Adventures in Hologram Space: Exploring the Design Space of Eye-to-eye Volumetric Telepresence

Rafael Kuffner dos Anjos[1,2,*], Maurício Sousa[2], Daniel Medeiros[1,2], Daniel Mendes[2], Mark Billinghurst[3], Craig Anslow[4], Joaquim Jorge[2]
[1]CMIC, Victoria University of Wellington, [2]INESC-ID / Universidade de Lisboa, [3]ITMS, University of South Australia, [4]School of Engineering and Computer Science, Victoria University of Wellington
[*]rafael.dosanjos@vuw.ac.nz

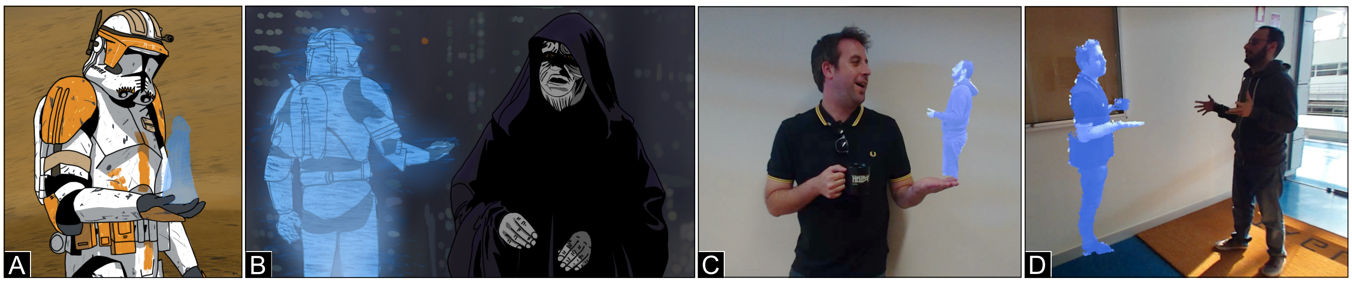

Figure 1: Volumetric projection-based telepresence in fictional visions of the future do not guarantee eye-contact. For instance, the character in (A) is looking down, while the character in (B) is looking forward, thus, their eye gaze does not match. We present a design space where (C) and (D) can communicate with eye-contact.

ABSTRACT

Modern volumetric projection-based telepresence approaches are capable of providing realistic full-size virtual representations of remote people. Interacting with full-size people may not be desirable due to the spatial constraints of the physical environment, application context, or display technology. However, miniaturizing remote people can create eye-gaze matching problems. Eye-contact is essential to communication as it allows for people to use natural nonverbal cues and improves the sense of "being there". In this paper we discuss the design space for interacting with volumetric representations of people and present an approach to dynamically manipulate scale, orientation and position of holograms that guarantees eye-contact. Results garnered from a 14 participant test of our Augmented Reality prototype show improved collaboration and mutual participant awareness.

CCS CONCEPTS

• **Human-centered computing → Computer supported cooperative work**; *User centered design*;

VRST, 2019

ACM ISBN 123-4567-24-567/08/06…$15.00
https://doi.org/10.475/123_4

KEYWORDS

Holograms, Volumetric Projection, Augmented Reality, Eye-to-eye

ACM Reference Format:
Rafael Kuffner dos Anjos[1,2,*], Maurício Sousa[2], Daniel Medeiros[1,2], Daniel Mendes[2], and Mark Billinghurst[3], Craig Anslow[4], Joaquim Jorge[2]. 2019. Adventures in Hologram Space: Exploring the Design Space of Eye-to-eye Volumetric Telepresence. In *Proceedings of VRST*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

In this paper we report on research into sharing gaze cues in volumetric teleconferencing applications. Virtual volumetric projections (VVPs), popularly known as holograms[1] of remote people, are a familiar image in fictional visions of the future. They have been widely used in sci-fi pop culture (e.g. *Forbidden Planet*, *Star Wars*, *Total Recall*, *Matrix Reloaded*) as an example of future communication technology. As opposed to phone calls or 2D video conferencing systems, VVPs can offer a full body portrayal of remote people [3, 14] and create an approximation of a co-located experience [15]. This allows the sharing of non-verbal visual cues including gaze, posture, proxemics and deitic gestures in addition to normal speech and facial expressions [4, 6].

