Check for updates

Industrial Metaverse

Supporting remote maintenance with avatars and digital twins in collaborative XR environments

Leif Oppermann
Fraunhofer FIT, Sankt Augustin, Germany
leif.oppermann@fit.fraunhofer.de

Florian Buchholz
Fraunhofer FIT, Sankt Augustin, Germany
florian.buchholz@fit.fraunhofer.de

Yücel Uzun
Fraunhofer FIT, Sankt Augustin, Germany
yuecel.uzun@fit.fraunhofer.de

ABSTRACT

We present a 5G mixed reality toolbox that supports hands-free remote assistance in industrial settings. It provides mixed reality and virtual reality views for on-site and office workers linked via a shared digital space. Working on actual machines in a real production line, our system uses the actual CAD-data of those machines to provide for a realistic prototyping-environment. We focus on data-scarcity with cloud-services to protect intellectual property, while embracing the possibilities offered by new technology, such as remote rendering over wireless networks. The presented prototype exhibits several key characteristics of an industrial metaverse application.

CCS CONCEPTS

• **Human-centered computing**; • **Mixed / augmented reality (500) Computer systems organization**; • **Cloud computing (300)**;

KEYWORDS

Remote Maintenance, Mixed Reality, VR/AR, Metaverse, 5G, Industry, User

ACM Reference Format:

Leif Oppermann, Florian Buchholz, and Yücel Uzun. 2023. Industrial Metaverse: Supporting remote maintenance with avatars and digital twins in collaborative XR environments. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3544549.3585835

1 INTRODUCTION

Much has been talked about "Metaverse" since 2021 – a vision for a future, collaborative, spatial web that is enabled by XR-technologies, stems from science-fiction literature, and was triggered by the renaming of Facebook Inc. into Meta Platforms in October that year. Its definition is unclear and differs with parties and use-cases. For some parties it is mostly a network of collaborative virtual environments (CVEs) or Virtual Reality (VR), while other emphasize the links to real environments in a "mixed reality", including digital twins and mirror worlds [20]. Still, both camps entail the social aspect of collaborating with other people supported by spatial (3D) interaction technology that is ideally wireless and facilitates remote collaboration.

While the possible existence of the Metaverse might be perceived somewhere between "already there" and "it will never happen", not least due to confusion stemming from big-tech marketing haze, we support an active engagement with this vision and its supporting technologies to shape future workplaces for white- and blue-collar workers. We subscribe to the notion of having different "metaverses" for different scenarios [2] rather than having a single all-encompassing Metaverse that is the Internet [14]. While the latter might partially happen for predominantly entertainment purposes, there are serious drawbacks for company use linked to the platform capitalistic nature of currently dominant Internet companies [10]. The transatlantic dataflow between Europe and the US to support Internet businesses has been enabled by treaties such as "Safe-Harbor" or "Privacy Shield". Yet, different conceptions of data-protection and privacy in the aftermath of 9/11 and the 2013 revelations of Edward Snowden [21] showed a legal divide between Europe and the US. Subsequently, those treaties have been cancelled by the Court of Justice of the European Union when challenged in the historic "Schrems I" (2015) and "Schrems II" (2020) rulings which subsequently led to realignments of transatlantic data protection policies in light of the surveillance ambitions of the Chinese government [17].

Considering these developments, we present an industrial metaverse application that exhibits many of the characteristic aspects of a metaverse and does that in a data-privacy conserving way. It stems from a national research project that aims to produce tangible use-cases such as this "5G mixed reality toolbox" (as is its original name) that could be send around the world. It could be employed in the near future by small and medium enterprises to reap the benefits of remote collaboration supported by different XR-technologies without having to fear immediate loss of their intellectual property such as CAD-data in a legally uncertain international Internet-environment. In this paper we show some of the practical concerns and complexities when implementing such metaverse-like interactions today. We describe the system design, what the prototype does, and briefly discuss it. Due to the late breaking nature of this work, the full evaluation will be covered in a later paper.

