IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 28 January 2023, accepted 9 February 2023, date of publication 15 February 2023, date of current version 23 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3245833

APPLIED RESEARCH

Multimodal Multi-User Mixed Reality Human–Robot Interface for Remote Operations in Hazardous Environments

KRZYSZTOF ADAM SZCZUREK[1,2], RAUL MARIN PRADES[2], ELOISE MATHESON[1], JOSE RODRIGUEZ-NOGUEIRA[1], AND MARIO DI CASTRO[1]

[1]European Organization for Nuclear Research (CERN), 1211 Geneva, Switzerland
[2]Department of Computer Science and Engineering, Jaume I University of Castellon, 12071 Castelló de la Plana, Spain

Corresponding author: Krzysztof Adam Szczurek (krzysztof.adam.szczurek@cern.ch)

ABSTRACT In hazardous environments, where conditions present risks for humans, the maintenance and interventions are often done with teleoperated remote systems or mobile robotic manipulators to avoid human exposure to dangers. The increasing need for safe and efficient teleoperation requires advanced environmental awareness and collision avoidance. The up-to-date screen-based 2D or 3D interfaces do not fully allow the operator to immerse in the controlled scenario. This problem can be addressed with the emerging Mixed Reality (MR) technologies with Head-Mounted Devices (HMDs) that offer stereoscopic immersion and interaction with virtual objects. Such human-robot interfaces have not yet been demonstrated in telerobotic interventions in particle physics accelerators. Moreover, the operations often require a few experts to collaborate, which increases the system complexity and requires sharing an Augmented Reality (AR) workspace. The multi-user mobile telerobotics in hazardous environments with shared control in the AR has not yet been approached in the state-of-the-art. In this work, the developed MR human-robot interface using the AR HMD is presented. The interface adapts to the constrained wireless networks in particle accelerator facilities and provides reliable high-precision interaction and specialized visualization. The multimodal operation uses hands, eyes and user motion tracking, and voice recognition for control, as well as offers video, 3D point cloud and audio feedback from the robot. Multiple experts can collaborate in the AR workspace locally or remotely, and share or monitor the robot's control. Ten operators tested the interface in intervention scenarios in the European Organization for Nuclear Research (CERN) with complete network characterization and measurements to conclude if operational requirements were met and if the network architecture could support single and multi-user communication load. The interface system has proved to be operationally ready at the Technical Readiness Level (TRL) 8 and was validated through successful demonstration in single and multi-user missions. Some system limitations and further work areas were identified, such as optimizing the network architecture for multi-user scenarios or high-level interface actions applying automatic interaction strategies depending on network conditions.

INDEX TERMS Augmented Reality, facility maintenance, hand tracking, hazardous environment, human–robot interaction, mixed reality, mobile robotic manipulator, mobile network, multi-user, safe operations, point cloud, spatial perception, telerobotics, voice control.

I. INTRODUCTION

The design of human-machine interfaces for mobile robots is a multidisciplinary challenge that requires taking into account the limitations of the complete control chain. To start with the robot's side, there are constraints of the mechatronic systems installed in the robot (i.e. maximum motor torques, limited types of control algorithms in drivers, material strength limit or singularities in manipulators) or processing power availability in the robot's control units that can be a bottleneck

The associate editor coordinating the review of this manuscript and approving it for publication was Zheng Chen.

for sensory data on-board processing. Furthermore, the communication link between the robot and the human operator must be established, which demands an investigation of the available network bandwidth, delays, volatility, security, and adapted protocols for dynamic environmental conditions and control strategies [1]. Finally, on the operator's side, the human-centred factors such as cognitive limits, psychology and ergonomics play a significant role in addition to the hardware and software design requirements [2]. During the human-robot interface specification, the level of automation is decided, which impacts how safeguarded teleoperation functionalities [3] or artificial intelligence (AI) should be [4]. The choice of the interface hardware is made by function of the required sensory feedback from the robot. For example, a 2D screen is enough for a video stream, and most digital information is presented as text or 2D graphics. Still, the 3D feedback from spatial sensors such as LiDAR, stereo cameras, and 3D model recognition can be naturally perceived only with stereo glasses, Virtual Reality (VR) or AR HMDs. Similarly, to allow force feedback from the robot, a hand-held controller can be used by the operator, or if the orientation of the robot must be conveyed (i.e. for a teleoperated aeroplane or a lunar lander [5]), a fully rotating system with the operator inside is required. Field operations, such as underwater robot control from a boat, underground tunnel robot control from nearby safe zones, search & rescue, emergency response, or space robot operation from a space shuttle, may require easily deployable and reliable interfaces with backup solutions. In the literature, numerous concepts aiding the design of the interfaces can guide the specification of required functionalities and making choices. Starting from the automation levels [6], which define how much decision-making authority the robot has and how much the human should supervise these actions. Furthermore, as there is always a difficulty with translating human intentions into a complex system, the terms ''Gulf of Execution'' and ''Gulf of Evaluation'' were proposed [7], which help create a user interaction system that brings closer the user high-level operation goals and low-level inputs, and makes evaluation of the system's state by the user easier. In the state-of-the-art, several studies targeted problems encountered in specific applications. The authors of [8] presented a broad examination of such issues and their mitigation proposed in 150 papers. The problems such as bandwidth, delay, frame rate limitations, absence of prioperception and frame of reference, 2D views, attention switches, distance estimation, understanding of the robot's position and orientation, predicting motion, or completing concurrent tasks were investigated. Some solutions were proposed as mitigation: stereoscopic displays, overlays, multimodality, decision, and predicting systems. In [9], which focused on AR control of unmanned aerial vehicles (UAVs), it was emphasized that teleoperation hid the mapping between the operator's input and robot's dynamics, which can be learned only by experience, and presents a dangerous situation for untrained operators. While evaluating the MR human-robot interface in [10], the problems of predicting motion, spatial awareness in an unstructured fragile environment were addressed with preview and collision avoidance functionalities.

The Human-Robot Interaction (HRI) in field robotics (e.g. in radiation or fire zones, space or underwater) is highly affected by the communication media, which enables telemetry from the remote robot and control signals sent by the operator. According to the operation scenario, a specific communication link must be used, which affects the telerobotic interaction mode (i.e. from manual to supervised control). Using a constrained communication link requires giving the robot more intelligence and a higher level of interaction, which does not require constant fast feedback. The operator must be able to specify high-level missions, such as waypoints and trajectories, as well as the activation of semi-autonomous behaviours. Nevertheless, it is essential to understand the limits of the network performance in the case of manual remote control in case the robot encounters a situation that cannot be solved in an autonomous or supervisory way. Combining simulated and actual data while performing an intervention enables the user to follow the mission steps, confirm them, and receive telemetry information while the robot is performing the corresponding task [11]. In the underwater communication domain, sonar modems are used in large areas [12], and Visual Light Communication (VLC) [13], and Radio Frequency (RF) are considered more for small scenarios such as industrial underwater facilities. Sonar modems offer communication links at low bandwidth and higher delays, while RF [14] gives low and constant latencies at short distances (e.g. <12 m). The VLC can be used in dark scenarios with higher communication requirements (e.g. 2 Mbps) and distances around 15 meters maximum. Moreover, underwater robots are usually supported by surface vehicles to bring the communication link to the air, so that long distances can be established from the control station to the target scenario (e.g. <50 km). In underwater robotics, significant efforts are made to have the robots ready for use in terms of mechatronics, software, and communication. It is necessary to use a realistic VR [15] and simulation tools [16] such as UWSIM [17] and Stone Fish [18], which allow testing the onboard robotic algorithms in controlled virtual scenarios. Also, it is usual to first test the robotic operations in a controlled water pool with the real robots [19] before bringing them to a more realistic scenario, such as the open sea [20].

In a complex teleoperated system, the user interaction with the robot greatly influences the mission result. In a collaborative space with humans or during a surgery [21], any harm to a person due to a command misunderstanding would be considered a mission failure. Similarly, any not avoided or even not perceived collision during manipulation, causing damage to unique scientific equipment in a particle accelerator or in a space station, would have critical negative consequences even if other mission goals were achieved.

A. STATE-OF-THE-ART OF WEARABLE AR TECHNOLOGY FOR MULTI-USER HUMAN-ROBOT TELEOPERATION

There have been multiple recent studies evaluating wearable AR technology for human-robot interaction while sharing a virtual or physical workspace of a single operator with a single mobile robot [9], [22], [23], [24], [25] or a fixed manipulator [26], [27]. Some studies were done for a single operator controlling multiple robots [23], [28], [29]. Table 1 presents an analysis of selected MR human-robot interfaces relevant to the research and development of the work presented in this paper. The study focused on the elements applicable to the compound challenges for the interfaces for robots operating in hazardous environments (explained in Section I-C), which is:

- The **environment where the robot operates**; if it is a laboratory or real scenario; if the scenario presents hazards for the robot or the operator; if the interface controls a real robot or it is done only in simulation.
- The **type of the user interface** (2D, 3D, and if it is based on VR, AR or MR).
- The **type of interaction** (joystick, gamepad, keyboard, hands tracking, voice commands, eyes tracking).
- The **collision avoidance or detection** methods.
- The **operator-robot communication link** between the operator, the interface server and the robot, and if it is adapted for shared or dynamic networks.
- The **AR, VR HMD, game engines, technologies** that were used to create the interface product.
- The **robot type**, e.g. wheeled, underwater, mobile or stationary.
- The **human-robot placement**, e.g. direct collaboration in the workspace with a robot; Line of Sight (LoS); long-distance teleoperation.
- If the robot uses a **manipulator**, what its type and complexity are, or if it has trajectories control.
- The **perception capabilities**, models of the robot or the environment available to the operator, 2D video or point cloud feedback, sensor fusion, interaction with force feedback.
- The **multi-user operation capabilities** integrated into the VR, AR, or MR, if it allows collaboration in a shared real workspace or if users can be distant. In remote collaboration, if the users can see other users' positions, hands, gestures, voice, video, point cloud or mesh.
- **Estimated TRL**.

Regarding the multi-user MR human-robot teleoperation, there has been a preliminary investigation at a conceptual level of multiple users sharing an AR workspace and operating a single robot or multiple robots, for example in [23]. On the other hand, there have been extensive studies of products for remote MR teleconferencing, which do not offer teleoperation capabilities:

- The authors of [30] proposed the Virtual Monitors methodology to overlay the real world with virtual images of other users in an AR conferencing system, where one user wore an optical see-through HMDs and other users used traditional webcam-based computer stations. The user could interact with the virtual objects using a Virtual Shared Whiteboard and communicate by voice and gestures.
- A mixed collaboration between AR and VR environments was discussed in [31]. The concept of a transitional augmented reality interface was introduced, where the interface could transition from the real world to the virtual world and then to the augmented real world, and provide collaboration capabilities between users or allow an individual immersion into a problem. The authors emphasized that different viewpoints in collaborative tasks generally improved the efficiency of manipulation or navigation tasks in VR, but the applications needed high-level control and a well-designed communication layer.
- The authors of [32] presented an AR system where a remote collaborator was rendered into the scene. The collaborator was surrounded by 15 cameras, which allowed the construction of the 3D-rendered model. A fiducial marker provided a stable anchor in the scene where the collaborator was drawn.
- Volumetric avatars based on 3D capture were proposed in [33]. In [34], 3D avatars represented other players or conference interlocutors.
- Commercial products developed by enterprises play a big role in driving the MR collaboration research: Microsoft Dynamics 365 [35], Imverse [36], Meta Oculus [37], Magic Leap Social [38], Gixel [39] or High Fidelity [40].

Table 2 presents an analysis of selected MR conferencing systems.

B. CERN HUMAN-ROBOT INTERFACES EVOLUTION AND STATE-OF-THE-ART

This section describes the evolution of CERN human-robot interfaces or related domains that extended the capabilities of CERN robotic teleoperation, such as GUI-robot networking, point cloud processing for spatial feedback, SLAM, simulations, collision avoidance, objects recognition, pose estimation, 3D information presentation, VR, AR, operator training or master-slave bilateral systems. For a complete overview, Table 3 presents the timeline and the domains to each reference contributed. The list below explains the key contributions and added value of each reference in chronological order:

- The preliminary CERN multimodal human-robot interface [42] allowed multiple types of input devices (keyboard, joypad, haptic device) and adapted dynamically to the configuration of the robot's components.
- A stereo vision system based on point cloud acquisition and simulation was studied in [43].
- Research on object pose estimation for precise robotic manipulation in unstructured and dynamic CERN underground environment was presented in [44].

TABLE 1. The analysis of selected MR human-robot interfaces regarding the aspects crucial for the teleoperation of robots in hazardous unstructured environments.

Interface product	Environment where the robot operates	User interface	Interaction	Collisions avoidance or detection	Operator-robot communication link	AR, VR HMD, game engines, technologies	Robot type	Manipulator	Human-robot placement	Perception capabilities	Estimated TRL	Multi-user operation capabilities
Mixed Reality for Robotics [23]	Laboratory scenarios, combinations of virtual and real spaces, and real or simulated robots	MR on computer screens	-	-	-	Unity 3D, V-REP and Gazebo	UAV (Crazyfly) and wheeled robots (TurtleBot)	-	Sharing the same local or remote workspace, with virtual and real elements	Real or virtual cameras of modelled environments	2-3 (laboratory, conceptual level)	Concepts of collaboration in real and virtual spaces between robots and humans
Designing planning and control interfaces to support user collaboration with flying robots [41]	Laboratory indoor scenario with free-flyers operations	MR on computer screens	Gamepad (safeguarded teleoperation), hands tracking (waypoint delegation and collaborative interfaces)	Safeguarded teleoperation	Custom-written network communication (no details)	Windows Presentation Foundation	UAV	Waypoints by using active markers	Remote teleoperation	Reconstructed real environment (point cloud and video)	2-3	Collaborative interfaces (3D supervisory control, map for waypoints planning, interactive timeline for managing interdependencies and sequential task order)
Communicating Robot Motion Intent with Augmented Reality [25]	Laboratory indoor scenario with free-flyers operations	AR with HMD	-	-	-	HoloLens, Unity	UAV	-	LoS	-	2-3	-
Robot Teleoperation with Augmented Reality Virtual Surrogates [9]	Laboratory scenario, real robot	AR with HMD	Gamepad	Safeguarded motion in state-of-the-art	UDP communication with the real robot	HoloLens, Unity	UAV (quadrocopter)	Trajectory control	LoS, shared environment between a human and a robot	-	3 (physical quadcopter teleoperated in laboratory environment)	-

TABLE 2. The analysis of selected MR conferencing systems in terms of providing multi-user capabilities, local/remote collaboration, avatar representations, and the limitations of the systems.

MR teleconferencing product	Multi-user capabilities	Avatar visualization	Limitations
3D live: real time captured content for mixed reality [32]	An AR uni-directional video conferencing system where a remote collaborator is rendered into the scene.	A full 3D model visible from any viewpoint in the scene.	The system is too complex (14 cameras) to be easily deployed in an operational scenario, and not symmetrical (one collaborator is captured, the other one is observing).
General-purpose telepresence with head-worn optical see-through displays and projector-based lighting [33]	Virtual Monitors methodology to overlay the real world with virtual images of other users in an AR conferencing system, where one user wears an optical see-through HMD and other users use traditional webcam-based stations. The user can interact with the virtual objects using a Virtual Shared Whiteboard and communicate by voice and gestures.	Virtual images of other users in an AR conferencing system.	The system is not symmetrical (only one user can use an AR HMD, and the others see 2D webcam video feedback). The latency due to the wireless video link was high, and there were problems with the AR display with too much lighting in the room.
The Effect of Avatar Appearance on Social Presence in an Augmented Reality Remote Collaboration [34]	Bi-directional interaction with a remote user represented as different types of avatars.	Realistic or cartoon head, hands, upper, whole or no body.	The collaborator is seen as an avatar, which cannot convey information from facial expressions or unrecognized hand gestures.

- AR spatial visualization of physical quantities, such as radiation, superposed with a 3D point cloud environment, was integrated with the human-robot interface [45].
- A robotic teleoperation training simulator in a virtual environment for teaching beginner operators was created [46].
- A collision avoidance system prototype [47], which uses Infrared (IR) Time-of-Flight (ToF) sensors for the Radio-Protection (RP) arm, was integrated into the Train Inspection Monorail (TIM) in the LHC tunnel.
- An idea of an AR display in an operator's glasses of environmental measurements (oxygen, radiation, temperature) during robot-assisted interventions was presented [48].
- A novel real-time object recognition and tracking system, which enters the teleoperation loop and helps the operator achieve goals, was introduced [49].
- A CERNTAURO framework with a modular architecture [50] covering all aspects of CERN's robotic remote facility maintenance and interventions has been in use. It covers the elements from the specification and operator training, robot selection, material choice according to radiological contamination risks, until the realization of the mission, taking into account procedures and recovery scenarios. It synthesized, for example, the novel bilateral master-slave control, user-friendly multimodal human-robot interfaces, and offline operator training.
- A robotic platform with dual arms and modular configuration for complex multi-arm telemanipulation

handling tasks on old CERN accelerator equipment was designed [51].
- An accelerator construction structural inspection system using a robotic platform was built [52]. The images and point clouds captured by the robot's on-board cameras were used to reconstruct the environment, visualize and process in VR.
- The Intelligent Train Inspection Monorail (i-TIM) system was introduced [53]. It provided measures to increase operation safety, such as collision avoidance systems, increased perception with sensor fusion, arms for manipulation, or lost communication procedures.
- A novel vision system tracking and estimating the depth of metallic target [54] was designed for specific CERN robotic intervention.
- Autonomous communication relay mobile stations [55] were designed to extend the communication range for robots operating in remote and harsh environments, which do not have enough network coverage.
- The multimodal human-robot interface for remote robotic intervention [11] supporting all robotic platforms operating in CERN hazardous environments was commissioned. It dealt with various practical issues, such as adaptation to varying network delays, frequent reconfiguration of robots, and multiple types of input devices or sensors installed on robots. The system also supported multi-robot and multi-tasking scripting.
- A module in the human-robot interface, which allowed a single user to teleoperate or simulate multiple robots' cooperation, was added [56].
- The VR was applied to preparing a robotic intervention, where the environment was fully modelled, and the operators used an immersive VR HMD [57]. The system was used to estimate dangers in the maintenance planning, the radiation dose received or checking approach feasibility.
- An algorithm for the robust 6D pose estimation with an RGB-D camera in harsh and unstructured environments using object detection was proposed [58].
- An original algorithm of graph SLAM for robot localization in accelerator tunnels was used [59].
- The CERN MR human-robot interface prototype was studied in [60]. It used Augmented Virtuality for real-time video feedback from the robot and entire scene modelling, which was used for Beam Loss Monitor robotic measurements with a redundant manipulator. The study also introduced operator vital parameters (heartbeat and galvanic skin response) monitoring.
- For educational, testing, and prototyping purposes, a minimized version of CERNBot with similar capabilities (i.e. perception with sensors, a manipulator, omnidirectional propulsion) was built [61]. The MiniCERNbot used additional human-robot interfaces that could be used with a simple browser or on a portable device, such as a smartphone.
- A visual servoing control of a robot that had to pass a narrow gate in the CERN SPS accelerator was implemented [62].
- The MR human-robot interface with full robotic model representation, real-time camera video and 3D point cloud feedback, collision detection and avoidance, inverse kinematics, trajectories planning, and novel adaptive congestion control based on network conditions, was implemented and introduced in operation [10].