Different technologies have been proposed to achieve volumetric projections, including: see-through HMDs (head-mounted displays) [14], 3D stereoscopic displays [3], spinning 2D displays [10], perspective-corrected projections [16], balls of super-heated plasma in 3D space [11] or particle systems combined with lasers [18]. Among the proposed technologies, see-through HMDs stand out

[1]Not to be confused with the sophisticated printing technology that has already laid claim to the term hologram [17].

as being those that more closely resembles the science fiction depictions of holographic remote communication, and recent work has successfully applied this approach to create co-located experiences [14].

However, holograms for remote conferencing do not address one primordial aspect of face-to-face communication, *gaze-matching*. This is where one person's gaze in a conversation matches that of their conversation partner. Take the interaction depicted in Figure 1 as an example. In Figure 1A, the *soldier* has a small virtual projection of the remote *hooded figure* on his hand, forcing him to look down. In contrast, in Figure 1B, the *hooded figure* sees a projection of the *soldier* that allows him to look him directly forward. From the *hooded figure's* point of view, the *soldier* is not talking to him, but to his own hand. This exact same problem happens in current hologram-based teleconferencing applications, not promoting eye-contact (EC) between both speakers, unless when dealing with talking heads [5, 7, 13] or using real-size projections [14, 16]. When miniaturization of remote people is necessary, eye-contact is lost.

In the present work, we address the gaze mismatch problem for holographic-based communication: how to manipulate the position, orientation and scale of a remote person in order to guarantee eye-contact between both speakers. We propose a design space to promote eye contact in different scenarios and conditions, presenting a real-time approach to manipulate both local and remote holograms. Using the gaze direction from each user towards their visualized hologram, we transform the virtual volumetric projection in order to guarantee eye-contact.

To demonstrate this, we developed an augmented reality-based prototype [2] to assess the correct functioning of the algorithm. Therefore, the main contributions of the paper are: 1) detailed discussion of the design space and 2) an algorithm to achieve gaze matching.

In the following, we discuss eye-contact in remote communication, its importance and challenges, describe the proposed hologram-space, and the implemented prototype to validate it. Finally, we discuss the implications of this research for future telepresence-based communications.

2 EYE-CONTACT IN REMOTE COMMUNICATION

Previous research [1, 2] describes in detail the different benefits EC offers to human communication. This sentiment is shared by previous work on teleconferencing, citing eye contact as a key part of any communication system [9]. Moreover, research on virtual avatars showed that informed eye contact improved the perceived sense of presence [8], which is our goal when representing remote people in such communication systems.

Argyle and Dean [2] describe functions and determinants of EC which can have a direct impact in creating the desired type of communication. While there are extensive reviews of the functions and purposes of EC in human relationship, we can highlight four key points that can have a direct impact on developing teleconferencing applications.

Turn-taking: when EC is broken it may indicate a third party is present, meaning the conversation partner is not ready to receive the message. This is relevant for remote-collaboration, given that one is not co-located with the other speaker having no awareness of the environment and possible distractions. Moreover, studies show that there is more EC when one is listening instead of speaking, and people glance up and away before and after starting to speak.

Information-seeking: Assessing how the transmitted message is received by the partner. A wide array of emotions can be detected through the gaze direction of the other speaker. Naturally, this is interleaved with individual differences regarding the amount of EC utilized during conversations, which can differ due to culture, gender, or disabilities.

Cooperation, Likeness, and social status: Different levels of hierarchy, submission, or intimacy can be expressed through the amount of EC established. EC is also higher if both elements are cooperating in a task instead of competing, and is also related to the likeness and intimacy one speaker has with the other.

Nature of the topic: there is more EC when less personal topics are discussed and when the content is straightforward. Correct EC can quickly inform speakers about the nature of the discussion.

Only by promoting these four points in holographic-based teleconferencing applications we will be able to value these experiences the same as physically-present communications, and apply them to novel scenarios (e.g. job interviews, dating, interrogation, discussion and brain storming). As mentioned previously, current teleconferencing applications either do not correctly promote EC between speakers, or do so by strict restrictions on the practical applications of the systems (e.g. do not support miniaturization or movement of the speakers). The following discussion will motivate and guide the design of unrestricted holographic-based applications where EC is guaranteed.