2 RELATED WORK

Computer supported cooperative work (CSCW) and mixed reality are long-standing research topics that are now brought together with recent advances in and availability of fitting devices and high-bandwidth, low-latency networks. Monitor, keyboard, mouse and networked computing became commonplace since at least the (later

CHI EA '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9422-2/23/04.
https://doi.org/10.1145/3544549.3585835

aptly named) mother of all demos which was conceived in Engelbart's conceptual framework "Augmenting Human Intellect" [3]. The seminal head-mounted display dates back to 1968 [23]. The first data-glove appeared in 1983 and former Atari-developer Jaron Lanier founded the first VR company in 1984, combining all of the above and conducting government-funded research that paved the way for the first wave of virtual reality of the 1980s and 1990s [16, 29]. This line of VR-, or collaborative virtual environments (CVE) research ran mostly in parallel to classic office-oriented CSCW/groupware research, with the first academic CVE-workshop taking place in 1996 [22]. Windows on the world [5], and the concept of mixed reality [12] allowed substituting various parts of the virtual environments with elements of the real environment, therefore paving the way for mobile mixed reality, or augmented reality applications [1]. This also changed the notion of presence perceived in such systems [26]. The concept of ubiquitous computing defined todays popular form-factors, like tablets, and the general (ubiquitous) availability of human-computer interfaces in the environment of the users [27]. While these ubicomp interfaces are not necessarily three dimensional, or even graphic, the concept has much in common with the aspect of the "metaverse" that makes most of the environment virtual, interactive, and thereby part of the ubiquitous interface.

Going beyond mixed reality interfaces, the linking of actions in reality and in virtual environments to exploit relationships was conceived and demonstrated in projects like Tower [15] that were located at the crossing of the real and the virtual. "New work environments" projects of the European Commission sought to provide the conceptual foundation for future mobile collaborative workplace scenarios, concluding that the provision of cheap telecommunications bandwidth, ubiquitous mobile access, flexible business organisations, societal requirements and new technology will result in the network becoming the workplace for the knowledge worker [19]. Projects from the late 2000s started to link the (white-collar) knowledge worker with blue-collar workers on-site in "unforeseen problem" and remote maintenance scenarios [7]. Over the last decade, smartphone technology became truly ubiquitous, and wearables and head-mounted displays based on that technology matured with increasingly desirable off-the-shelf devices being readily available. This led to a surge of apps and projects using those devices. Consequently, Ens et al. recently summarised that the future of CSCW and mixed reality lies in the combination of these fields [4].

Research in human-computer interaction toolboxes is a broad field that frequently uses custom-built prototypes, which typically can only be operated by their developers. One particular area of research centres on enhancing accessibility for designers through authoring tools that permit content to be linked to specific locations with an underlying infrastructure, thereby democratizing the creative process [6, 8, 9, 13]. A related strand of research focused on packing all required components into a product-like box that would be easy to send around and give into the hands of users [18, 28].

Regarding this papers system, there have been similar systems being developed in the recent years that focused on enabling remote collaborators to view and interact with physical objects in a shared virtual environment [11] or expert instructors to guide and instruct remote users in real-time using mixed-reality telepresence technology [24, 25]. These systems mostly focused on very specific tasks or just basic interactions with the real world in lab(-like) environments. For industrial purposes, there are still practical challenges, such as data privacy and security regarding the 3D-models that are being used, and general workflow issues. In an industrial environment, the emphasis is on creating a simple and robust setup that can be activated quickly without requiring a lengthy setup process involving external sensors. A simple real-time visualization of the real environment is insufficient for repairing or constructing a system that mostly occurs inside a machine. Instead, a complete digital twin of the machine is preferable and can be utilized in a see-through mode.

3 SYSTEM

Our system builds upon the related work in several ways. It offers both mixed reality and virtual reality interfaces that are linked over wireless networks to create a shared space for remote collaboration. In this section, we provide an overview of the system's capabilities, followed by a description of its technical implementation.

3.1 Overview

The goal of our industrial prototype is to enable experts independently of their actual workplace to work together with on-site technicians. Our application makes it possible for them to share their expertise to solve problem situations more effectively. Usually, in case of malfunctions in industrial operations, the biggest time factor is the trip of an expert to the place of the malfunction - which might be anywhere in the world - since troubleshooting by other means like a mere phone call often proves to be too difficult. This also causes not only expensive downtime for the industry but also lots of stress for the experts. Our application provides a "shared world" for such scenarios, where the expert works on a digital version of the machine in VR and the on-site employee, who reported the problem, operates on the real machine. The employee can see a mixed reality view that is shared with the expert, including a digital twin of the machine.

