

Compositing real-time media streams for groupware panoramic browsing, situation awareness, and enriched user experience

BEKTUR RYSKELDIEV

A DISSERTATION

SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

IN COMPUTER SCIENCE AND ENGINEERING

Graduate Department of Computer and Information Systems

The University of Aizu

2018



© Copyright by Bektur Ryskeldiev
All Rights Reserved.

The thesis titled

Compositing real-time media streams for groupware panoramic
browsing, situation awareness, and enriched user experience

by

Bektur Ryskeldiev

is reviewed and approved by:

Chief referee

Professor

Michael Cohen



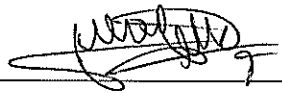
Professor

Ihor Lubashevsky



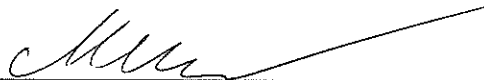
Associate Professor

Julián Villegas



Associate Professor

Maxim Mozgovoy



The University of Aizu

2018

TO MY PARENTS

Contents

LIST OF FIGURES	iii
LIST OF ABBREVIATIONS	v
ABSTRACT	1
1 INTRODUCTION	3
1.1 Live media streaming	3
1.2 Motivation and research goals	4
1.3 Main contributions	5
1.4 Thesis outline	8
2 BACKGROUND	11
2.1 Mobile mixed reality displays	12
2.1.1 Classification of mixed reality displays	12
2.1.2 Related virtual and mixed reality displays	14
2.1.3 Conclusion	16
2.2 Social livestreaming systems	17
2.2.1 Social livestreaming system taxonomy and relevant works	17
2.2.2 Extent of Social Interaction Space	18
2.2.3 Extent of User Communication	19
2.2.4 Reproduction Fidelity	20
2.2.5 Conclusion	20
3 DESIGN AND IMPLEMENTATION OF SPATIAL LIVESTREAM COMPOSITION (SLC)	
METHOD	23
3.1 Requirements	23
3.2 Method design	24
3.2.1 Spatial background	25
3.2.2 Spatialized live media streams	25
3.2.3 User interactions	26
3.3 Examples of possible applications of SLC method	26
4 COLLABORATIVE APPLICATIONS (STREAMSPACE)	29
4.1 Introduction	29
4.2 Implementation	31
4.2.1 System overview	31

4.3	User evaluation	33
4.3.1	Experiment design	33
4.3.2	Setup	36
4.3.3	Analysis	37
4.4	Conclusion	40
5	SOCIAL APPLICATIONS (REACTSPACE)	43
5.1	Introduction	43
5.2	Implementation	45
5.3	User evaluation	46
5.3.1	Experiment design	46
5.3.2	Analysis and results	48
5.4	Conclusion	49
6	EXTENDING THE SLC METHOD TO OTHER APPLICATIONS	51
6.1	Introduction	51
6.2	Background	52
6.3	Implementation	53
6.4	Conclusion	54
6.4.1	Benefits	54
6.4.2	Limitations	55
7	CONCLUSIONS	57
7.1	Theoretical contributions	57
7.2	Empirical contributions	59
7.3	Future work	60
7.3.1	Theoretical applications	61
7.3.2	Empirical applications	62
	ACKNOWLEDGMENTS	63
	REFERENCES	70

List of Figures

1.1	Thesis structure and suggested reading order	8
2.1	RV continuum and extended mixed reality taxonomy	12
2.2	The distribution of mixed reality systems based on extended taxonomy	16
2.3	Live streaming system taxonomy	18
3.1	SLC method implementation example	24
4.1	StreamSpace interface example	30
4.2	Examples of 3D drawings	31
4.3	StreamSpace connections and dataflow diagram	32
4.4	Experiment setup	34
4.5	User evaluation scenario for StreamSpace	35
4.6	On-screen view example of a testing scenario for StreamSpace	36
4.7	RTLX scores for viewers	38
4.8	RTLX scores for streamers	38
4.9	Spatial awareness scores for viewers	39
4.10	Spatial awareness scores for streamers	39
4.11	Situational awareness scores for viewers	40
4.12	Situational awareness scores for streamers	40
4.13	Elapsed time	40
5.1	Example of a live session	45
5.2	ReactSpace network dataflow diagram	46
5.3	Examples of Likert items	48
5.4	User evaluation results	49
6.1	Example of multiple mixed reality spaces in a single metaverse	53
6.2	Blockchain example and block content outline	53
6.3	Blockchain synchronization protocol	53

This page intentionally left blank.

List of Abbreviations

2D	Two-dimensional
3D	Three-dimensional
3G	Third generation network
4G	Fourth generation network
5G	Fifth generation network
4K, 8K	Orders of horizontal screen resolution 4,000 pixels and 8,000 pixels respectively
AR	Augmented Reality
AV	Augmented Virtuality
EWK	Extent of World of Knowledge
EPM	Extent of Presence Metaphor
ESIS	Extent of Social Interaction Space
EUC	Extent of User Communication
FoV	Field of View
IP	Internet Protocol
HDR	High Dynamic Range
HFR	High Frame Rate
HD	High Definition
HMD	Head-Mounted Display
LTE	Long-Term Evolution
LAN	Local Area Network
NAT	Network Address Translation
MR	Mixed Reality
QoE	Quality of Experience
RF	Reproduction Fidelity
RV	Reality–Virtuality
SLAM	Simultaneous Location and Mapping
SLC	Spatial Livestream Composition
UHD	Ultra High Definition
VoD	Video on Demand
VR	Virtual Reality
WebRTC	Web Real-Time Communication Protocol
WLAN	Wireless Local Area Network
XR	Extended Reality

This page intentionally left blank.

Compositing real-time media streams for groupware panoramic browsing, situation awareness, and enriched user experience

ABSTRACT

According to the Cisco report published in June 2017 [1], by year 2021 82% of the world's IP traffic will be taken over by video streaming services. Within that segment, live video streaming (often referred to as "livestreaming") would represent 13% of the world's video traffic, with mobile users being the fastest growing group of consumers. It is evident that video streaming is growing in popularity, which perhaps can be explained by its relative ease of use: often in video streaming applications users are a single "tap" away from experiencing the media content. Such availability makes livestreaming, or more specifically, mobile livestreaming, ubiquitous in cases where a quick sharing of a remote situation with multiple users in real time is needed, including social and collaborative applications such as Facebook Live, Periscope, Skype, and FaceTime.

However, such convenience comes with a cost: usually a streaming session consists of a low-resolution single-viewpoint video stream, which provides limited information about a remote situation. In scenarios where spatial context is important, for instance when a streaming user constantly changes locations and interacts with different environments, such limitations can negatively affect user experience: it is hard for viewers to understand where a streaming user is located, and for streamers it is equally hard to react when viewers refer to a certain part of the environment surrounding the streamer.

Therefore, this dissertation investigates and proposes a new approach to how mobile live media streams are presented and interacted with in social and collaborative applications. The main research problem can be formulated in the following question: *"Given the spatial data inferred from mobile devices, how can multiuser live video streaming sessions be improved?"* In this case the definition of "improvement" is dispersed and discussed within the scope of the following contributions.

In the first contribution, this thesis approaches the livestream composition problem from the perspectives of mixed reality displays and social livestreaming systems. It proposes new and updates existing taxonomies, upon which the related publications and projects were categorized and qualitatively compared.

Based on the observed literature, in the second contribution this thesis outlines and proposes a new spatial livestream composition (SLC) method, which organizes spatial and media information into two categories: *spatial background*, which provides a generalized spatial context, and *spatialized live media streams* which are embedded within the spatial background, creating a composited mixed reality space in which users can experience multiple live media streams and interact with each other in real time.

The proposed SLC method was implemented and evaluated in several proof-of-concept applications. In the third contribution, this thesis discusses an implementation of SLC method in a collaborative mixed reality application, which was tested in a user study that detected a statistically significant decrease in mental workload and increase in spatial and situational awareness among viewers in comparison with a regular video streaming application. Based on these results, in the fourth contribution this dissertation applies the SLC method to social interactive applications, and investigates whether the proposed method also affects user engagement in a user study that compares the developed application with Periscope (a popular social livestreaming platform). Although the user comments favored the social livestreaming application with SLC method, the statistical results were inconclusive, and therefore a set of recommendations for similar studies was formed.

In the final contribution, this dissertation discusses how SLC method can be combined with other mixed reality applications by proposing a decentralized “metaverse,” a blockchain-based method for organization and sharing of virtual spaces for mixed reality systems. It discusses how this approach can be integrated in both SLC and non-SLC based systems, and their possible future applications.

1

Introduction

1.1 Live media streaming

According to the Cisco report published in June 2017 [1], by year 2021 more than eighty percent of the world's Internet traffic will be taken over by video streaming. In that segment, live video streaming would represent thirteen percent of video traffic, with mobile users being the fastest growing group of video consumers. While these numbers might seem ambitious, even now video streaming takes over more than a half of total Internet traffic, with the second most visited website on the Internet being YouTube [2], a video streaming platform.

Such popularity can be explained by how video streaming technology is implemented. Video streaming, or more broadly, media streaming refers to multimedia content being delivered to

users continuously throughout a streaming session by a content provider. Such implementation, in comparison with having to download a complete media file first, allows almost instantaneous display of streamed media to consumers. Video streaming can be further divided into video on demand (VoD), where streamed media is represented by a complete file stored on a content provider's server, and live video streaming (often referred to as "livestreaming") where the media is captured and displayed to users in real time.

1.2 Motivation and research goals

Livestreaming is also widely adopted as the main form of interaction in social and collaborative applications. Mobile applications such as Instagram Live, Facebook Live, and Skype feature livestreaming as a way of quickly sharing a remote situation with distributed session participants "on the go." However, in such cases the applications are either not taking full advantage of spatial data provided by mobile devices, or use it only in specific cases, such as photospherical video streaming. Therefore, quite often the only source of information for remote viewers is a monoscopic video stream, which, without any spatial context, makes it hard or even impossible for users to understand the remote situation, or navigate around the remote location in cases where a streamer is constantly moving from one location to another and interacts with different environments.

Therefore, this dissertation investigates and proposes a new approach to how mobile live media streams are presented and interacted with in social and collaborative applications. The main research problem can be formulated in the following question: *"Given the spatial data inferred from mobile devices, how can multiuser live video streaming sessions be improved?"* In this case the interpretation of "improvement" is reflected by the following research goals:

1. Would implementing a different multi-viewpoint live media stream composition method decrease users' cognitive workload, and increase spatial and situational awareness in collaborative applications?
2. Would such change also affect user engagement in social livestreaming applications? In

this case the definition of user engagement is comprised of users' opinion on whether they could influence and interact with the remote media streams.

3. How can other applications benefit from the proposed change in the live media stream composition approach?

1.3 Main contributions

Based on the defined research goals, the main contributions of this dissertation can be summarized in the following items:

1. *Updated extended mixed reality taxonomy and proposed categorization of social live media streaming systems.* In the first contribution, this thesis updates the existing and proposes new taxonomies, and uses them for qualitative comparison of relevant works. Based on the observed literature, the approaches to media stream composition problem can be viewed from two different aspects: mixed reality displays, and social livestreaming systems. Using such categorization and comparison, this work outlines a set of desirable features such as fast access to full spatial data, pervasiveness, and multiple forms of user interactions.
2. *Spatial live media stream composition method.* Building on the outlined features, in the second contribution this study proposes a novel spatial live media stream composition (SLC) method. The proposed method combines both spatial and media information, and divides them into two categories: *spatial background*, a generalized information source that provides a spatial context in which real-time media streams and users can be placed; and *spatialized live media streams*, real-time media (audio, video, or other) feeds coming from streaming users, rotationally oriented within the spatial background. Both categories are composited into a single interactive mixed reality space, in which all users can communicate and collaborate together. The spatial information is synchronized, and viewers can observe and interact with streamers' remote environments and their viewpoints in real time.

3. *SLC method for pervasive telepresence and remote collaboration on mobile devices.* The SLC method is then implemented in three proof-of-concept applications. In the third contribution this dissertation investigates the merits of SLC method in remote collaboration context. Named as “StreamSpace,” the proposed collaborative mobile application creates a mixed reality space from a web-based photospherical imagery (spatial background), and places a set of rotationally tracked real-time video streams that represent connected streaming users (spatialized live media streams). Aside from video streams, both viewing and streaming users can interact in mixed reality space through real-time audio and 3D annotations. This application has been tested in an experiment where users, placed in two separate locations, were asked to collaborate together and find an object hidden within one of the locations. In comparison with a regular monoscopic live video streaming system, StreamSpace, and therefore this specific application of SLC method, has shown a statistically significant decrease in cognitive workload, and increase in spatial and situational awareness among viewers.

4. *SLC method for interactive mobile social live video streaming systems.* In the fourth contribution the study investigates whether SLC method could also affect user engagement in social livestreaming systems by building and evaluating a second application called “ReactSpace.” The application is based on top of the media stream composition interface developed in StreamSpace, and it adds spatialized reaction buttons on top of audio and video modes of interactions already available to users. ReactSpace was evaluated in a test, in which users viewed and interacted with remote livestreams both by using Periscope (a popular mobile livestreaming platform) and ReactSpace. Afterwards, the users filled out a questionnaire that measured whether they felt that they could influence a remote video stream using different interaction modalities presented in ReactSpace, and whether they enjoyed using the application in comparison with Periscope. Although according to the user feedback participants preferred using ReactSpace, the statistical results were inconclusive, since a ceiling effect was detected. Therefore, the study discusses the issues in

conducted experiment and outlines a set of suggestions for future similar studies.

5. *Distributed metaverse: extending the SLC method for other mixed reality applications.* Moving forward from the social and collaborative proof-of-concept applications, in the final contribution this dissertation also investigates how SLC method can be reapplied and connected with other existing mixed reality systems. The study proposes an approach that combines spatial data (such as geospatial coordinates), used when creating a mixed reality space in SLC-based applications, with a blockchain-based distributed database. Such combination creates a “metaverse,” a set of virtual spaces attached to real geospatial coordinates that can be shared and reused among other virtual and mixed reality systems. This work discusses the use-cases of such approach both within the context of StreamSpace, and other applications that might not necessarily employ the SLC method.

