VROOM: Virtual Robot Overlay for Online Meetings

Brennan Jones
University of Calgary &
Microsoft Research
Cambridge
bdgjones@ucalgary.ca

Priscilla Wong
University College London &
Microsoft Research
Cambridge
ngoi.wong.13@ucl.ac.uk

Yaying Zhang
Microsoft Vancouver
yayzhang@microsoft.com

Sean Rintel
Microsoft Research
Cambridge
serintel@microsoft.com

CHI 2020 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6819-3/20/04.
DOI: https://doi.org/10.1145/3334480.XXXXXXX
*update the above block & DOI per your rightsreview confirmation (provided after acceptance)

Abstract

Telepresence robots allow remote users to freely explore a space they are not in, and provide a physical embodiment in that space. However, they lack a compelling representation of the remote user in the local space. We present *VROOM* (*Virtual Robot Overlay for Online Meetings)*, a two-way system for exploring how to improve the social experience of robotic telepresence. For the local user, an augmented-reality (AR) interface shows a life-size avatar of the remote user overlaid on a telepresence robot. For the remote user, a head-mounted virtual-reality (VR) interface presents an immersive 360° view of the local space with mobile autonomy. The VR system tracks the remote user's head pose and hand movements, which are applied to the avatar. This allows the local user to see the remote's head direction and hand gestures, and the remote user to identify with the robot as an identifiable embodiment of self.

Author Keywords

Telepresence; remote collaboration; video communication; mixed reality; augmented reality; virtual reality; remote embodiment; avatar; awareness

CSS Concepts

- **Human-centered computing~Mixed / augmented reality**

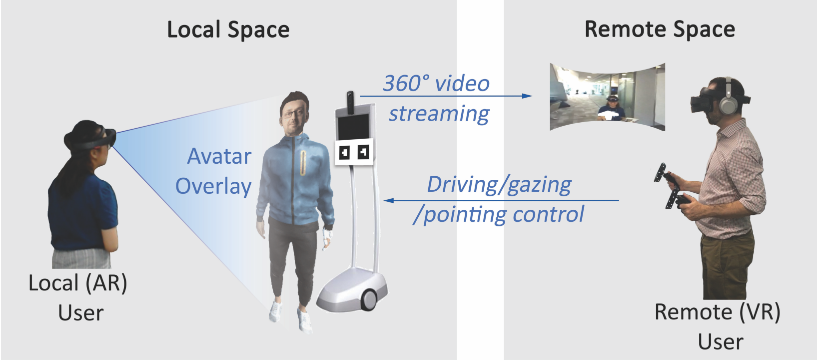

Figure 1: An overview of the VROOM system.

Introduction

Video communication has enabled global work and personal life and distributed access to education, healthcare, professional services, and more. However, traditional 2D video calling is inherently asymmetrical, constraining users' abilities to achieve common ground [18], maintain awareness and control [12,18], and share experiences [22]. People can certainly work around these constraints, but they will still be physically and spatiality limited.

There are two notable ways to add physical and spatial experiences into video communication. One is to use telepresence robots (e.g., [23,34,35])—effectively video-chat-on-wheels. While not yet widely adopted in domestic contexts, they are increasingly common in the workplace (e.g., [36]). Another is to use an augmented-reality (AR) system (e.g., [37]), in which remote and local users wear head mounted devices to see one another as avatars in their respective local spaces. These avatars can be cartoon-like or more photorealistic, providing something approaching a parametric representation of the user.

Both methods are a step ahead of traditional video communication in terms of physical and spatial mobility, autonomy, and bodily identification. However, they both still have limitations. Telepresence robots still lock users into 2D screens showing constrained fields of view (FOVs) from a remote camera, so while they allow for more autonomous mobility, they suffer from the same limitations on conveying remote users' body language and expressions as traditional video communication [21]. In addition, low FOVs can result in reduced task performance [11]. While current technologies limit the fidelity of parametric avatars, AR systems enable free use of arm and hand gestures and spatial bodily arrangements. However, mobile autonomy in a remote location is limited to a current shared meeting instance with another person—an AR avatar cannot roam around a remote location.

