Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 128 (2024) 591–596

www.elsevier.com/locate/procedia

34th CIRP Design Conference

Designing Immersive Tools for Expert and Worker Remote Collaboration

Sergio Camilo Medina Galvis[a,*], Damien Mazeas[b], Frédéric Noël[a], and John Ahmet Erkoyuncu[b]

[a]Grenoble INP, Grenoble 38000, France
[b]Cranfield University, Cranfield MK43 0AL, United Kingdom

* Corresponding author. *E-mail address:* sergio-camilo.medina-galvis@grenoble-inp.fr

Abstract

Remote collaboration between on-site workers and remote experts is challenging due to physical limitations and lack of context sharing. Immersive technologies like Augmented Reality (AR) offer innovative interaction methods. However, workers still need wearable or handheld devices, and restricted viewpoints limit remote experts and overall efficiency. This paper aims to design a framework that enables expert presence and improves worker usability, fostering greater collaboration. Our methodology involves using a collaborative robot (cobot) with a camera, projector, smartphone, and a Virtual Reality (VR) interface. The result enables images and 3D captures of the workspace to be shared with the remote expert, who can freely manipulate the cobot to project inspection guidance cues. This autonomy enhances the expert's involvement and frees the worker to focus on other essential tasks. We detail the proposed architecture design and setup. Finally, we evaluate the system using time delay and latency technical analysis.

Keywords: Remote collaboration; Human interfacing; Design methodology, tools and technologies

1. Introduction

Augmented Reality (AR) technologies have gained interest in providing remote collaboration and guidance to workers in various industries [1]. AR helps visualize the real environment with enhanced virtual elements and provides tracking methods to interactively manage worker tasks with information rendered in multiple forms [2]. Different devices can achieve this, including Optical See-Through (OST) with AR head-mounted displays [3,4], Video See-Through (VST) solutions with computers, phones, or tablets, and Spatial AR (SAR) thanks to projection systems in the real environment [3].

Despite the advances in AR-based assistance, the majority of existing solutions necessitate those workers wear or hold devices, a requirement that can be cumbersome and diminish operational efficiency. Commercial AR-based remote assistance tools, such as PTC's Vuforia Chalk[1], offer real-time 3D annotations and object tracking, all overlaid on the live view of an on-site worker's environment, complemented by audio communication systems. However, these tools have a significant limitation: the restricted viewpoint resulting from the worker's need to hold the device can adversely impact the effectiveness of remote assistance. Addressing this challenge, our study introduces a novel approach to AR-based remote assistance, enabling workers to receive guidance without the encumbrance of wearing or holding equipment.

We address the following research question: **How can remote guidance enable worker-expert collaboration without carrying any device?** To address this question, we

[1]https://www.ptc.com/en/products/vuforia/vuforia-chalk

2212-8271 © 2024 The Authors. Published by Elsevier B.V.

10.1016/j.procir.2024.07.055

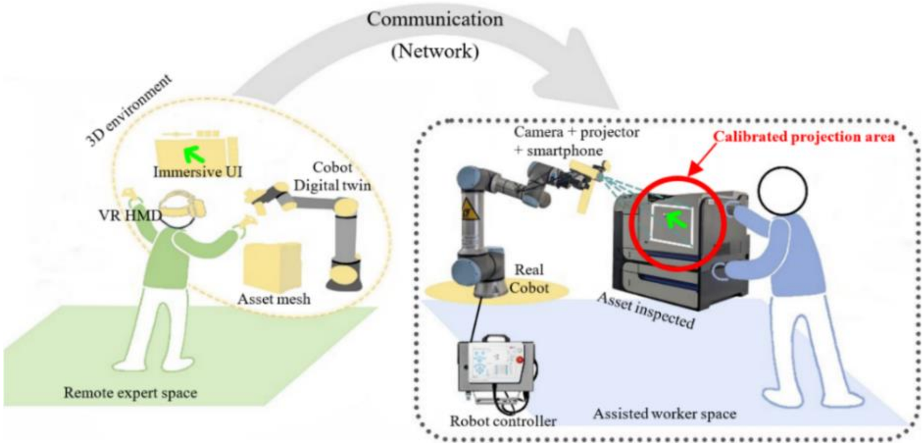

Fig. 1. Proposed setup and approach enabling remote collaboration.