The presented main contributions were published in a refereed journal and international conferences. The list of publications that comprise the work carried out within the scope of this dissertation is as follows:

- *Spatial Social Media: Towards Collaborative Mixed Reality Telepresence “On The Go”* (Chapters 2 and 3). In Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. Montreal, 2018 [3].
- *StreamSpace: Pervasive Mixed Reality Telepresence for Remote Collaboration on Mobile Devices* (Chapter 2 and 4). In IPSJ Journal of Information Processing, Special issue of “Advances in Collaboration Technologies.” Tokyo, 2018 [4].
- *ReactSpace: Spatial-Aware User Interactions for Collocated Social Live Streaming Experiences* (Chapters 2 and 5). In 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Miyazaki, Japan, 2018 [5].
- *Applying Rotational Tracking and Photospherical Imagery to Immersive Mobile Telepresence and Live Video Streaming Groupware* (Chapter 5). In ACM SIGGRAPH Asia

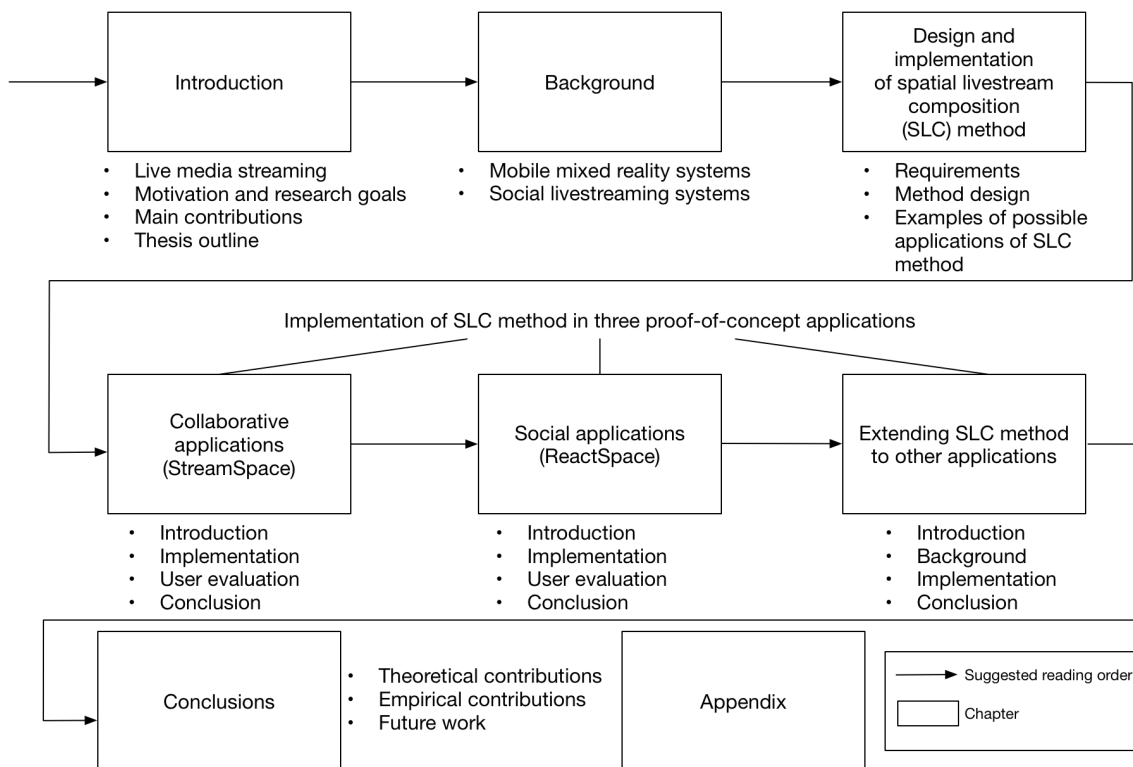


Figure 1.1: Thesis structure and suggested reading order

2017 Mobile Graphics & Interactive Applications. Bangkok, 2017 [6].

- *Distributed metaverse: creating decentralized blockchain-based models for peer-to-peer sharing of virtual spaces for mixed reality applications* (Chapter 6). In the 9th Augmented Human International Conference. Seoul, Republic of Korea, 2018 [7].

1.4 Thesis outline

The condensed outline of this thesis can be seen on Fig. 1.1. The content of the rest of the chapters is as follows:

- **Chapter 2:** Proposal of the updated mixed reality display taxonomy and the new social livestreaming system taxonomy. Overview of the relevant literature and projects, outline of the features that are used in design and implementation in the next chapters.
- **Chapter 3:** Design and implementation of spatial live media stream composition (SLC) method.

- **Chapter 4:** Implementation of SLC method in collaborative applications, investigation of effects of SLC method on user cognitive workload, spatial, and situational awareness in comparison with a regular video streaming application.
- **Chapter 5:** Implementation of SLC method in social media applications and investigation of its effects on user engagement in comparison with Periscope, a popular video streaming platform.
- **Chapter 6:** Extension of developed SLC method for unified archiving, mapping, and sharing of virtual spaces in social and collaborative mixed reality applications.
- **Chapter 7:** Summary of the main contributions achieved within the scope of this work, discussion of possible impacts and future iterations of the developed method and its proof-of-concept applications.

This page intentionally left blank.

2

Background

There are multiple studies on improving livestream viewing and interaction experience through introduction of different viewing modes and enriched spatial information about remote locations. The approaches can be generally divided into two, often connected, categories: mobile mixed reality displays, and social interactive systems. This chapter updates old and proposes new taxonomies for both approaches, based on which the relevant publications and projects are qualitatively compared. As a result of such comparison, this study outlines desirable features and recommendations that are used in design and implementation of SLC method and its proof-of-concept applications.

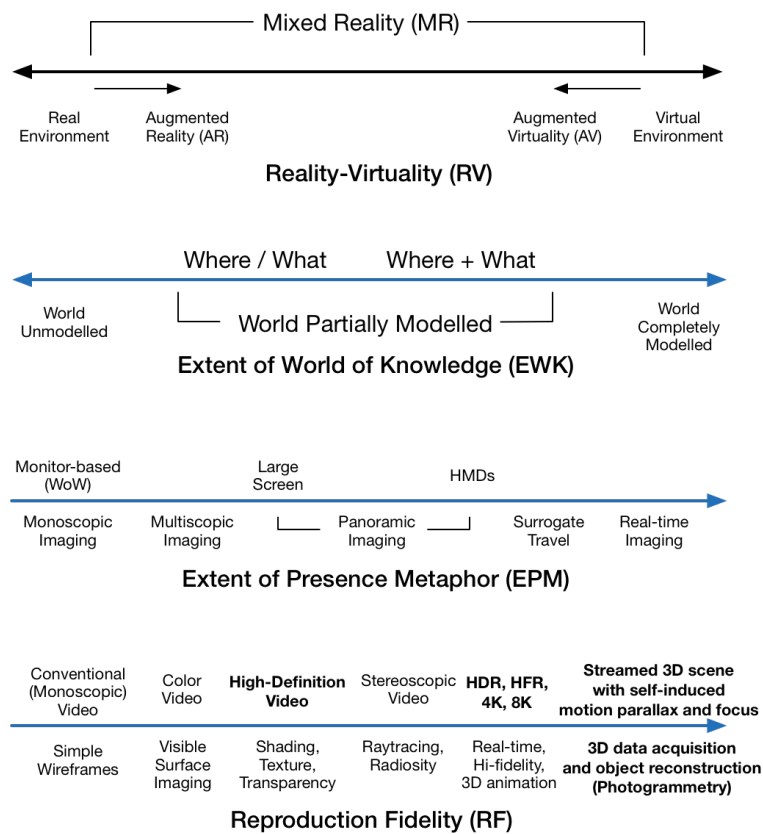


Figure 2.1: RV continuum and extended mixed reality taxonomy, as presented in [8] and [9]

2.1 Mobile mixed reality displays

2.1.1 Classification of mixed reality displays

Mixed reality was first introduced by Paul Milgram and Fumio Kishino in [10], represented in the form of Reality–Virtuality (RV) continuum (Fig. 2.1, top). The RV continuum placed mixed reality applications on a spectrum between real and virtual environments, and classified mixed reality experiences as augmented reality (AR) or augmented virtuality (AV).

As discussed by Mark Billinghurst [8], the RV continuum was further expanded by Milgram et al. [9] to provide a more detailed classification of mixed reality displays. The extended mixed reality display taxonomy includes three continua (Fig. 2.1, middle and 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” entails complete understanding of real world.
- *Extent of Presence Metaphor (EPM)*, i.e., the level of user immersion in a scene, is described on a spectrum between monoscopic imaging (no immersion) and real-time imaging (full immersion). It should be also noted that the use of “presence” here is different from the more recent and now broadly accepted interpretation as subjective impression [11], 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, to 3D high-definition reproductions as the highest. Since the initial definition of this continuum was introduced over twenty years ago, in this dissertation it was adjusted to reflect the recent changes in mixed reality display technology. Namely, the adjustments (in bold font) are: the high dynamic-range imaging (HDR), high frame rate (HFR), 4K and 8K video standards, introduced photogrammetry and high-definition 3D scene streaming. Also the high-definition video was changed to appear earlier in the spectrum due to it being more common nowadays.

Using these spectra it is possible to estimate and compare the quality of real world capture, immersion, and reproduction among different mixed reality displays.

It should be also noted that at the moment of writing several new terms were observed in description and classification of mixed reality systems: extended reality (often referred to as XR) and diminished reality. Extended reality is an umbrella term that includes both mixed reality and virtual reality interactions [12], whereas diminished reality represents a special case of augmented reality that digitally “removes” objects from real world instead of adding or complementing it [13]. Although this dissertation does not specifically interact with XR or diminished

reality systems, it is important to include this distinction in order to avoid any possible confusions that might arise from mixed reality research terminology.

2.1.2 Related virtual and mixed reality displays

One of the earliest virtual reality displays that used photospherical imagery was described by Hirose in “Virtual Dome” and its extending revisions [14]. The system presented a set of images captured by a rotated camera, arranged in a sphere (bottom end of EWK, middle of EPM), which could be viewed through a Head-Mounted Display (“Stereoscopic Video” of RF). Virtual Dome extensions included the introduction of motion parallax and GPS tracking.

The advantage of mixed reality systems for remote collaboration was affirmed by Billinghurst and Kato in the “Collaborative Augmented Reality” survey [15], which reported improved sense of presence and situational awareness compared to regular audio- and videoconferencing solutions. The authors also acknowledged the necessity of handheld displays for wider adoption of mixed reality techniques in collaborative scenarios.

Cohen et al. [16] also noted the limitations of tethered Head-Mounted Displays (HMD) in such mixed reality systems as Virtual Dome and developed a motion platform for navigation around panoramic spaces. It used a regular laptop screen (“Colour Video” of RF) for panoramic display (bottom end of EWK, middle of EPM) and a rotating chair for navigation around a panoramic space. 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. [17]. 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 latter also means that users’ viewpoints were bound together, and a viewer cannot freely explore the mixed reality space without being attached to the streaming user’s viewpoint.

Similarly, the overlaid video annotation aspect was also investigated in Skype for HoloLens

in a study by Chen et al. [18], and for mobile and desktop platforms in a study by Nuernberger et al. [19]. Although both studies featured monoscopic high-definition video similarly to Chili, their tracking approach partially modeled the world around them which puts both studies in the middle of EWK spectrum.

Free navigation around panoramas was presented in Bing Maps by Microsoft [20] where a user could overlay a real-time video stream over a photospherical panorama of a location, but the system was limited only to photospherical imagery provided by Microsoft and at the time it did not fully support mobile devices. On mobile platform it was explored by Billingham et al. in “Social Panoramas” [21], 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 further demonstrated in systems by Gauglitz et al. [22] and Kasahara et al. in “JackIn” [23]. JackIn used 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. was using a handheld mobile tablet for image capture. Both solutions, however, required large motion parallax for stable tracking and creating a complete panoramic image.

This issue was discussed by Nagai et al. in LiveSphere [24], 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” [25] used a HMD with embedded eye tracker to switch between blended stereoscopic video streams originating from cameras on robots’ heads. Both studies fall at the bottom of EWK, at “Stereoscopic Video” point of RF, and “Surrogate Travel” of EPM.

Such solutions, however, still required computational power that was relatively high for mobile devices, which was addressed in PanoVC by Müller et al. [26]. Instead of a set of head-mounted cameras, PanoVC used a mobile phone to create a continuously updated cylindrical panorama (bottom of EWK, “High Definition Video” point of RF, and middle of EPM), in

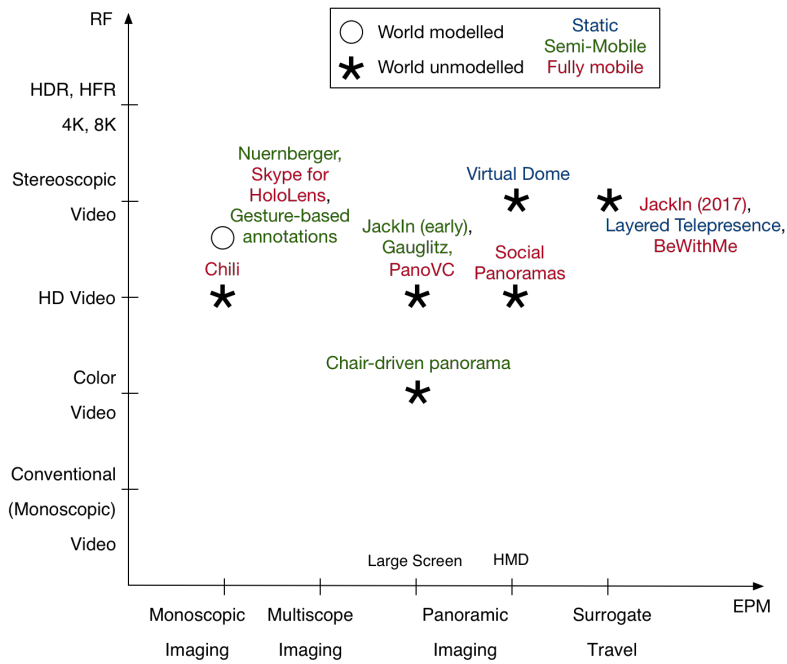


Figure 2.2: The distribution of mixed reality systems based on extended taxonomy. Since most of the presented examples are at the bottom end of EWK, the solutions are sorted only along RF and EPM axes. The several studies at the middle of EWK spectrum are shown as a circle. For the sake of convenience, all studies were marked in different colors depending on the range of mobility available to users.

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 [27] used 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’s surroundings, but the resulting experience was still limited to only two users at once. Similar approach was also implemented in the updated version of JackIn project by Kasahara et al. [28], moving the display towards the “Stereoscopic Video” of RF, and “Surrogate Travel” of EPM.

2.1.3 Conclusion

As a result of mixed reality display literature overview, this dissertation outlines a series of key features and limitations in media stream composition methods, such features include:

- High-immersion approaches (rightmost side of EPM) tend to provide near-instantaneous access to full spatial information (e.g., realtime spherical video streams),

- however it is done at a cost of decreased mobility (“Layered Telepresence”) or limited interactivity (in JackIn 2017 and BeWithMe, the viewers can experience only one media stream at a time).
- Stitching-based methods (e.g., PanoVC, “Social Panoramas,” early version of JackIn) are inefficient since they require either a large motion parallax or high computational power to represent a remote space. Furthermore, nowadays stitching is already provided in either software or hardware used for mixed reality displays.
- Mobile systems ensure low computational power and higher pervasiveness (availability of the system to a wide range of users without requiring specific or high-end hardware) by operating with basic spatial information, such as rotational tracking (Chili, PanoVC) and edge detection (Chili, Nuernberger, Gauglitz).