Both users (VR and AR/MR) are represented by avatars and can communicate with each other face-to-face, i.e. avatar-to-avatar, including audio with bone conducting headsets to provide the audio-link in noisy environments. The VR-expert's view includes a live video stream from the on-site employee's perspective. VR-controllers and MR-hand-recognition allow for pointing at, highlighting, annotating, moving, and hiding parts of the virtual machine to best guide the on-site user from remote while leaving his hands free. Figure 1 shows the VR- and MR-users, their avatar representations, and how a video-chat looks like for the VR-user. Figure 2 shows their respective views in the shared space. In this system, the expert has full control of the digital twin, while the employee focuses on the real machine and uses the digital twin and the information from the expert more as a visual guidance system. Due to the potential dangers associated with working on complex machinery in the industrial setting, manually editing the digital world from an MR-perspective could lead to accidents. Therefore, it is important to limit this process or even automate it in the future. Additionally, our system currently has limitations with respect to hardware capabilities for scanning the real world. As a result, the

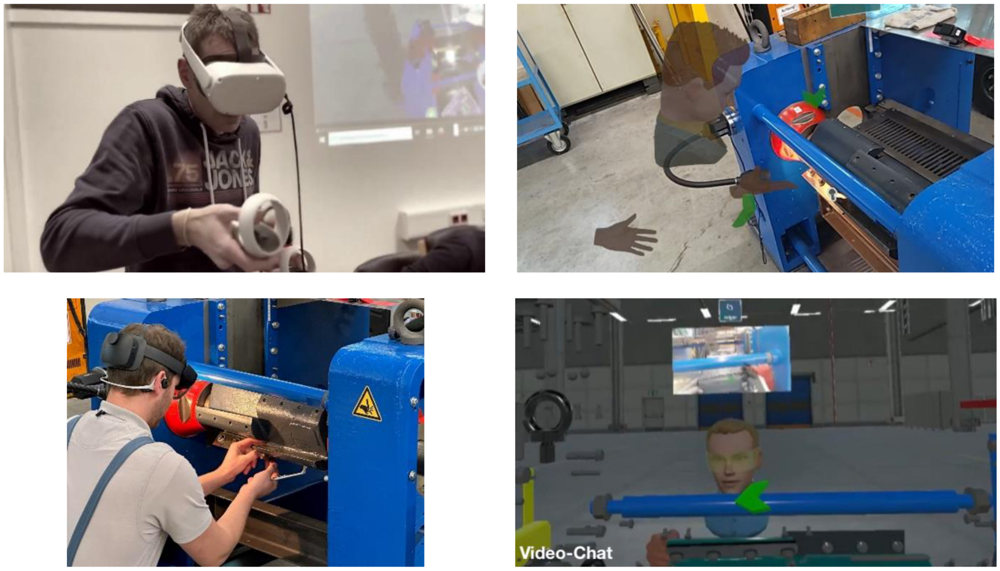

Figure 1: VR- and MR-users (left) and their avatar representations in the shared space