C. MOTIVATION

The particle physics accelerators and experimental facilities present a risk for humans due to radiation hazards, gas leaks, oxygen deficiency, confined spaces, electrical shocks, or magnetic fields. Therefore, any intervention in such places should be done with remotely controlled robots (Figure 1). Moreover, the operators usually cannot be in close vicinity of the robot, and the only received feedback is the sensory information sent from the robot. The research, developments, and experimental results of the CERN multimodal multi-user MR human-robot interface with multimodal and multi-user capabilities, which are presented in this publication, were motivated by the following:

FIGURE 1. The figure shows two CERN robotic intervention scenarios in the Large hadron Collider (LHC) tunnel. On the left, the remotely controlled CERNBot performed an inspection of equipment in the radioactive environment. It was equipped with a pan-tilt-zoom camera and a manipulator with a radiation sensor. On the right, two CERNBots transported a metallic beam collaboratively.

- The safety of the operation must be assured by providing reliable hardware solutions and by creating intuitive interfaces that take profit of the robot's perception and display the synthesized information.
- Up to now, the operational CERN human-robot interfaces used screens for visualization and keyboard, gamepad, space mouse, or a master-slave control with haptic feedback for control [10], [11], [56]. However, in a complex teleoperated system extended mapping of key bindings and actions are difficult to remember, and a more intuitive input system should be designed.
- There have been attempts to use VR HMDs to present the information in robotic intervention scenarios without success due to the complexity (cabling, base stations, tracking units, power supply) and intrusiveness of such

TABLE 3. The state-of-the-art of CERN human-robot interfaces and telerobotics with a timeline. The ✓ points the areas that the reference extended the state-of-the-art. The HRI means a general contribution to the human-robot interfaces, while the other areas focus on a particular aspect.

Reference	Year	HRI	Net-working	Point cloud proce-ssing	SLAM	Simu-lation	Collision avoidance	Object recogni-tion	Object pose estima-tion	3D	VR	AR	Operator training	Master-slave bilateral system
[42]	2016	✓	✓											
[43]				✓		✓								
[44]	2017			✓				✓	✓					
[45]					✓							✓		
[46]	2018			✓		✓				✓	✓		✓	
[47]						✓	✓							
[48]												✓		
[49]		✓						✓						
[50]		✓	✓	✓		✓		✓					✓	✓
[51]		✓	✓		✓		✓						✓	
[52]				✓		✓				✓	✓			
[53]		✓	✓	✓			✓	✓		✓	✓			
[54]	2019	✓						✓	✓					
[55]		✓	✓		✓									
[11]		✓	✓	✓						✓				
[56]	2020	✓				✓		✓	✓	✓	✓			✓
[57]		✓				✓				✓	✓		✓	
[58]								✓	✓	✓				
[60]	2021	✓				✓				✓	✓			
[59]				✓	✓					✓				
[61]		✓	✓											
[62]	2022							✓						
[10]		✓	✓	✓		✓	✓			✓	✓			

headsets and used VR controllers, which often caused motion sickness and were difficult to deploy in the field.

- Robotic interventions performed at CERN often involve multiple experts in the scenario:
  1) a robotic operator that controls the robot,
  2) an expert of the scenario that guides the operator on how the task should be executed or gives advice if any unexpected event happens,
  3) a monitoring operator who is checking the feedback from the robot to avoid collisions, guide during movements or give advice on actions,
  4) an operator who is an operator in training,
  5) a senior expert operator teaching the operator in training.

  As shown in Figure 2, the collaboration may not always be convenient or achievable because of constrained space or limited visibility of screens in an ad-hoc created workspace. Therefore, a solution based on MR workspace sharing locally or remotely was necessary.

The experiments performed in this work were conducted to verify the following hypotheses:

- Hypothesis 1 (H1): The designed and implemented (at the TRL 8) 3D MR human-robot interface with the AR HMD can be qualified for remote operations in particle accelerators (the validation conditions are defined in Section III) with their particular limitation of the available communication networks, and the requirements to efficiently navigate a robot, manipulate a robotic arm and avoid collisions.
- Hypothesis 2 (H2): The system can provide AR multi-user remote operation capabilities with collaborators' local or remote presence. The communication architecture can support multiple users who can collaboratively perform a real robotic intervention.

D. NOVELTY AND CONTRIBUTION

We want to extend and contribute to two domains in this publication. The first one is the domain of operational MR human-robot interfaces using AR HMD. No such interfaces have yet been qualified for real operation scenarios, especially in hazardous unconstrained environments and field operations. The second is the interface's capability for collaborative remote teleoperation and supervision from local or remote workspaces, which has not yet been proposed and released in industrial use cases with mobile robots and constrained communication networks.

As a result of the motivation described in Section I-C, the detailed novelty and contributions presented in the paper are:

- The MR human-robot interface was fully designed, implemented, and deployed in the operational scenarios in particle accelerator facilities. The multimodality of the proposed solution is manifested in interaction types (using human senses, speech, and bare hands, or can be used with physical controllers such as a gamepad, a keyboard, or a joystick), or connection options (4G, Wi-Fi), or combination of 2D or 3D content on screens and in the HMD. It uses sight and hearing to receive feedback from the remote robot; recognized speech for commands; eyes tracking to point or select; hands tracking to interact with virtual elements; and gestures for precise control. The only wearable device is the

FIGURE 2. The figure presents the problematics of multi-user scenarios, where the operator's workspace in the field does not provide enough capabilities for other experts to comfortably and efficiently advise or guide the operator in the environment or task execution. The pictures were taken during robotic interventions at CERN.

wireless HMD used by the operator, an AR headset (in this publication, Microsoft HoloLens 2 [63] use-case is presented) that has hands-tracking capabilities, as well as integrated microphone and speakers. It eliminates the need to set up a station with a table, multiple screens, base station for tracking the operator and controllers. This interface allows visualization of the robot's model flexibly positioned in the workspace, independently of the operator's location. As the feedback, spacial point clouds, or multiple camera feedback are visualized. The control of the robot is done in an intuitive, natural, and multimodal way. The human locomotion or rotating the model is used to change viewpoints. This interface was developed at the TRL 8 for reliable operations at CERN in an unstructured environment, and additional constraints are described below.

- The spatial awareness, collision avoidance, trajectories preview and execution supervision, and adaptive communication network congestion control for visual feedback were adapted to the MR technology with the AR HMD. The interface provided better environmental understanding and facilitated collision avoidance presented in the MR environment by automatic movement termination when a collision was detected or when an imminent collision could happen if a trajectory were executed. The interface raised the remote operation level from manual teleoperation to supervisory control for such tasks as the approach, trajectory following, or automatic adaptation to dynamically changing network communication constraints.
- For the collaboration of multiple operators, a multi-user architecture was designed, deployed, and tested. It preserves all the control functionalities of a single robot operator with the MR human-robot interface with the AR HMD, while allowing other operators to see the robot's status and feedback or take control. Multiple operators can collaborate in the same workspace, where collaborators can see each other. Or, it can be done remotely, where the users can see virtual hands with precise gestures and finger joints movements, as well as the user point clouds that can provide an even better interaction feeling. The multi-user environment contributes to the MR conferencing state-of-the-art and allows full remote control of a robot in a multimodal multi-user collaborative way.
- In the underground environment, such as the CERN accelerators and experimental halls, there are specific communication limitations (i.e. availability and fixed type of network; its limitation of bandwidth; large round-trip time, often varying over time and locations). Therefore, flexible and adaptive communication architecture and solutions must have been designed to enable MR telerobotics. Already with a single operator, a continuous sent data stream can easily reach available bandwidth limits. Moreover, to facilitate multi-user collaboration and spatial feedback from the robot, an even larger amount of data must be continuously sent from the robot to the operators. In this paper, we measured what was the required communication load to provide useful feedback for a multi-user and multi-camera AR context and tested the proposed architecture of the communication system with the Adaptive Communications Congestion Control [10]. This publication extends the experimental results described in the previous publication.

E. PAPER STRUCTURE

The paper is structured as follows:

- Section II describes the developed MR interface system, in which the robot operator interacts with 3D holograms of the robot's model and environment representation, and uses hands, voice, eyes tracking and locomotion to send control commands.
- Section III describes the setup of the experiments in which the interface was tested in robotic intervention scenarios at CERN. The validation conditions for task execution, network performance, and feedback quality are specified.

- Results of tasks execution in single and multi-operator scenarios, operators' feedback, and network performance are presented in Section IV and discussed in Section V.
- Section VI concludes the findings and proposes further work.

Moreover, this paper should be read together with our previous publication [10], as it references its multiple sections and figures. Also, the experiments performed here use several scenarios the Adaptive Communications Congestion Control paper fully described and characterized. Therefore, it is recommended to read the previous publication first to understand better the results, which became more complex in the multi-camera and multi-user applications with the AR HMD presented here.

II. SYSTEM DESCRIPTION

The system architecture of the solution presented in this publication extends the MR human-robot interface described in Section B of [10]. Specifically, this system implements all the functionalities available for that interface on screens, and extends it in the following aspects:

1) The use of AR HMD for control and visualization of the operated robot and its environment.
2) As depicted in Figure 3, the operator inputs were replaced by natural human interaction inputs: hands, eyes, and user locomotion tracking, as well as speech recognition. In the standard screen-based interfaces, there was always a physical device to be manipulated to obtain input signals. The 3D holographic output placed in the real operator's workspace for visualization is now used, as well as spatial audio feedback from the robot was added.
3) The system now supports two types of robot bases: the CERNBot with an omnidirectional wheels base (Figure 4) and the LHC Train Inspection Monorail with a robotic arm (Figures 5, 6 and 9 in [10]). The CERNBot can be equipped with a scissor lift to increase the task space and two manipulators with end-effectors. The train robotic wagon contains a 9 degrees of freedom (DOF) manipulator for the Beam Loss Monitor measurements (explained in Section IE of [10]).
4) Multiple users can cooperate and share control at the same time while using AR HMDs.

The AR interface was primarily designed to operate CERNBots (Figure 4) and all functionalities presented in this section are presented with it. However, the interface is also under test with other types of robots shown in Figure 5 of [10].

The interface was developed with the Unity 2021 game engine with Microsoft's Mixed Reality Toolkit (MRTK) 2 library for HoloLens 2 interaction and visualization. The interface processing is localized in a server or a portable computer. The visualization and input signals are streamed to and from the HMD via the Holographic Remoting Player application [64].

A. INTERFACE INTERACTIONS AND FUNCTIONALITIES

The operator is provided with a multimodal interaction with the robot controls and sensory feedback acquisition settings. There are six types of interaction:

1) Hands near interaction: the fingers are in "contact" with holographic elements, for example, pressing a button, grasping and moving an object; the element must be within arm's reach.
2) Hands far interaction: with a pointer controlled by the hand's position, the elements can be interacted with from a distance, for example, to move a video canvas or the robot's model; the element does not have to be within the arm's reach.
3) Hand gesture, position, and orientation tracking: a hand acts as a remote controller. A robot's base or a manipulator follows the movement of the hand; the hand's or both hands' gestures can also act as a confirmation key to launch movement or activate a control mode.
4) Voice command: a sequence of words recognized by the system launches a command. For example, "base control" activates the control of a mobile base, and "save waypoint" creates a waypoint in a planning mode.
5) Eyes tracking + voice command: a holographic element can be pointed with eyes tracking and a voice command launches an action on this element, for example, pointing a waypoint with eyes and saying "go to target" brings the planning arm to the waypoint, or an element in a hand menu can be pointed with eyes and saying "select" activates it.
6) Eyes tracking + dwell: only eyes tracking can be used to interact with a holographic element, for example, eyes can look at a button, stop for 500 ms of dwell time, and then the element is activated, or eyes can point an arrow that controls the robot's base movement. This mode only works when hands are within their tracking region to avoid conflict.

Table 4 presents a mapping between functionalities and available interaction types. Multiple interaction types for functionality allow the choice according to an operator's preference and are also helpful when an external condition prevents the usage of a particular interaction. For example, in noisy environments, the voice command may not work reliably, but it is possible to use the hand menu and press a button with a finger. Or, a particular eye anatomy, glasses or imprecise eyes tracking calibration may offset the tracked point. In this case, a hand pointer will be more suitable with a finger pinch confirmation. From experience, hand interaction is the most reliable interaction type, although it requires a learning phase, as the holographic elements can only give visual and audible feedback when activated.

B. NETWORKING

The remote-controlled robot in unstructured hazardous environments at CERN requires flexible and multimodal networking solutions. Due to specific hazards in accelerator complex

TABLE 4. Interface interaction inputs and functionalities mapping. In most cases, each functionality can be used multimodally with 2-4 input types at the operator's convenience (hands near or far interaction, hand gestures and tracking, voice command, eyes tracking + dwell/voice command). Each functionality has a graphical example if the interface implements it, which can be consulted in its corresponding figure or the video demonstration available in [65].

Functionality / input		Hands near interaction	Hands far interation	Hand gesture, position and orientation tracking	Voice command	Eye tracking + voice command	Eyes tracking + dwell
Independent interaction with hand menu	Robot and arm speed settings	✓	✓ Figure 30			✓	✓
	Communication parameters measurements	✓	✓ Figure 31				
	Camera settings	✓ Demo [65]	✓ Figure 32, Demo [65]		✓	✓ Demo [65]	✓
	Launching commands	✓	✓		✓ Demo [65]	✓	✓
	Consult operator's vital parameters	✓	✓ Figure 30				
Moving, rotating and resizing the robot's model in workspace		✓	✓ Figure 16, Demo [65]				
Moving an arm joint	Moving the arm joint directly		✓ Figure 21				
	Moving the arm joint with an arrow		✓ Figure 20				✓
Cartesian control	Moving the arm in Cartesian mode		✓ Figure 23, Demo [65]	✓Figure 22			✓ Demo [65]
FABRIK inverse kinematics	Moving the arm in FABRIK inverse kinematics planning mode	✓ Demo [65]	✓ Figure 24				
	Move the planning arm to point cloud normal point				✓		
Trajectories management and execution	Select destination waypoint		✓ Figure 25, Demo [65]			✓ Demo [65]	
	Launch preview of movement				✓(Figure 25)		
	Launching the movement to selected waypoint			✓ Figure 26, Demo [65]			
Moving robot base	Moving the robot base with arrows		✓				✓
	Moving the robot base with hand gesture, position and orientation tracking			✓ Figure 19 Demo [65]			
Camera control	Interaction with camera settings hologram	✓ Figure 32, Demo [65]	✓ Demo [65]			✓ Demo [65]	✓
	Interaction with camera video stream canvas	✓ Demo [65]	✓				
Point cloud interaction	Moving normal point	✓ Figure 27, Demo [65]	✓				
	Select normal point	✓	✓			✓ Demo [65]	

or available infrastructure (described in our previous publication [10] in Section II-A-I), only the 4G network can be currently used in the radioactive underground areas. In some places, only direct LoS connection (e.g. Wi-Fi hotspot) is available because of environmental shielding, or only CERN Wi-Fi or cabled network infrastructure is available. Figure 5 presents the architecture of the communication system allowing the operator to wear the AR HMD and control the robot. The usage of each connection type has consequences on the communication performance: bandwidth, delays, fluctuations; and has requirements regarding the locations of the operator, interface server, and the robot. These requirements

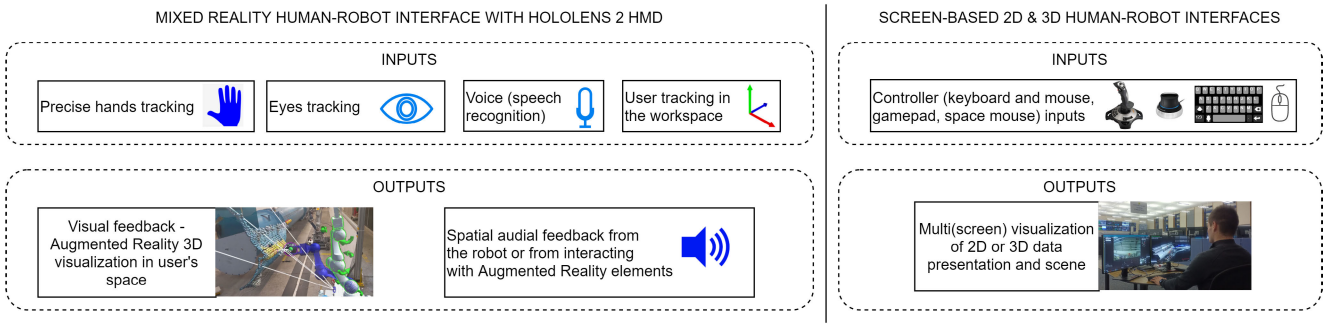

FIGURE 3. The figure presents the difference in inputs and outputs between an interface using the AR HMD and a screen-based 2D or 3D interface with standard controllers.

FIGURE 4. CERNBots with different configurations (dual or single manipulator, PTZ cameras, lifting stage, cutters, gripper or custom-made tools).