2.2 Social livestreaming systems

2.2.1 Social livestreaming system taxonomy and relevant works

Another approach to improvement of live media stream composition problem comes from the perspective of social livestreaming systems. Indeed, the term “livestreaming” in the modern context is mostly used in such services as Twitch, YouTube, or Facebook Live, that add a social element to real-time video stream viewing. Furthermore, since the elements presented in mixed reality displays are being implemented in the modern mobile livestreaming applications (e.g., “AR filters” feature in Snapchat), it is necessary to establish a taxonomy that combines both mixed reality displays and social interactive systems in a set of features upon which the relevant works can be qualitatively compared.

Therefore, this dissertation proposes a taxonomy (Fig. 2.3) based on the extended mixed reality classification discussed above. Similarly, this taxonomy builds on three continua: Extent of Social Interaction Space (ESIS), Extent of User Communication (EUC) and Reproduction Fidelity (RF).

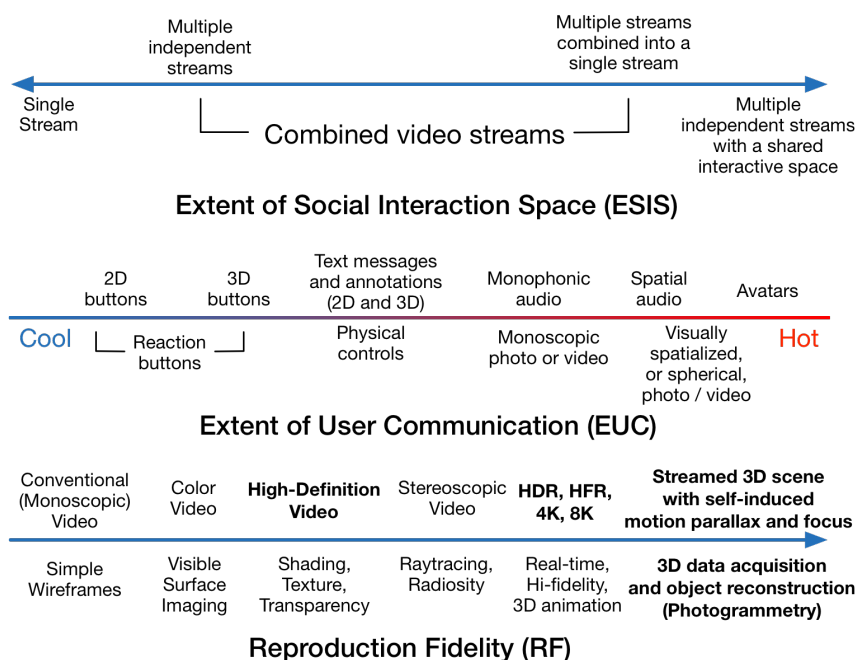


Figure 2.3: Proposed livestreaming system taxonomy

2.2.2 Extent of Social Interaction Space

The first proposed continuum measures the extent of the social interaction space available to the livestream viewers. For example, such livestreaming application as Periscope allows viewing only one stream at a time and thus minimizes the social interaction space to only a single stream. Stream aggregating services like multitch [29] or kbmod [30] allow combining multiple streams visually, however the interaction is still restricted only to each individual stream. Similarly, Google Hangouts allows streaming live group chats to YouTube [31], combining multiple streams into a single video stream, which allows interacting with all video streams at the same time, but not with each of them independently. Finally, there is a new emerging approach to live stream interaction presented in Rivulet by Hamilton et al. [32], which aggregates multiple streams and allows users to interact both within a single stream and across multiple streams simultaneously.

2.2.3 Extent of User Communication

The second continuum classifies different means of communication between viewers and streamers. It is based on McLuhan's "hot" and "cool" media concept [33] and was used for classification of user interactions in [32] and [34]. The hot and cool concept proposes a spectrum, where different media is classified based on amount of mental workload. Cool media requires users to apply a considerable amount of mental effort to understand it, unlike hot media, that is more understandable and thus easily "digestible." In McLuhan's example, the books would be classified as cool media, while movies would be on the hot side of the spectrum.

Similarly, this concept is employed for categorization of available modes of user interaction observed in similar studies:

- *Reaction buttons*, which are also sometimes referred to as "hearts" and "like buttons," are implemented in almost any social live streaming service, including Periscope, Instagram and YouTube. Reaction buttons are one of the "coolest" forms of media, because they do not provide a complete representation of a person's reaction. Although in most services such buttons are represented by 2D icons, in the recent project by Facebook [35], the reaction buttons can be placed around the virtual 3D space and thus could theoretically provide more context for the participants.
- *Text messages and annotations / Physical controls*. Similarly to reaction buttons, almost all services mentioned above provide support for text messaging. Some researchers have even extended the messaging capabilities to include specific commands. For example, Yonezawa and Tokuda [36] integrated a set of commands that allows users changing the camera orientation and light conditions in the live video stream, and Nassani et al. tested both 2D and 3D text annotations for social applications [37].
- *Monophonic audio and monoscopic photo or video* is perhaps the most common form of hot media streaming and is available on all live streaming services.

- *Spatial audio and visually spatialized, or spherical, photo / video* is supported on services like Periscope, YouTube and Facebook Live in the form of 360° video streaming, providing a spherical overview of the scene, and thus becoming a hotter form of media in comparison with regular monoscopic video streaming. Another example would be the previously discussed project LiveSphere by Nagai et al. [24] that used a set of head mounted cameras to present a “first-person” spherical live video feed. Similarly, Kasahara et al. in JackIn [23] streamed from a single head-mounted monoscopic camera, but through simultaneous localization and mapping (SLAM) algorithm the authors were able to create a constantly updating spherical panorama from different video frames and spatially arrange the live video stream inside of it.
- *Avatars* is an emerging form of communication in live streamed media. For instance in Facebook Spaces [35], users are represented as virtual avatars that can navigate around the scene and interact with everybody through other cooler forms of media. Most likely the future forms of live streamed media would also include different forms of haptic interactions that would fall in this category.

2.2.4 Reproduction Fidelity

As in the extended mixed reality taxonomy, this continuum describes the level of fidelity of live streamed media. Similarly to the updated extended mixed reality taxonomy, the scale reflects the recent changes in live streamed media displays. The continuum starts at standard definition color video which is supported by systems like Periscope [38] or Rivulet [32], and moves towards high-definition, stereoscopic (HMD-based) and spherical video. Finally, this spectrum includes volumetric video showcased in Facebook Surround 360 [39] and by Zhou in Visibit [40].

2.2.5 Conclusion

Based on the observed social livestreaming systems, this study outlines key features and limitations of the current state-of-the-art solutions:

- User engagement (active participation in livestreaming sessions) can be increased via high ESIS cross-stream interactivity [32, 34, 41], therefore a user's ability to freely explore multiple real-time video streams and virtual interactive spaces is desirable.
- Users might favor 3D annotations as a form of interaction in social livestreaming applications [37], however it is unclear whether the presence of other hotter forms of media also affects user interest in livestreaming sessions.
- The most advanced social livestreaming system [35] includes the full spectrum of EUC modalities, however it is unclear whether it is beneficial for users. Furthermore, [41] indicates that increasing available interactive modalities might overwhelm users and can negatively affect their cognitive workload.

This page intentionally left blank.

3

Design and implementation of spatial livestream composition (SLC) method

3.1 Requirements

Based on the observed literature, its key features and limitations, this work outlines a set of requirements that define the developed media stream composition method:

- *Multi-viewpoint or multi-stream viewing and interaction.* As demonstrated in [32, 41], the media stream composition solutions that feature multiple real-time streams within a single interface, sharing combined and individual interactive spaces, are potentially beneficial in supporting interest among viewers interacting with remote video streams.

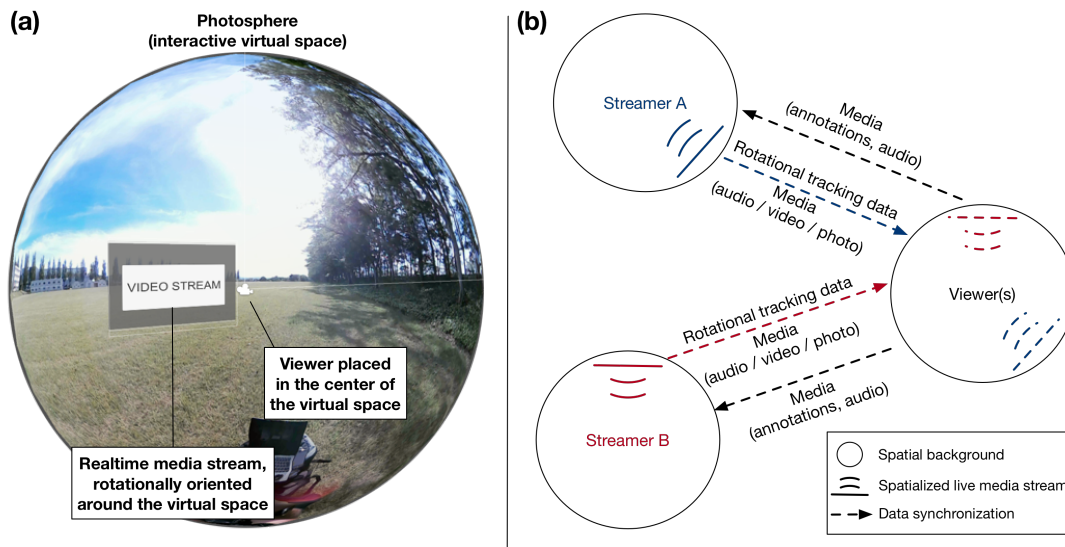


Figure 3.1: (a) Example of a virtual space created in an application via SLC method. (b) Data synchronization flow among separate mixed reality spaces within a single streaming session in a SLC-based application.

- *Fast access to full spatial context.* As demonstrated in [35] and discussed in [26], simplifying access to full spatial context can improve the sense of social presence and co-presence.
- *Untethered viewpoints.* As demonstrated and discussed in Chapter 2, higher immersion systems tend to allow users to freely explore the virtual environment.
- *Support for different modes of user interaction.* Finally, the proposed method should also support extensions with different interaction modalities, including, but not limiting to, text messages and realtime annotations, monoscopic and stereoscopic photo, audio, and video.

3.2 Method design

Building on the established requirements this study proposes the spatial live media stream composition (SLC) method that employs mixed reality, a combination of virtual spaces and real-world live media feeds, to create a pervasive interactive space that provides fast access to rich spatial information and different modes of interaction.

Similarly to common livestreaming applications, the proposed method assumes two types of users: *streamers* who are physically present in a remote location and who start a streaming session, and *viewers* who can be located anywhere, but are virtually transported to a remote location. Although such definitions are more appropriate within a video streaming context, these groups of users could be also referred to as “source” (streamer), and “sink” (viewer). However, for the sake of convenience, the rest of this thesis would use “viewer” and “streamer” to disambiguate two roles of users within the proposed method.

The SLC method organizes both spatial and media information into two following categories (Fig. 3.1, b).

3.2.1 Spatial background

Spatial background represents a generalized information source that provides a spatial context in which users can be placed. In such case it could be a spherical virtual space, or a spatial audio display.

In applications that employ SLC method it is assumed that a collaborative session starts with a streamer creating a virtual space (e.g., from a photospherical image in case with photo or video streaming), in which all users are being placed throughout the session (Fig. 3.1, a). The created space allows users to almost immediately experience spatial context and interact with remote environment.

3.2.2 Spatialized live media streams

Spatialized live media streams represent live media (audio, video, photo, or text) feed coming from a streaming user, rotationally oriented within the spatial background. In case with video streaming applications, a spatialized live media stream would be represented by a monoscopic video rectangle, rotationally oriented within the virtual space.

In the proof-of-concept applications that implement SLC method, the feeds are spatialized through rotational tracking, however the media streams can also be oriented by other means employed in modern mobile systems, such as visual-inertial odometry [42], or any other form

of external tracking.

Furthermore, in SLC method all media streams are always synchronized. Therefore, viewers can perceive streamers' live viewpoints (in case with photo or video streaming), and all users can interact with each other in real time.

3.2.3 User interactions

SLC method was designed to include different modes of interaction, both spatialized and two-dimensional, as the availability of virtual space provides a multimodal way of communication for connected users. In the implemented proof-of-concept applications, users can communicate through regular and spatialized audio and video streams, and real-time text and annotations.

3.3 Examples of possible applications of SLC method

However, the possible ways of application of SLC method extend beyond the proof-of-concept applications presented in the next chapters. For example, considering the tree commonly used types of modalities in mixed reality applications: visual (photo / video), aural (audio), and haptics (touch, vibration), the SLC method can fit into the following use-cases:

- *Real-time viewpoint sharing.* Spatial background can be represented both by static and dynamic media. In such case, as demonstrated in [43], the virtual space can be comprised of a live photospherical video stream, and users' viewpoints can be represented by real-time spatially oriented rectangle-shaped markers (spatialized live media streams).
- *"Ventriloquism."* In SLC-based systems, spatialized media streams do not necessarily have to display an objective representation of a remote situation. Instead, some applications can employ a technique similar to ventriloquism, where a real-world media source is being intentionally displaced in the composited virtual space in order to compliment the user experience. An example of such technique can be observed in [44] where virtual space is composited of a spatial background, represented by a binaural audio stream, and

a haptic sphere placed in user's hands, vibrations of which are synchronized with spatial background. In such case the sphere compliments user experience through directionalized haptic feedback, while not necessarily being an objective representation of the real space from which the virtual one was generated.

- *“Portals.”* When a spatial background is created, an application can assign a certain set of coordinates (e.g., a set of real-world geospatial coordinates) to a virtual space in order to distinguish it from other virtual spaces. In such case, SLC-based systems can use spatialized media streams to represent a “portal” through which users can observe or travel to other virtual spaces (e.g., by sorting virtual spaces against each other based on the assigned coordinates). An example of such approach was presented in [45], where authors suggested to model mixed reality spaces after physical rooms, and distinguish available modes of interaction based on the types of connections between spaces. For instance, two spaces could be connected by a “door,” that allows users to both enter the other space and observe it if the door is open; or two spaces could be connected by a “window,” which allows users to observe the other space, but not enter it.

In the next chapters this thesis gives an in-depth examples of some of the use-cases presented in this section and discusses their benefits for user experience.

This page intentionally left blank.

4

Collaborative applications (StreamSpace)

4.1 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. This proof-of-concept application investigates how such technology in combination with SLC method can be applied to remote collaboration, creating a quick way of sharing snapshots of real environments so that distributed users can work together.