propose an approach that underscores the use of a cobot to facilitate collaborative work safely by mitigating the risk of collisions and other potential hazards. Furthermore, the proposal includes the development of a coherent environment illustrated in Fig. 1 where a cobot is controlled by a remote expert—a form of human intelligence. This system enables the execution of collaborative tasks without necessitating physical presence and instructions that could be in real-time [4].

This work is limited to the proposal of the system's architecture and its functionalities, accompanied by a proof of concept. System usability tests are beyond the scope of this study and remain a long-term goal.

2. Related work

2.1. Virtual reality for control interfaces

Virtual Reality (VR) relies on a 3D virtual environment, a real-time interface with this virtual environment, and self-projection to ensure real-virtual coincidence and synchronization [4]. Telexistence refers to a human's ability to feel present in a location other than their physical surroundings and engage in a remote environment. VR redefines how the operator interacts with the robot thanks to direct natural interaction in virtual space. VR as a control interface for robotic manipulators has been studied for decades. It remains a popular area of research due to the increasing need for efficient remote working solutions and the improvement of VR solutions available. Teleoperation of robots using VR has found use cases in numerous areas, including military applications [5], robotic surgery [6], robotic-assisted rehabilitation [7], industrial settings [8], space exploration [9], and others. VR allows more human-like interaction with robotic systems, increasing intuitiveness for operators with little or no training. Operators can view and interact with the stereoscopic vision of the 3D model of the robot and the remote physical environment thanks to sensors. Even though automation solves many complex repetitive tasks, many scenarios need humans in the loop due to their subjectivity.

2.2. Collaborative systems and remote guidance

Developing remote collaborative systems for robotics entails the consideration of three key components [10]: electronic devices (such as cameras, robots, projectors, and sensors), display modalities (including 2D, 3D, AR, VR, mixed reality, and 360-degree panoramas, or a combination thereof), and communication cues (e.g., annotations, pointing, gestures, audio, and gaze) [11]. Means of guidance have been investigated, such as remote pointer displays on the local user's HMD view tracked by the gaze from a remote user [12] or using 360-video sharing to give the remote user the freedom to look where they want [13]. Other works have investigated volumetric shared scene capture with hand tracking [14]. Researchers investigating view sharing primarily focus on 3D as it increases immersion, improves scene understanding and provides better collaboration. Many studies combine different means, such as 360 panoramic cameras with 3D volumetric streaming or VR with AR systems [15]. Another related topic being explored is the representation of a remote user. [16] have investigated virtual hand models with different levels of realism for remote collaboration and concluded that low polygon should be considered for moderate social presence needs. Therefore, a simple remote pointer should be efficient enough to provide remote guidance in industrial and maintenance contexts. The *StickyLight system* [17] allows remote experts to draw annotations directly into the worker's workspace using a small projector, with the ability to stick annotations to their original positions despite worker head movement. There is still scope for enhancement, especially in addressing the physical burden on local workers using wearable or handheld devices, the restricted viewpoint of the remote expert, and the lack of the remote expert's presence within the local worker's operating space. *TeleAdvisor* [18] offers remote assistance through a teleoperated robotic arm with a camera and projector, enabling interaction with the worker's environment. *TeleAdvisor* has a restricted motion in a 2D plan to align the robotic arm with the helper's 2D perspective. We, therefore, propose a new telexistence-based remote guidance system with projection capabilities that offer more advanced motion freedom than *TeleAdvisor* and more advanced spatial awareness with mesh reconstruction from the worker space.