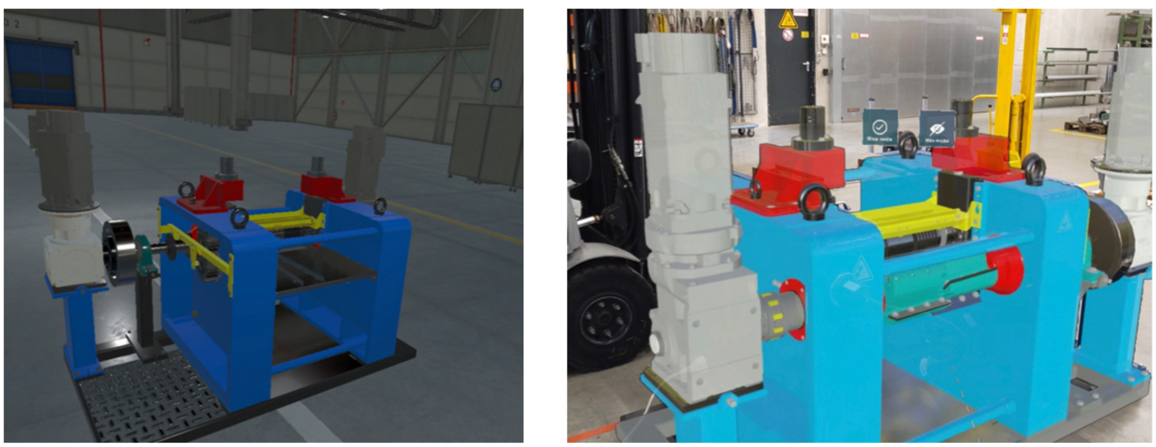

Figure 2: VR- and MR-views of the machine in the shared space

system does not track changes on the real machine and synchronize those changes to the digital twin.

3.2 Description

The system consists of three main components: a Mixed-Reality-headset, in this case a HoloLens 2, a Meta Quest 2 as a VR-Head Mounted Display (HMD) that is connected to a PC for better performance via Oculus Link, and a server-system that connects both devices. With this server, various data can be shared and synchronized between the HoloLens 2 and the Quest 2 (see Figure 3).

In the following, we look at the various services that enable the on-site employee and the remote expert to work together as efficiently as possible. We pay particular attention to both the real-time change of the digital twin and the presence of the respective persons in virtual and real space. Complementary to this, the question of "decision-making power" in the digital space has been relevant for us. That is, how much change to the digital twin should be possible by the MR user and how much by the VR user.

The application is designed to store as little private and contextual information as possible and it acquires all the necessary information at runtime after the authentication to a server. After this, it downloads local rendered models, their metadata and process guidelines that are stored on a server, the so-called "BSCW server" (Basic Support for Collaborative Work).

In order to achieve the real time synchronization and communication, as shown in Figure 3, both HoloLens 2 and Quest 2 are permanently connected to the server system during a session. At the beginning of a session, the devices register with the main server

