing gaze awareness widgets is one way of addressing this challenge (e.g. [6,3,15,11]), though care must be taken in how to visualize gaze when there are a large number of viewers without overwhelming the view.

People need the ability to point, gesture or otherwise reference objects in the video. We observed people simply reaching over to point at others' screens; however, this does not scale if there are multiple viewers. Furthermore, touching the screen is overloaded with changing the view on the screen, which was not always the intent of pointing.

Of interest is how participants gravitated towards verbal cues that made use of "forward" to mean the "direction" of the cyclist who captured the video. This suggests that such verbal shortcuts are useful and desirable, and that we can either provide these kinds of cues (e.g. a compass to indicate which direction one is pointed), or introduce subtle movement into such a captured video to allow participants to orient one another.

Head-Mounted & Tracked Displays. A common way for people to view 360° videos is to use head-mounted and tracked displays. While appropriate perhaps for a single user experience, it is unclear how this translates in a multi-person scenario. As we saw in the study, people observe other's physical posture to infer the direction (when they are using the gyroscope interface). This strategy would be obviated with head-mounted displays that obscure peripheral vision.

Temporal Exploration and Annotations. Just as people can deviate in their view of the space, it may make sense to allow a limited amount of deviation in time, but then be able to "snap back" to where they were looking together before. Space and time are connected in viewing these kinds of videos, and while this freedom might be desirable, we need to give people additional affordances that they do not have with normal videos. Several participants suggested providing the ability to annotate the videos with markers (thereby allowing others to view the landmark later, even if the landmark is out of view).

Joint Views. As we scale up to larger groups or heterogeneous groups (e.g. educational contexts where a teacher guides a group), we need to consider how to give people the freedom to explore, and then to come back to the views of others. This means being able to understand where others are (in terms of orientation or time), to snap to it, and/or to observe where "the majority" are looking etc. This depends a lot on the specific context (e.g. educational vs. casual), and this warrants a careful consideration of the design space of viewing freedom (e.g. being able to orient freely vs. with certain constraints; being able to slow down, pause or speed up playback on an individual basis vs. being tied to a common playback timestamp and speed). While people enjoy the freedom of having their own view [19], this may not necessarily be the best approach for group work [7].

CONCLUSION

360° videos have opened up an entirely new way for people to immerse themselves in new places and experiences. Yet, by studying how pairs view 360° videos in an observational lab study, we have demonstrated that current interfaces do provide desirable experience for shared viewing. Because a core aspect of the 360° video immersion is the freedom to explore the scene on one's own, we need to provide effective mechanisms for people to be aware of one another's perspectives, as this impairs effective communication in these experiences.

REFERENCES

1. Karolina Buchner, Roman Lissermann, and Lars Erik Holmquist. 2014. Interaction techniques for co-located collaborative TV. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '14). ACM, New York, NY, USA, 1819-1824. DOI=http://dx.doi.org/10.1145/2559206.2581257
2. 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
3. Jeff Dyck and Carl Gutwin. 2002. Groupspace: a 3D workspace supporting user awareness. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '02). ACM, New York, NY, USA, 502-503. DOI=http://dx.doi.org/10.1145/506443.506450
4. Hasan Shahid Ferdous, Bernd Ploderer, Hilary Davis, Frank Vetere, Kenton O'Hara, Geremy Farr-Wharton, and Rob Comber. 2016. TableTalk: integrating personal devices and content for commensal experiences at the family dinner table. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp '16). ACM, New York, NY, USA, 132-143. DOI: http://dx.doi.org/10.1145/2971648.2971715
5. 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
6. Carl Gutwin, and Saul Greenberg. 2002. A Descriptive Framework of Workspace Awareness for Real-Time Groupware. *Computer Supported Cooperative Work 11*(3-4): 411-446.
7. Carl Gutwin and Saul Greenberg. 1998. Design for individuals, design for groups: tradeoffs between power and workspace awareness. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work* (CSCW '98). ACM, New York, NY, USA, 207-216. DOI=http://dx.doi.org/10.1145/289444.289495
8. Jon Hindmarsh, Mike Fraser, Christian Heath, Steve Benford, and Chris Greenhalgh. 1998. Fragmented interaction: establishing mutual orientation in virtual environments. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work* (CSCW '98). ACM, New York, NY, USA, 217-226. DOI=http://dx.doi.org/10.1145/289444.289496
9. Jon Hindmarsh, Mike Fraser, Christian Heath, Steve Benford, and Chris Greenhalgh. 2000. Object-focused interaction in collaborative virtual environments. *ACM Trans. Comput.-Hum. Interact.* 7, 4 (December 2000), 477-509. DOI=http://dx.doi.org/10.1145/365058.365088
10. 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
11. 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
12. 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
13. Andrés Lucero, Matt Jones, Tero Jokela, and Simon Robinson. 2013. Mobile collocated interactions: taking an offline break together. *interactions* 20, 2 (March 2013), 26-32. DOI=http://dx.doi.org/10.1145/2427076.2427083
14. Koji Miyauchi, Taro Sugahara, and Hiromi Oda. 2009. Relax or study? A qualitative user study on the usage of live mobile TV and mobile video. *Comput. Entertain.* 7, 3, Article 43 (September 2009), 20 pages. DOI=http://dx.doi.org/10.1145/1594943.1594955
15. 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.
16. Kenton O'Hara, April Slayden Mitchell, and Alex Vorbau. 2007. Consuming video on mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07). ACM, New York, NY, USA, 857-866. DOI=http://dx.doi.org/10.1145/1240624.1240754
17. Erick Oduor, Carman Neustaedter, William Odom, Anthony Tang, Niala Moallem, Melanie Tory, and Pourang Irani. 2016. The Frustrations and Benefits of

Mobile Device Usage in the Home when Co-Present with Family Members. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (DIS '16). ACM, New York, NY, USA, 1315-1327. DOI: http://dx.doi.org/10.1145/2901790.2901809

18. Martin Porcheron, Joel E. Fischer, and Sarah Sharples. 2016. Using Mobile Phones in Pub Talk. In*Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16). ACM, New York, NY, USA, 1649-1661. DOI: http://dx.doi.org/10.1145/2818048.2820014
19. 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
20. Sherry Turkle. 2012. *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books, New York.
21. Nelson Wong and Carl Gutwin. 2014. Support for deictic pointing in CVEs: still fragmented after all these years'. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (CSCW '14). ACM, New York, NY, USA, 1377-1387. DOI=http://dx.doi.org/10.1145/2531602.2531691
22. Nicola Yuill, Yvonne Rogers, and Jochen Rick. 2013. Pass the iPad: collaborative creating and sharing in family groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13). ACM, New York, NY, USA, 941-950. DOI: http://dx.doi.org/10.1145/2470654.2466120