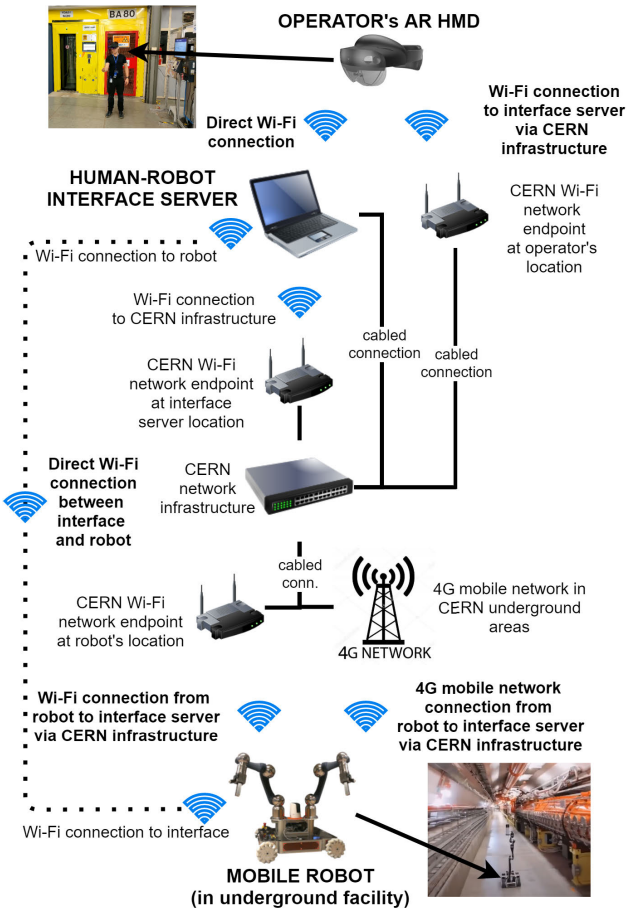

FIGURE 5. The diagram summarizes multimodal connections between the operator wearing the AR HMD, the interface server (streaming the holographic display contents, translating the operator's commands into robot control commands, and communicating with the robot), and the physical mobile robot. Starting from the operator's side, the HMD has two options to connect to the interface server: 1) direct Wi-Fi connection to the server (i.e. using Wi-Fi hotspot); 2) a chain of CERN infrastructure of Wi-Fi and cable connections to the interface server. The server can be connected to the infrastructure by Wi-Fi or a cabled connection. And finally, the mobile robot has 3 options to connect to the interface server: 1) direct Wi-Fi connection; 2) via CERN Wi-Fi endpoint at the robot's location and then infrastructure; 3) via 4G mobile network at the robot's location and then infrastructure.

are presented in Table 5, which also describes their use cases, and advantages or disadvantages.

C. MULTI-USER OPERATION AND COLLABORATION

As motivated in Section I-C, the intervention may require more than one person operating a robot or supervising a mission. Therefore the multi-user scenario extends the single-user scenario by introducing other users in the AR workspace (Figure 6). In the single-operator scenario, one robot is controlled by one operator, who is the expert in robot control and knows well the intervention scenario. If the scenario is complicated, an expert can join to help navigate the unstructured environment or provide task execution guidance. Similarly, if another viewpoint, expertise, or complexity of the robot manipulation requires another operator to join to monitor or take control, that is also possible in the multi-user scenario.

The multi-user operation can be executed in a shared local workspace or remotely. In the local workspace, users collaborate in the same physical space. The holographic scene is visualized precisely at the same place for all users (e.g. at a round table in a control centre conference room in Figure 7).

TABLE 5. Connection types used in robotic interventions at CERN, their network capabilities and limitations, use-cases, advantages, disadvantages, and the required infrastructure.

Connection type	Use-cases	Advantages	Disadvantages
Direct Wi-Fi connection between interface server and robot	Robot located in areas without 4G or Wi-Fi coverage; used for transporting or testing.	The highest bandwidth and shortest delay.	Interface server and robot have to be in the vicinity. Practically used only for deployment or commissioning.
Wi-Fi connection from CERN infrastructure to mobile robot	Robot is in less radioactive environment with Wi-Fi coverage.	Shorter delay than with 4G network.	Wi-Fi network coverage is not provided in highly radioactive areas, outdoors and outside CERN.
Wi-Fi connection via CERN infrastructure from AR HMD to interface server	HMD connection used when CERN Wi-Fi is available.	The most convenient connection of HMD. The distributed interface server can be remotely connected to HMD.	The Wi-Fi infrastructure delay adds to the total delay. CERN Wi-Fi network coverage required where the operator is.
Cabled connection from interface server to CERN infrastructure	Distributed control with interface server in a fixed control centre and operator in the field.	The best performance on the server side, the network and processing are practically limited only on the robot's or HMD sides.	Cabled infrastructure is required, available only for a stationary server.
4G network connection from robot to CERN infrastructure	Teleoperation in highly radioactive areas (i.e. LHC or SPS), outdoor operation far from Wi-Fi network.	The only robot connection in highly radioactive areas.	The longest delays, lowest bandwidth and highest variations in time and location.
4G network connection from interface server to CERN infrastructure	Operator is in the area with only 4G coverage, AR HMD connected to the server via Wi-Fi hotspot	Teleoperate possibility from experimental, underground or outside open-air areas without LoS with the robot.	The worst network performance on the operator's side.

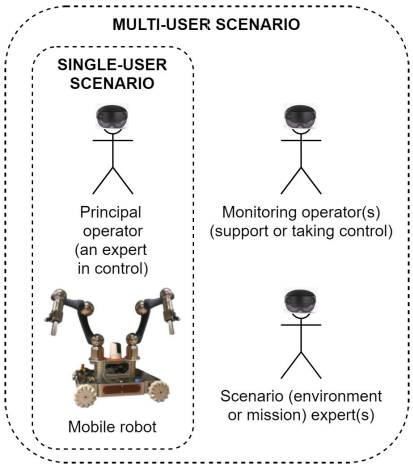

FIGURE 6. Multi-user scenario extending the single-user teleoperation scenario.

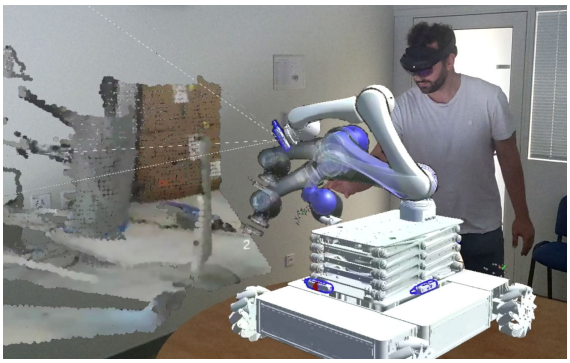

FIGURE 7. Local collaboration scenario with two operators in the same room. The figure shows the viewpoint of one operator on the scene and the second operator that controls the robot. In this example, the controlling operator created a trajectory with waypoints to approach a target in the real environment.

The scene positioning and scale can be done using spatial anchors or manually adjusted. Direct pointing and discussions are facilitated. The actions performed by the controlling operator are synchronized in the display of the monitoring operator. The video demonstration [65] shows the process and collaboration between operators.

Remote collaboration can be done without any restrictions on location in the world. The only requirement is sharing the same network, for example, by using a VPN connection, which allows communication between the users and the robot. Facial, hand, or body expressions help better understand the intentions, interactions and messages despite the physical distance. To visualize other collaborators and their gestures, a streamed point cloud of the person (Figure 8) or a digital hands representation, with all hand joints being tracked, can be used. The point cloud or hands representation positions are tracked and placed in the remote workspace used by other operators. These functionalities enable seeing where a person is in the scene and what actions are being performed.

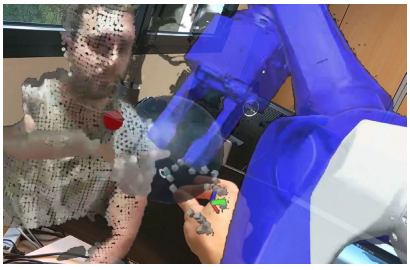

FIGURE 8. Remote workspace collaboration scenario of two operators. In the example here, a remote user pointed with a point cloud finger a position where the robot should be moved, and the controlling operator moved the arm to that position.

The collaboration between multiple users requires a protocol established between them to avoid conflicting

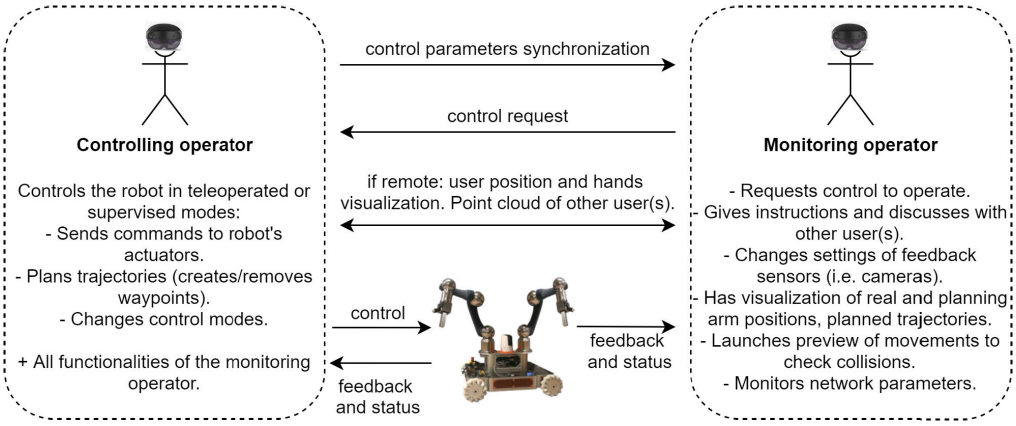

FIGURE 9. Multi-user operation workflow and data exchange. In the multi-user scenario with two or more operators, one operator controls the robot, and the other(s) is/are monitoring. The roles can be exchanged when a monitoring operator requests control. The robot's position is synchronised if the operators are located in the same workspace (Figure 7). If the operators are remote, the hands or a point cloud of the operator are shown (Figure 8). The robot's status and camera feedback are sent to all operators, and the control commands (i.e. mode change, preview, trajectories generation) are synchronised with all operators.

commands. A similar strategy to the aircraft control by two pilots was adopted, where one pilot is controlling, and the other is monitoring. The controlling operator sends commands to the robot that move the actuators or change control modes and planning trajectories. The monitoring operator can give advice, change feedback acquisition settings, visualise, preview planned movements, and monitor communication network situations. At any time, the roles can be reversed automatically or with necessary approval by the currently controlling operator. The automatic role change could be available to expert operators, while approval is required for spectators or operators in training. The control parameters, modes, and robot status are synchronised for all users. The information exchange was implemented with the use of the Photon Unity Networking framework [66].

D. ARCHITECTURE, INTERFACE MULTIMODALITY, SETUP, AND FUNCTIONALITIES

Before starting the intervention, the operator performs an interface setup according to the workflow presented in Figure 10. According to the mission objectives and environment, available infrastructure for the operator, and robot's configuration, a choice between 2D, 2D+3D, or 3D interface is made. The interface can be visualised on computer screens, in the AR HMD, or mixed (e.g. AR 2D screens in Figure 11). Next, the input devices are selected. Currently, supported robots are CERNBot and TIM, which can be equipped with manipulators, end-effectors, multiple cameras, and other accessories.

The 3D interface visualized with the AR HMD (HoloLens 2) is described in Figures 12, where the robot was in Line Of Sight, and the planning mode was used to move the planning arm, preview, check collisions and move the real arm. The operator saw the interactive camera video (2D) and point cloud (3D) feedback on the left side. In Figure 13, the robot was located underground and controlled from a remote control room. The manipulator and the base were moved by interacting with the arrows in front of the end-effector and next to the base. The collisions were checked continuously with camera video and point cloud feedback.

The operator's workspace appears empty for a person who is not part of the intervention and does not wear the HMD (Figure 14). However, for the operator with the HMD, the workspace is full of detailed information and interactive objects (Figure 15). The figures show a real intervention scenario: the robot was teleoperated from a safe workspace next to the accelerator's entrance while the robot was underground in the radioactive zone.

1) MODEL PLACEMENT IN AR WORKSPACE

The model of the robot can be moved, rotated, and scaled flexibly in the operator's environment. It can be done with one or two hands, with near and far interaction (Figure 16). The model moving is enabled/disabled by voice command or its menu command button (Figure 30).

The robot can be operated on a 1-to-1 scale, which gives the most realistic perception of distances and allows one to see all the robot's details and the environment closely, or it can be minimized as shown in Figures 17 and 18. The robot can be downscaled and placed on the table so the operator can walk around it to check the environment from different viewpoints. The feedback can be composed of video streams from cameras (Figure 17) or point clouds (Figure 18). When all the point clouds are enabled, the operator can have a broad spatial awareness of the environment to avoid collision and drive through tight passages. The video feedback canvases, by default, are projected in front of the camera to indicate the origin of the video stream. However, they can be moved and scaled if that provides a better perspective for the operator or obstructs point clouds.

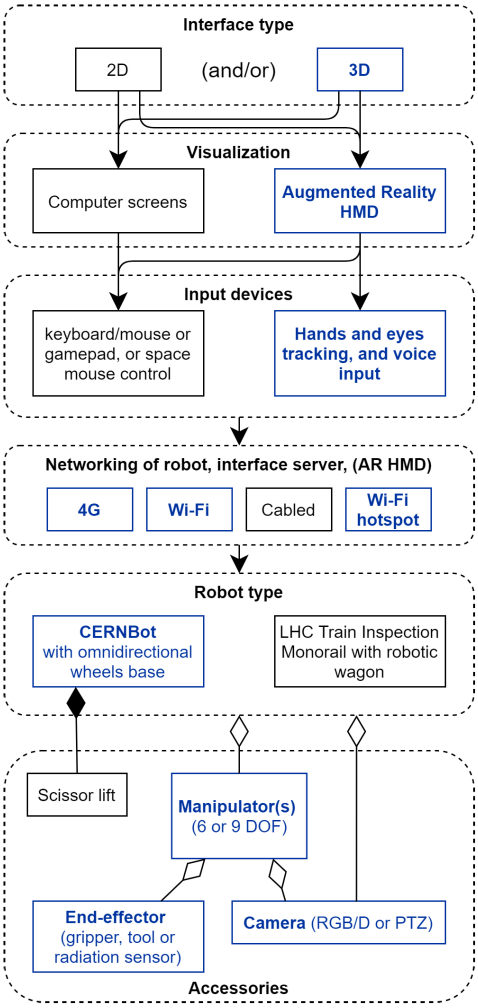

FIGURE 10. The diagram presents the workflow of the operator's interface setup. In this publication, as highlighted in the diagram in blue, the 3D scene is used, visualised in the AR HMD with hand, eyes tracking, and voice input, with 4G and Wi-Fi networking options for CERNBot with omnidirectional wheels, scissor lift, 6 DOF manipulator, RGB-D camera and end-effector.

FIGURE 11. A virtual control centre projected in the AR workspace. The 2D and 3D interface type, computer screen visualisation, keyboard, and mouse are used.

2) ROBOT BASE CONTROL

The robot's base platform can be moved with hand tracking (Figures 15 and 19) or by activating arrows next to the model with hands or eyes tracking with dwell (video demonstration [65]). The hand tracking algorithm is explained in Section II-D4.

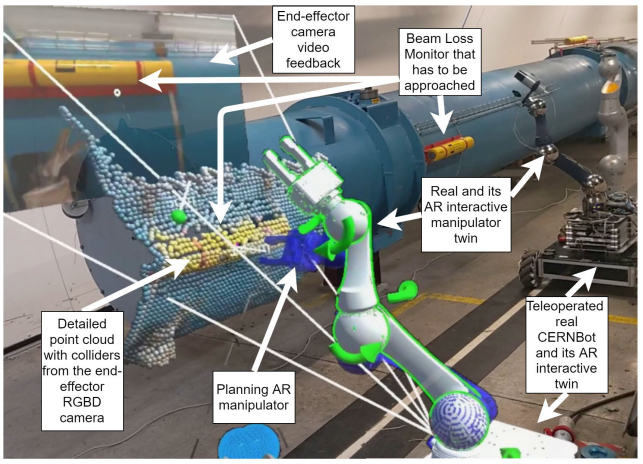

FIGURE 12. The overview of the scene where the operator and the robot were in Line Of Sight, the robot was located in a tunnel with the dipole magnet and Beam Loss Monitor device that the robot's end-effector had to approach. The operator used the robot's hologram (the Holographic Twin) in the workspace next to him. In the foreground, there was the CERNBot manipulator with the planning arm (blue colour) and the real arm (white colour) displayed.

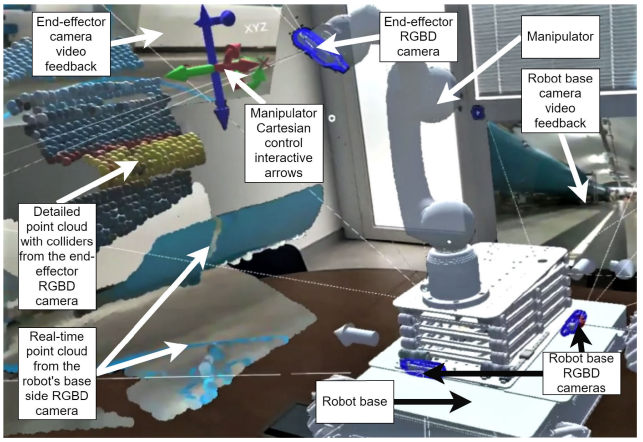

FIGURE 13. The scene overview of the remote control of the arm manipulated in the Cartesian space, the base control, and camera video and point cloud feedback.

FIGURE 14. The external view of the AR HMD operator's workspace. The teleoperated robot was in the CERN North Area radioactive underground zone. The operator's viewpoint is in Figure 15.

