Tourgether360: Exploring 360° Tour Videos with Others

Contemporary 360° video players do not provide ways to let people explore the videos together. Tourgether360 addresses this problem for 360° tour videos using a pseudo-spatial navigation technique that provides both an overhead "context" view of the environment as a minimap, as well as a shared pseudo-3D environment for exploring the video. Collaborators appear as avatars along a track depending on their position in the video timeline and can point and synchronize their playback. In this work, we describe the intellectual precedents for this work, our design goals, and our implementation approach of Tourgether360. Finally, we discuss future work based on this prototype.

CCS CONCEPTS • Human-centered computing • Collaborative and social computing • Collaborative and social computing systems and tools

Figure 1: Tourgether360 allows collaborators, represented by avatars, to tour together "inside" a 360-degree video.

1 INTRODUCTION

360° tour videos are an increasingly popular way of exploring remote destinations and environments. Such videos are typically shot using an omnidirectional camera mounted atop a tripod as the cameraperson continuously moves through an environment (e.g., by walking or driving). The videos provide viewers with the ability to freely look around, independent of the direction that the cameraperson was moving. Because of this freedom, they provide users with a rich sense of immersion—particularly when coupled with head mounted displays (e.g. [4,26]). 360 tour videos for a wide range of environments, from urban landscapes, museums, to college campuses, are abundant on online video platforms, such as YouTube. Some of these videos have already attracted over one million views. *Collaborative* viewing is also becoming increasingly important. Collocated or remote friends may want to watch 360° videos to experience immersive and social entertainment together, or such videos may be used in the educational context, with a class of students going on a virtual museum tour or virtual trip to cultural locations.

The problem is that current 360° interfaces do not provide effective support for collaborative navigation and exploration of 360° videos (e.g. [25]). With only a handful of exceptions (e.g.[16,22]) 360° video players are intended for single-person use; in addition to normal video playback controls, such video players need to provide a special, separate means for controlling the view orientation. On a desktop, orientation is controlled by grabbing the scene and moving it with a mouse; on tablets, this is augmented through gyroscopic sensors, and on a head-mounted display, orientation can be controlled by turning or tilting one's head. Yet, there is little to no support for collaborative viewing of these immersive videos where two or more users watch the same video simultaneously and explore interesting parts of the video together.

We propose a pseudo-spatial navigation metaphor for collaborative exploration of 360 videos inspired by a focus+context approach [23] that allows both high-level understanding of video content and detailed focus on specific parts of interest while being aware of other user's position, actions, and intentions. We realize this approach in a prototype system called Tourgether360, which allows several collaborators to explore a 360 video together. Figure 1 demonstrates the system with two users watching the video. With Tourgether360, the video tour context is visualized as a path on a 2D map of the environment. This eases the coordination between users when finding points in the video with spatial locations visited in the video. When multiple users view the same video, their viewing position (time, space, and orientation) is embodied by an avatar directly in the video, and they can position and mark different points of interest for one another. This allows collaborators to experience the video as if they were embodied together in the video.

2 RELATED WORK

Navigation of 360 videos. Existing work on interaction with 360° videos has focused on supporting *orientation navigation* (directing one's view in the video), and *temporal navigation* (controlling playback or directing one to interesting moments in the video). Prior work proposed supporting orientation navigation within 360 degree videos by automating this through computational measures (e.g. [9,18]), while other researchers have designed mechanisms to signal to viewers where the view should be oriented [11,18]. Mäkelä et al. [12] show that such mechanisms can improve the experience, even if they are subtly distracting. Several researchers have also proposed new techniques for temporal navigation of videos. For instance, Petry & Huber [19] explore multimodal gestures for playback controls for 360° videos within a head-mounted display viewing context. Similarly, Ruiz et al. [21] apply this approach within a multi-person viewing context. Neng and Champbell present a 360° video player that augments the traditional timeline with cues representing points of interest, and regular thumbnails for some frames in the 360 video [14]. For scrubbing through videos, VRmiere provides a "Little Planet" navigation technique for these videos, which can provide some spatial awareness [15]. Li et al. explores a unique temporal navigation approach that blurs the boundary between navigating time and space using a visual tapestry constructed of "slits" of the video [10]. Our approach centers on the insight that navigation through video may be better supported through a semantic, contextual understanding of the content (i.e., what is in the video) rather than time leading to navigation to a particular *place* in addition to a *particular time* in the video. In the context of 360° videos, we explored a visual map-based approach that provides this context view, combined with the normal view of the scene, which is the focus view.