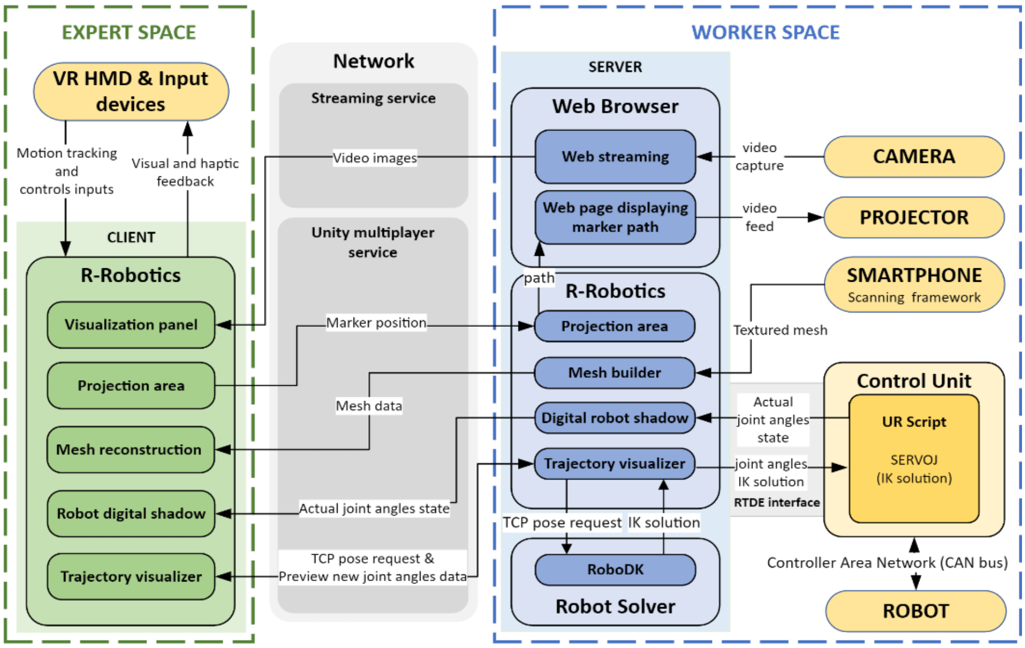

Fig. 2. High-level system architecture and data flow.

3. Design method

3.1. System architecture

The architecture proposed in this study is visualized in Fig. 2. and is integral to our application, R-Robotics, which we have developed using the Unity 3D platform. Known for its extensive use in creating interactive applications and simulations, Unity 3D has been instrumental in organizing R-Robotics into two distinct spaces: one representing the worker's location and the other, the expert's. These spaces communicate via a network platform, utilizing Unity 3D's high-level networking library, Netcode for GameObjects (Netcode)[2], to facilitate this interaction.

Netcode functions as a server in the worker's space, reserving a slot on the Unity 3D server and creating a public endpoint. This configuration allows remote experts to connect to the worker's space using a unique joint code, fostering real-time assistance.

3.2. Worker space

R-Robotics incorporates RoboDK, a robust robotic simulation tool for programming and optimizing robots. RoboDK is crucial as an inverse kinematics solver in our system, processing end-effector target pose requests from the expert control and avoiding potential cobot singularities. After determining the appropriate joint angle values to achieve the desired target pose, RoboDK[3] communicates this solution to RRobotics to ensure safeness, which then sends it to the cobot controller unit via a Real-Time Data Exchange (RTDE) interface. This interface operates over a standard TCP/IP connection, transmitting data at 125Hz.

Our system leverages URScript, a programming language, to control the cobot at the script level. It accesses specific registers storing the calculated joint angles solution, and the *SERVOJ()* command, a command for realtime control of joint positions, is used to prompt the cobot to execute the necessary movements.

The proposed framework captures and projects images via a web browser by accessing two specific web pages: one streams the video captured by the camera, and the other displays visual cues from the marker path for projection. Mesh acquisition is executed through a scanning application built with Niantic's Lightship Augmented Reality Developer Kit[4], enabling high-fidelity scanning with a smartphone. Upon scanning completion, the application forwards the Textured Mesh data to R-Robotics using a TCP connection.

