Mixed Reality Human-Robot Interface for Robotic Operations in Hazardous Environments

Report submitted by Krzysztof Adam Szczurek in order to be eligible for a doctoral degree awarded by Universitat Jaume I
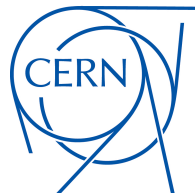
Author: *Signatures:*

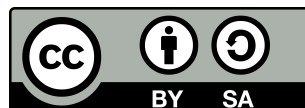MSc. Eng. Krzysztof Adam SZCZUREK

 Supervisors:

Dr. Raúl MARÍN PRADES
Universitat Jaume I Doctoral School

Dr. Mario DI CASTRO
CERN - European Organization for Nuclear Research

Castelló de la Plana, June 2023

ii

Abstract

Interventions in high-risk hazardous environments often require teleoperated remote systems or mobile robotic manipulators to prevent human exposure to danger. The need for secure and effective teleoperation is growing, demanding enhanced environmental understanding and collision prevention. Therefore, the human-robot interfaces must be designed with reliability and safety in mind to enable the operator to perform remote inspections, repairs, or maintenance. Modern interfaces provide some degree of telepresence for the operator, but they do not allow full immersion in the controlled situation. Mixed Reality (MR) technologies with Head-Mounted Devices (HMDs) can address this issue, as they allow for stereoscopic perception and interaction with virtual and real objects simultaneously. However, such human-robot interfaces were not showcased in telerobotic interventions in hazardous environments, and the work done within this thesis intended to address this challenge. The research was done at the European Organization for Nuclear Research (CERN) for mobile robots operated remotely in particle accelerators and experimental areas.

During the thesis progression, three subsequent goals were achieved. Firstly, the teleoperator was provided with immersive interactions while still ensuring the accurate positioning of the robot. These techniques had to be adapted to accommodate delays, bandwidth restrictions, and fluctuations in the 4G shared network of the realistic underground particle accelerator environment. A developed network optimization framework enabled Mixed Reality technologies, such as 3D collision detection and avoidance, trajectory planning, real-time control, and automated target approach. A novel application-layer congestion control with automatic settings was applied to the video and point cloud feedback with adaptive algorithms based on the camera frame rate, resolution, point cloud subsampling, network round-trip time, and throughput-to-bandwidth ratio.

Secondly, the MR human-robot interface was designed to function with Augmented Reality (AR) HMDs in wireless network environments. The multimodal interface provided efficient and precise interaction through hand and eye tracking, user motion tracking, voice recognition, and video, 3D point cloud, and audio feedback from the robot. Furthermore, the interface allowed multiple experts to collaborate locally and remotely in the AR workspace, enabling them to share or monitor the robot's control. The interface was tested in real intervention scenarios at CERN to evaluate its performance. Network characterization and measurements were conducted to assess if the interface met the operational requirements and if the network architecture could support single and multi-user communication loads.

Finally, the 3D MR human-robot interface was compared with a well-validated 2D interface to ensure it was safe and efficient. The 3D MR interface brought multiple valuable functionalities, which may have added to the operator's workload and stress while increasing system complexity. The CERN 3D MR and operational 2D interfaces were compared using the NASA TLX assessment method, custom questionnaires, task execution time curves, and by measurement of the heart rate (HR), respiration rate (RR), and skin electrodermal activity (EDA) evaluated by the developed Operator Monitoring System (OMS). The system was designed to measure the

physiological parameters of a teleoperator during robotic interventions.

Limitations and further research areas for improvement were identified, such as optimizing the network architecture for multi-user scenarios and applying automatic interaction strategies depending on network conditions for higher-level interface actions. The practical use of OMS revealed the necessity of applying machine learning techniques in signal interpretation to detect non-standard situations and utilizing contactless monitoring technology. The developed interface systems demonstrated operational readiness, achieving a Technical Readiness Level (TRL) 8, through successful single and multi-user missions.

Resumen

Las intervenciones en entornos peligrosos de alto riesgo a menudo requieren sistemas remotos teleoperados o manipuladores robóticos móviles para evitar la exposición humana al peligro. La necesidad de una teleoperación segura y eficaz es cada vez mayor, lo que exige una mejor comprensión ambiental y prevención de colisiones. Por lo tanto, las interfaces hombre-robot deben diseñarse teniendo en cuenta la confiabilidad y la seguridad para permitir que el operador realice inspecciones, reparaciones o mantenimiento remotos. Las interfaces modernas brindan cierto grado de telepresencia para el operador, pero no permiten una inmersión total en la situación controlada. Las tecnologías de realidad mixta (MR) con dispositivos montados en la cabeza (HMD) pueden abordar este problema, ya que permiten la percepción estereoscópica y la interacción con objetos virtuales y reales simultáneamente. Sin embargo, tales interfaces hombre-robot no se encontraron en intervenciones telerrobóticas en entornos peligrosos, y el trabajo realizado dentro de esta tesis pretendía abordar este desafío. La investigación se realizó en la Organización Europea para la Investigación Nuclear (CERN) para robots móviles operados de forma remota en aceleradores de partículas y áreas experimentales.

Durante el desarrollo de la tesis, se lograron tres objetivos principales. En primer lugar, al teleoperador se le proporcionaron interacciones inmersivas mientras se aseguraba el posicionamiento preciso del robot. Estas técnicas tuvieron que mejorarse para adaptarse a los retrasos, las restricciones de ancho de banda y las fluctuaciones en la red compartida 4G del entorno realista del acelerador de partículas subterráneo. Un marco de optimización de red desarrollado permitió tecnologías de realidad mixta, como la detección y prevención de colisiones en 3D, la planificación de trayectorias, el control en tiempo real y el enfoque automatizado de objetivos. Se aplicó un novedoso control de congestión de la capa de aplicación con configuraciones automáticas al video y la retroalimentación de la nube de puntos con algoritmos adaptativos basados en la velocidad de captura de la cámara, la resolución, el submuestreo de la nube de puntos, el tiempo de ida y vuelta de la red y la relación rendimiento-ancho de banda.

En segundo lugar, la interfaz hombre-robot MR fue diseñada para funcionar con HMD de realidad aumentada (AR) en entornos de red inalámbrica. La interfaz multimodal proporcionó una interacción eficiente y precisa a través del seguimiento de manos y ojos, seguimiento de movimiento del usuario, reconocimiento de voz y video, nube de puntos 3D y retroalimentación de audio del robot. Además, la interfaz permitió que varios expertos colaboraran local y remotamente en el espacio de trabajo de AR, lo que les permitió compartir o monitorizar el control del robot. La interfaz se probó en escenarios de intervención reales en el CERN para evaluar su desempeño. Se realizaron caracterizaciones y mediciones de la red para evaluar si la interfaz cumplía con los requisitos operativos y si la arquitectura de la red podía admitir cargas de comunicación de uno o varios usuarios.

Finalmente, la interfaz 3D MR hombre-robot se comparó con una interfaz 2D bien validada para garantizar que fuera segura y eficiente. La interfaz 3D MR trajo múltiples funcionalidades útiles, que pueden haber aumentado la carga de trabajo

y el estrés del operador al tiempo que aumentaba la complejidad del sistema. La MR 3D del CERN y las interfaces 2D operativas se compararon utilizando el método de evaluación TLX de la NASA, cuestionarios personalizados, curvas de tiempo de ejecución de tareas y mediante la medición de la frecuencia cardíaca (HR), la frecuencia respiratoria (RR) y la actividad electrodérmica de la piel (EDA) evaluado por el Sistema de Monitorización del Operador (OMS) desarrollado. El sistema fue diseñado para medir los parámetros fisiológicos de un teleoperador durante intervenciones robóticas.

Se identificaron limitaciones y otras áreas de investigación para mejorar, como la optimización de la arquitectura de red para escenarios multiusuario y la aplicación de estrategias de interacción automática según las condiciones de la red para acciones de interfaz de nivel superior. El uso práctico de OMS reveló la necesidad de aplicar técnicas de aprendizaje automático en la interpretación de señales para detectar situaciones no estándar y utilizar tecnología de monitorización sin contacto. Los sistemas de interfaz desarrollados demostraron preparación operativa, logrando un nivel de preparación técnica (TRL) 8, a través de misiones exitosas de uno o varios usuarios.

Contents

List of Abbreviations

API	Application Programming Interface
AR	Augmented Reality
CERN	Conseil Européen pour la Recherche Nucléaire
	European Organization for Nuclear Research
CRF	CERN Robotic Framework
EAC	European Astronaut Centre
ESA	European Space Agency
ESOC	European Space Operations Centre
ESS	European Spallation Source
HMD	Head-Mounted Device
HRI	Human-Robot Interaction
GUI	Graphical User Interface
IRSLab	Interactive and Robotic Systems Lab
JET	Joint European Torus
LHC	Large Hadron Collider
LiDAR	Light Detection And Ranging
LoS	Line of Sight
MR	Mixed Reality
NASA	National Aeronautics and Space Administration
OMS	Operator Monitoring System
RGBD	Red Green Blue Depth
ROV	Remotely Operated Vehicle
R&D	Research & Development
SLAM	Simultaneous Localization And Mapping
SPS	Super Proton Synchrontron
TIM	Train Inspection Monorail
TRL	Technical Readiness Level
UAV	Unmanned Aerial Vehicle
UDP	User Datagram Protocol
UX	User Experience
VR	Virtual Reality
XR	Extended Reality

Chapter 1

Introduction

Robots are often necessary for interventions in hazardous environments due to dangers such as dust, fire, pressurized water, or radioactivity. To ensure the reliability and safety of telerobotics, appropriate user interfaces should be provided to the operator for remote inspections, repairs, or maintenance tasks. The amount of effort an operator must contribute must be considered, and cognitive fatigue avoided, as these factors affect precision, efficiency, and safety. Typically, out-of-the-box robotic solutions, such as tEODor [1] or telemax [2] Remotely Operated Vehicles (ROVs) are used for telerobotics in hazardous scenarios, which limits the number of tools, sensors, and devices that can be integrated and controlled. These solutions have specific communication systems and closed-source user interfaces. In extremely harsh environments, such as in the space around the International Space Station [3], fully customized solutions must be used. Controlling robots on the Moon from Earth requires overcoming a round-trip ~3-second communication delay. Radiation and temperature could also damage electronics and reduce computation efficiency. Furthermore, exploring the Mars surface is challenging due to the 8-42 minutes communication latency and low bandwidth between the Earth and Mars stations, making some teleoperation tasks impossible [4] and require predictive and autonomous control.

FIGURE 1.1: The MASCOT master arms are operated to replace components in the Joint European Torus (JET) tokamak [5]. The interface is composed of multiple screens with camera video feedback and 3D models visualisation.

In interventions with radioactive hazards, such as the Fukushima Daiichi NPS survey of radioactive water leakage [6], specific aspects of robot's design and use

have to be taken into account. In such environments, the challenges include hardware resistance to radiation, communication link limitations, and the need for portability and decontamination easiness of the emergency response robots [7]. Remote handling in fusion research [8] is motivated by radiation near installation, repair or replacement of components (Figure 1.1). Robotic arms are mounted on fixed positions or mobile platforms and connected to a high-bandwidth wired or wireless network to overcome bandwidth limitations. Particle accelerator complexes challenge strong safety protocols, limited access, and interventions with mobile robots next to equipment positioned with micrometre precision. The communication systems are limited by radiation in normal operating conditions.

Designing human-machine interfaces for mobile robots is a complex task considering numerous technical characteristics and human factors [9]. These include the limitations of the robot's mechatronic systems (e.g. motor torques, control algorithms, material strength), the processing power available for onboard sensory data processing, and the communication link between the robot and the operator (bandwidth, delays, volatility, security). On the operator's side, human-centred factors such as cognitive limits, psychology, and ergonomics must also be evaluated [9]. Deciding on the level of automation is critical, as it will influence the required teleoperation functionalities and Artificial Intelligence (AI) [10]. The type of sensory feedback from the robot determines the choice of interface hardware. For example, a 2D screen may be sufficient for a video stream, while 3D feedback may require the use of stereo glasses, Virtual Reality (VR) or Augmented Reality (AR) head-mounted displays (HMDs). A hand-held controller may be used to enable force feedback from the robot.

For field operations such as underwater and underground robot control [11], search and rescue, and space robot operation [12], easily deployable interfaces are needed. 2D-based Graphical User Interfaces (GUIs) are reliable but limited to flat screens. The user experience (UX) can be improved by providing more accessible information with Virtual, Augmented, and Mixed Reality techniques and 3D perception. However, they require a larger data transmission from the robot to the operator, which can cause issues in high-latency, low-bandwidth scenarios [13][14][15]. Communication system optimization is therefore crucial for improved user interaction. High-bandwidth communication is required for the quality of environmental perception and precise robot localization in tunnels.

Human-Robot Interaction (HRI) in field robotics requires the combination of different communication media, depending on the mission and the environment, to decrease the operator's workload and increase the robot's autonomy. Using a high-level command structure requires the operator to be able to specify missions and trajectories, as well as activate semi-autonomous behaviours. Simulated and actual data can be combined to give the user a better understanding of the remote robot's performance, follow the mission steps, and receive telemetry information.

1.1 State of the art

The state of the art of research presented in the thesis can be divided into five domains:

1. Mixed Reality displays taxonomy;

2. Adaptive communications congestion control;
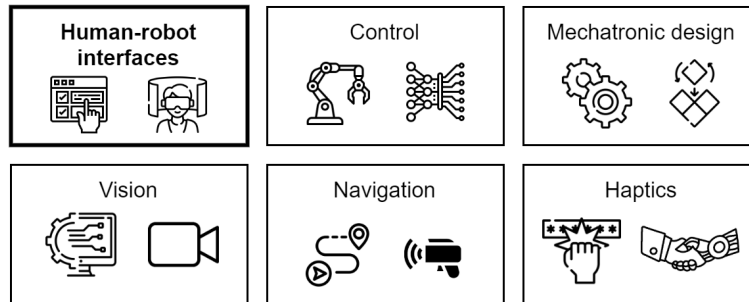
3. Mixed Reality telerobotics;

FIGURE 1.2: The diagram shows the current main research lines in the CERN robotics team's R&D. In **bold**, the work in this thesis is highlighted.

4. Multi-user collaboration in Mixed Reality;

5. Vital physiological parameters measurement and stress assessment;

Table 1.1 summarises the domains and their most representative references. The three journal papers, published within this thesis framework, presented as Chapters 2, 3, and 4 extend, interconnect and present integrally a comprehensive state of the art. Chapter 2 focuses on domains 1, 2, and 3, Chapter 3 on domains 2, 3, and 4, and Chapter 4 on domains 3 and 5.

The work presented in this thesis has been completed at the European Organization for Nuclear Research (CERN) within the Mechatronics, Robotics and Operation (MRO) section that is part of the Controls, Electronics and Mechatronics (CEM) group which is under the Beams Department (BE). The previous research done at CERN significantly contributed to the success of the developments. The most relevant references are presented in Table 1.2.

At CERN, multiple technologies have been investigated and applied in operational robotic scenarios. Figure 1.2 presents the main research lines. *Human-robot interfaces* covers teleoperation and supervisory control aspects, particularly the 2D, 3D and stereoscopic visualisation techniques, operator's tasks execution assistance and automation, or other solutions provided in the interface application layer. *Control* covers the robot's dynamics, inverse kinematics, trajectory generation, or machine learning. *Mechatronic design* focuses on robots' modular mechanical, electrical and electronic design. *Vision* concentrates on using control and feedback acquisition from visible or infrared spectrum imagery, or LiDARs. *Navigation* implements the SLAM and path planning algorithms. *Haptics* research utilises force-feedback input devices and environmental sensors.

1.2 Motivation

The novelty and contribution pursued in this thesis were motivated by the state of the art's lack of reliable MR human-robot interface solutions applicable to CERN telerobotics. Figure 1.3 illustrates the key challenges in mobile robotics in CERN's hazardous environments, which must be considered for successful interventions. The presented work concentrates on mitigating these difficulties by investigating and applying novel MR techniques in the interfaces, addressing the network communication constraints by adaptive communication congestion protocols, and monitoring the operator's vital parameters for the user's inclusion as a critical part of the telerobotic system.

TABLE 1.1: This table presents the most representative references of the five distinctive domains of the general state-of-the-art: MR displays, adaptive communication congestion control, MR telerobotics, multi-user MR collaboration, and the operator's vital physiological parameters measurement and stress assessment.
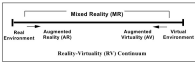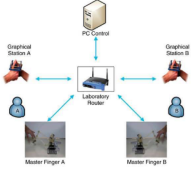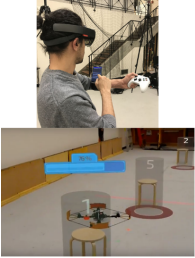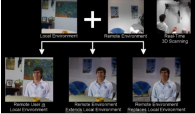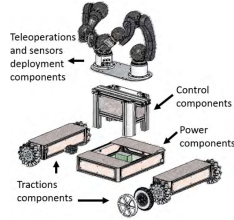
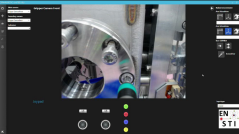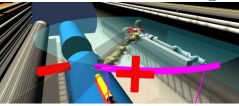| Domains	The most relevant aspects of the references
Mixed Reality displays [16][17][18]	The authors proposed a taxonomic framework classifying MR visual displays using three-dimensional space of Extent of World Knowledge, Extent of Presence Metaphor and Reproduction Fidelity. The Extent of World Knowledge relates to the amount of real world modelled in MR. The Extent of the Presence Metaphor relates to the user's immersion or presence in the Mixed Reality. Lastly, the Reproduction Fidelity relates to how realistic the graphics are rendered.
Adaptive communications congestion control [19]	The authors explored transport protocol solutions for Internet Telerobotics, which were necessary to manage an IP protocol due to the connection over the Internet. It was found that TCP protocol was more suitable for high-level commands sent with low frequency, as reliability was the primary concern, while UDP was better for low-level control, where transmission delay was more critical. As a result, the paper demonstrated the need to develop a specific UDP-based teleoperation protocol for low latency and optimised congestion/flow end-to-end control.
Mixed Reality telerobotics [20]	The research concentrated on AR control of UAVs, emphasising that teleoperation conceals the mapping between the operator's input and the robot's dynamics, which can only be acquired through experience and is difficult for inexperienced operators. The UDP communication protocol was used to communicate with the robot. The development was done with the Unity game engine and the HoloLens AR HMD. The robot was controlled with waypoint and trajectory control. The system incorporated safeguarded teleoperation. The operator shared the same environment with the robot and operated in LoS.
Multi-user collaboration in Mixed Reality [21]	A Virtual Monitors system combined the real world with virtual images of other participants in an Augmented Reality (AR) conference, with one user wearing an optical head-mounted display (HMD) and the others using webcams. The interaction was carried out using a Virtual Shared Whiteboard, and communication was done by voice and gestures. This system permitted virtual representations of other conference participants to be seen. However, the system was not symmetrical, as only one user can use an augmented reality head-mounted display (AR HMD) while the others view 2D webcam video feedback. Additionally, the wireless video link experienced high latency, and the AR display had difficulty functioning in an overly lit room.
Vital physiological parameters measurement and stress assessment [22]	This system aimed to adjust a robotic arm's control strategy based on the operator's emotional state, as measured through cardiac and electrodermal activity, to enable a cooperative human-machine interaction.

TABLE 1.2: This table shows four references that were the foundations of the R&D done within this thesis. They cover the modular architecture of the robotic framework developed at CERN for harsh semi-structured environments; multimodal human-robot interfaces; cooperative robotics; and a pilot project studying the advantages of 3D Mixed Reality for robotic interventions.

| References	The most relevant aspects of the references
CERNTAURO: A modular architecture for robotic inspection and telemanipulation in harsh and Semi-Structured environments [23]	The CERNTAURO framework has provided a modular architecture encompassing all aspects of the CERN facility's remote robotic maintenance and interventions. This includes elements such as the specifications and training of operators, the selection of robots, the selection of materials based on radiological contamination risks, the successful completion of the mission, and the development of recovery strategies and procedures. This framework includes innovative features such as bilateral master-slave control, user-friendly multimodal human-robot interfaces, and offline operator training.
Multimodal Human-Robot Interface for Accessible Remote Robotic Interventions in Hazardous Environments [24]	This system was created to assist robotic platforms in dangerous CERN settings. It addressed practical concerns such as adapting to changing network latencies, continuous robot reconfiguration, and many different kinds of input devices or sensors mounted on the robots. Furthermore, it enabled multi-robot and multi-tasking scripting.
Cooperative and Multimodal Capabilities Enhancement in the CERNTAURO Human-Robot Interface for Hazardous and Underwater Scenarios [25]	The system was based on a multimodal user interface enabling the user to activate assisted cooperative behaviours according to a mission plan. The cooperative teleoperation could be coupled with additional assisted tools such as vision-based tracking and grasping determination of metallic objects, and communication protocols design.
From 2D to 3D Mixed Reality Human-Robot Interface in Hazardous Robotic Interventions with the Use of Redundant Mobile Manipulator [26]	This study was the CERN MR human-robot interface prototype, which employed Augmented Virtuality to provide real-time video feedback from the robot and complete scene modelling. This technology was studied to facilitate robotic measurements with a redundant manipulator. Additionally, the study introduced monitoring of the operator's vital parameters (heartbeat and galvanic skin response).
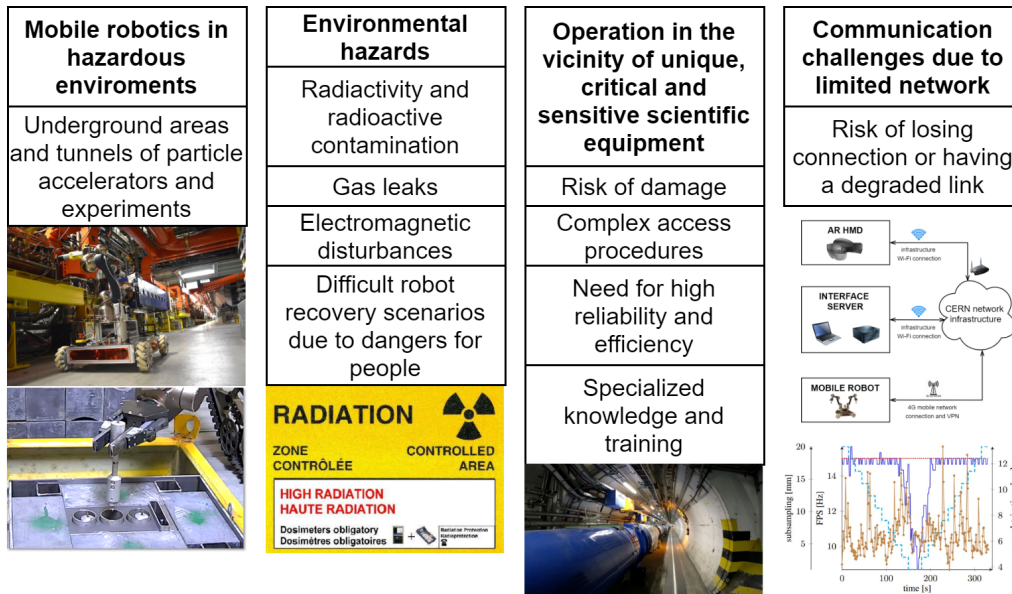
FIGURE 1.3: The diagram highlights the main aspects in the mobile robotics context in the CERN hazardous environments that must be addressed to perform successful interventions.

The objectives of the work are defined as follows:

- Evaluate the risks and challenges of telerobotics in hazardous environments;

- Address the risks of collisions and damage to the unique scientific equipment by providing spacial collision detection and avoidance mechanisms based on the robot's geometry, joint torques, and environmental feedback;

- Increase environmental awareness by real-time 3D point cloud and video feedback;

- Mitigate volatility of network bandwidth limitation and increased delays by automatically recognising network quality and adjusting video and point cloud feedback parameters;

- Increase the interaction level by transiting from manual teleoperation to the supervisory control of a complex redundant manipulator;

- Evaluate available MR HMDs and their applicability to human-robot interfaces;

- Address the problem of a too complex and not intuitive input system by applying novel control with hand, locomotion, eye tracking, and voice inputs;

- Create a multi-user system in the MR workspace for the collaboration of multiple experts during an intervention and robot's remote control;

- Perform a methodical assessment of user experience, workload, and vital physiological parameters and its comparison between the standard 2D interface and the developed 3D MR interface;

- Apply the research in operation by providing an expert system with a high TRL and reliability for robotic interventions at CERN.

The work presented in this thesis has multiple interconnections with other research lines (Figure 1.2 in Section 1.1). For example, the point clouds captured by multiple RGBD cameras and displayed in the human-robot interfaces are also used for localisation and navigation in the SPS or LHC tunnels. The AR environment visualisation and interface can surround the force feedback input manipulator. The interface plans and previews the inverse kinematics solutions developed for arm's control.

1.3 Publications

During the work, the following manuscripts were published[1]:

- Szczurek., K. A., Prades., R., Matheson., E., Perier., H., Buonocore., L., di Castro., M. (2021). "From 2D to 3D Mixed Reality Human-Robot Interface in Hazardous Robotic Interventions with the Use of Redundant Mobile Manipulator". Proceedings of the 18th International Conference on Informatics in Control, Automation and Robotics - ICINCO, 388–395.

- Szczurek, K. A., Prades, R. M., Matheson, E., Rodriguez-Nogueira, J., di Castro, M. (2022). "Mixed Reality Human–Robot Interface With Adaptive Communications Congestion Control for the Teleoperation of Mobile Redundant Manipulators in Hazardous Environments". IEEE Access, 10, 87182–87216.

- Szczurek, K. A., Prades, R. M., Matheson, E., Rodriguez-Nogueira, J., di Castro, M. (2023). "Multimodal Multi-User Mixed Reality Human–Robot Interface for Remote Operations in Hazardous Environments". IEEE Access, 11, 17305–17333.

- Szczurek, K. A., Cittadini, R., Prades, R. M., Matheson, J., di Castro, M. (2023). "Enhanced Human-Robot Interface with Operator Physiological Parameters Monitoring and 3D Mixed Reality". IEEE Access, 11, 39555-39576.

Moreover, the results of the thesis have been disseminated in the following public events:

- The CERN Academic Training, lecture delivered on 26 Jan 2022 on "Robotics activities at CERN - Enhanced reality, user interfaces and artificial intelligence" for CERN community and open for external participants, the recording and presentation are available online: https://indico.cern.ch/event/1055752/.

- CERN IT Lightning Talks, lecture delivered on 13 May 2022 on "Telerobotics at CERN: Mixed-Reality Human-Robot Interface" for CERN community and open for external participants, the recording and presentation are available online: https://cds.cern.ch/record/2811231?ln=en.

1.4 Professional and personal background

The choice of doctoral thesis research was made based on CERN robotics necessities and the professional and personal interests of the author. The author's educational

---

[1]This thesis has been accepted by the co-authors of the publications listed above, that have waived the right to present them as a part of another PhD thesis.

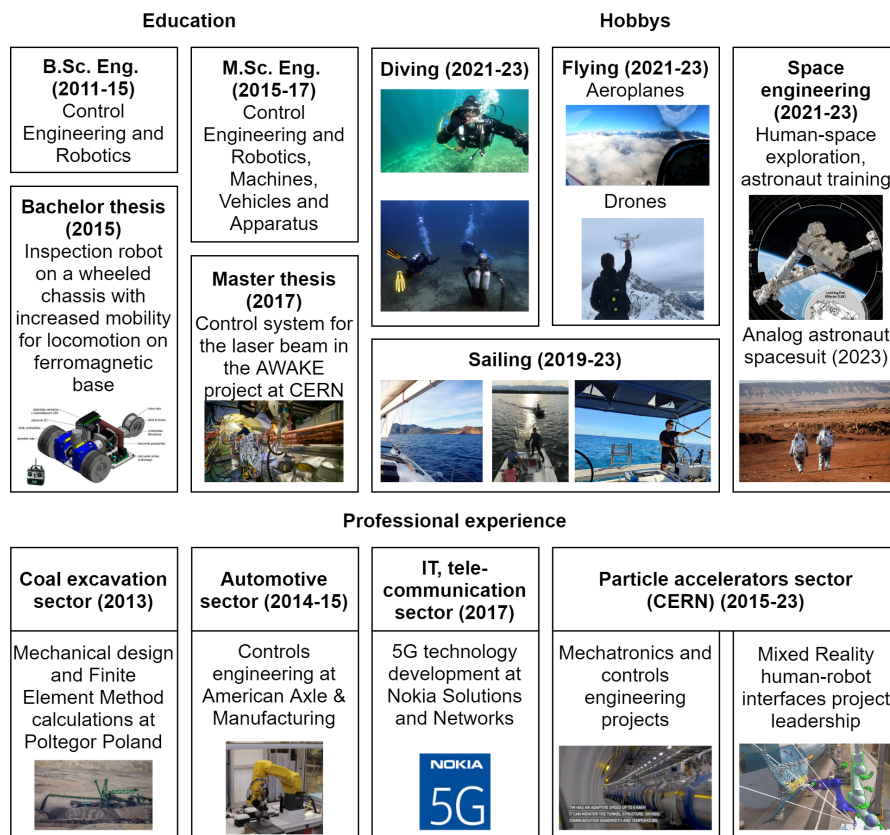**Education**   **Hobbys**

FIGURE 1.4: Author's background that created a synergy for the R&D
of the MR human-robot interfaces.

background was in control engineering and robotics, specialising in machines, vehicles and apparatus automation. The previous professional experiences were in 5G technology software engineering; automotive production lines automation and robotic systems engineering; coal excavation machines design; and mechatronics and controls engineering at CERN. The author's bachelor thesis covered all aspects of a mobile robot climbing ferromagnetic surfaces: mechanical and electrical design and its practical production, electronics, microcontroller programming, and integration with a 2.4 GHz aviation remote control. The master thesis was a project at CERN covering all control layers (hardware, low-level programming, real-time systems, middleware and interfaces) for more than 20 mechatronic devices controlling and monitoring a laser beam in the Advanced Proton Driven Plasma Wakefield Acceleration Experiment.

The multidisciplinary understanding gathered during studies, work, and bachelor and master theses were crucial during the execution of the doctoral thesis. It helped with the execution and management of the challenging project on the Mixed Reality human-robot interfaces. It required telecommunication expertise to master the network behaviour, its limited resource optimisation, and industrial reliability assurance for cutting-edge technology. A good understanding of mechanical components, motors, gearboxes, and electrical systems was necessary to develop appropriate control methods.

The personal interests in diving, sailing, aeroplane piloting, remote drone operations, human space exploration, analogue astronaut spacesuit design, and MR astronaut training contributed to finding methodologies, new approaches and solutions in the research presented in this thesis. For example, the interaction protocol between the pilot flying and the pilot monitoring was implemented for the multiuser MR interface command exchange between multiple operators. Diving in harsh underwater environments contributed to forward-thinking of backup technical solutions and potential reaction to stress. Practical experience with autonomous drone operations helped to understand the options of automatic recovery scenarios in case of lost communication. The human space exploration fascination, experience during an astronaut selection and further collaboration at ESA for applying Extended Reality (XR) to astronaut training contributed to the broader MR outlook on technical solutions for MR interfaces for robots at CERN. Participating in a spacesuit design equipped with AR HUD gave a better perspective on the most optimal information display in stressful situations. Together, diving, flying and sailing taught what the physiological reactions of the machine operator to workload and stress could be, and that human is more important than the technical perfection of a human-machine system. Thus, the vital physiological parameters monitoring system became an important part of the research.

1.5 Outline of the thesis

The material presented in this thesis is organized, with colour coding[2], as follows:

Chapter 2 presents, through the publication "*Mixed Reality Human-Robot Interface with Adaptive Communications Congestion Control for the Teleoperation of Mobile Redundant Manipulators in Hazardous Environments*", the application of MR interface in robotic intervention scenarios in the LHC at CERN, which offered such techniques as trajectories planning, real-time control, 3D collision detection and avoidance, and

---

[2]The text colour coding links each publication with its results in Chapter 5, and the diagram in Figure 5.1.

automatized target approach based on point cloud representation of the environment. Since these MR functionalities require higher throughput due to point cloud streaming, and the available network in the underground particle accelerator and the experimental areas has limited bandwidth and high delays, the Adaptive Communications Congestion Control framework was proposed to optimize the sensor feedback depending on network characteristics and task. The framework offered twelve automatic setting modes for video and point cloud feedback with algorithms based on: network round-trip time and throughput to bandwidth rate, camera frame rate, resolution, and point cloud subsampling. As a result, the higher network use was optimized, and the drawback was mitigated, simultaneously allowing to take advantage of additional MR functionalities increasing the teleoperator's spatial awareness, efficiency and safety.

Chapter 3 proposes, through the publication "*Multimodal Multi-user Mixed Reality Human-Robot Interface for Remote Operations in Hazardous Environments*", a solution for multi-user MR interface for collaborative remote robot control using stereoscopic HMDs with a holographic display. It addressed the problem of screen-based interfaces not fully allowing the robotic teleoperator to immerse in the controlled scenario. This solution extended the previous interface's functionalities with multimodal operation using hands, eyes and locomotion tracking, gestures and voice commands. It proposed an architecture for multiple users to cooperate and share the robot's control in local or remote MR workspaces. The experimental work determined whether operational requirements were met and whether the network architecture could support single and multi-user communication loads. The Adaptive Communications Congestion Control described in the previous publication was also taken advantage of and applied to multi-camera and multi-user setups.

Chapter 4 considers, through the publication "*Comparative study of 3D Mixed Reality and 2D Human-Robot Interfaces with Operator Vital Parameters Monitoring for Redundant Mobile Manipulator in Hazardous Environment*", the robotic teleoperator as a part of the human-robot interface system, contributing to the mission's success. The Operator Monitoring System (OMS) was proposed to assess the teleoperator's physiological stress. For the evaluation, the heart rate (HR), respirate rate (RR), and skin electrodermal activity (EDA) were measured and analysed. The OMS was used to compare 3D MR and 2D interfaces and evaluate the physiological response of operators during teleoperation. The comparative methodology also used the NASA TLX assessment method, custom questionnaires, and execution time curves. The experimental results found that the workload did not increase with the 3D MR interface use, while the interface provided multiple functionalities increasing safety and more supervisory control. The OMS trials during real interventions indicated the need for applying machine learning techniques in signal interpretation to detect non-standard physiological situations, and for contactless monitoring technology. Moreover, critical improvements for both interfaces were identified.

Chapter 5 discusses the obtained results, and Chapter 6 summarises the thesis achievements and proposes further future work.

Chapter 2

Mixed Reality Human-Robot Interface with Adaptive Communications Congestion Control

APPLIED RESEARCH

Mixed Reality Human–Robot Interface With Adaptive Communications Congestion Control for the Teleoperation of Mobile Redundant Manipulators 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 Robotic interventions with redundant mobile manipulators pose a challenge for telerobotics in hazardous environments, such as underwater, underground, nuclear facilities, particle accelerators, aerial or space. Communication issues can lead to critical consequences, such as imprecise manipulation resulting in collisions, breakdowns and mission failures. The research presented in this paper was driven by the needs of a real robotic intervention scenario in the Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN). The goal of the work was to develop a framework for network optimisation in order to help facilitate Mixed Reality techniques such as 3D collision detection and avoidance, trajectories planning, real-time control, and automatized target approach. The teleoperator was provided with immersive interactions while preserving precise positioning of the robot. These techniques had to be adapted to delays, bandwidth limitation and their volatility in the 4G shared network of the real underground particle accelerator environment. The novel application-layer congestion control with automatic settings was applied for video and point cloud feedback. Twelve automatic setting modes were proposed with algorithms based on the camera frame rate, resolution, point cloud subsampling, network round-trip time and throughput to bandwidth ratio. Each mode was thoroughly characterized to present its specific use-case scenarios and the improvements it brings to the adaptive camera feedback control in teleoperation. Finally, the framework was presented according to which designers can optimize their Human-Robot Interfaces and sensor feedback depending on the network characteristics and task.

INDEX TERMS Collision avoidance, collision detection, communication, human-robot interaction, mixed reality, mobile robots, network bandwidth, point cloud, redundant manipulator, telerobotics.