Collaboration in Virtual Environments. When people watch 360° videos together, several communication emerge [12,20,25]: users do not experience social presence of other people during collaborative view of the 360 video [20], and they experience challenges coordinating and synchronizing between the users [1,2,5,6]. Many of the problems outlined by researchers studying collaborative viewing of 360° videos [20,24,25] are reminiscent of early CSCW research focused on Collaborative Virtual Environments (e.g. [26,27]), where designers were forced to contend with basic

awareness issues: (1) Where are my collaborators? (2) What can my collaborators see? (3) What are they looking at? (4) How can I draw someone's attention to what I am talking about? While we can provide additional cues for awareness in 3D workspaces [7,8], most design approaches we see in collaborative video have not pushed the boundaries of the embodiments first envisioned in the mid-1990s. Particularly for experiences where viewers can watch simultaneously, there is a strong need for systems to provide an awareness of where others are viewing, and potentially gesture support to support communication and coordination [27,28].

3 TOURGETHER360: USER EXPERIENCE DESIGN

Building on prior work, we identified four major design goals for our system:

- *DG1: Support semantic navigation of the video space:* users should be able to navigate based on landmarks and elements in the video rather than only using a timeline.
- *DG2: Support awareness of collaborators' perspectives and temporal position*: users should know what others are looking at, and where they are [13].
- *DG3: Enable smooth engagement and disengagement with collaborators' perspectives*: collaborators should be able to smoothly move between independent and synchronized modes of interaction [24,25].
- *DG4: Support deictic reference with a semantic understanding of the environment:* collaborators should be able to point and refer to things in the video and environment.

Tourgether360 reconceptualizes viewing a 360° video as a shared virtual 3-dimensional space, and implements a number of unique features that increase users spatial understanding of the environment and enable *allocentric* navigation. We support each of the four design goals through separate features built into the system.

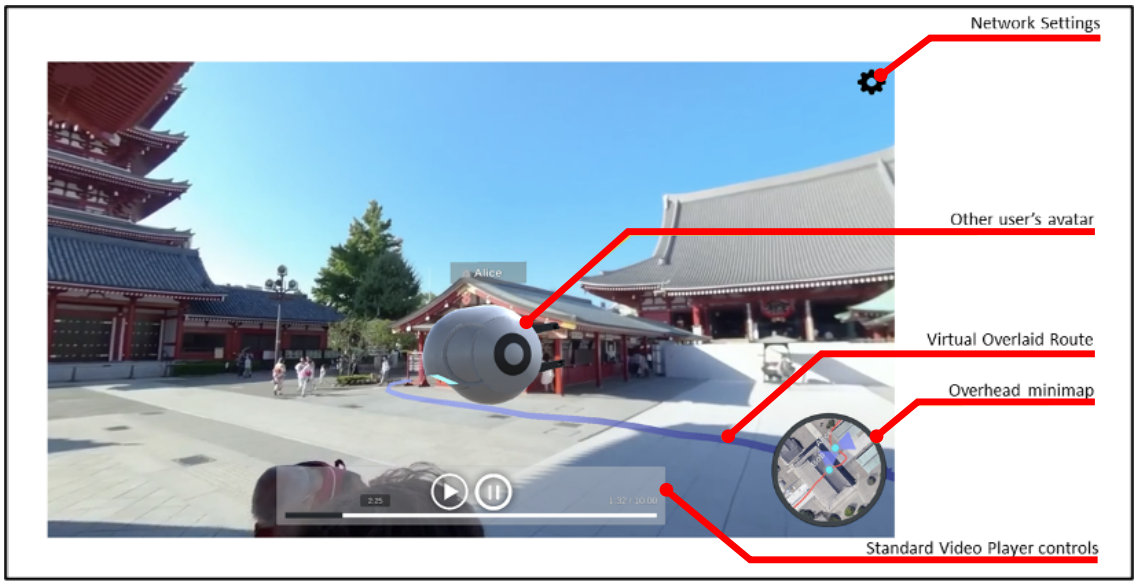

Figure 2: Full Interface of Tourgether360 from the first-person perspective of a user (Bob), looking at the video and seeing the avatar of the other user (Alice).