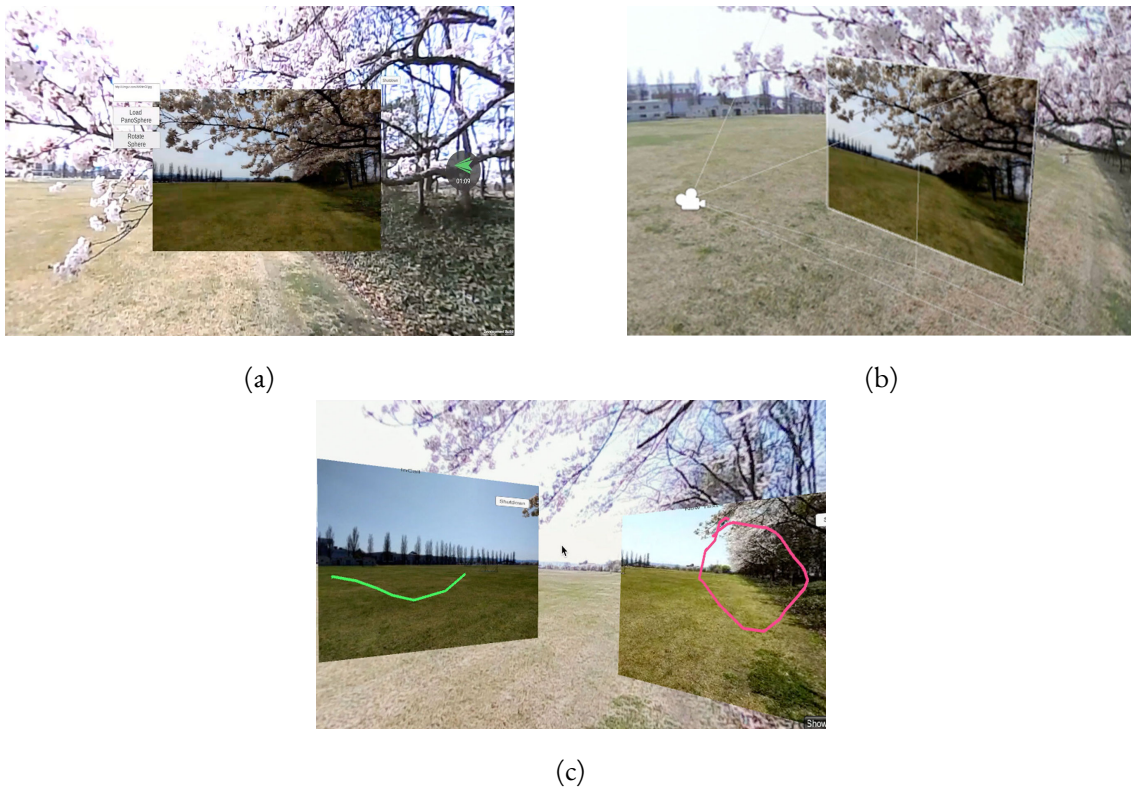


Figure 4.1: (a) Streaming mode user interface, (b) Viewing mode scene overview, (c) Live snapshot of a collaborative session with multiple streamers and a viewer

“StreamSpace” is a proof-of-concept system that uses mobile video streaming and mixed reality spaces for remote collaboration. It takes advantage of photospherical imagery (captured by a user or downloaded from elsewhere) to create a spherical background, i.e., a snapshot of a real location, in 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-based 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, this application works on multiple Android devices and does not

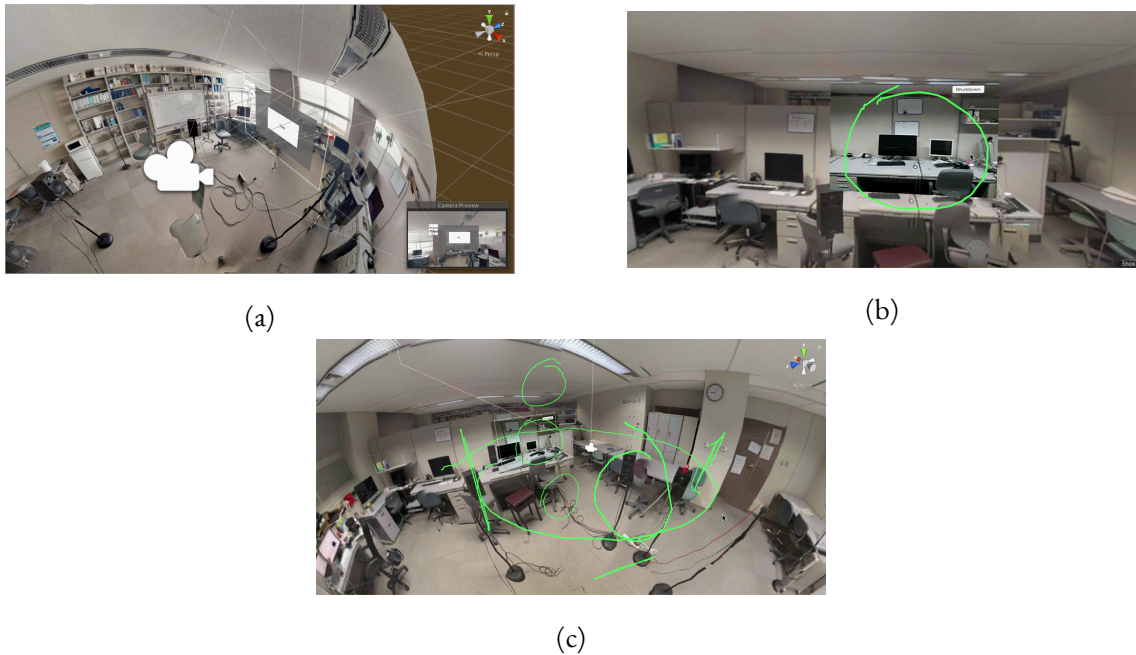


Figure 4.2: Examples of 3D drawings. (a) Close-up of a Unity scene, the camera in the center of the scene representing the viewer's position. (b) Drawing by a streamer. (c) Scene overview after a typical collaborative drawing session.

need excessive computational power for a collaborative session. Such features help StreamSpace to be more “pervasive,” i.e. applicable to a wide range of various users in different scenarios, advancing the state of mobile collaboration groupware.

4.2 Implementation

Based on the proposed model for live media stream composition, StreamSpace is designed to support multiple users, provide panoramic background with realtime updates, and be able to adapt to both fixed and mobile scenarios. Furthermore, compared with similar solutions using the extended taxonomy (Fig. 2.2), StreamSpace is among the most immersive and high fidelity mixed reality displays, as it supports surrogate travel, and can work in stereoscopic video mode via Google Cardboard SDK.

4.2.1 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 com-

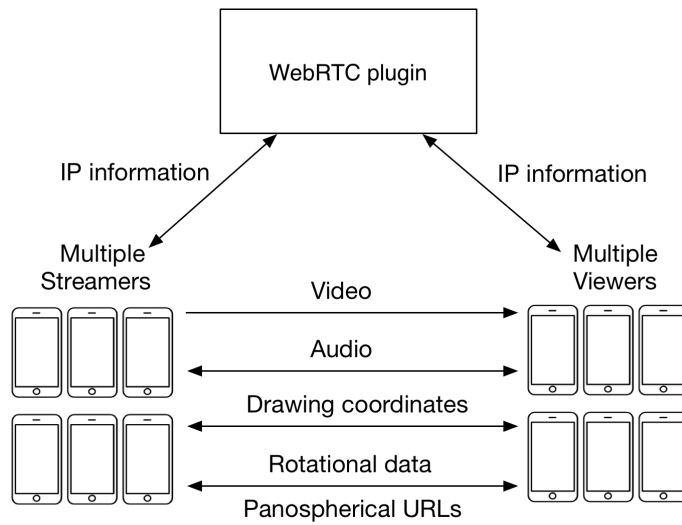


Figure 4.3: StreamSpace connections and dataflow diagram

patible 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 realtime 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, the recently announced Google VPS (Visual Positioning Service) [46], 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. 4.1): 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 rotational state 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 elsewhere) for all users, whereas viewers can only passively receive new panoramic images.

The user interaction features not only audio and video streaming, but also drawing. Since the virtual space is a 3D scene (built in Unity), the 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. 4.2).

Networking is handled through the Web Real-time Communication (WebRTC) protocol [47]. WebRTC is used due to its ability to establish peer-to-peer connections among remote users via network address translation (NAT) traversal technologies, i.e., connecting users without prior knowledge of each others' IP addresses. Furthermore, WebRTC protocol design ensures low-latency connection, supports multiple simultaneous users, and works both over mobile networks (3G, 4G, LTE) and wireless LAN (WLAN).

The WebRTC implementation is provided through the mobile version of the "WebRTC Videochat" plugin for Unity [48]. It sends and receives audio from all connected users, sends streamers' video feeds to viewers in a native resolution, and handles the synchronization of drawing coordinates, users rotational data, links to panoramic images, and photosphere's rotational offset (Fig. 4.3).

Finally, even though the system was tested with panoramas captured through Insta360 Air camera [49] 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.

4.3 User evaluation

4.3.1 Experiment design

To test the feasibility of SLC-based approach implemented in StreamSpace, a user evaluation was conducted. Its experimental hypotheses conject that, compared with regular video-conferencing systems, StreamSpace imposes less cognitive workload on users and increases their spatial and situational awareness.

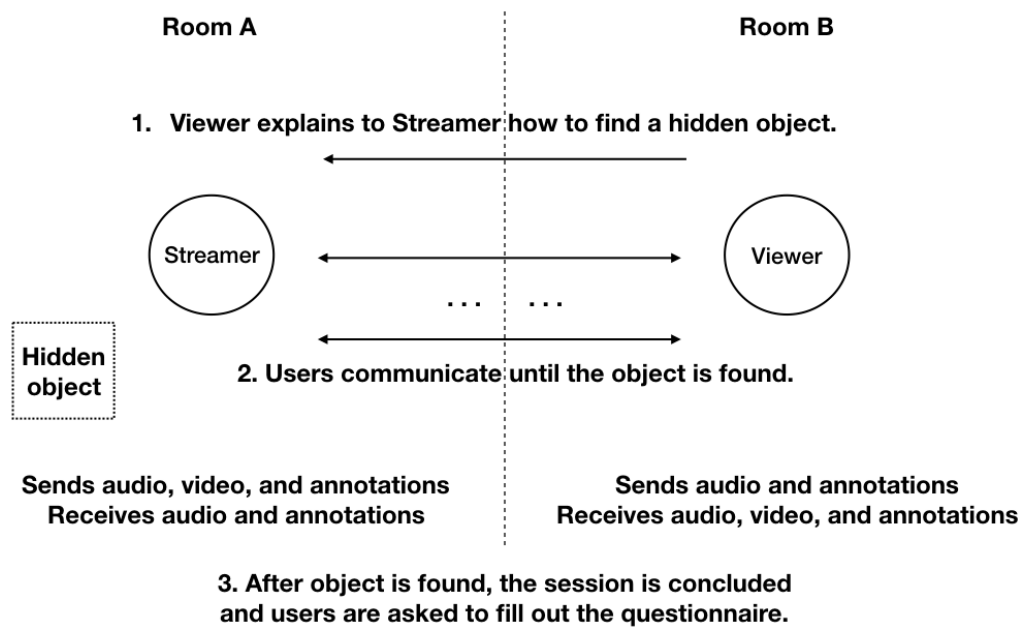


Figure 4.4: Experiment setup

Since differences in the user interface between the proposed application and commercially available solutions might confuse participants, a separate regular videoconferencing mode within StreamSpace was developed, which was called “flat,” since it does not use a mixed reality space (and StreamSpace is abbreviated 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).

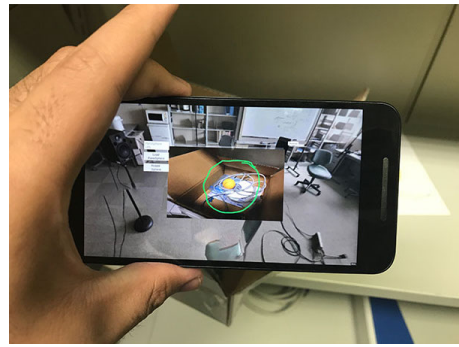
Furthermore the capture of photospherical imagery was intentionally excluded from the experiment, because StreamSpace was designed to support panoramas captured through third-party applications. Before each trial, a panorama of the room in which the experiment was conducted was uploaded, captured with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

The experiment itself had the following steps (a condensed illustration of which can be seen on Fig. 4.4):

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.



(a)



(b)

Figure 4.5: User evaluation scenario for StreamSpace. (a) Streamer walking around the location, (b) Streamer successfully finding the ball, highlighting it, and concluding an experiment’s session

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 a uniform complexity of performed tasks.
5. The Viewer received an explanation about where the target was hidden, and he or she needed to explain it to the Streamer using the 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 items on spatial understanding introduced by Kasahara et al. [23] 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, the last two questions regarding cognitive workload were replaced with ques-

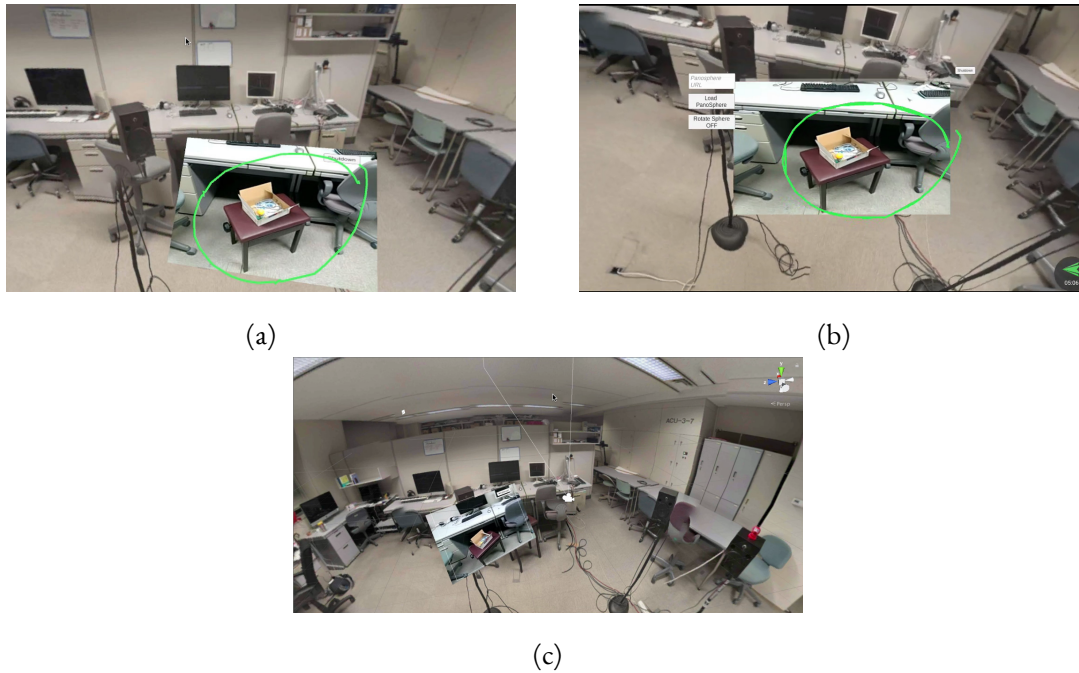


Figure 4.6: On-screen view example of a testing scenario for StreamSpace. (a) Viewer, (b) streamer, (c) and exocentric view of the scene

tions from the unweighted NASA Task Load Index test [50], 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 a faster execution of the experiment while still providing valid results that are highly correlated with traditional TLX scores [51].

