Electronic Preprint for Journal of Information Processing Vol.26

Regular Paper

StreamSpace: Pervasive Mixed Reality Telepresence for Remote Collaboration on Mobile Devices

Bektur Ryskeldiev[1,a)] Michael Cohen[1,b)] Jens Herder[2,c)]

Received: May 24, 2017, Accepted: November 7, 2017

Abstract: We present a system that exploits mobile rotational tracking and photospherical imagery to allow users to share their environment with remotely connected peers "on the go." We surveyed related interfaces and developed a unique groupware application that shares a mixed reality space with spatially-oriented live video feeds. Users can collaborate through realtime audio, video, and drawings in a virtual space. The developed system was tested in a preliminary user study, which confirmed an increase in spatial and situational awareness among viewers as well as reduction in cognitive workload. Believing that our system provides a novel style of collaboration in mixed reality environments, we discuss future applications and extensions of our prototype.

Keywords: mixed reality, telepresence, remote collaboration, live video streaming, mobile technologies, ubiquitous computing, photospherical imagery, spatial media

1. Introduction

In the past two years, photospherical imagery has become a popular format for still photo and live video streaming on both fixed and mobile platforms. With social networks such as Facebook, Twitter (via Periscope), and YouTube, users can quickly share their environment with connected peers "on the go." When captured, panoramas are typically geotagged with information, which allows using such imagery for the reconstruction of real locations in virtual spaces. We are interested in how such technology can be applied to remote collaboration, creating a quick way of sharing snapshots of real environments so that distributed users can work together.

We propose "StreamSpace": a system that uses mobile video streaming and omnidirectional mixed reality spaces for remote collaboration. It uses photospherical imagery (captured by a user or downloaded from elsewhere) to create a spherical background, i.e., a snapshot of a real location, upon which local (streaming) and remote (viewing) users can collaborate. The local users' viewpoints are represented by live video streams, composited as moving video billboards (rectangles that always "face" a viewer), spatially distributed around the photosphere, providing realtime updates of the local scene. Remote users are virtually placed in the center of the sphere, and can freely look around the location. Both local and remote users can collaborate through audio and video streams, as well as realtime drawing in a virtual space.

StreamSpace's use of web-served photospherical imagery and mobile rotational tracking makes it highly adaptive to different streaming scenarios, as it can work both with web-served and user-captured photospheres, and does not require external objects or additional steps for tracking calibration. Furthermore, our application is supported on multiple Android devices and does not draw excessive computational power for a collaborative session. Such factors make our application "pervasive," i.e., applicable to a wide range of various users in different scenarios, advancing the state of mobile collaboration groupware.

Finally, StreamSpace provides a novel approach to mixed reality collaboration on mobile devices, and in the rest of this paper we discuss similar state-of-the-art solutions, implementation, preliminary testing, current status of the application, and possible future extensions of our solution.

2. Background

2.1 Classification of Mixed Reality Systems

Before discussing mixed reality systems, it is necessary to establish a taxonomy in which mixed reality experiences can be qualitatively compared. The mixed reality (MR) classification was originally proposed by Milgram and Kishino [3] in the form of a Reality–Virtuality (RV) continuum (**Fig. 2**, top). The RV continuum locates mixed reality applications on a spectrum between real and virtual environments, and classifies mixed reality experiences as augmented reality (AR) or augmented virtuality (AV).

As discussed by Billinghurst [1], the RV continuum was further expanded by Milgram et al. [2] to provide a more detailed classification of mixed reality displays. The extended mixed reality display taxonomy includes three continua (Fig. 2, bottom):

- *Extent of World Knowledge* (*EWK*) represents the amount of real world modeled and recognized by a mixed reality system. "World Unmodelled" means that the system knows nothing about the real world, and "World Modelled" implies a complete understanding of the real world.

[1] University of Aizu, Aizu-Wakamatsu, Fukushima 965–8580, Japan
[2] Hochschule Düsseldorf: University of Applied Sciences, 40476 Düsseldorf, Germany
[a)] d8171101@u-aizu.ac.jp
[b)] mcohen@u-aizu.ac.jp
[c)] jens.herder@hs-duesseldorf.de

(a) (b) (c)

Fig. 1 (a) Streaming mode user interface, (b) Viewing mode scene overview, (c) Live snapshot of collaborative session with multiple streamers and a viewer.

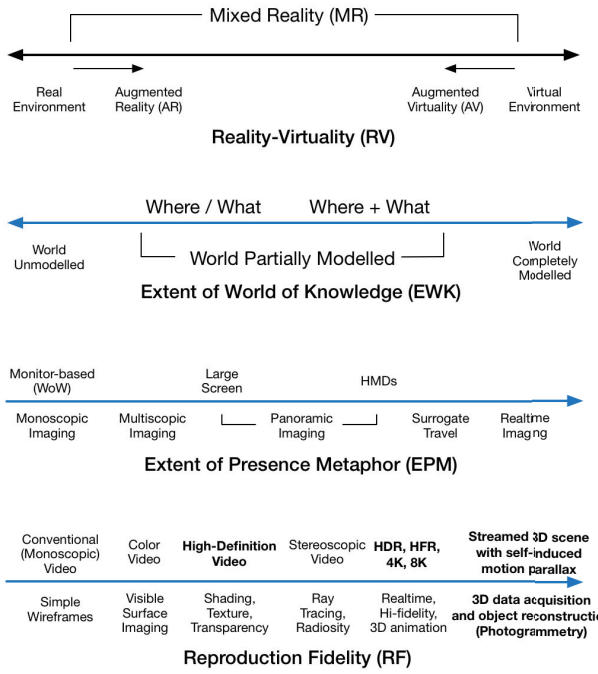


Fig. 2 RV continuum and extended mixed reality taxonomy, as presented in Refs. [1] and [2].