The question, then, is how to combine robotic telepresence with mixed-reality avatars to provide the best of both worlds to each endpoint of a video-call experience? We designed a system called *VROOM* (*Virtual Robot Overlay for Online Meetings*; Figure 1) that enhances the experience of being embodied by a telepresence robot as a remote user, as well as the experience of being co-located with a virtual avatar overlaid on a telepresence robot as a local user.

Background

Given the evolution of video communication and improvements in allied technologies such as wireless-communications infrastructure, more opportunities to collaborate with people not sharing a physical space are arising. This is enabling a greater number of remote and distributed work meetings and activities. As a result, such technologies are empowering, providing inclusion and new opportunities to those who would otherwise not have them; e.g., people with disabilities who cannot leave the home, people who live far away and cannot afford to travel, or need to be far away to take care of family members, etc.

However, video communication is rife with asymmetries. For example, in hybrid meetings [30] a remote person calling in to a room with several people in it cannot control their viewpoint in the room on their own, thus restricting them to seeing only what the laptop camera sees. A remote user's lack of ability to control their view can lead to frustration and provide an unequal experience, making it difficult for the user to contribute to the activity at hand [12]. While there are some activities that are not hampered by these asymmetries, remote participation becomes more difficult as physicality increases (e.g., where groups of people share ideas through sketching on whiteboards [10], refer to whiteboard drawings and objects in the room [7], or express ideas through body language [7]).

Designers have proposed various solutions to such asymmetries. These include giving users the ability to refer to objects in the other space [2,4,5,6,8,15], giving people richer or wider views into the space [11,13,31], and giving users more control of their viewpoint into the space through mechanical movable cameras [16,17] or telepresence robots [23,34,35].

A telepresence robot (e.g., [23,34,35]) is a remotely-controlled movable robot with a screen, speakers, a microphone, and a camera. Such a robot provides an experience that is akin to 'video-chat on wheels,' allowing a user to drive around and interact with people in another space. The main benefits of these robots are twofold: (1) they give remote users the ability to freely explore an environment that they are not physically in [28], and (2) they give remote users a *physical* presence, or *physical* embodiment, in that environment [28]; thus giving them a *place* in that environment, making that space *belong* to them (just as much as a spot occupied by another physically-present collaborator belongs to them). Usage of telepresence robots has been studied by researchers in HCI and CSCW, in collaborative and social contexts such as museum visits [29], remotely attending academic conferences [21,27], outdoor activities [9], and long-distance relationships [32,33]. Such research concentrates on the nature of the telepresence experiences more than the technical needs of creating wholly new robotic telepresence experiences.

Past research has looked at mixed reality (MR), including AR and virtual reality (VR), as a tool for supporting remote collaboration. Piumsomboon et al. explored adding a miniature avatar of the remote user to the local space, viewable through an AR headset [24]. They also explored attaching the mini avatar to a 360° camera with a tracker, and using the 360° camera to provide an immersive view of the local environment to the remote user, who views the environment through an immersive VR headset [25]. We expand on these concepts by attaching the remote user's avatar to a telepresence robot, thus providing the remote user a means of locomoting through the environment, as well as an additional physical embodiment which others not wearing the AR headset can see.

Previous work has explored the use of 360° cameras for video conferencing [13,19,20,31] and even on telepresence robots [9]. Viewing a 360° live video

Figure 2: Avatar representation in the local (AR) user's view.

Figure 3: Avatar representation in the remote (VR) user's view.

through a VR HMD can lead to a higher sense of immersion and emotional investment in the remote location [3], which could result in a remote user contributing more to or getting more out of a shared activity in that space. While immersive 360° video shows promise, simply adding a larger field of view does not provide an easy way of gesturing to other users or indicating which direction they are looking in the remote space [31]. We aim to address this through the remote user's avatar, viewable by the local user.