I. INTRODUCTION

A. HUMAN-ROBOT INTERACTION AND TELEROBOTICS IN HAZARDOUS ENVIRONMENTS

Robotic interventions in hazardous environments are often required due to the presence of dust, fire, pressurized water or radioactivity. The robotic platforms must be reliable, and

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

appropriate user interfaces should be present when an operator is performing remote inspection, maintenance or repair tasks. A consideration of the effort made by the operator and the necessity to avoid their cognitive fatigue must be taken, as this affects the safety and precision. Interventions in this kind of scenario are often performed using out-of-the-box robotic solutions, which cannot be adapted and improved according to the current requirements. An example are Remotely Operated Vehicles (ROVs), which have their
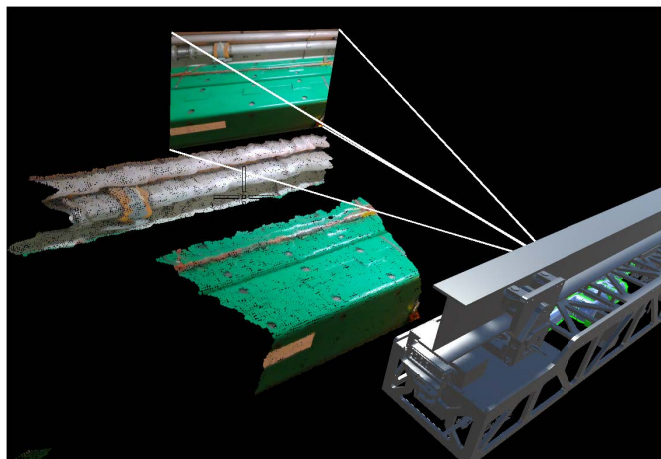
FIGURE 1. 3D perception of the environment, the robot model and video stream, combined in one interface. A fragment of a particle accelerator is captured by a RGBD camera installed on a robotic platform.

FIGURE 2. Virtual reality head-mounted device tested in the CERN teleoperation framework.

own communication system and a very specific user interface, which limits the number of tools, sensors, and devices that can be integrated and controlled.

Most systems rely on 2D-based Graphical User Interfaces, which have been demonstrated as very reliable when used by expert robotic operators, but should be improved when the expertise of the operator controlling the robot must be focused on the complexity of the environment instead [1]. This normally occurs in facilities where the real value of the operator is their knowledge of the machines and procedures that the robot will intervene on, and not their familiarity of the software interface. Interfaces could be enhanced by providing more information in a simple manner, for example, using Virtual, Augmented, and Mixed Reality techniques, as well as a 3D perception of the environment (see Figure 1, where additionally to a video stream, a point cloud of the environment and its relation to the robot can be seen). However, this enhancement requires the transmission of a larger quantity of information from the remote robot to the operator. The communication system behaviour can be affected enormously, especially when using wireless techniques in scenarios with a limited access to low-latency and high-bandwidth network, eg. underground, underwater or search and rescue missions in remote areas. The optimization and improvement of communication system is necessary in order to improve the user interaction experience.

Numerous telerobotic solutions have been used in hazardous intervention scenarios. In nuclear facilities, a critical mission which required a robotic teleoperation in a radioactive environment was the Fukushima Daiichi NPS survey of a radioactive water leakage [2]. During the mission, apart from dealing with radiation and using sufficiently resistant hardware components, the challenges also included the communication link characteristics. The need for portability, transportability and decontamination easiness of emergency response robots have been emphasized in [3] during nuclear facility accidents. The paper listed multiple robots that have been taken into account for such scenarios, however some

of them were not suitable for the conditions at Fukushima. Regarding the user interface, they were controlled with a standard 2D GUI with video feedback.

An even more challenging endeavor is that of Mars surface exploration, where the Mars-Earth communication latency and low bandwidth are key obstacles for teleoperation, and for some tasks only supervised or autonomous control are possible. As written in [4], the round-trip communication delay ranged from 8 to 42 minutes for the Mars Sample Return mission. Space exploration also poses radiation and temperature risks which decrease electronics and computation efficiency. Meanwhile, the teleoperation of robots on the Moon from the Earth has to overcome the delay of 3 s. A new control approach for kinesthetic coupling under this extreme condition was studied in [5].

Remote handling in fusion research [6] is motivated by the radiation present near components that have to be installed, repaired or replaced. The robotic arms are mounted in a fixed position or on well defined movable platforms. With such structured systems, the network limitations are easier to overcome by connecting it with a well planned, high-bandwidth wired or wireless designated network.

Particle accelerator complexes present a challenge where strong safety protocols must be kept, access is limited and the installed equipment is aligned with micrometer precision. Any telerobotic operation must not only be able to execute tasks normally done by a person, but also do them safely for the accelerator itself, not to cause any damage to fragile devices. It means that the robotic systems require high-bandwidth wireless communication for the quality of the environmental perception, as well as precise localization of the robots in the tunnels. The techniques can be applied together in order to improve the whole performance of mission, and be combined with on-board artificial intelligence techniques for the situations where the robots might get constrained communication links [7], [8].

In Table A.1 in Appendix A, there is a non-exhaustive state-of-the-art of Human-Robot Interface products in terms of

hazardous environments, 2D and 3D interfaces, operational state, communication type, complexity of manipulator, level of interaction, collision avoidance or detection, adaptive communication for shared and dynamic networks.

B. MIXED REALITY INTERFACES FOR TELEROBOTICS

Mixed Reality can provide added value to telerobotic interfaces and control, for instance by providing more intuitive collision avoidance strategies. The main advantage of collision avoidance is improved safety, however this functionality must be properly managed and integrated to avoid overcomplicating teleoperations. As described in [10], there are multiple methods of presenting real objects in Virtual Reality. They can be depicted with the Chaperone method with boundaries, as a point cloud, or as Virtual Reality objects if they are recognized. The first method displayed only the boundaries of the objects, therefore only partial surface information was available. The second method provided more information but gave less immersion to the operator, because it could significantly block the virtual content display. However, this inconvenience could be mitigated by setting a maximum distance of the point cloud. The third method depended on the precision, availability of model and recognition of real objects and placing them correctly as virtual objects.

The underwater environment is another hazardous environment that presents multiple risks for humans: lack of oxygen if the life support fails or decompression sickness in the case of a rapid ascent. Therefore, underwater tasks such as infrastructure maintenance, recovery of benthic stations or disarming post-war explosives are often addressed with remotely operated submersibles. However, complex tasks require precise manipulation, localization and advanced teleoperation techniques. Some advancements in this field were presented in [9], where underwater robots cooperation and supervision with Virtual Reality system were presented.

The use of Mixed Reality can also help with the execution of complex tasks which require redundant robotic arms. The teleoperation of such arms becomes more complicated without any additional visual 3D or model representation. Moreover, the inverse kinematics management, pose awareness and control or understanding of the arm's movement may be possible only for highly experienced operators. Therefore, as presented in [11], the use of Mixed Reality could greatly improve the efficiency of teleoperation and provide more intuitive and immersive control methods. The inverse kinematics problem may not be solvable with standard techniques, therefore alternative methods could be used specifically for redundant arms, such as the Forward And Backward Reaching Inverse Kinematics (FABRIK) heuristic method introduced in [12]. However, in this method the subsequent movements of the teleoperator arm have the impact on the final pose. This is why the continuous visual and spatial feedback presented to the operator is important.

Mixed Reality Interfaces are still in their early development and research phase, although there have been attempts to standardize them and their design process. As an example,

an extensive study of Virtual Reality Human-Robot Interface for underwater robot operation was done in [9], where types of controllers, headsets, simulation engines or frameworks and computer components were tested and characterized. A link between the interface usability and ISO usability standards was established, which lays the foundation for more methodical approach of the Human-Robot Interface design which must meet specific requirements of performance and user satisfaction.

The authors of [13], [14], and [15] proposed a taxonomic framework classifying Mixed Reality Visual Displays. In Figure 3, the currently used CERN Robotic 2D Graphical User Interface (Figure 4) and the Mixed Reality Human-Robot Interface have been classified according to this (black points). Additionally, two next generation interfaces (Mixed Reality using Virtual Reality Head-Mounted Device - Figure 2, and Mixed Reality using holographic Head-Mounted Device) currently being developed are classified (white points). The Virtual Reality and holographic Head Mounted Devices improve the Presence Metaphor. Finally, an ideal Mixed Reality interface using holographic Head-Mounted Devices, with master-slave control and force feedback, as well overlaying the fully modelled environment with a fully registered point cloud, would achieve an interface of unmediated reality, as described in Section 5.3 of [14].

C. COMMUNICATION PROTOCOLS AND POINT CLOUD STREAMING

Mixed Reality Interfaces are based on a 3D representation of the environment, therefore a reliable and fast communication becomes a very important success factor. In this section the topic of communication protocols and point cloud streaming is explained.

The authors of [16] described end-to-end congestion control transport protocol solutions designed for Internet Telerobotics, where time-varying transmission delay and non-guaranteed bandwidth were problems. Since the devices needed to connect over the Internet, they had to manage an IP protocol. It was found that the TCP protocol was better for high-level commands sent with low frequency, where reliability was more important, while the UDP protocol was better for low-level control, where transmission delay played a bigger factor. Therefore, the paper shows the necessity to build a specific UDP-based teleoperation protocol for low latency and adjusted congestion/flow end-to-end control.

Maritime search and rescue communication solution for Unmanned Aerial Vehicles (UAVs) described in [17] was based on Long Term Evolution (LTE) technology for high volume traffic and LoRa for low-rate telemetry. Due to the very high latency and reliability requirements for teleoperated UAVs, the proposed solution was application-aware for mission-critical traffic and users. A prioritizing scheduling strategy was a trade-off between network latency, reliability, and capacity.

An adaptive point cloud streaming compression algorithm was proposed in [18], which adapted to the current
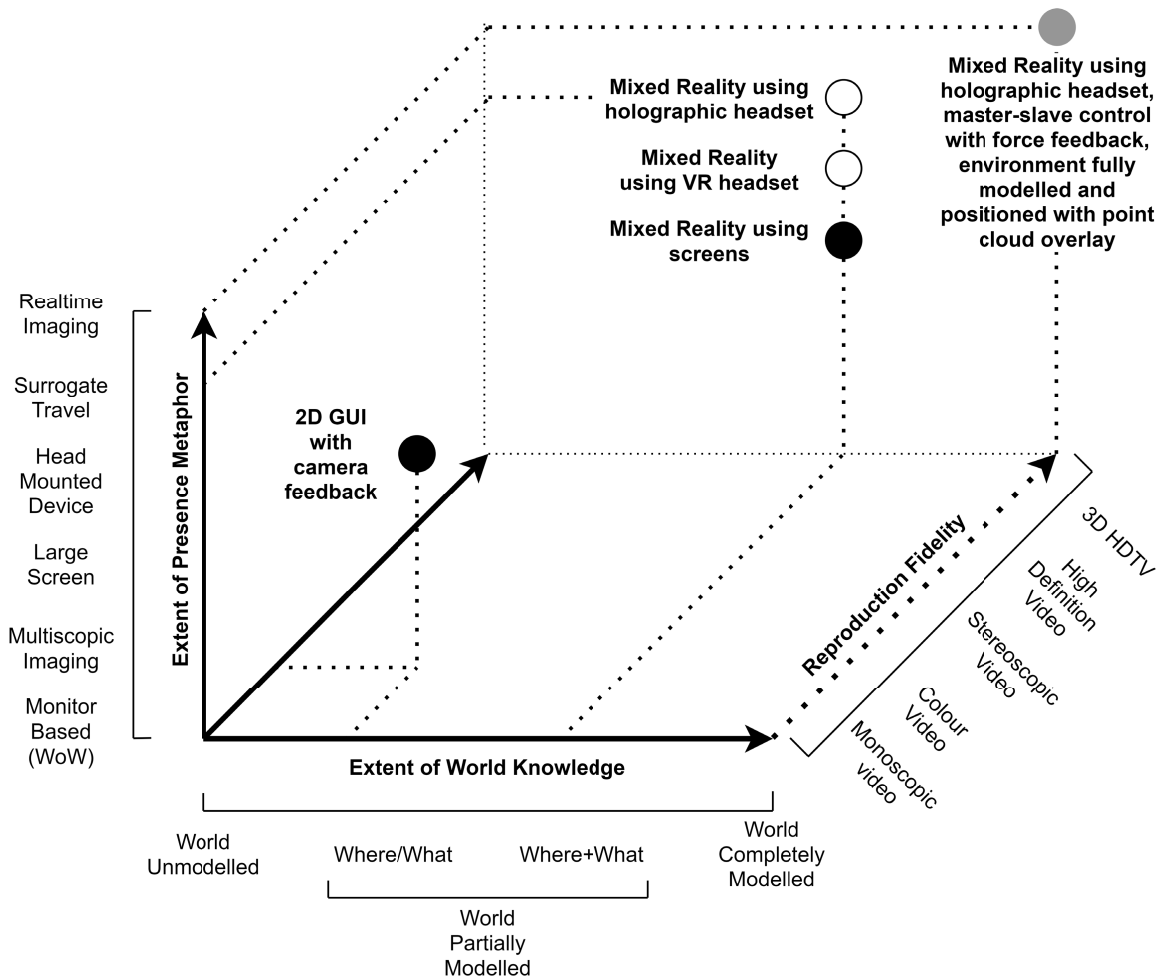
FIGURE 3. Placement of currently used, or being developed, or an ideal CERN mixed reality human-robot interface in the three dimensional taxonomic framework for classifying Mixed Reality displays.

FIGURE 4. CERN Robotic 2D Graphical User Interface used in a real intervention scenario.

based approaches. Point cloud streaming ideas were also described. The first one was the DASH-PC [23], which sent different parts of a point cloud depending on the user's current view and the available network data rate. The second one was adapting voxel length according to the available network data rate [24], [25].

A method using layered structures of point cloud transmission depending on communication channel conditions was described in [26]. The source encoded point cloud was converted into a layered structure where deeper layers hold finer point cloud representations. If the channel conditions were bad, only the upper layers were sent, which gave less detailed representation, while if the conditions were good, all the point cloud without any loss (represented by all layers) was sent.

available network data rate. The algorithm operated on an on-going point cloud stream instead of compressing a full point cloud, which avoided additional delay. In the paper, current state-of-the-art point cloud compression solutions were mentioned: Google Draco, MPEG codecs [19], machine learning approaches [20], [21] and the traditional octree [22] or kd-tree

D. CERN HUMAN-ROBOT INTERFACES EVOLUTION

Robotic interventions require a great level of modularity to achieve solutions in different scientific, hazardous and semi-structured scenarios. This modularity should be provided in both the mechatronics design (i.e. various mechanical configurations, as can be seen in Figure 5) and also
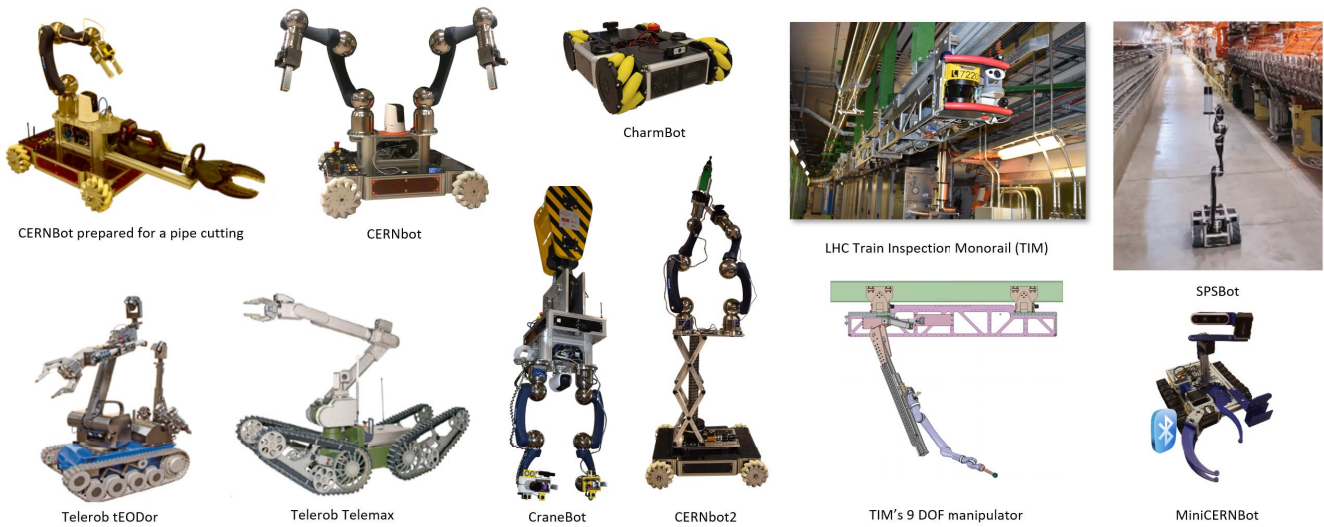
FIGURE 5. Robots used at CERN, consisting of industrial commercial robots (Telerob) and custom-made solutions (CERNBots [27], CharmBot, CraneBot, TIM [28], SPSBot, MiniCERNBot [29]).

the software architecture. The CERN CERNTAURO robotic framework [30] and [31] provides multiple autonomous and supervised teleoperation techniques. It has been tested during the last 8 years in more than 150 real interventions, 500 performed tasks and 500 hours of operation, that required multimodality [32] and flexibility. The framework contains functionalities such as master-slave interaction with time-delay monitoring, multimodal user interface, passivity control when the delay exceeds limits, vision and SLAM-based operation. The system was extended with cooperative multi-robot teleoperation [33].

The cooperative behaviour was tested with the CERN-TAURO project at CERN and with the TRIDENT/TWINBOT projects [34], [35] in an underwater scenario, and proved that one operator can safely handle an intervention with several robots simultaneously. The paper introduced a scripting feature, which allows the operator to supervise vision and artificial intelligence semi-automated behaviours to complement the manual teleoperation. It was especially helpful when multiple robots had to perform a complex task together, such as transporting or assembling a big object. As described in [36], the system was enriched with a depth estimation algorithm to be used with monocular cameras, such as a laparoscopic camera, which have been used in the design of grippers and tools for specific interventions.

Various phenomena affecting the operator were studied in [32]. For example, during the intervention, an operator could get distracted and lose focus on the relevant part of the screen. To track the operator focus, an eye-tracking system was employed. An eye-tracking camera checked the position of the pupils and sent the information to a software interlocking movement if the operator was looking elsewhere. Moreover, in the paper it was verified that the real-time feedback from the robot's environment was crucial in remote teleoperation. It concluded that crossing
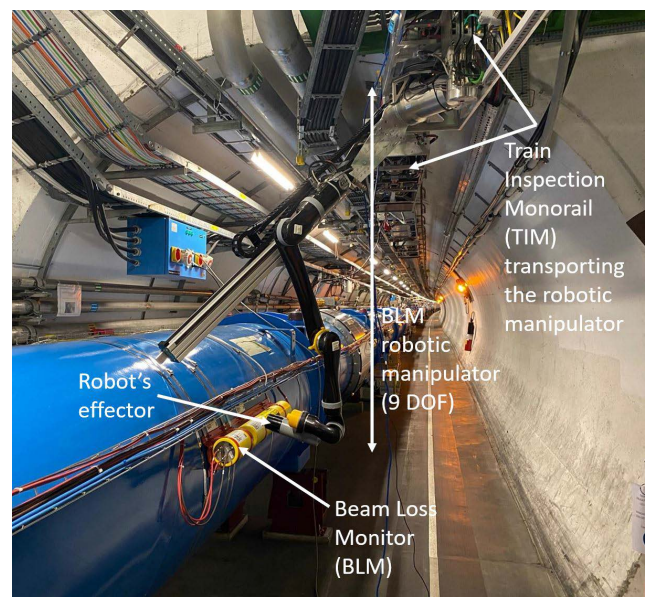

FIGURE 6. Beam loss monitor robotic arm installed on the train inspection monorail in the LHC.

a 300 ms delay threshold should decouple the commands from the robot feedback. As presented in [37], a preliminary study of Mixed-Reality Human-Robot Interface was done and a pilot project was tested in a real intervention scenario inside the LHC accelerator testing the particle beam at CERN. In the scenario, the Beam Loss Monitors robotic measurements were performed [38], where a redundant manipulator (Figure 6) using different trajectories required full perception of the robot's pose. The redundant manipulator was installed on a mobile robotic platform - the Train Inspection Monorail - mounted on a ceiling rail in the LHC (Figure 5).

E. MOTIVATION AND PROBLEM FORMULATION

Virtual, Augmented and Mixed Reality techniques in Human-Robot Interfaces have been used, however the problems of high positioning precision, collision avoidance and network delay/bandwidth constraints for mobile robots with redundant manipulators in hazardous intervention scenarios have not been fully addressed using these techniques.

Currently, there are no accepted standards or frameworks according to which designers can optimize their Human Robot Interfaces depending on the network characteristics. This paper seeks to present a framework that can be adapted to the needs of the interface. The application is given to redundant manipulators in hazardous intervention scenarios where there is a need for high positioning precision and collision avoidance while actively managing network delays/bandwidth constraints.

Much of the state-of-the-art research in Mixed Reality Human-Robot Interface controlling robots is done in controlled laboratory scenarios. However, in reality, a bigger margin of uncontrolled parameters can be expected. In this paper, emphasis is put firstly on the communication network variability, limitations and accessibility and secondly on the consequences of collisions, prioritizing safe operation with maximum possible collision detection and prevention. Both self-collisions and collisions with the environment must be checked. Thirdly, the interface, although providing much more immersion and functionalities in the teleoperation scenario, must still provide very precise positioning of the robotic arm in the industrial and scientific scenario. The positioning precision should be in the range of 1 mm for contact interaction with small elements, such as screws, or in the range of 10 mm for non-contact tasks that avoid collisions and operate, for example, a camera.

The most important aspect of the teleoperation in a hazardous environment is safety. In the Beam Loss Monitors (BLM) measurements project (Figure 6), the manipulation is performed in a very specific environment containing crucial equipment used for the LHC operation, and any collision could cause a major failure of the Collider. The robotic manipulator also carriers a radioactive source which is used to trigger and validate the Beam Loss Monitor sensors. This part is delicate, and contact with the environment should be avoided. The requirements for such missions are as follows:

- Multiple waypoints with complex manipulator configurations that can be recorded and modified;
- Collisions or dangerous close passes check during motions between waypoints. Warning about manipulator's self-collision or a collision with the environment;
- Information on the distance from the end effector to any captured point of the environment displayed during planning or real-time movements;
- The verification of all these conditions from any operator viewpoint in the scene;
- Feedback: internal status of the robot, its 3D pose visualization, and real-time environment view.

Although the LHC is partially modelled, it is not practically possible to position the robot in the modelled environment due to the localization error of the robot and the fact that not all the areas or details have been modelled. This is why the most robust 3D feedback is the real-time point cloud. The necessity to transmit 3D information increases required bandwidth, which is challenging in the 4G shared network with variable delays and bandwidth parameters. The available bandwidth depends on the position in the accelerator and current network use by different devices.

In terms of point cloud transmission, the reliability of the transmission is important (e.g. TCP protocol is more suitable than UDP protocol, as discussed in the state-of-the-art). If a part of the transmission is corrupted, the fragment or even the whole point cloud area will not be received. Some compression mechanisms can be applied and binary format of point cloud can be used to lower the transmitted size, however the biggest impact on throughput is the size of the point cloud (3D area acquired by the camera), the precision (distance between points) and the update frequency (point cloud frame rate). Therefore, application level of the end-to-end congestion control is needed. This was the main motivation of the study presented in this paper regarding point cloud acquisition and automatic setting algorithms that adapt to network conditions. Since the robot shares a public network and the point cloud communication parameters are related to the point cloud acquisition and transmission, where the network cannot control it (the parameters are beyond TCP stream), the interface needs an application-level, end-to-end and multimodal (from manual to supervisory) congestion control of point cloud streaming.

F. PAPER STRUCTURE

The paper is structured as follows: Section II describes the developed Mixed Reality Human-Robot Interface system, Section III describes the experimental setup. The results are presented in Section III-B and discussed in Section IV. Finally, Section V concludes the work.

II. SYSTEM DESCRIPTION

The system uses multimodal teleoperation techniques, from low-level manual control to high-level commands. It includes velocity control of the manipulator in real-time; trajectory specification and motion planning; target-centered task specification (e.g. aligning the end-effector to a normal line from a point cloud perceived object); and collision avoidance or detection. These strategies are applied to the Mixed Reality Human-Robot Interface for a mobile redundant manipulator operated under challenging communication conditions in hazardous intervention scenarios, characterized in Sections I-E and III-A1. As the video and point cloud feedback are crucial, an adaptive congestion control of this feedback was implemented in the interface to provide reliable automatic behaviour during operations.
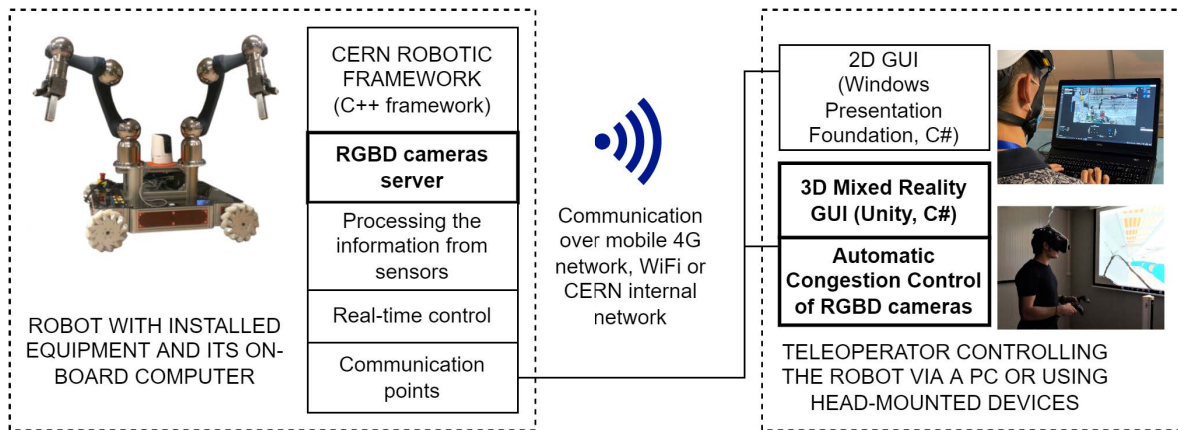
FIGURE 7. Robot-GUI architecture presenting processes that run on the robot and how the connection is established with the user interfaces. The parts which are focused on in this paper are highlighted.

A. OVERALL SYSTEM ARCHITECTURE

The robot is located in a remote area which is usually hazardous for a human. It has an onboard computer that communicates with all physical devices (e.g. motor drivers, cameras, sensors), and runs the CERN robotic framework that processes all the sensory information, controls and internal/external communication. The operator sits in a control centre and connects to the robot over the available network from a Graphical User Interface (GUI). Currently there are two types of interfaces developed for CERN purposes, the 2D GUI and 3D, Mixed Reality GUI. The 2D GUI has been developed using the Windows Presentation Foundation Technology and programmed in C# programming language. The 3D GUI has been developed in Unity, also in C#. Using the same language allowed the core communication and abstraction layers between the two GUIs to be shared. The two GUIs can and usually run in parallel as they provide different functionalities. In this paper, the CERN Mixed Reality Human-Robot Interface is presented and all the experimental data has been produced with its use. A diagram presenting the robot-GUI architecture is shown in Figure 7.

Depending on the operational setup, the robot and GUIs can communicate via a 4G network, WiFi or the cabled network infrastructure. However, the 4G network is mostly used for the reasons described in Section III-A1. Standard robotic intervention use cases and sequences of the functionalities are presented in Appendix B.

B. MIXED-REALITY HUMAN-ROBOT INTERFACE ARCHITECTURE

The global software architecture is shown in Figure 8. It specifies modules responsible for distinct functionalities and how they interact with each other. The main modules of the software are:

1) Arm control (architecture view in Figure 12 and description in Section II-D);
2) Trajectory planning (description in Section II-E);

3) Closest end effector position (description in Section II-H);
4) Target approach with a normal point (description in Section II-G);
5) Collisions (collision integration in the arm control is shown in Figure 12 and description is in Section II-F);
6) Menu (description in Section II-C4);
7) Camera control and automatic settings algorithms (description in Section II-I);
8) Camera hologram (description in Section II-I3);
9) Player movement;
10) Head-Up Display (architecture view in Figure 10 and description in Section II-C1);
11) Network measurements (description in Section II-J).

C. INTERACTION MODES

The user has the possibility to interact with the following interfaces:

- Head-Up Display;
- Holograms in the scene;
- Scene objects interaction;
- Menu interaction.

The executions of commands continues only if a corresponding confirmation key is kept pressed. This ensures a degree of safety and allows fast cancellation of a given command. For example, during a joint control mode, when a key is pressed, the program periodically sends a command to move the joint with a selected speed. As soon as the key is released a stop command is sent. Similarly, an execution of trajectory following is done as long as the key is pressed, and is stopped if key is released.

1) HEAD-UP DISPLAY

The Head-Up Display (HUD) allows the operator to have critical information constantly available during teleoperation. The HUD is shown in Figure 9 and its components in Figure 10. It displays the current connection status to the robot, current control mode of the manipulator and speed
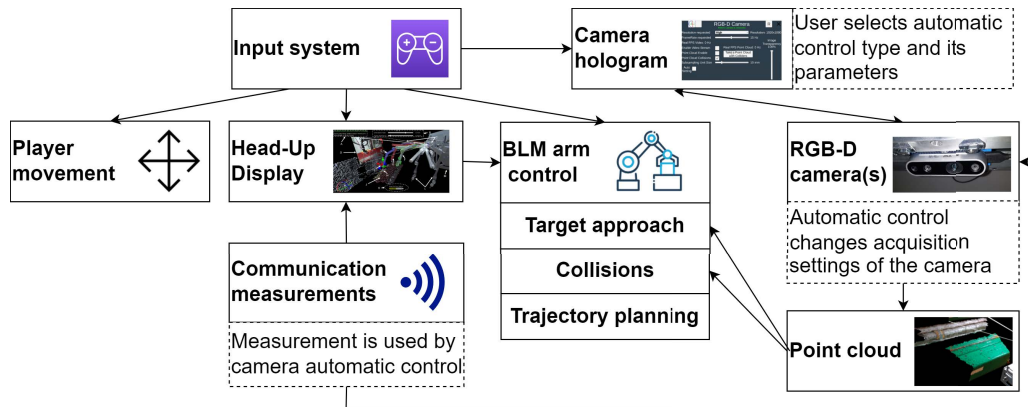
FIGURE 8. CERN mixed reality human-robot interface architecture overview. The components taking part in the camera automatic control has additional description of their role in the end-to-end application level congestion control.

settings (control modes are described in Section II-D), current usage of available bandwidth or round-trip time to estimate the control delay to be expected (network measurements are described in Section II-J), or the Mixed Reality simulation FPS (i.e. frames per second).

2) HOLOGRAMS IN THE SCENE

Some controlled elements of the robot have detailed and numerous settings that the operator must set during the intervention. An example is a camera which has ~20 parameters to set or display simultaneously. Showing them on the Head-Up Display would take too much space and the information could overload the operator. Using a menu screen would hide the scene and camera video or point cloud feedback. Therefore, a solution of holograms placed in the scene was implemented. The hologram can be seen in Figures 23 and 24. The operator can interact with the hologram from any position in the scene.

3) SCENE OBJECTS INTERACTION

Various objects in the scene can be interactive. The arm target position (Point I in Figure 9) can be moved and the arm follows it. A hologram can be enabled by clicking on the 3D model of the camera. This type of interaction is more intuitive and plays a major role while using a Virtual Reality or Augmented Reality headsets with hand controls.

4) MENU INTERACTION

Certain functionalities require complex structures of information or parametrization. As an example, for the current interface, the menu presents a network measurements screen (Figure 11), which can request bandwidth measurements or observe network behaviour and values during operations.

D. MANIPULATOR CONTROL

The robot arm can be controlled via 2 methods. The motion can be planned and then executed, or the arm can be controlled directly in real-time. Moreover, the control in each

of these methods is also multimodal. The arm can use inverse kinematics to approach a desired point by the end-effector or each joint can be controlled separately. The inverse kinematic used in planning is based on the FABRIK algorithm [12] which allows certain joints to be blocked, and during real-time control, the Jocobian-based inverse kinematic is used and computed directly by the on-board computer. For moving between trajectory positions, the position PID with velocity feedforward control is used.

The arm can be controlled in 4 control modes:

1) Real-time direct joint control;
2) Real-time inverse kinematic control;
3) Planning with separate joint control. The speed of angular and linear movement can be adjusted;
4) Planning with FABRIK inverse kinematic [12] control, where the last 6 joints follow the target. One or more of the joints can be blocked to change the behaviour of the inverse kinematics and approach.

E. TRAJECTORY SPECIFICATION

As presented in Figures 9 and 13, the operator can create a trajectory to reach the target. The 3D representation of the environment given by the point cloud feedback facilitates the manipulator's path planning by the operator or any other path planner. Path planner specifications are generalized for this framework. The waypoints can be saved, modified or removed on demand. Afterwards, a waypoint can be selected, a movement preview can be launched to check any potential collisions (more in Section II-F1), and finally the movement to the selected waypoint can be initiated.

F. COLLISIONS

The collision avoidance or detection plays a very important role in the CERN Mixed Reality Human-Robot Interface. The safety during the intervention in the LHC and with the use of the radioactive source is the most important aspect, and any collision or unplanned contact between the source and the environment should be avoided. While self-collisions
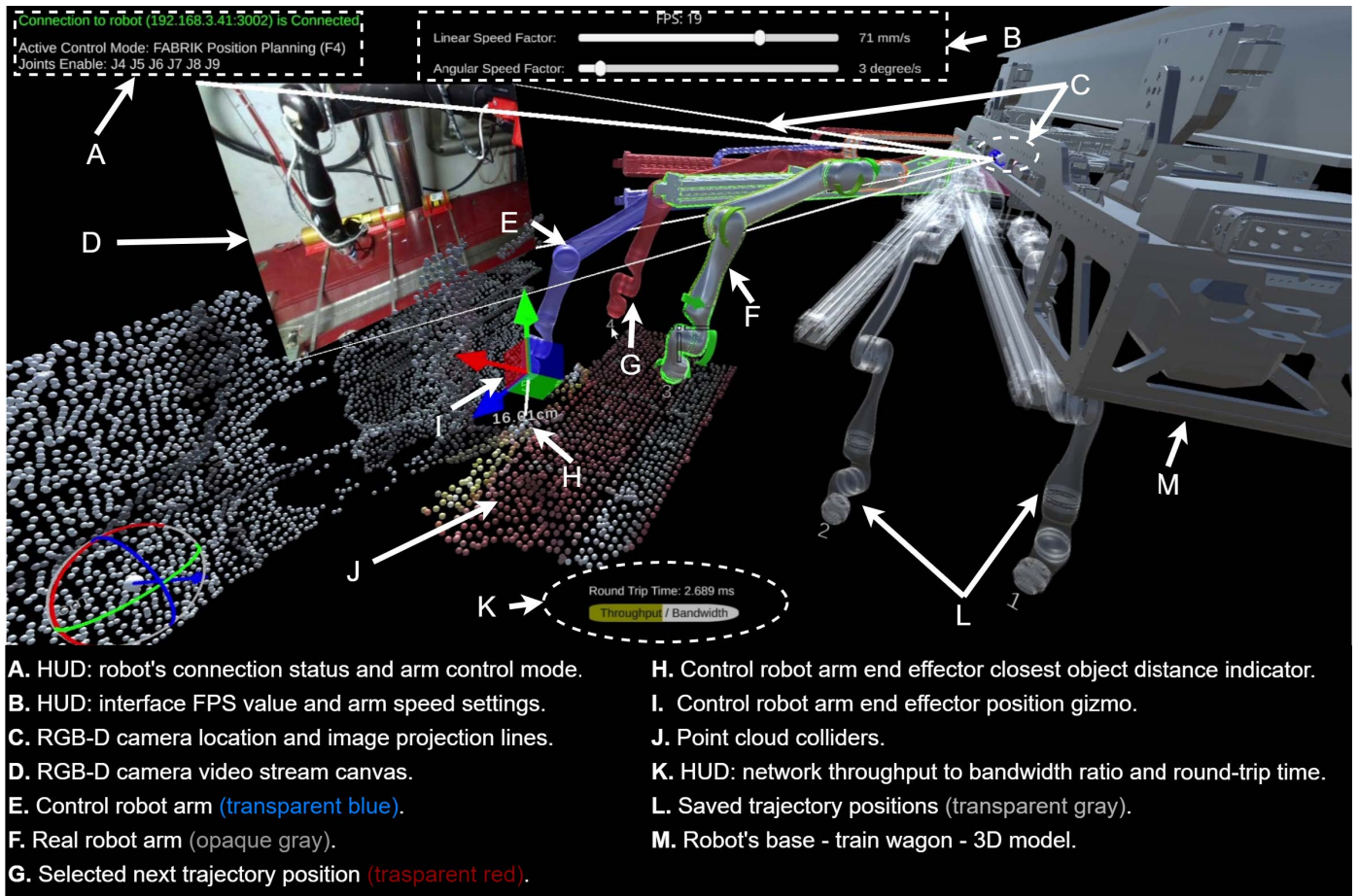
A. HUD: robot's connection status and arm control mode.