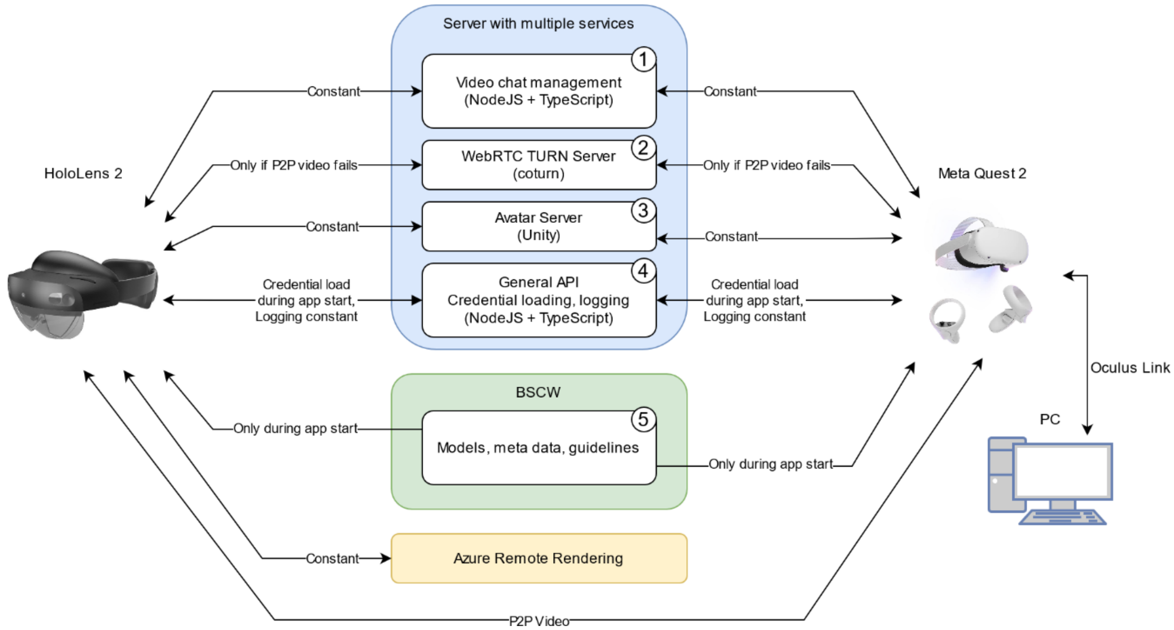

Figure 3: System setup for industrial metaverse application

(see #4 in Figure 3), which matches them and assigns them to related metadata. This server starts also instances a so-called "avatar server" which is used for synchronizing end-user Unity apps. Video-audio chat has been implemented using "Web Real Time Communication" (WebRTC) protocol and this server also manages the signaling process of WebRTC, where different peers negotiate a direct communication path between each other.

In many cases in industrial premises, a direct connection between the two devices is not possible because of the security restrictions imposed due to security concerns or other technical reasons. In these cases, an additional WebRTC TURN server is needed to forward the information for the video chat over the Internet.

The avatar server tracks and synchronizes both the tracking of the people in the room and any changes to the digital twin. As soon as a person makes a change to the digital twin, for example moving a sub-object, this information is forwarded to the other devices. In this way, the status of the digital model always remains the same for all users.

The digital twins used in this project are engineering models used during the production of the said machines. These models are not optimized for rendering real-time on a device with limited processing capacity, like HoloLens 2. Moreover, their overall structure – which is optimized for real assembly processes- causes algorithmic mesh rendering tools to generate unusable variants and optimizing these models by hand is extremely time-consuming process. In order to render these models with higher accuracy and in a less-time-consuming manner, Azure Remote Rendering is used on HoloLens 2.

4 DISCUSSION

Prior to our work, our industrial partners had already utilized Remote Assist on HoloLens 2. This was beneficial, as they possessed a strong understanding of the use case and had already undergone IT-security clearance. Our primary contribution, therefore, is the addition of remote collaboration within the digital twin, which is directly exported from the CAD software. In order to achieve this, numerous challenges needed to be tackled. First, the best ex- and import routes for model-data needed to be found. Then, the high polygon count of these models needed to be rendered at interactive framerates on the available hardware. For example, with one machine used during development, the model had 5 million triangles, but the HoloLens 2 could only locally render 100k at acceptable speeds. So local rendering on that device was ruled out.

Using available highspeed-networks like 5G and WiFi6 enabled remote rendering the machine-data in the cloud and then streaming the resulting frame down to the end-user device. Remote rendering thus moves the computationally intensive tasks from the HoloLens 2 to the cloud and allows using more complex models. We tested at least three different remote rendering solutions and found that they had all different pros and cons. While we cannot provide full details due to NDAs, we can say that one such solution did not actually work with HoloLens 2, albeit the website saying otherwise. Another solution did not to allow to use video-chat and was quite expensive but provided for rendering on own servers. Our third and currently selected solution provides pay-per-use plans and two different polygon count limits (20 million or "hundreds" of millions). One potential disadvantage is the inability to install the application on personal servers. However, on the positive side, this allows the application to move closer to the on-site user in terms of network distance, regardless of their location on earth, which aligns well with the use-case. It could be argued that using professionally managed cloud services is not only easier (e.g., no need for administration), but also more secure than hosting one's own servers. Finally, using remote rendering also provides a neat security improvement, as the confidential model-data will never be stored on the mixed reality device, which can easily get lost or stolen. Only the necessary bitstreams to draw the current frame are transmitted to the device, ensuring that no other data is ever exposed. With this method, it is not possible to reconstruct the source model from these bitstreams. In comparison to similar systems discussed in related work, this approach allows for high-level, detailed 3D models that are necessary in industrial settings. While related use cases may illustrate similar methods in laboratory environments, remote assistance in the industry and working with complex machinery requires a degree of flexibility that enables the addition of new or more detailed models without requiring a deeper understanding of the system architecture.

For the office VR setup, a dedicated rendering PC is connected to the Quest 2 via Oculus Link, which is not an issue as PCs and office/home-office environments are well-matched. Moreover, replacing it with a more capable HMD when needed would be a straightforward process.

Although maintaining data synchronization can be challenging, the current setup works quite well. As mentioned before, tracking changes to the real-world object could be beneficial, allowing for corresponding updates to the digital model. This type of bi-directional sync would align with proposed characteristics of industrial metaverse applications [2], but its implementation could prove difficult and may not necessarily offer significant improvements over the current state of the art.