System Design

We built VROOM to enhance the experience of using a telepresence robot for both the local (AR) and remote (VR) users. In the local space, we give the local user a life-size view of a compelling representation of the remote user and their gestures, head direction, and other non-verbal behaviour. This is provided through a virtual-avatar representation of the remote user, viewed through a HoloLens headset worn by the local user and overlaid over the telepresence robot. In the remote space, we give the remote user an immersive and autonomous view into the activity space, allowing them to look around more freely in 360°. This is provided through a 360° live video image, viewable through a VR headset, from a 360° camera attached to the top of the telepresence robot. We also provide a first-person view of their avatar

Local Space

Local (AR) User Interface: The local space is the location where the activity is taking place. In there, the local user wears Microsoft HoloLens, through which they can see the remote user's avatar representation overlaid on the telepresence robot.

Avatar Representation: In VROOM, we created an avatar representing the remote user's appearance. This avatar is displayed in both the local (AR) user's view (Figure 2), and remote (VR) user's view (Figure 3). In the local user's view, the avatar is overlaid on the telepresence robot using marker tracking. In the remote user's view, it can be seen in first-person (e.g., when the user look down or moves their hands), and thus becomes like a view of the user's own body. The avatar's appearance and actions are mapped to the remote user. This mapping includes two aspects:

(1) **Appearance**: we used an 2D image of the remote user's face to create the 3D avatar's face, so they have the same facial appearance (Figure 4).
(2) **Actions**: the avatar is rigged to respond to the remote user's actions. The head pans and tilts as the remote user's head does, the mouth flaps in time with speech, and blink and idle animations are included to give the avatar some natural movement. The hands and arms are articulated to move as the user moves the controllers around. So where the remote user looks or points maps to where the avatar looks and points. The avatar's movements in both the local and remote views are synchronized (Figure 5). Finally, driving the robot triggers a full body walk animation.

This full-body avatar is meant to heighten the local user's sense that the remote user is present in the space with them. At the same time, it is also intended to increase the remote user's sense that *they* are present in the local space, through allowing them to see in first-person their own avatar body (which they would hopefully identify with) surrounded by and immersed in a 360° view of the space.

Figure 4: Avatar's appearance is made from the remote user's face.

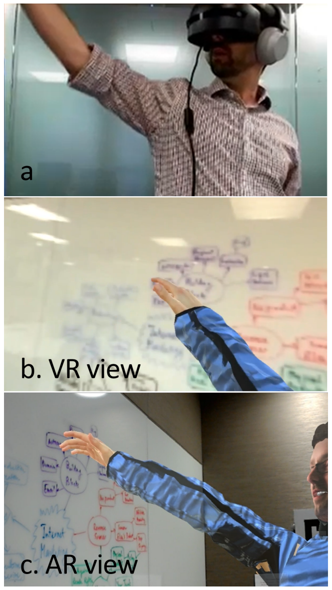

Figure 5: The remote user pointing to an area on white board. **a.** Remote (VR) user's action. **b.** Remote (VR) user's view. **c.** Local (AR) user's view.

***Robot**:* The telepresence robot has fiducial markers for tracking [14] and a 360° camera attached to it. The markers are tracked by the HoloLens app to overlay the avatar to the robot in the AR view. In this version we used front and rear markers to enable to avatar to be facing in the same direction as the robot. This worked reasonably well for this first version, but future versions might include more markers for smoother tracking. The 360° camera is used to stream live 360° video to the remote user's view. We used the built-in audio and driving capabilities of the telepresence robot.

Although we used a telepresence robot with a screen so that we could run a comparison between standard robotic telepresence and VROOM (to be reported in a future paper), the screen would be unnecessary in a space where all local users wore a headset. Thus, a future iteration could use any driveable robot with a 360° camera on a pole reaching head-height.

Remote Space

Remote User (VR) Interface: The remote user wears a Windows Mixed Reality headset displaying the 360° view from the perspective of the robot in the local space. The user also holds a Windows Mixed Reality controller in each of their hands. They use the thumbstick to control the driving direction of the robot. When the remote user talks, looks around, and uses their hands to point in a direction, these movements synchronize with the avatar's movements.