4.3.2 Setup

The experiments were conducted on campus at both the University of Aizu (UoA) and Hochschule Düsseldorf: University of Applied Sciences (HSD). Forty participants (or twenty pairs) in total were recruited, 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 respective institutions (UoA and HSD) 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, however 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 fourteen sessions at the UoA and six at HSD (three each in HSD-A and HSD-B) were conducted.

4.3.3 Analysis

Table 4.1: Z- and p-value table scores, all $p < 0.05$ are in highlighted in **bold font**

Name	Z-value	p-value
RTLX-Viewer	2.0121	0.0442
RTLX-Streamer	-0.1959	0.8446
Q1-Viewer	-2.7424	0.0061
Q1-Streamer	-0.6653	0.5058
Q2-Viewer	-2.3749	0.0175
Q2-Streamer	-1.4747	0.1403
Elapsed time	1.1014	0.2707

Before conducting the experiment, a sensitivity analysis was performed in order to determine minimal effect size that this experiment could possibly detect. The analysis included the following factors:

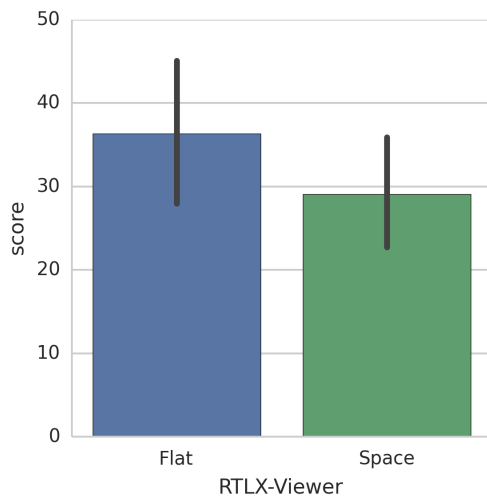


Figure 4.7: RTLX scores for viewers (all error bars in the following figures in chapters 4 and 5 represent 95% confidence intervals)

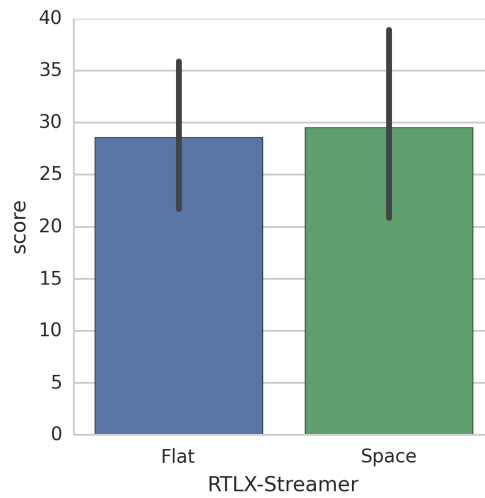


Figure 4.8: RTLX scores for streamers

- *Limited number of participants:* given the physical constraints of experimental setup (it needed to be conducted in pair, in specific locations), and a small pool of available participants, the total sample size was limited to 20 pairs (40 participants).
- *Ordinality of collected data:* due to the fact that the main tools for estimating the user experience were comprised of Likert items and a RTLX questionnaire, Wilcoxon Signed Rank Test was used to find whether there is a statistically significant improvement in using StreamSpace over a regular videoconferencing application.

Thus based on the sample size of 20, $\alpha = 0.05$, and power = 0.8 (as suggested in [52]), the minimal effect size that this experiment could possibly detect was $d_z = 0.59$, which according to [52] can be considered as large.

In 12 pairs out of 20, RTLX Viewer scores were lower for the space than for the flat mode, which is also reflected in the RTLX scores (Fig. 4.7). This trend is confirmed by Wilcoxon Signed Rank Test results with $p < 0.05$ for Viewers (for precise Z- and p-values, please refer to Table 4.1). 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 (Fig. 4.8).

For spatial and situational awareness (Fig. 4.9 to Fig. 4.12) a strong improvement was ob-

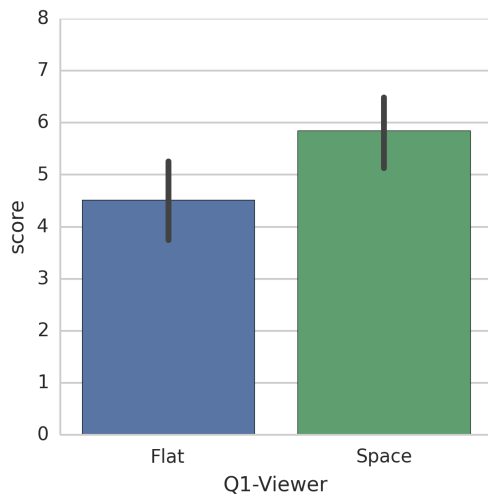


Figure 4.9: Spatial awareness scores for viewers

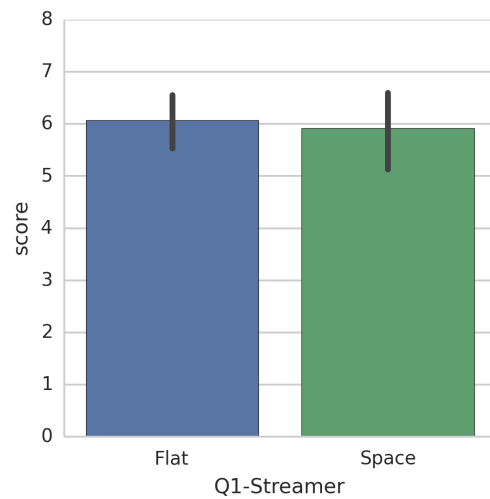


Figure 4.10: Spatial awareness scores for streamers

served in scores for Viewers with $p < 0.05$ for Q1 and Q2, however the results for Streamers were not statistically significant ($p > 0.05$ for Q1 and Q2). The time scores (Fig. 4.13) showed a reduction in elapsed time in space mode as compared to flat, but the results were not statistically significant ($p > 0.05$).

Some other interesting effects of environment on user performance were also observed. Although the sample size of UoA participants was twice as large as that of HSD (14 and 6 respectively), a similar mean RTLX score was noted, as well as Q1 and Q2 scores between rooms at UoA and HSD-A. This consistency could be explained by the fact that HSD-A and UoA environments were of similar size and layout (although in UoA the lack of visual confirmation was guaranteed, while in HSD-A it was not).

HSD-B test, however, was held 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 increase in elapsed time. Such differences in HSD-B results can be explained by location, unfamiliarity with which disoriented test participants.

The conditions in HSD-B are perhaps the closest to how StreamSpace is expected to be used in real life scenarios, and therefore more experiments in similar environments are recommended.

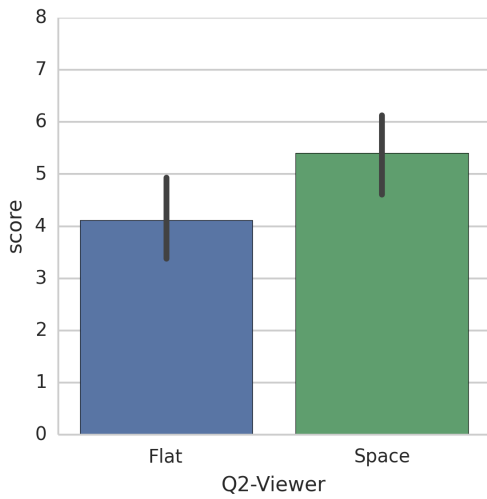


Figure 4.11: Situational awareness scores for viewers

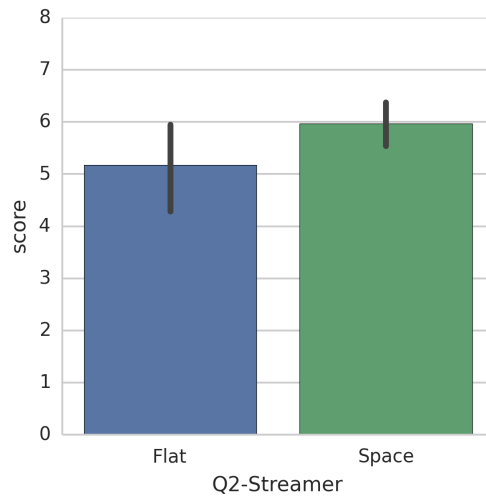


Figure 4.12: Situational awareness scores for streamers

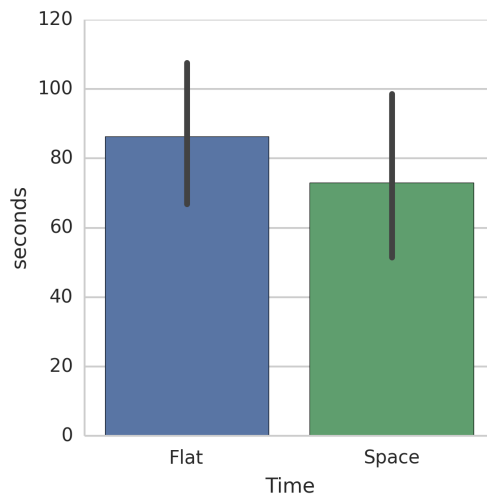


Figure 4.13: Elapsed time

4.4 Conclusion

StreamSpace 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, StreamSpace, and thus SLC method, does improve spatial and situational awareness, and reduces cognitive workload.

For streamers, however, it 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 confusion, as it seems that although they did not encounter any issues, users took more time to adjust to the interface and unknown environment.

For future revisions it would be interesting to replace the streaming interface by a full-screen live video feed with embedded three-dimensional drawings, as in “Chili” [17], or studies by Gauglitz et al. [22] and Chen et al. [18].

Users seemed to like the introduced photospherical aspect of the mixed reality interaction, as they could navigate around a panorama without being tethered to a streamer’s viewpoint, which indicates that having the application to operate at a higher level of the EPM spectrum could indeed improve collaborative aspects. Such 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 [28].

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. Users also requested to add such features as a haptic feedback and an HMD integration to improve the immersion.

In the future, StreamSpace can also include markerless tracking through such systems as Kudan [53], and Google Visual Positioning Service [46], or HMDs like Google Daydream View [54] and Microsoft HoloLens [55]. 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 the environment and send a three-dimensional map of real space to viewers, who could “touch” its surfaces through haptic controls, as demonstrated in different human-computer interaction studies, such as, for example, by Lopes et al. in [56]. Furthermore, since the panoramic background can feature different synaesthetic displays such as IR heat-maps, the haptic interaction could be extended to feature thermoception.

Inclusion of advanced tracking and mapping in StreamSpace could also help in address-

ing the issue of field-of-view (FoV) matching. Currently the 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 devices’ cameras and screens, and the wide variety of photospherical images available on the web. Since a recent study indicates that FoV differences have a strong effect on collaboration in mixed reality environments [57], an improvement of FoV management is necessary in future iterations. One of the possible solutions for that could be using markerless tracking systems such as Apple ARKit [58] or Android ARCore [42] to determine user displacement in a scene, or alternatively, implement a machine learning approach that should automatically readjust either the photosphere or a video feed to create a matching image.

Even though the words “streamer” and “viewer” were used to distinguish the two peer–peer modes driven by 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 [59], to project spatial soundscape recordings, but even a simple rendering such as lateral intensity panning could be used. 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 sources’ and sinks’ virtual standpoints, such aural separation could flatter groupware sessions and enhance the situation awareness.

Another interesting extension would be the implementation of stereoscopic video streaming. StreamSpace 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.

5

Social applications (ReactSpace)

5.1 Introduction

Live video streaming is becoming an increasingly popular medium for social interactions. Through such applications like Twitch, YouTube, Periscope, Snapchat, Instagram, and Facebook Live users can quickly share their video feed with multiple remote users online. Video feeds can cover a wide range of activities: from video game streaming to live coverage of different events by streamers located onsite. The latter kind of live streaming has become especially popular on mobile devices due to their social availability and relative ease of use: most of the time mobile streaming applications do not require prior setup or special equipment to start streaming.

At the same time, however, the simplicity of mobile live streaming is also one of its major drawbacks. First, the applications listed above only support viewing of a one livestream at a time, and for instance in case if users would want to watch the same event from different viewpoints simultaneously, they would have to manually close one video stream and open another. Secondly, the applications do not take full advantage of mobile devices' spatial data. For example, such applications as Periscope and Facebook Live use mobile rotational tracking for navigation around spherical video streams, but the rest of the interactions, such as text messages and "Like" buttons, remain two-dimensional without any attachment to specific locations around the spherical scene. Finally there is a noticeable interest among users for more complex, three-dimensional modes of interactions with captured video streams. For instance, Instagram has introduced a set of icons, also known as "stickers," which can be overlaid on top of captured photos or videos. Since such icons can be moved, rotated and resized, some users have been applying them as pseudo-3D markers, arranged around the captured scene. Similar functionality was implemented in Facebook Spaces for tethered head-mounted displays (HMDs), where users could put "Likes" around the virtual space and annotate it using rotationally tracked controllers [35].

Based on such observations, this study proposes a SLC-based system that creates an interactive livestreaming experience for multiple users while exploiting mobile devices' rotational sensing. Named as "ReactSpace," a livestreaming system uses a spherical image (captured by streaming user or downloaded from elsewhere) to support a shared virtual space for multiple collocated live video streams. ReactSpace presents several novel modes of live stream viewing and interaction:

- Viewers can watch multiple simultaneous streams arranged in the form of video billboards, spatially oriented rectangles texture-mapped with live streamed content that are always facing the viewer.
- Each video billboard is oriented through synchronizing mobile devices' rotational tracking and represents streamer's current viewpoint.