3 EYE-TO-EYE HOLOGRAM SPACE

Given a face-to-face remote interaction scenario where holograms are being used to display both interlocutors' remote conversation partner, a generic scenario can be described in order to clearly discuss the challenges posed to guarantee eye contact between the participants. This can be seen in Figure 2A.

In the described scenario, the main elements of the hologram spaces A and B are S_A and S_B, which represent each of the speakers. P_A and P_B, represent the position where the holograms$S_B \rightarrow A$ and $S_A \rightarrow B$ will be displayed, respectively. Finally, we assume two main areas F_A and F_B, representing the floor of the space where S_A and S_B are reliably tracked and captured by the chosen technology.

Our goal is to guarantee eye-contact between both interlocutors without making the hologram float above the positions P_A and P_B where the display technology is set, thus only manipulating the 3D representation's orientation, scale, and position in the XZ plane defined by areas F_A and F_B.

3.1 Challenges

Given the described scenario and its components, the real elements in space A may differ from B in their size, positions, and relation to each other. This section will discuss each one of the challenges that arise when this happens and their possible solutions, which will incrementally contribute to a final approach.

[2]Figures of our prototype depict the remote person as blue to be more clearly distinguished. Such effect was not present in the prototype.

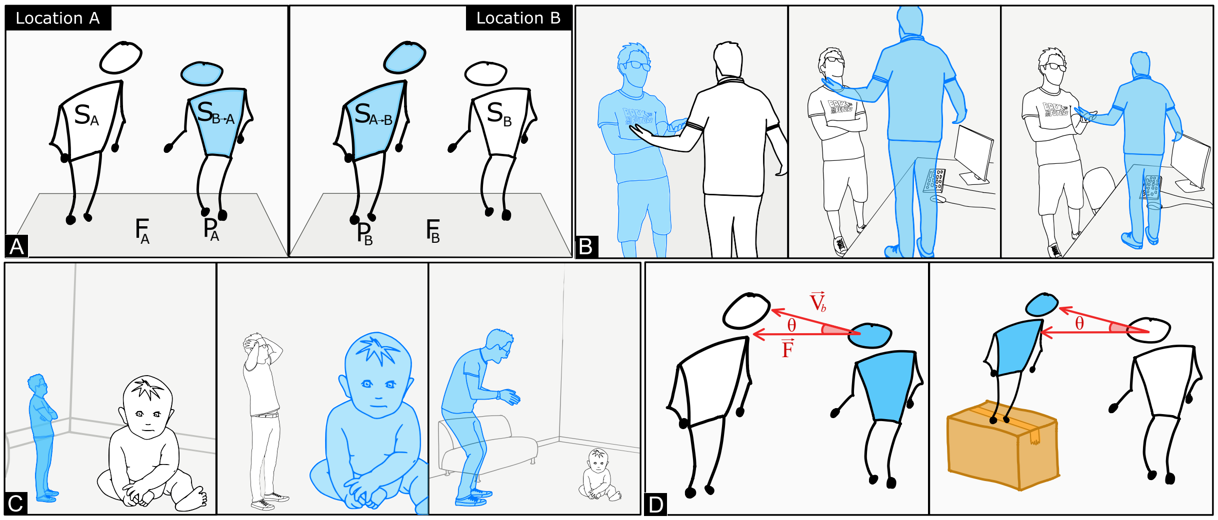

Figure 2: Definition of hologram space (A), Problems A B and C, with the final proposed view-matching algorithm representation (B to D). $\vec{V_b}$ is the view vector from H_b to S_a, and $\vec{F}$ forward vector for H_b

3.1.1 Difference between surfaces. Given two hologram spaces, one aspect that will be highly variable between scenarios is the position where the holograms will be displayed (P_A, P_B). Be it due to different display technologies, difference in height between desks supporting such displays, or the context where this technology is being used. For this first scenario, we assume equal heights between speakers S_A and S_B. A simple example can be seen in Figure 2A, where P_A is located in the floor, and P_B on top of a desk. If the holograms are rendered using their real-world scale, speakers will naturally break eye contact, as S_A will be looking straight ahead, and S_B will be looking up. One solution in this case, is to downscale $S_{A\rightarrow B}$, and eye-contact is re-established.