Implementation

In the local space, we used a Microsoft HoloLens (version 1) AR headset, running a Unity application that we built. This application tracks the robot, and overlays the avatar onto it. The avatar's head is made from a front-facing photo of the user, using the *Avatar Maker Pro* Unity library [38], and attached to an animated human-body model available as a standard asset in Unity. To track the robot, the HoloLens app uses the HoloLensARToolKit library [1,26] to track fiducial marker patterns that we printed and placed on the robot (Figure 7). The robot we use is a BeamPro telepresence robot [35]. On this robot, we attached a RICOH Theta V 360° camera [39], connected to a small laptop attached to the base of the robot. This laptop runs another application that streams the 360° video from the camera to the VR application running on the remote side. On the remote side, we implemented the VR application with Unity and using an HP Windows Mixed Reality headset and controller set [40] connected to a Windows desktop PC. This application displays the 360° live video in the headset, and a first-person view of the avatar. The VR application also sends the remote user's head orientation and hand position data to the AR application via HTTP polling. In addition, we implemented another application, running on the PC in the remote space, allowing the user to drive the robot using the thumbsticks on the Windows Mixed Reality hand controllers. This app sends the controller commands to the Beam's normal controller app [41].

Usage Scenario

Amy is a design director in a motorcycle manufacturing company. She is located in Seattle, USA, but she has teammates in Shanghai, China. With VROOM, she can be autonomously present in the Shanghai studio. She has a virtual 'key to the door' of the Shanghai office, engaging with the office on her own timetable and without the need to organize meeting with a particular Shanghai colleague. Her colleagues, each wearing a HoloLens, can see her avatar present in any room,

Figure 6: The remote user can express meaning with gazing and pointing.

Figure 7: The robot with the marker pattern (bottom) and 360° camera (top).

moving around the building, and if they call to her they can see her look back over her shoulder. Amy can 'walk' around the studio to check project progress from team to team, hold a 1:1s in her manager's office, and engage in ad hoc 'watercooler' conversations with people she comes across. In a specific design session, one team shows five life size clay maquettes of new motorcycle designs. Amy and her colleagues are all able to move around the room discussing the designs, huddling around each model and pointing at various design elements. As people move, Amy is always able to know whether others are looking at her, where others are looking, and can also direct her attention to anything in the room's context. At the end of the session, as everyone exits the room and goes back to their open office space, Amy is able to continue conversations with a couple of colleagues on the move. At the end of that days' visit, Amy docks the robot ready for another user. An hour later Red, who works in the Brisbane, Australia, office, calls into the robot and can move around like Amy. Although the robot itself is identical, all of Robert's Shanghai colleagues are able to know he is there at a glance because they can see that his life-size avatar is different to Amy's.

Future Work

We are interested in understanding how VROOM affects social and spatial presence for remote and local users, and how users interact through VROOM in comparison to standard robotic telepresence. The specific questions we are interested in include:

- How does VROOM affect users' collaborative and social interactions, compared to standard robotic telepresence?
- How do remote users make use of VROOM to explore and make sense of the environment?
- How do local users understand and perceive the remote user while using VROOM?

To expand on VROOM further, we are interested in providing the local and remote users with a *shared mixed-reality workspace*, (similar to [37]). That is, both remote and local users would be able to spawn virtual objects (such as 3D models, documents, etc.) that they could pin to locations in the real environment, and both users would be able to see and interact with the objects. The local user would see the objects overlaid on the real environment through the AR headset, while the remote user would see the same objects but in VR. This could be applied in areas like home planning (e.g., discussing furniture choices in an apartment) or remote education/training (e.g., trainer and trainee adding virtual arrows, notes, tags in the training space).

Lastly, although we only illustrated one local user and one remote user, VROOM can also be used in scenarios with multiple local and remote users. More explorations can be done looking at how multiple users interact with each other, how to have multiple remote users see each other's avatars, and how to reduce cost (e.g. with cheap remotely-controlled robots, instead of telepresence robots).

Acknowledgements

We thank our fellow Microsoft colleagues who have helped us tremendously with this work: James Scott, Xu Cao, He Huang, Minnie Liu, Zhao Jun, Matthew Gan, and Leon Lu.