3.3. Expert space

Upon initialization, R-Robotics operates in client mode, synchronizing all Unity 3D scene elements preset with network parameters, such as the cobot joint angles. The expert interacts with an immersive virtual environment composed of the cobot control model and a UI for visual cues, placed for optimal visibility during cobot operation and live camera feed monitoring.

We utilize the XR Interaction Toolkit (XRI) for interaction within this space[5]. Experts can teleoperate the cobot by grabbing its virtual end-effector through VR controllers and subsequently relocating it to the desired location using the IK solution calculated by RoboDK. This interaction emulates the cobot's 'Freedrive' mode, which allows for manual

[2]https://docs-multiplayer.unity3d.com/netcode/current
[3]https://robodk.com/
[4]https://lightship.dev/products/ardk
[5]https://docs.unity3d.com/Packages/com.unity.xr.interaction.toolkit@0.9

manipulation. During teleoperation, the system provides the expert with two cobot visualizations:

- The initial, transparent representation shows the cobot's proposed pose derived from the inverse kinematics solution.
- A solid representation displays the cobot's real-time joint angles.

This dual visualization strategy allows experts to preview the cobot's intended pose, making real-time adjustments efficient and precise. Additionally, experts can initiate environmental scans, enriching their spatial comprehension of depth and distances at the remote site in the virtual environment. This approach is ideal without a pre-existing 3D model of the remote location, facilitating more intuitive hand-driven teleoperation.

To engage the system's visual cues, the expert must first perform a calibration process, delineated in Fig. 3(2), aligning the drawing area with the projection area. This alignment is visualized on the camera's 2D feed display, as shown in Fig. 3(1).

The necessity for this manual calibration leads to our system's current limitations: it does not incorporate dynamic projection mapping algorithms, primarily due to the existing resolution differences between the camera and the video projector. To achieve accurate calibration, the expert must position two calibration cubes, highlighted in Fig. 3(3), precisely at the corners of a predefined projected square. This procedure allows the camera's size and orthographic positioning, restoring the visual cues with dimensions and placements mirroring those of the original projection area.

Once calibration is complete, the expert can initiate freehand drawing by activating specific controls on the VR HMD controller. Customization of the visual markers is available, allowing adjustments to the marker's size and color through interactive tools depicted in Fig. 3(4) for size (slider) and Fig. 3(5) for color (color palette).

After creating a visual cue, the system enables its projection via a dedicated button, shown in Fig. 3(6). This feature remains flexible, offering a 'clear' option when the visual cue becomes redundant and needs to be removed.

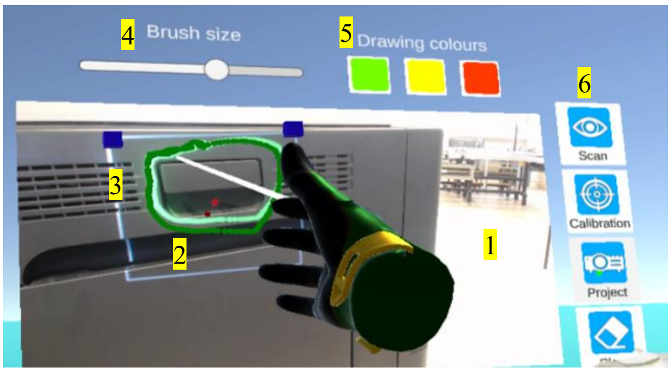

Fig. 3. Expert immersive user interface: (1) visualization panel of the camera 2D feed, (2) drawing area, (3) calibration cubes, (4) brush size controller slider, (5) colors palette, and (6) buttons.

3.4. Technical setup

The demonstration occurred at the S.MART platform within Grenoble INP. For the telexistence-based guidance tasks, we deployed a Universal Robot UR16e cobot, shown in Fig. 4(3). This industrial cobot arm is equipped with six degrees of freedom, can handle payloads up to 16 kg, and has a reach of 900 mm.