- *Extent of Presence Metaphor* (*EPM*), i.e., the level of user immersion in a scene, is described on a spectrum between monoscopic imaging (minimal immersion) and real-time imaging (deep immersion). We also note that the use of "presence" here is different from the more recent and now broadly accepted interpretation as subjective impression [4], reserving "immersion" to basically mean objective richness of media.
- *Reproduction Fidelity* (*RF*) describes how detailed is the reproduced world in a mixed reality system, with monoscopic video as the lowest fidelity reproduction, ranging to 3D high-definition reproductions as the highest. Because the initial definition of this continuum was introduced over twenty years ago, we have adjusted it to reflect recent changes in the mixed reality display technology. Namely we added (in bold type) high dynamic-range imaging (HDR), high frame rate (HFR), 4K and 8K video standards; introduced photogrammetry and high-definition 3D scene streaming; and shifted high-definition video to appear earlier in the spectrum due to it being more common nowadays.

Using these spectra it is possible to estimate and compare

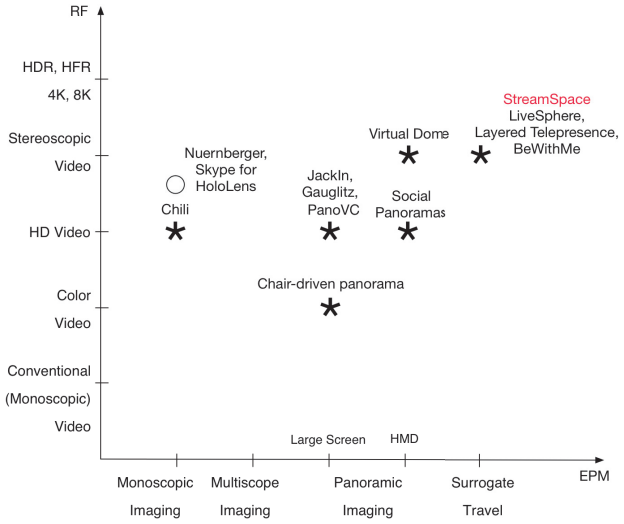


Fig. 3 Distribution of mixed reality systems based on an extended taxonomy. Since most of the presented examples are at the bottom end of EWK, the solutions are distinguished only along RF and EPM axes. The two studies at the middle of the EWK spectrum are shown as a circle.

the quality of real world capture, immersion, and reproduction among different mixed reality displays. For particular instance, StreamSpace is at the "World Unmodelled" end of the EWK continuum, because the system uses only rotational tracking; the upper end of the EPM spectrum, since it allows "Surrogate Travel"; and the middle of RF spectrum, due to stereoscopic video support (via Google Cardboard SDK) (**Fig. 3**).

2.2 Related Works

One of the earliest virtual reality systems using photospherical imagery was described by Hirose in "Virtual Dome" and its extending refinements [5]. The system presented a set of images captured asynchronously by a rotated camera, buffered in a sphere (bottom end of EWK, middle of EPM), which could be browsed through a head-mounted display ("Stereoscopic Video" of RF). Virtual Dome extensions included the introduction of motion parallax and GPS tracking. While similar to StreamSpace, Virtual Dome required specific camera equipment for capture, and a specific viewing environment for reproduction, making the system semi-tethered and non-pervasive (i.e., unavailable to regular users).

The advantage of mixed reality systems for remote collaboration was affirmed by Billinghurst and Kato in their "Collaborative

Augmented Reality" survey [6], which reported improved sense of presence and situational awareness compared to regular audio- and videoconferencing solutions. They also noted the necessity of handheld displays for wider adoption of mixed reality techniques in collaborative scenarios.

Cohen et al. [7] also noted the limitations of tethered head-mounted displays in such mixed reality systems as Virtual Dome and developed a motion platform system for navigation around panoramic spaces. It used a regular laptop screen ("Color Video" of RF) for panoramic display (bottom end of EWK, middle of EPM) and a rotating swivel chair. Although the system aimed to be fully untethered, it still used chair-driven gestures for interaction.

Fully untethered collaborative mixed reality telepresence was presented in "Chili" by Jo et al. [8]. The application allowed users to control the viewpoint of a remote scene with gestures and on-screen drawings on their mobile devices. Using the extended MR taxonomy, Chili would be at the bottom end of EWK since its feature detection algorithm was used only for drawing stabilization, at the "High Definition Video" point of RF, and the "Monoscopic Imaging" end of EPM. The last characterization also implies that users' viewpoints were bound together, and a viewer could not freely explore a mixed reality space without being attached to the streaming user's viewpoint.

Similarly, such overlaid video annotation aspect was also investigated in Skype for HoloLens in a study by Chen et al. [9], and for mobile and desktop platforms in a study by Nuernberger et al. [10]. Although both studies featured a monoscopic high-definition video similarly to Chili, their tracking approach partially modeled the world around them, which puts both studies in the middle of the EWK spectrum.

Free navigation around panoramas was presented in Bing Maps by Microsoft [11], where a user could overlay a realtime video stream over a photospherical panorama of a location, but the system was limited to photospherical imagery provided by Microsoft and at the time it did not fully support mobile devices. Mobile platform deployment was explored by Billinghurst et al. in "Social Panoramas" [12], with which users could access static panoramic images (bottom of EWK, "High Definition Video" of RF, middle of EPM) and collaborate by drawing on them in real-time.

Panoramic realtime updates were demonstrated in systems by Gauglitz et al. [13] and Kasahara et al. in "JackIn" [14]. JackIn used a head-mounted camera and simultaneous localization and mapping (SLAM) to create an updated photospherical image from stitched photos (bottom end of EWK, middle of EPM) which could be viewed and interacted with through a high definition display ("High Definition Video" end of RF). Similarly, the system proposed by Gauglitz et al. used a handheld mobile tablet for image capture. Both solutions, however, required a large motion parallax for stable tracking and creating a complete panoramic image.