B. HUD: interface FPS value and arm speed settings.

C. RGB-D camera location and image projection lines.

D. RGB-D camera video stream canvas.

E. Control robot arm (transparent blue).

F. Real robot arm (opaque gray).

G. Selected next trajectory position (trasparent red).

H. Control robot arm end effector closest object distance indicator.

I. Control robot arm end effector position gizmo.

J. Point cloud colliders.

K. HUD: network throughput to bandwidth ratio and round-trip time.

L. Saved trajectory positions (transparent gray).

M. Robot's base - train wagon - 3D model.

FIGURE 9. CERN mixed reality human-robot interface - operator's view from an LHC intervention scenario.
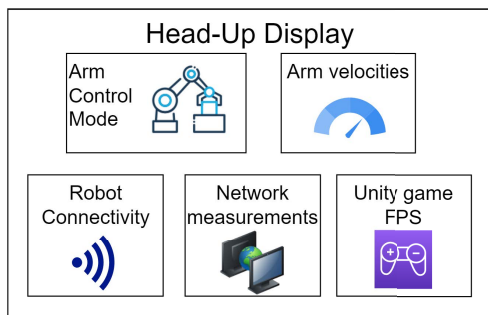

FIGURE 10. Head-Up display components overlaid on the scene content. It is visible at all times during the teleoperation.
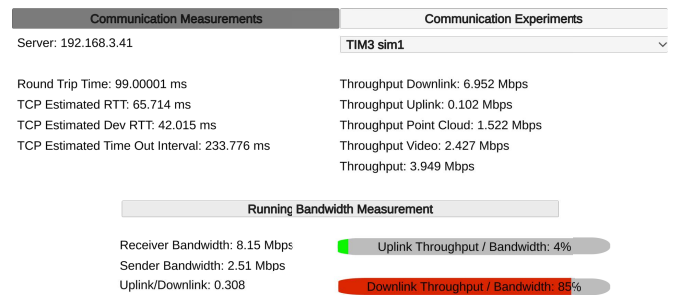

FIGURE 11. Communication measurements screen. It shows all network parameters related to the general communication with the robot and the camera-specific acquisition data transfer information.

can be recognized using the robot model, the collisions with the environment require environmental spatial information. The functionalities described in this section require 3D point cloud feedback, whose parameters are controlled by the adaptive congestion control. Therefore, this automatic control facilitates the reliable collision avoidance or detection. The interface implements the following functionalities:

- Collision avoidance during planning and preview of movements (Section II-F1);

- Real-time collision detection with joint torques (Section II-F2);
- Real-time collision detection of real arm with itself or point cloud (Section II-F3).

They assist the operator by preventing the movement if any recognized collision may happen or limiting the consequence of a collision if it occurs.
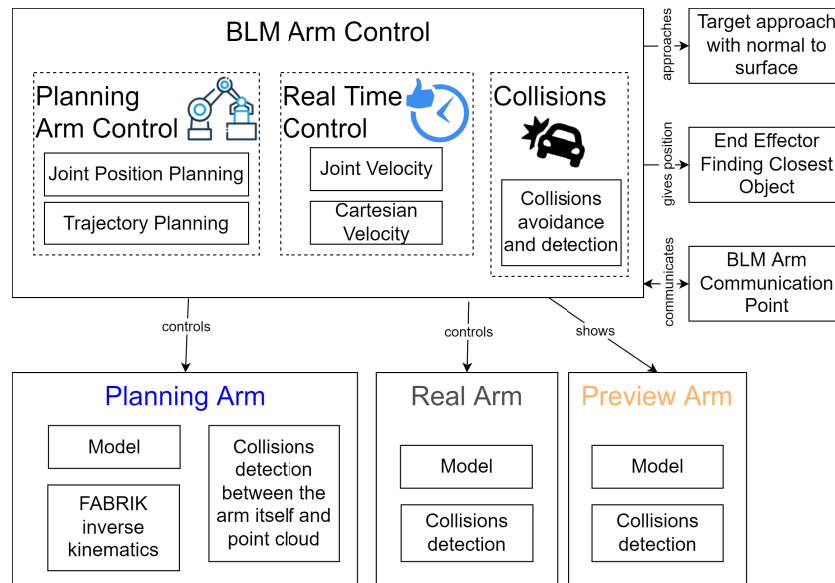
FIGURE 12. BLM arm control architecture. It shows how the main functionalities, responsible for manipulator movement planning, real-time control, preview, collision avoidance and detection, or target approach, are structured.
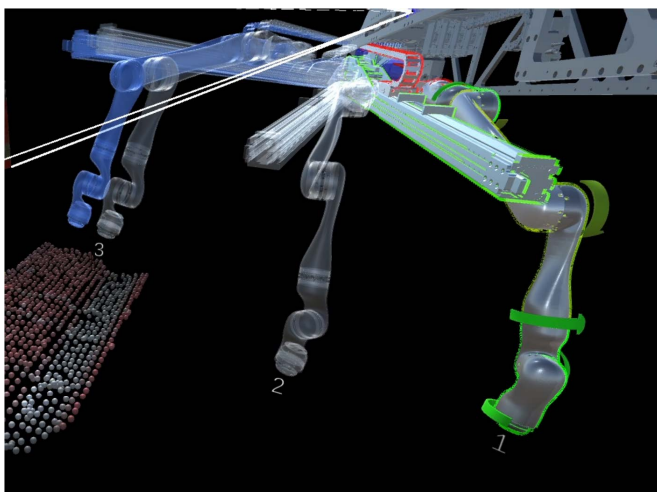

FIGURE 13. Trajectory planning. There are 3 types of arms visible: opaque real arm, transparent waypoints with numbers and blue planning arm.
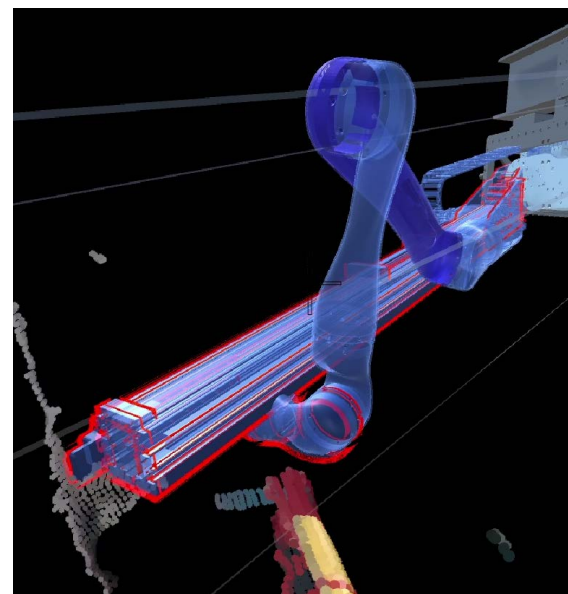

FIGURE 14. Planning arm self-collision detection.

1) COLLISION AVOIDANCE

If a collision is detected by the system during planning control mode, the operator should modify the waypoint or current arm position to avoid the collision. An example of the self-collision of the planning arm is shown in Figure 14, where the end-effector collided with another part of the arm. An example of the preview arm collision is shown in Figure 15, where the arm will self-collide if the movement is executed. The last example of a collision of the planning arm with point cloud is shown in Figure 16, where the end-effector collided with the Beam-Loss Monitor device's point cloud.

2) SENSORY COLLISION DETECTION

The sensory collision detection is represented in the interface by directional arrows and colors indicating the value of a torque or a force in each joint. The movement of the arm is stopped when the torque value exceeds a specified limit. An example of a situation when the arm collided and two joints started to have higher torques is shown in Figure 17.

3) VIRTUAL COLLISION DETECTION

Apart from the sensory collision detection using real joint torque feedback, a virtual real-time collision detection has been implemented. It detects when the real arm model collides with itself or with point cloud. It can happen during an execution of a movement to a waypoint or during real-time direct arm control. When a collision is detected, the
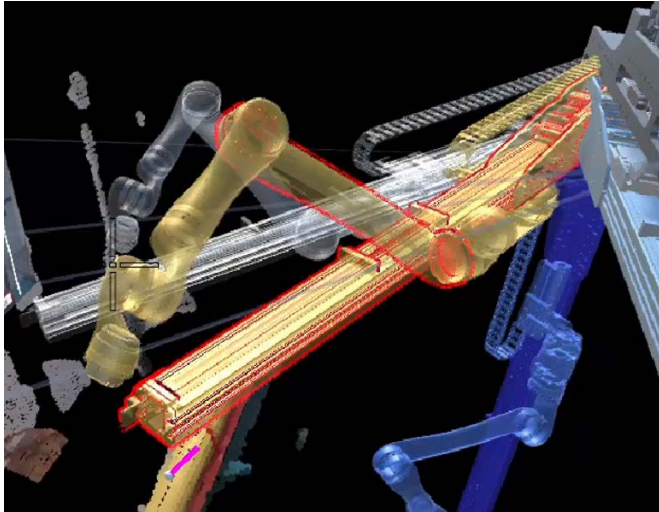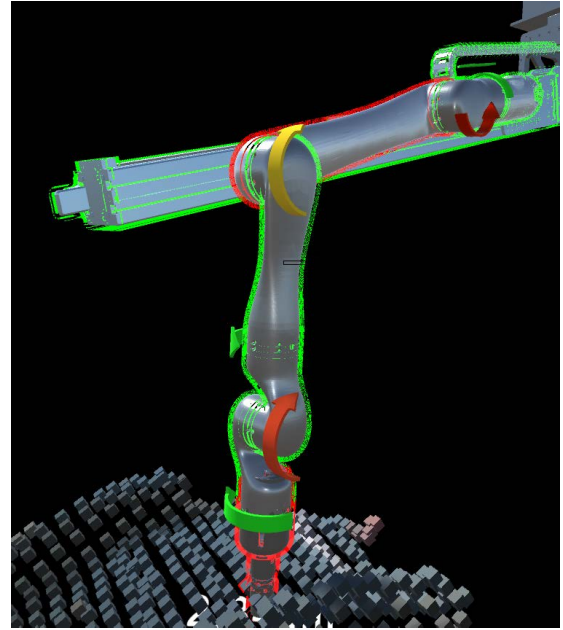
FIGURE 15. Preview arm self-collision detection.

FIGURE 16. Planning arm point cloud collision detection.

FIGURE 17. Real arm torques during a collision, with point cloud collision visualization.
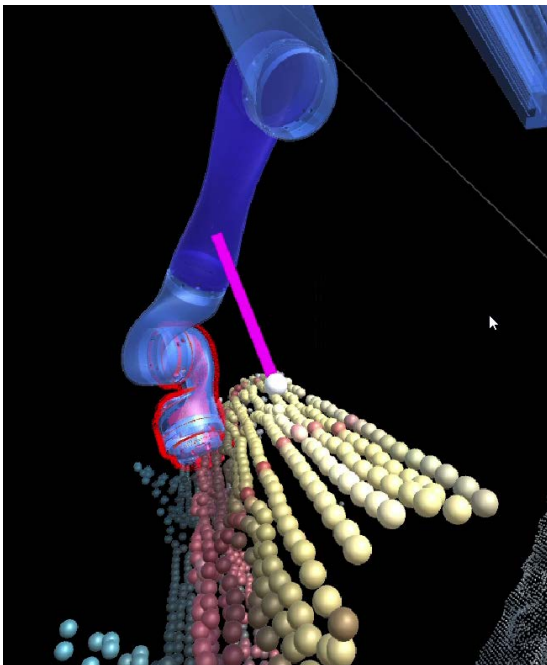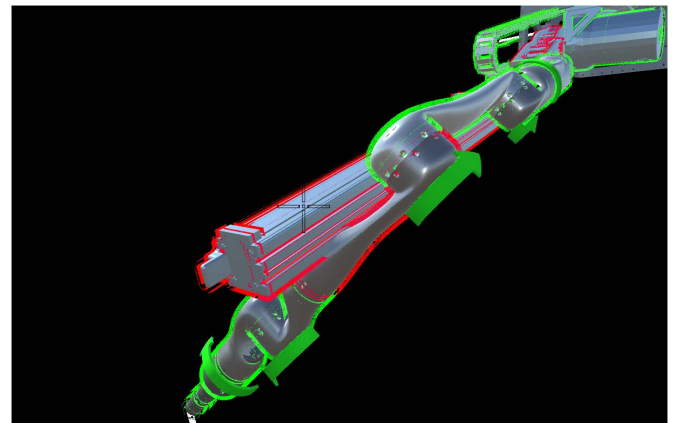

FIGURE 18. Real arm self-collision detection.

movement of the arm is stopped. The situations when the real arm collided can be seen in Figures 17 and 18.

G. TARGET APPROACH WITH A POINT CLOUD NORMAL

The point cloud acquisition allows the interface to visualize the environment in 3D. Additionally, on most point cloud points, a surface normal can be calculated. Using this, a functionality of a simplified target approach has been implemented. The operator can select any point cloud point and a normal will be added in the environment. The direction of the surface normal is based on the surrounding points. However, the direction and distance can be adjusted by using a gizmo (Figure 19). Furthermore, in the planning mode with the inverse kinematics, the arm can automatically move to

this point on demand of the operator. This makes any target approach much faster and it requires only 4 quick operations:

1) Select a point cloud point for the normal vector;
2) Adjust the direction/distance of a point on the normal;
3) Command the end-effector to move to this point;
4) Launch movement after the preview arm collision check.

H. CLOSEST OBJECT DISTANCE INDICATOR

For further assistance with environment awareness and early prevention of collisions, or for precise end-effector manipulation, a functionality of showing the closest object distance to the end-effector has been implemented. In Figure 20, indicators for real and waypoint arms are presented. The distance is calculated between the closest point cloud point and the end-effector tip. Analogously, the closest object distance is indicated for the planning and preview arms.
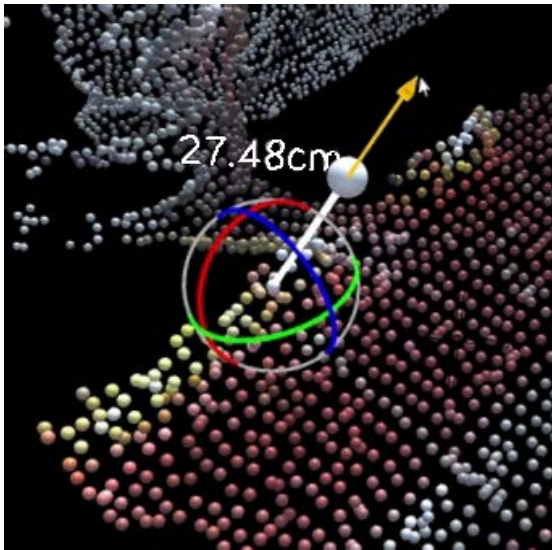
FIGURE 19. Target approach point normal to point cloud surface with gizmo interaction.

I. VIDEO AND POINT CLOUD FEEDBACK

During teleoperations, operators require real-time feedback from the robot. The feedback signals can be a single piece of information such as speed, force or position, or it can have a complex structure, such as a camera feedback. In the presented Mixed Reality Interface, in addition to the 2D video stream, the 3D feedback has a form of point cloud, which resembles reality as much as it is technically possible. The video feedback provides a clearer feedback due to its 2D form, however it does not provide precise 3D spatial information. When approaching an object, only the point cloud can provide direct distance information. The video and point cloud are acquired in real-time according to the parameters specified by the teleoperator. In the presented application an RGB-D camera was used as the point cloud source, but it is universal and can also be acquired from a different sensor type, for example, a LiDAR. The initial processing of the point cloud is done on the onboard computer, including filtering, subsampling, compression and serialization as needed by the task. After processing, it is sent to the interface.

1) RGB-D CAMERA SETTINGS AND ACQUISITION STATUS

Basic camera parameters are shown in Figure 21. Three of them (requested resolution, resolution feedback and requested FPS) are shared between the point cloud and video streams, while real video FPS and video stream enable are specific to video, and real point cloud FPS, point cloud enable and subsampling unit size are specific for point cloud. The subsampling unit size is the distance between point clouds points. This parameter can decrease the size of the transmitted point cloud by decreasing the point cloud density. The real video and point cloud FPS are estimated as the difference between the current and the previous received video or point cloud frame counter divided by a constant interval. The Equation 1 shows this estimation, where $C_{current}$ is the current
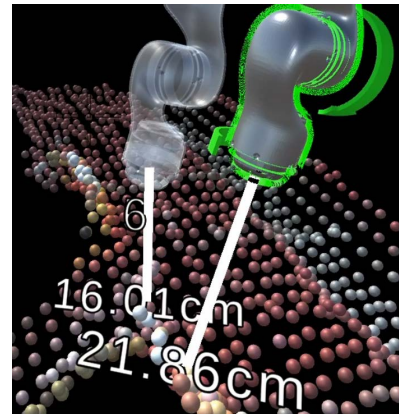

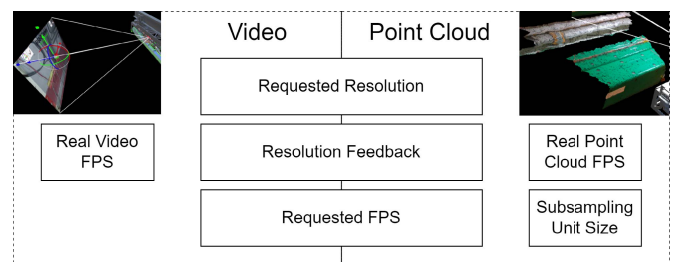**FIGURE 20.** Closest object indicators for real and waypoint arms.

FIGURE 21. Point cloud and video settings. The settings in the middle are shared between video and point cloud acquisition.

video or point cloud frame counter, $C_{previous}$ is the previous video or point cloud frame counter, FPS_{real} is the real video or point cloud FPS and Interval is the real FPS update interval that in this project it is set to 3 seconds but can be changed.

$$FPS_{real} = \frac{C_{current} - C_{previous}}{Interval} \qquad (1)$$

Depending on the model of the RGB-D camera, the available resolutions and corresponding allowed FPS settings may vary. That is why the choice of resolution is simplified to 3 options: LOW, MEDIUM, HIGH which are automatically selected from the available resolutions. The requested FPS and resolution are parameters which can be changed as inputs.

2) UNITY RGB-D CAMERA ARCHITECTURE

The Unity RGB-D camera structure is shown in Figure 22. The component controls the following functionalities:

- Camera control, communication with the camera process in the robot, setting the requested parameters, enabling video or point cloud streams;
- Managing the operator's interaction with the camera hologram;
- Display of a video canvas with projection lines and a point cloud area in the scene;
- Automatic camera settings controlled by camera internal state or network measurements (i.e. the focus of the paper).
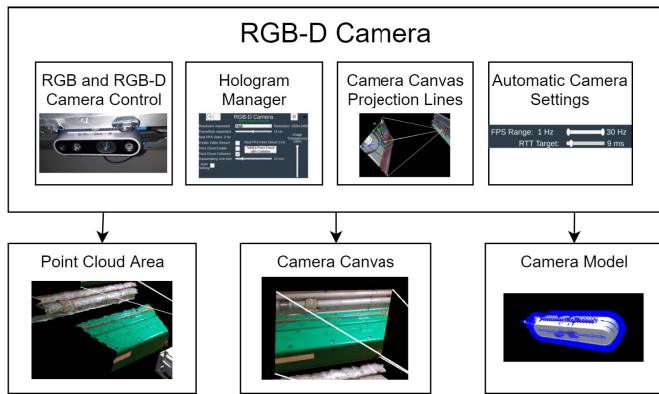
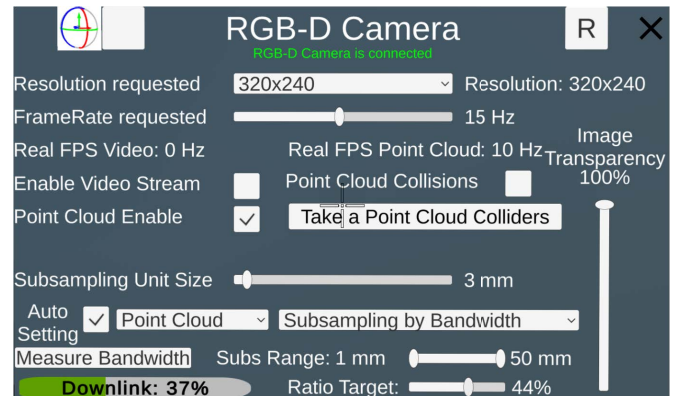**FIGURE 22.** RGB-D camera architecture in unity.

FIGURE 23. Camera hologram with automatic settings enabled. All the parameters of camera acquisition and automatic settings can be controlled here.

3) CAMERA HOLOGRAM

The teleoperator can modify the parameters of the camera acquisition (Section II-I1), video and point cloud streams, or automatic settings via a hologram. The hologram is a transparent 3D object displayed in the interface (Figures 23 and 24).

4) AUTOMATIC SETTINGS CONTROL

Automatic settings control (accessible to the teleoperator in the hologram shown in Figure 23) adjusts the camera settings depending on a chosen parameter. The operator can decide to use only 2D feedback, or both video and point cloud, or only point cloud. The automatically controlled settings and monitored parameters are camera-related (FPS, point cloud subsampling or resolution) but the monitored parameters can also be network-related (throughput, bandwidth, round-trip time). The synthesis of the automatic setting modes is in Table 1. The algorithms and hologram visual controls are grouped in Appendix C. A selected automatic setting algorithm is executed once per specified amount of time (e.g. 5 seconds). There are 4 distinct types of 12 automatic setting modes:

- **Settings controlled by throughput to bandwidth ratio:** During an intervention a robot can be located in areas where the network coverage is poor with highly limited bandwidth. Furthermore, if it happens that the available bandwidth is used in full capacity, it causes problems of even temporarily lost connection or stopping any video or point cloud stream. Moreover, the available bandwidth can vary over time and even nearby locations. This is why the automating setting of FPS, resolution or subsampling were introduced to allow the system to adapt to current conditions and facilitate the teleoperation. The setting adapts to the current throughput to bandwidth ratio, and the operator has the possibility to launch the available bandwidth measurement at any time. When the ratio is above the setpoint set by the operator, the FPS or resolution decreases or the subsampling increases. Accordingly, they change if the ratio is below the setpoint to give
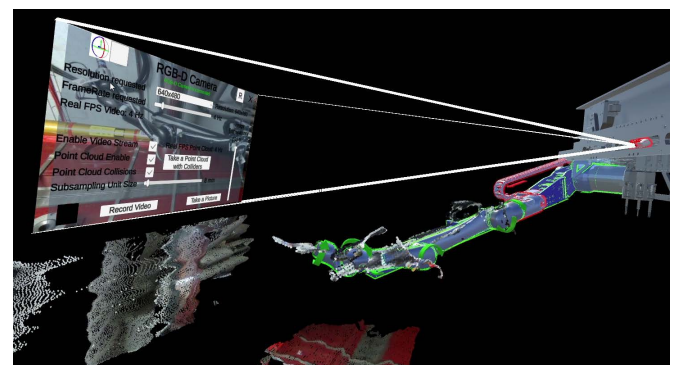
FIGURE 24. Transparent camera hologram view overlaid on the video stream in the 3D environment.

better quality of control. The operator can set the limits of FPS or subsampling within the range that the automatic control algorithm can adjust them. The algorithms using throughput to bandwidth ratio for point cloud FPS and subsampling control are described in Algorithms 1, 2, and for video FPS and resolution control in Algorithms 1, 3 in Appendix C. The holographic control is shown in Figures C.1, C.2, C.3 and C.4.

- **Settings controlled by round-trip time:** During a precise teleoperation it is important to have fast-enough and stable feedback from the robot to avoid collisions and to allow delicate manipulation, particularly if haptic feedback should be implemented. The frequency of camera feedback can be controlled with a setting, however the round-trip time depends on the used network and load. This is why the automatic setting mode monitoring round-trip time allows automatic control of the FPS, resolution and point cloud subsampling. The algorithms using round-trip time for point cloud FPS and subsampling control are described in Algorithms 4, 5, and for video FPS and resolution control in Algorithms 4, 6 in Appendix C. The holographic control is shown in Figures C.5, C.6, C.7 and C.8.

- **FPS controlled by point cloud subsampling or resolution:** Subsampling and resolution settings impact

the size of a single frame and allow the control of feedback precision required by a teleoperator. However, network bandwidth constraints might not allow high enough frame rates. That is why the requested FPS setting is automatically lowered in order to avoid buffering frames and a collapse of the stream. On the other hand, if the requested frame rate is achieved, the automatic setting algorithm tries to slowly increase the setting to increase the frame rate, which improves control quality. The automatic control is described in Algorithm 7 in Appendix C. The holographic control is shown in Figures C.9 and C.10.

- **Resolution or point cloud subsampling controlled by FPS:** In certain operation scenarios the teleoperator needs to have high-enough feedback frequency which is task and environment specific. The frequency defines the delay added to network delay that the teleoperator feels while controlling a robot. However, in this case the maximum size of a single feedback frame (point cloud or video) can be limited by network bandwidth. This is why in this automatic setting mode the resolution or point cloud subsampling are automatically adjusted to achieve the requested FPS. Similarly to the previous point, overloading the network can lead to a total collapse of stream due to buffering. On the other hand, if the network allows, the algorithm tries slowly to increase the precision (by increasing resolution or lowering subsampling). It is important to note that this mode does not use any direct network monitoring, it is based only on the camera behaviour. The automatic control is described in Algorithms 8 and 9 in Appendix C. The holographic control is shown in Figures C.11 and C.12.

J. NETWORK MEASUREMENTS

During an intervention, the operator can check the current network parameters and measured bandwidth, as shown in the menu screen in Figure 11 in Section II-C4 which provides a full overview of the network performance situation.

The camera acquisition throughput is proportional to the FPS and resolution, and inversely proportional to the subsampling (applicable for point cloud only), as presented in the Equation 2. The formulas calculating the round-trip time and deviation were inspired by the TCP procedures [39].

$$\text{Throughput} \sim \text{FPS} \cdot \frac{1}{\text{subsampling}} \cdot \text{resolution} \quad (2)$$

Another parameter is the round-trip time, which is calculated according to the Equation 3, and standard values are shown in the Equation 4.

$$\text{RTT} = (1 - \alpha) \cdot \text{PreviousRTT} + \alpha \cdot \text{CurrentRTT} \quad (3)$$

where $\alpha = 0.125$, then

$$\text{RTT} = 0.875 \cdot \text{PrevRTT} + 0.125 \cdot \text{CurrRTT} \quad (4)$$

The round-trip time oscillations are also important to assess the network behaviour; the Equations 5 and 6 present

how it is calculated.

$$\begin{aligned} \text{DevRTT} = {} & (1 - \beta) \cdot \text{PrevDevRTT} \\ & + \beta \cdot |\text{CurrRTT} - \text{PrevRTT}| \end{aligned} \quad (5)$$

where $\beta = 0.25$, then

$$\begin{aligned} \text{DevRTT} = {} & 0.75 \cdot \text{PrevDevRTT} \\ & + 0.25 \cdot |\text{CurrRTT} - \text{PrevRTT}| \end{aligned} \quad (6)$$

The timeout interval is an indicative parameter in the communication performance, it is presented and calculated according to the Equation 7.

$$\text{TimeoutInterval} = \text{PrevRTT} + 4 \cdot \text{PrevDevRTT} \quad (7)$$

III. EXPERIMENTS
A. EXPERIMENT SET-UP
1) NETWORK CHARACTERIZATION
During the CERN robotic BLM validation measurements the robot is connected to the operator via the 4G/LTE network present in the LHC tunnel - currently the only suitable communications link with large coverage of the LHC. A large number of equipment would be necessary to provide full Wi-Fi coverage and it is not feasible to adequately protect standard Wi-Fi active electronic devices from ionising radiation, as explained in [40]. The 4G network uses a coaxial antenna cable, which since it is a passive device, can be exposed to radiation. The authors registered download maximum speed of 70 Mbps and upload maximum speed 14 or 30 Mbps depending on location. The antenna is shared between other networks, such as the LoRA network used for other purposes. Therefore, the CERN network in the LHC is characterized by volatility, variable bandwidth and latency and lack of information of the current use of the network which can all have an impact on robotic operations [32], [33].

During the experiments, 4 connection types were tested (Table 3) to provide a wider set of behaviours. Network parameters measured for all connection types are shown in Table 4. The Ethernet cable connection was expected to be a reference value as a maximum bandwidth and minimum round-trip-time that could be achieved. The connection of Ethernet over CERN General Purpose Network (GPN) is characterized by relatively very good bandwidth
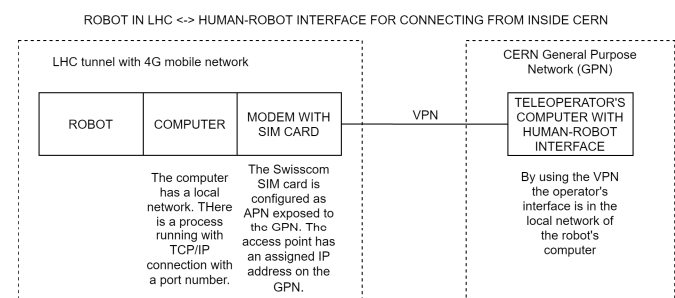

FIGURE 25. Connection between the camera acquisition system installed in a robot and the interface using automatic setting. A connection type using VPN 4G modem in the LHC tunnel is shown here.

TABLE 1. Point cloud and video automatic setting modes. All automatic modes are listed, the automatically controlled setting and settings controlled by operator are specified for each algorithm.

	Automatic setting mode	Automatically controlled settings	Settings controlled by operator
Point cloud	FPS by throughput to bandwidth ratio	FPS	Throughput to bandwidth ratio target, subsampling, FPS limits
	FPS by round-trip time		Round-trip time limit, subsampling, FPS limits
	FPS by subsampling		Subsampling, FPS limits
	Subsampling by throughput to bandwidth ratio	Subsampling	Throughput to bandwidth ratio target, FPS, subsampling limits
	Subsampling by round-trip time		Round-trip time limit, FPS, subsampling limits
	Subsampling by FPS		FPS, subsampling limits
Video	FPS by throughput to bandwidth ratio	FPS	Throughput to bandwidth ratio target, resolution, FPS limits
	FPS by Round-trip time		Round-trip time limit, resolution, FPS limits
	FPS by resolution		Resolution, FPS limits
	Resolution by throughput to bandwidth ratio	Resolution	Throughput to bandwidth ratio target, FPS
	Resolution by round-trip time		Round-trip time limit, FPS
	Resolution by FPS		FPS

TABLE 2. Automatic settings behaviours and dependencies of parameters. The arrows represent changes, i.e.: in the round-trip time dependent algorithms, when the round-trip time increases, an automatic setting algorithm will decrease the requested FPS or resolution, or increase subsampling to maintain the requested round-trip time target.

Parameter	Parameter change	Automatic settings controlled		
		Requested FPS	Subsampling	Resolution
Throughput / Bandwidth	↗	↘	↗	↘
	↘	↗	↘	↗
Round-trip time	↗	↘	↗	↘
	↘	↗	↘	↗
Real FPS	↗	↗	↘	↗
	↘	↘	↗	↘

and round-trip time values, however the bandwidth starts to vary over time-dependent infrastructure load. The Wi-Fi over CERN GPN significantly lowers the available bandwidth (74 Mbps) and the round-trip time is much longer (20-30 ms) and has a jitter of 5 ms. The VPN with 4G modem connection type had the worst results, with significant round-trip time (43-132 ms), depending highly on the bandwidth usage and had a jitter of 26 ms), and a low bandwidth value (12 Mbps) although the deviation of the bandwidth was low (< 1 Mbps). The VPN with 4G modem connection type is largely used for real intervention scenarios at CERN.

2) PARAMETERS OF VIDEO AND POINT CLOUD

During the experiments, all relevant settings and measurements were logged at the moment of acquisition of the point cloud or video frame to understand and recreate the state of the user interface and the network. The parameters presented in the graphs in Section III-B are:

- requested resolution, FPS and subsampling unit size of point cloud;
- real FPS for video or point cloud - calculated as number of received frames per last second;
- downlink throughput / bandwidth ratio;
- measured round-trip time.

3) EXPERIMENTS OVERVIEW

The network and camera video or point cloud streaming parameters and behaviour were tested. The experiments were undertaken separately for both the point cloud and video streams. In each experiment, there was one setting (i.e. FPS, resolution, subsampling, throughput to bandwidth ratio or round-trip time limit) that was the experiment setpoint, changing in time according to the experiment sequence. Other settings were either constant, not used or automatically controlled by the algorithms. In total, there were 16 types of experiments. During the first half of the experiment the demand for throughput was increasing, and in the second part decreasing. There were 4 experiments in which the automatic setting algorithms were not enabled to test the behaviour and encountered problems, and 12 experiments where there was one automatic setting algorithm enabled to test if and how the problems were solved or what was the improvement in the behaviour. The overview of experiments, showing which parameter was the setpoint and which were automatically controlled or constant, is shown in Table 5.

Additionally, the times of point cloud processing on the robot/server and the operator/client sides were measured to provide the range of the processing delay in addition to the network delay.

4) HARDWARE

As shown in Figure 25, the connections were established between 2 machines: the operator interface and the robot computer. All the variations of interfaces were tested. The

TABLE 3. Connection types between a robot and an operator's computer.

Network connection type	Robot side's connection	Interconnection	Operator's computer
Ethernet cable directly	Ethernet cable connector	Single cable	Ethernet cable connector
Ethernet over CERN GPN	Ethernet cable connection to CERN General Purpose Network	CERN General Purpose Network cabled indrastructure	Ethernet cable connection to CERN General Purpose Network
VPN 4G modem in the LHC tunnel	4G modem	Network operator 4G connection in the LHC tunnel, VPN connection to CERN General Purpose Network cabled infrastructure	Ethernet cable connection to CERN General Purpose Network
Wi-Fi over CERN GPN	Wi-Fi connection to CERN General Purpose Network Wi-Fi infrastruture	CERN General Purpose Network cabled infrastructure	Wi-Fi connection to CERN General Purpose Network Wi-Fi infrastructure

TABLE 4. Bandwidth and its deviation, round-trip time and its jitter measurements for all connection types.

Network connection type	Downlink bandwidth [Mbps]	Round-trip time [ms]		Jitter [ms]	Bandwidth standard deviation [Mbps]
		Bandwidth usage = 0%	Bandwidth usage = 100%		
VPN with 4G modem	11.99	43.4	131.5	26.00	0.61
Wi-Fi over CERN GPN	73.95	20.33	30.3	5.64	17.88
Ethernet over CERN GPN	885.8	0.04	1.61	0.19	38.1
Ethernet cable directly	941.8	0.24	1.47	0.28	0.42

FIGURE 26. Teltonika RUTX12 used for experiment with the VPN 4G modem over CERN GPN connection type experiments.

experiments were done with a hardware configuration shown in Table 6. The client was a Windows machine running the Unity project, while the robot onboard computer (Intel NUC) ran Linux Ubuntu. In the experiments with the VPN connection over 4G network, a mobile modem was used (Figures 26 and 27) and the onboard computer was the server. In the experiments over Wi-Fi or cabled connection over CERN GPN, a DELL computer was the server. The Intel RealSense D415 and D435 camera were used for the experiments (Figures 28, 29 and 30).