The setup included an MSI GT63 Titan featuring an Intel Core i7-8750H processor, an NVIDIA GeForce GTX 1080 graphics card, and 16GB of DDR4 RAM. The virtual reality (VR) interface was facilitated through the HTC VIVE Cosmos HMD, presented in Fig. 4(1), set up in a clear space measuring 4 m x 4 m. This headset offers a resolution of 2880 x 1700 pixels, a refresh rate of 90Hz, and a field of view spanning 110 degrees. Accompanied by two 6DoF (degrees of freedom) controllers, the VIVE Cosmos allows interaction within the virtual space.

We used a Logitech Webcam C925e for imaging, referenced in Fig 4(5). Our projections were carried out using a ViewSonic M1+ projector, which emits 300 lumens and is shown in Fig. 4(4), offering a resolution of 854x480. Additionally, an iPhone 13 Pro, seen in Fig. 4(6), was used for environmental scanning and mesh transmission.

Fig. 4. Technical setup for the demonstration: (1) VR HMD, (2) asset inspected, (3) UR16e, (4) projector, (5) camera, and (6) smartphone.

4. Prototyping and technical analysis

4.1. Assisting operator by addressing in-situ instructions

As illustrated in Fig. 5 and the accompanying video[6], we successfully demonstrated the proposed system's ability to remotely guide a local worker by assisting with the maintenance of a printer. The immersive interaction provided by VR within the virtual environment enables individuals unfamiliar with the cobot to manipulate it easily, comfortably, and intuitively without prior knowledge of robotics. The 3D scanning step aided the expert in accurately perceiving depth and distances within a worker's remote space, thereby allowing precise positioning of the cobot toward the area of interest. Utilizing a 3D representation is imperative, as 2D cameras, while facilitating scene visualization, fail to offer depth perception. This absence of depth perception complicates precise navigation and elevates the likelihood of erroneous cobot maneuvers.

[6]https://www.youtube.com/watch?v=TUMJmiIG-CI

STEP 1
Acquisition of the mesh in VR

STEP 2
Positioning the cobot in VR

STEP 3
Visual cues creation in VR

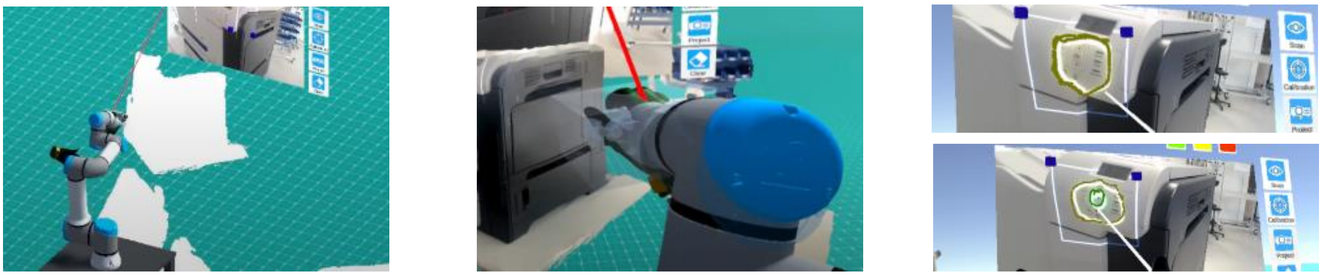

Fig. 5. Expert delivering remote guidance procedure in 3 steps.

Building on the proposed proof of concept, our system showcases that remote guidance utilizing telexistence can enable worker-expert collaboration compared to traditional methods. Conventional approaches often impose limitations on the worker, who must handle or be equipped with devices, and on the expert, who faces constraints regarding freedom of action, spatial understanding, and system usability. In this context, we have identified several aspects of our system that merit further assessment and enhancement to augment overall efficiency; Section 4.3 will explore these aspects in detail.