Pseudo-spatial navigation. Inspired by prior work, such as research by Noronha and colelagues [17], and the UI of the 3D video games, navigation affordances are supported by an overhead schematic interactive minimap of the 360°

video tour environment (Figure 2). The minimap provides a virtual path that represents the route on which the tour takes place, where the user's position in the video is represented by a blue dot in space. Each collaborator's gaze direction is represented on the minimap by the conical light beam. The minimap allows users to navigate through the video using spatial landmarks visible from the minimap (DG1) by clicking and dragging the mouse along the virtual path on the minimap (Figure 3). The metaphor of the 3D space is also realized in the main video view. Here, the path of the video is represented by a virtual overlaid route seen in Figure 4, showing the forward and backwards route of the tour from the first-person perspective of the user in the environment, and helps enhance users' spatial understanding.

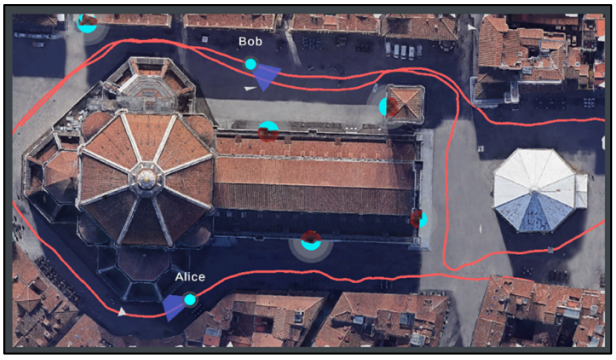

Figure 3: Minimap of the Florence Cathedral environment shown in one of the 360 videos we used in the study. Here, the path taken in the 360 video is represented by the red line. Alice and Bob are at different parts of the video, where their viewing orientations are represented by a cone. Finally, spheres represent marked "points of interest" that were placed by the collaborators.

Figure 4: Representation of a virtual route overlayed on top of the video. Taken from the user's point of view, the path (highlighted blue) illustrates how the video tour will take the user around the cathedral. Because Tourgether360 understands the architecture of the space represented in the video, the line path is cropped at the edge of the cathedral.

Collaborator Embodiment. As illustrated in Figure 5, each collaborator is embodied by an avatar in the 3D video tour scene. The avatar is a flying spherical robot with four antennas indicating the "face" part of the robot. This "face" is synchronized with the user's camera's forward vector to indicate the gaze orientation. This embodiment approach in

the 3D scene provides awareness of others' temporal position and view orientation (DG2). When collaborators are watching the video together, the apparent distance between two avatars in the 360° scene is equivalent to the temporal distance between two collaborators. To maintain the illusion that collaborators are navigating a 3D environment rather than a video (DG1), collaborators' avatars are presented using a silhouette representation if they would normally be occluded by the architecture of the space (Figure 6).

Collaborators can use the embodiments to engage and disengage smoothly with each other through view synchronization. A user can assume a spectator role by double clicking on another collaborator's avatar, which synchronizes both collaborators so that both playback and view orientation are synchronized to the guide. Users can regain control by simply moving their orientation or explicitly navigating once again.

Figure 5: Representation of the user avatar overlayed on the top of the played video on the virtual route line

Figure 6: A collaborator's avatar is rendered as a silhouette if they would normally be occluded by the environment (here, by the wall of the building).