Rather than only focusing on the technological aspects, the project also analyses how to embed these innovations into existing work processes from a social science perspective and provide for good future working conditions.

ACKNOWLEDGMENTS

This work was funded by the German Federal Ministry for Digital and Transport BMDV as part of the national 5x5G strategy. We thank our project partners from “IndustrieStadtpark” (www.5gtroisdorf.de).

REFERENCES

[1] Ronald Azuma. 1997. A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments* 6, 4 (1997), 355–385. Retrieved from http://www.cs.unc.edu/~azuma/ARpresence.pdf

[2] Florian Buchholz, Leif Oppermann, and Wolfgang Prinz. 2022. There's more than one Metaverse. *i-com* 22, 3 (2022), 313–324. DOI:https://doi.org/10.1515/icom-2022-0034

[3] Doug Engelbart. 1962. Augmenting Human Intellect: A Conceptual Framework. Retrieved February 19, 2016 from http://www.dougengelbart.org/pubs/augment-3906.html

[4] Barrett Ens, Joel Lanir, Anthony Tang, Scott Bateman, Gun Lee, Thammathip Piumsomboon, and Mark Billinghurst. 2019. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies* 131, (November 2019), 81–98. DOI:https://doi.org/10.1016/j.ijhcs.2019.05.011

[5] Steven Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon. 1993. Windows on the world: 2D windows for 3D augmented reality. In *Proceedings of the 6th annual ACM symposium on User interface software and technology* (UIST '93), Association for Computing Machinery, New York, NY, USA, 145–155. DOI:https://doi.org/10.1145/168642.168657

[6] Martin Flintham. 2005. Painting the Town Red: Configuring Location-Based Games by Colouring Maps. In *Advances in Computer Entertainment Technology (ACE)*, Valencia, Spain.

[7] Elke Hinrichs, May Bassanino, Colin Piddington, Gilles Gautier, Farzad Khosrowshahi, Terrence Fern, and Jens Ove Skjærbæk. 2008. *Mobile Maintenance Workspaces: Solving Unforeseen Events on Construction Sites More Efficiently.*

[8] Richard Hull, Ben Clayton, and Tom Melamed. 2004. Rapid Authoring of Mediascapes. In *Ubicomp*, Springer, Nottingham, England, 125–142. Retrieved from http://link.springer.com/chapter/10.1007%2F978-3-540-30119-6_8#page-1

[9] A. Jurgelionis, R. Wetzel, L. Blum, and L. Oppermann. 2013. TOTEM.Scout: A mobile tool for in-situ creation of location-based content. In *Games Innovation Conference (IGIC), 2013 IEEE International*, 89–96. DOI:https://doi.org/10.1109/IGIC.2013.6659153

[10] Jaron Lanier. 2014. Who Owns The Future? - Du bist nicht der Kunde der Internetkonzerne. Du bist ihr Produkt. Penguin, London.

[11] Yongjae Lee, Byounghyun Yoo, and Soo-Hong Lee. 2021. Sharing Ambient Objects Using Real-time Point Cloud Streaming in Web-based XR Remote Collaboration. In *The 26th International Conference on 3D Web Technology*, ACM, Pisa Italy, 1–9. DOI:https://doi.org/10.1145/3485444.3487642

[12] Paul Milgram and Fumio Kishino. 1994. A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information Systems* E77-D, 12 (1994), 1321–1329. DOI:https://doi.org/10.1.1.102.4646

[13] Leif Oppermann, Martin Flintham, Stuart Reeves, Steve Benford, Chris Greenhalgh, Joe Marshall, Matt Adams, Ju Row Farr, and Nick Tandavanitj. 2011. Lessons from touring a location-based experience. In *Pervasive Computing*, Springer, San Francisco, CA, 232–249. Retrieved from https://dl.acm.org/doi/10.5555/2021975.2021996

[14] Tony Parisi. 2021. The Seven Rules of the Metaverse. *Metaverses*. Retrieved January 13, 2022 from https://medium.com/meta-verses/the-seven-rules-of-the-metaverse-7d4e06fa864c

[15] Wolfgang Prinz, Andrew McGrath, Alan Penn, Peter Schickel, and Fredrick Wilhelmsen. 2000. TOWER - Theatre of Work Enabling Relationships. In *eBusiness and eWork*, Madrid.