B. RESULTS STRUCTURE
The results are organized in the order presented in Table 5 and divided in two groups; the first where the automatic control of settings was not used and the second where the automatic

control was used. Although the experiments were done with all the network connection types (Table 3), in each experiment section there are results and graphs with chosen connection type(s) that best present the situation, problem or solution.

C. EXPERIMENTS WITH NO AUTOMATIC SETTINGS
The first group of experiments with no automatic control of settings is presented in this section. The behaviours and problems shown here are often encountered during real robotic interventions. The consequences for teleoperation caused by these behaviours are synthesized in Table 8 in Section III-G.

1) POINT CLOUD EXPERIMENT: CHANGING FPS
The FPS setpoint was changing over time, while the point cloud subsampling and resolution settings were constant. In Figures 31 and 32 the requested FPS could not be achieved due to the available bandwidth limit. After reaching the limit, the real FPS started to oscillate around its maximum, and the round-trip time increased 2 times instantaneously.

2) POINT CLOUD EXPERIMENT: CHANGING SUBSAMPLING
The subsampling setpoint was changing over time and the point cloud requested FPS and resolution settings were constant. In Figures 33 and 34 the acquisition FPS decreased significantly after reaching the bandwidth limit. The used connection type was Ethernet over CERN GPN. An analogous situation is shown for the Wi-Fi connection type (Figures 35 and 36), which has a lower bandwidth, and which

TABLE 5. Overview of experiments.

Setting control mode	Point cloud or video	N°	Experiment name	Settings			Network settings	
				FPS	Resolution	Subsampling	Throughput to bandwidth ratio	Round-trip time
Manual control of settings	Point cloud	1	Changing FPS	Experiment setpoint	Const	Const	Not used	
		2	Changing subsampling	Const		Experiment setpoint		
	Video	3	Changing FPS	Experiment setpoint	Const	Not used		
		4	Changing resolution	Const	Experiment setpoint			
Automatic settings	Point cloud	5	FPS by throughput to bandwidth ratio	Controlled by automatic setting algorithm	Const	Const	Experiment setpoint	Not used
		6	FPS by round-trip time				Not used	Experiment setpoint
		7	FPS by subsampling			Experiment setpoint		Not used
		8	Subsampling by throughput to bandwidth ratio	Const		Controlled by automatic setting algorithm	Experiment setpoint	Not used
		9	Subsampling by round-trip time				Not used	Experiment setpoint
		10	Subsampling by FPS					Not used
	Video	11	FPS by throughput to bandwidth ratio	Controlled by automatic setting algorithm	Const	Not used	Experiment setpoint	Not used
		12	FPS by round-trip time				Not used	Experiment setpoint
		13	FPS by resolution		Experiment setpoint			Not used
		14	Resolution by throughput to bandwidth ratio	Const	Controlled by automatic setting algorithm		Experiment setpoint	Not used
		15	Resolution by round-trip time				Not used	Experiment setpoint
		16	Resolution by FPS	Experiment setpoint				Not used

caused the problem to start earlier and for a lower requested FPS setting.

3) VIDEO EXPERIMENT: CHANGING FPS

When using VPN 4G modem connection in the LHC tunnel, the available bandwidth was limited and affected the video acquisition. After consuming all the available bandwidth, the requested FPS could not be achieved. The increase of the round-trip time, which doubled instantly by exceeding 200 ms, and its oscillation was very significant (100%), when the throughput reached the bandwidth. The observed behaviour is shown in Figures 37 and 38.

4) VIDEO EXPERIMENT: CHANGING RESOLUTION

The video resolution setpoint was changing over time, while the requested FPS was constant. The VPN 4G modem in the LHC tunnel connection type was used. As presented in Figures 39 and 40, for HIGH resolution, the requested FPS was not achieved due to bandwidth limit. The round-trip time increased by 4 times when the resolution was switched from MED to HIGH, and the throughput and round-trip time became unstable and started significant oscillations. The temporary real FPS falls were caused by switching resolution.

D. EXPERIMENTS WITH AUTOMATIC MODES

In this section, the experiments using automatic setting modes are presented. A synthesis of scenarios and solved problems or improvements due to the automatic settings is presented in Table 8 in III-G Section.

1) POINT CLOUD FPS CONTROLLED BY THROUGHPUT TO BANDWIDTH RATIO

The point cloud throughput to bandwidth ratio setpoint was changing over time, while the requested FPS was controlled

TABLE 6. The Human-Robot Interface's and robot's machines specification.

|	Human-Robot Interface machine (client)	Robot's machine (server)	
Device model	MSI GL75 Leopard 10SFR, rev 1.0	DELL Latitude E6230, 089JM2, rev. A00	Intel NUC11TNKv70QC
CPU	Intel(R) Core(TM) i7-10750H	Intel(R) Core(TM) i7-3520M	Intel(R) Core™ i7-1185G7
GPU	NVIDIA GeForce RTX 2070	Integrated, Intel	Integrated, Intel
Memory	16 GB, DDR4, 1333 MHz	4 GB, DDR3, 1600 MHz	64GB, DDR4, 3200 MHz
System	Windows 10 Pro, version 20H2	Linux Ubuntu 20.04.2 LTS	Linux Ubuntu 20.04.2 LTS
Benchmark	Graphics 7435, CPU 6622 (3DMark TimeSpy v1.2), 523 with 1 thread, 3747 with 12 threads (CPU-Z v.17.01.64)	35855 with 1 thread, 64978 with 4 threads (CPU-X v.3.2.4, slow prime numbers)	68000 with 1 thread, 127358 with 8 threads (CPU-X v.3.2.4, slow prime numbers)
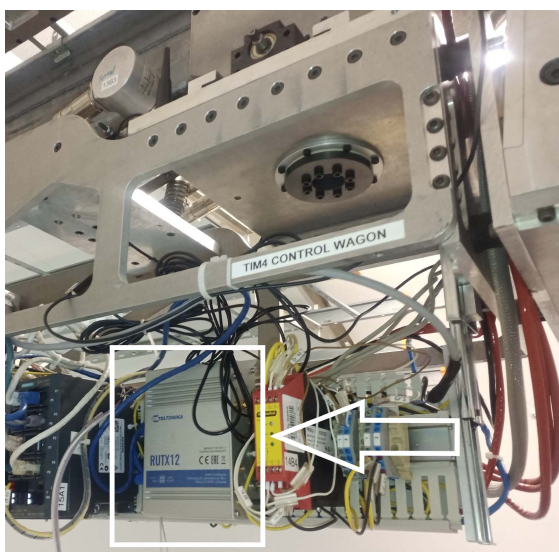
FIGURE 27. Teltonika RUTX12 installed in the Train Inspection Monorail wagon. The position of the router device in the robot is highlighted.

FIGURE 28. Intel RealSense D415 used for experiments, mounted on a tripod. The camera consists of a pair of depth sensors, an infrared projector, and an RGB sensor. The D415 model has a rolling shutter.
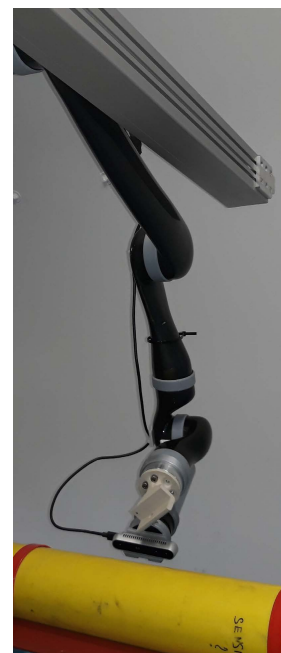

FIGURE 29. Intel RealSense D435, also used in the experiments, installed in the train inspection monorail wagon. The D435 model, compared to D415, has a wider field of view (FOV) and a global shutter.

FIGURE 30. The D415 camera mounted on the effector of the BLM manipulator that approached a beam loss monitor.

by the automatic setting algorithm to follow the setpoint, and the resolution and subsampling were constant. The Wi-Fi over CERN GPN connection type is presented. In Figure 41, the requested FPS adjustments are changing the throughput correctly, and the real FPS value is following it precisely. The requested FPS adapts automatically to the size of the point cloud and the specified ratio and subsampling, while maintaining a stable acquisition stream.

2) POINT CLOUD FPS CONTROLLED BY ROUND-TRIP TIME

The point cloud round-trip time setpoint was changing over time, the requested FPS was controlled by the automatic setting algorithm to follow the setpoint, and the resolution and subsampling were constant. The Wi-Fi over CERN GPN connection type is presented. In Figures 42 and 43, the requested round-trip time was automatically followed by adjusting the requested FPS setting. Despite major oscillations of the round-trip time characterizing this connection type, the setpoint was followed correctly.
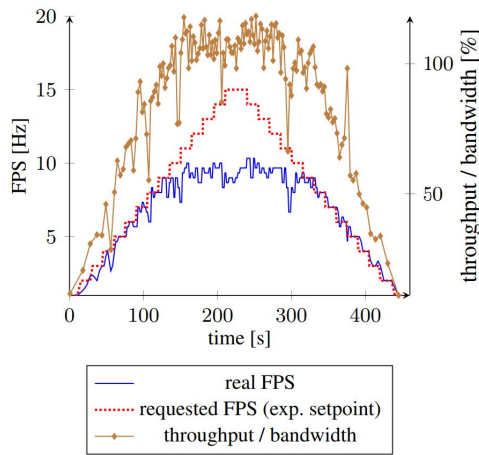
FIGURE 31. Point cloud experiment without automatic settings (exp. n° 1 in Table 5). Requested FPS as setpoint. Throughput to bandwidth ratio, real FPS behaviour presented. Connection type: Wi-Fi over CERN GPN. Constant parameters: resolution = MEDIUM, subsampling = 10 mm.
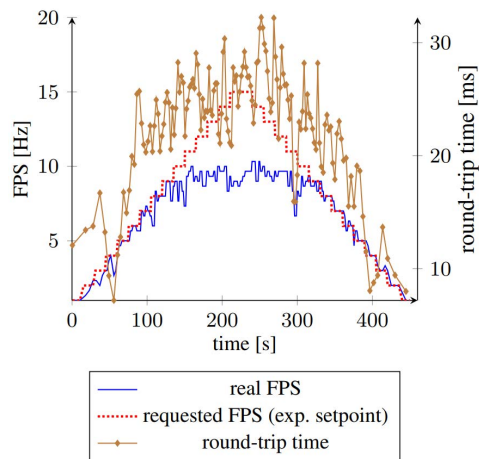

FIGURE 32. Point cloud experiment without automatic settings (exp. n° 1 in Table 5). Requested FPS as setpoint. Round-trip time, real FPS behaviour presented. Connection type: Wi-Fi over CERN GPN. Constant parameters: resolution = MEDIUM, subsampling = 10 mm.

3) POINT CLOUD FPS CONTROLLED BY SUBSAMPLING

The point cloud subsampling setpoint was changing over time, the requested FPS was controlled by the automatic setting algorithm to follow the setpoint, and the resolution was constant. The Wi-Fi over CERN GPN connection type is presented. In Figures 44 and 45, by changing the subsampling, the requested FPS automatically adjusted to maintain the used bandwidth at high level. This mode does not need any network measurement and is based only on the internal acquisition parameters. The bandwidth limit was probed by increasing the requested FPS until the acquisition failed to achieve the setting, then decreasing it accordingly. The round-trip time was maintained at a similar level with natural oscillations for this connection type.

4) POINT CLOUD SUBSAMPLING CONTROLLED BY THROUGHPUT TO BANDWIDTH RATIO

In this experiment, the point cloud throughput to bandwidth ratio setpoint was changing over time, the subsampling was
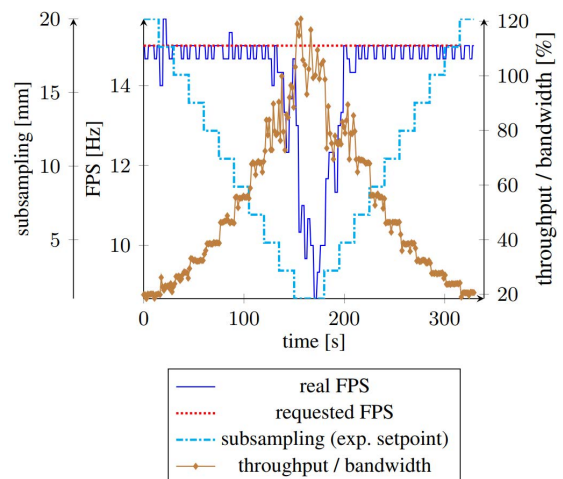

FIGURE 33. Point cloud experiment without automatic settings (exp. n° 2 in Table 5). Subsampling as setpoint. Throughput to bandwidth ratio, requested and real FPS behaviour presented. Connection type: Ethernet over CERN GPN. Constant parameters: resolution = MEDIUM, requested FPS = 15 Hz.

FIGURE 34. Point cloud experiment without automatic settings (exp. n° 2 in Table 5). Subsampling as setpoint. Round-trip time, requested and real FPS behaviour presented. Connection type: Ethernet over CERN GPN. Constant parameters: resolution = MEDIUM, requested FPS = 15 Hz.
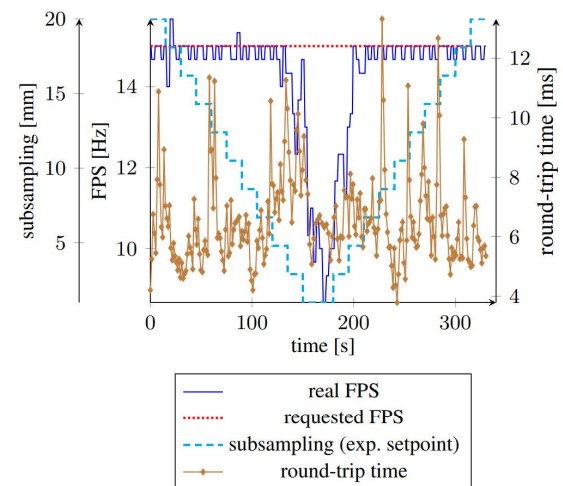
controlled by the automatic setting algorithm to follow the setpoint, and the resolution and requested FPS were constant. The Wi-Fi over CERN GPN connection type is presented. In Figures 46 and 47, the subsampling setting was adapting to the throughput to bandwidth ratio target, while maintaining the real FPS at the value set by the requested FPS setting, even for the highest setpoint values. The round-trip time approximately doubled when the throughput to bandwidth ratios had minimum and maximum values. There was an occasional spike, which was observed for this connection type.

5) POINT CLOUD SUBSAMPLING CONTROLLED BY ROUND-TRIP TIME

The point cloud round-trip time setpoint was changing over time, the subsampling was controlled by the automatic setting algorithm to follow the setpoint, and the resolution and requested FPS were constant. The Wi-Fi over CERN GPN
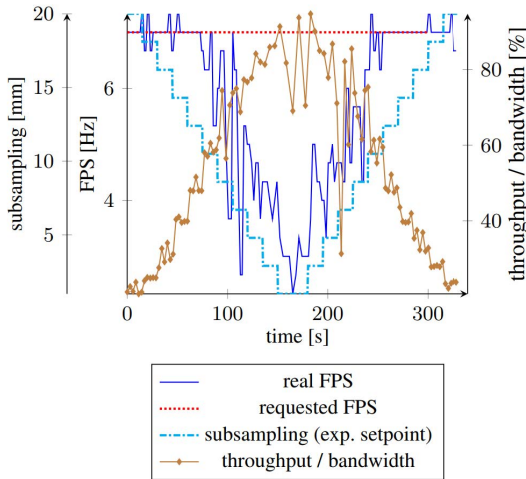
FIGURE 35. Point cloud experiment without automatic settings (exp. n° 2 in Table 5). Subsampling as setpoint. Throughput to bandwidth ratio, requested and real FPS behaviour presented. Connection type: Wi-Fi over CERN GPN. Constant parameters: resolution = MEDIUM, requested FPS = 7 Hz.

FIGURE 37. Video experiment without automatic settings (exp. n° 3 in Table 5). Requested FPS as setpoint. Throughput to bandwidth ratio, real FPS behaviour presented. Connection type: VPN 4G modem in the LHC tunnel. Constant parameters: resolution = HIGH.

FIGURE 36. Point cloud experiment without automatic settings (exp. n° 2 in Table 5). Subsampling as setpoint. Round-trip time, requested and real FPS behaviour presented. Connection type: Wi-Fi over CERN GPN. Constant parameters: resolution = MEDIUM, requested FPS = 7 Hz.
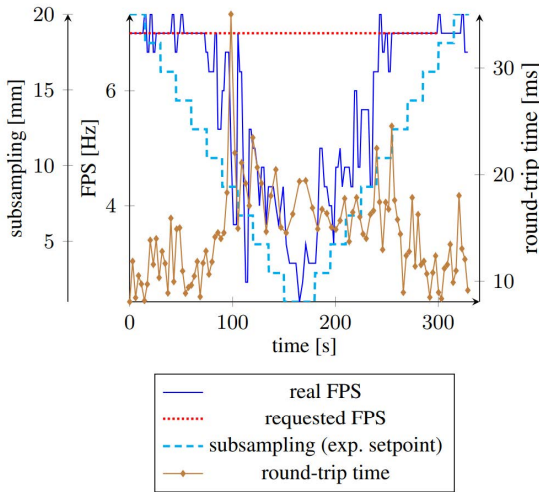

FIGURE 38. Video experiment without automatic settings (exp. n° 3 in Table 5). Requested FPS as setpoint. Round-trip time, real FPS behaviour presented. Connection type: VPN 4G modem in the LHC tunnel. Constant parameters: resolution = HIGH.

connection type is presented. In Figures 48 and 49, the round-trip time target was followed, although the achieved round-trip time had significant oscillations, which characterize the used network. Because of these oscillations, the subsampling had to change often by up to 30%. Occasionally, when the throughput to bandwidth ratio was close to 100%, the real FPS value decreased by a few frames per second below the requested FPS.

6) POINT CLOUD SUBSAMPLING CONTROLLED BY FPS

The point cloud requested FPS setpoint was changing over time, the subsampling was controlled by the automatic setting algorithm to follow the setpoint, and the resolution was constant. The Wi-Fi over CERN GPN connection type is presented. In the Figures 50 and 51, the real FPS followed the requested FPS target by automatically adjusting the subsampling setting. The bandwidth use was indirectly maximized whenever possible to have a minimum subsampling

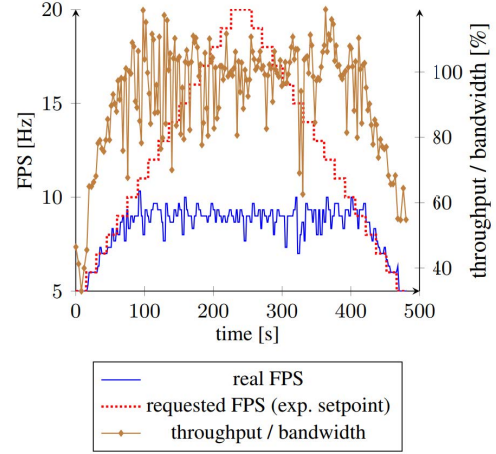**FIGURE 39.** Video experiment without automatic settings (exp. n° 4 in Table 5). Resolution as setpoint. Throughput to bandwidth ratio, requested and real FPS behaviour presented. Connection type: VPN 4G modem in the LHC tunnel. Constant parameters: FPS = 20 Hz.
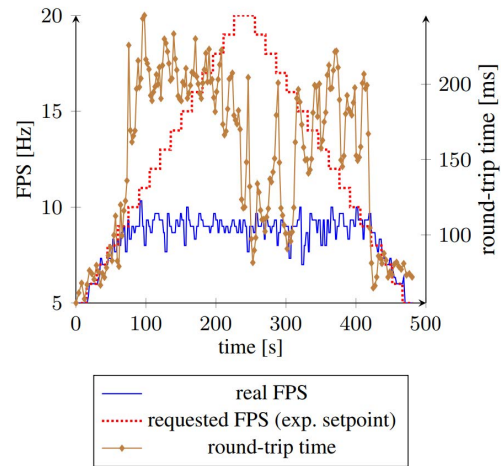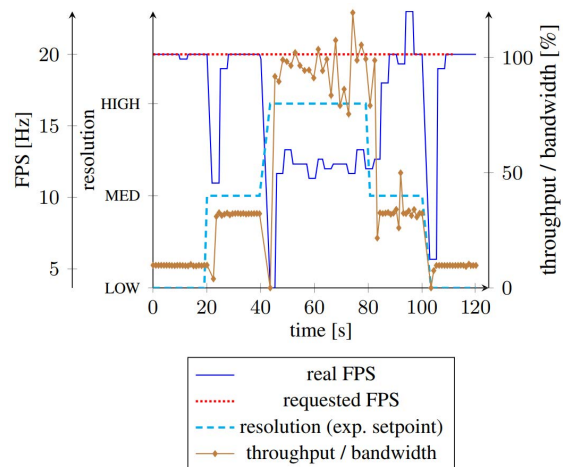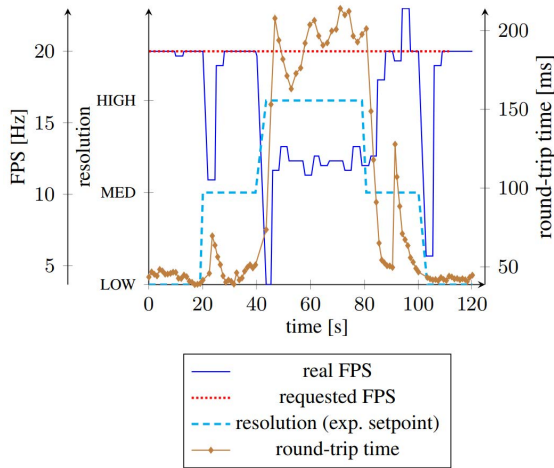
FIGURE 40. Video experiment without automatic settings (exp. n° 4 in Table 5). Resolution as setpoint. Round-trip time, requested and real FPS behaviour presented. Connection type: VPN 4G modem in the LHC tunnel. Constant parameters: FPS = 20 Hz.
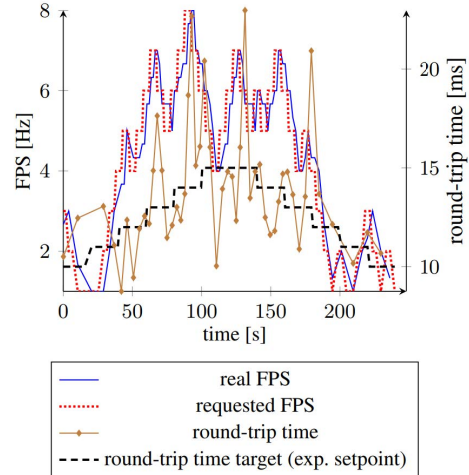

FIGURE 42. Point cloud experiment with automatic settings (exp. n° 6 in Table 5). FPS controlled by round-trip time target as setpoint. Round-trip time, requested and real FPS behaviour presented. Connection type: Wi-Fi over CERN GPN. Constant parameters: resolution = MEDIUM, subsampling = 15 mm.
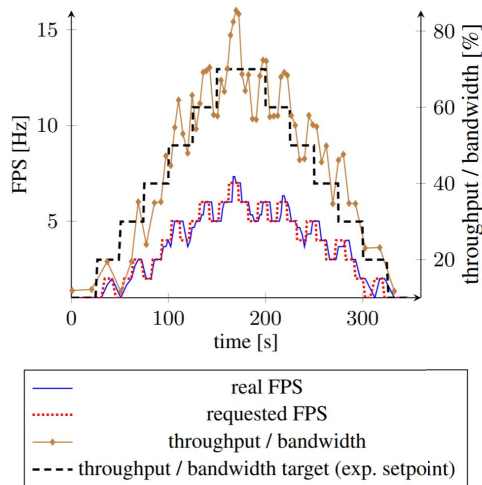
FIGURE 41. Point cloud experiment with automatic settings (exp. n° 5 in Table 5). FPS was controlled by throughput to bandwidth ratio target as the setpoint. The throughput to bandwidth ratio, requested and real FPS behaviour are presented with the onnection type of Wi-Fi over CERN GPN and with constant parameters: resolution = MEDIUM, subsampling = 10 mm.

FIGURE 43. Point cloud experiment with automatic settings (exp. n° 6 in Table 5). FPS controlled by round-trip time target as setpoint. Throughput to bandwidth ratio, requested and real FPS behaviour presented. Connection type: Wi-Fi over CERN GPN. Constant parameters: resolution = MEDIUM, subsampling = 15 mm.
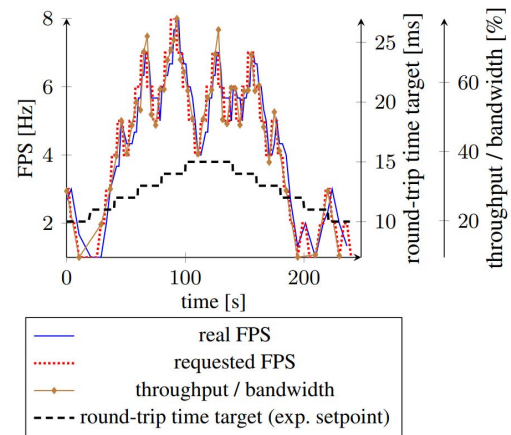
8) VIDEO FPS CONTROLLED BY ROUND-TRIP TIME

The video round-trip time setpoint was changing over time, the requested FPS was controlled by the automatic setting algorithm to follow the setpoint, and the resolution was constant. The VPN 4G modem in the LHC tunnel connection type is presented. In Figures 54 and 55, the round-trip time correctly followed the requested target, though it was characterized with significant oscillations (up to 80% temporary increase), which are characteristic to this connection type. The throughput to bandwidth ratio was proportional to the real FPS setting that was automatically controlled.

9) VIDEO FPS CONTROLLED BY RESOLUTION

The video resolution was changing over time, and the FPS was controlled by the automatic setting algorithm to follow the setpoint. The VPN 4G modem in the LHC tunnel connection type is presented. In Figures 56 and 57, the requested FPS adapted to the active resolution by probing the highest

for a given requested FPS setting. The round-trip time was proportional to the throughput to bandwidth ratio with an occasional spike.

7) VIDEO FPS CONTROLLED BY THROUGHPUT TO BANDWIDTH RATIO

The video throughput to bandwidth ratio setpoint was changing over time, the requested FPS was controlled by the automatic setting algorithm to follow the setpoint, and the resolution was constant. The VPN over CERN GPN connection type is presented. In Figures 52 and 53, the requested FPS setting automatically adapted to the throughput to bandwidth ratio target changes. The round-trip time increased when more throughput was used, and had significant oscillations characteristic to this connection type.
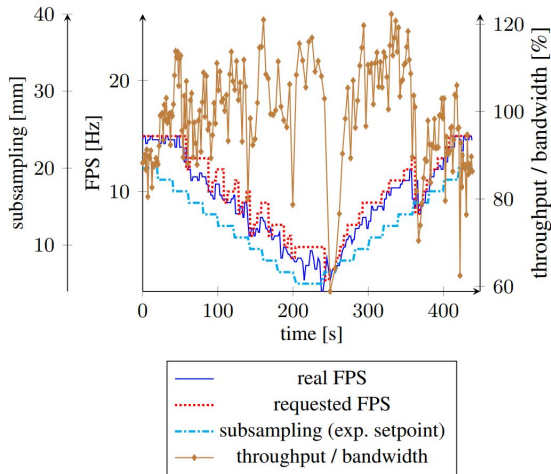
FIGURE 44. Point cloud experiment with automatic settings (exp. n° 7 in Table 5). FPS controlled by subsampling as setpoint. Throughput to bandwidth ratio, requested and real FPS behaviour presented. Connection type: Wi-Fi over CERN GPN. Constant parameters: resolution = MEDIUM.
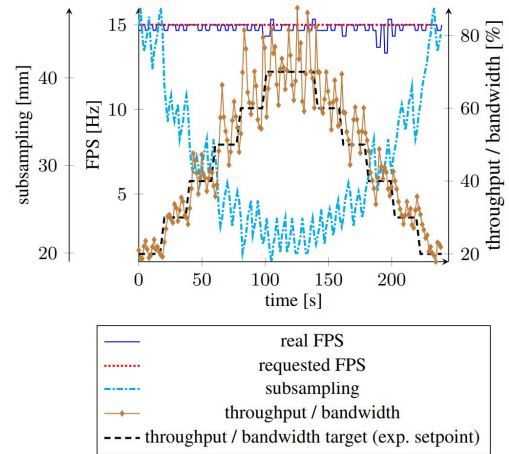

FIGURE 46. Point cloud experiment with automatic settings (exp. n° 8 in Table 5). Throughput to bandwidth ratio target as setpoint. Throughput to bandwidth ratio, subsampling, requested and real FPS behaviour presented. Connection type: Wi-Fi over CERN GPN. Constant parameters: resolution = MEDIUM, requested FPS = 15 Hz.

FIGURE 45. Point cloud experiment with automatic settings (exp. n° 7 in Table 5). FPS controlled by subsampling as setpoint. Round-trip time, requested and real FPS behaviour presented. Connection type: Wi-Fi over CERN GPN. Constant parameters: resolution = MEDIUM.

FIGURE 47. Point cloud experiment with automatic settings (exp. n° 8 in Table 5). Throughput to bandwidth ratio target as setpoint. Round-trip time, subsampling, requested and real FPS behaviour presented. Connection type: Wi-Fi over CERN GPN. Constant parameters: resolution = MEDIUM, requested FPS = 15 Hz.

possible value allowed by the bandwidth. With the HIGH resolution, the bandwidth limit was reached and the FPS had to be limited to around 3 times from the value that was used for the MED resolution. Temporary drops of the real FPS value were caused by resolution changes in the camera acquisition process.

10) VIDEO RESOLUTION CONTROLLED BY THROUGHPUT TO BANDWIDTH RATIO

The video throughput to bandwidth ratio was changing over time, and the resolution was controlled by the automatic setting algorithm to follow the setpoint. The VPN 4G modem in the LHC tunnel connection type is presented. In Figures 58 and 59, the resolution setting correctly followed the throughput to bandwidth ratio setpoint. Temporary drops of the real FPS value come from the resolution change in the camera acquisition process.

FIGURE 48. Point cloud experiment with automatic settings (exp. n° 9 in Table 5). Subsampling unit size controlled by round-trip time target as setpoint. Round-trip time, subsampling, requested and real FPS behaviour presented. Connection type: Wi-Fi over CERN GPN. Constant parameters: resolution = MEDIUM, requested FPS = 15 Hz.

FIGURE 49. Point cloud experiment with automatic settings (exp. n° 9 in Table 5). Subsampling unit size controlled by round-trip time target as setpoint. Throughput to bandwidth ratio, subsampling, requested and real FPS behaviour presented. Connection type: Wi-Fi over CERN GPN. Constant parameters: resolution = MEDIUM, requested FPS = 15 Hz.
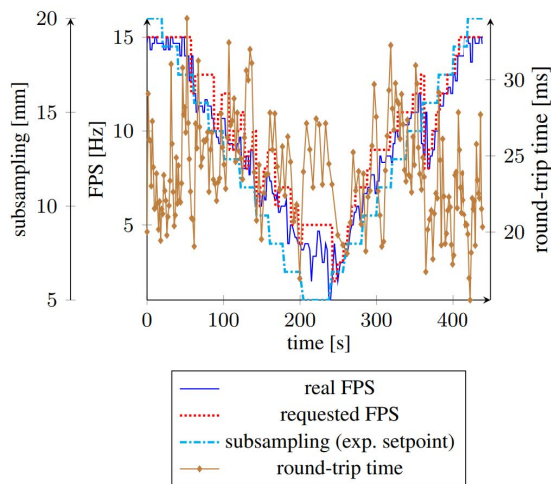

FIGURE 50. Point cloud experiment with automatic settings (exp. n° 10 in Table 5). Subsampling unit size controlled by FPS as setpoint. Throughput to bandwidth ratio, requested and real FPS behaviour presented. Connection type: Wi-Fi over CERN GPN. Constant parameters: resolution = MEDIUM.

FIGURE 51. Point cloud experiment with automatic settings (exp. n° 10 in Table 5). Subsampling unit size controlled by FPS as setpoint. Round-trip time, requested and real FPS behaviour presented. Connection type: Wi-Fi over CERN GPN. Constant parameters: resolution = MEDIUM.
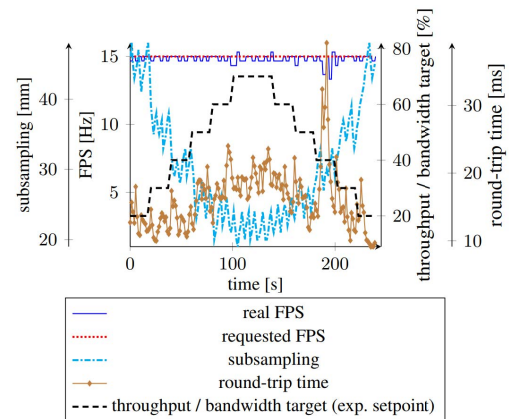

FIGURE 52. Video experiment with automatic settings (exp. n° 11 in Table 5). FPS controlled by throughput to bandwidth ratio as setpoint. Throughput to bandwidth ratio, requested and real FPS behaviour presented. Connection type: VPN 4G modem in the LHC tunnel. Constant parameters: resolution = HIGH.

FIGURE 53. Video experiment with automatic settings (exp. n° 11 in Table 5). FPS controlled by throughput to bandwidth ratio as setpoint. Round-trip time, requested and real FPS behaviour presented. Connection type: VPN 4G modem in the LHC tunnel. Constant parameters: resolution = HIGH.

FIGURE 54. Video experiment with automatic settings (exp. n° 12 in Table 5). FPS controlled by round-trip time target as setpoint. Round-trip time, requested and real FPS behaviour presented. Connection type: VPN 4G modem in the LHC tunnel. Constant parameters: resolution = HIGH.
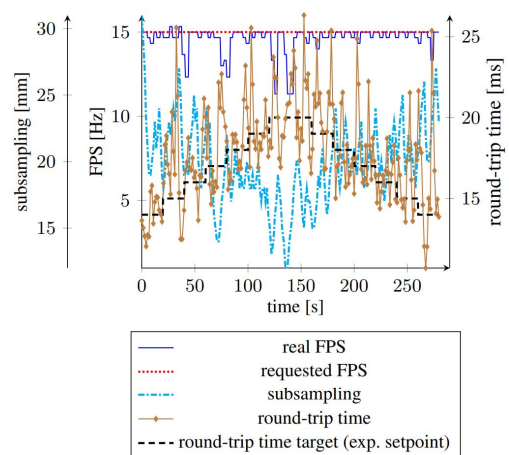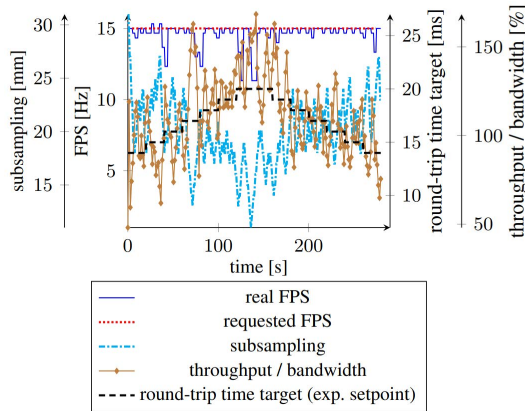
FIGURE 55. Video experiment with automatic settings (exp. n° 12 in Table 5). FPS controlled by round-trip time target as setpoint. Throughput to bandwidth ratio, requested and real FPS behaviour presented. Connection type: VPN 4G modem in the LHC tunnel. Constant parameters: resolution = HIGH.

FIGURE 57. Video experiment with automatic settings (exp. n° 13 in Table 5). FPS controlled by resolution as setpoint. Round-trip time, requested and real FPS behaviour presented. Connection type: VPN 4G modem in the LHC tunnel.