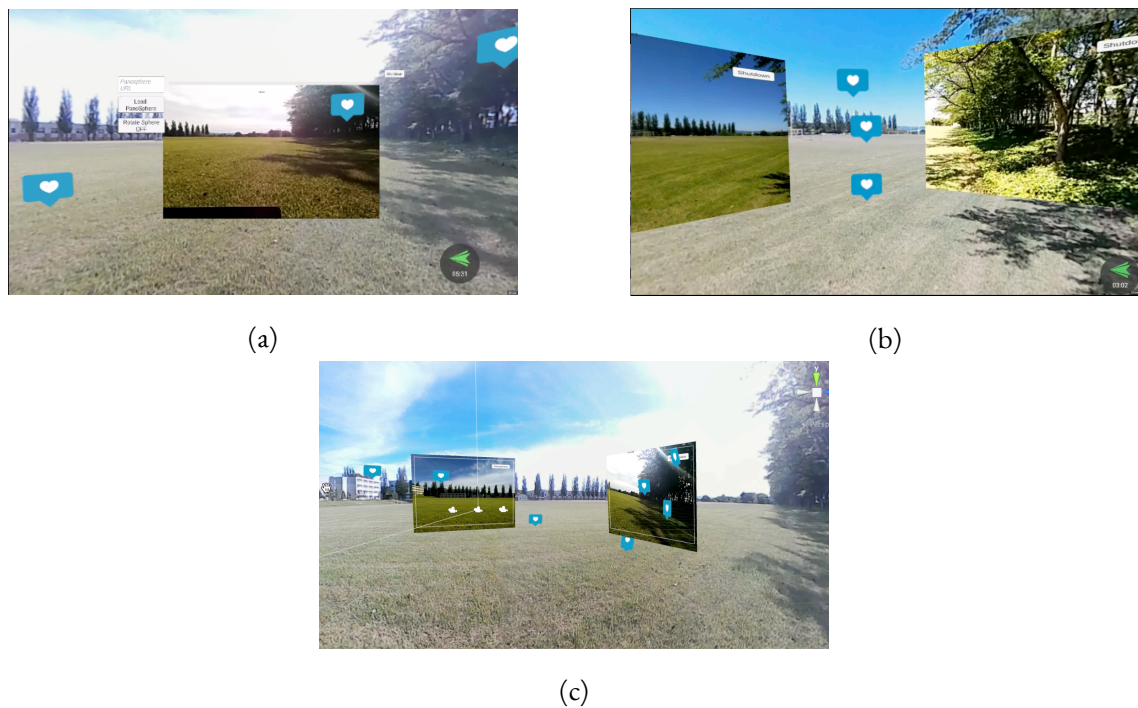


Figure 5.1: Example of a live session featuring two streamers and a viewer: (a) streamer’s perspective, (b) viewer’s perspective, (c) scene overview

- Viewers can place “Like” buttons around a virtual scene, which are mapped in 3D space using mobile rotational tracking. They can be viewed both by streamers and viewers at the same time, and could influence livestreaming experience for all involved users.
- All viewers and streamers can communicate through real-time voice chat.

5.2 Implementation

ReactSpace was built to feature both within-stream and across-stream interactions. The application uses Unity game engine running on Android devices, and achieves multi-stream interaction by creating a single virtual space where all streamers and viewers can communicate with each other. The space is represented by a photosphere (taken by one of the streamers or downloaded from elsewhere) of a real space in which all streamers are supposedly collocated (Fig. 5.1). The spherical background can be uploaded by a streamer at any time throughout the session, and it provides a spatial background for both viewers and streamers.

Spatialized live video streams are represented by video billboards (rectangles that are always

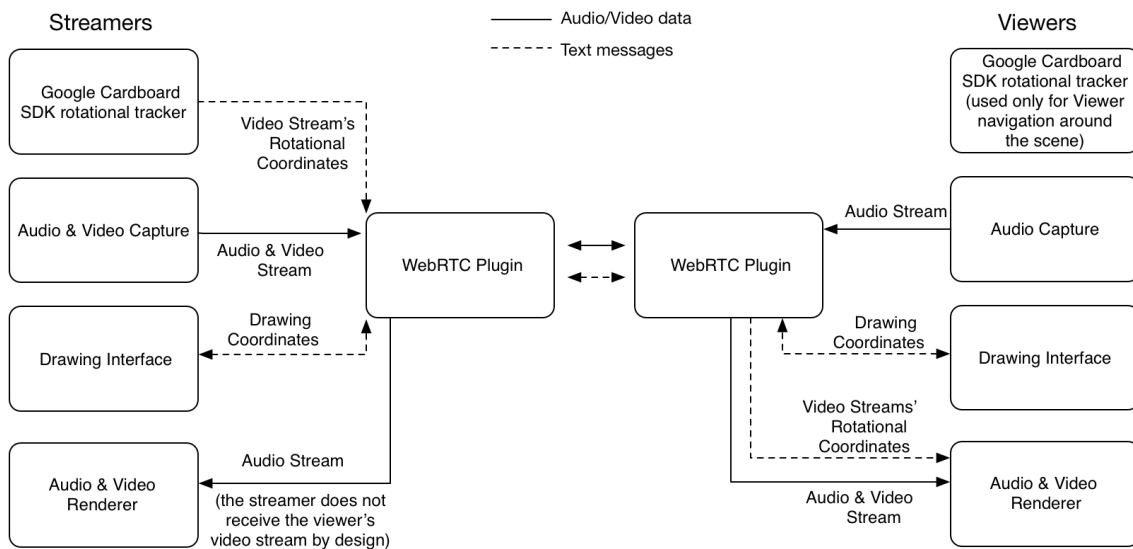


Figure 5.2: ReactSpace network dataflow diagram

facing the viewer), spatially arranged using rotational tracking data from Google Cardboard SDK (which allows the application to run both in handheld and HMD modes). The billboard's rotation around the sphere shows the streamer's current viewpoint and is being updated in real time.

The reaction button functionality is implemented by ray tracing in front of the camera (user's current viewpoint in virtual space), which converts touchscreen coordinates into 3D coordinates within the virtual space.

The network communication between viewers and streamers is handled through Web Real-Time Communication protocol (WebRTC) [47], which is implemented through WebRTC Videochat [48] plugin for Unity3D. WebRTC plugin handles both the transfer of media (audio and video) and text (reaction button coordinates, rotational tracking) data (Fig. 5.2).

5.3 User evaluation

5.3.1 Experiment design

In order to test the effect of SLC method in social livestreaming applications on user engagement, a user evaluation was conducted. In this case by user engagement the study assumes

Table 5.1: ReactSpace user evaluation questionnaire

Q1	I enjoyed being able to view multiple streams in a single virtual space
Q2	I was able to understand what was happening in each stream
Q3	I enjoyed interacting with multiple live streams at once
Q4	I felt like I was able to influence the live streams using the voice feature
Q5-R	I felt like I was able to influence the live streams using hearts in ReactSpace
Q5-P	I felt like I was able to influence the live streams using hearts in Periscope
Q6-R	Using ReactSpace was fun
Q6-P	Using Periscope was fun

the subjective opinion of users on whether they could influence a remote situation using interaction modalities in ReactSpace, and whether they felt if ReactSpace was more interesting an interactive than Periscope.

Each trial consisted of two treatments. In one treatment the users were asked to watch a mobile spherical live video stream located in the experiment location over Periscope. In the other treatment the users were asked to watch multiple streams from the same location using ReactSpace. In both cases the participants were encouraged to use available means of interaction in each streaming service, including text messages and heart buttons in Periscope, spatial reaction buttons and audio in ReactSpace, and navigate around the scene using their phone in both applications. The order of the treatments was randomized before each session.

At the end of each trial, the users were asked to fill out a questionnaire (Table 5.1) which subjectively compared their experiences of using ReactSpace and Periscope. The questionnaire was based on a study by Hamilton et al. [32], it included eight questions, in each one the participants needed to grade their answer based on Likert items (1 - Disagree, 5 - Agree). After the questionnaire, participants were also encouraged to provide any freeform comments and feedback.

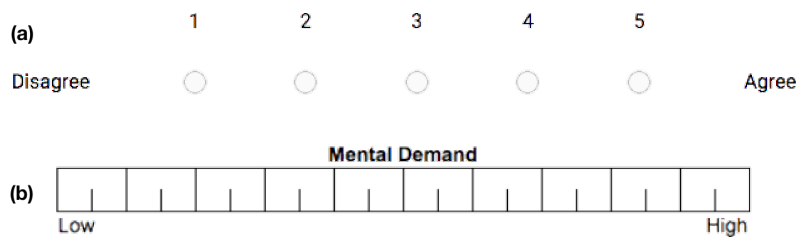


Figure 5.3: Examples of Likert items in questionnaires administered in: (a) ReactSpace, (b) StreamSpace

Table 5.2: ReactSpace user evaluation Z-value and p-value results, p-values that are <0.05 are highlighted in **bold font**.

Name	Z-value	p-value
Q1	2.9191	0.0035
Q2	2.1651	0.0303
Q3	2.7386	0.0061
Q4	2.0702	0.0384
Q5 (P vs R)	-1.8071	0.0707
Q6 (P vs R)	-1.8411	0.0656

5.3.2 Analysis and results

The participants were recruited through social media networks, 9 men and 1 woman, aged between 20 and 35 years old, 10 participants total. Similarly to StreamSpace study, due to the limited pool of participants, a sensitivity analysis, with sample size of 10, $\alpha = 0.05$ and power = 0.8, was conducted, showing minimal effect size that this experiment could possibly detect of $d_z = 0.9$, which according to [52] can be considered as large.

Due to ordinal nature of collected data, the results were analyzed through one-sample (Q1-Q4) and paired (Q5-Q6) Wilcoxon Signed Rank tests. Statistically significant differences ($p < 0.05$) were noticed over an average score of 3 in Q1-Q4 (for exact Z- and p-value scores, please refer to the Table 5.2), but pairwise comparison with the Periscope application did not show any improvement ($p > 0.05$).

Furthermore, upon closer inspection, a possible ceiling effect was detected in questions Q5 and Q6, thus a change in experiment methodology is needed. The effect could have been caused by the inappropriate gradation of Likert items in the questionnaire (users could only choose integer values on the scale from 1 to 5).

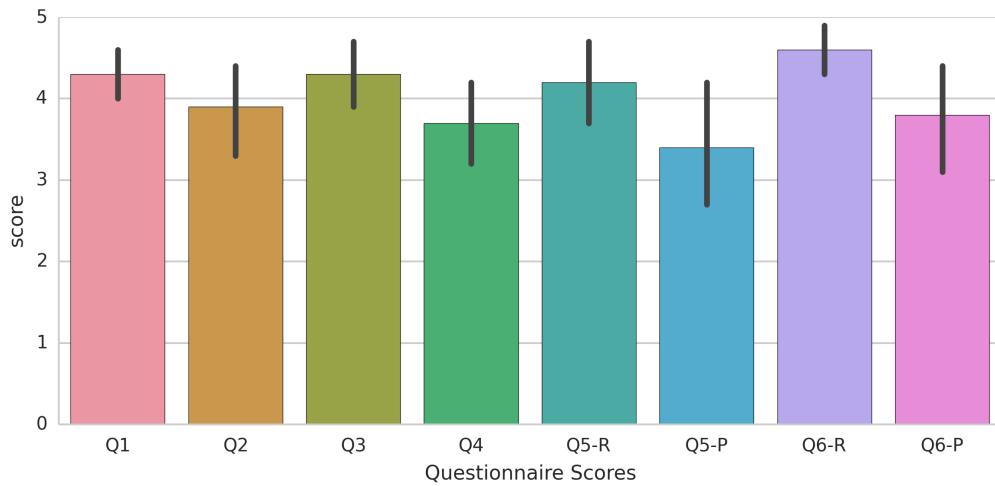


Figure 5.4: User evaluation results

Based on these observations, this work proposes a set of recommendations for future studies on SLC-based social livestreaming applications:

- *Application comparison.* Instead of using commercial applications such as Periscope, it is recommended to use a simplified version of the same SLC application instead. The reasons for that are possible confusion of users by graphical interfaces (e.g., while Periscope and ReactSpace offer similar sets of features, the implementation of user interfaces is vastly different), and the fact that users could have personal opinions (like or dislike) towards commercial social media platforms (e.g., some people dislike Facebook, therefore they might favor any other application that is not Facebook in comparison)
- *Gradation of Likert items.* In this experiment the ceiling effect could have been caused by a crude visual representation of Likert items (Fig. 5.3, a), a more fine-grained gradation, similar to the one implemented in StreamSpace study, could be introduced instead (Fig. 5.3, b). Alternatively, a continuous scale as shown in [60] can be used.

5.4 Conclusion

ReactSpace shows an example of how SLC method can be integrated into social livestreaming applications. Although it is unclear whether there is a statistically significant improvement

in user engagement from the results of user evaluation, the application has received multiple positive comments in user feedback.

Aside from comments, the users also requested such features as haptics (e.g., introducing a vibration feedback when a reaction button is placed in a virtual space), automatic stream rearrangement in cases if multiple streams coincide in one location within a spatial background, and support for spatialized audio (e.g., spatializing streamer voices depending on their rotational orientation).

In the next iterations of ReactSpace system it would be interesting to introduce support for photospherical video, and, depending on the availability of more complex tracking solutions, volumetric video streaming, which can possibly make the livestream viewing experience more interactive in comparison with a currently implemented static photospherical background.

6

Extending the SLC method to other applications

6.1 Introduction

One of the possible approaches to ensuring the further adoption and support of SLC method, is to introduce an ability to share the data used in and generated through SLC-based applications with other mixed reality systems that might not necessarily use SLC method.

Since it is most likely that collaborative applications would be commonplace in the near future, sharing of virtual spaces (or, in SLC terminology, spatial backgrounds) could be beneficial for other applications, as the availability of “metaverse,” a persistent and constantly up-

dated collection of mixed reality spaces mapped to different geospatial locations, could decrease the computational costs for mobile mixed reality applications and expand available interactive space.

This study proposes a decentralized blockchain-based peer-to-peer model of distribution, where spaces are represented as blocks containing necessary information (such as links to photospherical imagery, geospatial data, timestamps, etc.), synchronized among connected users. To demonstrate the implementation of proposed approach, the blockchain-based backend was integrated into a collaborative mobile mixed reality application presented in Chapter 4. This chapter discusses the relevant blockchain-based systems, implementation, and the possible benefits and limitations of such approach.

6.2 Background

The proposed solution combines several key concepts: remote collaboration through mobile mixed reality telepresence and decentralized blockchain-based storage.

The collaborative mixed reality studies observed in Chapter 2 revealed several limitations. For instance, since the applications often did not save the imagery captured in collaborative sessions, each mixed reality space had to be recreated from scratch, requiring additional processing power. Furthermore, the presented applications worked only with one mixed reality space per session, and did not allow users to traverse among multiple active spaces, although studies indicate that remote collaboration can benefit from multi-space and multi-viewpoint [32, 41, 61] interactions with enriched spatial context (e.g., combining video streams with geospatial updates).

However, introducing such mapping functionality poses an architectural challenge: a public system that stores metaverse should be resilient, in case of a large amount of requests, and immutable, to prevent alteration of previously archived spaces by third parties. These issues were partially addressed in social virtual reality network Decentraland [62], which used distributed storage paired with blockchain, a continuous immutable ledger of unique transactions, to ensure the delivery of a single virtual space to multiple users. Similarly, benefits of blockchain



Figure 6.1: Example of multiple mixed reality spaces in a single metaverse

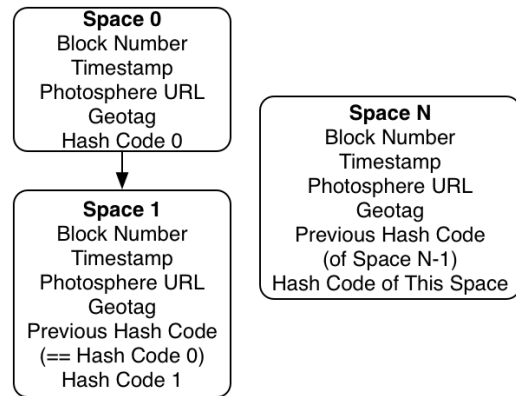


Figure 6.2: Blockchain example and block content outline

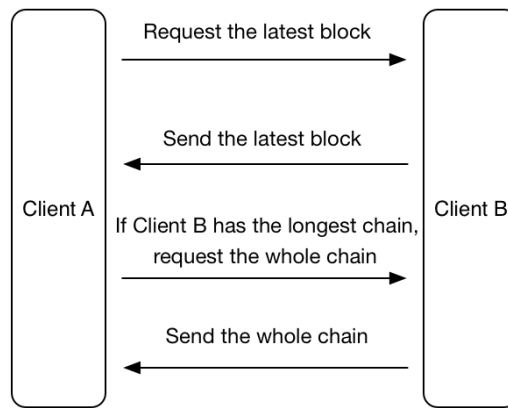


Figure 6.3: Blockchain synchronization protocol

technology for media content delivery have been suggested in [63, 64].