This issue was addressed by Nagai et al. in LiveSphere [15], which used a set of head-mounted cameras, eliminating the need for movement around the location to capture a complete panorama. Similarly, Saraiji et al. in "Layered Telepresence" [16] used a HMD with an embedded eye tracker to switch between blended stereoscopic video streams originating from cameras on robot heads. Both studies fall at the bottom of EWK, at the "Stereoscopic Video" point of RF, and "Surrogate Travel" of EPM.

Such solutions, however, still required a relatively high computational power for mobile devices, which was addressed in PanoVC by Müller et al. [17]. Instead of a set of head-mounted cameras, PanoVC used a mobile phone to render a continuously updated cylindrical panorama (bottom of EWK, "High Definition Video" point of RF, and middle of EPM), in which users could see each others' current viewpoints and interact through overlaid drawings.

Finally, while PanoVC used a set of static images to create a live panorama, Singhal et al. in BeWithMe [18] used a panoramic video streaming (bottom end of EWK, "Stereoscopic Video" of RF, and "Surrogate Travel" of EPM) for mobile telepresence. It allowed immediate capture of user surroundings, but the resulting experience was still limited to only two users at a time.

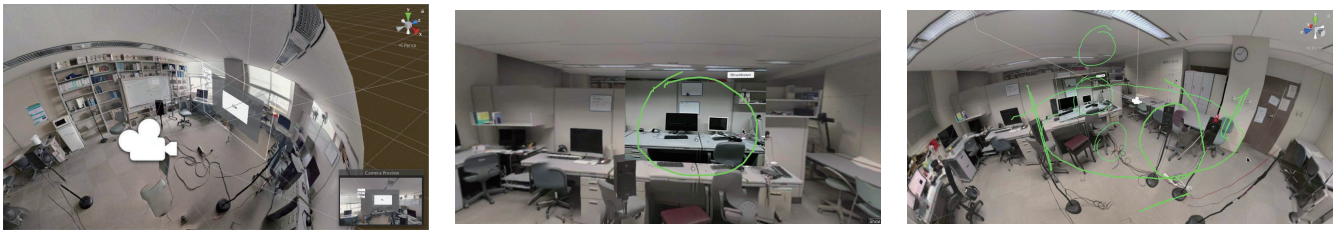
3. Implementation

Addressing such limitations, we designed StreamSpace to support multiple users, provide panoramic background with realtime updates, and be able to adapt to both fixed and mobile scenarios. Furthermore, compared using the extended taxonomy (Fig. 3) with similar solutions, ours is among the most immersive and high fidelity mixed reality displays.

3.1 System Overview

StreamSpace is based on the Unity game engine and supports Android mobile devices. For rotational tracking, it uses the Google Cardboard SDK, which also makes the application compatible with both handheld and HMD modes. An environment-mapping photosphere might be captured just before a realtime session, but its asynchrony invites alternative juxtapositions. For instance, temporal layers could alternate among different times of day, seasons, or conditions (like "before & after" comparisons). Synaesthetic displays such as IR heat-maps or arbitrary info-viz contextual renderings can interestingly complement real-time overlays. The panorama itself is mapped onto a sphere with a web texture, which allows integrating such backgrounds with external sources, including indoor positioning systems such as iBeacon, Google VPS (Visual Positioning Service) [19], or public navigation services such as Google Street View. If a photosphere and a user's current viewpoint are misaligned, the user can manually rotate the photosphere and the offset will be synchronized with other users simultaneously.

The system operates in two modes (**Fig. 1**): streaming and viewing. In both cases users can browse and interact within a mixed reality space. When a user is in a viewing mode, their space features multiple video stream billboards, while a streaming mode features only one fixed billboard, the user's local video feed. The streaming user's orientation is used to adjust the corresponding video stream billboards in viewing clients in real-time. Furthermore, streaming users can also change the photospherical image (either by uploading their own or by sharing one from else-



(a) (b) (c)

Fig. 4 Examples of 3D drawings. (a) Close-up of a Unity scene, the camera in the center of the scene representing the viewer's position. (b) Drawing by a streamer. (c) Scene overview after a typical collaborative drawing session.

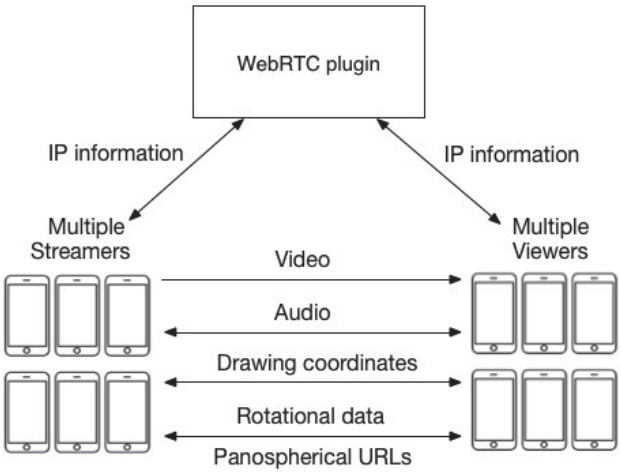


Fig. 5 StreamSpace connections and dataflow diagram.

where) for all users, whereas viewers can only passively receive new panoramic images.

StreamSpace user interaction features not only audio and video streaming, but also drawing. Since the virtual space is a 3D scene (built in Unity), users' touchscreen coordinates are converted to virtual space by ray-casting onto a transparent plane in front of the camera (i.e., the user's viewpoint in the virtual space). Since the camera rotation is adjusted through mobile rotational tracking, drawings can be three-dimensional, and are shared among streaming and viewing users simultaneously (**Fig. 4**).

Networking is handled through the Web Real-time Communication (WebRTC) protocol [20]. We chose WebRTC due to its ability to establish connections among remote peers using network address translation (NAT) traversal technologies, i.e., connecting users without prior knowledge of each others' IP addresses. Furthermore, WebRTC supports multiple simultaneous connections, and works both over mobile networks (3G, 4G) and wireless LAN (Wi-Fi), and is future-proof, since it will also run on WiGig and 5G.