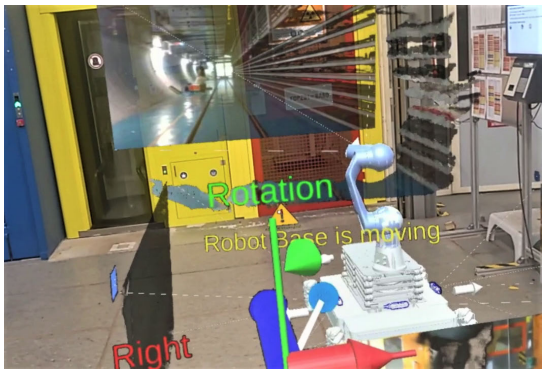

FIGURE 15. The operator's viewpoint, where the video and point cloud feedback are used to traverse the tunnel area, and the hand tracking mode for the robot's base control. The external person's view is in Figure 14.

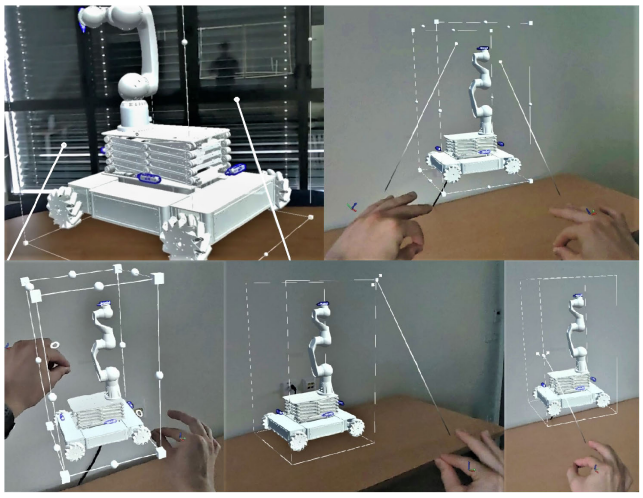

FIGURE 16. Holographic robot model movement, rotation, and scaling with hand near and far interactions.

FIGURE 17. The robot surrounded by video feedback from 5 cameras.

3) MANIPULATOR CONTROL

The manipulator can be controlled in 4 control modes (described in detail in Section II-D of [10]). For the AR

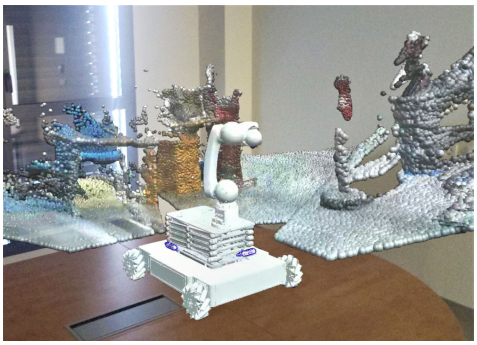

FIGURE 18. The robot surrounded by point clouds feedback from 3 cameras.

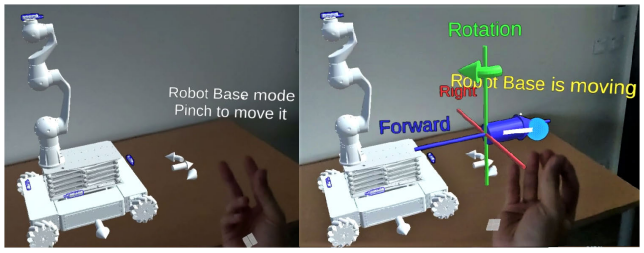

FIGURE 19. Movement of the robot's base with hand tracking. For double confirmation, the palm has to be orientated upwards, and the control is enabled when three fingertips are pinched together. Then the coordinate system appears and the base moves and rotates with speed proportional to the hand displacement and rotation.

control, the following interactions were added to facilitate the use of each of the control modes:

1) In real-time and planning joint mode, the operator can use the hand in near or far interaction (Figure 20), or eye tracking with dwell to activate the arrow that is moving the joint. During the hand-tracking interaction, the joint follows the hand rotation. The hand tracking algorithm is explained in Section II-D4. The joint can also be directly dragged, which causes the rotation around its axis (Figure 21).

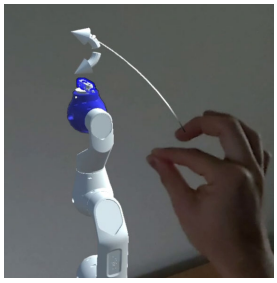

FIGURE 20. A single joint can be controlled with arrows clicked with a hand or eyes and dwell. The arrow corresponds to the direction where the end-effector moves when the joint is rotated.

2) In real-time inverse kinematics mode, the hand translations and rotations can be followed by the arm simultaneously in 6 DOF (Figure 22). This algorithm is explained in Section II-D4. The individual Cartesian

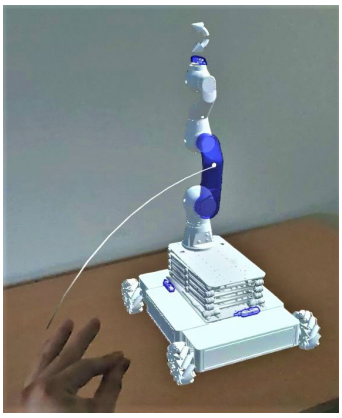

FIGURE 21. Instead of using the arrows (Figure 20), the joint can be rotated directly by clicking on the model and dragging it around its axis.

coordinate system arrows can also be activated by hand or eye tracking with dwell to actuate the movements (Figure 23).

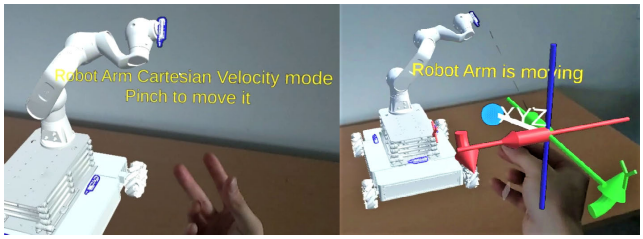

FIGURE 22. The manipulator in the Cartesian real-time velocity mode can be controlled with hand tracking. Similar to the base control (explained in Figure 19, there is the double confirmation system, and the manipulator's end-effector moves and rotates at a speed proportional to the hand displacement and rotation. In the video demonstration [65], it is presented how the robot is operated this way.

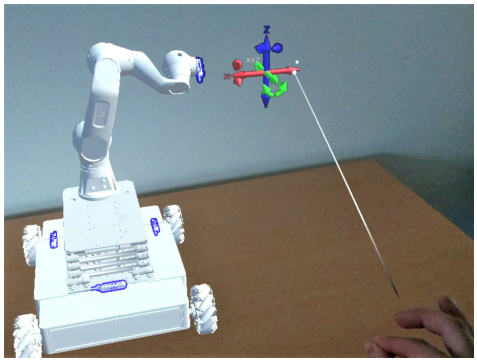

FIGURE 23. In the Cartesian real-time velocity mode, the arm can be controlled by clicking on the arrows or by pointing the arrow with eyes and using dwell (shown in the video demonstration [65]).

3) In the Planning Forward And Backward Reaching Inverse Kinematics (FABRIK) mode, the end-effector target position is moved with hand near or far interaction (Figure 24).

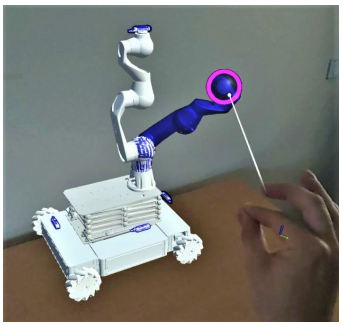

FIGURE 24. The planning inverse kinematic mode allows approaching a target using the FABRIK inverse kinematics. The end-effector target position can be changed by hand interaction.

In each control mode, the speeds are adjusted in the hand menu (Figure 30).

In the planning modes, the trajectory can be specified by creating, replacing, removing waypoints, and collisions can be avoided by launching previews of the movements. The full explanation of trajectories specification, collisions avoidance, sensory and virtual collision detection is provided in Sections II-E, II-F of [10]. The interaction with these functionalities is facilitated by hand or eye pointing, voice commands and gesture recognition (Figure 25).

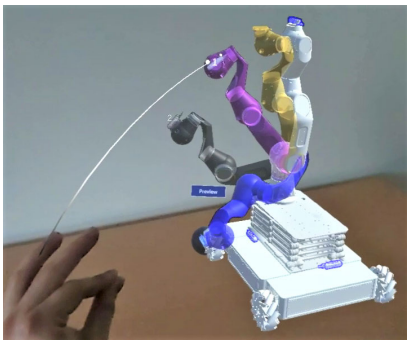

FIGURE 25. Selecting a waypoint and launching preview. In the planning modes, a trajectory with waypoints is created. Then before moving the real arm, the movement preview can be launched. If there is no collision with the environment or self-collision, the movement is started (Figure 26). Explanation of colours: dark grey -> waypoint, violet -> currently pointed (by eye tracking or hand pointer) waypoint, blue -> planning arm, yellow -> preview arm, white -> current real arm position.

The sequence of the automated approach of the end-effector to a normal point in relation to the acquired point cloud of the environment (described in detail in Sections II-G of [10]) is initiated by selecting the point cloud point by hands or eyes tracking and voice command. The normal point direction is based on the surrounding points' positions, and a specified distance offsets the normal point. Then, the point can be adjusted with a hand. These two actions are presented in Figure 27. And finally, in the planning inverse kinematics mode, the arm is automatically positioned at that selected target point. The sequence of selection, and planning arm automatic movement is shown in video demonstrations [65].

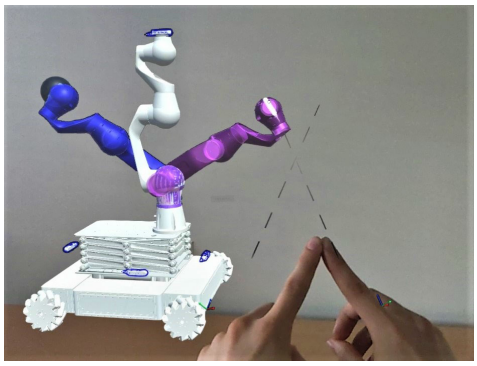

FIGURE 26. Moving the arm to a selected waypoint (violet colour). The movement of the real arm (white color) is executed with a dual hand gesture of touched index fingers. If the gesture is stopped, the robot stops the movement immediately.

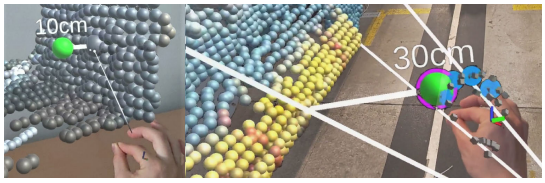

FIGURE 27. In the left picture, the selection of point cloud normal point is done with hand pointing and click or with eyes tracking and voice command. In the right picture, the generated normal point can be moved with hand interaction. The next step is to select the point for the target approach with the planning arm. The whole sequence is shown in the video demonstration [65]).

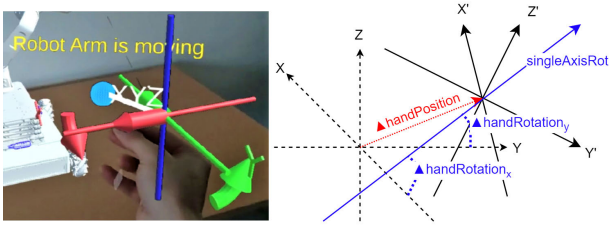

FIGURE 28. The diagram shows the hand tracking algorithm principle. When the user activates the tracking, the initial position and rotation of the hand are saved. Then the user moves the hand, causing the $\Delta handPosition$ linear displacement and a rotation that can be expressed as *singleAxisRot* projected on individual axes and translated into speed commands. The X, Y, Z are the initial hand positioning and X', Y', Z' represent how the hand's coordinate system moved and rotated.

4) HAND TRACKING ALGORITHM

An algorithm was developed to calculate the hand displacements and rotations and then translate them into speed control signals. It is used to control the manipulator with the hand tracking interaction in joint control mode (1 DOF), Cartesian control mode (6 DOF simultaneously), or to control the base (3 DOF simultaneously). The algorithm principle is explained graphically in Figure 28. Below is a full explanation of how the algorithm works with 6 DOF (3 translations and 3 rotations). For 3 DOF or 1 DOF, only the significant axes are taken into account, and the algorithm works similarly. The implementation was done in Unity with the use of quaternions. The position of the tip of the index finger is taken as the hand position, and the palm rotation is taken as the hand rotation. The hand and fingers pose capture and recognition are done by the HMD and calculated using the MRTK 2 library functions. The calculations are continuously performed and the control signals are sent to the robot as long as the hand is rotated up and the fingers are pinched. Otherwise, the movement stops. The hand tracking is initiated when the fingers are pinched, and this defines the starting position and orientation of the coordinate system, which represents the coordinate system of the base or the manipulator (Figures 19 and 22). Any further hand displacement and rotation are input as $\overrightarrow{\Delta handPosition}$ (Equation 1) and $\overrightarrow{\Delta handRotation}$ (Equation 3).

$$\overrightarrow{\Delta handPosition} = \begin{bmatrix} \Delta handPosition_x \\ \Delta handPosition_y \\ \Delta handPosition_y \end{bmatrix} \tag{1}$$

After the initiation, the linear movement of the hand $\overrightarrow{\Delta handPosition}$ is projected ($projOnAxisPos_i$) on each coordinate system axis, according to Equation 2, where $coordSysAxis_i$ are the x, y and z axes of the base or manipulator coordinate system. Then, the projection values are normalized and translated into speed commands (+/−0-100%) according to the graph in Figure 29.

$$projOnAxisPos_i = \overrightarrow{coordSysAxis_i} \cdot \overrightarrow{\Delta handPosition} \tag{2}$$

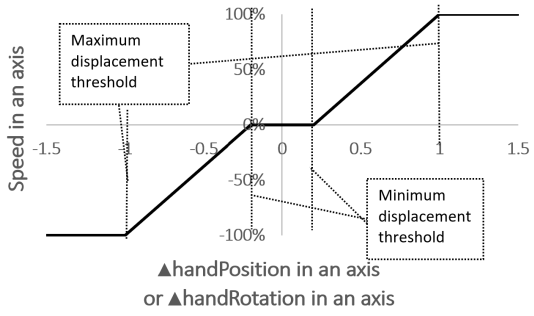

FIGURE 29. Translation of $\Delta handPosition_i$ or $\Delta handRotation_i$ into speed commands in *i* (x, y or z) axis. The movement starts after the displacement threshold (e.g. +/−4 cm or +/−15° and achieves maximum +/−100% at, for example, +/−20 cm of hand displacement or +/−90° rotation. The threshold can be changed in user settings.

The $\Delta handRotation$ shown in Equation 3 is obtained by first finding the $\overrightarrow{singleAxisRot}$ rotation vector and its *singleRotationAngle* magnitude, and then projecting this axis vector on all coordinate system axes. The single axis is obtained by calculating the *handRotation* quaternion according to Equation 4, where *handRotEnd* is the final hand rotation quaternion and *handRotInit* is the initial hand rotation quaternion. Then, by using the Unity function *ToAngleAxis* from *Quaternion* library, the $\overrightarrow{singleAxisRot}$ vector and *singleRotationAngle* value are obtained. The projections on each axis are calculated according to Equation 5. The rotational speed is calculated similarly to

linear speed in Figure 29.

$$\overrightarrow{\Delta handRotation} = \begin{bmatrix} \Delta handRotation_x \\ \Delta handRotation_y \\ \Delta handRotation_y \end{bmatrix} \quad (3)$$

$$handRotation = handRotEnd \cdot handRotInit^{-1} \quad (4)$$

$$projOnAxisRot_i = \overrightarrow{coordSysAxis_i} \cdot \overrightarrow{singleAxisRot} \quad (5)$$

5) HAND MENU, OPERATOR VITAL PARAMETERS, CAMERA ACQUISITION CONTROL, NETWORK MEASUREMENTS

Several parameters of the robot are adjusted using the hand menu (Figure 30). The menu opens when the left hand is turned palm up. Then the interaction can be done with the right hand, eyes tracking and dwell, or eyes tracking and selecting by voice command. The operator's vital parameters (heartbeat, respiration rate, and skin electrodermal activity) are visible in the menu. The menu has a few tabs with different functionalities, the network measurements (Figure 31), camera(s) settings (Figure 32), or commands (video demonstration [65]). The camera settings panel provides with the acquisition parameters, for example, resolution, frames per second (FPS), video and point cloud enable, subsampling, and automatic settings parametrization [10]. The interaction with this panel is possible with hands, eyes tracking and dwell, or eyes tracking and select command.

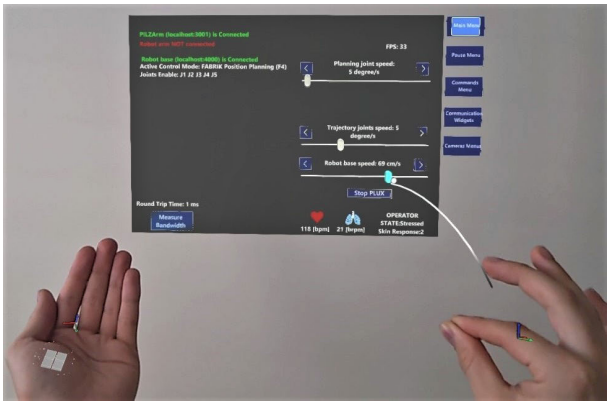

FIGURE 30. The hand menu opened by rotating the hand up and interacting with the right hand to change the speed of the robot base. The interaction can also be done using voice recognition or eye tracking and dwell.

III. EXPERIMENTAL SETUP

The experimental validations of the Interface in real intervention scenarios were performed, and conclusions were drawn based on recorded cameras' acquisition and network parameters measurements while the operator(s) were executing tasks necessary for the intervention. The tasks were done by several operators, both experts, and beginners, to provide broad user feedback. The experiments workflow is shown in Figure 33. The quantitative and qualitative validation conditions are explained in Section III-A. The network setup for single-user validation is presented in Section III-B and for multi-user

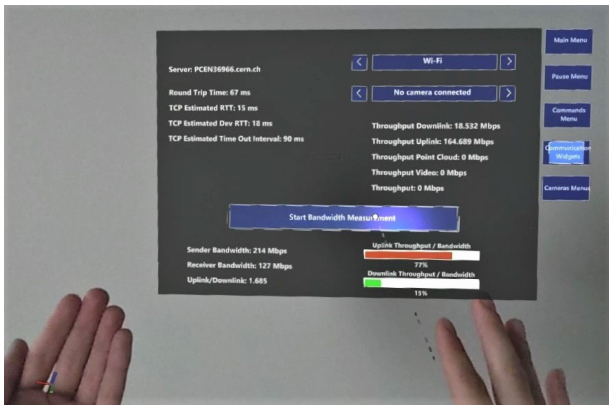

FIGURE 31. Network and communications menu showing parameters related to the network (i.e. bandwidth use and its measurements, delays, throughput related to camera feedback over selected network interfaces.