6.3 Implementation

Based on such observations this study proposes a model for decentralized mixed reality space storage and distribution. Since in this case a metaverse does not need to ensure validity of user transactions, it was decided to move the space-related data into the blockchain payload itself. Such approach provides the following advantages: the blockchain can provide unique identifiers for each created space via generated hash codes; it would be immutable, since changing the payload would invalidate the chain; and it is relatively easy to store, update, and share the spaces since it is stored in a form of a plain text JavaScript Object Notation (JSON) array.

The proposed blockchain-based storage was integrated into a mobile collaborative mixed

reality application introduced in Chapter 4. In this application all session participants connect in a peer-to-peer fashion via Web Real-Time Communication (WebRTC) protocol. Each time a new space is created (Fig. 6.1), the application generates a block containing the URL to the photospherical image, its geographical coordinates, and the time it was created (Fig. 6.2). Then it sends the created block to all connected users in a session, updating the blockchain. Whenever a new user joins the session, their application requests blockchains from all connected peers and downloads the longest valid chain (Fig. 6.3).

6.4 Conclusion

6.4.1 Benefits

The proposed approach to extension and sharing of data used in and generated through SLC-based applications can be beneficial both in SLC and non-SLC applications. In the first case, the availability of virtual metaverses allows faster access to spatial backgrounds. For instance, initially in StreamSpace, a user would have had to take a photospherical image and upload it to a web server, however with blockchain-based approach, a user would have access to all previously generated virtual spaces including (depending on availability) the ones appropriate for user's current spatial context.

In the second case with non-SLC based applications, such approach offers access to a more resource-efficient environment mapping solution. For instance, currently such navigational services as Google or Bing Maps do not offer access to creation of personalized photospherical maps (e.g., Google Street View would not necessarily be able to cover a user's own apartment), and a more accessible service that could provide a functionality of creating personal virtual maps and metaverses is desirable. However, creating a centralized service is resource-intensive – users would have to set up and support a personalized online storage service. In such case, a blockchain-based approach offers a simpler alternative, where the metaverse is distributed among all participants and stored on their devices.

6.4.2 Limitations

The current implementation is rather simplistic and does not provide such additional layers of protection against “spamming” attacks as proof-of-work or proof-of-stake. Furthermore, with increasing number of available virtual spaces in a metaverse, eventually adding new blocks could become inefficient: both because of the transferred blockchain size, and because of the hash calculation function. Unless significant changes in storage strategy are introduced, the system might become unscalable in the long run. Possible solutions to this issue could include division of metaverses by regions similarly to [65] (“a metaverse of metaverses”) or applying a different, non blockchain-based approach to distributed storage (e.g., decentralized peer-to-peer storage).

This page intentionally left blank.

7

Conclusions

This dissertation investigated, proposed, and implemented a novel approach to livestream composition for social and collaborative applications. The scope of the conducted work includes both theoretical (taxonomies, SLC model) and empirical (developed applications, user studies) contributions. This chapter gives an overview of completed work in the context of the impact on the industry and the overall body of science, and discusses how other researchers and engineers can build on the conducted work.

7.1 Theoretical contributions

This study has two main theoretical contributions: the introduction of new and update of old taxonomies for social livestreaming systems and mixed reality displays, and design and

implementation of SLC method.

The papers on RV continuum [10] and extended mixed reality taxonomy [9] are often referred to as one of the most cited papers in the fields of mixed reality and human-computer interaction research, which makes the extended mixed reality taxonomy one of the most common (and in some cases the only available) methods for categorization and qualitative comparison of mixed reality displays. However, even in the recent discussions such as [8] the extended taxonomy is used in its initially published form which, at the moment of writing, was introduced over twenty years ago. Therefore, by updating the extended mixed reality taxonomy, this dissertation contributes to the field of mixed reality research by providing a more updated and fine-tuned classification of mixed reality displays. In comparison with the more detailed standards for categorization of MR / XR research that are still in the active development by IEEE VRAR [66] and Khronos groups [67], the updated taxonomy presented in this dissertation can be readily applied by scientists and engineers in the relevant fields of research.

Creating the new social livestreaming systems taxonomy also contributes to the mixed reality and social media research. Given the fact that social media applications are integrating a rising number of mixed reality features (e.g., Facebook Spaces [35]), social media system taxonomies similar to the one proposed in this work would be needed for further categorization and comparison of relevant work. The interest in such has already been shown in the recently released studies such as [32, 61], and the author of this dissertation believes that the interest in such classification would only grow in subsequent years. Therefore by introducing the new social livestreaming systems taxonomy, this work supports the discussion on the future of social livestreaming systems and helps shape the unified classification upon which such systems can be compared.

Finally, the introduction of SLC method also contributes to the relevant fields of research discussed within the scope of this work. Although the introduction of spatial models for virtual environments is not a new concept (in fact, some of the proposals are even older than the mixed reality taxonomy [68]), the author of this dissertation believes that restarting and advancing the dialogue on collaborative mixed reality telepresence models through introduction of SLC

method is beneficial for development of future collaborative systems. Furthermore, considering that some of the recently published studies already fit into the SLC method [43, 44], such formalization of live media stream composition can already influence and benefit the existing and ongoing research projects.

7.2 Empirical contributions

This study also adds to the overall body of science through empirical contributions, by developing proof-of-concept prototype applications and conducting user evaluations.

The methods of interactions presented in the proof-of-concept applications have been discussed and showcased at ACM SIGGRAPH Asia 2017 Symposium on Mobile Graphics and Interactive Applications, where the presented approach received praise for novelty and practicality both during peer-review process and during the on-site live demonstration. The author of this dissertation believes that such feedback can indicate tangible interest among industry professionals in the work conducted within the scope of this thesis. Furthermore, the received commentaries (both from conference and journal publications) have prompted discussion on several aspects presented within the developed applications. For instance, the issue of field-of-view (FoV) matching (the de-synchronization between a photospherical image and a picture in live media stream within the same virtual space), has prompted the author of this work to conduct an investigation, results of which will be used in the future iterations of developed proof-of-concept applications.

Aside from discussions, the developed applications follow module-based architecture, and thus can be reused by other scientists and developers in corresponding projects. For example, the author sees future revisions of commercial videoconferencing applications having a “regular”, monoscopic, mode, and a “spatial” mode such as presented in StreamSpace (including spatialized live media streams and annotations), which can be switched on or off depending on the context (e.g., users need richer spatial context to complete a task, but do not need it for regular video calls). Such features are already finding their way in specialized commercial applications such as Skype for HoloLens [18] and vuforia chalk [69], and therefore by having

the developed application projects available online, the SLC method can be further adopted in other mixed reality systems.

Finally, conducted user studies also contribute to the fields of human-computer interaction and mixed reality research. First of all, based on the StreamSpace evaluation results, it was possible to confirm the benefits of SLC method in collaborative applications. As the matter of fact, the lack of improvement in streamer scores poses an even higher research interest, because it indicates that the user interface provided for streamers was insufficient, and in the future similar projects redesigning such interface would be necessary. However the problem is that fitting a redesigned streamer interface within the SLC method paradigm could be challenging: for instance, instead of having the video feed fixed in the center, streamers could have an interface element that would “maximize” or “minimize” their video feeds depending on the context, but at the same time it is unclear how in this case a spatial background should be implemented. Perhaps an interface with dynamic photospherical video for a spatial background similar to [43] could be more beneficial. Secondly, besides implementation challenges, the user evaluation results have demonstrated a need for different experiment setup for social livestreaming systems. Although the study was based on the existing literature, it was still insufficient to reliably detect statistically significant improvement in user engagement when using ReactSpace. The author hopes that the formulated set of recommendations could help other researchers in designing their own experiments, especially considering the high occurrence of relatively small (10-30 participants) sample size experiments employing questionnaires with Likert items in observed relevant literature (e.g., [23, 26, 37, 61]).

7.3 Future work

This study can be described as a part of emerging trend of mobile mixed reality telepresence studies, named as “the future of collaboration” by Mark Billinghurst [70]. The author of this dissertation sees following tangentially related concepts and technologies as the ones that can be successfully combined with SLC method and SLC-based applications.

7.3.1 Theoretical applications

- *SLC as a pointing framework.* The act of pointing poses a considerable interest in human-computer interaction research, as it helps understanding how participants share and interact with visible information within the same real or virtual area of interaction. Charles Goodwin has proposed the pointing framework [71], which was further discussed by Streeck[72]. In this framework, pointing is defined by the visibility of participants' bodies and their orientation in space, communication between participants, space which is being pointed at and the context within which this activity is being performed. In essence, the SLC-based applications are implementing a similar approach by creating a shared interaction area (spatial background), which contextualizes the user actions through the act of rotationally tracked media streams and annotations. SLC method and SLC-based applications could be useful in the further advancement of understanding how pointing is being used in mixed reality, in order to form user interface design recommendations for future collaborative applications.
- *Proxemic interactions.* Defined by Greenbert et al.[73], the concept of proxemic interactions discusses how multiple displays (e.g., laptops, phones, tablets, projectors), located within the short distance from each other, could be connected in a single collaborative space, creating a more efficient way of sharing the information among connected participants. Similarly, SLC-based applications could benefit from proxemic interaction frameworks as it is important to understand how locally and remotely-connected users could effectively share information together in a single mixed or extended reality collaborative environment.
- *Extending the Quality of Experience (QoE) model.* In video streaming applications, a QoE function is used to determine users' quality of video stream viewing experience, which depends on multiple different factors such as network connectivity, video quality, and available bandwidth. Based on QoE function, for instance, a video streaming

service can determine the quality of delivered media content via adaptive streaming algorithms. It would be interesting to combine the content categories defined within the SLC method with this concept, in order to see how QoE function can be adapted to real-time video streaming in mixed reality.

- *Cross-modal interactions in neurological studies.* Recently, mixed and virtual reality interfaces have gained popularity in neurological projects. SLC method and SLC-based applications could, for instance, be applied to cross-modal interactions research projects such as [74], or extended reality brain-computer interfaces such as [12].

7.3.2 Empirical applications

- *WebXR.* It would be interesting to expand a collaborative application such as StreamSpace to support WebAR / WebVR. The advantage of WebXR technology is that it does not require a prior installation on users' mobile devices, and thus could be used as a drop-in collaborative solution (for instance, on-demand remote help in public spaces such as libraries or supermarkets). The URLs to WebXR applications could be distributed via QR codes or near-field communication (NFC) markers.
- *Low-latency, "tactile" 5G network.* The WebRTC network backend implementation in proof-of-concept applications was designed to ensure low latency, however it might still perform imperfectly on mobile networks such as 3G or 4G. The next generation 5G networks support extremely low-latency, "tactile," connections which can benefit interactive applications. It would be interesting to combine the multimodal interaction approaches compatible with SLC method with such advanced network connectivity.

Finally, the author hopes that regardless of the type of application, the conducted work would benefit future scientists and engineers in the fields of virtual reality and human-computer interaction.

Acknowledgments

I would like to thank my advisors, Professors Michael Cohen, Jens Herder, and Julián Villegas for guiding me throughout the doctoral program, and my dissertation committee members, Professors Maxim Mozgovoy and Ihor Lubashevsky for providing their insight into my thesis.

I would also to thank my friends and colleagues, including, but not limiting to, Daniel Drochert for encouragement and help throughout the past three years, Arkady Zgonnikov for help with statistical data, Professor Yoichi Ochiai and Digital Nature Group of University of Tsukuba for assisting me throughout the last year of doctoral program, and Professor Evgeny Pyshkin who supported development of the blockchain-based mapping project via JSPS funding. Furthermore, I would like to thank students and staff at University of Applied Sciences Düsseldorf, MEXT, IEEE Student Chapter of University of Aizu, and ACM SIGGRAPH for supporting my work.

Finally, this thesis would not have been possible without the unconditional support and encouragement from my parents, to whom I dedicate this work.

This page intentionally left blank.

References

- [1] Cisco, “The Zettabyte Era: Trends and Analysis,” Jun. 2017.
- [2] Alexa Ranking. (Accessed: 2017-10-08) Top Sites. [Online]. Available: <https://www.alexa.com/topsites>
- [3] B. Ryskeldiev, “Spatial Social Media: Towards Collaborative Mixed Reality Telepresence “On The Go,” in *CHI’18 Extended Abstracts*. New York, NY, USA: ACM, 2018. [Online]. Available: <https://doi.org/10.1145/3170427.3173020>
- [4] B. Ryskeldiev, M. Cohen, and J. Herder, “StreamSpace: Pervasive Mixed Reality Telepresence for Remote Collaboration on Mobile Devices,” *IPSJ Journal of Information Processing, Special issue of “Advances in Collaboration Technologies”*, vol. 26, no. 1, pp. 1–9, Jan. 2018. [Online]. Available: <http://www.ipsj.or.jp/english/index.html>
- [5] B. Ryskeldiev, M. Cohen, J. Herder, and Y. Ochiai, “ReactSpace: Spatial-Aware User Interactions for Collocated Social Live Streaming Experiences,” in *IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC)*. IEEE, 2018 (Accepted).
- [6] B. Ryskeldiev, M. Cohen, and J. Herder, “Applying Rotational Tracking and Photospherical Imagery to Immersive Mobile Telepresence and Live Video Streaming Groupware,” in *SIGGRAPH Asia 2017 Mobile Graphics and Interactive Applications*, ser. SA ’17. New York, NY, USA: ACM, 2017, pp. 5:1–5:2. [Online]. Available: <http://doi.acm.org/10.1145/3132787.3132813>
- [7] B. Ryskeldiev, Y. Ochiai, M. Cohen, and J. Herder, “Distributed metaverse: creating decentralized blockchain-based models for peer-to-peer sharing of virtual spaces for mixed reality applications,” in *The 9th Augmented Human Int. Conf.* New York, NY, USA: ACM, 2018. [Online]. Available: <https://doi.org/10.1145/3174910.3174952>
- [8] M. Billingham. (Accessed: 2017-10-08) What is Mixed Reality? [Online]. Available: <https://medium.com/startup-grind/what-is-mixed-reality-60e5cc284330>
- [9] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, “Augmented reality: A class of displays on the reality-virtuality continuum,” in *Photonics for industrial applications*. Int. Society for Optics and Photonics, 1995, pp. 282–292.
- [10] P. Milgram and F. Kishino, “A taxonomy of mixed reality visual displays,” *IEICE Trans. on Information and Systems*, vol. 77, no. 12, pp. 1321–1329, 1994.