The WebRTC implementation is provided through the mobile version of the "WebRTC Videochat" plugin for Unity [21]. It sends and receives audio from all connected users, sends streamers' video feeds to viewers in a native resolution, and handles synchronization of drawing coordinates, user rotational data, links to panoramic images, and the photosphere's rotational offset (**Fig. 5**).

Even though we tested the system with panoramas captured through Insta360 Air camera [22] and Android's built-in camera application, StreamSpace assumes that a streaming user has a URL of a captured panoramic image prior to the beginning of a session.

4. Preliminary User Testing

4.1 Experiment Design

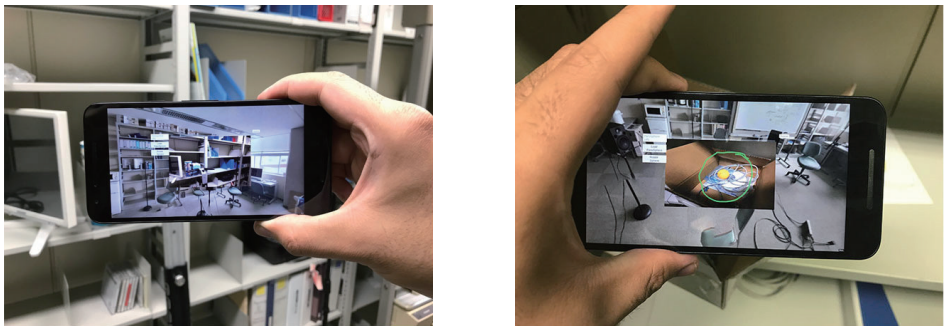
To confirm the feasibility of our approach, we conducted preliminary user testing (**Figs. 6** and **7**). Our experimental hypotheses conject that, compared with regular teleconferencing systems, our solution imposes less cognitive workload and increases spatial and situational awareness.

Since we suppose that differences in the user interface between our solution and commercially available applications would confuse participants, we developed a separate regular videoconferencing mode within StreamSpace, which we call "Flat," since it does not use a mixed reality space (and we abbreviate StreamSpace as "Space" for convenience). The flat videoconferencing mode projects a simple video rectangle with a connected peer's video stream and two buttons that start or stop the connection. In this mode the application supports only one viewing and one streaming user, and provides only audio and video streaming (with no drawing).

Furthermore we intentionally excluded the capture of photospherical imagery from the experiment, because our system is designed to support panoramas captured through third-party applications. Before each trial we uploaded a panorama of the room in which the experiment was conducted, captured with Insta360 Air or Ricoh Theta S cameras.

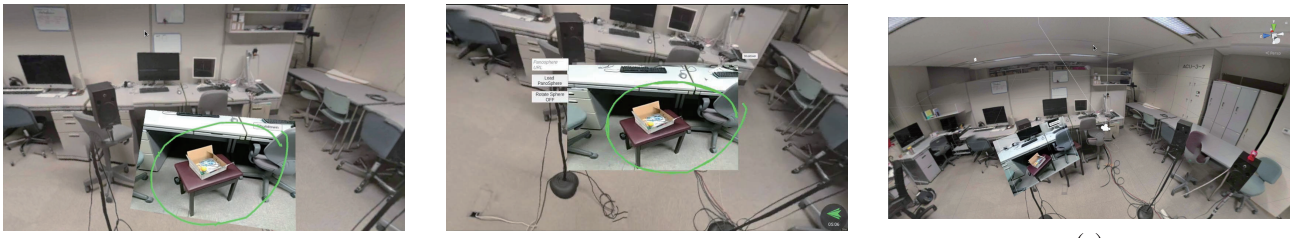
The experiment itself had the following steps:

(1) Each trial consisted of two sessions: one running StreamSpace in Flat mode, and another in photospherical Space mode.
(2) The session order was determined randomly before the start of each trial.
(3) At the start of each session, two users were located in two different rooms. One user was the designated Viewer, the other was the Streamer (and these roles were retained until the end of the trial).
(4) In each room was hidden an object of the same type (e.g., an orange table tennis ball). The hiding locations were relatively similar to ensure uniform complexity of performed tasks.
(5) The Viewer received an explanation about where the target was hidden, and he or she was requested to explain it



(a) (b)

Fig. 6 Preliminary user testing scenario for StreamSpace. (a) Streamer walking around the location, (b) Streamer successfully finding the ball, highlighting it, and concluding an experiment's session.



(a) (b) (c)

Fig. 7 On-screen view example of a testing scenario for StreamSpace. (a) Viewer, (b) streamer, (c) and exocentric view of the scene.

to the Streamer using our application in different modes (depending on the session).

(6) Each session ended when the Streamer found the hidden object, and the time taken to completed each session was recorded.

(7) After both sessions users completed a questionnaire, one for each session, and provided additional comments.

The questionnaire was based on Likert-like questions on spatial understanding introduced by Kasahara et al. [14] in JackIn, namely: "Q1: Ease in finding the target," and "Q2: Ease in understanding of the remote situation," where the scale ranged between 1 (disagree) and 7 (agree) points. However, we replaced the last two questions regarding cognitive workload with questions from the unweighted NASA Task Load Index test [23], also known as "raw" TLX or RTLX.

The choice of RTLX over traditional TLX testing was deliberate. On the participant side, the traditional NASA TLX test requires two steps: measuring participant workload on six subscales presented by the questionnaire, and then creating an individual weighting for each subscale through pairwise comparison regarding their perceived importance. RTLX omits the second part, which allows faster execution of the experiment while still providing valid results that are highly correlated with traditional TLX scores [24].