3.1.2 Height differences: "Big Baby" effect. In a two way conversation, it is also likely that both participants have different heights. Lets consider again the previously presented scenario, where P_A differs from P_B, their relative position to the visualized hologram is the same, but both speakers have different heights. This would mean that in order to maintain eye-contact between both speakers, we would also need to upscale $S_B \rightarrow A$. If the height difference was small, it would not severely affect the holographic representation. However, lets consider the situation in Figure 2C, where a parent S_A is talking to his baby S_B. In order to maintain eye contact, we could downscale $S_{A\rightarrow B}$ to S_B's height, and then upscale $S_{B\rightarrow A}$ to S_A's height. Although eye-contact would be maintained, we would have a very unrealistic representation of the baby, which we call the "Big Baby" effect.

This up-scaling transformation would guarantee that both speakers would be at an equal eye height, but would also end up causing a mismatch in the perceived scale of $S_{B\rightarrow A}$ by S_a, which could mean that eye contact would be broken. Moreover, as mentioned by Sherdoff and Noesel [17], enlarged holographic representations of one of the speakers can be interpreted as a sign of superior hierarchy. Social hierarchy would not be a problem in the father baby scenario, but could be an issue during a meeting between two co-workers with a considerable height difference or if one of them is seated.

This scenario indicates that the scaling factor applied to each one of the speakers must consider their height difference, so that only $S_{A\rightarrow B}$ is down-scaled as much as needed to keep the height difference between him and S_B, as indicated by Figure 2C. Such transformation would ensure that eye-contact is not lost during the communication.

3.1.3 Position related to the surfaces. Finally, let us consider the fact that people move around the holographic display, while still inside F_A and F_B. Assuming the camera system that captures S_A and S_B knows their relative position from the center of the tracking area, it is a straightforward process to keep their holographic representation always in the center of P_A and P_B, regardless of their location.

Another question that can be posed is of their bodily orientation. In order to guarantee eye-contact between both speakers, the holograms should be facing each other. Moreover, if one decides to turn their backs from the hologram they are talking to, this should be correctly represented remotely, so eye-contact is not artificially created. Each representation can be rotated according to the angle described by $\angle\vec{E_b}\vec{E_h}$, where $\vec{E_b} = S_r - P_r$ is the gaze direction of the remote speaker S_r when looking at his hologram at P_r, and $\vec{E_h} = P_l - S_l$ the desired gaze direction of local hologram at P_l in the direction of local speaker S_l.

Finally, the proposed scaling in the previous scenario was based solely on height differences. Naturally, if the speaker's position relatively to the hologram is different, so is the angle from which one speaker looks to the other. In order to maintain eye contact, we must scale each holographic representation according not only to the height difference, but matching the viewing angle of both speakers. Figure 2D describes the applied scaling transformation, where the viewing angle θ is used to calculate the proper height of the remote speaker.

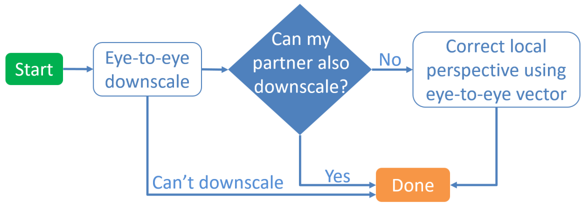

Figure 3: Flowchart summarizing the scaling component of our algorithm.

3.2 Scaling Algorithm

To deal with aforementioned challenges, we propose an algorithm to automatically scale holographic representations in order to ensure eye-contact. It runs locally for each one of the hologram spaces, requiring access to all of the previously mentioned variables on both sides (S_A, S_B, P_A, P_B, F_A and F_B).

A simplified version of the scaling component, which is the most key aspect of the design space, can be seen in Figure 3, which can be summarized by the following three scenarios:

The simplest case: If we are able to downscale (i.e. hologram is taller than local user as in Figure2B) the remote user's representation locally and know that his/her new scale will enable the remote speaker to downscale our representation to provide eye-contact, we end the algorithm.

A second case: If we are able to downscale the remote user's representation locally, but we know that the remote user will have to upscale our representation to provide eye contact, we use the viewing vector-based correction to calculate a local scale that will not force the remote participant to upscale.

The final case: If we are forced to upscale locally to provide eye contact, we just use the current scale we have for the remote person, since they will downscale using the viewing vector on their end to provide eye-contact.