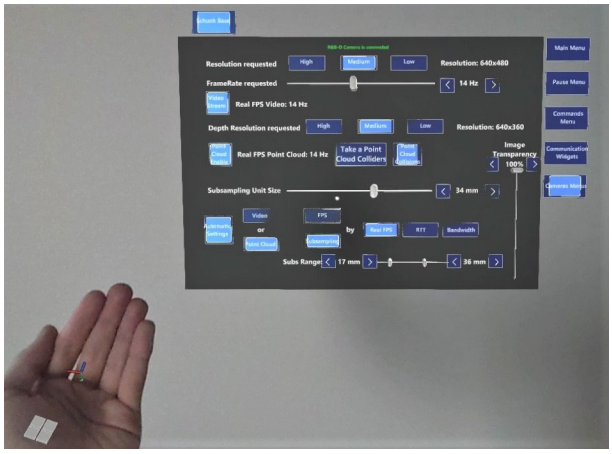

FIGURE 32. The camera settings can be accessed in the hand menu (Figure 30), which allows reaching the settings panels quickly.

validation in Section III-C. The network connections characterization had been performed in the previous work [10] in Section III-A-I, which presented a practical comparison between using 4G modem, Wi-Fi over CERN network, cabled CERN network connection and direct Ethernet connection in terms of bandwidth, round-trip time and jitter. However, these values might depend on location, signal strength, and quality fluctuations due to antenna orientation or obstacles. Additionally, the CERN network infrastructure is used by other users and shared among hundreds of thousands of communication nodes. The 4G network is publicly shared with transceivers also outside CERN premises. These effects were expected and registered during experiments. Before each experiment was started, the bandwidth was measured to best represent the prevailing conditions, and the value was presented in the results. During each experiment, network parameters, such as round-trip time, throughput, all camera parameters, and 4G network signal strength, were recorded to study the behaviour of the delays, feedback acquisition stability, and bandwidth availability. From these network measurements, it was concluded if the CERN infrastructure could support

AR operations in different scenarios with selected network connections, as well as if additional network protocols, automatic congestion control mechanisms, or other communication architecture are needed in future work (with the focus on the multi-user collaborative operation).

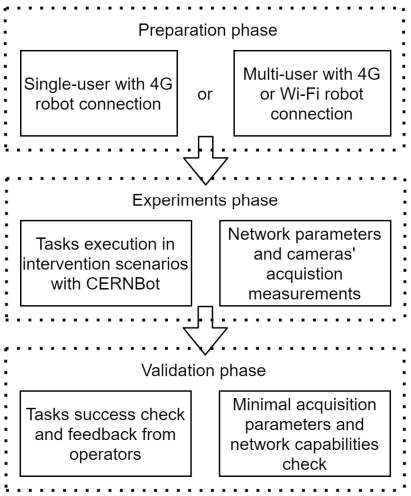

FIGURE 33. The workflow of each experiment. The single or multi-user interface was used with 4G or Wi-Fi connections to the robot. Then the tasks were executed with parallel measurements. Finally, the validations of tasks and network behaviour were done.

A. VALIDATION CONDITIONS FOR SINGLE AND MULTI-USER OPERATION

The teleoperation tasks in an intervention, that had to be successfully executed, are:

- Navigating the CERNBot robotic base in an unstructured environment (e.g. in a tunnel or close to particle accelerator equipment).
- Manipulator operation and performing a detailed inspection of a piece of equipment or a location, or a task requiring physical action with a gripper. Real-time and planning modes choice, as well as the use of collision avoidance mechanisms, were at the operator's convenience.
- The operator controlled the video and point cloud feedback settings according to the task's perception needs or network limitations. Manual or automatic control of camera parameters was available.

The operator could use any available interaction modality (hands, eyes, voice). During each experiment, a video recording of the intervention from the operator's HMD perspective was saved for the AR interface feasibility study and task completion analysis. Based on these results, it was qualitatively concluded if the operation was satisfactory and if the tasks were completed.

Minimum conditions were set for cameras' acquisition and network parameters quantitative validation:

1) The FPS of the interface processing and streaming to the HMD must be minimum 25 Hz.
2) The FPS of the main camera, which is used for moving the robot's base or the manipulator and on which the operator is focused, is minimum 5 Hz.
3) The FPS of secondary cameras, used for periodic collision checks or peripheral view, is minimum 1 Hz.
4) The point cloud subsampling must be maximum 40 mm for moving the robot's base and minimum 25 mm for the manipulator approach.
5) The round-trip time must be below 200 ms during any movements or manipulation.

B. NETWORK SETUP FOR SINGLE-USER OPERATION VALIDATION SCENARIOS

The single-user experiments were focused on the most common setup during a real intervention, where the robot was connected to the 4G network (Figure 34). This connection offers the lowest bandwidth and highest delays and is more prone to interferences or variations than other connection types. The remote operator's HMD and the interface server connected via CERN Wi-Fi infrastructure, and they were always less limiting than the robot's 4G connection.

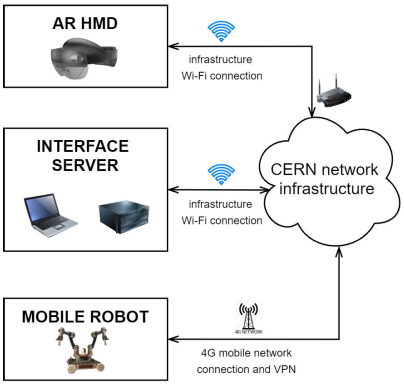

FIGURE 34. Single-user validation scenario: robot (4G), HMD and interface (CERN Wi-Fi).

C. NETWORK SETUP FOR MULTI-USER OPERATION VALIDATION SCENARIOS

In the multi-user experiments, two collaborating operators performed an intervention together. The operator stations were deployed in the field. The operators used the same physical workspace. The validation of the two following network connections was done and presented:

1) **The robot connected to the 4G network, the interface servers and the HMDs connected via CERN Wi-Fi infrastructure**. It is the most common scenario, where the robot is deployed in a hazardous unstructured environment and the operators are deployed and cooperating in the field with Wi-Fi coverage.
2) **The robot, the interface server and the HMD connected to CERN Wi-Fi**. This scenario is optimized for portability and flexible locations of servers and operators. However, it requires the robot to have access

to the Wi-Fi network available only in less radioactive areas.

The architecture of the two connections is shown in Figure 35. The operators connect separately to their interface servers via CERN network infrastructure, and the servers connect independently to the robot.

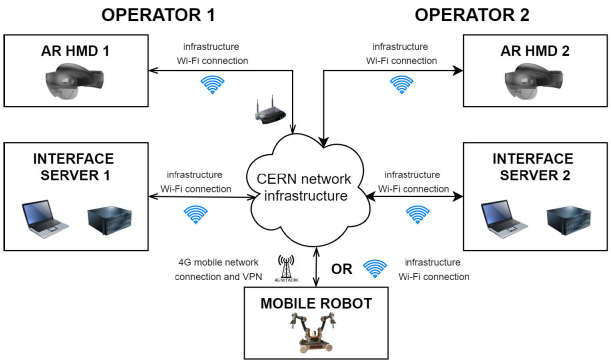

FIGURE 35. Multi-user network connections validation scenarios: robot (4G or CERN Wi-Fi), HMD and interface (CERN Wi-Fi).

D. HARDWARE

The experiments were undertaken using portable gaming laptops as interface servers (one for single-user and two for multi-user) with the characteristics shown in Table 6. The used HMD was Microsoft HoloLens 2. The robot's computer characteristics are shown in Table 7.

TABLE 6. Interface server characteristics.

Device model	MSI GL75 Leopard 10SFR, rev 1.0
CPU	Intel® Core™ i7-10750H
GPU	NVIDIA GeForce RTX 2070
Memory	32 GB, DDR4, 1333 MHz
System	Windows 10 Pro, version 20H2
Benchmark	Graphics 7435, CPU 6622 (3D Mark TimeSpy v1.2

TABLE 7. Robot's computer characteristics.

Device model	Intel NUC11TNKv70QC
CPU	Intel® Core™ i7-1185G7
GPU	Integrated, Intel
Memory	64GB, DDR4, 3200 MHz
System	Linux Ubuntu 20.04.2 LTS
Benchmark	68000 with 1 thread, 127358 with 8 threads (CPU-X v3.2.4, slow prime numbers)

The used robot was the CERNBot equipped with one manipulator and five RGB-D cameras (one in the end effector and four on the base - one each side). For communication on the 4G network and Wi-Fi, the robot used the mobile router Teltonika RUT955. The robot is depicted in Figure 36.

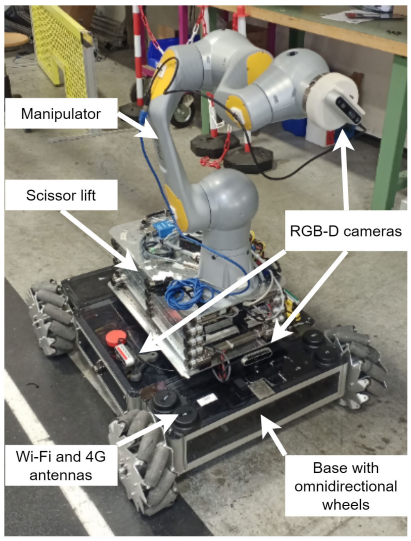

FIGURE 36. The CERNBot used for experiments.

IV. RESULTS

This section presents the results of the experiments. The operators' feedback is described in Section IV-A. The measurements are divided into two groups: single-user validation scenarios (IV-B) and multi-user validation scenarios (IV-C). Each group was analyzed according to the qualitative and quantitative conditions specified in Section III. For each scenario, a set of graphs shows the behaviour of the acquisition of cameras' feedback and the network in critical moments (such as driving the robot; changing from base control to manipulator control; overload and collapse of the network - saturation of bandwidth and increased delays), or when automatic congestion control [10] was used to judge its efficiency in multi-camera and multi-user contexts. Since the amount of recorded data gathered during all experiments is extensive (lasting up to 30 minutes per each of 6 network connection pathway testing, and per each of 12 automatic modes, repeated for single multi-user scenarios), only the relevant validation sections of recordings are presented. For example, the presented recordings show the most limiting network or operational situations, a problem, or a relevant event that occurred.

A. OPERATORS FEEDBACK

In total, the interface was tested by ten (10) operators, who performed several tasks for 45-240 minutes in total per operator. They provided detailed feedback on functionalities, interactions, inconveniences, or potential improvements, which are discussed in Section V. Five (5) operators had previous teleoperation experience using standard screen interfaces (defined as more than 30 hours of operation experience). Thus, they could compare the user experience with the screen-based and the AR HMD Mixed Reality versions. Based on the feedback, several conclusions were made:

1) The environment awareness increases thanks to the point cloud representation and the freedom to walk

around it to see from different viewpoints were most appreciated in the AR compared to standard 2D video feedback or MR on screens. The sound feedback was also considered a significant improvement because it added another feedback dimension that was not available in previous CERN interfaces.

2) The hand-tracking interaction is convenient, although it requires training to understand the boundaries of the hand-tracking space captured by the HMD. For example, while moving the base or the manipulator and observing the environment with head movements, the HMD sometimes lost track of the hand. Also, the wrist is limited in rotations (yaw, pitch and roll angles), and the limits differ depending on the operator's wrist anatomy. For two operators, the rotations were uncomfortable or beyond the wrist joint limit in certain rotation planes, and for others, they were acceptable. The linear movements were intuitive and comfortable.
3) Interaction with buttons and sliders initially posed a problem because it has only audio and visual feedback and lacks haptic feedback. Some operators had a problem with pinching the slider or pressing a button. But it became less problematic with training and when the buttons and sliders were bigger.
4) The eyes calibration had to be performed individually for each operator. Otherwise, pointing with eyes for most users had an unacceptable offset. The calibration requires a few minutes of pause in operation. Therefore, in time-pressured operations, it was more convenient not to exchange the HMD between operators and to have multiple HMDs calibrated individually.
5) The voice command recognition worked well. However, the commands must have been selected carefully not to be easily confused with similar words or sequences but still be quick to repeatedly pronounce and easy to remember.
6) The multi-user collaboration was efficient. The operators could easily communicate intentions or discuss the best way to execute tasks. There were events when a monitoring operator noticed a potential collision and warned the controlling operator, which as a result, increased safety.

B. SINGLE-USER EXPERIMENTAL MEASUREMENTS

1) ROBOT (4G), HMD AND INTERFACE SERVER (CERN WI-FI)

- **Minimally required parameters for cameras acquisition validation**. In Figure 37, there is an experiment, where the minimally required camera acquisition parameters are requested. The main camera used video and point cloud feedback with 5 Hz FPS, and four cameras with video feedback of 1 Hz FPS. The subsampling was fixed at a constant level to not influence the point cloud point number. The cameras' resolutions were also constant at medium levels (e.g. 640 × 360), enough for base driving and a coarse manipulator approach. During the experiment, the robot was moving, which resulted in a varying point cloud throughput due to changing point cloud points number of the environment captured by the depth sensor of the camera. The requested FPS of all cameras was achieved and the acquisition was stable. The interface FPS was stable at around 30 Hz, which provided a smooth interaction. This experiment confirmed that the minimum parameters could be achieved for a single-user operation with the 4G connection of the robot.

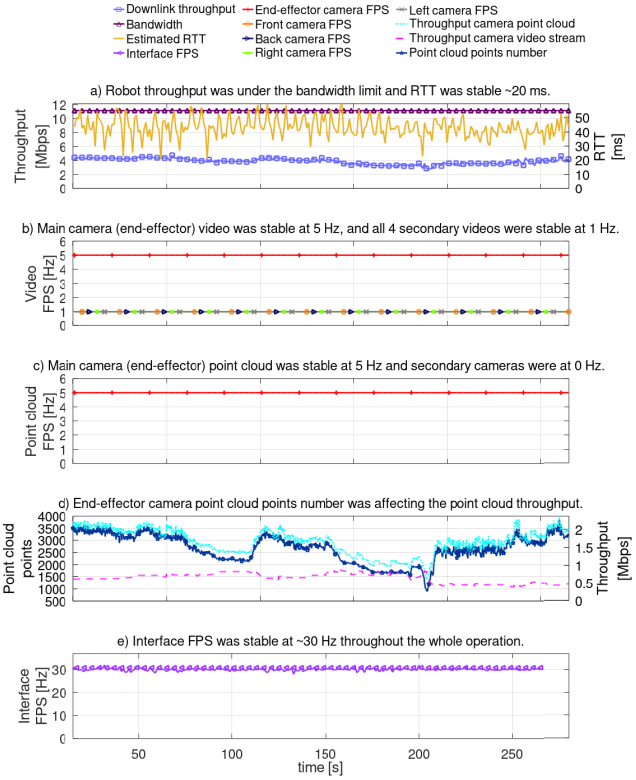

FIGURE 37. Single-user, 4G robot connection measurements to test the minimally required acquisition parameters for a single-user operation. In graphs b and c, there are stable FPS of the main camera (5 Hz) and four secondary cameras (1 Hz) for point clouds and video. In graph a, the RTT was between ~25 and ~50 Mbps, and the throughput was at a stable value ~4 Mbps. The bandwidth limit was 11 Mbps. In graph e, the Unity game rendering FPS was stable at 30 Hz. The FPS parameters were stable despite varying point cloud point numbers, which as an example of one camera, is shown in graph d.

- **Network congestion situation of reaching the maximum throughput**. In Figure 38, there is a recording presenting a network congestion situation when the maximum bandwidth usage was reached. Compared to the previous example, the throughput was slightly higher and there was a temporary bandwidth limitation. Each of the five cameras requested a video stream of 5 Hz FPS. The video resolution was constant. It resulted in oscillations of cameras FPS and video throughput as the transmission was irregular due to frames queuing. As a result, the round-trip time (RTT) was also fluctuating. The operator sees in this situation a warning about delayed frames and manually adjusts or uses the automatic

settings. Although the FPS could be achieved on average, the oscillations were unpredictable and uncomfortable for operation. This presents a boundary acquisition state, which is essential to notice.

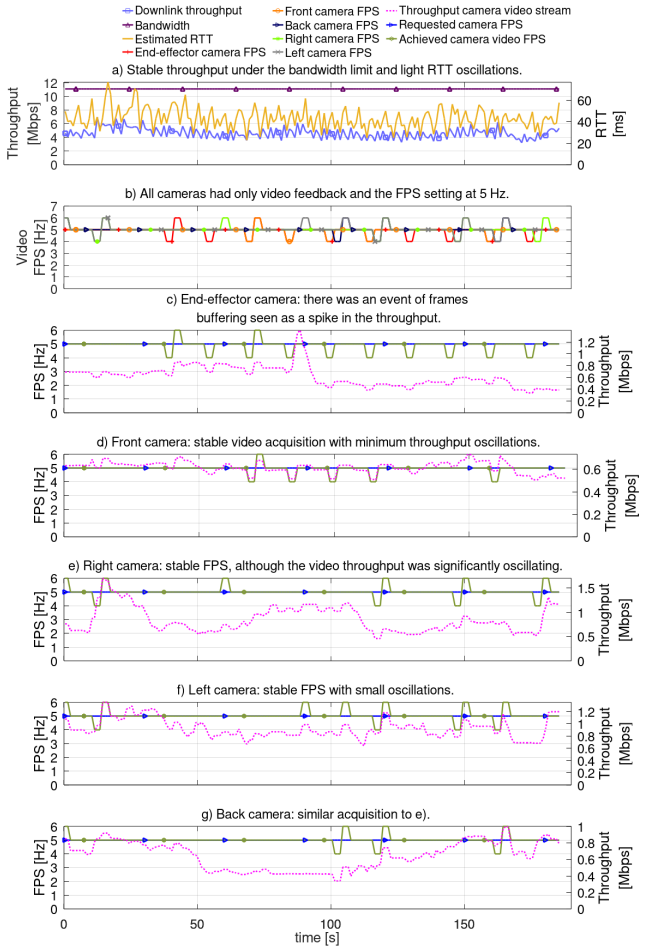

FIGURE 38. Single-user, 4G robot connection measurements of network congestion situation of reaching the maximum throughput (graph a). Throughout the whole period, all cameras were oscillating between 4 Hz and 6 Hz due to frames queuing (seen in graph b for all cameras and individually for each camera in graphs c-g). This example presents that the limit in that configuration was five cameras with video feedback of 5 Hz. In each camera measurement, it is visible that the video throughput was oscillating due to buffering (especially in graphs c and e).