4.2. Time delay and latency

Fig. 6 shows the comparative visualization of teleoperation trajectories of the TCP (Tool Center Point) for the virtual robot of the real robot over time during a 21-second teleoperation sequence. Direct observations from the graph justify the choice of control command parameters, as the positions of the real robot TCP align closely with the virtual counterpart. Nevertheless, more noticeable discrepancies manifest at distinct intervals, especially in the rotation around the y-axis. These intervals suggest that the chosen parameters might hinder the robot's rapid adaptation to positional shifts. MacKenzie and Ware demonstrated significant increases in movement times and error rates with latency increases, suggesting that even small delays can impact performance drastically [19]. Addressing these delays is crucial for high-precision applications and effective remote guidance. The average communication delay ranges between 151ms to 328ms, with the expert in Singapore and the worker in France. The average discrepancies across the x, y, and z coordinates are approximately 0.024 ± 0.022 m, 0.038 ± 0.034 m, and 0.022 ± 0.031 m respectively. In terms of rotation, the average angular discrepancies around the x, y, and z axes are roughly 1.48 ± 2.34°, 0.84 ± 1.22°, and 8.03 ± 6.99°, respectively. The literature emphasizes the negative impacts of delay on both performance and operator state [20]. Our measured time delay is acceptable as previous studies found that from 1500 ms, a simple task of maneuvering a robotic end effector becomes substantially more difficult than at lesser time delays [21].

4.3. Limitations

In the proposed demonstration, several limitations and improvements can be highlighted:

- Employing a robotic arm, specifically the UR16e, restricts the collaborative workspace to a maximum range of 900mm.
- The drawing area's calibration is currently done manually with two calibration cubes in Fig. 3 (3); we could automate that with computer vision techniques. A more complex upgrade would be dynamic projection mapping, which adapts the projected cues to real-time changes in the object’s position or rotation [22].
- Usage of an improved projector device: While our system used a small, short-focal-length projector adequate for our needs, brighter environments require a higher resolution and brightness projector. Adaptive focus would also help maintain sharp images at varying distances.

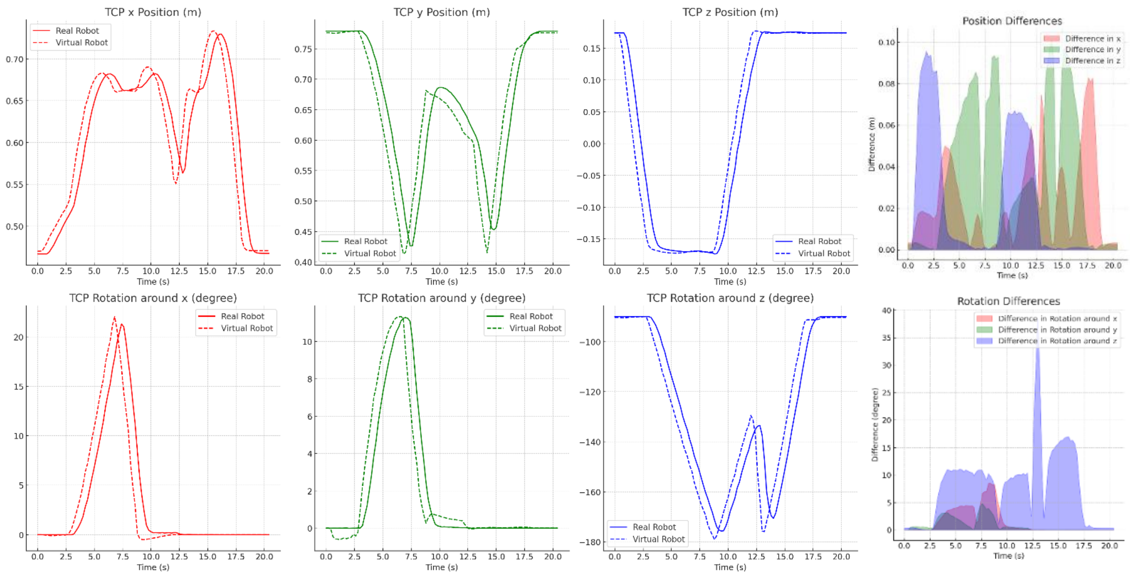

Fig. 6. Interpretation of teleoperation trajectories for a given expert operation sequence of 21 seconds.