4 VALIDATION AND DISCUSSION

We implemented a Unity3D-based[3] prototype to validate our proposed gaze-matching approach, showing that communication is possible maintaining eye contact throughout different workspace settings. We used the *Creepy Tracker Toolkit* [19] to capture and broadcast a point cloud representation of each participant, which was then rendered through the Meta2[4] AR headset as a hologram. Headset position was obtained using OptiTrack markers on the HMDs, as well as the hologram position. The prototype can be seen in Figure 4. For this matter, we had pairs of subjects communicate using our prototype with a given riddle to be solved, followed by a semi-structured interview. Our experiment had 7 pairs of participants who were acquainted to each other, with ages from 23 to 45 years old (average 27), 9 male and 5 female. Most participants had rarely or never used augmented reality setups, but the large majority frequently was familiar with video-chat systems.

Participants reported that our described hologram space promotes EC between remote people, enabling seamless turn-taking, and rendering the remote participant at an adequate size to allow

[3]Unity3D: http://unity3d.com
[4]Meta2 Glasses: https://www.metavision.com

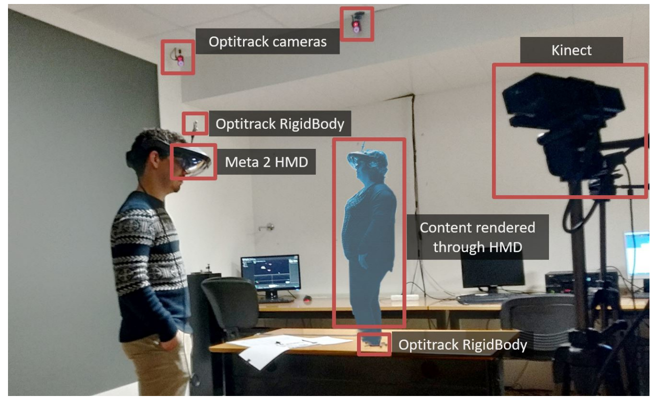

Figure 4: Full setup for the prototype. Remote user in blue just to illustrate the concept. Actual rendering had realistic colors.

face-to-face communication including body language expression. During the experiments, we observed that the participants' behaviour was adequate to a joint problem solving task. They shared less EC during information sharing phase, and a higher amount when engaged in discussion, specially when posing questions or in humorous moments. Overall, participants had no difficulty adapting to this novel type of communication, being able to consistently locate their partner, and solve the proposed riddles.

The use of an HMD made some participants misinterpret the existence of eye-contact, as they did not directly see their conversational partner's eyes, as pointed by Orts et al. [14]. Alternative types of visualization (specialized displays, techniques [3, 10, 16], or real hologram rendering [11, 18]) can be used to overcome it. Future research must adapt these visualization techniques to be as flexible as AR regarding visualization position. Alternatively, work on facial reconstruction to remove HMD's can be included in this type of application [21] as well as eye gaze correction [12, 20].

5 CONCLUSIONS AND FUTURE WORK

In this paper, we address the gaze matching problem introducing a design space for future Augmented Reality conferencing approaches using volumetric projections of remote people. We contribute with an approach to dynamically scale remote people by considering differences in height, positioning, orientation and the characteristics of the surrounding space. We validated this concept by presenting an AR-based prototype where eye-contact is guaranteed, discussed the observed results of our experiment, and what they indicate us about where should future work in this design-space be focused. The most recent advances on rendering people as avatars provide a good foundation for researchers to focus on the next AR challenge, that is, on how volumetric projections of remote people can react and relate to the physical environment where they are depicted.

ACKNOWLEDGEMENTS

This project was supported by the Entrepreneurial University Programme funded by TEC, also by FCT through grants IT-MEDEX PTDC/EEISII/6038/2014 and UID/CEC/50021/2019.

REFERENCES

[1] Michael Argyle and Mark Cook. 1976. Gaze and mutual gaze. (1976).

[2] Michael Argyle and Janet Dean. 1965. Eye-contact, distance and affiliation. *Sociometry* (1965), 289–304.

[3] Stephan Beck, Andre Kunert, Alexander Kulik, and Bernd Froehlich. 2013. Immersive group-to-group telepresence. *IEEE Transactions on Visualization and Computer Graphics* 19, 4 (2013), 616–625.