FIGURE 56. Video experiment with automatic settings (exp. n° 13 in Table 5). FPS controlled by resolution as setpoint. Throughput to bandwidth ratio, requested and real FPS behaviour presented. Connection type: VPN 4G modem in the LHC tunnel.
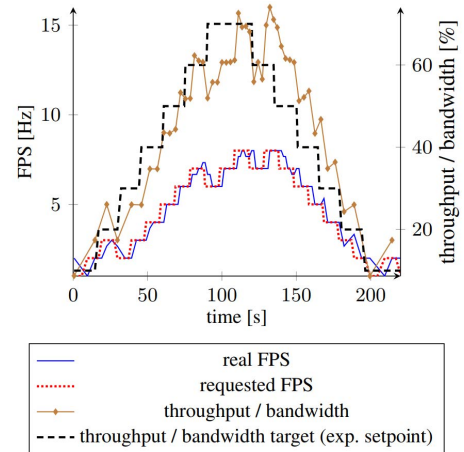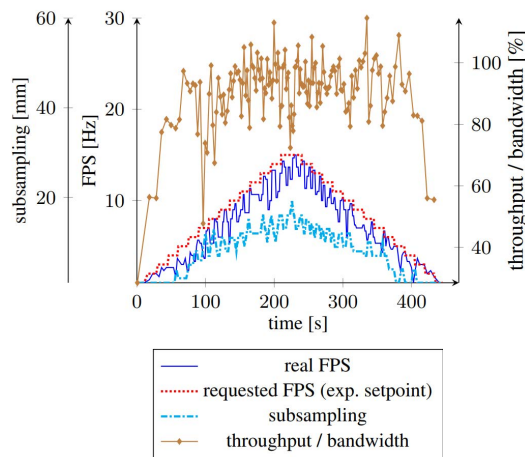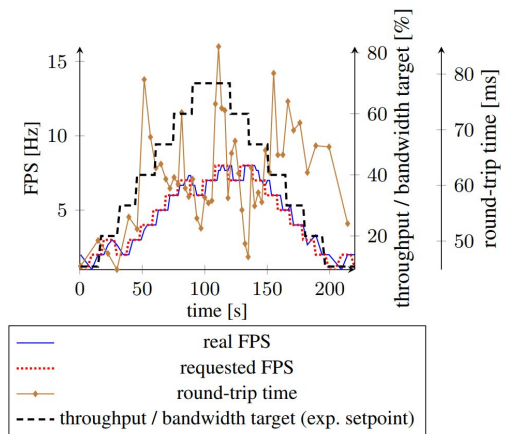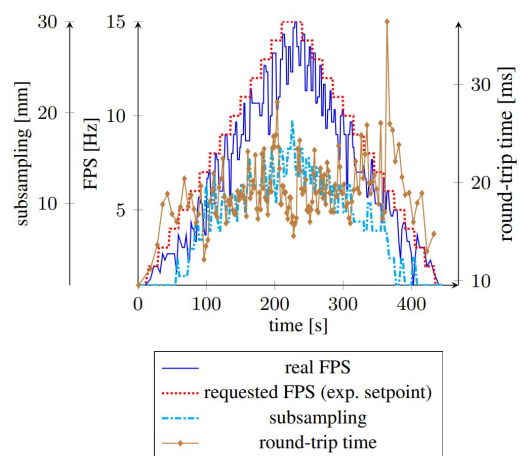

FIGURE 58. Video experiment with automatic settings (exp. n° 14 in Table 5). Resolution controlled by throughput to bandwidth ratio target as setpoint. Throughput to bandwidth ratio, requested and real FPS behaviour presented. Connection type: VPN 4G modem in the LHC tunnel. Constant parameters: FPS = 13 Hz.

11) VIDEO RESOLUTION CONTROLLED BY ROUND-TRIP TIME

The video round-trip time was changing over time, and the resolution was controlled by the automatic setting algorithm to follow the setpoint. The VPN 4G modem in the LHC tunnel connection type is presented. In Figures 60 and 61, the round-trip time setpoint was changing the resolution setting to achieve the target value. The round-trip time increased especially when the resolution changed from MED to HIGH.

12) VIDEO RESOLUTION CONTROLLED BY FPS

The video requested FPS was changing over time, and the resolution was controlled by the automatic setting algorithm to follow the setpoint. The VPN 4G modem in the LHC tunnel connection type is presented. In Figures 62 and 63, the resolution was changing in the function of the requested FPS. If the requested FPS could not be achieved, the resolution

dropped, and correspondingly, if the FPS was achieved for a given resolution, an attempt to increase the resolution was made.

E. POINT CLOUD PROCESSING TIME MEASUREMENTS

The measured times of point cloud processing are shown in Table 7. The processing time increases with the size of the transmitted point cloud, and can even reach ∼245 ms for a 1 mm resolution point cloud. The total time does not include the network delay, which is communication type dependent, and the camera internal latency (ranging, for example, from 25 ms to 127 ms for the D435 camera and 848 × 100, according to [41]), which generally depends on the camera capture mode, frame rate and resolution. The execution time of automatic setting algorithms was less than 1 μs.

TABLE 7. Times of point cloud processing on the robot/server side (capturing, subsampling, serializing) and the operator/client side (deserializing, rendering, automatic settings). The times are presented in function of subsampling unit size for a fixed number of the initially acquired point cloud points.

Number of point cloud points before subsampling	Subsampling unit size [mm]	Number of point cloud points after subsampling	Capturing [ms]	Subsampling [ms]	Serializing [ms]	Deserializing [ms]	Rendering [ms]	Automatic settings [ms]	Total time [ms]
230400	1mm	230400	1.2	9.9	52.9	136.5	44.4	0.0	244.8
	5mm	87938		16.8	16.0	52.4	19.8		106.1
	10mm	39117		15.8	6.7	21.2	7.9		52.8
	20mm	10411		16.4	2.8	6.2	2.4		29.1
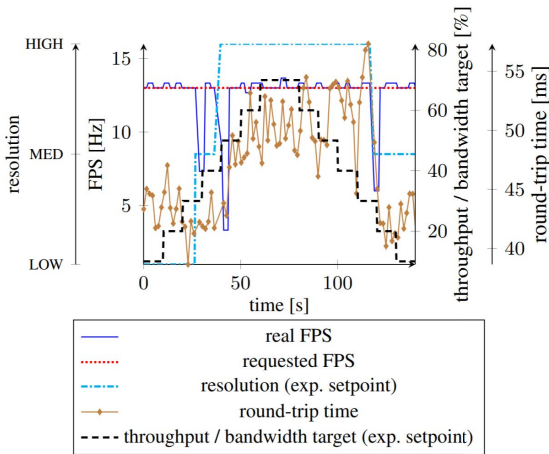
FIGURE 59. Video experiment with automatic settings (exp. n° 14 in Table 5). Resolution controlled by throughput to bandwidth ratio target as setpoint. Round-trip time, requested and real FPS behaviour presented. Connection type: VPN 4G modem in the LHC tunnel. Constant parameters: FPS = 13 Hz.

FIGURE 61. Video experiment with automatic settings (exp. n° 15 in Table 5). Resolution controlled by round-trip time target as setpoint. Throughput to bandwidth ratio, requested and real FPS behaviour presented. Connection type: VPN 4G modem in the LHC tunnel. Constant parameters: FPS = 8 Hz.
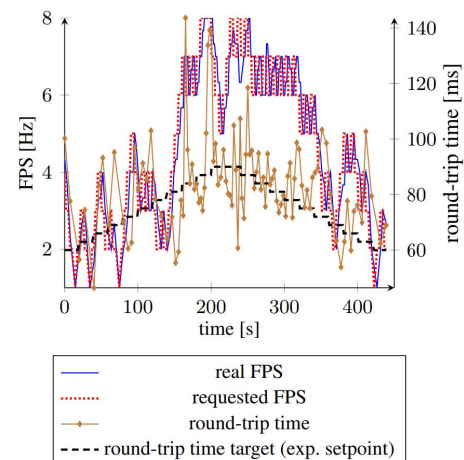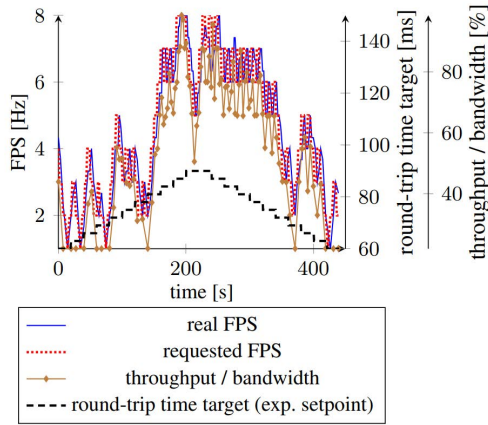

FIGURE 60. Video experiment with automatic settings (exp. n° 15 in Table 5). Resolution controlled by round-trip time target as setpoint. Round-trip time, requested and real FPS behaviour presented. Connection type: VPN 4G modem in the LHC tunnel. Constant parameters: FPS = 8 Hz.

FIGURE 62. Video experiment with automatic settings (exp. n° 16 in Table 5). Resolution controlled by FPS as setpoint. Throughput to bandwidth ratio, requested and real FPS behaviour presented. Connection type: VPN 4G modem in the LHC tunnel.

F. RELATION OF ROUND-TRIP TIME AND THROUGHPUT TO BANDWIDTH RATIO

The main difference between the bandwidth and round-trip time measurements is that the bandwidth measurement interrupts the teleoperation, taking some time and using the full bandwidth which sometimes might not be acceptable. The round-trip time measurement is done on-the-fly and it is not intrusive. However, when the camera settings are constant, the throughput-to-bandwidth values are more stable over time than the round-trip time values, which are characterized by a jitter. The jitter magnitude depends on the connection type, as presented in Table 4. As described in Section II-J, the presented round-trip time was calculated with the Equation 3, which is a paramaterizable filter. Therefore, it is possible to change the weights, which would increase/decrease the response time though the oscillations would be minimized/reinforced accordingly. In the presented graphs, the standard weights were used (Equation 4).
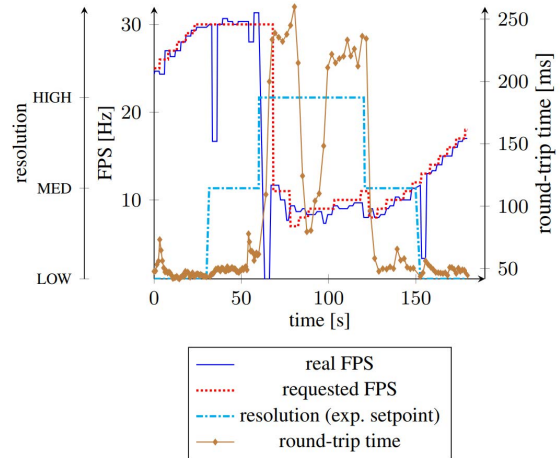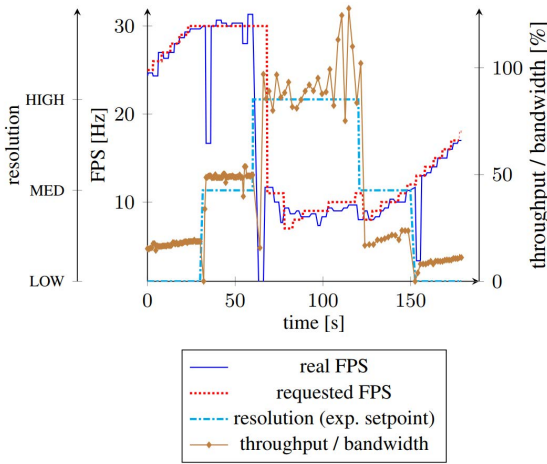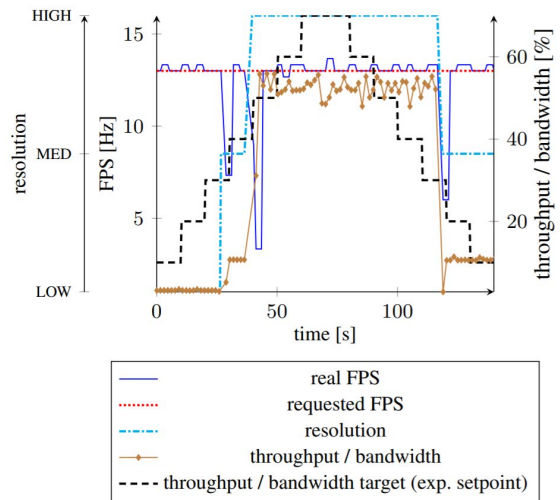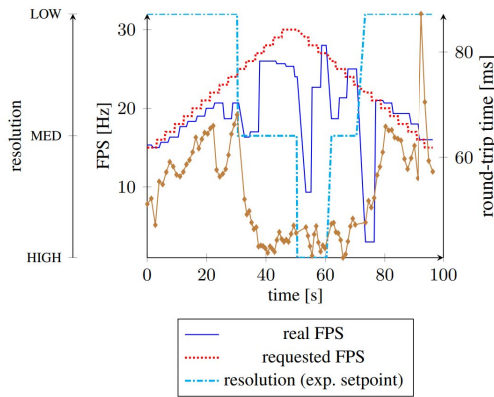
FIGURE 63. Video experiment with automatic settings (exp. n° 16 in Table 5). Resolution controlled by FPS as setpoint. Round-trip time, requested and real FPS behaviour presented. Connection type: VPN 4G modem in the LHC tunnel.
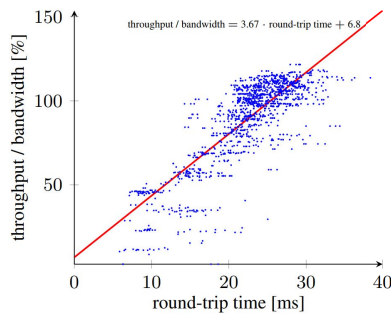
FIGURE 64. Point cloud experiment without automatic settings (exp. n° 1 in Table 5). Requested FPS as setpoint. Throughput to bandwidth ratio vs. round-trip time behaviour presented. Connection type: Wi-Fi over CERN GPN. Constant parameters: resolution = MEDIUM, subsampling = 10 mm.
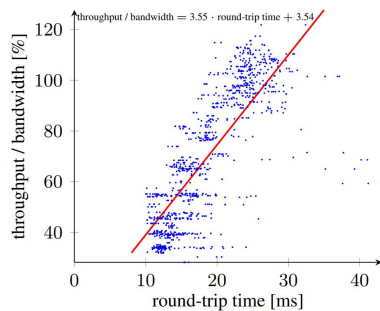
FIGURE 65. Point cloud experiment without automatic settings (exp. n° 2 in Table 5). Subsampling as setpoint. Throughput to bandwidth ratio vs. round-trip time behaviour presented. Connection type: Wi-Fi over CERN GPN. Constant parameters: resolution = MEDIUM, requested FPS = 7 Hz.

FIGURE 66. Point cloud experiment without automatic settings (exp. n° 1 in Table 5). Requested FPS as setpoint. Throughput to bandwidth ratio vs. round-trip time behaviour presented. Connection type: Ethernet over CERN GPN. Constant parameters: resolution = MEDIUM, subsampling = 3 mm.
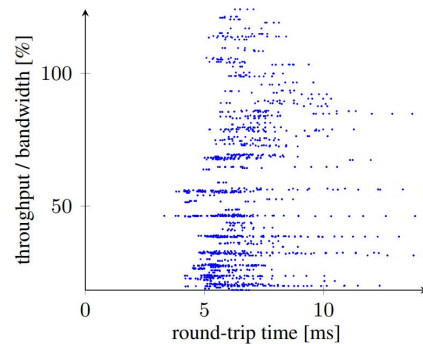
FIGURE 67. Point cloud experiment without automatic settings (exp. n° 2 in Table 5). Subsampling as setpoint. Throughput to bandwidth ratio vs. round-trip time behaviour presented. Connection type: Ethernet over CERN GPN. Constant parameters: resolution = MEDIUM, requested FPS = 15 Hz.
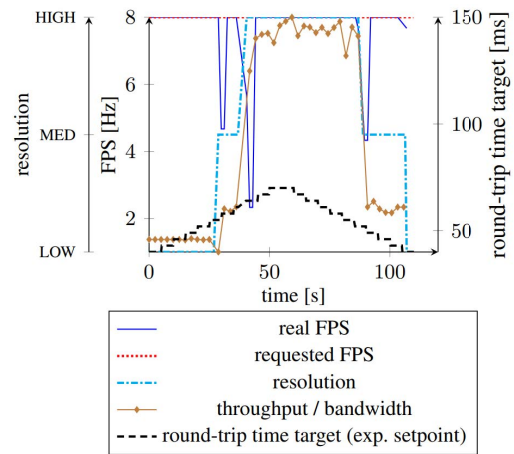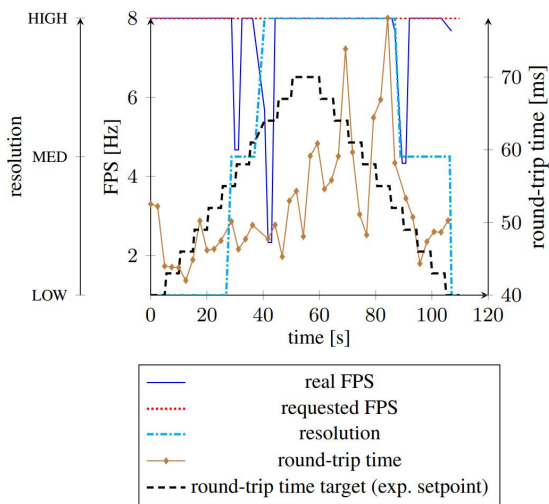

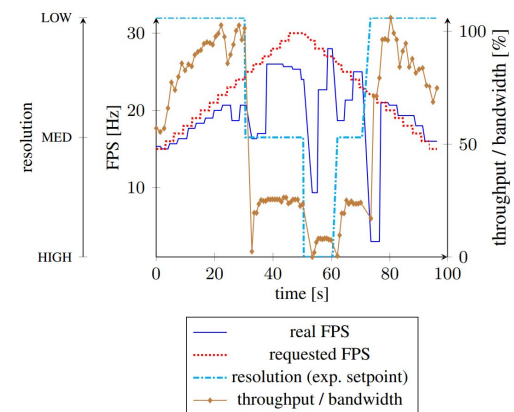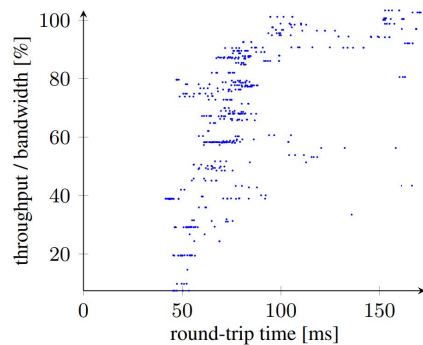**FIGURE 68.** Video experiment without automatic settings (exp. n° 3 in Table 5). Requested FPS as setpoint. Throughput to bandwidth ratio vs. round-trip time behaviour presented. Connection type: VPN 4G modem in the LHC tunnel. Constant parameters: resolution = HIGH.

During the experiments, the round-trip time was found to be correlated to the throughput to bandwidth ratio. The round-trip time generally increases when the throughput increases. Moreover, for certain situations, the values are proportional with a certain precision, indicating that the bandwidth measurement could be approximated by the round-trip time measurement. In Figures 64 and 65, the proportional correlations are sufficient. The used connection type was Wi-Fi and the point cloud acquisition was tested. The scattering depends on the variability of the network bandwidth and round-trip time jitter. However, as shown in Figures 66 and 67, the round-trip time changes with respect to the throughput to

bandwidth ratio might be relatively big and with a significant jitter. The connection type used here was the Ethernet over CERN GPN and the point cloud was acquired. For the VPN 4G modem in the LHC tunnel connection type, the correlation is visible (Figure 68), however when almost full bandwidth is used, the round-trip time started oscillating with large amplitudes. Therefore, in the last 3 examples, the correlation cannot be used practically in real-time for the approximation of throughput-to-bandwidth ratio by the round-trip time.

TABLE 8. The summary of automatic settings modes: their desired intervention scenario uses, problems that are encountered when the automatic settings are not used and the improvement or solutions to problems when the automatic settings are used.

Scenarios of use	Problems for teleoperation	Adaptive congestion control type	Improvement thanks to the the adaptive congestion control	Limitation
Poor network coverage with highly limited bandwidth and its variation over time or in nearby locations. E.g. traversing a tunnel with scarcely distributed access points.	Because of variation, the network might get overloaded unexpectedly which leads to temporarily losing connection or stopping video or point cloud streams. Full bandwidth use increases delay multiple times which also impacts command safety.	Settings controlled by throughput to bandwidth ratio	The system adapts to environment change, especially the different number of point cloud points, while keeping the network load at the same value. The streaming is not collapsed because of overload, and the round-trip time is more stable.	If the bandwidth decreases significantly in different location, it has to be measured again. However, if the robot is stationary it does not require frequent measurements.
Precise teleoperation with fast-enough feedback from the robot to avoid collisions and to allow delicate manipulation. A strict round-trip time limitation is required. E.g. bilateral, force feedback control; operating in real-time close to abstacles or dynamic environment.	The round-trip time depends on the used network and varying load, often depending on external and uncontrolled factors. Varying round-trip time undermines the predictability and confuses the operator.	Settings controlled by round-trip time	This mode is most useful when the round-trip time parameter is crucial. It adapts fast to network parameters change and does not require bandwidth measurement.	Fast reaction of the algorithm keeping the round-trip timat at a level can change subsampling, resolution or FPS by a significant factor.
Constant rate of feedback is required, which translates into responsiveness of the robot. E.g. supervising the robot when it is performing an autonomous task.	Overloading the network can lead to a total collapse of stream due to buffering. Responsivess of the robot is highly deteriorated.	Resolution or point cloud subsampling controlled by real FPS	The resolution or point cloud subsampling are automatically adjusted to achieve requested FPS, while maintaining maximum available resolution or point cloud size. The collapse of stream is avoided. This mode does not use any network measurement.	The settings are controlled so that the bandwidth is used in full, therefore if there are other robots in the vicinity, it may create problems.
The teleoperator requires specific precision of point cloud or resolution of video with compromised but maximum available framerate. E.g. avoiding collisions with small obstacles such as cables, which are only visible with high resolution video or 1-3mm point cloud spacing resolution.	Network bandwidth constraints might not allow high enough framerates, thus a buffering of frames and a collapse of the stream may follow.	Requested FPS controlled by point cloud subsampling or video resolution	The buffering and stream collapse is avoided, while maintaining the required feedback precision. This mode does not use any network measurement and adapts fast to network changes.	As above, the drawback when the bandwidth is used in full.

These observations and conclusions are also important for the use of the automatic settings modes. From experience, the operator might choose the round-trip time based modes in the networks which have more stable parameters and the presented correlations are approximately linear, in order to have more stable behaviour of the camera acquisitions. While using a network which has variable round-trip times, the preferred modes can be based on the throughput measurement or even only on the internal acquisition settings and the real FPS feedback.

G. RESULTS SUMMARY

Table 8 summarizes in which scenarios the automatic settings modes should be used, what problems they solve, and what the improvements are due to the application of the adaptive congestion control.

H. VIDEO DEMONSTRATIONS

Various videos can further extend the understanding of the interface used in these experiments. At [42], there is a video showing how the teleoperation has been performed with the CERN 2D GUI for an intervention, as a reference and as a comparison to the solution of the Mixed Reality Human-Robot Interface presented in this paper. Secondly, at [43],

there is a video showing the use of the interface in the same real scenarios executed by the same robot in the LHC. Lastly, at [44], there is a video showing the use of all the automatic settings modes by a teleoperator.

IV. DISCUSSION AND FUTURE WORK

The automatic settings provide an easier way of camera control for a non-expert operator who might not know exactly which parameter should be changed in case of, for example, network latency variation. Each control mode gives a different way of control, mitigating control effort. However, there is still room for simplification based on the control scenario that is used. For example, the scenario can be focused on precise manual manipulation, master-slave control, short or long distance between the robot and a target or operating in a very network-volatile environment. Thus, the setting would be rather a robotic scenario than a certain camera parameter that is controlled. The automatic settings selection can also depend on the interaction mode. In manual control, more FPS is needed; in a supervisory mode, a compromise between point cloud resolution and FPS is sufficient for collision avoidance and movement observation; in an autonomous mode the highest precision might be required.

The processing time of the point cloud (Section III-E) can change ∼8 times (from ∼245 ms to ∼29 ms) for the subsampling unit size increase from 1 mm to 20 mm. This delay has to be added to already affected network round-trip time by saturated bandwidth (Section III-D9, the round trip time can change from 50 ms to 250 ms when the throughput increases to the available maximum). In such a situation, the action of the automatic settings algorithm, based on network parameters, to increase the subsampling will have a positive impact on lowering the processing time. However, an automatic control based also on processing times measured on client and server sides may be added to the framework for delay-critical applications.

For further immersion increase using the Mixed Reality Human-Robot Interface for teleoperation in hostile environments, multiple advancements are proposed:

1) Apply the proposed framework to other operational scenarios, including those with very long and varying delays, such as the space telerobotics scenarios.

2) Study how the presence of multiple robots and/or other devices influences the available bandwidth while using the 4G network at CERN, particularly at different times of the day and what can cause a bottleneck.

3) In the higher level of telepresence with master-slave (bilateral) control and force feedback, the round-trip time is crucial. Therefore, a study using Augmented Reality with point cloud feedback should be done to verify if the proposition presented in this paper of automatic setting of subsampling and resolution based on round-trip time will be enough in the hazardous scenario.

4) Point cloud area streaming can be customized according to the current operator's viewpoint and look direction. Limiting which part of the cloud is sent according to the operators focus should be explored. Optionally, the point density can be increased in the observed area.

5) Explore point cloud stream density based on the distance of the operator's view to the point cloud area. As the rendering resolution is limited, saving on bandwidth can be achieved by lowering the density when the operator's position is far from point cloud.

6) The proposed collision detection and avoidance techniques greatly increase the operator's confidence and perception abilities to prevent collisions. However, for precise contact manipulations, a dedicated end effector force and torque sensor should be integrated.

V. CONCLUSION

The communication/network system is a key aspect of hazardous teleoperation scenarios where an operator requires a high amount of environmental awareness. An increased redundancy and complexity of the mobile manipulator can also increase the risk of collisions. These collisions can be avoided by the implemented Mixed Reality techniques, either by a spacial collision model and point cloud based avoidance, or detection of torque changes. However, these techniques

require a significant amount of point cloud data sent from the robot to the control interface, which can be much larger than that of video streams. This can greatly impact the round-trip time and used bandwidth, and is why the automatic setting framework has been proposed and implemented in the CERN Mixed Reality interface.

The point cloud parameters need to be adjusted continuously to control the congestion of the network and flow. The adaptive congestion control in the application level has been designed to optimize the available bandwidth or to fulfil the specialized needs of the teleoperator. In the presented application context, it is possible to adjust the video and point cloud settings. These methods can be combined and used on top of the different compression methods which have been widely explored and are continuously improved. The automatic setting modes also enable the user to switch to the interaction mode that is best suited to the task.

The interface can improve the operator's environmental awareness and achieve better collision avoidance by providing immersion in the scene and automatic mechanisms to prevent or stop the movement when a collision is detected. Two types of collisions were taken into account: arm self-collision or collision with the robot's environment, both executed in real-time. The first type was based on the knowledge of the exact robot's model. The second type was based on the information taken from 3D point cloud acquisition or joints torque feedback.

Full network information was made available to the operator to facilitate the assessment of the current network conditions. The higher network bandwidth requirements due to the camera acquisition were met with algorithms of real-time video and 3D point cloud settings adaptation to the network. This allowed the optimization of the communication load and fulfilled the teleoperation goal with regards to camera feedback; whether a higher precision was needed, or a maximum delay to encounter, or maximum bandwidth use to manage. Twelve automatic setting modes were proposed with algorithms based on FPS, resolution, point cloud subsampling, round-trip time or throughput to bandwidth ratio. Each mode was thoroughly characterized and tested to present its specific use-case, advantages and how it lowers the operator's workload.

A framework was presented according to which designers can optimize their Human-Robot Interfaces and environmental video or point cloud feedback depending on the network characteristics and intervention scenarios. Based on the experience learnt during the development and experiments of Mixed Reality techniques for the described application and constraints, multiple further advancements or research were described. The solutions presented in this paper can also be applied in many other fields where network parameters vary and any Mixed Reality control using 3D environmental feedback is needed. These fields can be automotive environment sensing or remote driving, teleoperation in space or underwater, or terrestrial teleoperation over large distances.

APPENDIX A HUMAN-ROBOT INTERFACES – STATE-OF-THE-ART

TABLE A.1. Non-exhaustive state-of-the-art of Human-Robot interfaces in terms of hazardous environments, 2D and 3D interfaces, if it is operational, communication type, complexity of manipulator, level of interaction, collision avoidance or detection, adaptive communication for shared and dynamic networks, and references.

Interface product	Hazardous environment	2D interfaces	3D user interface	Operational	Communication	Complexity of manipulator	Level of interaction	Collisions avoidance or detection	Adaptive communication for shared or dynamic network	References
Telerob tEODor, Telemax	Military applications, difficult terrain, used at CERN in radioactive zones	Video stream	X	Yes	Long-range radiofrequency	Telemax: 7 degrees of freedom	Manual teleoperation	X	X	[1], [2]
Fukushima robots: Monorobo, RESQ, SMERT, MARS	Nuclear plant radioactively contaminated areas	Video stream	X	Yes	Cabled	Up to 6 degrees of freedom	Manual teleoperation	X	X	[3]
Rescue robot Quince mobile redesigned for Fukushima missions	Nuclear plant radioactively contaminated areas	Video stream	3D laser scan view	Yes	Cabled, inter-robot wireless 2.4 GHz	2 degrees of freedom	Manual teleoperation	X	X	[4]
Mascot 4	Fusion energy tokamak JET	Video stream	Precisely modelled and simulated environment	Yes	Cabled	7 degrees of freedom of each manipulator	Manual operation with force feedback	Force feedback	X	[5]
TWINBOT	Underwater	Video stream in 2D interface	3D modelled robot and environment with 2D real-time perception	Yes, in lab scenario	Radiofrequency and sonar	4 degrees of freedom	Teleoperation, high-level supervision and autonomous	X	X	[6], [7], [8], [9]
Perserverance rover	Mars	Video stream	Spatial mapping of the environment	Yes	Communication with through Mars orbiters, UHF and X-band	5 degrees of freedom manipulator	Supervised and autonomous control	Manipulator-robot collisions avoidance in simulation before execution	The Mars Relay Operations Service for communication control	[10], [11]
CERN 2D GUI	Underground particle accelerator	Video stream	X	Yes	2G, 3G, 4G network, Wi-Fi	Up to 9 degrees of freedom	Manual teleoperation	Torque information display	X	[12], [13]
CERN 3D GUI	Underground particle accelerator	Video stream shown in 3D interface	Fully modelled robot controlled in real tim in 3D space with 3D point cloud feedback	Yes	2G, 3G, 4G network, Wi-Fi	9 degrees of freedom	Real-time teleoperation, supervised position and trajectory control, target-based task specification	Collisions with robot, point cloud and self-collision avoidance in planning and real-time, additionally real-time collisions detection with torques	Automatic Congestion Control of camera feedback	Pilot project [13]

APPENDIX B
ROBOTIC INTERVENTION USE CASES

A standard teleoperation sequence using a Mixed Reality Interface during the Beam Loss Monitor measurement intervention scenario in the LHC consists of the following phases:

1) Before setting the camera(s) control parameters and moving the manipulator, the maximum bandwidth parameter must be tested to enable automatic control mode by throughput to bandwidth ratio (more in Section II-J). The bandwidth measurements can be repeated at any time on demand.

2) Setting up the camera(s) control parameters (Section II-I1) to achieve the desired point cloud or video resolution, frame rate, or automatic behaviour (Section II-I4). The position or transparency of the video feedback can be changed if it is obstructing the view.

3) Planning the manipulator's trajectory from the current real position to the desired position. The operator moves the planning manipulator using a convenient planning control mode (explained in Section II-D) and saves one or more waypoints. During planning, the operator will be notified if a self- or point cloud collision happens.

4) Before executing the movement to a selected waypoint, a collision check and preview are done to check if any self- or point cloud collisions occur (Section II-F).

5) The operator decide to execute the movement, acts on a control button and observes the environment with real-time point cloud and video feedback. The movement can be quickly interrupted by releasing the button.

6) Collision check and execution of movement are repeated for each waypoint. At any time the waypoint can be modified or removed.

7) The manipulator can be also controlled directly without planning (real-time direct control, Section II-D). The direct control allows adjusting a single joint or use the robot's inverse kinematics.

8) The final approach to the BLM device can also be accelerated by using the Target Approach feature (Section II-G) that allows any point cloud point to be selected, which creates an object which is normal to the surface and which can be adjusted using a gizmo (rotation or distance). Then, in the planning control mode the end effector can be moved exactly to this point with a single click.

9) At any time, there is an indicator showing a distance from the end effector of real, planning, preview or waypoint manipulators to any object recorded with point cloud in the scene. This helps the operator to predict any potential collisions or plan the movement precisely.

The use case diagram is shown in Figure B.1, which links graphically the phases described above.
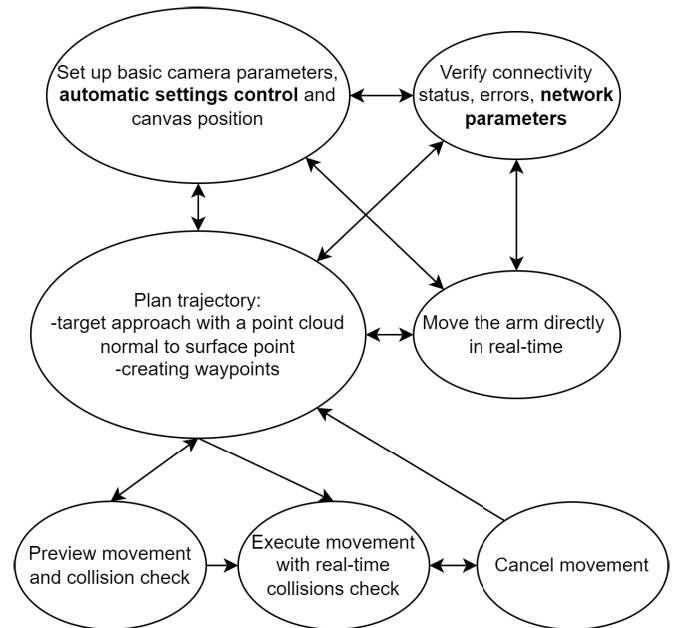
FIGURE B.1. Robotic intervention use cases with the Mixed Reality Interface. The parts taking part in the automatic end-to-end congestion control protocol for the video and point cloud feedback are highlighted.