4.2 Setup

The experiments were conducted on campus at both the University of Aizu (UoA) and Hochschule Düsseldorf: University of Applied Sciences (HSD). We recruited forty participants (or twenty pairs) in total, including university students and staff. The participants' age range was from twenty to fifty years old, and included ten women and thirty men. Some subjects were financially compensated, while others refused payment.

The devices used for testing were provided by us and consisted of:

- UoA: Samsung Galaxy S7 running Android 6.0.1, LG Nexus 5X with Android 7.1.2
- HSD: Samsung Galaxy Note 3 with Android 5.1.1, ASUS Zenfone AR Prototype with Android 7.0.

All devices were connected over local 2.4 GHz and 5 GHz Wi-Fi networks supporting the IEEE 802.11n wireless-networking standard.

The rooms in which the experiments were conducted were different as well. In the UoA, the room was separated by a cubicle partition into two smaller rooms, and while the two users could not see each other, they could hear each other if they spoke loudly, although users preferred to use the voice communication functionality provided by the application. At HSD the first pair of rooms (HSD-A) was similar to those at UoA, except the rooms were separated by desks, so the users could occasionally see each other, but in the second (HSD-B) the users were placed in completely different locations. In total we conducted fourteen sessions at the UoA and six at HSD (three each in HSD-A and HSD-B).

4.3 Analysis

First we calculated results including the data obtained from HSD-A, but since after conducting the experiment we could not ensure that in HSD-A the participants could not see each other completely, we decided to exclude its data for reanalysis. However, since the sample size of HSD-A was relatively small (only 3 test cases), the exclusion did not significantly affect the overall outcome. We also noticed several outlying results in Q1, Q2, and time measurements that differed significantly from the main re-

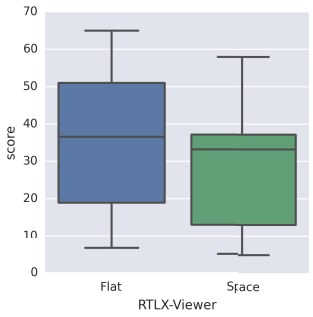


Fig. 8 RTLX Viewer Scores (with HSD-A data).

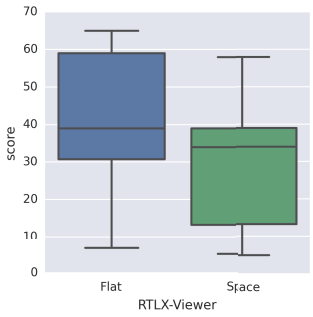


Fig. 9 RTLX Viewer Scores (without HSD-A data).

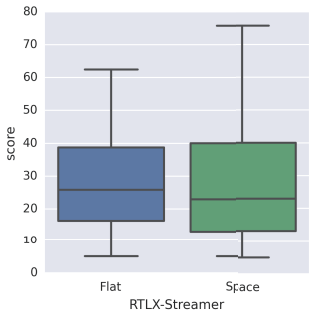


Fig. 10 RTLX Streamer Scores (with HSD-A data).

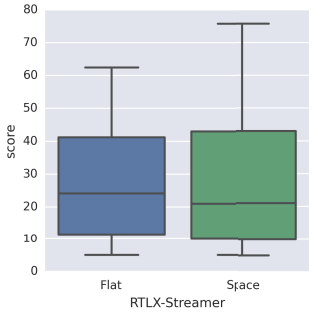


Fig. 11 RTLX Streamer Scores (without HSD-A data).

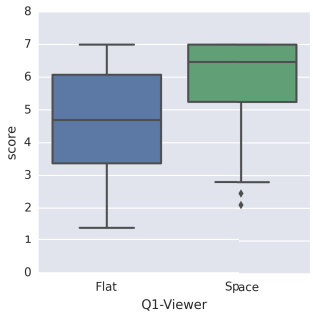


Fig. 12 Spatial awareness question scores for viewers (with HSD-A data).

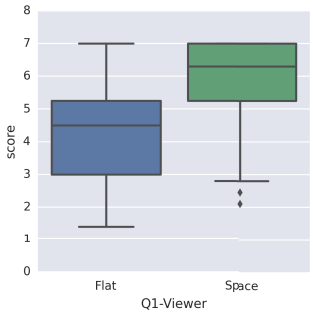


Fig. 13 Spatial awareness question scores for viewers (without HSD-A data).

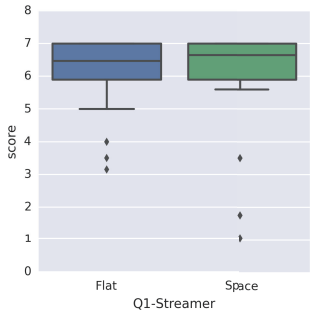


Fig. 14 Spatial awareness question scores for streamers (with HSD-A data).

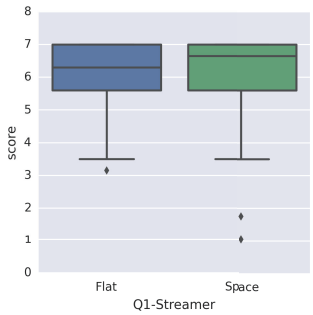


Fig. 15 Spatial awareness question scores for streamers (without HSD-A data).

sults. We investigated the relevant test cases, but did not discover abnormalities (all experiments were conducted properly and there were no major issues reported), and therefore decided to retain them in calculations.

In 12 (10 without HSD-A) pairs out of 20, RTLX Viewer scores were lower for the Space mode than for the Flat mode, which is also reflected in the RTLX scores (**Figs. 8** and **9**). This trend is confirmed by Student's paired t-test results (df = 20 and 17) with $p < 0.05$ for Viewers. Streamers, on the other hand, did not show any significant improvement, with $p > 0.05$, with most of the scores similar for both the flat and space modes (**Figs. 10** and **11**).

For the spatial and situational awareness (**Fig. 12** to **Fig. 19**) we observed a strong improvement in scores for Viewers with $p < 0.05$ for Q1 and $p < 0.05$ for Q2, but with unfortunately no statistically significant results for Streamers ($p > 0.05$ in both cases).