Pseudo-spatial annotation and allocentric navigation. As illustrated in Figure 7, users can annotate and navigate the environment with artificially created landmarks – pseudo-spatial markers placed by directly in the environment of 360° video tour. The markers are visible to everyone and, from the user perspective, diminish or increase in size depending on the closeness of the user to them. This allows users to communicate about the environment via deictic reference (DG4) and reinforces the notion that the annotations are about the semantic space (DG1). Users can instantiate

markers by double clicking at the point of interest in the environment where they want to place them. Clicking on a marker will teleport the users to the point of interest and time in the video when this marker was instantiated. Users also can delete the existing markers by double clicking on them.

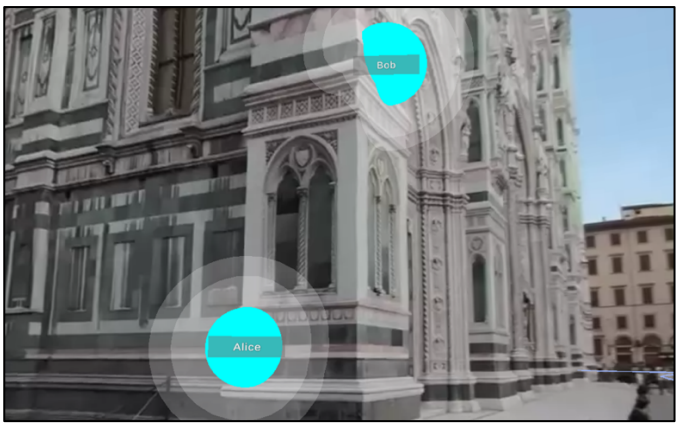

Figure 7: Representation of pseudo-spatial markers placed on the walls of the Florence Cathedral by two users (Alice and Bob).

4 SYSTEM ARCHITECTURE

We created Tourgether360 using the Unity game engine environment. The multi-user functionality was supported by Unity's Multiplayer Networking library (MLAPI). We employ vision techniques to understand geometric features of the space, the route, and then place this within a 3D environment.

Route Extraction from 360° Tour Video. The virtual route was extracted using Simultaneous Localization and Mapping technique (SLAM), specifically through its open-source implementation in the package ORB-SLAM [13] run on monocular 360° videos using the omnidirectional camera model on a per-frame basis. The algorithm generated the camera's spatial coordinates (x, y & z) and rotational orientation (quaternions) relative to a central coordinate system, and the virtual route was constructed by sampling from these spatial coordinates at an interval of 0.5 seconds and joining the resulting points.

Invisible 3D layer Overlaid on Top of The Video. In addition to the virtual route, to support all our pseudo-spatial design affordances, we overlaid the video environments played in Tourgether360 with the invisible 3D models of the environments depicted in the video. We extracted the models of the locations from Google Maps geo-information system and aligned the models with the 360° video of the location through a manual calibration process, in which the models were scaled, positioned and rotated to reflect their size, position and angle in the actual video. We used RenderDoc[1] to extract Google Map's 3D buffer cache of the location, removing the unnecessary parts of the 3D models that were not visible in the video using Blender[2].

The 360° video is rendered on a sphere around each user, which engulfs the 3D models, the virtual route, and the avatars of other users. On each client, only the sphere corresponding to the local player is rendered. Each sphere along with the parented user's avatar at their center, move independently along the fixed route in the virtual unity space (Figure 8). For aligning the video with the stationary invisible 3D model while the video is playing, a rotation correction is subsequently

[1] https://renderdoc.org/
[2] https://www.blender.org/

applied to the video sphere. As in the case of computation of virtual route, the exact values of camera rotation for the rotation correction of the video sphere are computed through the SLAM algorithm, which reports these values in terms of quaternions for each frame in the video.