- Another observation we made is the high workload required when operating the system, including mental and physical demands. The expert testing our system felt tired after just a few minutes.

5. Conclusion and future work

In conclusion, this paper presents a telexistence-based remote guidance system that addresses the limitations of existing remote collaboration solutions, enabling work-er-expert collaboration efficiency. Integrating a cobot with a camera, projector, and smartphone, our approach provides an immersive interface for remote experts, enhances the sense of presence, and eliminates the need for local workers to wear or carry devices. We have demonstrated the system's general use with a maintenance task, highlighting its potential applicability across various professional contexts.

Future work will address areas for improvement, such as automating the calibration process, upgrading the projector device, and integrating additional elements to enhance the expert's understanding of the remote location.

We also plan to conduct a comprehensive user study to assess the cognitive load associated with our system compared to traditional AR-based guidance systems and to evaluate the impact of human factors on usability and effectiveness.

Ultimately, our telexistence-based remote guidance system promises to transform expert and worker collaboration by eliminating physical limitations, enabling more effective and efficient remote interactions.

Acknowledgments

The National Research Agency supported this work under the COLLABORATION 4.0 ANR-18-CE10-0009 and by French government funding managed by the National Research Agency, grant ANR-21-ESRE-0030 (CONTINUUM project). This work is also supported by the joint Anglo-French Ph.D. scheme, organized by the respective nations' Ministry of Defense with the Defence Science and Technology Laboratory (Dstl) acting as the technical partner.

References

[1] R. Palmarini, J.A. Erkoyuncu, R. Roy, H. Torabmostaedi, A systematic review of augmented reality applications in maintenance, Robot Comput Integr Manuf 49 (2018) 215–228. https://doi.org/10.1016/j.rcim.2017.06.002.

[2] F. Herpich, R.L.M. Guarese, L.M.R. Tarouco, A Comparative Analysis of Augmented Reality Frameworks Aimed at the Development of Educational Applications, Creat Educ 08 (2017) 1433–1451. https://doi.org/10.4236/ce.2017.89101.

[3] M. Mengoni, S. Ceccacci, A. Generosi, A. Leopardi, Spatial Augmented Reality: an application for human work in smart manufacturing environment, Procedia Manuf 17 (2018) 476–483. https://doi.org/10.1016/j.promfg.2018.10.072.

[4] S. Tachi, Telexistence, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8844 (2015) 229–259. https://doi.org/10.1007/978-3-319-17043-5_13.

[5] T. Kot, P. Novák, Application of virtual reality in teleoperation of the military mobile robotic system TAROS, Int J Adv Robot Syst 15 (2018). https://doi.org/10.1177/1729881417751545.

[6] J.D. Bric, D.C. Lumbard, M.J. Frelich, J.C. Gould, Current state of virtual reality simulation in robotic surgery training: a review, Surg Endosc 30 (2016) 2169–2178. https://doi.org/10.1007/s00464-015-4517-y.

[7] M. Al-Sada, K. Jiang, S. Ranade, M. Kalkattawi, T. Nakajima, HapticSnakes: multi-haptic feedback wearable robots for immersive virtual reality, Virtual Real 24 (2020) 191–209. https://doi.org/10.1007/s10055-019-00404-x.

[8] D. Mazeas, J.A. Erkoyuncu, F. Noël, A Telexistence Interface for Remote Control of a Physical Industrial Robot via Data Distribution Service, in: IFIP International Conference on Product Lifecycle Management, Springer, Cham, 2023: pp. 388–398. https://doi.org/10.1007/978-3-031-25182-5_38.