- [11] M. Slater, “A note on presence terminology,” *Presence connect*, vol. 3, no. 3, pp. 1–5, 2003.
- [12] J. Jantz, A. Molnar, and R. Alcaide, “A brain-computer interface for extended reality interfaces,” in *ACM SIGGRAPH 2017 VR Village*. ACM, 2017, p. 3.
- [13] T. Morozumi, S. Mori, S. Ikeda, F. Shibata, A. Kimura, and H. Tamura, “[POSTER] Design and Implementation of a Common Dataset for Comparison and Evaluation of Diminished Reality Methods,” in *Mixed and Augmented Reality (ISMAR-Adjunct), 2017 IEEE Int. Symp. on*. IEEE, 2017, pp. 212–213.
- [14] M. Hirose, “Image-based virtual world generation,” *IEEE Multimedia*, vol. 4, no. 1, pp. 27–33, 1997.
- [15] M. Billinghurst and H. Kato, “Collaborative augmented reality,” *Communications of the ACM*, vol. 45, no. 7, pp. 64–70, 2002.
- [16] M. Cohen, K. Doi, T. Hattori, and Y. Mine, “Control of Navigable Panoramic Imagery with Information Furniture: Chair-Driven 2.5D Steering Through Multistandpoint QTVR Panoramas with Automatic Window Dilation,” in *Proc. CIT: 7th Int. Conf. on Computer and Information Technology*, T. Miyazaki, I. Paik, and D. Wei, Eds., Aizu-Wakamatsu, Japan, Oct. 2007, pp. 511–516.
- [17] H. Jo and S. Hwang, “Chili: viewpoint control and on-video drawing for mobile video calls,” in *CHI’13 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2013, pp. 1425–1430.
- [18] H. Chen, A. S. Lee, M. Swift, and J. C. Tang, “3D collaboration method over HoloLens™ and Skype™ end points,” in *Proc. of the 3rd Int. Workshop on Immersive Media Experiences*. ACM, 2015, pp. 27–30.
- [19] B. Nuernberger, K.-C. Lien, T. Höllerer, and M. Turk, “Interpreting 2d gesture annotations in 3d augmented reality,” in *3D User Interfaces (3DUI), 2016 IEEE Symp. on*. IEEE, 2016, pp. 149–158.
- [20] B. Agüera. (Accessed: 2017-10-08) TED Talks: Microsoft Augmented-Reality Maps. [Online]. Available: https://www.ted.com/talks/blaise_aguera
- [21] M. Billinghurst, A. Nassani, and C. Reichherzer, “Social Panoramas: Using Wearable Computers to Share Experiences,” in *SIGGRAPH Asia Mobile Graphics and Interactive Applications*, ser. SA ’14. New York: ACM, 2014, pp. 25:1–25:1.
- [22] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer, “World-stabilized annotations and virtual scene navigation for remote collaboration,” in *Proc. of the 27th annual ACM Symp. on User interface software and technology*. ACM, 2014, pp. 449–459.
- [23] S. Kasahara and J. Rekimoto, “JackIn: integrating first-person view with out-of-body vision generation for human-human augmentation,” in *Proc. of the 5th Augmented Human Int. Conf.* ACM, 2014, p. 46.

- [24] S. Nagai, S. Kasahara, and J. Rekimoto, “Livesphere: Sharing the surrounding visual environment for immersive experience in remote collaboration,” in *Proc. of the Ninth Int. Conf. on Tangible, Embedded, and Embodied Interaction*. ACM, 2015, pp. 113–116.
- [25] M. Y. Saraiji, S. Sugimoto, C. L. Fernando, K. Minamizawa, and S. Tachi, “Layered Telepresence: Simultaneous Multi Presence Experience Using Eye Gaze Based Perceptual Awareness Blending,” in *ACM SIGGRAPH 2016 Posters*, ser. SIGGRAPH ’16. New York: ACM, 2016, pp. 20:1–20:2. [Online]. Available: <http://doi.acm.org/10.1145/2945078.2945098>
- [26] J. Müller, T. Langlotz, and H. Regenbrecht, “PanoVC: Pervasive telepresence using mobile phones,” in *Pervasive Computing and Communications (PerCom)*. IEEE, 2016, pp. 1–10.
- [27] S. Singhal and C. Neustaedter, “BeWithMe: An Immersive Telepresence System for Distance Separated Couples,” in *Companion of the 2017 ACM Conf. on Computer Supported Cooperative Work and Social Computing*. ACM, 2017, pp. 307–310.
- [28] S. Kasahara, S. Nagai, and J. Rekimoto, “JackIn Head: Immersive Visual Telepresence System with Omnidirectional Wearable Camera,” *IEEE trans. on visualization and computer graphics*, vol. 23, no. 3, pp. 1222–1234, 2017.
- [29] MultiTwitch.tv. (Accessed: 2017-10-08) This site that enables watching any number of twitch.tv streams at the same time. [Online]. Available: <http://multitwitch.tv/>
- [30] kbmod. (Accessed: 2017-10-08) The KBMOD Gaming Community. [Online]. Available: <https://kbmod.com/>
- [31] Google. (Accessed: 2017-10-08) Hangouts On Air with YouTube Live. [Online]. Available: <https://support.google.com/youtube/answer/7083786?hl=en>
- [32] W. A. Hamilton, J. Tang, G. Venolia, K. Inkpen, J. Zillner, and D. Huang, “Rivulet: Exploring participation in live events through multi-stream experiences,” in *Proc. of the ACM Int. Conf. on Interactive Experiences for TV and Online Video*. ACM, 2016, pp. 31–42.
- [33] M. McLuhan, *Understanding media: The extensions of man*. MIT press, 1994.
- [34] W. A. Hamilton, O. Garretson, and A. Kerne, “Streaming on Twitch: Fostering Participatory Communities of Play Within Live Mixed Media,” in *Proc. of the 32Nd Annual ACM Conf. on Human Factors in Computing Systems*, ser. CHI ’14. New York, NY, USA: ACM, 2014, pp. 1315–1324. [Online]. Available: <http://doi.acm.org/10.1145/2556288.2557048>
- [35] Facebook. (Accessed: 2017-10-08) Facebook Spaces. [Online]. Available: <https://www.facebook.com/spaces>

- [36] T. Yonezawa and H. Tokuda, *Enhancing communication and dramatic impact of online live performance with cooperative audience control*. UbiComp'12 - Proc. of the 2012 ACM Conf. on Ubiquitous Computing, 2012, pp. 103–112.
- [37] A. Nassani, H. Kim, G. Lee, M. Billinghamurst, T. Langlotz, and R. W. Lindeman, “Augmented reality annotation for social video sharing,” in *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications*. ACM, 2016, p. 9.
- [38] Periscope. (Accessed: 2017-10-08) What is Periscope Producer? [Online]. Available: <https://help.pscp.tv/customer/en/portal/articles/2600293-what-is-periscope-producer>
- [39] Facebook. (Accessed: 2017-10-08) Introducing Facebook Surround 360: An open, high-quality 3D-360 video capture system. [Online]. Available: <https://code.facebook.com/posts/1755691291326688/introducing-facebook-surround-360-an-open-high-quality-3d-360-video-capture-system/>
- [40] C. Zhou. (Accessed: 2017-10-08) Making 12K 360°VR Streaming a Reality: Why and How We Did It. [Online]. Available: <https://medium.com/visbit/making-12k-360%C2%BA-vr-streaming-a-reality-why-and-how-we-did-it-ce65e9aa0bc3>
- [41] S. Kim, S. Junuzovic, and K. Inkpen, “The Nomad and the Couch Potato: Enriching Mobile Shared Experiences with Contextual Information,” in *Proc. of the 18th Int. Conf. on Supporting Group Work*, ser. GROUP '14. New York, NY, USA: ACM, 2014, pp. 167–177. [Online]. Available: <http://doi.acm.org/10.1145/2660398.2660409>
- [42] Google. (Accessed: 2017-10-08) Android ARCore. [Online]. Available: <https://developers.google.com/ar/>
- [43] G. A. Lee, T. Teo, S. Kim, and M. Billinghamurst, “Sharedsphere: MR Collaboration Through Shared Live Panorama,” in *SIGGRAPH Asia 2017 Emerging Technologies*, ser. SA '17. New York, NY, USA: ACM, 2017, pp. 12:1–12:2. [Online]. Available: <http://doi.acm.org/10.1145/3132818.3132827>
- [44] H. Fushimi, D. Kato, Y. Kamiyama, K. Yanagihara, K. Minamizawa, and K. Kunze, “atmoSphere: designing cross-modal music experiences using spatial audio with haptic feedback,” in *ACM SIGGRAPH 2017 Emerging Technologies*. ACM, 2017, p. 4.
- [45] H. Schnädelbach, A. Penn, and P. Steadman, “Mixed reality architecture: a dynamic architectural topology,” 2007.
- [46] Google. (Accessed: 2017-10-08) Google Visual Positioning Service overview. [Online]. Available: https://developers.google.com/tango/overview/concepts#visual_positioning_service_overview
- [47] A. Bergkvist, D. C. Burnett, C. Jennings, and A. Narayanan, “WebRTC 1.0: Real-time communication between browsers,” *Working draft, W3C*, vol. 91, 2012.
- [48] C. Kutza. (Accessed: 2017-10-08) WebRTC Video Chat. [Online]. Available: <https://www.assetstore.unity3d.com/en/#!/content/68030>

- [49] Insta360. (Accessed: 2017-10-08) Insta360 Air. [Online]. Available: <https://www.insta360.com/product/insta360-air>
- [50] S. G. Hart, “NASA-task load index (NASA-TLX); 20 years later,” in *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9. Sage Publications Sage CA: Los Angeles, CA, 2006, pp. 904–908.
- [51] A. Cao, K. K. Chintamani, A. K. Pandya, and R. D. Ellis, “NASA TLX: Software for assessing subjective mental workload,” *Behavior Research Methods*, vol. 41, no. 1, pp. 113–117, 2009.
- [52] K. Yatani, “Effect Sizes and Power Analysis in HCI,” in *Modern Statistical Methods for HCI*. Springer, 2016, pp. 87–110.
- [53] Kudan. (Accessed: 2017-10-08) Kudan Computer Vision. [Online]. Available: <https://www.kudan.eu/>
- [54] Google. (Accessed: 2017-10-08) Google Daydream. [Online]. Available: <https://vr.google.com/daydream/>
- [55] Microsoft. (Accessed: 2017-10-08) Microsoft HoloLens. [Online]. Available: <https://www.microsoft.com/en-us/hololens>
- [56] P. Lopes, S. You, L.-P. Cheng, S. Marwecki, and P. Baudisch, “Providing Haptics to Walls; Heavy Objects in Virtual Reality by Means of Electrical Muscle Stimulation,” in *Proc. of the 2017 CHI Conf. on Human Factors in Computing Systems*, ser. CHI ’17. New York: ACM, 2017, pp. 1471–1482. [Online]. Available: <http://doi.acm.org/10.1145/3025453.3025600>
- [57] D. Ren, T. Goldschwendt, Y. Chang, and T. Höllerer, “Evaluating wide-field-of-view augmented reality with mixed reality simulation,” in *Virtual Reality (VR), 2016 IEEE*. IEEE, 2016, pp. 93–102.
- [58] Apple. (Accessed: 2017-10-08) Apple ARKit. [Online]. Available: <https://developer.apple.com/arkit/>
- [59] Google. (Accessed: 2017-10-08) Use spatial audio in 360-degree and VR videos. [Online]. Available: <https://support.google.com/youtube/answer/6395969>
- [60] G. M. Sullivan and A. R. Artino Jr, “Analyzing and interpreting data from Likert-type scales,” *Journal of graduate medical education*, vol. 5, no. 4, pp. 541–542, 2013.
- [61] A. Nassani, G. Lee, M. Billingham, T. Langlotz, S. Hoermann, and R. W. Lindeman, “[POSTER] The Social AR Continuum: Concept and User Study,” in *Mixed and Augmented Reality (ISMAR-Adjunct), 2017 IEEE Int. Symp. on*. IEEE, 2017, pp. 7–8.
- [62] E. Ordano, A. Meilich, Y. Jardi, and M. Araoz, “Decentraland: A blockchain-based virtual world,” 2017. [Online]. Available: <https://decentraland.org/whitepaper.pdf>

- [63] N. Herbaut and N. Negru, “A model for collaborative blockchain-based video delivery relying on advanced network services chains,” *IEEE Communications Magazine*, vol. 55, no. 9, pp. 70–76, 2017.
- [64] A. Chakravorty and C. Rong, “Ushare: user controlled social media based on blockchain,” in *Proc.s of the 11th Int. Conf. on Ubiquitous Information Management and Communication*. ACM, 2017, p. 99.
- [65] Mastodon.social. (Accessed: 2017-10-08) Mastodon social network. [Online]. Available: <https://mastodon.social/about>
- [66] IEEE Computer Society. (Accessed: 2017-10-08) IEEE VRAR - Virtual Reality and Augmented Reality Working Group: Standards for Virtual and Augmented Reality. [Online]. Available: <https://standards.ieee.org/develop/wg/VRAR.html>
- [67] Khronos Group. (Accessed: 2017-10-08) OpenXR standard. [Online]. Available: <https://www.khronos.org/openxr>
- [68] S. Benford and L. Fahlén, “A spatial model of interaction in large virtual environments,” in *Proc. of the Third European Conf. on Computer-Supported Cooperative Work 13–17 September 1993, Milan, Italy ECSCW’93*. Springer, 1993, pp. 109–124.
- [69] Vuforia. (Accessed: 2017-10-08) Vuforia Chalk. [Online]. Available: <https://chalk.vuforia.com/>
- [70] M. Billingham. (Accessed: 2017-10-08) Will Mixed Reality Replace Phone Calls? [Online]. Available: <https://medium.com/super-ventures-blog/will-mixed-reality-replace-phone-calls-29b1feb2c62a>
- [71] C. Goodwin, “Pointing as situated practice,” *Pointing: Where language, culture and cognition meet*, pp. 217–241, 2003.
- [72] J. Streeck, “Embodiment in human communication,” *Annual Review of Anthropology*, vol. 44, pp. 419–438, 2015.
- [73] S. Greenberg, N. Marquardt, T. Ballendat, R. Diaz-Marino, and M. Wang, “Proxemic Interactions: The New UbiComp?” *interactions*, vol. 18, no. 1, pp. 42–50, Jan. 2011. [Online]. Available: <http://doi.acm.org/10.1145/1897239.1897250>
- [74] V. J. Harjunen, I. Ahmed, G. Jacucci, N. Ravaja, and M. M. Spapé, “Manipulating bodily presence affects cross-modal spatial attention: A virtual-reality-based ERP study,” *Frontiers in human neuroscience*, vol. 11, 2017.