APPENDIX C
AUTOMATIC CAMERA SETTING ALGORITHMS AND HOLOGRAM CONTROL VIEWS
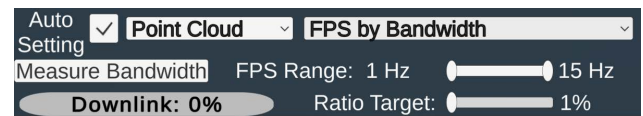
FIGURE C.1. Hologram automatic setting for point cloud FPS controlled by throughput to bandwidth ratio mode.
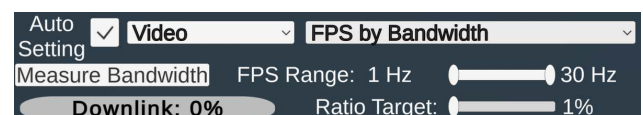
FIGURE C.2. Hologram automatic setting for video FPS controlled by throughput to bandwidth ratio mode.
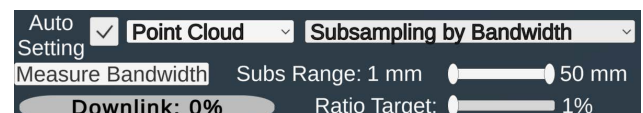
FIGURE C.3. Hologram automatic setting for point cloud subsampling unit size controlled by throughput to bandwidth ratio mode.
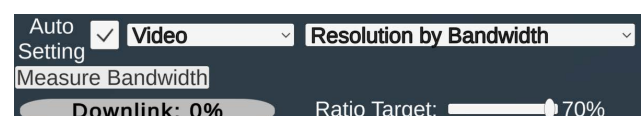
FIGURE C.4. Hologram automatic setting for video resolution controlled by throughput to bandwidth ratio mode.
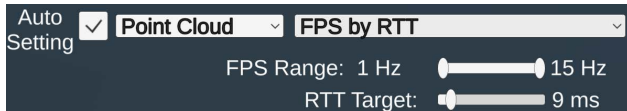
FIGURE C.5. Hologram automatic setting for point cloud FPS controlled by round-trip time mode.
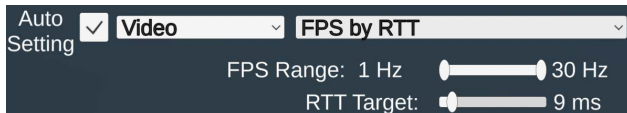
FIGURE C.6. Hologram automatic setting for video FPS controlled by round-trip time mode.
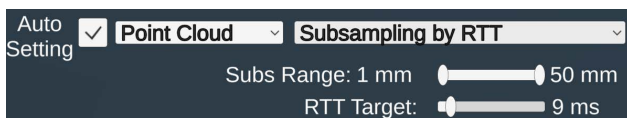
FIGURE C.7. Hologram automatic setting for point cloud subsampling unit size controlled by round-trip time mode.
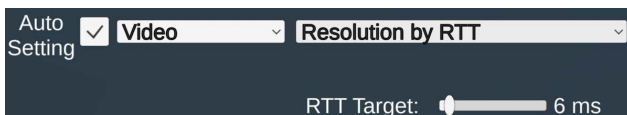
FIGURE C.8. Hologram automatic setting for video resolution controlled by round-trip time mode.
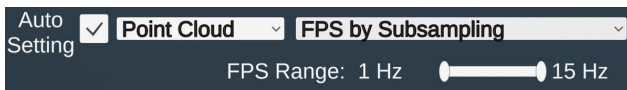
FIGURE C.9. Hologram automatic setting for point cloud FPS controlled by subsampling unit size mode.
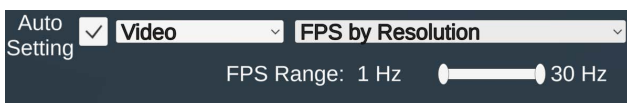
FIGURE C.10. Hologram automatic setting for video FPS controlled by resolution mode.
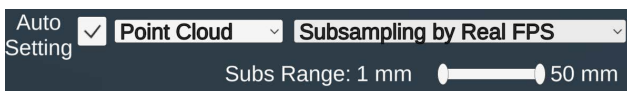
FIGURE C.11. Hologram automatic setting for point cloud subsampling unit size controlled by FPS mode.

FIGURE C.12. Hologram automatic setting for resolution controlled by FPS mode.

Algorithm 1 Algorithm of Automatic Point Cloud or Video FPS Setting Controlled by Throughput to Bandwidth Ratio

Inputs: Throughput: downlink throughput [Mbps],
   Bandwidth: downlink bandwidth [Mbps],
   RatioTarget: downlink throughput to downlink bandwidth ratio target,
   ReqFPS: requested FPS [Hz],
   MinFPS: minimum FPS limit [Hz],
   MaxFPS: maximum FPS limit [Hz].
Output: NewFPS: new requested FPS [Hz].
Temp: CalcRatio: calculated throughput to bandwidth ratio.
Constants: DECR = 1 Hz: FPS decrement value,
   INCR = 1 Hz: FPS increment value.
 1: CalcRatio = Throughput / Bandwidth
 2: **if** CalcRatio > RatioTarget **then**
 3:   **if** ReqFPS + INCR ≤ MaxFPS **then**
 4:     **if** ReqFPS + INCR ≤ MinFPS **then**
 5:       NewFPS = MinFPS
 6:     **else**
 7:       NewFPS = ReqFPS + INCR
 8:     **end if**
 9:   **else**
10:     NewFPS = MaxFPS
11:   **end if**
12: **else if** ReqFPS − DECR ≥ MinFPS **then**
13:   **if** ReqFPS − DECR ≥ MaxFPS **then**
14:     NewFPS = MaxFPS
15:   **else**
16:     NewFPS = ReqFPS − DECR
17:   **end if**
18: **else**
19:   NewFPS = MinFPS
20: **end if**

Algorithm 2 Algorithm of Automatic Settings Subsampling Unit Size Controlled by Throughput to Bandwidth Ratio

Inputs: Throughput: downlink throughput [Mbps],
   Bandwidth: downlink bandwidth [Mbps],
   RatioTarget: downlink throughput to downlink bandwidth ratio target.
   CurrSubs: current subsampling unit size [mm],
   MinSubs: minimum subsampling limit [mm],
   MaxSubs: maximum subsampling limit [mm].
Output: NewSubs: new subsampling unit size [mm].
Temp: CalcRatio: calculated throughput to bandwidth ratio.
Constants: DECR = 1 mm: subsampling decrement value.
   INCR = 1 mm: subsampling increment value.
 1: CalcRatio = Throughput / Bandwidth
 2: **if** CalcRatio > RatioTarget **then**
 3:   **if** CurrSubs + INCR ≤ MaxSubs **then**
 4:     **if** CurrSubs + INCR ≤ MinSubs **then**
 5:       NewSubs = MinSubs
 6:     **else**
 7:       NewSubs = CurrSubs + INCR
 8:     **end if**
 9:   **else**
10:     NewSubs = MaxSubs
11:   **end if**
12: **else if** CurrSubs − DECR ≥ MinSubs **then**
13:   **if** CurrSubs − DECR ≥ MaxSubs **then**
14:     NewSubs = MaxSubs
15:   **else**
16:     NewSubs = CurrSubs − DECR
17:   **end if**
18: **else**
19:   NewSubs = MinSubs
20: **end if**

Algorithm 3 Algorithm of Automatic Settings Video Resolution Controlled by Throughput to Bandwidth Ratio

---

Inputs: Throughput: downlink throughput [Mbps],
Bandwidth: downlink bandwidth [Mbps],
RatioTarget: downlink throughput to downlink bandwidth ratio target.
CurrResolution: current video resolution,
Output: NewResolution: new video resolution.
Constants: HIGH: high video resolution.
MEDIUM: medium video resolution.
LOW: low video resolution.
DIFF = 20%: difference between target and current throughput to bandwidth ratio.
1: CalcRatio = Throughput / Bandwidth
2: **if** CalcRatio − DIFF > RatioTarget **then**
3:   **if** CurrResolution == HIGH **then**
4:     NewResolution = MEDIUM
5:   **else if** CurrResolution == MEDIUM **then**
6:     NewResolution = LOW
7:   **else if** CurrResolution ≠ LOW **then**
8:     NewResolution = LOW
9:   **end if**
10: **else if** CalcRatio + DIFF < RatioTarget **then**
11:   **if** CurrResolution == LOW **then**
12:     NewResolution = MEDIUM
13:   **else if** CurrResolution == MEDIUM **then**
14:     NewResolution = HIGH
15:   **else if** CurrResolution ≠ HIGH **then**
16:     NewResolution = HIGH
17:   **end if**
18: **end if**

Algorithm 4 Algorithm of Automatic Point Cloud or Video FPS Setting Controlled by Round-Trip Time

---

Inputs: CurrRTT: current round-trip time [ms],
RTTTarget: round-trip time target [ms].
ReqFPS: requested FPS [Hz],
MinFPS: minimum FPS limit [Hz],
MaxFPS: maximum FPS limit [Hz].
Output: NewFPS: new requested FPS [Hz].
Constants: DECR = 1 Hz: FPS decrement value.
INCR = 1 Hz: FPS increment value.
1: **if** CurrRTT > RTTTarget **then**
2:   **if** ReqFPS + INCR ≤ MaxFPS **then**
3:     **if** ReqFPS + INCR ≤ MinFPS **then**
4:       NewFPS = MinFPS
5:     **else**
6:       NewFPS = ReqFPS + INCR
7:     **end if**
8:   **else**
9:     NewFPS = MaxFPS
10:   **end if**
11: **else if** ReqFPS − DECR ≥ MinFPS **then**
12:   **if** ReqFPS − DECR ≥ MaxFPS **then**
13:     NewFPS = MaxFPS
14:   **else**
15:     NewFPS = ReqFPS − DECR
16:   **end if**
17: **else**
18:   NewFPS = MinFPS
19: **end if**

Algorithm 5 Algorithm of Automatic Settings Subsampling Unit Size Controlled by Round-Trip Time

---

Inputs: CurrRTT: current round-trip time [ms],
RTTTarget: round-trip time target [ms].
CurrSubs: current subsampling unit size [mm],
MinSubs: minimum subsampling limit [mm],
MaxSubs: maximum subsampling limit [mm].
Output: NewSubs: new subsampling unit size [mm].
Constants: DECR = 1 mm: subsampling decrement value,
INCR = 1 mm: subsampling increment value.
1: **if** CurrRTT > RTTTarget **then**
2:   **if** CurrSubs + INCR ≤ MaxSubs **then**
3:     **if** CurrSubs + INCR ≤ MinSubs **then**
4:       NewSubs = MinSubs
5:     **else**
6:       NewSubs = CurrSubs + INCR
7:     **end if**
8:   **else**
9:     NewSubs = MaxSubs
10:   **end if**
11: **else if** CurrSubs − DECR ≥ MinSubs **then**
12:   **if** CurrSubs − DECR ≥ MaxSubs **then**
13:     NewSubs = MaxSubs
14:   **else**
15:     NewSubs = CurrSubs − DECR
16:   **end if**
17: **else**
18:   NewSubs = MinSubs
19: **end if**

Algorithm 6 Automatic Settings Video Resolution Controlled by Round-Trip Time Algorithm

---

Inputs: CurrRTT: current round-trip time [ms],
RTTTarget: round-trip time target [ms].
CurrResolution: current video resolution,
Output: NewResolution: new video resolution.
Constants: HIGH: high video resolution.
MEDIUM: medium video resolution.
LOW: low video resolution.
DIFF = 3 ms: difference between target and current round-trip time.
1: **if** CurrRTT − DIFF > RTTTarget **then**
2:   **if** CurrResolution == HIGH **then**
3:     NewResolution = MEDIUM
4:   **else if** CurrResolution == MEDIUM **then**
5:     NewResolution = LOW
6:   **else if** CurrResolution ≠ LOW **then**
7:     NewResolution = LOW
8:   **end if**
9: **else if** CurrRTT + DIFF < RTTTarget **then**
10:   **if** CurrResolution == LOW **then**
11:     NewResolution = MEDIUM
12:   **else if** CurrResolution == MEDIUM **then**
13:     NewResolution = HIGH
14:   **else if** CurrResolution ≠ HIGH **then**
15:     NewResolution = HIGH
16:   **end if**
17: **end if**

Algorithm 7 Algorithm of Automatic Point Cloud FPS Setting Controlled by Subsampling Unit Size or Video FPS Setting Controlled by Resolution

Inputs: ReqFPS: requested FPS [Hz],
    RealFPS: real FPS [Hz],
    MinFPS: minimum FPS limit [Hz],
    MaxFPS: maximum FPS limit [Hz].
Output: NewFPS: new requested FPS [Hz].
Temp: CalcFPS: calculated new requested FPS [Hz].
Constants: DIFF = 2 Hz: difference between new and current FPS.
    INCR = 1 Hz: FPS increment value.
 1: **if** ReqFPS − RealFPS ≥ DIFF **then**
 2:     **if** RealFPS ≥ MaxFPS **then**
 3:         NewFPS = MaxFPS
 4:     **else if** RealFPS ≤ MinFPS **then**
 5:         NewFPS = MinFPS
 6:     **else**
 7:         NewFPS = RealFPS
 8:     **end if**
 9: **else if** ReqFPS − RealFPS ≤ 0 **then**
10:     **if** ReqFPS + INCR ≤ MaxFPS **then**
11:         **if** ReqFPS + INCR ≤ MinFPS **then**
12:             NewFPS = MinFPS
13:         **else**
14:             NewFPS = ReqFPS + INCR
15:         **end if**
16:     **else**
17:         NewFPS = MaxFPS
18:     **end if**
19: **end if**

Algorithm 8 Algorithm of Automatic Settings Point Cloud Subsampling Unit Size Controlled by FPS

Inputs: ReqFPS: requested point cloud FPS [Hz],
    RealFPS: real point cloud FPS [Hz],
    CurrSubs: current subsampling unit size [mm],
    MinFPS: minimum FPS limit [Hz],
    MaxFPS: maximum FPS limit [Hz].
Output: NewSubs: new subsampling unit size [mm].
Temp: CalcSubs: calculated new subsampling unit size [mm].
Constants: DIFF = 2 Hz: difference between new and current FPS.
    DECR = 1 mm: subsampling decrement value.
 1: **if** ReqFPS − RealFPS > DIFF **then**
 2:     CalcSubs = CurrSubs · (1 + 0.1 · (ReqFPS − RealFPS) / RealFPS)
 3:     **if** CalcSubs ≤ MaxFPS **then**
 4:         **if** CalcSubs ≤ MinFPS **then**
 5:             NewSubs = MinFPS
 6:         **else**
 7:             NewSubs = CalcSubs
 8:         **end if**
 9:     **else**
10:         NewSubs = MaxFPS
11:     **end if**
12: **else if** ReqFPS − RealFPS ≤ 0 **then**
13:     CalcSubs = CurrSubs - DECR
14:     **if** CalcSubs ≥ MinFPS **then**
15:         **if** CalcSubs ≥ MaxFPS **then**
16:             NewSubs = MaxFPS
17:         **else**
18:             NewSubs = CurrSubs − DECR
19:         **end if**
20:     **else**
21:         NewSubs = MinFPS
22:     **end if**
23: **end if**

Algorithm 9 Automatic Settings Video Resolution Controlled by Video FPS Algorithm

Inputs: ReqFPS: requested point cloud FPS [Hz],
    RealFPS: real point cloud FPS [Hz],
    CurrResolution: current video resolution,
Output: NewResolution: new video resolution.
Constants: HIGH: high video resolution.
    MEDIUM: medium video resolution.
    LOW: low video resolution.
    DIFF = 4: difference between real and requested FPS.
 1: **if** ReqFPS − RealFPS > DIFF **then**
 2:     **if** CurrResolution == HIGH **then**
 3:         NewResolution = MEDIUM
 4:     **else if** CurrResolution == MEDIUM **then**
 5:         NewResolution = LOW
 6:     **else if** CurrResolution ≠ LOW **then**
 7:         NewResolution = LOW
 8:     **end if**
 9: **else if** ReqFPS − RealFPS ≤ 0 **then**
10:     **if** CurrResolution == LOW **then**
11:         NewResolution = MEDIUM
12:     **else if** CurrResolution == MEDIUM **then**
13:         NewResolution = HIGH
14:     **else if** CurrResolution ≠ HIGH **then**
15:         NewResolution = HIGH
16:     **end if**
17: **end if**

REFERENCES

[1] J. Carff, M. Johnson, E. M. El-Sheikh, and J. E. Pratt, "Human-robot team navigation in visually complex environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2009, pp. 3043–3050.

[2] T. Sakaue, S. Yoshino, K. Nishizawa, and K. Takeda, "Survey in Fukushima daiichi NPS by combination of human and remotely-controlled robot," in *Proc. IEEE Int. Symp. Saf., Secur. Rescue Robot. (SSRR)*, Oct. 2017, pp. 7–12.

[3] S. Kawatsuma, R. Mimura, and H. Asama, "Unitization for portability of emergency response surveillance robot system: Experiences and lessons learned from the deployment of the JAEA-3 emergency response robot at the Fukushima Daiichi nuclear power plants," *ROBOMECH J.*, vol. 4, no. 1, pp. 1–7, Dec. 2017, doi: 10.1186/s40648-017-0073-7.

[4] M. Bajracharya, M. W. Maimone, and D. Helmick, "Autonomy for Mars rovers: Past, present, and future," *Computer*, vol. 41, no. 12, pp. 44–50, Dec. 2008.

[5] M. Panzirsch, H. Singh, T. Kruger, C. Ott, and A. Albu-Schaffer, "Safe interactions and kinesthetic feedback in high performance earth-to-moon teleoperation," in *Proc. IEEE Aerosp. Conf.*, Mar. 2020, pp. 1–10.

[6] R. Buckingham and A. Loving, "Remote-handling challenges in fusion research and beyond," *Nature Phys.*, vol. 12, no. 5, pp. 391–393, May 2016, doi: 10.1038/nphys3755.

[7] C. P. Sesmero, S. V. 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. [Online]. Available: https://www.mdpi.com/1424-8220/21/16/5340

[8] M. Di Castro, G. Lunghi, A. Masi, M. Ferre, and R. M. Prades, "A multi-dimensional RSSI based framework for autonomous relay robots in harsh environments," in *Proc. 3rd IEEE Int. Conf. Robotic Comput. (IRC)*, Feb. 2019, pp. 183–188.

[9] 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

[10] K. Kanamori, N. Sakata, T. Tominaga, Y. Hijikata, K. Harada, and K. Kiyokawa, "Obstacle avoidance method in real space for virtual reality immersion," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Oct. 2018, pp. 80–89.

[11] A. Martín-Barrio, J. J. Roldán-Gómez, I. Rodríguez, J. del Cerro, and A. Barrientos, "Design of a hyper-redundant robot and teleoperation using mixed reality for inspection tasks," *Sensors*, vol. 20, no. 8, p. 2181, Apr. 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/8/2181

[12] A. Aristidou and J. Lasenby, "FABRIK: A fast, iterative solver for the Inverse Kinematics problem," *Graph. Models*, vol. 73, no. 5, pp. 243–260, Sep. 2011. 10.1016/j.gmod.2011.05.003

[13] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE Trans. Inf. Syst.*, vol. 77, no. 12, pp. 1321–1329, Dec. 1994.

[14] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, "Augmented reality: A class of displays on the reality-virtuality continuum," *Proc. SPIE*, vol. 2351, pp. 282–292, Dec. 1994.

[15] P. Milgram and H. W. Colquhoun, "A framework for relating head-mounted displays to mixed reality displays," *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, vol. 43, no. 22, pp. 1177–1181, Sep. 1999, doi: 10.1177/154193129904302202.

[16] R. Wirz, R. Marín, J. M. Claver, M. Ferre, R. Aracil, and J. Fernández, "End-to-end congestion control protocols for remote programming of robots, using heterogeneous networks: A comparative analysis," *Robot. Auto. Syst.*, vol. 56, no. 10, pp. 865–874, Oct. 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0921889008000845

[17] P. Gorczak, C. Bektas, F. Kurtz, T. Lubcke, and C. Wietfeld, "Robust cellular communications for unmanned aerial vehicles in maritime search and rescue," in *Proc. IEEE Int. Symp. Saf., Secur., Rescue Robot. (SSRR)*, Sep. 2019, pp. 229–234.

[18] D. Laniewski and N. Aschenbruck, "On the potential of rate adaptive point cloud streaming on the point level," in *Proc. IEEE 46th Conf. Local Comput. Netw. (LCN)*, Oct. 2021, pp. 49–56.

[19] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuća, S. Lasserre, Z. Li, J. Llach, K. Mammou, R. Mekuria, O. Nakagami, E. Siahaan, A. Tabatabai, A. M. Tourapis, and V. Zakharchenko, "Emerging MPEG standards for point cloud compression," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 9, no. 1, pp. 133–148, Mar. 2019.

[20] T. Huang and Y. Liu, "3D point cloud geometry compression on deep learning," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 890–898.

[21] M. Quach, G. Valenzise, and F. Dufaux, "Learning convolutional transforms for lossy point cloud geometry compression," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 4320–4324.

[22] J. Kammerl, N. Blodow, R. B. Rusu, S. Gedikli, M. Beetz, and E. Steinbach, "Real-time compression of point cloud streams," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2012, pp. 778–785.

[23] M. Hosseini and C. Timmerer, "Dynamic adaptive point cloud streaming," in *Proc. 23rd Packet Video Workshop*, Jun. 2018, pp. 25–30, doi: 10.1145/3210424.3210429.

[24] C. Moreno, Y. Chen, and M. Li, "A dynamic compression technique for streaming Kinect-based point cloud data," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Jan. 2018, pp. 550–555.

[25] C. Moreno and M. Li, "A progressive transmission technique for the streaming of point cloud data using the Kinect," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Mar. 2018, pp. 593–598.

[26] A. Akhtar, B. Kathariya, and Z. Li, "Low latency scalable point cloud communication," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 2369–2373.

[27] M. Di 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 Experim. Phys. Control Syst.*, 2018, pp. 1–5.

[28] M. Di 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," in *Proc. IEEE Int. Symp. Saf., Secur., Rescue Robot. (SSRR)*, Aug. 2018, pp. 1–6.

[29] J. M. Garcés, C. V. Almagro, G. Lunghi, M. Di Castro, L. R. Buonocore, R. M. 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

[30] 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.

[31] M. Di Castro, M. Ferre Pérez, and A. Masi, "A novel robotic framework for safe inspection and telemanipulation in hazardous and unstructured environments," Ph.D. dissertation, Dept. Automat., Elect. Electron. Eng. Comput. Sci., Universidad Politécnica de Madrid, Madrid, Spain, 2019. [Online]. Available: http://cds.cern.ch/record/2708835

[32] 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.

[33] C. V. Almagro, G. Lunghi, M. Di 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

[34] D. Ribas, P. Ridao, A. Turetta, C. Melchiorri, G. Palli, J. J. Fernández, and P. J. Sanz, "I-AUV mechatronics integration for the TRIDENT FP7 project," *IEEE/ASME Trans. Mechatronics*, vol. 20, no. 5, pp. 2583–2592, Oct. 2015.

[35] A. Solis, R. Marin, J. Marina, F. J. Moreno, M. Ávila, M. D. L. 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*, San Diego, CA, USA, Sep. 2021, pp. 1–6.

[36] C. V. Almagro, M. Di 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, no. 14, pp. 1–28, 2019.

[37] K. Szczurek, R. Prades, E. Matheson, H. Perier, L. Buonocore, and M. Di Castro, "From 2D to 3D mixed reality human-robot interface in hazardous robotic interventions with the use of redundant mobile manipulator," in *Proc. 18th Int. Conf. Informat. Control, Autom. Robot.*, 2021, pp. 388–395.

[38] R. Schmidt, R. Assmann, E. Carlier, B. Dehning, R. Denz, B. Goddard, E. B. Holzer, V. Kain, B. Puccio, B. Todd, and J. Uythoven, "Protection of the CERN large hadron collider," *New J. Phys.*, vol. 8, p. 290, 2006.

[39] J. F. Kurose and K. W. Ross, *Computer Networking: A Top-Down Approach*, 7th ed. Boston, MA, USA: Pearson, 2016.

[40] S. Agosta, R. Sierra, and F. Chapron, "High-speed mobile communications in hostile environments," *J. Phys. Conf.*, vol. 664, no. 5, p. 52001, 2015, doi: 10.1088/1742-6596/664/5/052001.

[41] T. Sonoda, J. N. Sweetser, T. Khuong, S. Brook, and A. G.-J. Rev. *High-Speed Capture Mode of Intel RealSenseT Depth Camera D435*. Accessed: Jun. 28, 2022. https://dev.intelrealsense.com/docs/high-speed-capture-mode-of-intel-realsense-depth-camera-d435

[42] (2022). CERN Robotics Team BE-CEM-MRO. *Video Demonstration: Beam Loss Monitors Robotic Measurements With 2D GUI*. [Online]. Available: https://videos.cern.ch/record/2295957

[43] K. A. Szczurek. (2022). *Video Demonstration: Summary of Mixed Reality Human-Robot Interface With Adaptive Communications Congestion Control for the Teleoperation of Mobile Redundant Manipulators in Hazardous Environments*. [Online]. Available: https://videos.cern.ch/record/2295954

[44] K. Szczurek. (2022). *Video demonstration: Adaptive Communications Congestion Control for Mixed Reality Human-Robot Interface*. [Online]. Available: https://videos.cern.ch/record/2295956

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 the CERN, working on automation, control software and robotics projects. He is passionate about mixed reality human-robot interfaces, and specialized in design and implementation of 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 working as an Associate Professor at the Department of Computer Science, Jaume I University of Castellon, where he lectures computer networking, and robotics. He has been appointed as a Visiting Scientist at Blaise Pascal University, in 2002, the University Polithecnique of Madrid, in 2007, the University Federal of Brasilia, in 2016, and European Organization for Nuclear Research (CERN) (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 and 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 and 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 of 2020, he started the Automatic and Robotic Master with the Politechnic University of Madrid, which he is still finishing. He realized an internship with the Vision & Aerial Robotics Group, Center of Automatics and Robotics, Madrid. There he implemented a unity UAV simulator into a ROS framework called Aerostack. He also worked on the implementation of an 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.

MARIO DI CASTRO received 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 the 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 the 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.

• • •

Chapter 3

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

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 (Turtle-Bot)	-	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:
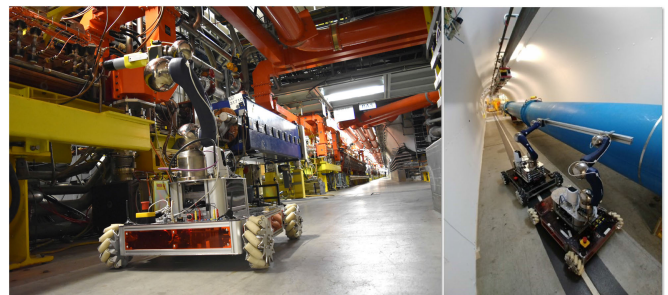
FIGURE 1. The figure shows two CERN robotic intervention scenarios in the Large hadron Collider (LHC) tunnel. On the left, the remotely controlled CERNBot performed an inspection of equipment in the radioactive environment. It was equipped with a pan-tilt-zoom camera and a manipulator with a radiation 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 processing	SLAM	Simu-lation	Collision avoidance	Object recognition	Object pose estimation	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 CERN-Bots (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
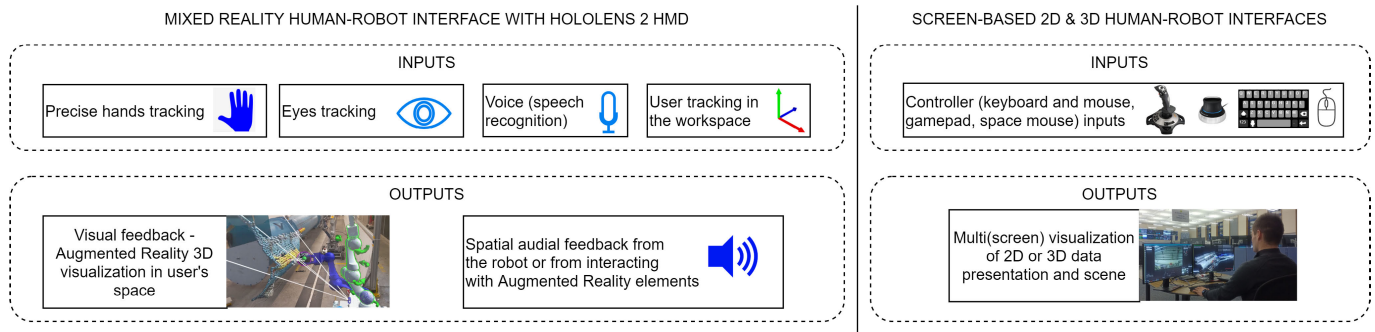
FIGURE 3. The figure presents the difference in inputs and outputs between an interface using the AR HMD and a screen-based 2D or 3D interface with standard controllers.

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
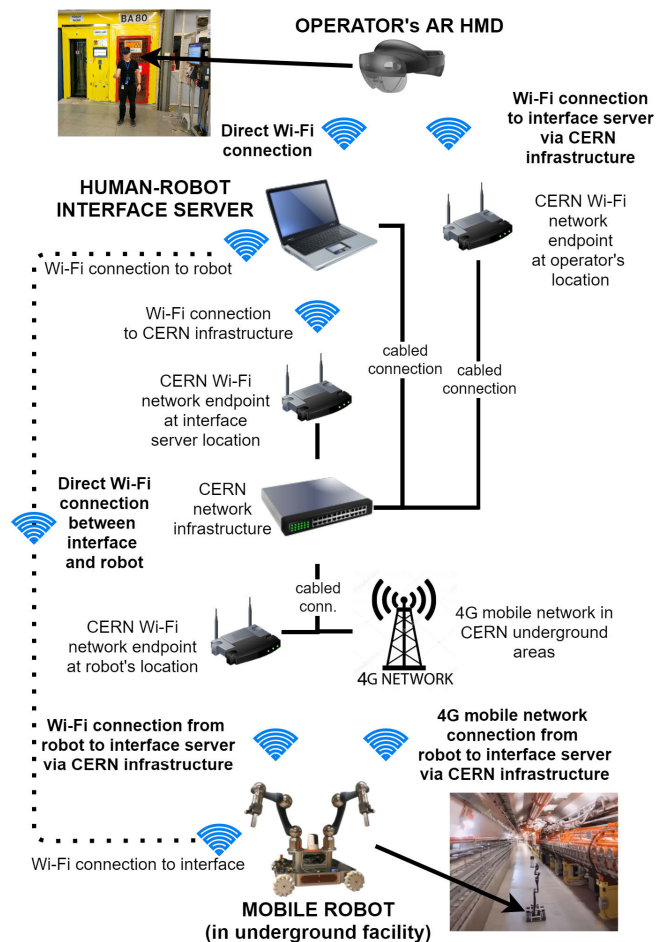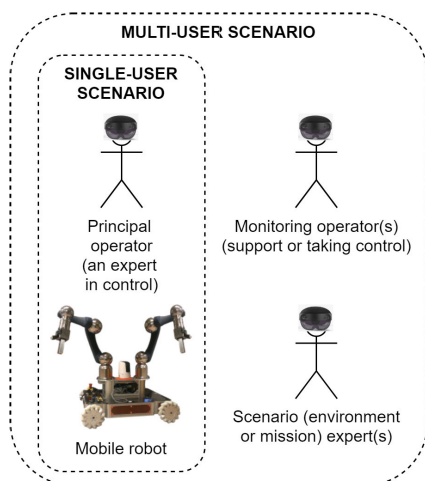
visualized precisely at the same place for all users (e.g. at a round table in a control centre conference room in Figure 7).
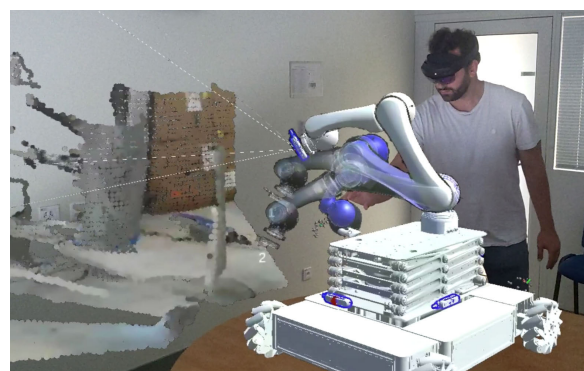
TABLE 5. Connection types used in robotic interventions at CERN, their network capabilities and limitations, use-cases, advantages, disadvantages, and the required infrastructure.

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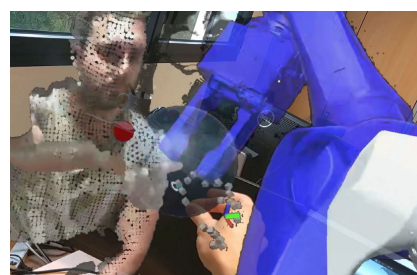
FIGURE 8. Remote workspace collaboration scenario of two operators. In the example here, a remote user pointed with a point cloud finger a position where the robot should be moved, and the controlling operator moved the arm to that position.

The collaboration between multiple users requires a protocol established between them to avoid conflicting
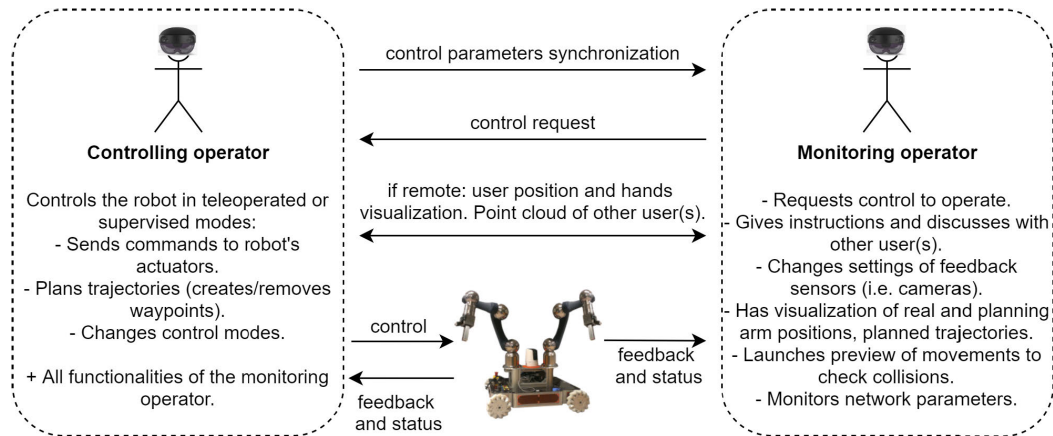
FIGURE 9. Multi-user operation workflow and data exchange. In the multi-user scenario with two or more operators, one operator controls the robot, and the other(s) is/are monitoring. The roles can be exchanged when a monitoring operator requests control. The robot's position is synchronised if the operators 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.
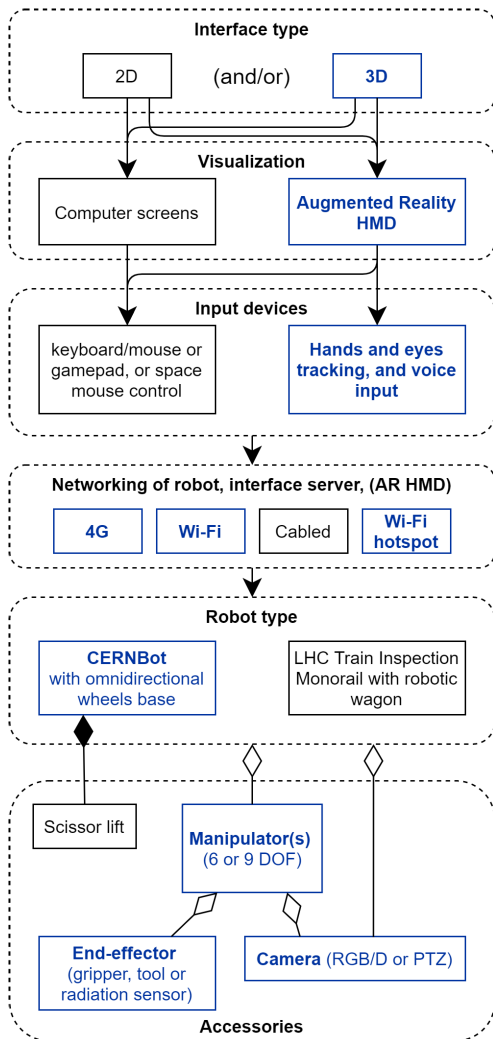
FIGURE 10. The diagram presents the workflow of the operator's interface setup. In this publication, as highlighted in the diagram in blue, the 3D scene is used, visualised in the AR HMD with hand, eyes tracking, and voice input, with 4G and Wi-Fi networking options for CERNBot with omnidirectional wheels, 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
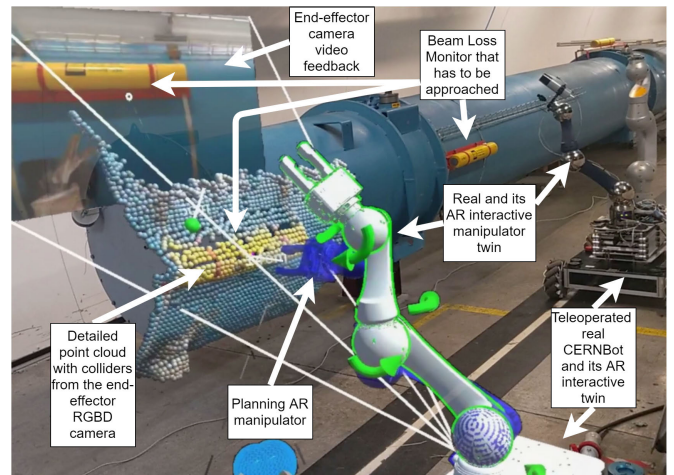
FIGURE 12. The overview of the scene where the operator and the robot were in Line Of Sight, the robot was located in a tunnel with the dipole magnet and Beam Loss Monitor device that the robot's end-effector had to approach. The operator used the robot's hologram (the Holographic Twin) in the workspace next to him. In the foreground, there was the CERNBot manipulator with the planning arm (blue colour) and the real arm (white colour) displayed.
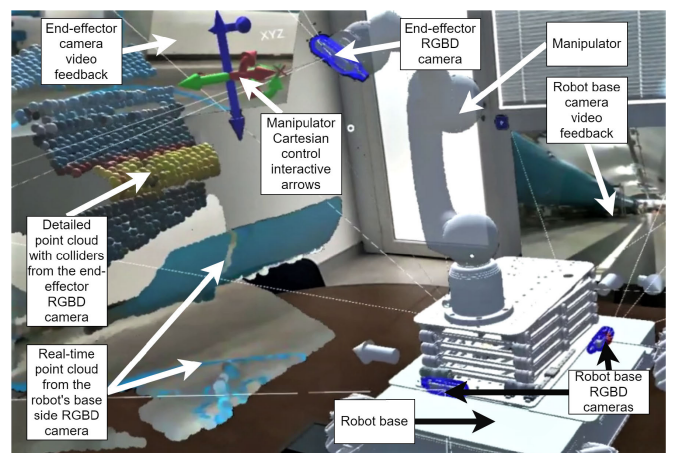
FIGURE 13. The scene overview of the remote control of the arm manipulated in the Cartesian space, the base control, and camera video and point cloud feedback.