[9] H. Chen, P. Huang, Z. Liu, Mode Switching-Based Symmetric Predictive Control Mechanism for Networked Teleoperation Space Robot System, IEEE/ASME Transactions on Mechatronics 24 (2019) 2706–2717. https://doi.org/10.1109/TMECH.2019.2946197.

[10] A.K. Inkulu, M.V.A.R. Bahubalendruni, A. Dara, S. K., Challenges and opportunities in human robot collaboration context of Industry 4.0 - a state of the art review, Industrial Robot: The International Journal of Robotics Research and Application 49 (2022) 226–239. https://doi.org/10.1108/IR-04-2021-0077.

[11] R. Druta, C. Druta, P. Negirla, I. Silea, A Review on Methods and Systems for Remote Collaboration, Applied Sciences 11 (2021) 10035. https://doi.org/10.3390/app112110035.

[12] K. Gupta, G.A. Lee, M. Billinghurst, Do You See What I See? The Effect of Gaze Tracking on Task Space Remote Collaboration, IEEE Trans Vis Comput Graph 22 (2016) 2413–2422. https://doi.org/10.1109/TVCG.2016.2593778.

[13] G.A. Lee, T. Teo, S. Kim, M. Billinghurst, Mixed reality collaboration through sharing a live panorama, in: SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications on - SA '17, ACM Press, New York, New York, USA, 2017: pp. 1–4. https://doi.org/10.1145/3132787.3139203.

[14] X. Zhang, X. Bai, S. Zhang, W. He, P. Wang, Z. Wang, Y. Yan, Q. Yu, Real-time 3D video-based MR remote collaboration using gesture cues and virtual replicas, The International Journal of Advanced Manufacturing Technology 121 (2022) 7697–7719 https://doi.org/10.1007/s00170-022-09654-7.

[15] S. Kim, G. Lee, W. Huang, H. Kim, W. Woo, M. Billinghurst, Evaluating the Combination of Visual Communication Cues for HMD-based Mixed Reality Remote Collaboration, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA, 2019: pp. 1–13. https://doi.org/10.1145/3290605.3300403.

[16] B. Yoon, H. Kim, S.Y. Oh, W. Woo, Evaluating Remote Virtual Hands Models on Social Presence in Hand-based 3D Remote Collaboration, in: 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), IEEE, 2020: pp. 520–532. https://doi.org/10.1109/ISMAR50242.2020.00080.

[17] M. Adcock, C. Gunn, Using Projected Light for Mobile Remote Guidance, Computer Supported Cooperative Work (CSCW) 24 (2015) 591–611. https://doi.org/10.1007/s10606-015-9237-2.

[18] P. Gurevich, J. Lanir, B. Cohen, R. Stone, TeleAdvisor: A Versatile Augmented Reality Tool for Remote Assistance, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA, 2012: pp. 619–622. https://doi.org/10.1145/2207676.2207763.

[19] I.S. MacKenzie, C. Ware, Lag as a determinant of human performance in interactive systems, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '93, ACM Press, New York, New York, USA, 1993: pp. 488–493. https://doi.org/10.1145/169059.169431.

[20] W. KUHMANN, Experimental investigation of stress-inducing properties of system response times, Ergonomics 32 (1989) 271–280. https://doi.org/10.1080/00140138908966087.

[21] C. Korte, S. Sudhakaran Nair, V. Nistor, T.P. Low, C.R. Doarn, G. Schaffner, Determining the Threshold of Time-Delay for Teleoperation Accuracy and Efficiency in Relation to Telesurgery, Telemedicine and E-Health 20 (2014) 1078–1086. https://doi.org/10.1089/tmj.2013.0367.

[22] A. Grundhöfer, D. Iwai, Recent Advances in Projection Mapping Algorithms, Hardware and Applications, Computer Graphics Forum 37 (2018) 653–675. https://doi.org/10.1111/cgf.13387.