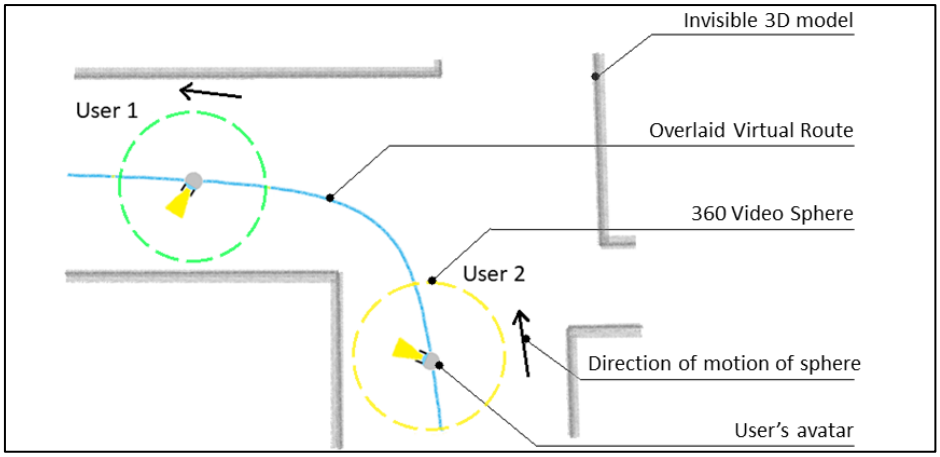

Figure 8: Overhead diagram of two users (and their video spheres) moving along the virtual route in the environment. Here for user 1 only the green sphere and the avatar of user 2 is rendered and vice versa. Note: The spheres are made small for visualization purposes, in the actual system they engulf all of the 3D model and the route.

This combination of the video and 3D models provides dynamic occlusion of users' avatars, and the virtual route. For instance, if the other user's avatar moves behind a wall, the avatar changes to a silhouette-like appearance. The implementation of this function was via a custom shader, which although is transparent (to allow for the unobstructed view of the video), tints the shaders of specific objects like avatars to a red fresnel (silhouette) shader when occluded (e.g., Figure 9). Similarly, the path is occluded when it is obstructed by any solid spatial entity.

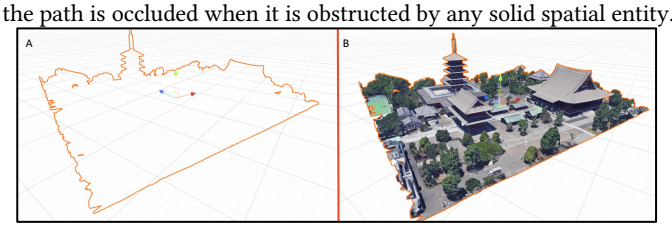

Figure 9: 3D model of Asakusa Shrine Complex that we used as a virtual overlay for the corresponding video used in the study. (a) The model with a custom transparent shader that is used as a direct video overlay (traced for clarity), (b) The textured model used as an overlay for the minimap.

Pseudo-Spatial Markers. We implemented the ability to mark the spots in the environment by using pseudo-spatial markers, represented by the flashing sphere model. The pseudo-spatial positioning of the markers in space was implemented via a ray-casting technique, where a ray from the mouse cursor points on the screen determined the position of the marker in the location where this ray hit the 3D model of the environment. When markers are created, the markers are instantiated and positioned directly on the 3D model. Users perceive the elements as if they are synchronized with the actual video, appearing to stick to the place where they were instantiated.

5 FUTURE WORK AND CONCLUSIONS

The current design of 360 video players makes it hard to comfortably enjoy and navigate such videos with others—particularly when the goal is to communicate, coordinate and socializing with other people. Tourgether360 extends prior work on providing spatial means of understanding and watching 360° videos [17]. It supports spatial navigation on an architectural minimap and simulates a co-habited space with other collaborators. Yet, our work to this point has been to design and build a prototype that we (as designers) are comfortable to use and play with. There are a number of important future directions, including both user experience studies and technical work, that we are currently exploring in follow-up work.

Evaluation of Tourgether360 as a Social Experience. Our aim in designing and building Tourgether360 was to make 360 tour video viewing a social experience. We will conduct a study that explicitly explores how the awareness and communication affordances support this. In particular, our expectation is that people will find the use of the avatars, the minimaps and the landmarks useful; however, this may depend a lot on the task. It seems likely that the availability of these affordances, since they are not normally available, will create entirely new patterns of behaviour and conversation which we have yet to observe. To what extent, for instance, do the avatars support this social experience versus the icons on minimap, or how does the landmarking functionality support interaction and discussion about the environment and video when collaborators may be viewing a landmark at different points in time? While our intention is to smooth interaction for such experiences, it is unclear whether Tourgether360 does this better than a conventional interface (e.g. sitting side-by-side in front of a computer).