The time scores (**Figs. 20** and **21**) showed a significant reduc-

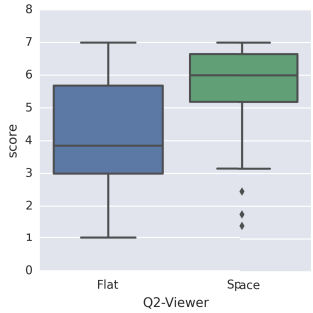


Fig. 16 Situational awareness question scores for viewers (with HSD-A data).

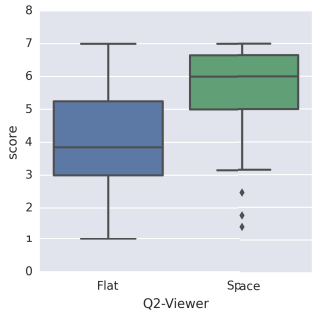


Fig. 17 Situational awareness question scores for viewers (without HSD-A data).

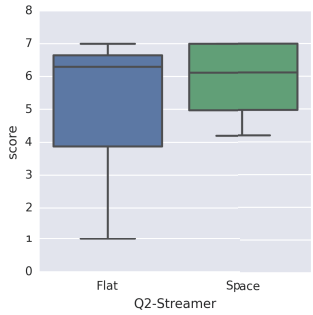


Fig. 18 Situational awareness question scores for streamers (with HSD-A data).

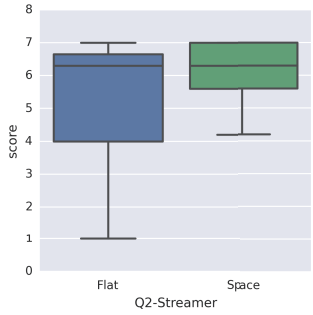


Fig. 19 Situational awareness question scores for streamers (without HSD-A data).

tion in elapsed time in Space mode as compared to Flat, but the results were not statistically significant ($p > 0.05$).

We were also able to observe some interesting effects of environment on the user performance. Although the sample size of UoA participants was more than twice as large as that of HSD (14 and 6 respectively), we noted similar mean RTLX, Q1 and Q2 scores between rooms at UoA and HSD-A. We can understand this consistency by the fact that HSD-A and UoA environments were of similar size and layout (although in UoA we could guarantee the lack of visual confirmation, whereas in HSD-A we could not). HSD-B, however, was conducted in two completely different rooms, and expectedly showed an increase in mean RTLX scores. It also had the lowest mean Q1 score among panoramic Streamers, and halving of mean elapsed time. We think that such differences in HSD-B results can be explained by location, unfamiliarity with which disoriented test participants.

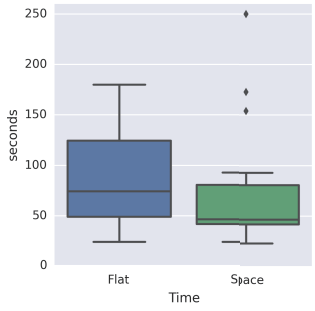


Fig. 20 Elapsed time (with HSD-A data).

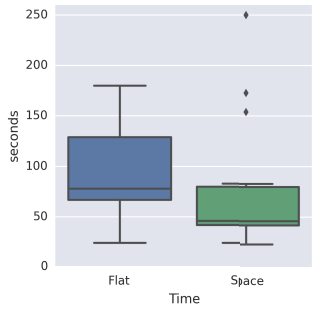


Fig. 21 Elapsed time (without HSD-A data).

We found the conditions in HSD-B to be the closest to how we expect our application to be used in real life scenarios, and are hoping to conduct more experiments in similar environments.

5. Conclusion

5.1 Discussion and User Feedback

We have developed an application that allows sharing photospherical imagery of real environments with remotely connected peers and using it for mixed reality collaboration on the go. Preliminary testing has shown that among viewing users, our approach does improve spatial and situational awareness, and reduces cognitive workload.

For streamers, however, our approach did not provide statistically significant improvement, which could be explained by user interface issues. For instance, on the streaming side a user can see the real environment, its photospherical snapshot, and the same environment again in the centered live video feed from the user's mobile camera. This could cause some confusion, which also explains the outlier results in measurements, as it seems that although they did not encounter any serious issues, users took more time to adjust to the interface and unknown environment.

For future revisions we plan to replace the streaming interface by a full-screen live video feed with embedded three-dimensional drawings, as in "Chili" [8], or studies by Gauglitz et al. [13] and

Chen et al. [9].

Users seemed to like the introduced photospherical aspect of our mixed reality interaction, as they could navigate around a panorama without being tethered to a streamer's viewpoint, which indicates that having the application operate at a higher level of the EPM spectrum could indeed improve collaborative aspects. Our assumption is also confirmed by the latest update of the JackIn project by Kasahara et al., which switched from SLAM-based panoramas to spherical video streaming [25].

Aside from the panoramic aspect, all users commented that they found the application interesting, and the collaborative drawing aspect to be flattering for groupware sessions. We were also requested to add such features as a haptic feedback and HMD integration to improve the immersion.

5.2 Future Work

Aside from user feedback, we are also exploring integrating markerless tracking through such systems as Kudan [26] and Google Visual Positioning Service [19], or HMDs like Google Daydream View [27] and Microsoft HoloLens [28]. Such integration would allow the system to move into the "World Partially Modelled" range of the EWK spectrum, providing more interesting modes of user interaction. For example, by using markerless feature detection of a scene, a streaming user could recreate an environment and send a three-dimensional map of real space to viewers, who could "touch" its surfaces through haptic controls, as demonstrated in different Virtual Reality studies, such as Lopes et al. [29]. Furthermore, since the panoramic background can feature different synaesthetic displays such as IR heat-maps, haptic interaction could be extended to feature thermoception.