[16] Howard Rheingold. 1991. Virtual Reality, Exploring the brave new technologies of artificial experience and interactive worlds from Cyberspace to Teledildonics (Book Club Edition ed.). QPD, London.

[17] Marc Rotenberg. 2020. Schrems II, from Snowden to China: Toward a new alignment on transatlantic data protection. *European Law Journal* 26, 1–2 (2020), 141–152. DOI:https://doi.org/10.1111/eulj.12370

[18] Hannu Salmi, Arja Kaasinen, and Veera Kallunki. 2012. Towards an open learning environment via augmented reality (AR): Visualising the invisible in science centres and schools for teacher education. *Procedia-Social and Behavioral Sciences* 45, (2012), 284–295.

[19] Hans Schaffers, Torsten Brodt, Marc Pallot, and Wolfgang Prinz. 2006. *The Future Workspace. Mobile and Collaborative Working Perspectives.* Retrieved from https://www.researchgate.net/publication/36382216_The_Future_Workspace_Mobile_and_Collaborative_Working_Perspectives

[20] John Smart, Jamais Cascio, and Jerry Paffendorf. 2007. Metaverse Roadmap: Pathways to the 3D Web. *Metaverse: a cross-industry public foresight project* (2007). Retrieved from https://www.metaverseroadmap.org/MetaverseRoadmapOverview.pdf

[21] Edward Snowden. 2019. *Permanent Record* (1st Edition ed.). Metropolitan Books, New York.

[22] Dave Snowdon, Elizabeth Churchill, and Jolanda Tromp. 1996. CVE '96 (Collaborative Virtual Environments) Workshop Proceedings, 1996, Snowdon, Dave, Churchill, Elizabeth., Tromp, Jolanda G., (eds.), University of Nottingham, Nottingham, United Kingdom, 19-20 September.

[23] Ivan E. Sutherland. 1968. A Head-mounted Three Dimensional Display. In *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I* (AFIPS '68 (Fall, part I)), ACM, New York, NY, USA, 757–764. DOI:https://doi.org/10.1145/1476589.1476686

[24] Balasaravanan Thoravi Kumaravel, Fraser Anderson, George Fitzmaurice, Bjoern Hartmann, and Tovi Grossman. 2019. Loki: Facilitating Remote Instruction of Physical Tasks Using Bi-Directional Mixed-Reality Telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, ACM, New Orleans LA USA, 161–174. DOI:https://doi.org/10.1145/3332165.3347872

[25] Huayuan Tian, Gun A. Lee, Huidong Bai, and Mark Billinghurst. 2023. Using Virtual Replicas to Improve Mixed Reality Remote Collaboration. *IEEE Trans. Visual. Comput. Graphics* (2023), 1–11. DOI:https://doi.org/10.1109/TVCG.2023.3247113

[26] Ina Wagner, Wolfgang Broll, Giulio Jacucci, Kari Kuutti, Rod Mccall, Ann Morrison, Dieter Schmalstieg, and Jean-Jacques Terrin. 2009. On the Role of Presence in Mixed Reality. *Presence* 18, (August 2009), 249–276. DOI:https://doi.org/10.1162/pres.18.4.249

[27] Mark Weiser. 1991. The Computer for the 21st Century. *Scientific American* 265, 3 (1991), 66–75. DOI:https://doi.org/10.1145/329124.329126

[28] Richard Wetzel, Annika Waern, Staffan Jonsson, Irma Lindt, Peter Ljungstrand, and Karl-Petter Åkesson. 2009. Boxed pervasive games: an experience with user-created pervasive games. In *International Conference on Pervasive Computing*, Springer, 220–237.

[29] Thomas G. Zimmerman, Jaron Lanier, Chuck Blanchard, Steve Bryson, and Young Harvill. 1986. A hand gesture interface device. *SIGCHI Bull.* 18, 4 (May 1986), 189–192. DOI:https://doi.org/10.1145/1165387.275628