We plan to objectively assess participants' performance and experience in a formal user evaluation. We will incorporate objective quantitative measurements of system usability and participants performance, as well as use formal techniques for evaluation of user experience, particularly in the aspects of embodiment, co-presence, and collaboration. Further, we will compare our system with a control condition, where the participants will collaborate over 360 videos using conventional video-based navigation and awareness techniques. In addition, we will create several experimental conditions with different combinations of system features, to assess their effectiveness and come up with recommendations for the design of future systems.

Extensions for Head-Mounted Displays. This work explores 360° video viewing from the perspective of desktop computer use, where providing a minimap in the periphery of the display is an accepted practice (from video games). Yet, it is unclear how to modify this approach for head-mounted displays. We are actively pursuing how to provide awareness of others' activities in 3D workspace when collaborators are wearing head-mounted displays. This type of viewing apparatus should create an even more immersive one (albeit possibly nausea-inducing).

Extensions for Non-Tour 360° Videos. Our approach relies on 360° video tours, where a single camera moves through a relatively fixed architectural space. Yet, while many videos on popular platforms are recorded as tour videos, not all 360° videos are recorded in this way. We need to investigate how non-tour 360° videos are explored and watched to understand what kinds of approaches would be appropriate for viewing collaboratively with others.

This work presents Tourgether360, a prototype designed for collaborative viewing of 360° video tours. It presents a novel interface for exploring 360° video tours that is grounded spatial navigation (as opposed to temporal navigation), and is designed to support collaborative exploration of such videos between multiple viewers.

REFERENCES

[1] Steve Benford, John Bowers, Lennart E. Fahlén, Chris Greenhalgh, and Dave Snowdon. 1995. User embodiment in collaborative virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '95), ACM Press/Addison-Wesley Publishing Co., USA, 242–249.

[2] Steve Benford and Lennart E. Fahlén. 1993. Awareness, focus, and aura: A spatial model of interaction in virtual worlds. *ADVANCES IN HUMAN FACTORS ERGONOMICS* 19, (1993), 693–693.

[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), Association for Computing Machinery, New York, NY, USA, 502–503.

[4] Diana Fonseca and Martin Kraus. 2016. A comparison of head-mounted and hand-held displays for 360° videos with focus on attitude and behavior change. In *Proceedings of the 20th International Academic Mindtrek Conference* (AcademicMindtrek '16), Association for Computing Machinery, New York, NY, USA, 287–296.

[5] Mike Fraser, Steve Benford, Jon Hindmarsh, and Christian Heath. 1999. Supporting awareness and interaction through collaborative virtual interfaces. In *Proceedings of the 12th annual ACM symposium on User interface software and technology* (UIST '99), Association for Computing Machinery, New York, NY, USA, 27–36.

[6] Chris Greenhalgh and Steven Benford. 1995. MASSIVE: a collaborative virtual environment for teleconferencing. *ACM Transactions on Computer-Human Interaction (TOCHI)* 2, 3 (September 1995), 239–261.

[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), Association for Computing Machinery, New York, NY, USA, 207–216.

[8] Carl Gutwin and Saul Greenberg. 2002. A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work* 11, 3–4 (2002), 411–446.

[9] Kyoungkook Kang and Sunghyun Cho. 2019. Interactive and automatic navigation for 360° video playback. *ACM Transactions on Graphics (TOG)* 38, 4 (July 2019), 1–11.

[10] Jiannan Li, Jiahe Lyu, Mauricio Sousa, Ravin Balakrishnan, Anthony Tang, and Tovi Grossman. 2021. Route tapestries: Navigating 360° virtual tour videos using slit-scan visualizations. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, ACM, New York, NY, USA. DOI:https://doi.org/10.1145/3472749.3474746

[11] Yen-Chen Lin, Yung-Ju Chang, Hou-Ning Hu, Hsien-Tzu Cheng, Chi-Wen Huang, and Min Sun. 2017. Tell me where to look. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ACM, New York, NY, USA. DOI:https://doi.org/10.1145/3025453.3025757

[12] Ville Mäkelä, Tuuli Keskinen, John Mäkelä, Pekka Kallioniemi, Jussi Karhu, Kimmo Ronkainen, Alisa Burova, Jaakko Hakulinen, and Markku Turunen. 2019. What Are Others Looking at? Exploring 360° Videos on HMDs with Visual Cues about Other Viewers. In *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video* (TVX '19), Association for Computing Machinery, New York, NY, USA, 13–24.