- **Automatic video FPS for cameras to reach a bandwidth usage target**. In Figure 39, there is a recording where each of the cameras used the automatic adjustment of FPS to achieve a requested bandwidth target of 20% per camera individually (∼2 Mbps), which in total gave 100% use of the initially measured bandwidth. This situation allowed maximising the network's use to have as frequent feedback as the bandwidth and dividing the throughput of cameras equally. Also, when the camera bandwidth setpoint was correctly selected, it helped to avoid queuing. This mode should not be used in a variable bandwidth environment because the bandwidth would need to be measured often, but it requires disabling all camera feedback. The alternative is the automatic mode in which the RTT controls the FPS setting, and the RTT is measured continuously.

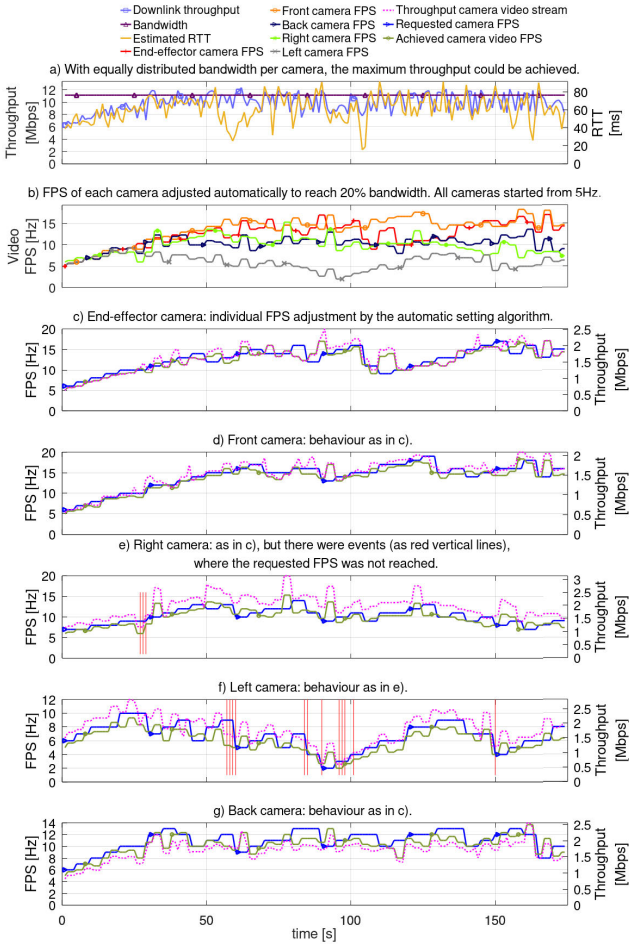

FIGURE 39. Single-user, 4G robot connection measurements of automatic video FPS for cameras to reach a bandwidth usage target. Each camera had an activated automatic mode to reach a 20% bandwidth usage target (total throughput shown in graph a, and individual camera throughputs in graphs c-g). All cameras started initially from 5 Hz and were adapted to achieve a throughput of around 2 Mbps. Due to the different resolutions of cameras, the finally achieved FPS per camera were different (graph b). The measurements also present a situation when 100% bandwidth is reached, they started competing for the bandwidth, and that caused oscillations.

- **Main camera automatic optimization of point cloud FPS to achieve a bandwidth target**. In Figure 40, there is an experiment, where the main camera had point cloud and video feedback enabled, and four other cameras requested video-only feedback of 1 Hz FPS. The main camera used the point cloud automatic FPS adaptation to bandwidth target. During the robot's movement, the operator decreased subsampling to increase the resolution of the point cloud. However, that resulted in increased throughput. The adaptation algorithm automatically reduced the FPS setting to keep the throughput at the same level. With smaller subsampling than 30 mm, there were more events of delayed FPS due to network

congestion, and with 18 mm subsampling, the FPS was only 2-3 Hz.

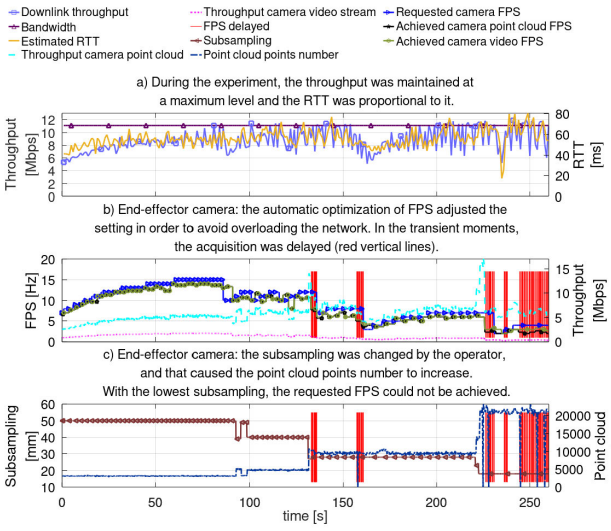

FIGURE 40. Single-user, 4G robot connection measurements of automatic optimization of FPS of 1 camera to achieve a bandwidth target (graph a). The other four cameras had fixed 1 Hz video feedback. During the recording, the robot was moving, and the operator changed the subsampling setting for a better point cloud resolution, significantly changing the point cloud point number captured by the depth sensor (graph c). That triggered the automatic optimisation to lower the requested FPS (graph b) to maintain the throughput. The vertical lines represent events when the real FPS is lower than the requested FPS by at least 2 Hz, and this is due to the delay of the algorithm and a temporary network overload.

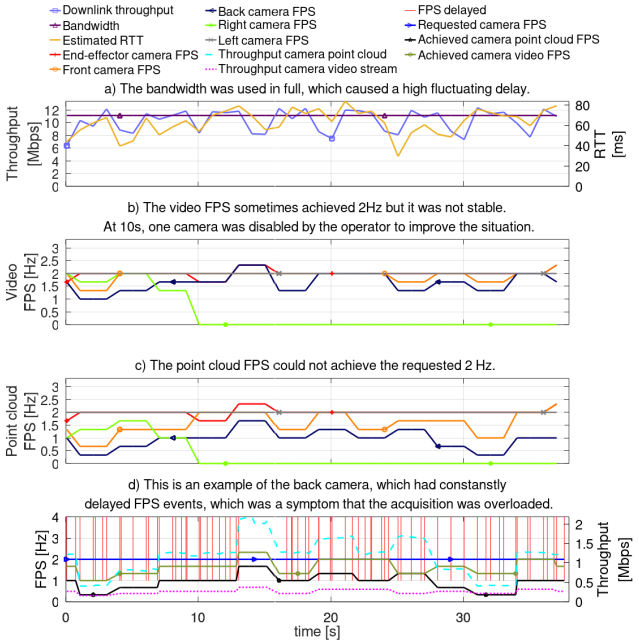

FIGURE 41. Single-user, 4G robot connection measurements when the system was overloaded by requesting 2 Hz of point cloud and video (graphs b and c). The example presents a situation when the requested parameters were too demanding for a given network connection. The RTT was 40-80 ms, following the throughput fluctuations (graph a). In graph d, there is an example of one of the cameras. The FPS was constantly delayed (shown as events with red vertical lines), and the throughput was significantly oscillating, especially the one of the point cloud.

- **Overloaded system by requesting too demanding acquisitions**. In Figure 41, there is an example of an overloaded system, where each camera was requested to send point cloud and video with 2 Hz FPS. At 10 s, one camera was disabled, which slightly improved the situation, but still, the bandwidth was not enough to support four other cameras with the requested FPS, as their acquisition struggled to keep up. In the throughput and RTT graphs, it is visible that as soon as throughput exceeded the bandwidth, there were collapses of the streaming, and the situation was regularly repeated. These circumstances should be avoided in teleoperation because it introduces high delay fluctuation, instabilities, and lost control packages.

C. MULTI-USER EXPERIMENTAL MEASUREMENTS

1) ROBOT (4G), HMD AND INTERFACE (CERN WI-FI)

- **The maximum FPS of a camera with point cloud and video feedback**. In Figure 42, the experiment tested the maximum FPS of one camera with point cloud and video feedback. The requested FPS was gradually increased to check if the network could support it. As shown in the graphs, the value of 9 Hz for both users was achieved while maintaining the throughput between 75% and 100% of the bandwidth of each operator. During the recordings, there were a few events of delayed FPS, which could be caused by an abrupt change of the setting (the camera acquisition in the robot had to be reconfigured) or network overload.
- **The maximum FPS of one camera with video feedback**. Similarly to the previous example, the experiment sought the maximum FPS value for both operators by gradually increasing the video FPS until the network support limit. In the test, stable 15 Hz was achieved for both operators. The resolution of 640 × 360 was used in the camera settings.
- **A temporary collapse of the network**. In Figure 43, there is an example of a collapsed network communication for both operators due to a spurious event (such as a temporary degradation of 4G signal or interference). The most characteristic symptom is the highly increased RTT value, which in this example rose four times from 15 ms to 60 ms. For 5 seconds, the first operator completely lost the feedback from one camera, while the second operator lost the point cloud streaming but could still receive 2 Hz of video. This situation may happen unexpectedly at any time. Having two operators introduces a redundancy, where one operator can focus on controlling the robot while the second monitors the network state and the acquisition of environmental feedback from sensors and cameras.

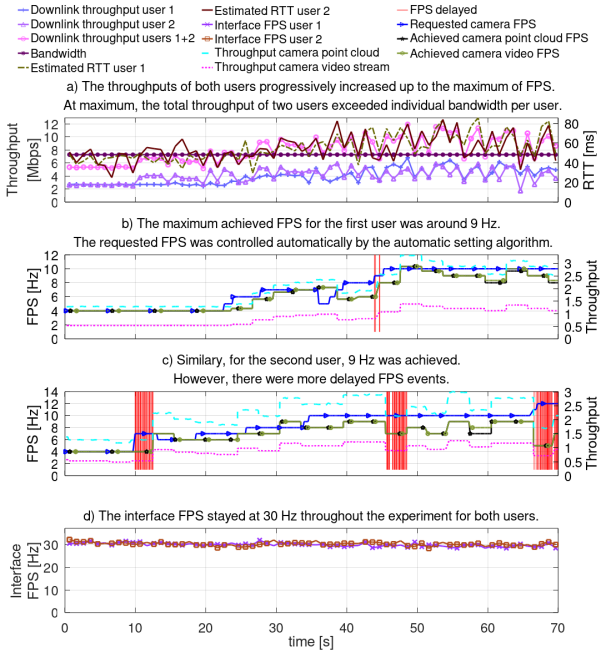

FIGURE 42. Multi-user, 4G robot connection measurements of the maximum FPS of one camera with point cloud and video used by two users. In graph a, the individual and total throughput progressively increased, and the total throughput exceeded the bandwidth per user. In graphs b and c, the FPS progressively increased, and 9 Hz was the maximum setting for both users. The interface FPS was stable at 30 Hz for both users (graph d). In graphs b and c, the evolution of the settings, real achieved FPS, and video/point cloud throughput of the camera is also shown. Also, the delayed FPS warnings (vertical red lines) indicate events when the real FPS was lower than 2 Hz than the requested setting, which was more frequent for user 2 (graph c).

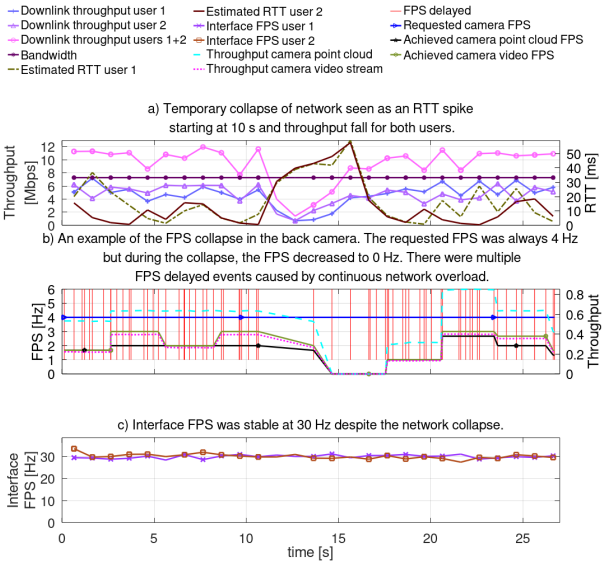

FIGURE 43. Multi-user, 4G robot connection measurements of a temporary event of network collapse for both users. It is characterized by the drop of throughput (graph a) and FPS to zero (graph b) and a high increase of RTT starting at 10 s and lasting 5 s (graph a). Before and after the event, the network was overloaded, which resulted in delayed FPS represented by red vertical lines in graph b. During the event, the game FPS was kept at a stable 30 Hz value (graph c).

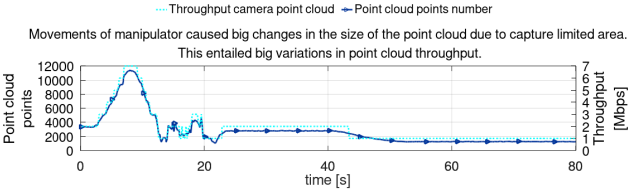

FIGURE 44. Multi-user, Wi-Fi robot connection measurements of dynamic point cloud size changes while driving the robot. In the first phase, the camera captured maximum ∼11000 points, and then the number decreased to ∼1300 points due to the move. Accordingly, the throughput used for this point cloud acquisition with a constant FPS was at maximum 7 Mbps and decreased to ∼1 Mbps, which was of a similar value to the video throughput with the same FPS.

2) ROBOT, INTERFACE AND HMD CONNECTED TO CERN WI-FI

- **Dynamic point cloud size during manipulator's movement**. In Figure 44, there is a specific example of how the point cloud throughput could change ∼8 times just by moving the robot's manipulator without changing any setting related to point cloud acquisition. In a limited network bandwidth scenario, it could cause a slowdown or a collapse of the network. In this example, the automatic settings modes were not used, but they could adapt the FPS or subsampling according to the situation for network use optimization and avoid collapses. Such an example is presented in the "Automatic FPS adaptation to point cloud size" experiment and Figure 46. The automatic behaviour also can lower the operator's workload of continuously monitoring the acquisition status.
- **Throughput and RTT linear relation**. In Figure 45, a direct relation between throughput and RTT can be observed, which is important for telemanipulation, especially with force feedback. When the throughput was decreasing, the RTT followed accordingly in a linear relation for both operators.
- **Automatic FPS adaptation to point cloud size**. As shown in the previous example in Figure 44, the size of point cloud can change significantly during the robot's movement while keeping the same settings. In Figure 46, the automatic FPS setting of the point cloud was used to adapt to the point cloud points number. When the number of points increased, that caused a delayed FPS flag and a decrease of the requested FPS not to overload the network. Accordingly, when the number of points decreased, the requested FPS setting increased to provide more responsive feedback with the available bandwidth.

D. VIDEO EXPERIMENTAL DEMONSTRATIONS AND MEASUREMENTS RECORDINGS

The video recordings from the multi-user operation scenario, from both operators' HMD's point of view, can be found in [65]. In the example video, the first operator set up the cameras' acquisition, drove the robot's base to a target location to

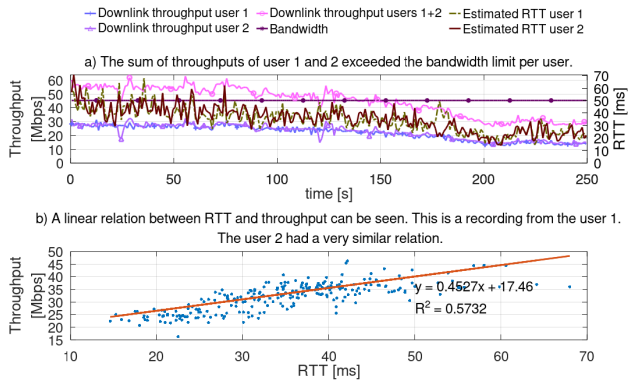

FIGURE 45. Multi-user, Wi-Fi robot connection measurements of the relation between throughput and RTT. In graphs a and b, the proportional relation between throughputs and RTT can be noticed. When the throughput declined, the RTT also proportionally decreased, for example, when the total throughput (both users summed) changed from ∼55 Mbps to ∼28 Mbps, and the RTT lowered from ∼40 ms to ∼20 ms.

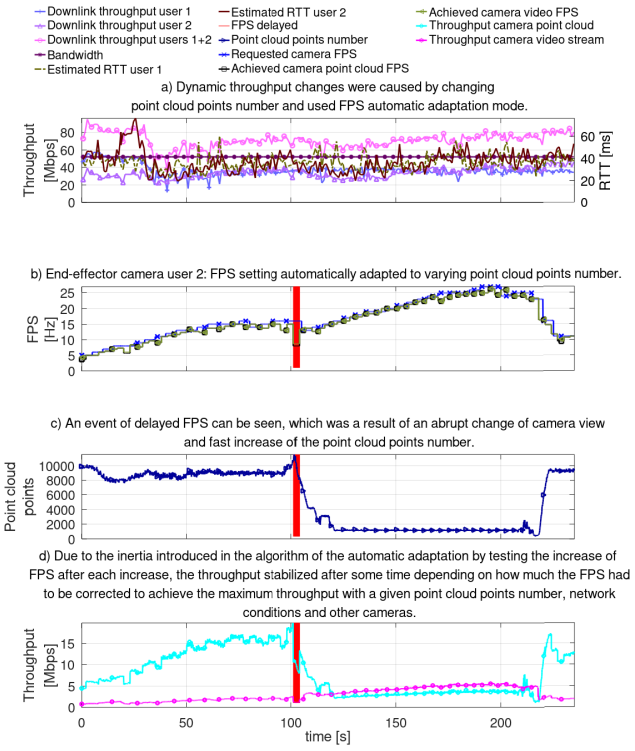

FIGURE 46. Multi-user, Wi-Fi robot connection measurements of automatic FPS adaptation to point cloud size. As seen in graph c, the point cloud points numbers were changing between ∼1000 and ∼13000 when the manipulator was moving and changing points of view. During these changes, the requested FPS was automatically adapted to values between 5 Hz and 25 Hz to stabilize and maximize the throughput (graph a). It is visible that abrupt changes in the point number caused delayed FPS events (vertical red lines in graphs b-d), but after a short time, the acquisition followed the requested FPS.

approach a magnet dipole (a piece of equipment in the particle accelerator that bends the particle beam) with a manipulator to read an inscription on it. In another example video, there is a recording in which the minimally required parameters for cameras' acquisition validation were checked with the 4G robot connection, described in Section IV-B. The network measurements and camera acquisition parameters recordings of all experiments presented in this publication are available in [67].