Inclusion of advanced tracking and mapping in our system could also help address the issue of field-of-view (FoV) matching. Currently our system uses a "naïve" approach to FoV management, and hopes that the video feed and the photosphere "fit together." However, this is not always the case, given the variety of Android device cameras and screens, and the wide variety of photospherical images available on the web. Since recent studies indicate that FoV differences have a strong effect on collaboration in mixed reality environments [30], we hope to improve the FoV management aspect of our application. One of the possible solutions for that could be to use markerless tracking systems such as Apple ARKit [31] or Android ARCore [32] to determine the user displacement in a scene, or alternatively, implement a machine learning approach that could automatically readjust either the photosphere or a video feed to create a matching image.

Even though we have used the words "streamer" and "viewer" to distinguish the two peer-to-peer modes in StreamSpace, the feeds are actually multimodal, and currently also include audio, so better descriptions that generalize to such multimodal media would be "source" and "sink." Such voice streams could be directionalized from their respective projected realtime video rectangles. YouTube uses FOA, first-order Ambisonics [33], to project spatial soundscape recordings, but we could use even simple rendering such as lateral intensity panning. Monaural streams, capturable by smartphone proximity microphones, can be lateralized into stereo pairs at each terminal that encode the azimuthal direction of each streaming source's visual contribution. Even though such rendered soundscapes are not veridical, in the sense that such displaced auditory rendering deliberately ignores the logical colocation of source and sink virtual standpoints, we think that such aural separation would flatter groupware sessions and enhance the situation awareness.

Another interesting extension would be the implementation of stereoscopic video streaming. Our system allows streaming video in the original resolution, and supports such mobile HMDs as Google Cardboard. Due to coherent rotational tracking data, mobile device cameras can operate as a single stereoscopic camera when paired side-by-side, sending binocular video streams to viewers.

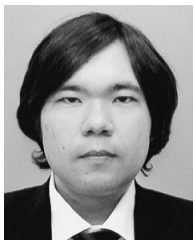
Finally, we are also working on the integration of StreamSpace with publicly available services for panoramic imagery, such as Google Street View, Facebook Live, YouTube, and Periscope [34]. We find current modes of interaction in social networks with panoramic video streams to be mostly two-dimensional (e.g., plain text messages or "Like" buttons). With rotational tracking and live video billboards, users would be able to highlight interesting details or viewing angles in panoramic streaming scenarios.

Acknowledgments We thank the IEEE Student Group, and students and staff at both the University of Aizu and University of Applied Sciences, Düsseldorf for assisting us with user tests. We also thank Arkady Zgonnikov of the University of Aizu for assisting us with the analysis of data gathered throughout this user study.

References

[1] Billinghurst, M.: What is Mixed Reality?, Medium Blogs (online), available from ⟨https://goo.gl/3aKt3o⟩ (accessed 2017-05-08).
[2] Milgram, P., Takemura, H., Utsumi, A. and Kishino, F.: Augmented reality: A class of displays on the reality-virtuality continuum, *Photonics for Industrial Applications*, pp.282–292, International Society for Optics and Photonics (1995).
[3] Milgram, P. and Kishino, F.: A taxonomy of mixed reality visual displays, *IEICE Trans. Inf. Syst.*, Vol.77, No.12, pp.1321–1329 (1994).
[4] Mel, S.: A note on presence terminology, *Presence Connect*, Vol.3, No.3, pp.1–5 (2003).
[5] Hirose, M.: Image-based virtual world generation, *IEEE Multimedia*, Vol.4, No.1, pp.27–33 (1997).
[6] Billinghurst, M. and Kato, H.: Collaborative augmented reality, *Comm. ACM*, Vol.45, No.7, pp.64–70 (2002).
[7] Cohen, M., Doi, K., Hattori, T. and Mine, Y.: Control of Navigable Panoramic Imagery with Information Furniture: Chair-Driven 2.5D Steering Through Multistandpoint QTVR Panoramas with Automatic Window Dilation, *Proc. CIT: 7th Int. Conf. **C**omputer and **I**nformation **T**echnology*, Miyazaki, T., Paik, I. and Wei, D. (Eds.), Aizu-Wakamatsu, Japan, pp.511–516 (online), DOI: 10.1109/CIT.2007.140 (2007).
[8] Jo, H. and Hwang, S.: Chili: Viewpoint control and on-video drawing for mobile video calls, *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pp.1425–1430, ACM (2013).
[9] Chen, H., Lee, A.S., Swift, M. and Tang, J.C.: 3D collaboration method over HoloLens and Skype end points, *Proc. 3rd Int. Workshop Immersive Media Experiences*, pp.27–30, ACM (2015).
[10] Nuernberger, B., Lien, K.-C., Höllerer, T. and Turk, M.: Interpreting 2D gesture annotations in 3D augmented reality, *3D User Interfaces (3DUI)*, pp.149–158, IEEE (2016).
[11] Aguera, B.: Microsoft Augmented-Reality Maps, TED Talks (online), available from ⟨https://goo.gl/Xn5vtr⟩ (accessed 2017-05-08).
[12] Billinghurst, M., Nassani, A. and Reichherzer, C.: Social Panoramas: Using Wearable Computers to Share Experiences, *SIGGRAPH Asia Mobile Graphics and Interactive Applications*, p.25:1, New York, ACM (2014).
[13] Gauglitz, S., Nuernberger, B., Turk, M. and Höllerer, T.: World-