References

1. Ehsan Azimi, Long Qian, Nassir Navab, and Peter Kazanzides. 2018. Alignment of the Virtual Scene to the 3D Display Space of a Mixed Reality Head-Mounted Display. *arXiv preprint arXiv:1703.05834*.
2. Omid Fakourfar, Kevin Ta, Richard Tang, Scott Bateman, and Anthony Tang. 2016. Stabilized Annotations for Mobile Remote Assistance. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM, 1548–1560.
3. Diana Fonseca and Martin Kraus. 2016. A Comparison of Head-mounted and Hand-held Displays for 360° Videos with Focus on Attitude and Behavior Change. *Proceedings of the 20th International Academic Mindtrek Conference*, ACM, 287–296.
4. Susan R. Fussell, Leslie D. Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer, and Adam D. I. Kramer. 2004. Gestures over Video Streams to Support Remote Collaboration on Physical Tasks. *Hum.-Comput. Interact.* 19, 3: 273–309.
5. Steffen Gauglitz, Cha Lee, Matthew Turk, and Tobias Höllerer. 2012. Integrating the Physical Environment into Mobile Remote Collaboration. *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services*, ACM, 241–250.
6. Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. World-stabilized Annotations and Virtual Scene Navigation for Remote Collaboration. *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, ACM, 449–459.
7. William W. Gaver, Abigail Sellen, Christian Heath, and Paul Luff. 1993. One is Not Enough: Multiple Views in a Media Space. *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, ACM, 335–341.
8. Carl Gutwin and Reagan Penner. 2002. Improving Interpretation of Remote Gestures with Telepointer Traces. *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work*, ACM, 49–57.
9. Yasamin Heshmat, Brennan Jones, Xiaoxuan Xiong, et al. 2018. Geocaching with a Beam: Shared Outdoor Activities Through a Telepresence Robot with 360 Degree Viewing. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ACM, 359:1–359:13.
10. Hiroshi Ishii and Minoru Kobayashi. 1992. ClearBoard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, 525–532.
11. Steven Johnson, Irene Rae, Bilge Mutlu, and Leila Takayama. 2015. Can You See Me Now?: How Field of View Affects Collaboration in Robotic Telepresence. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ACM, 2397–2406.
12. Brennan Jones, Anna Witcraft, Scott Bateman, Carman Neustaedter, and Anthony Tang. 2015. Mechanics of Camera Work in Mobile Video Collaboration. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ACM, 957–966.
13. Shunichi Kasahara and Jun Rekimoto. 2015. JackIn Head: Immersive Visual Telepresence System with Omnidirectional Wearable Camera for Remote Collaboration. *Proceedings of the 21st ACM*

Symposium on Virtual Reality Software and Technology, ACM, 217–225.

14. Anuroop Katiyar, Karan Kalra, and Chetan Garg. 2015. Marker based augmented reality. *Advances in Computer Science and Information Technology (ACSIT)* 2, 5: 441–445.
15. David Kirk and Danae Stanton Fraser. 2006. Comparing Remote Gesture Technologies for Supporting Collaborative Physical Tasks. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 1191–1200.
16. Sven Kratz, Daniel Avrahami, Don Kimber, Jim Vaughan, Patrick Proppe, and Don Severns. 2015. Polly Wanna Show You: Examining Viewpoint-Conveyance Techniques for a Shoulder-Worn Telepresence System. *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI 2014), Industrial Case Studies*, ACM.
17. Sven Kratz, Don Kimber, Weiqing Su, Gwen Gordon, and Don Severns. 2014. Polly: Being there through the parrot and a guide. *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, ACM, 625–630.
18. Robert E. Kraut, Susan R. Fussell, Susan E. Brennan, and Jane Siegel. 2002. Understanding effects of proximity on collaboration: Implications for technologies to support remote collaborative work. *Distributed work*: 137–162.
19. Zhengqing Li, Shio Miyafuji, Toshiki Sato, Hideki Koike, Naomi Yamashita, and Hideaki Kuzuoka. 2018. How Display Shapes Affect 360-Degree Panoramic Video Communication. *Proceedings of the 2018 Designing Interactive Systems Conference*, ACM, 845–856.
20. Zhengqing Li, Shio Miyafuji, Erwin Wu, Hideaki Kuzuoka, Naomi Yamashita, and Hideki Koike. 2019. OmniGlobe: An Interactive I/O System For Symmetric 360-Degree Video Communication. *Proceedings of the 2019 on Designing Interactive Systems Conference*, ACM, 1427–1438.
21. Carman Neustaedter, Gina Venolia, Jason Procyk, and Daniel Hawkins. 2016. To Beam or Not to Beam: A Study of Remote Telepresence Attendance at an Academic Conference. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, ACM, 418–431.
22. Kenton O'Hara, Alison Black, and Matthew Lipson. 2006. Everyday Practices with Mobile Video Telephony. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 871–880.
23. Eric Paulos and John Canny. 2001. Social Tele-Embodiment: Understanding Presence. *Autonomous Robots* 11, 1: 87–95.
24. Thammathip Piumsomboon, Gun A. Lee, Jonathon D. Hart, et al. 2018. Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ACM, 46:1–46:13.
25. Thammathip Piumsomboon, Gun A. Lee, Andrew Irlitti, Barrett Ens, Bruce H. Thomas, and Mark Billinghurst. 2019. On the Shoulder of the Giant: A Multi-Scale Mixed Reality Collaboration with 360 Video Sharing and Tangible Interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ACM, 228:1–228:17.
26. Long Qian. 2019. *qian256/HoloLensARToolKit*. .