E. RESULTS SUMMARY

Table 8 specifies if each requirement set in Sections III-B and III-C was achieved and under which conditions, if any.

TABLE 8. Summary of the experimental results and fulfilment of the interface requirements for single and multi-user teleoperation.

Validation condition	Single and multi-user scenarios with 4G robot connection
Movement of the CERNBot robotic base in an unstructured environment	✓(all operators successfully moved the robotic base through a tight passage with obstacles)
Manipulator operation to perform a detailed inspection of equipment or location, or a task requiring physical action with a gripper	✓(all operators successfully moved the manipulator to inspect a piece of equipment)
The operator controlled the video and point cloud feedback settings according to the task's perception needs or network limitations.	✓(all combinations of video and point clouds acquisition of 5 cameras and manual or automatic settings control were used)
The FPS of the interface processing and streaming to the HMD must be minimum 25 Hz	✓(if the laptop was power supplied. It could go down to 15 Hz if battery powered)
The FPS of the main camera that is used for moving the robot's base or the manipulator and on which the operator is focused, is minimum 5 Hz	✓(multi-user for both users equally 9 Hz of point cloud and 15 Hz of video with medium resolution)
The FPS of secondary cameras, that are used for intermittent collision check or peripheral view, is minimum 1 Hz	✓(in multi-user main camera 5 Hz, and 4 secondary cameras of 1 Hz, all with point cloud, accordingly main 11 Hz, and 4 secondary 1 Hz, all with video)
The point cloud subsampling must be minimum 40 mm for moving the robot's base and minimum 25 mm for a precise manipulator approach	✓(1 mm precision is available for a small region of point clouds, however usually 20-30 mm can be achieved with sufficient FPS, min. 5 Hz)
The round-trip time must be below 200 ms during any movements or manipulation	✓(maximum 90 ms in single and multi-user)

A phenomenon of increased total throughput (on average by 20%) when summed for two users compared to a single-user connection was observed. The intensity of this increase varied in time and location, but it was always present. The average values are shown in Table 9. The reason may be related to the communication links created for each user separately from the interface server to the robot.

A basic comparison of bandwidth achieved with different connection types from the AR HMD (HoloLens 2) was made for 2.4 GHz and 5 GHz Wi-Fi hotspot created by the laptop, CERN Wi-Fi infrastructure and by cable Ethernet connection with a USB-C connector, which is presented in Table 10. It can be seen that the hotspot connection provides only limited bandwidth, so it was impossible to use the 2.4 GHz network band. However, the 5 GHz hotspot was enough to send

the streaming to the HMD via the Holographic Remoting and receive fast inputs from the HMD. The other connections were not limiting.

TABLE 9. Comparison of maximum throughput for single and multi-user measurements.

Comparatory parameter	Single-user	Multi-user (2 users)
Average maximum throughput in 4G robot connection	7 Mbps	10 Mbps
Average maximum throughput in CERN Wi-Fi robot connection	45 Mbps	55 Mbps

TABLE 10. Bandwidth measurements between HoloLens 2 AR HMD and an Ookla server in Zurich, done via the website www.speedtest.com, launched in the HMD. The measurements can be compared between different connections, but direct values should not be used.

Connection	Ethernet cable with USB-C adapter	CERN Wi-Fi 5 GHz	Wi-Fi hotspot 2.4 GHz	Wi-Fi hotspot 5 GHz
Downlink bandwidth	188 Mbps	176 Mbps	10 Mbps	28 Mbps
Uplink bandwidth	270 Mbps	176 Mbps	10 Mbps	15 Mbps

V. DISCUSSION

All the functional requirements for the interface were achieved, and each operator performed successful teleoperation. The network performance in the worst case - 4G robot connection - was sufficient to provide enough point cloud and video feedback. However, as provided in the experimental data, situations and events may require more adaptive behaviours and high-level acquisition management. Therefore, if the adaptive behaviour was necessary due to dynamic changes in the network, the use of the automatic settings was demonstrated. The interface minimum FPS was lower than expected when the interface server (with parameters specified in Table 6) was powered by its internal battery. Otherwise, it was sufficient and stable. This inconvenience can be overcome by using a portable computer with higher graphical processing parameters. Also, the minimum expected FPS for cameras was achieved in the multi-user scenarios with a 4G connection to the robot. Still, spurious network effects could temporarily lower it, which was mitigated by using the camera's automatic settings.

In the multi-user scenarios, the operators were equipped with the same functionalities as in the single-user scenarios and the additional possibility for the other operator to observe or take control. Therefore, it could be implied that if the teleoperation task was achieved in the single-user operation, it could be completed in the same functional way in the multi-user process. Also, multi-user control capability provides a degree of redundancy, if the connection is temporarily degraded for one operator or the person is tired - the other operator can easily take over and continue the intervention.

In the used multi-user network architecture, a higher total bandwidth was sometimes achieved than each user's bandwidth due to multiplied connection links, especially with the complex Wi-Fi architecture and when both users were in different locations or used various infrastructure links. It was also observed that the network behaviour (variations in bandwidth, spurious delays, temporary streaming collapses) varied depending on the time of the day for the public 4G, Wi-Fi, or cable infrastructure, which depends on infrastructure clients' number and their usage.

It can be seen that the usage of point clouds streamed with a specified frequency cause significantly higher throughput than video feedback of the same frequency. Therefore, we can assume that if the bandwidth is below 3-5 Mbps, the current interface cannot support any more point cloud feedback with sufficient FPS and precision for an average-sized captured point cloud. However, there are still measures to increase the bandwidth by selecting a more performant modem/router or providing more optimized streaming of the point cloud, which is proposed as future work in Section VI. Also, specific network configurations could be used, for example, the added Access Control List (ACL) to a socket or destination address in an infrastructure router configuration or a Quality of Service (QoS) change could lead to an increased bandwidth in a publicly shared Wi-Fi network. Moreover, in the areas where the 4G coverage is poor or unavailable, multiple relay robots acting as communication nodes could be used to extend the range [55].

The field-of-view (FOV) problem of RGBD cameras on the CERNBot's setup base was noticed during the tests. Due to the camera's narrow FOV angle - 87 degrees in horizontal axis - and placement offsets from the centre of the robotic base, there were blind areas at the corners of the robot, even with four cameras around the base. It can be seen especially in the point clouds environment visualization (Figure 47). The problem could be mitigated by placing the cameras closest to the centre of the robot, which was not possible for CERNBot as the camera view would be covered by other equipment. It would also be possible by switching to two LiDAR sensors for point cloud in two opposite corners of the base for full point cloud view; and 360° cameras for full video view. As a workaround, in the setup used for experiments, one of the blind areas could be covered by the end-effector RGB-D camera (as shown in Figure 18), which was sufficient for safe operation.

In the experiments, all commands were input with hands, voice, and eye tracking. In the multi-user context, each operator could choose the input modality individually. However, some operators indicated that a physical controller would still be better and more reliable for specific tasks. These tasks could be precise manipulator movement with haptic or force feedback or driving the base with a gamepad while looking around at point clouds. Such a solution is available in the interface, and other input devices can be integrated and used simultaneously in single and multi-user operations.

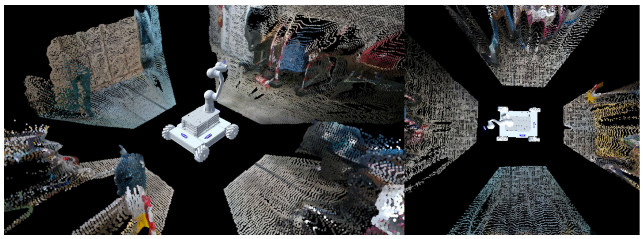

FIGURE 47. Problem of RGBD cameras placement to have 360 degrees point cloud view of the environment. Due to the offsets from the centre of the base and limited horizontal FOV <90° per camera, there are blind (black) areas at the corners of the base. In the figure, there are two views shown: on the left, from a side as usually seen by the operator, and on the right, from above the robot.

A. STUDY LIMITATIONS

The experiments were performed using the CERN network infrastructure and 4G network available at its ground or underground premises, which may differ from other networks in other hazardous environments and their network characteristics, interferences, and limitations. The requirements set for the interface validations were derived from robotic activities in the particle accelerator environment present at CERN and the experience of operators. Other dangerous environments may have different requirements depending on robot type, used hardware, control algorithms and tasks. The study presented here does not discuss operator workload evaluation, efficiency comparison between different interfaces, or ergonomics. It is focused on qualitative and quantitative assessment of the functionalities of the interface and the feasibility of its usage in the available network infrastructure. In the study, the network hardware (especially the mobile router and 4G modem) was not compared with other models. However, initial investigation showed that using more advanced mobile router models could increase the bandwidth (e.g. by improving Carrier Aggregation). Also, during the experiments, only the 4G technology (which covers the majority of underground areas at CERN) was used, and 2G, 3G, or 5G technologies were not tested. The experiment did not study delays between the interface server and the AR HMD using the Holographic Remoting to stream the holographic content or send input signals. The used RTT values were the delays between the interface server and the robot.

VI. CONCLUSION AND FUTURE WORK

Hypotheses H1 and H2 were confirmed by providing extensive tests and measurements of the interface in the CERN intervention scenarios. Despite challenging conditions, especially related to multifunctional user interface capabilities, network infrastructure limitations, and the interface's high reliability, an operator accompanied by other operators and scenario experts could perform a stable AR remote teleoperation.

Future work has to be performed in the field of multi-user collaboration network architecture. In the architecture presented in this paper, each user created a separate connection to the robot, which multiplied the throughput proportionally to the number of users. As observed, the achieved total bandwidth is slightly bigger with two connections than with a single one, but it may not be enough in low-bandwidth cases with more than two users. An architecture with a single link from the robot to a single interface multi-server, which further distributes data to each user, is necessary to increase feedback quality. Data distribution would usually be done within the CERN infrastructure network, which offers much higher bandwidth than the robot's connection. Figure 48 presents a comparison of the two architectures. In the optimized acquisition architecture, the required bandwidth will be lowered by a factor of a maximum of 2. Although there will be an additional delay due to the proxying and processing of the stream in the server to adapt to each user's needs, the network RTT is expected to be significantly lower due to the lower load for the same feedback quality.

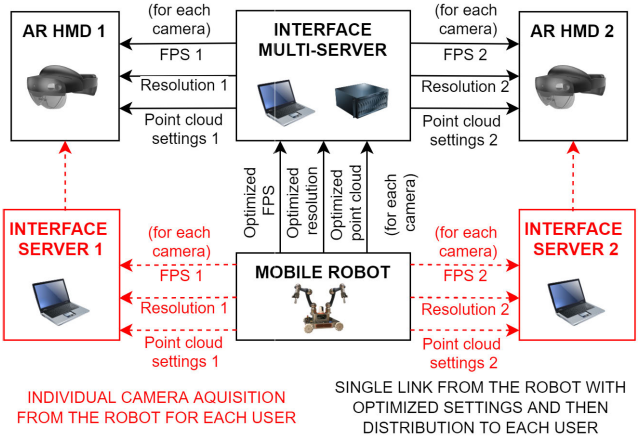

FIGURE 48. The optimized architecture for the multi-user acquisition of cameras feedback. In red, the architecture used for the validations (Figures 34 and 35) is presented, where each user could request individual camera acquisition from the robot. In black, the improved architecture optimizes the acquisition by a single connection to the robot with the camera parameters that will suit all users' most demanding setting requests.

In the multi-user scenarios tested in experiments, both users used the AR HMD. The system description explained the workflow of selecting 2D/3D, AR, or a combination (Figure 10). Accordingly, in multi-user collaboration, each user could use a different interface to achieve the goals most efficiently according to the executed task. For example, one operator could be focused on robot navigation and manipulation in AR with the best perception. Another expert could monitor measurements, mission status, or a global operation area in a screen-based 2D interface at a higher interaction level. Such evaluation will be the subject of subsequent research and experimentation comparing 2D, 3D and AR interfaces, operator workload and task execution efficiency.

The potential future work will focus on a high-level autonomous behaviour used for applying different interactions with the robot depending on network conditions (e.g. moving from an area with good bandwidth and RTT to a worse one). That would recognize and adapt to a situation and give a choice to the operator at a more supervisory level. Artificial Intelligence (AI) could also use other information

like a position in an accelerator, based on advanced simultaneous localization and mapping (SLAM) techniques or radiation levels, gathered in real-time or based on historical data. The passivity could be automatically controlled in position control (trajectories) when the RTT is above a specified limit.

The gaze information could be further utilised for future optimization of camera acquisition and improved perception. A camera at the moment looked at could switch to a high resolution or higher FPS. Accordingly, the point cloud region where the gaze is focused could be denser, peripheral areas could be less dense, and the not rendered areas could not even be sent from the robot. Other techniques of the foveated rendering could also be applied [68]. All would significantly reduce the demand for bandwidth and improve perception.

Currently, the interface supports multi-user operations with one robot at a time. The next step would integrate collaborating multi-robot scenarios, already tested at CERN [56], with multiple operators in the AR workspace.

REFERENCES

[1] N. Mohamed, J. Al-Jaroodi, and I. Jawhar, "Middleware for robotics: A survey," in *Proc. IEEE Conf. Robot., Automat. Mechtron.*, Sep. 2008, pp. 736–742.

[2] J. Adams, "Critical considerations for human–robot interface development," in *Proc. AAAI Fall Symp.*, Dec. 2002, pp. 1–8.

[3] A. Zacharaki, I. Kostavelis, A. Gasteratos, and I. Dokas, "Safety bounds in human robot interaction: A survey," *Saf. Sci.*, vol. 127, Jul. 2020, Art. no. 104667. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925753520300643

[4] S. H. Alsamhi, O. Ma, and M. S. Ansari, "Survey on artificial intelligence based techniques for emerging robotic communication," *Telecommun. Syst.*, vol. 72, pp. 483–503, Mar. 2019, doi: 10.1007/s11235-019-00561-z.

[5] A. Seefried, A. Pollok, R. Kuchar, M. Hellerer, M. Leitner, D. Milz, C. Schallert, T. Kier, G. Looye, and T. Bellmann, "Multi-domain flight simulation with the DLR robotic motion simulator," in *Proc. Spring Simul. Conf. (SpringSim)*, Apr. 2019, pp. 1–12.

[6] T. B. Sheridan and W. L. Verplank, "Human and computer control of undersea teleoperators," Cambridge Man-Mach. Syst. Lab, Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep., Jul. 1978. [Online]. Available: https://apps.dtic.mil/sti/citations/ADA057655

[7] D. A. Norman and S. W. Draper, *User Centered System Design; New Perspectives on Human-Computer Interaction*. USA: L. Erlbaum Associates, 1986, doi: 10.5555/576915.

[8] J. Y. C. Chen, E. C. Haas, and M. J. Barnes, "Human performance issues and user interface design for teleoperated robots," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 37, no. 6, pp. 1231–1245, Nov. 2007.

[9] M. E. Walker, H. Hedayati, and D. Szafir, "Robot teleoperation with augmented reality virtual surrogates," in *Proc. 14th ACM/IEEE Int. Conf. Hum.-Robot Interact. (HRI)*, Mar. 2019, pp. 202–210.

[10] K. A. Szczurek, R. M. Prades, E. Matheson, J. Rodriguez-Nogueira, and M. D. Castro, "Mixed reality human–robot interface with adaptive communications congestion control for the teleoperation of mobile redundant manipulators in hazardous environments," *IEEE Access*, vol. 10, pp. 87182–87216, 2022.

[11] G. Lunghi, R. Marin, M. Di Castro, A. Masi, and P. J. Sanz, "Multimodal human–robot interface for accessible remote robotic interventions in hazardous environments," *IEEE Access*, vol. 7, pp. 127290–127319, 2019.

[12] H. Dol, M. Colin, P. van Walree, and R. Otnes, "Field experiments with a dual-frequency-band underwater acoustic network," in *Proc. 4th Underwater Commun. Netw. Conf. (UComms)*, 2018, pp. 1–5.

[13] T. Sawa, N. Nishimura, and S. Ito, "Wireless optical EtherNet modem for underwater vehicles," in *Proc. 15th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2018, pp. 1–4.

[14] B. Kelley and K. Naishadham, "Rf multicarrier signaling and antenna systems for low SNR broadband underwater communications," in *Proc. IEEE Topical Conf. Power Amplif. Wireless Radio Appl.*, Jan. 2013, pp. 340–342.

[15] M. de la Cruz, G. Casañ, P. Sanz, and R. Marín, "Preliminary work on a virtual reality interface for the guidance of underwater robots," *Robotics*, vol. 9, no. 4, p. 81, Oct. 2020. [Online]. Available: https://www.mdpi.com/2218-6581/9/4/81

[16] A. Solis, R. Marin, J. Marina, F. J. Moreno, M. Ávila, M. de La Cruz, D. Delgado, J. V. Marti, and P. J. Sanz, "An underwater simulation server oriented to cooperative robotic interventions: The educational approach," in *Proc. OCEANS*, 2021, pp. 1–6.

[17] D. Centelles, A. Soriano, J. V. Martí, R. Marin, and P. J. Sanz, "UWSIM-Net: An open-source framework for experimentation in communications for underwater robotics," in *Proc. OCEANS*, 2019, pp. 1–8.