[4] Steve Benford, John Bowers, Lennart E. Fahlén, Chris Greenhalgh, and Dave Snowdon. 1995. User Embodiment in Collaborative Virtual Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '95)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 242–249. https://doi.org/10.1145/223904.223935

[5] Bill Buxton. 2009. *Mediaspace – Meaningspace – Meetingspace*. Springer London, London, 217–231. https://doi.org/10.1007/978-1-84882-483-6_13

[6] William Buxton. 1992. Telepresence: Integrating shared task and person spaces. In *Proceedings of graphics interface*, Vol. 92. 123–129.

[7] 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 (Sept. 2004), 273–309. https://doi.org/10.1207/s15327051hci1903_3

[8] Maia Garau, Mel Slater, Simon Bee, and Martina Angela Sasse. 2001. The Impact of Eye Gaze on Communication Using Humanoid Avatars. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*. ACM, New York, NY, USA, 309–316. https://doi.org/10.1145/365024.365121

[9] Hiroshi Ishii and Minoru Kobayashi. 1992. ClearBoard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '92)*. ACM, New York, NY, USA, 525–532. https://doi.org/10.1145/142750.142977

[10] Andrew Jones, Magnus Lang, Graham Fyffe, Xueming Yu, Jay Busch, Ian McDowall, Mark Bolas, and Paul Debevec. 2009. Achieving eye contact in a one-to-many 3D video teleconferencing system. *ACM Transactions on Graphics (TOG)* 28, 3 (2009), 64.

[11] H. Kimura, A. Asano, I. Fujishiro, A. Nakatani, and H. Watanabe. 2011. True 3D Display. *ACM SIGGRAPH 2011 Emerging Technologies, SIGGRAPH'11* (2011). https://doi.org/10.1145/2048259.2048279 cited By 10.

[12] Daniil Kononenko, Yaroslav Ganin, Diana Sungatullina, and Victor Lempitsky. 2017. Photorealistic monocular gaze redirection using machine learning. *IEEE transactions on pattern analysis and machine intelligence* 40, 11 (2017), 2696–2710.

[13] David Nguyen and John Canny. 2005. MultiView: Spatially Faithful Group Video Conferencing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. ACM, New York, NY, USA, 799–808. https://doi.org/10.1145/1054972.1055084

[14] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchyn, Cem Keskin, and Shahram Izadi. 2016. Holoportation: Virtual 3D Teleportation in Real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 741–754. https://doi.org/10.1145/2984511.2984517

[15] Tomislav Pejsa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew Wilson. 2016. Room2Room: Enabling Life-Size Telepresence in a Projected Augmented Reality Environment. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 1716–1725. https://doi.org/10.1145/2818048.2819965

[16] Tomislav Pejsa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew Wilson. 2016. Room2Room: Enabling life-size telepresence in a projected augmented reality environment. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 1716–1725.

[17] Nathan Shedroff and Christopher Noessel. 2012. *Make it so: Interaction design lessons from science fiction*. Rosenfeld Media.

[18] DE Smalley, E Nygaard, K Squire, J Van Wagoner, J Rasmussen, S Gneiting, K Qaderi, J Goodsell, W Rogers, M Lindsey, et al. 2018. A photophoretic-trap volumetric display. *Nature* 553, 7689 (2018), 486.

[19] Maurício Sousa, Daniel Mendes, Rafael Kuffner Dos Anjos, Daniel Medeiros, Alfredo Ferreira, Alberto Raposo, João Madeiras Pereira, and Joaquim Jorge. 2017. Creepy Tracker Toolkit for Context-aware Interfaces. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces (ISS '17)*. ACM, New York, NY, USA, 191–200. https://doi.org/10.1145/3132272.3134113

[20] Justus Thies, Michael Zollhöfer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. 2018. FaceVR: Real-time gaze-aware facial reenactment in virtual reality. *ACM Transactions on Graphics (TOG)* 37, 2 (2018), 25.

[21] Shih-En Wei, Jason Saragih, Tomas Simon, Adam Harley, Stephen Lombardi, Michal Perdoch, Alexander Hypes, Dawei Wang, Hernan Badino, and Yaser Sheikh. [n. d.]. VR Facial Animation via Multiview Image Translation. *ACM Trans. Graph.* ([n. d.]).