stabilized annotations and virtual scene navigation for remote collaboration, *Proc. 27th Annual ACM Symp. User Interface Software and Technology*, pp.449–459 (2014).
[14] Kasahara, S. and Rekimoto, J.: JackIn: Integrating first-person view with out-of-body vision generation for human-human augmentation, *Proc. 5th Augmented Human Int. Conf.*, p.46, ACM (2014).
[15] Nagai, S., Kasahara, S. and Rekimoto, J.: Livesphere: Sharing the surrounding visual environment for immersive experience in remote collaboration, *Proc. 9th Int. Conf. Tangible, Embedded, and Embodied Interaction*, pp.113–116, ACM (2015).
[16] Saraiji, M.Y., Sugimoto, S., Fernando, C.L., Minamizawa, K. and Tachi, S.: Layered Telepresence: Simultaneous Multi Presence Experience Using Eye Gaze Based Perceptual Awareness Blending, *ACM SIGGRAPH 2016 Posters*, New York, pp.20:1–20:2 (online), DOI: 10.1145/2945078.2945098 (2016).
[17] Müller, J., Langlotz, T. and Regenbrecht, H.: PanoVC: Pervasive telepresence using mobile phones, *Pervasive Computing and Communications* (*PerCom*), pp.1–10, IEEE (2016).
[18] Singhal, S. and Neustaedter, C.: BeWithMe: An Immersive Telepresence System for Distance Separated Couples, *Companion of the 2017 ACM Conf. Computer Supported Cooperative Work and Social Computing*, pp.307–310 (2017).
[19] Brennan, D.: Watch Google's 'Visual Positioning Service' AR Tracking in Action, RoadToVR (online), available from ⟨https://goo.gl/2La1u8⟩ (accessed 2017-05-24).
[20] Bergkvist, A., Burnett, D.C., Jennings, C. and Narayanan, A.: Webrtc 1.0: Real-time communication between browsers, *Working draft, W3C*, Vol.91 (2012).
[21] Kutza, C.: WebRTC Video Chat plugin, Unity Asset Store page (online), available from ⟨https://www.assetstore.unity3d.com/en/#!/content/68030⟩ (accessed 2017-05-08).
[22] Insta360: Insta360 Air, product page (online), available from ⟨https://www.insta360.com/product/insta360-air⟩ (accessed 2017-05-08).
[23] Hart, S.G.: NASA-task load index (NASA-TLX); 20 years later, *Proc. Human Factors and Ergonomics Society Annual Meeting*, Vol.50, No.9, pp.904–908, Sage Publications Sage CA: Los Angeles (2006).
[24] Cao, A., Chintamani, K.K., Pandya, A.K. and Ellis, R.D.: NASA TLX: Software for assessing subjective mental workload, *Behavior Research Methods*, Vol.41, No.1, pp.113–117 (2009).
[25] Kasahara, S., Nagai, S. and Rekimoto, J.: JackIn Head: Immersive Visual Telepresence System with Omnidirectional Wearable Camera, *IEEE Trans. Visualization and Computer Graphics*, Vol.23, No.3, pp.1222–1234 (2017).
[26] Kudan Computer Vision: SLAM algorithms for all the future devices, homepage (online), available from ⟨https://www.kudan.eu/⟩ (accessed 2017-05-24).
[27] Google: Google Daydream, project page (online), available from ⟨https://vr.google.com/daydream/⟩ (accessed 2017-05-24).
[28] Microsoft: Microsoft HoloLens, product page (online), available from ⟨https://www.microsoft.com/en-us/hololens⟩ (accessed 2017-05-24).
[29] Lopes, P., You, S., Cheng, L.-P., Marwecki, S. and Baudisch, P.: Providing Haptics to Walls; Heavy Objects in Virtual Reality by Means of Electrical Muscle Stimulation, *Proc. 2017 CHI Conf. Human Factors in Computing Systems, CHI '17*, pp.1471–1482, New York, ACM (online), DOI: 10.1145/3025453.3025600 (2017).
[30] Ren, D., Goldschwendt, T., Chang, Y. and Höllerer, T.: Evaluating wide-field-of-view augmented reality with mixed reality simulation, *Virtual Reality* (*VR*), pp.93–102, IEEE (2016).
[31] Apple: ARKit, developer page (online), available from ⟨https://developer.apple.com/arkit/⟩ (accessed 2017-09-20).
[32] Google: Android ARCore, developer page (online), available from ⟨https://developers.google.com/ar/⟩ (accessed 2017-09-20).
[33] Google: Use spatial audio in 360-degree and VR videos, support page (online), available from ⟨https://goo.gl/R1r4cC⟩ (accessed 2017-05-24).
[34] Ryskeldiev, B., Cohen, M. and Herder, J.: Applying Rotational Tracking and Photospherical Imagery to Immersive Mobile Telepresence and Live Video Streaming Groupware, *SIGGRAPH Asia 2017 Mobile Graphics and Interactive Applications, SA '17*, No.2, pp.5:1–5:2, New York, ACM (online), DOI: 10.1145/3132787.3132813 (2017).



Bektur Ryskeldiev is a PhD student at University of Aizu. He received his Masters degree from the same university and is currently a Team Leader on the ACM SIGGRAPH International Resources Committee. His research interests include mixed reality, human-computer interaction, mobile development, and ubiquitous computing. He is a member of ACM SIGGRAPH, the IEEE Computer Society, and the IEEE Student Group at the University of Aizu.

Michael Cohen is a Professor in the Computer Arts Lab. at the University of Aizu, where he is also Director of the Information Systems Division. He has research interests in interactive multimedia; mobile computing, IoT, and ubicomp; audio windows and stereotelephony; computer music; digital typography and electronic publishing; hypermedia; and virtual and mixed reality. He is a member of the ACM, IEEE Computer Society, and the Virtual Reality Society of Japan.

Jens Herder was appointed in 2000 Professor at the Faculty of Media at the HS Düsseldorf, University of Applied Sciences. He is the founding Editor-in-Chief of the Journal of Virtual Reality and Broadcasting. At the Virtual Sets and Virtual Environments Laboratory, he experiments with interaction processes for new media applications. In 1993, he joined the Computer Industry Laboratory at the University of Aizu. He got his PhD from the University of Tsukuba in 1999 with the dissertation "A Sound Spatialization Resource Management Framework."