[18] P. Cieslak, "Stonefish: An advanced open-source simulation tool designed for marine robotics, with a ROS interface," in *Proc. OCEANS*, Jun. 2019, pp. 1–6.

[19] R. Pi, P. Cieslak, P. Ridao, and P. J. Sanz, "TWINBOT: Autonomous underwater cooperative transportation," *IEEE Access*, vol. 9, pp. 37668–37684, 2021.

[20] G. McDonald, "Operations to 11,000M: Nereus ceramic housing design and analysis," in *Proc. OCEANS*, 2013, pp. 1–5.

[21] S. K. Singh, L. Yang, and H. Ma, "Recent challenges for haptic interface and control for robotic assisted surgical training system: A review," in *Proc. IEEE World AI IoT Congr. (AIIoT)*, May 2021, pp. 240–249.

[22] W. P. Chan, G. Hanks, M. Sakr, H. Zhang, T. Zuo, H. F. M. van der Loos, and E. Croft, "Design and evaluation of an augmented reality head-mounted display interface for human robot teams collaborating in physically shared manufacturing tasks," *ACM Trans. Human-Robot Interact.*, vol. 11, no. 3, pp. 1–19, Sep. 2022.

[23] W. Hönig, C. Milanes, L. Scaria, T. Phan, M. Bolas, and N. Ayanian, "Mixed reality for robotics," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2015, pp. 5382–5387.

[24] T. Kot, P. Novák, and J. Bajak, "Using hololens to create a virtual operator station for mobile robots," in *Proc. 19th Int. Carpathian Control Conf. (ICCC)*, May 2018, pp. 422–427.

[25] M. Walker, H. Hedayati, J. Lee, and D. Szafir, "Communicating robot motion intent with augmented reality," in *Proc. ACM/IEEE Int. Conf. Hum.-Robot Interact.*, Feb. 2018, pp. 316–324. doi: 10.1145/3171221.3171253.

[26] D. Puljiz, E. Stohr, K. Riesterer, B. Hein, and T. Kroger, "General hand guidance framework using Microsoft hololens," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Apr. 2019, pp. 5185–5190.

[27] C. Liang, C. Liu, X. Liu, L. Cheng, and C. Yang, "Robot teleoperation system based on mixed reality," in *Proc. IEEE 4th Int. Conf. Adv. Robot. Mechtron. (ICARM)*, Jul. 2019, pp. 384–389.

[28] F. Ghiringhelli, J. Guzzi, G. A. D. Caro, V. Caglioti, L. M. Gambardella, and A. Giusti, "Interactive augmented reality for understanding and analyzing multi-robot systems," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2014, pp. 1195–1201.

[29] K. Chandan, V. Kudalkar, X. Li, and S. Zhang, "ARROCH: Augmented reality for robots collaborating with a human," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2021, pp. 3787–3793.

[30] H. Kato and M. Billinghurst, "Marker tracking and HMD calibration for a video-based augmented reality conferencing system," in *Proc. 2nd IEEE ACM Int. Workshop Augmented Reality (IWAR)*, Oct. 1999, pp. 85–94.

[31] R. Grasset, P. Lamb, and M. Billinghurst, "Evaluation of mixed-space collaboration," in *Proc. IEEE ACM Int. Symp. Mixed Augmented Reality (ISMAR)*, Oct. 2005, pp. 90–99.

[32] S. Prince, A. D. Cheok, F. Farbiz, T. Williamson, N. Johnson, M. Billinghurst, and H. Kato, "3D live: Real time captured content for mixed reality," in *Proc. Int. Symp. Mixed Augmented Reality*, Oct. 2002, pp. 7–317.

[33] A. Maimone, X. Yang, N. Dierk, A. State, M. Dou, and H. Fuchs, "General-purpose telepresence with head-worn optical see-through displays and projector-based lighting," in *Proc. IEEE Virtual Reality (VR)*, Mar. 2013, pp. 23–26.

[34] B. Yoon, H. il Kim, G. A. Lee, M. Billinghurst, and W. Woo, "The effect of avatar appearance on social presence in an augmented reality remote collaboration," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces (VR)*, Mar. 2019, pp. 547–556.

[35] (2023). *Microsoft Hololens Solutions*. [Online]. Available: https://www.microsoft.com/en-us/hololens/apps

[36] (2023). *Imverse*. [Online]. Available: https://www.imverse.com

[37] *Meta Oculus*. Accessed: Dec. 1, 2022. [Online]. Available: https://www.oculus.com/horizon-worlds/
[38] (Nov. 16, 2018). *Magic Leap Social*. [Online]. Available: https://www.magicleap.com/news/connect-with-friends-with-avatar-chat
[39] (2020). *Gixel*. [Online]. Available: https://www.gixel.de
[40] (2023). *High Fidelity*. [Online]. Available: https://www.highfidelity.com
[41] D. Szafir, B. Mutlu, and T. Fong, "Designing planning and control interfaces to support user collaboration with flying robots," *Int. J. Robot. Res.*, vol. 36, nos. 5–7, pp. 514–542, Jun. 2017, doi: 10.1177/0278364916688256.
[42] G. Lunghi, R. M. Prades, and M. D. Castro, *An Advanced, Adaptive and Multimodal Graphical User Interface for Human-Robot Teleoperation in Radioactive Scenarios*, vol. 2, O. Gusikhin, D. Peaucelle, K. Madani, Eds. Vienna, Austria: SciTePress, 2016, pp. 224–231.
[43] F. Chenf, B. Gao, M. Selvaggio, Z. Li, D. Caldwell, K. Kershaw, A. Masi, M. D. Castro, and R. Losito, "A framework of teleoperated and stereo vision guided mobile manipulation for industrial automation," Inst. Electr. Electron. Eng., Manhattan, NY, USA, Tech. Rep., 2016, pp. 1641–1648. [Online]. Available: https://ieeexplore.ieee.org/document/7558810
[44] M. D. Castro, J. C. Vera, A. Masi, and M. F. Perez, *Novel Pose Estimation System for Precise Robotic Manipulation in Unstructured Environment*, vol. 2, O. Gusikhin K. Madani, Eds. Vienna, Austria: SciTePress, 2017, pp. 50–55.
[45] G. Lunghi, R. M. Prades, M. D. Castro, M. Ferre, and A. Masi, *An RGB-D Based Augmented Reality 3D Reconstruction System for Robotic Environmental Inspection of Radioactive Areas*. Vienna, Austria: SciTePress, 2017, pp. 233–238.
[46] C. Saliba, M. K. Bugeja, S. G. Fabri, M. D. Castro, A. Mosca, and M. Ferre, *A Training Simulator for Teleoperated Robots Deployed at Cern*, vol. 2, K. Madani O. Gusikhin, Eds. Vienna, Austria: SciTePress, 2018, pp. 283–290.
[47] L. Grech, G. Valentino, M. D. Castro, and C. V. Almagro, "Collision avoidance system for the RP survey and visual inspection train in the CERN large hadron collider," in *Proc. IEEE 14th Int. Conf. Automat. Sci. Eng. (CASE)*, Aug. 2018, pp. 817–822.
[48] L. Angrisani, P. Arpaia, D. Gatti, A. Masi, and M. D. Castro, *Augmented Reality Monitoring of Robot-Assisted Intervention in Harsh Environments at Cern*, vol. 1065. Bristol, U.K.: Institute of Physics, 2018.
[49] M. Di Castro, C. V. Almagro, G. Lunghi, R. Marin, M. Ferre, and A. Masi, "Tracking-based depth estimation of metallic pieces for robotic guidance," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 5503–5508.
[50] M. Di Castro, M. Ferre, and A. Masi, "CERNTAURO: A modular architecture for robotic inspection and telemanipulation in harsh and semi-structured environments," *IEEE Access*, vol. 6, pp. 37506–37522, 2018.
[51] M. D. Castro, L. R. Buonocore, M. Ferre, S. Gilardoni, R. Losito, G. Lunghi, and A. Masi, "A dual arms robotic platform control for navigation, inspection and telemanipulation," in *Proc. 16th Int. Conf. Accel. Large Exp. Phys. Control Syst.*, 2018, Art. no. TUPHA127.
[52] L. Attard, C. J. Debono, G. Valentino, M. di Castro, J. A. Osborne, L. Scibile, and M. Ferre, "A comprehensive virtual reality system for tunnel surface documentation and structural health monitoring," in *Proc. IEEE Int. Conf. Imag. Syst. Techn. (IST)*, Oct. 2018, pp. 1–6.
[53] M. D. Castro, M. L. B. Tambutti, M. Ferre, R. Losito, G. Lunghi, and A. Masi, "I-tim: A robotic system for safety, measurements, inspection and maintenance in harsh environments," Inst. Elect. Electron. Eng., Piscataway, NJ, USA, Tech. Rep., 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8468661
[54] C. V. Almagro, M. D. Castro, G. Lunghi, R. M. Prades, P. J. S. Valero, M. F. Pérez, and A. Masi, "Monocular robust depth estimation vision system for robotic tasks interventions in metallic targets," *Sensors*, vol. 19, pp. 3220, Jul. 2019.
[55] M. D. Castro, G. Lunghi, A. Masi, M. Ferre, and R. M. Prades, "A multidimensional RSSI based framework for autonomous relay robots in harsh environments," Inst. Electr. Electron. Engineers, Piscataway, NJ, USA, Tech. Rep., 2019, pp. 183–188. [Online]. Available: https://ieeexplore.ieee.org/document/8675673
[56] C. V. Almagro, G. Lunghi, M. D. Castro, D. C. Beltran, R. M. Prades, A. Masi, and P. J. Sanz, "Cooperative and multimodal capabilities enhancement in the CERNTAURO human–robot interface for hazardous and underwater scenarios," *Appl. Sci.*, vol. 10, no. 17, p. 6144, Sep. 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/17/6144
[57] M. D. Castro, G. Lunghi, and A. Masi, *Use of Virtual Reality for Robotic Intervention Preparation in Unstructured and Hazardous Environments* R. K. Leach, D. Billington, C. Nisbet, D. Phillips, Eds. Cranfield, U.K.: EUSPEN, 2020, pp. 247–250.
[58] M. di Castro, J. C. Vera, M. Ferre, and A. Masi, *Object Detection and 6D Pose Estimation for Precise Robotic Manipulation in Unstructured Environments*, vol. 495, O. Gusikhin and K. Madani, Eds. New York, NY, USA: Springer-Verlag, 2020, pp. 392–403.
[59] C. Prados Sesmero, S. Villanueva Lorente, and M. Di Castro, "Graph SLAM built over point clouds matching for robot localization in tunnels," *Sensors*, vol. 21, no. 16, p. 5340, Aug. 2021.
[60] K. Szczurek, R. Prades, E. Matheson, H. Perier, L. Buonocore, and M. D. Castro, *From 2D to 3D Mixed Reality Human-Robot Interface in Hazardous Robotic Interventions With the Use of Redundant Mobile Manipulator*. Vienna, Austria: SciTePress, 2021, pp. 388–395.
[61] J. Marín Garcés, C. Veiga Almagro, G. Lunghi, M. Di Castro, L. R. Buonocore, R. Marín Prades, and A. Masi, "MiniCERNBot educational platform: Antimatter factory mock-up missions for problem-solving STEM learning," *Sensors*, vol. 21, no. 4, p. 1398, Feb. 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/4/1398
[62] D. Morra, E. Cervera, L. R. Buonocore, J. Cacace, F. Ruggiero, V. Lippiello, and M. D. Castro, "Visual control through narrow passages for an omnidirectional wheeled robot," in *Proc. 30th Medit. Conf. Control Automat. (MED)*, Jun. 2022, pp. 551–556.
[63] (2023). *Microsoft Hololens 2*. [Online]. Available: https://www.microsoft.com/en-us/hololens
[64] (May 4, 2022). *Holographic Remoting Player Overview*. [Online]. Available: https://learn.microsoft.com/en-us/windows/mixed-reality/develop/native/holographic-remoting-player
[65] K. A. Szczurek. (2023). *Video Demonstration: Multimodal Multi-User Mixed Reality Human-Robot Interface—Augmented Reality Head-Mounted Device*. [Online]. Available: https://videos.cern.ch/record/2297328
[66] *Photon Unity Networking*. Accessed: Nov. 1, 2022. [Online]. Available: https://www.photonengine.com/en-us/sdks#pun
[67] K. A. Szczurek. (2023). *Experimental Data for Multimodal Multi-User Mixed Reality Human-Robot Interface for Teleoperation in Hazardous Environments*. [Online]. Available: https://gitlab.cern.ch/kszczure/multi-user-mixed-reality-human-robot-interface-experimental-data
[68] B. Mohanto, A. T. Islam, E. Gobbetti, and O. Staadt, "An integrative view of foveated rendering," *Comput. Graph.*, vol. 102, pp. 474–501, Feb. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0097849321002211

KRZYSZTOF ADAM SZCZUREK received the M.Sc.Eng. degree in control engineering and robotics from the Wrocław University of Science and Technology, Poland, in 2017. He is currently pursuing the Ph.D. degree in computer science and robotics with the Jaume I University of Castellon, Spain. From 2013 to 2014, he worked with American Axle & Manufacturing on industrial controls for the automotive sector. In 2017, he worked with Nokia on the 5G communication technology. From 2015 to 2016, and since 2017, he has been with CERN, working on automation, control software, and robotics projects. He is passionate about MR human–robot interfaces, space robotics, and specializes in designing and implementing control systems consisting of complex and state-of-the-art solutions, based on PLC/SCADA and real-time systems (C++/C#/Unity and LabVIEW).

RAUL MARIN PRADES received the B.Sc. degree in computer science engineering and the Ph.D. degree in engineering from the Jaume I University of Castellon, Spain, in 1996 and 2002, respectively. The subject of his Ph.D. was the development of a supervisory controlled telerobotic system via web, by using object and speech recognition, 3D virtual environments, grasping determination, and augmented reality. In 1996, he worked with Nottingham University Science Park, U.K., studying multimedia and simulation techniques for human–computer interfaces. In 1997, he joined Lucent Technologies (Bell Labs Innovations Research and Development) and worked as a Researcher, a Software Developer, and a Software Architect at the Switching and Access Division. In 1999, he began to teach and research as a Professor at the Jaume I University of Castellon. Since 2009, he has been an Associate Professor at the Department of Computer Science, Jaume-I University of Castellon, Spain, where he lectures computer networking and robotics. He has been appointed as a Visiting Scientist at Blaise Pascal University, in 2002, the Polytechnic University of Madrid, in 2007, the University Federal of Brasilia, in 2016, and the European Organization for Nuclear Research (CERN), in 2015, 2018, 2019, 2020, and 2021–2022, studying new techniques for telemanipulation in hazardous environments. He has teaching experience in computer science engineering degree, the intelligent systems master, and the EU EMARO advanced robotics master, among others. He has participated in research projects, such as FP6 GUARDIANS (group of unmanned assistant robots deployed in aggregative navigation supported by scent detection), FP7 TRIDENT Project (marine robots and dexterous manipulation for enabling autonomous underwater multipurpose intervention missions), and H2020 El-Peacetolero (embedded electronic solutions for polymer innovative scanning tools using light emitting devices for diagnostic routines). His research interests include robotics, rescue, and underwater, including subjects such as localization, networks of sensors and actuators, object recognition, teler-obotics, and education. He is the author or coauthor of around 150 research publications on these subjects.

ELOISE MATHESON received the B.Sc./B.Eng. degrees in mechatronics (space) engineering from The University of Sydney, Australia, in 2010, the M.Sc. degree in advanced robotics from the École Central de Nantes, France, in 2014, and the Ph.D. degree in surgical robotics from Imperial College London, U.K., in 2021. The subject of the Ph.D., was the research, development, and clinical evaluation of the human–machine interface of a novel steerable soft catheter for neurosurgery. From 2014 to 2016, she was an Engineer at the Telerobotics and Haptics Laboratory, European Space Agency, The Netherlands, largely working on telerobotic activities under the METERON Project. In 2020, she joined the Mechatronics, Robotics and Operations Section, CERN, as a Mechatronics Engineer, working on beam intercepting device mechatronic systems and the development and integration of robotic solutions in the accelerator complex. Her research interests include tele-operation, supervisory control, autonomous operations, haptics, and human–machine interfaces.

JOSE RODRIGUEZ-NOGUEIRA received the B.Sc. degree in electronics, robotics, and mechatronics engineering from the University of Malaga, Spain, in 2020. The subject of his Degree Dissertation was about the implementation of visual odometry in a one to ten scale autonomous vehicle in ROS and GAZEBO. In September 2020, he started the Automatic and Robotic Master's with the Politechnic University of Madrid. He realized an internship with the Vision & Aerial Robotics Group, Center of Automatics and Robotics, Madrid. There, he implemented an unity UAV simulator into a ROS framework called Aerostack. He also worked on the implementation of a inertial and visual odometry algorithm to estimate the UAV trajectory. In July of 2021 he started a trainee internship at CERN developing a robot teleoperation mixed reality GUI. After that internship, in February 2022, he continued his work at CERN with a Technical Student contract improving the current mixed reality GUI elaborating a new way of teleoperation using the augmented reality headset HoloLens 2.

MARIO DI CASTRO receives the M.Sc. degree in electronic engineering from the University of Naples "Federico II," Italy, and the Ph.D. degree in robotics and industrial controls from the Polytechnic University of Madrid, Spain. From 2005 to 2006, he was an Intern and a Technical Student at CERN in charge of advanced magnetic measurements and studies for LHC superconducting magnets. From 2007 to 2011, he works at EMBL c/o DESY in charge of advanced mechatronics solutions for synchrotron beamlines controls. Since 2011, he has been working at CERN, where he leads the Mechatronics, Robotics, and Operation Section. The section is responsible for the design, installation, operation, and maintenance of advanced control systems based on different control platforms for movable devices characterized by few um positioning accuracy (e.g., scrapers, collimators, goniometers, and target) in harsh environment. Important section activities are the design, construction, installation, operation, and maintenance of robotic systems used for remote maintenance in the whole CERN accelerator complex and quality assurance. His research interests include modular robots, tele-robotics, human–robot interfaces, machine learning, enhanced reality, automatic controls, mechatronics, precise motion control in harsh environment, and advanced robotics also for search and rescue scenarios.