[13] Raúl Mur-Artal and Juan D. Tardós. 2017. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. *IEEE transactions on robotics* 33, 5 (October 2017), 1255–1262.

[14] Luís A. R. Neng and Teresa Chambel. 2010. Get around 360° hypervideo. In *Proceedings of the 14th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek '10*, ACM Press, New York, New York, USA. DOI:https://doi.org/10.1145/1930488.1930512

[15] Cuong Nguyen, Stephen DiVerdi, Aaron Hertzmann, and Feng Liu. 2017. Vremiere: In-Headset Virtual Reality Video Editing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 5428–5438.

[16] Cuong Nguyen, Stephen DiVerdi, Aaron Hertzmann, and Feng Liu. 2017. CollaVR: Collaborative In-Headset Review for VR Video. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (UIST '17), Association for Computing Machinery, New York, NY, USA, 267–277.

[17] Gonçalo Noronha, Carlos Álvares, and Teresa Chambel. 2012. Sight surfers: 360º videos and maps navigation. In *Proceedings of the ACM multimedia 2012 workshop on Geotagging and its applications in multimedia* (GeoMM '12), Association for Computing Machinery, New York, NY, USA, 19–22.

[18] Amy Pavel, Björn Hartmann, and Maneesh Agrawala. 2017. Shot Orientation Controls for Interactive Cinematography with 360 Video. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (UIST '17), Association for Computing Machinery, New York, NY, USA, 289–297.

[19] Benjamin Petry and Jochen Huber. 2015. Towards effective interaction with omnidirectional videos using immersive virtual reality headsets. In *Proceedings of the 6th Augmented Human International Conference* (AH '15), Association for Computing Machinery, New York, NY, USA, 217–218.

[20] Sylvia Rothe, Mario Montagud, Christian Mai, Daniel Buschek, and Heinrich Hußmann. 2018. Social Viewing in Cinematic Virtual Reality: Challenges and Opportunities. In *Interactive Storytelling*, Springer International Publishing, 338–342.

[21] Gustavo Alberto Rovelo Ruiz, Davy Vanacken, Kris Luyten, Francisco Abad, and Emilio Camahort. 2014. Multi-viewer gesture-based interaction for omni-directional video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14), Association for Computing Machinery, New York, NY, USA, 4077–4086.

[22] Mehrnaz Sabet, Mania Orand, and David W. McDonald. 2021. Designing Telepresence Drones to Support Synchronous, Mid-air Remote Collaboration: An Exploratory Study. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (CHI '21), Association for Computing Machinery, New York, NY, USA, 1–17.

[23] Ben Shneiderman. 2003. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *The Craft of Information Visualization*, Benjamin B. Bederson and Ben Shneiderman (eds.). Morgan Kaufmann, San Francisco, 364–371.

[24] Anthony Tang and Omid Fakourfar. 2017. Watching 360° Videos Together. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ACM, New York, NY, USA.

[25] Anthony Tang, Omid Fakourfar, Carman Neustaedter, and Scott Bateman. 2017. Collaboration in 360° Videochat: Challenges and Opportunities.

[26] Audrey Tse, Charlene Jennett, Joanne Moore, Zillah Watson, Jacob Rigby, and Anna L. Cox. 2017. Was I There? Impact of Platform and Headphones on 360 Video Immersion. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '17), Association for Computing Machinery, New York, NY, USA, 2967–2974.

[27] Nelson Wong and Carl Gutwin. 2010. Where are you pointing? the accuracy of deictic pointing in CVEs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1029–1038.

[28] 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), Association for Computing Machinery, New York, NY, USA, 1377–1387.