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.

demonstration [65]). The hand tracking algorithm is explained in Section II-D4.

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.

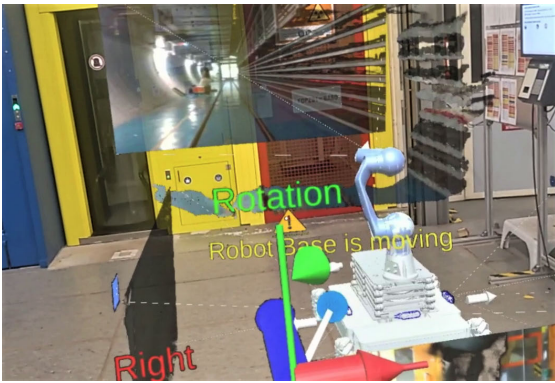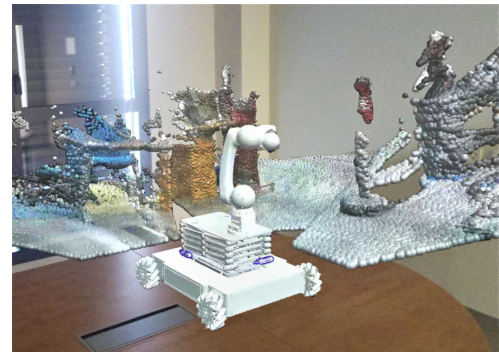**FIGURE 16.** Holographic robot model movement, rotation, and scaling with hand near and far interactions.

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

FIGURE 18. The robot surrounded by point clouds feedback from 3 cameras.
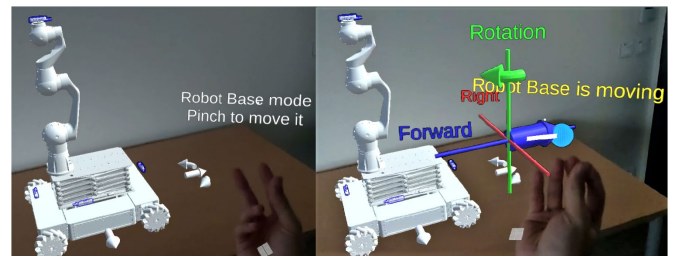
FIGURE 19. Movement of the robot's base with hand tracking. For double confirmation, the palm has to be orientated upwards, and the control is enabled when three fingertips are pinched together. Then the coordinate system appears and the base moves and rotates with speed proportional to the hand displacement 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

3) MANIPULATOR CONTROL

The manipulator can be controlled in 4 control modes (described in detail in Section II-D of [10]). For the AR
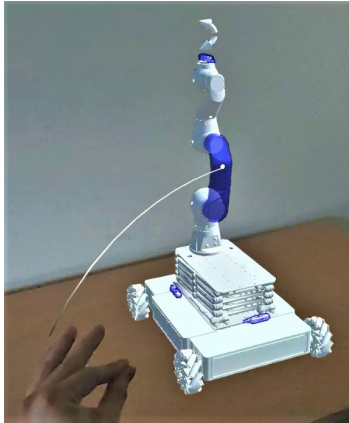
FIGURE 21. Instead of using the arrows (Figure 20), the joint can be rotated directly by clicking on the model and dragging it around its axis.

coordinate system arrows can also be activated by hand or eye tracking with dwell to actuate the movements (Figure 23).
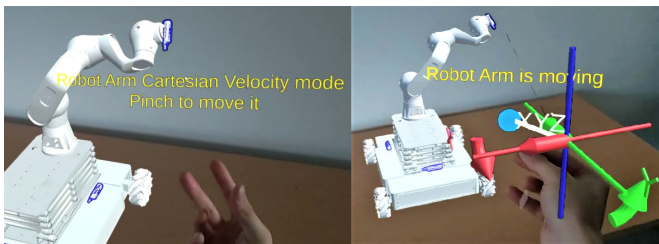
FIGURE 22. The manipulator in the Cartesian real-time velocity mode can be controlled with hand tracking. Similar to the base control (explained in Figure 19, there is the double confirmation system, and the manipulator's end-effector moves and rotates at a speed proportional to the hand displacement and rotation. In the video demonstration [65], it is presented how the robot is operated this way.
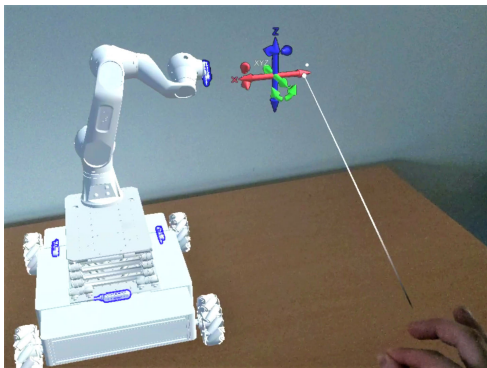
FIGURE 23. In the Cartesian real-time velocity mode, the arm can be controlled by clicking on the arrows or by pointing the arrow with eyes and using dwell (shown in the video demonstration [65]).

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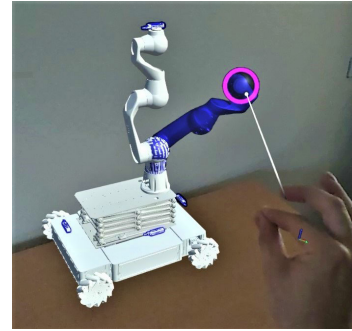
FIGURE 24. The planning inverse kinematic mode allows approaching a target using the FABRIK inverse kinematics. The end-effector target position can be changed by hand interaction.

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).
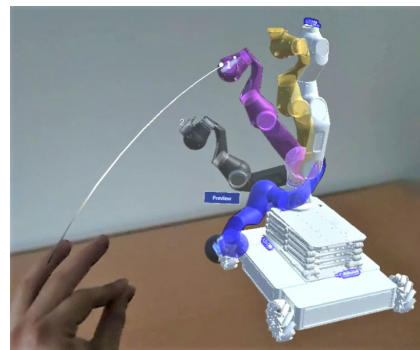
FIGURE 25. Selecting a waypoint and launching preview. In the planning modes, a trajectory with waypoints is created. Then before moving the real arm, the movement preview can be launched. If there is no collision with the environment or self-collision, the movement is started (Figure 26). Explanation 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].
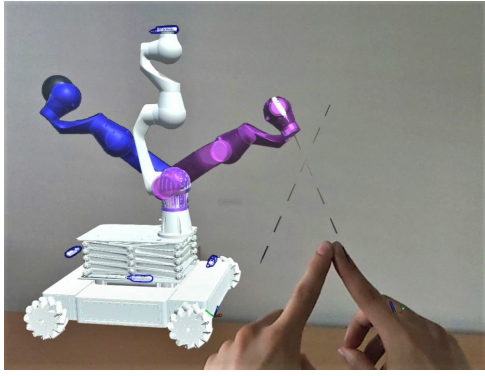
FIGURE 26. Moving the arm to a selected waypoint (violet colour). The movement of the real arm (white color) is executed with a dual hand gesture of touched index fingers. If the gesture is stopped, the robot stops the movement immediately.
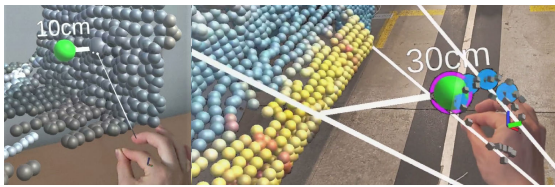
FIGURE 27. In the left picture, the selection of point cloud normal point is done with hand pointing and click or with eyes tracking and voice command. In the right picture, the generated normal point can be moved with hand interaction. The next step is to select the point for the target approach with the planning arm. The whole sequence is shown in the video demonstration [65]).
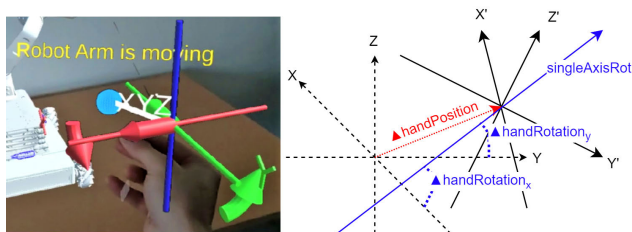
FIGURE 28. The diagram shows the hand tracking algorithm principle. When the user activates the tracking, the initial position and rotation of the hand are saved. Then the user moves the hand, causing the *ΔhandPosition* linear displacement and a rotation that can be expressed as *singleAxisRot* projected 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).

$$\overrightarrow{\Delta handPosition} = \begin{bmatrix} \Delta handPosition_x \\ \Delta handPosition_y \\ \Delta handPosition_y \end{bmatrix} \quad (1)$$

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.

$$projOnAxisPos_i = \overrightarrow{coordSysAxis_i} \cdot \overrightarrow{\Delta handPosition} \quad (2)$$
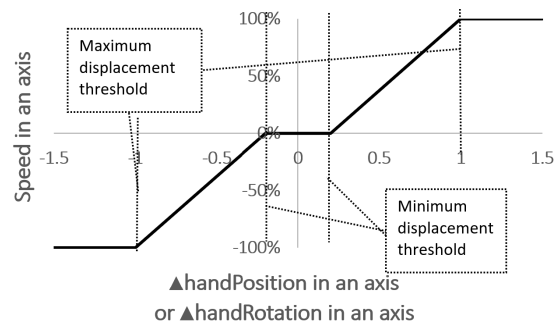
FIGURE 29. Translation of $\Delta handPosition_i$ or $\Delta handRotation_i$ into speed commands in *i* (x, y or z) axis. The movement starts after the displacement threshold (e.g. $+/-4$ cm or $+/-15°$ and achieves maximum $+/-100\%$ at, for example, $+/-20$ cm of hand displacement or $+/-90°$ rotation. The threshold can be changed in user settings.

The $\overrightarrow{\Delta handRotation}$ 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
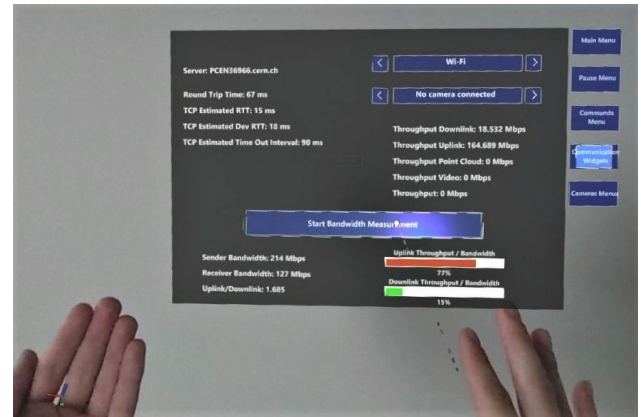
FIGURE 31. Network and communications menu showing parameters related to the network (i.e. bandwidth use and its measurements, delays, throughput related to camera feedback over selected network interfaces.
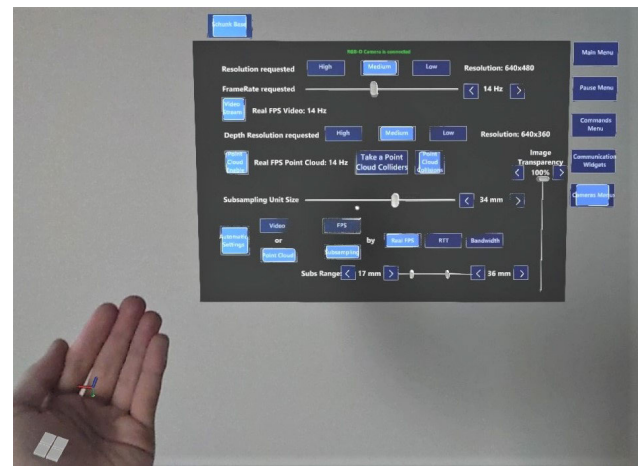
FIGURE 32. The camera settings can be accessed in the hand menu (Figure 30), which allows reaching the settings panels quickly.

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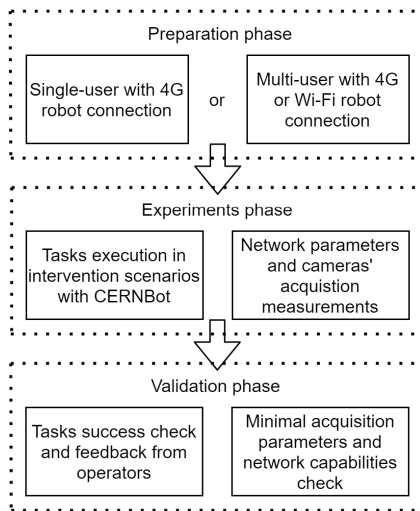
FIGURE 33. **The workflow of each experiment. The single or multi-user interface was used with 4G or Wi-Fi connections to the robot. Then the tasks were executed with parallel measurements. Finally, the validations of tasks and network behaviour were done.**

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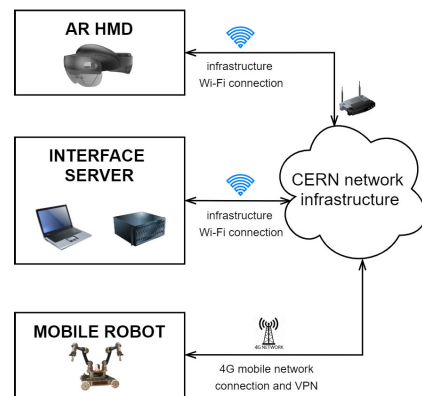
FIGURE 34. **Single-user validation scenario: robot (4G), HMD and interface (CERN Wi-Fi).**

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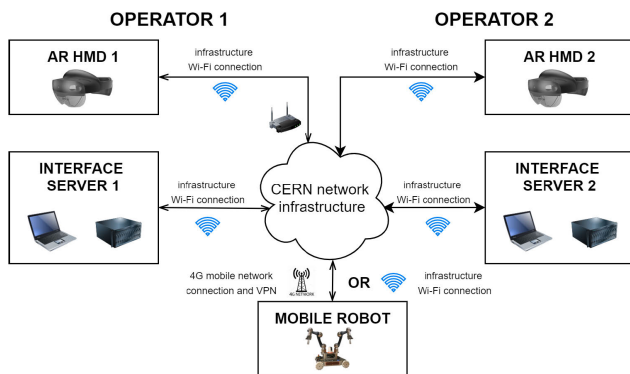
FIGURE 35. Multi-user network connections validation scenarios: robot (4G or CERN Wi-Fi), HMD and interface (CERN Wi-Fi).

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.
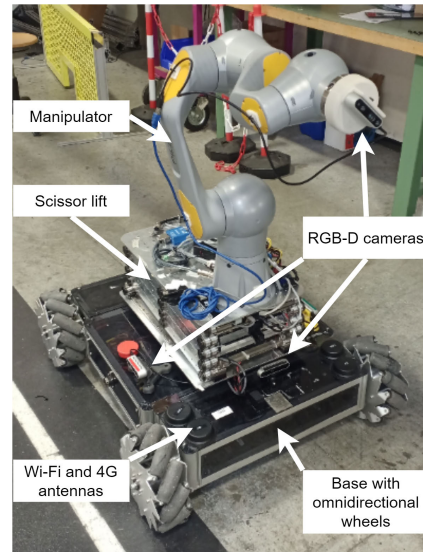
FIGURE 36. The CERNBot used for experiments.

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.
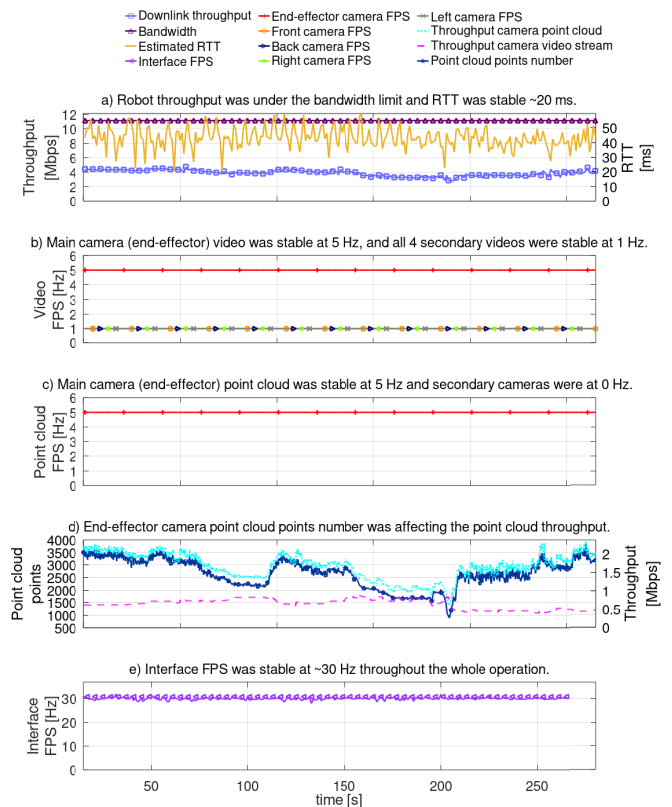
FIGURE 37. Single-user, 4G robot connection measurements to test the minimally required acquisition parameters for a single-user operation. In graphs b and c, there are stable FPS of the main camera (5 Hz) and four secondary cameras (1 Hz) for point clouds and video. In graph a, the RTT was between ~25 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.
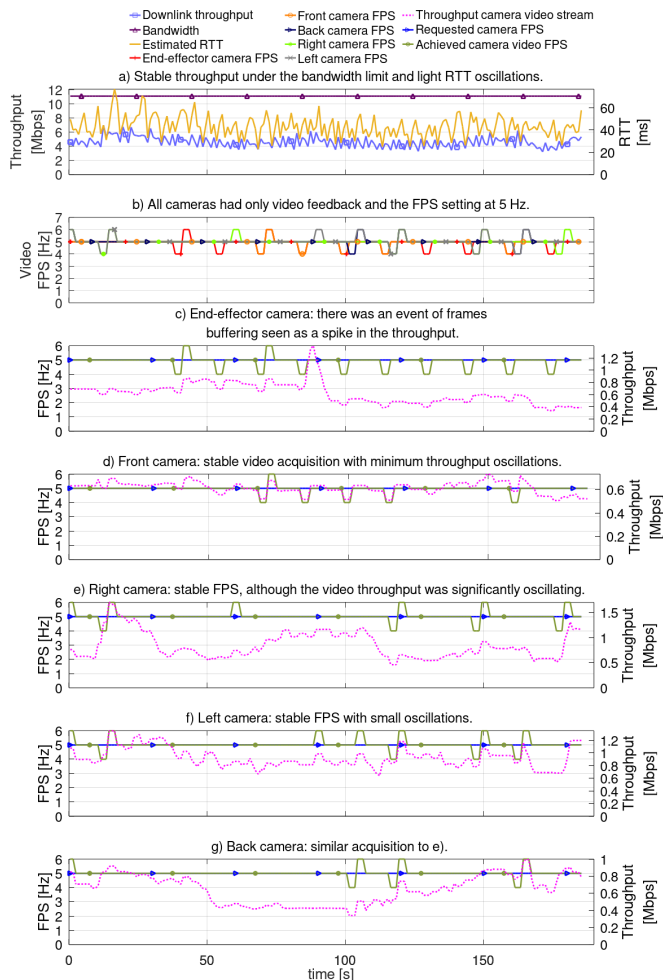
FIGURE 38. Single-user, 4G robot connection measurements of network congestion situation of reaching the maximum throughput (graph a). Throughout the whole period, all cameras were oscillating between 4 Hz and 6 Hz due to frames queuing (seen in graph b for all cameras and individually for each camera 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.
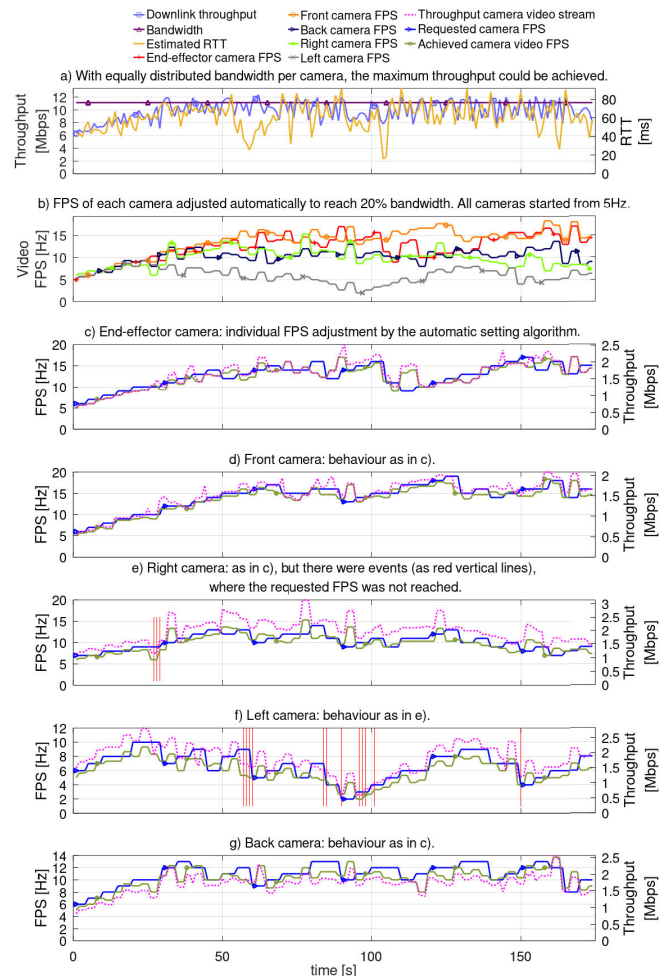
FIGURE 39. Single-user, 4G robot connection measurements of automatic video FPS for cameras to reach a bandwidth usage target. Each camera had an activated automatic mode to reach a 20% bandwidth usage target (total throughput shown in graph a, and individual camera throughputs in graphs c-g). All cameras 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 captur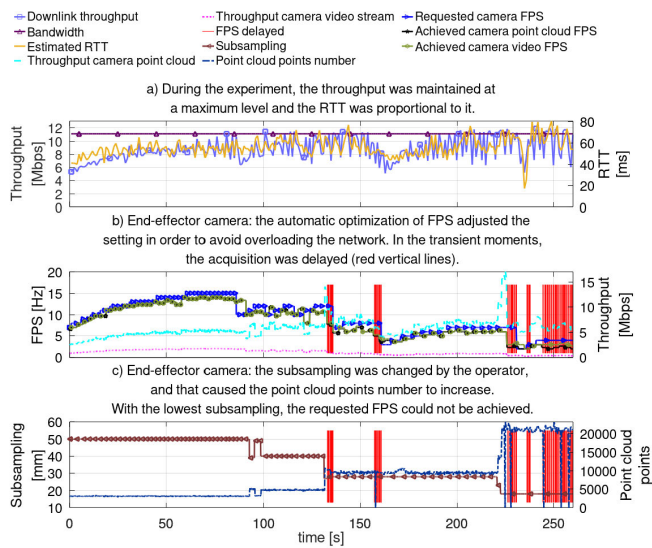ed by the depth sensor (graph c). That triggered the automatic optimisation to lower the requested FPS (graph b) to maintain the throughput. The vertical lines represent events when the real FPS is lower than the requested FPS by at least 2 Hz, and this is due to the delay of the algorithm and a temporary 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.
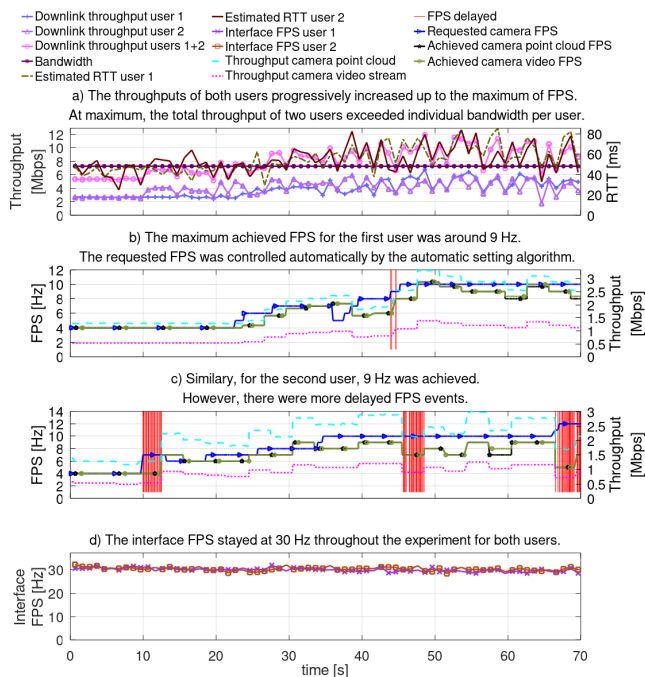
FIGURE 42. Multi-user, 4G robot connection measurements of the maximum FPS of one camera with point cloud and video used by two users. In graph a, the individual and total throughput progressively increased, and the total throughput exceeded the bandwidth per user. In graphs b and c, the FPS progressively 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).

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
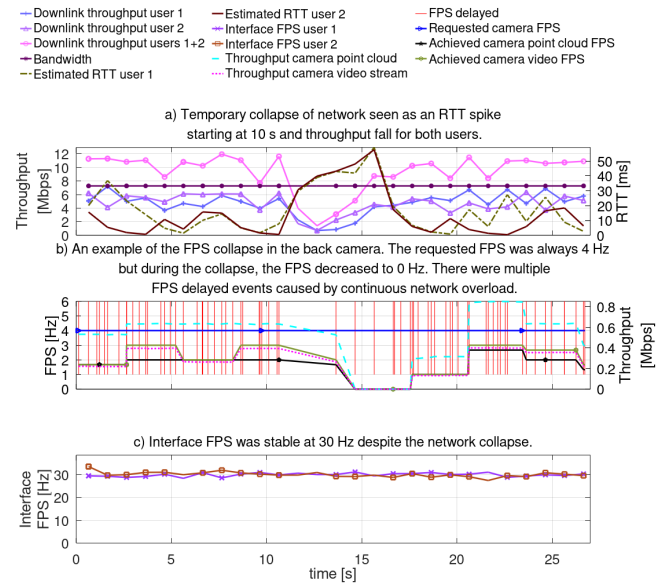
FIGURE 43. Multi-user, 4G robot connection measurements of a temporary event of network collapse for both users. It is characterized by the drop of throughput (graph a) and FPS to zero (graph b) and a high increase of RTT starting at 10 s and lasting 5 s (graph a). Before and after the event, the network 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).
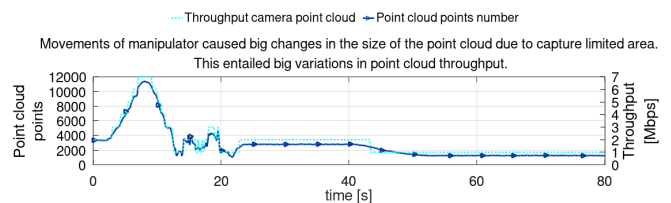
FIGURE 44. Multi-user, Wi-Fi robot connection measurements of dynamic point cloud size changes while driving the robot. In the first phase, the camera captured maximum ∼11000 points, and then the number decreased to ∼1300 points due to the move. According to the throughput used for this point cloud acquisition 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.

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
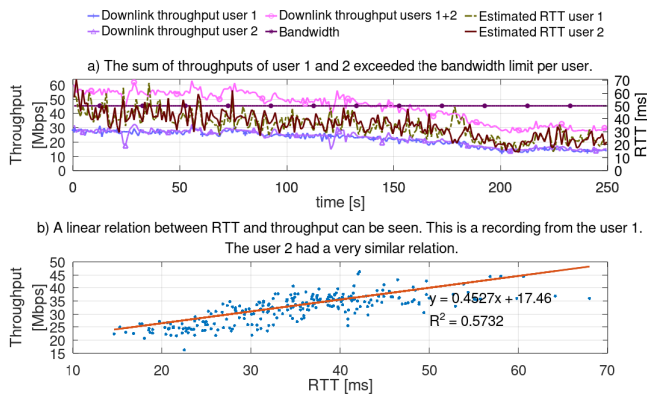
FIGURE 45. Multi-user, Wi-Fi robot connection measurements of the relation between throughput and RTT. In graphs a and b, the proportional relation between throughputs and RTT can be noticed. When the throughput declined, the RTT also proportionally decreased, for example, when the total throughput (both users summed) changed from ~55 Mbps to ~28 Mbps, and the RTT lowered from ~40 ms to ~20 ms.
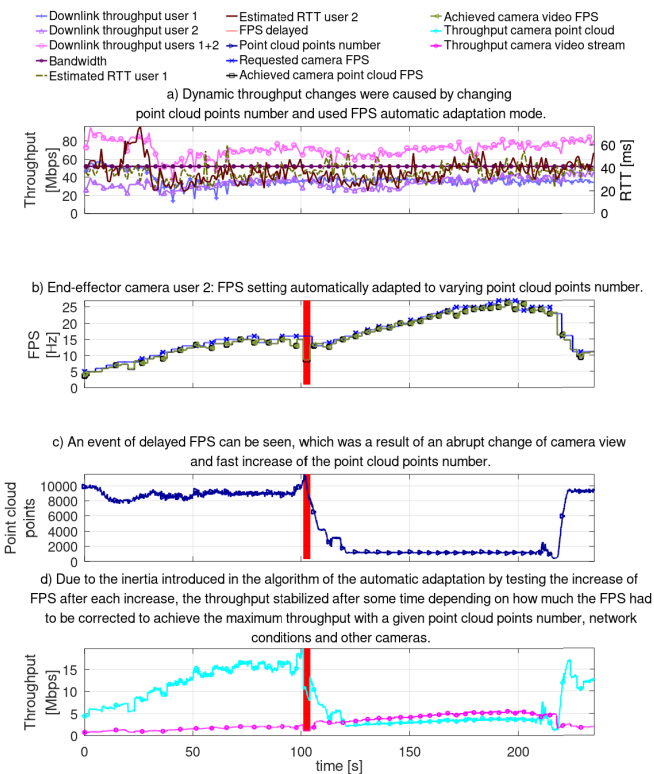
FIGURE 46. Multi-user, Wi-Fi robot connection measurements of automatic FPS adaptation to point cloud size. As seen in graph c, the point cloud points numbers were changing between ~1000 and ~13000 when the manipulator was moving and changing points of view. During these changes, the requested FPS was 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.
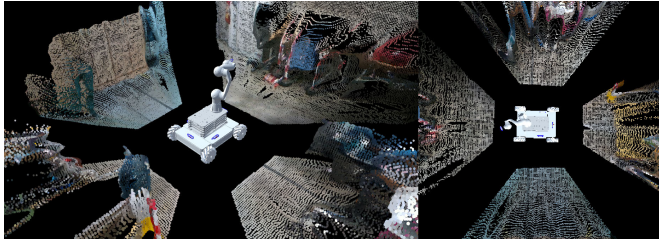
FIGURE 47. Problem of RGBD cameras placement to have 360 degrees point cloud view of the environment. Due to the offsets from the centre of the base and limited horizontal FOV <90° per camera, there are blind (black) areas at the corners of the base. In the figure, there are two views shown: on the left, 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.
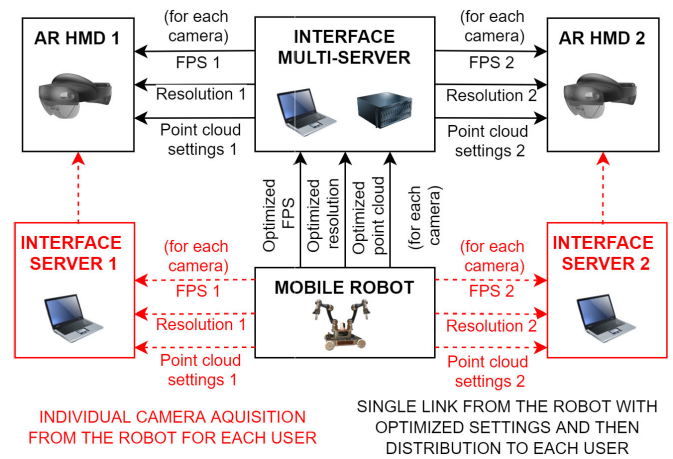
FIGURE 48. The optimized architecture for the multi-user acquisition of cameras feedback. In red, the architecture used for the validations (Figures 34 and 35) is presented, where each user could request individual camera acquisition from the robot. In black, the improved architecture optimizes the acquisition 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, telerobotics, 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.

• • •

Chapter 4

Enhanced Human-Robot Interface with Operator Physiological Parameters Monitoring and 3D Mixed Reality

Title:	Enhanced Human-Robot Interface with Operator Physiological Parameters Monitoring and 3D Mixed Reality
Authors:	Krzysztof Adam Szczurek, Roberto Cittadini, Raul Marin Prades, Eloise Matheson, Mario Di Castro
Journal:	IEEE Access
Year:	2023
Volume:	11
Pages:	39555-39576
DOI:	10.1109/ACCESS.2023.3268986
Quality index:	IF 3.476, Q1

APPLIED RESEARCH

Enhanced Human–Robot Interface With Operator Physiological Parameters Monitoring and 3D Mixed Reality

KRZYSZTOF ADAM SZCZUREK[1,2], **ROBERTO CITTADINI**[1,3], **RAUL MARIN PRADES**[2], **ELOISE MATHESON**[1], (Member, IEEE), **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, 12071 Castellón de la Plana, Spain
[3]Faculty of Engineering, Campus Bio-Medico University of Rome, 00128 Rome, Italy

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

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT Remote robotic interventions and maintenance tasks are frequently required in hazardous environments. Particularly, missions with a redundant mobile manipulator in the world's most complex machine, the CERN Large Hadron Collider (LHC), are performed in a sensitive underground environment with radioactive or electromagnetic hazards, bringing further challenges in safety and reliability. The mission's success depends on the robot's hardware and software, and when the tasks become too unpredictable to execute autonomously, the operators need to make critical decisions. Still, in most current human-machine systems, the state of the human is neglected. In this context, a novel 3D Mixed Reality (MR) human-robot interface with the Operator Monitoring System (OMS) was developed to advance safety and task efficiency with improved spatial awareness, advanced manipulator control, and collision avoidance. However, new techniques could increase the system's sophistication and add to the operator's workload and stress. Therefore, for operational validation, the 3D MR interface had to be compared with an operational 2D interface, which has been used in hundreds of interventions. With the 3D MR interface, the execution of precise approach tasks was faster, with no increased workload or physiological response. The new 3D MR techniques improved the teleoperation quality and safety while maintaining similar effects on the operator. The OMS worked jointly with the interface and performed well with operators with varied teleoperation backgrounds facing a stressful real telerobotic scenario in the LHC. The paper contributes to the methodology for human-centred interface evaluation incorporating the user's physiological state: heart rate, respiration rate and skin electrodermal activity, and combines it with the NASA TLX assessment method, questionnaires, and task execution time. It provides novel approaches to operator state identification, the GUI-OMS software architecture, and the evaluation of the 3D MR techniques. The solutions can be practically applied in mission-critical applications, such as telesurgery, space robotics, uncrewed transport vehicles and semi-autonomous machinery.