27. Irene Rae and Carman Neustaedter. 2017. Robotic Telepresence at Scale. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ACM, 313–324.
28. Irene Rae, Leila Takayama, and Bilge Mutlu. 2013. In-Body Experiences: Embodiment, Control, and Trust in Robot-Mediated Communication. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, 1921–1930.
29. Irene Rae, Gina Venolia, John C. Tang, and David Molnar. 2015. A Framework for Understanding and Designing Telepresence. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ACM, 1552–1566.
30. Banu Saatçi, Roman Rädle, Sean Rintel, Kenton O'Hara, and Clemens Nylandsted Klokmose. 2019. Hybrid Meetings in the Modern Workplace: Stories of Success and Failure. *Collaboration Technologies and Social Computing*, Springer International Publishing, 45–61.
31. Anthony Tang, Omid Fakourfar, Carman Neustaedter, and Scott Bateman. 2017. Collaboration with 360° Videochat: Challenges and Opportunities. *Proceedings of the 2017 Conference on Designing Interactive Systems*, ACM, 1327–1339.
32. Lillian Yang, Brennan Jones, Carman Neustaedter, and Samarth Singhal. 2018. Shopping Over Distance Through a Telepresence Robot. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW: 191:1–191:18.
33. Lillian Yang, Carman Neustaedter, and Thecla Schiphorst. 2017. Communicating Through A Telepresence Robot: A Study of Long Distance Relationships. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, ACM, 3027–3033.
34. Double Robotics - Telepresence Robot for Telecommuters. Retrieved December 17, 2019 from https://www.doublerobotics.com/.
35. Suitable Technologies Beam Smart Presence System. Retrieved March 18, 2016 from https://suitabletech.com/beam/.
36. R&D Lab - Microsoft Research. Retrieved December 21, 2019 from https://suitabletech.wistia.com/medias/eg82wq4x5b?embedType=iframe&seo=false&videoWidth=560?wtime=0.
37. Spatial - Collaborate from anywhere in Augmented Reality. Retrieved January 4, 2020 from https://spatial.is/.
38. Avatar Maker Pro - 3D avatar from a single selfie - Asset Store. Retrieved January 4, 2020 from https://assetstore.unity.com/packages/tools/modeling/avatar-maker-pro-3d-avatar-from-a-single-selfie-134800.
39. Product | RICOH THETA V. Retrieved December 21, 2019 from https://theta360.com/en/about/theta/v.html.
40. HP Windows Mixed Reality Headset | HP® Official Site. Retrieved December 21, 2019 from https://www8.hp.com/us/en/campaigns/mixedrealityheadset/overview.html.
41. Help Center - Desktop App. *Beam*. Retrieved December 21, 2019 from https://suitabletech.com/support/helpcenter/desktop-app-full-listing.