INDEX TERMS Electrodermal activity, hazardous environment, heartbeat, human–robot interfaces, mixed reality, operator workload, redundant mobile manipulator, respiration, safe operations, spatial perception, telerobotics, vital parameters.

I. INTRODUCTION

In hazardous environments, as long as it is technically and economically feasible, robots can execute actions in the

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

vicinity of danger to reduce risks for humans. However, depending on the conditions of such an environment and available resources, the autonomy level given to the robot must be decided. Suppose the task can be executed with high-enough reliability by following a set of formulated rules or using Artificial Intelligence (AI). In that case, an autonomous robot can be used without a constant connection with the operator. Such a task can be drone navigation in an open-space environment, recognition of objects, or manipulation of an assembly line robot. However, supervisory control or direct teleoperation is needed if the task is too unpredictable or critical decisions must be made based on a broad context. For example, during a robotic surgery executed by a remote doctor or a military robot control, the human operator should always be able to decide, interrupt and take control. Therefore, in a remote control system, the operators are a part of the system, and their state must be monitored. Mission success is inseparable from the human operator's performance and the efficiency of the tools [1]. From the operator's viewpoint, the factors contributing to the success can be divided into intrinsic and extrinsic, as explained in Figure 1.
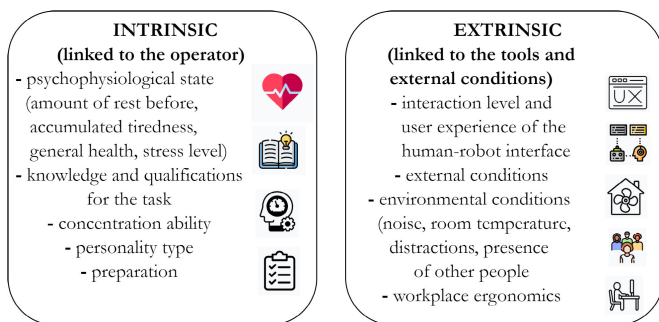

FIGURE 1. Factors contributing to the success of a telerobotic mission from the human operator's viewpoint.

Consequently, the success of a telerobotic mission can be considered in three categories to take into account the full system change impacted by an intervention and its consequences:

- The fulfilment of the intended goals (e.g. a remote inspection of a problem), the successful coping with potential unexpected events (e.g. obstacles, lost communication or additional tasks) that arose during the mission, and the time needed to accomplish it. The time can be measured only during the manipulation of the robot in such an intervention as surgery, or it can include the time needed for preparation, training and simulation for a space robotics mission [2]. Usually, the shorter the time, the better, or there is a time limit, such as a technical shutdown of an infrastructure where the mission is executed or when equipment is available.
- The state of the environment or the robot after completing the task. Suppose the environment's or the robot's value is high, for example, of the International Space Station [2] or the Large Hadron Collider (LHC)

equipment in a particle accelerator. In that case, any unintended collision or damage to the robot or the environment must be avoided and will contribute to telerobotic mission success. For example, if the robot's state is the most important, during operations with a Mars rover, any commands and navigation instructions must be sent after careful consideration.

- The state of the human operator during or after accomplishing the task. Maintaining the operator's good mental and physical state by eliminating stress factors, decreasing workload, or recognising the physiological overload of the person is crucial for avoiding manipulation errors and accidents, especially in long-lasting missions. The teleoperation can cause eye or muscle strain, stress, and temporary or even permanent harm to the operator (such as a headache, dry eyes or eye discomfort [3], or anatomical changes due to unnatural posture over extended time [4]). So, if the operator is in a bad state or can no longer operate, it may be considered a mission failure despite achieving other mission goals [5].

A. TRANSITION FROM 2D TO 3D MIXED REALITY HUMAN-ROBOT INTERFACES

Telerobotics in the particle accelerator complexes requires particular navigation techniques, knowledge of varied electrical, radioactive, magnetic or gas risks, and safety procedures. The controlled robot, such as a mobile platform with a redundant manipulator, also increases the necessary expertise. The risk of completely losing connection with the robot is present due to scarce communication resources or radiation that affects electronics. In case of such an event, another robot needs to be used to rescue the lost one, or special procedures must be thought of in advance. The lack of a high-bandwidth communication system affects the human perception of the remote environment. All these requirements and risks in the particle accelerator's hazardous environment impact the operator's emotional state. Therefore, improvements in the interface are necessary to provide the most appropriate tool that mitigates the stress and makes the executed actions easier. The training process is no less critical [2], and the simulators [6] are very helpful in training before the actual intervention.

Already 30 years ago, it was confirmed in [7] that adding virtual information in the perception channels during remote control greatly enhanced operator performance. In the cited work, virtual fixtures visually guiding the movement in specific directions or along certain shapes provided useful references that resulted in higher task efficiency. Moreover, the application of transparent Augmented Reality (AR) head-mounted devices (HMDs) for procedural guidance in space operations [8] showed improved performance and decreased workload during astronauts' tasks. The study in [9] confirmed that immersive interfaces enhance telepresence, efficiency and situational awareness, especially for hyper-redundant

robots, and were chosen by 94% of the research participants. The advantages of applying Augmented Reality Virtual Surrogates for aerial drone remote control were demonstrated in [10]. It eased distal operation and improved precise positioning and multitasking ability. Involving Mixed Reality in the robotic manipulator programming, as presented in [11], indicated the benefits of reducing program writing time and the number of errors due to virtual simulation in a virtual environment. Also, during a preliminary study and a 3D MR interface pilot project [12] at CERN, it was concluded that a transition from a 2D-based to 3D-based interfaces could bring multiple benefits, such as better spatial cognition, collisions avoidance or detection, motion planning, situational awareness, three-dimensional perception of the environment and spatial vision cues. Which, as a result, increased the capabilities and safety of teleoperation.

B. VITAL PARAMETERS MONITORING OF AN OPERATOR

The Autonomic Nervous System (ANS) [13] is a component of the peripheral nervous system that regulates involuntary physiologic processes, including heart rate, blood pressure, breathing, sweating, and digestion [14]. The ANS consists of two main branches:

1) Sympathetic Nervous System which activates body processes that help in stress or danger. It is responsible for your body's "fight-or-flight" response [15].
2) Parasympathetic Nervous System which has the opposite effect to the sympathetic nervous system and is responsible for the "rest-and-digest" body processes.

Therefore, these two parts of the ANS usually operate antagonistically: the former activates body processes, while the latter deactivates or lowers them. This balance is crucial for the body's well-being. Since dynamic changes in the ANS in response to a stressor cannot be controlled, certain physiological signals such as cardiac, respiration, and electrodermal activities can be used as reliable indicators of stress [16]. These biometric signals are valuable because they cannot be consciously controlled, falsified or kept hidden by a person and can reveal information about the unconscious state. Such a state can be frustration, a human response related to anger and disappointment, defined as an emotional state of no possibility of reaching a target. Stress is one of its consequences [17]. The most commonly used physiological signals to detect stress state are cardiac activity, respiratory activity, brain activity [18], body temperature [19], sweating, eye movements, facial expressions and gestures.

Previous studies about human health monitoring systems and the measurement of vital parameters have been considered for general purposes such as driving assistance and fatigue recognition [20], [21], office worker stress monitoring [22], coal mine worker safety [23], and remote video-mediated assistance [24]. In human-robot interaction, human emotions, especially negative ones, influence the performance of robotic interventions. For example, frustration could impact performance quality [25] and cause a waste

of time [26]. Several studies have been carried out on the monitoring of a person's vital parameters during activities involving the use of robots:

- Implementation of a cooperative human-machine interaction system with the primary objective of adapting the robotic arm control strategy according to the operator's emotional state, such as stress and fatigue using cardiac and electrodermal measurements [27];
- In human-robot cooperation, the robot was expected to recognise the psychological state of the human through the analysis of heart rate variability to deduce the mental stress of the user during collaboration actions [28];
- For wearable robotic equipment, such as exoskeletons, the interaction between humans and robots is paramount. User's physiological parameters are monitored to evaluate stress, reduce it to a minimal level and improve the applicability of assistive robotic devices [29];
- Operator emotions, physiological involvement, cognitive workload and usability in robotic teleoperation were investigated to design affect-aware robotic systems capable of adequately mitigating negative emotional states of the operator [30];
- In robotic surgery, surgeon electromyographic signal from muscles contraction was analysed for the validation of a novel approach to robot-aided pedicle screw fixation that guarantees comparable efficiency in the screw placement with lower muscular fatigue and more comfortable postures for the surgeon [31];

C. VITAL PARAMETERS MONITORING IN HAZARDOUS ENVIRONMENTS AT CERN

The health monitoring of workers in standard situations and emergencies in particle accelerators and experimental areas [32] is essential for personnel safety. A prototype Wireless Personnel Safety System (WPSS) [33] was developed to detect environmental conditions and monitor workers' health by measuring parameters such as heart rate and body temperature. In the context of search and rescue robots at CERN, an ultra-wideband radar for non-contact monitoring mounted on a mobile robotic platform was implemented following the autonomous detection of victims to classify survivors according to their need for medical assistance [34]. Robotics coupled with contactless monitoring was also applied to estimate the heart rate of workers during work activities in hazardous environments [35]. In telerobotics, when operations lasted longer than expected, the attention and concentration levels significantly dropped. Therefore, an eye-tracking system has been implemented in the human-robot interface [36], which prevented dangerous collisions by constraining the robot's movements and decreasing its speed when the operator became distracted or was not observing the robot's video feedback. A pilot project studying the benefits of the transition to the 3D MR interface and preliminary operator vital parameters monitoring assessment was conducted in [12], which motivated the further work presented here.
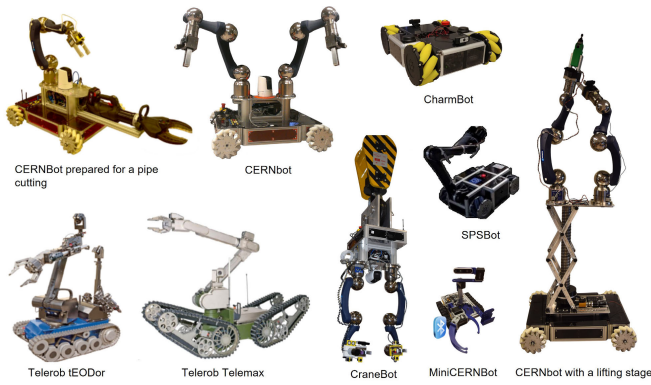
FIGURE 2. The robots equipped with specialized tools used for interventions in the CERN particle accelerators and experimental areas.

D. MOTIVATION AND PROBLEM FORMULATION

The CERN robots (Figure 2) have been used for more than 10 years for remote maintenance to increase the CERN accelerator complex's maintainability and availability [37]. These operations require reliable and well-adapted human-robot interfaces. Therefore, during the research and development process of the next-generation 3D Mixed-Reality human-centred interface, such validation criteria as the operator state, task execution efficiency and telemanipulation safety requirements needed to be taken into account. The 3D Mixed Reality (MR) brought new solutions, such as planning, automatic approach, point cloud 3D environment and stereoscopic view, improving efficiency and safety. However, the control complexity may have increased the operator's workload. Therefore, an appropriate method of assessing the workload had to be used. The fulfilment of functional specifications and the ability to execute predefined tasks were essential. However, human factors must also have been studied to deliver an optimal solution.

In the agile development and prototyping process, the users gave qualitative feedback that the 3D MR interface offered additional functionalities and advantages to the existing 2D interface. Still, a quantitative and detailed comparison with the previously used interface was needed. This quantitative data had to be gathered during the remote control of a real robot in nominal conditions, with environmental risks and stressors in real scenarios. In this study, the Train Inspection Monorail (TIM) platform [38] with a 9 degree of freedom (DOF) manipulator with a radioactive source in the end-effector was used to verify the correct functioning of Beam Loss Monitors (BLMs) in the LHC accelerator tunnels (Figure 3). Previously, humans must have done these verifications manually because of the complexity of the task, resulting in the person receiving a limited radioactive dose despite applying all safety procedures and personal protective equipment. The new robotic solution greatly enhanced human safety. However, the teleoperation risks involved damaging the accelerator equipment and the robot. Also, local rescue or repair interventions in case of robot accident or failure were limited due to the radiation hazard. Furthermore, the
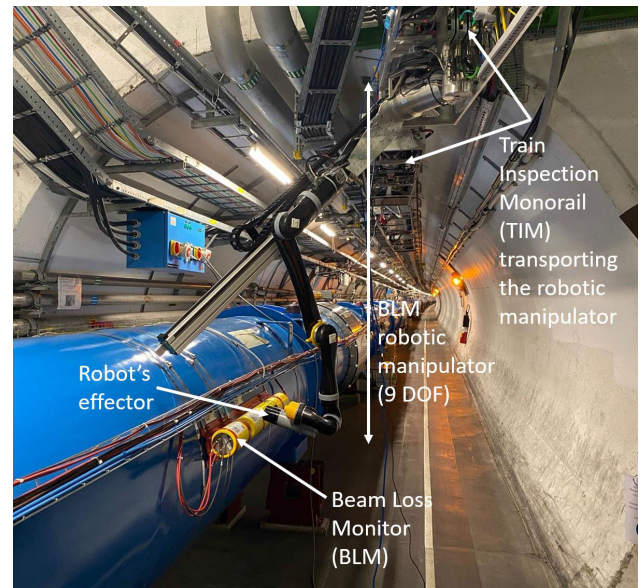

FIGURE 3. The 9 DOF robotic manipulator is installed on the TIM that operates in the LHC. In this picture [12], the arm is in the deployed state. A radioactive source is placed in the end-effector, and the arm must reach the BLM at a specified distance to verify the correct functioning of the device.

constrained communication link with the remote robot caused control signal or feedback delays. All these factors were the source of stress and required the total concentration of the operator.

The interface evaluation had to be based on objectively measured operator vital parameters, task execution times, failure or collision potential, precision and workload assessment techniques. The stress level of the operator was measured with the heartbeat, respiration and electrodermal activity to quantify how each interface's use affected the operator's physiological state. The quantitative data had to be supported by observations and qualitative responses given by the operators. The learning curves of the interfaces had to be compared. A study was done on how previous teleoperation and gaming experiences influenced learning and task execution efficiency. It was motivated by existing research [39], [40], [41], [42] showing that video game players perform significantly better on tasks requiring visual spatial attention, multiple object tracking, rapid processing of information and imagery, spatial resolution, visuomotor coordination and speed. Response to stressful situations, such as a collision, unexpected event or prolonged stress, was observed together with the measurements of the vital parameters to assess if the system could reliably recognize these situations. If vital parameters become abnormal, the system could warn the operator about the stress risk and potentially damaging effects or automatically reduce speed or stop the robot.

The experimental work presented in this publication studies the following hypotheses in the context of human-robot interfaces for mobile telerobotics in hazardous environments, based on the experimental data gathered in the particle accelerator robotic scenario with a redundant mobile manipulator:

1) Hypothesis 1: While providing a safer supervisory control and a better perception of the redundant manipulator and the environment, the 3D MR interface does not increase the operator's assessed workload compared to the previous operational 2D interface.
2) Hypothesis 2: The use of the 3D interface does not lead to an increase in the heart rate, respiration rate and electrodermal activity compared to the 2D interface.
3) Hypothesis 3: The task execution times with the 3D interface are faster than with the 2D interface.
4) Hypothesis 4: Operators with more gaming experience execute tasks faster.

The paper demonstrates that the 2D and 3D MR human-robot interfaces can be compared through the proposed evaluation techniques, which steer future interface developments. The paper also evaluates the Operator Physiological Parameters Monitoring System via a human study of 12 participants and concludes on its usability and limitations.

E. NOVELTY AND CONTRIBUTION

This publication extends and contributes to three domains. The first is the domain of physiological telerobotic operator vital parameters measurements applied to evaluate the workload and stress estimation during a telerobotic intervention in a harsh environment. The Operator Monitoring System (OMS) is proposed to continuously measure the heartbeat, respiration and electrodermal activity and assess the operator's state (relaxed, normal, stressed). The system had to be flexible enough to integrate sensors, including non-intrusive and wireless ones. These experiments gave valuable feedback on whether the system was well adapted to a person's anatomy and the differences in the physiological signals between operators under real operating conditions. The feedback allowed to fine-tune the algorithms and thresholds used for signal processing.

Secondly, in the experimental part of this work, a comparison of the 2D interface to the 3D Mixed Reality interface was provided in terms of task execution times, learning curves, the NASA TLX workload assessment, vital parameters measurements and detailed feedback questionnaires. The results were also compared regarding intrinsic human factors, such as robot teleoperation and gaming experience.

Lastly, improvements in the evaluation methodology were proposed based on the conclusions from the experiments, measurements and used methodologies. They should best suit the telerobotic use cases in particle accelerators, hazardous environments, or in general, applied to situations where the interfaces are used for remote control of manipulators or platforms. The methodology could also be applied to operator training and progress indication not only based on execution time but also on measured safety, operator's stress and focus, and calibrated feedback questions.

The experimental work was performed in a stressful and real scenario in the LHC at CERN, the world's biggest and most complex machine, which required the coordination of
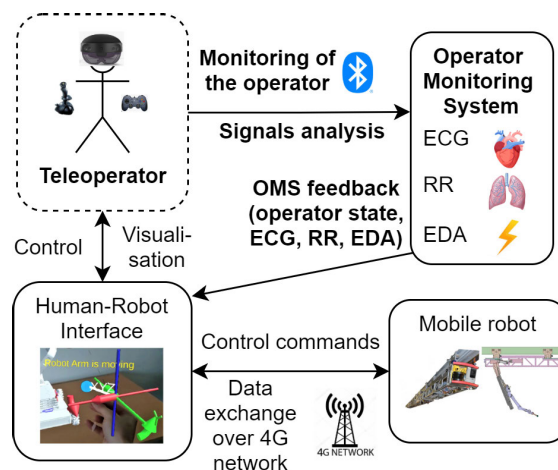

FIGURE 4. System overview presenting main components of the human-robot control chain with the OMS. The robot is controlled and exchanges data with the operator via the interface. The interface visualises the robot's state and interprets inputs from the operator. The OMS monitors the vital parameters, which are displayed in the interface. The elements highlighted in bold represent parts of the system this publication contributes.

multiple teams, accesses, and safety procedures. The results are a unique source of experience and ground for further human-robot interfacing advancement for robotics in harsh environments.

F. PAPER STRUCTURE

The paper is structured as follows:

- Section II describes the controlled robot, the 2D and 3D MR interface functionalities comparison and the OMS.
- Section III describes the experimental setup: tasks to be done, the use of interfaces, how the vital parameters were measured, what were the questionnaires, subjects, used hardware, and data post-processing.
- Section IV presents the results of the experiments, which are further discussed in Section V.
- Sections VI and VII conclude the work, summarise the findings, and define future work regarding the methodologies of interface evaluation and the OMS.

To better understand the problematics of telerobotics in particle accelerators and radioactive experimental areas, we recommend familiarisation with the CERNTAURO framework [37], the operational 2D interface description [36], and the detailed functionalities of the 3D MR interface [43].

II. SYSTEM DESCRIPTION

The system comprises four main elements that interact with each other: the robot, the interface, the operator and the OMS. The interrelations are shown in Figure 4. Section II-A details the characteristics of the robot and its mission tasks. Section II-B explains the differences between the two interfaces compared later in the study. Section II-C describes the OMS, physiological parameters, algorithms, signal processing, and its integration with the 3D MR interface.

FIGURE 5. The operator using the 2D interface and controlling a remote manipulator in the LHC. The cameras' view, control modes selection, keyboard input mapping, speed setting, and numeric joint positions can be seen on the laptop screen.

A. CONTROLLED ROBOT

The TIM mobile robot is used in the LHC's accelerator tunnel at CERN. The TIM comprises several principal wagons (control, drive, battery) and is responsible for carrying payload wagons, such as the robotic wagon (Figure 3). The train is suspended on a ceiling rail and has a wagon carrying a 9 DOF arm. The manipulator's task is to approach a BLM while holding a radioactive source at a specified distance. More explanation of the scenario, robot and task can be found in [12] in Section 1.1 and Section IE of [43].

B. 2D AND 3D INTERFACES FOR THE ROBOT'S CONTROL

The mission executed by the robot can be split into two phases. In the first phase, the monorail train approaches the inspection area, and safety is assured by clearing the passage and defining maximum speeds in the sections of the tunnel. The train also has a laser scanner that can detect any object in front of the train's front or back in the rail vicinity and stop movement. No obstacles are expected on the train's route, or automated doors open when the train must pass. When the train moves, the manipulator is folded and stored inside the wagon, and the radioactive source is sheltered in a radiation-blocking case. Therefore, an interface with video feedback has been sufficient for supervisory or manual control tasks. The radioactive source is extracted from the protective case for the manipulation phase and measurements. The folding and extraction are semi-automatised.

In the second phase, the manipulator is deployed in an environment that is not fully modelled, and obstacles are expected. The operator controls precisely the complex manipulator. This task requires high perceptual awareness due to the locations of the targets (the BLMs). Often, they are hidden behind other equipment or in the vicinity of fragile equipment. For these reasons, safety should be assured by additional means. In the standard 2D interface (Section II-B1), similarly to the train movement's supervision, safe intervention relies on the operator's experience and video feedback. There are no collision detection mechanisms,
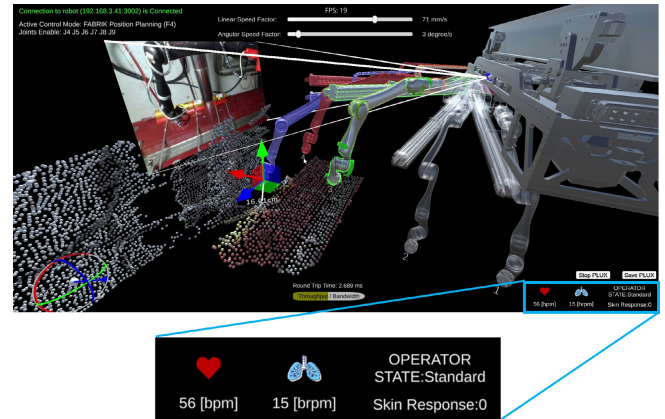

FIGURE 6. This overview shows the MR interface used with the trajectory specification to reach a target. On the upper right-hand side is the model of the train's wagon, where the arm is mounted. In the centre of the figure, there are waypoints displayed as transparent arms with numbers, the opaque grey arm showing the real arm's current position, the blue arm as the planning arm, and the red arm selected as the next waypoint. On the left are the 3D point cloud representation of the environment and the video camera 2D canvas, which, in this example, showed the equipment and the BLM in the LHC. Head-Up Display (HUD) elements show the robot's status and settings. In the lower right corner, the operator's vital parameters are displayed.

path planning, or obstacle recognition. These new functionalities have been developed for the 3D MR interface, described in Section II-B2.

1) THE 2D INTERFACE

In the 2D interface (Figure 5), the manipulator can be controlled in two real-time control modes:

1) Joints velocity control. The joints are moved separately at the desired speed.
2) Inverse kinematics. The arm can be moved or rotated along or around the axes of the end-effector or environment coordinate systems. The speed of the movement can be adjusted.

Several input devices, such as a joystick, keyboard, primary-secondary system or haptic controller, can be used. The feedback is composed of video camera streams, which can be displayed as multiple views on the screen. The operator can manually adjust the resolutions and frame rates of the video streams for the best compromise between necessary feedback quality and network bandwidth use and delays. Pan-tilt-zoom cameras can be rotated and zoomed to achieve the most convenient viewpoint, while standard cameras have fixed points of view. The 2D interface requires an operator with expertise in the robot's configuration and movement due to the manipulator's complexity and redundancy in the BLM robotic measurement project. A complete interface architecture description can be found in [36]. It also describes some other functionalities, such as autonomous behaviour scripting or a prototype of trajectory definitions.

2) THE 3D MIXED REALITY INTERFACE

The new functionalities provided by the 3D MR interface extend the previously used 2D interface manipulator control
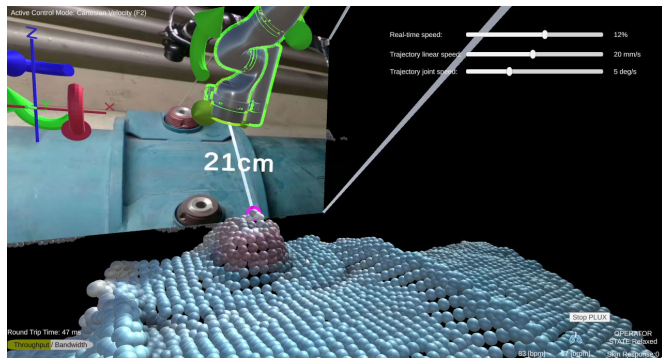
FIGURE 7. The precise spatial representation of the environment allows for calculating the distance between the end-effector and the nearest environmental point, which is an advantage in the approach task. In this figure, the video canvas can also be seen in the background, and torque information is shown as arrows next to the last joint of the arm. Here, the torque was minimal, displayed as green, but the arrow would turn orange or red during a collision.

capabilities with target-oriented and trajectory task specification, supervised position-based command, collision checks, movement preview, and a precise approach with distance measurement. A considerable portion of functionalities was developed thanks to the 3D representation of the robot and the perceived environment as a 3D point cloud. The extra functionalities are:

- Planning of the movement trajectories with joint or inverse kinematics control (Sections II-D and II-E, and Figure 13 in [43]).
- Preview of the manipulator behaviour, collision avoidance in planning and detection in real-time (Section II-F and Figures 14, 15, 16, 17, and 18 in [43]).
- 3D point cloud feedback (Section II-I and Figures 1, 9, 19, 20, and 24 in [43]).
- Automatic approach of the manipulator to a selected location of the point cloud (Section II-G and Figure 19 in [43]).

Figure 6 shows an overview of the interface during an operation in the LHC. A complete description of the architecture and functionalities in the 3D MR interface can be found in Section II of [43]. Figure 7 shows how the point cloud representation is used to approach an element of the robot's environment with a precise distance visualization. To cope with the network bandwidth limitations and high delays in the CERN facilities, especially when the voluminous point cloud is streamed, the Adaptive Communications Congestion Control was developed (Section II-I-4 of [43]).

C. THE OPERATOR MONITORING SYSTEM

The CERN Robot Operator Physiological Parameters Monitoring System, or shorter, the Operator Monitoring System (OMS), was designed to measure the physiological parameters of an operator during robotic interventions. The system has a modular architecture (Figure 8) and allows easy integration with physiological sensors. It can work as a standalone application to produce and export raw and post-processed
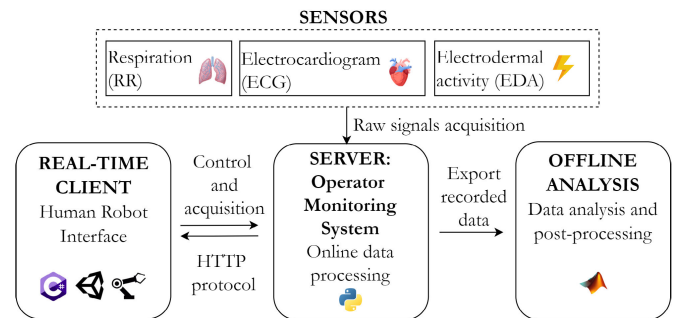

FIGURE 8. Architecture of the Operator Monitoring System. The OMS, a server, acquires the raw signals from the sensors. It processes them into vital physiological parameter values and assesses the operator state. A client (e.g. human-robot interface) subscribes to the process to receive the values and control the OMS (stop/start/recording/measure baseline). The exported recorded data can also be post-processed in offline analysis.

data recordings or connect to a client to provide a service. In our use-case of the service, the client was the MR human-robot interface communicating with the OMS. In this research, the system served the following two purposes:

1) Quantitative assessment of the operator's state during a mission by observing the vital parameters with the update rate of 5 s (which can be flexibly adjusted).
2) In the experiments performed in this study, to have a measurable indicator of stress or workload for comparing the user interfaces or assessing the workload during a task.

The OMS was a server for the MR human-robot interface client, as shown in Figure 8. The OMS was developed with Python 3.8.8, SciPy and Flask libraries, and the PLUX BioSignals API (Application Programming Interface). The source code and documentation of the OMS application can be found in [44]. On the other hand, the MR interface and its integration with the OMS were developed with C# in the Unity platform.

1) PHYSIOLOGICAL PARAMETERS ACQUISITION

The physiological signals and parameters considered in this study are related to the cardiac, respiratory and sweat glands' activities. These physiological parameters were chosen as the more substantial and accessible for inferring the operator's state of stress. The signal acquisition does not require invasive sensors disturbing telecontrol (e.g. helmets or sensory gloves) and does not significantly limit the operator or cause discomfort during prolonged use. The BioSignals PLUX toolset consists of an 8-channel central hub (Figure 9) designed for synchronous physiological data acquisition of up to 8 analogue sensors simultaneously. It supports up to 10 hours of signal streaming with Bluetooth communication at up to 3 kHz sampling rate and 16-bit resolution for each channel. In this work, the following physiological parameters were selected and acquired for evaluation and monitoring of the operator's status:
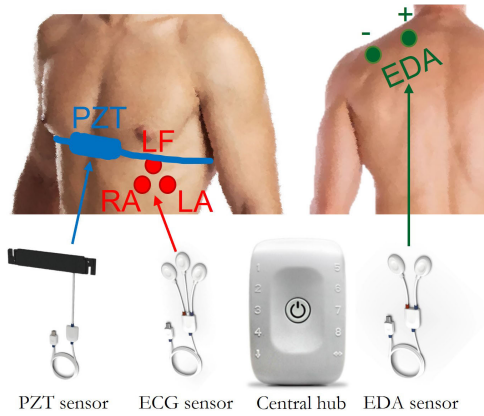
FIGURE 9. Body placement of the ECG electrodes on the chest, the EDA electrodes in the back shoulder region, and PZT belt around the chest at diaphragm level. The sensors are connected to the central hub.

- Heart rate (HR): heartbeats per minute (bpm). HR values are extracted directly from the ECG signal analysis, which measures the electrical activation leading to heart muscle contraction. This requires three electrodes to be applied on the operator's body at chest height (Figure 9), as the user's manual indicates. The recommended electrode placement described in the user manual was strictly followed for ECG signal acquisition in Einthoven configurations [45]. Cardiac activity was an excellent marker for detecting changes in the ANS activity [46]. HR value varies with gender, age, weight, medical conditions, medications, diet, and fitness level [47].

- Respiration rate (RR): number of breaths per minute (brpm). A complete breath combines two actions: inhalation, when the air is introduced into the lungs, and exhalation when the air leaves the lungs. The standard respiration rate for an adult at rest averages between 12 and 20 brpm. To detect respiration, the PZT sensor consists of a wearable chest belt with an integrated localized piezoelectric element that measures expansion changes caused by volume changes in the chest or abdomen during breathing cycles (Figure 9). The elastic chest belt can be adjusted in length to fit different anatomies, body positions, and chest and abdomen circumferences according to the scientific literature [48] [49].

- Electrodermal activity (EDA): refers to the variation of the skin's electrical conductance in response to sweat secretion due to changing sympathetic nervous system activity. Several works identified a strong correlation between EDA signal and emotional arousal [50]. A typical EDA signal results from two additive processes: the skin conductance level (SCL), which fluctuates slowly and represents the tonic base level, and the skin conductance response (SCR), a fast-varying phasic component. Therefore, the phasic activity can be identified as bursts with steep inclines and declines in the continuous data
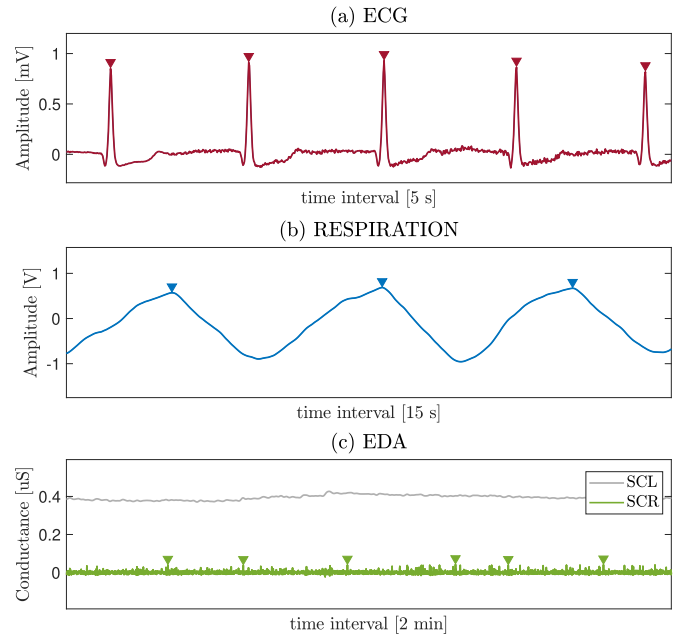

FIGURE 10. (a) ECG: characteristic electrocardiogram trace with the detection of QRS-complex. (b) Respiration: waves related to extending the sensory belt's embedded piezoelectric material due to the chest's expansion during breathing activity. (c) EDA: The electrodermal signal comprises the skin conductance level (SCL) related to the skin tonic level and the skin conductance response (SCR) related to the phasic response. The triangles represent detected peaks.

stream. The EDA signal peak represents an unexpected event that happened to the monitored person, and the peak level is proportional to the event's intensity. The EDA sensor electrodes were applied on the skin between the neck and the shoulder (Figure 9).

2) SIGNALS PROCESSING

The ECG signal tracing is characterized by the typical alternation of P wave, QRS complex, T wave and U wave. The QRS-complex is the most significant part of the ECG (Figure 10a), representing the electrical activation of the ventricles that contract and expel blood from the heart. Characteristic patterns of the QRS-complex are evident in the ECG tracing, from which the person's heartbeat can be tracked. R-peaks related to the QRS-complex are the points of the largest amplitude in the ECG signal and represent heartbeats. The recorded ECG time series feed the function identifying local maximums in the data set. The function is customized by setting the parameters of prominence, threshold, and minimum distance between two successive peaks with a twofold purpose: the former is to avoid the false detection of noise peaks unrelated to cardiac activity and triggered by the motion of the operator; the latter filters out the other characteristic waves of the ECG signal, but identifies only the peaks related to the QRS-complex.

The respiration pattern is characterized by a generic alternation of waves of varying amplitude depending on the depth of the operator's breath (Figure 10b). Deep breaths lead to a wider extension of the thoracic cavity and a higher stress on the sensitive piezoelectric part of the respiration belt.

On the contrary, short breaths result in less stress on the sensing part and lower the wave's amplitude. The respiration signal patterns can be affected by sudden movements, chest-abdominal muscle contraction or sometimes by the operator's speech.

The EDA signal is filtered with a low-pass filter with a cutoff frequency of 5 Hz, to remove noise or other artefacts and then decomposed in its tonic (SCL) and phasic (SCR) components (Figure 10c). A low-pass Butterworth filter is implemented for the SCL with a cutoff frequency of 0.1 Hz. The SCR can be extracted by filtering with a high-pass Butterworth filter with a cutoff frequency of 0.1 Hz [51].

The HR, RR and EDA processing is described by Algorithms 1, 2, and 3. Each algorithm takes the analogue-to-digital (ADC) measurements of electrical signals from ECG, PZT and EDA sensors, which were converted and scaled to their physical representations according to the PLUX manufacturer's guidelines. The constants in the algorithms were calibrated after extensive testing with several operators and set to values that worked for all of them. With the chosen sampling rate and constants, the maximum heart rate that can be calculated is 115 bpm. Accordingly, the RR limit is 42 brpm. The EDA events limit is 18 events per minute (epm), which is much above the physiological limit. The operators may need to operate a more physically demanding input device, which could increase their maximum heart rate above the algorithm limit. In this case, the sampling rate must be increased, or the *DISTANCE* parameter in the find_peaks() function be decreased.

Algorithm 1 The HR Calculation From the Raw ECG Acquisition Data

Inputs: *rawData*: sampled raw data from the acquisition unit with 250 Hz [mV],

Output: *HR*: calculated heart rate [bpm].

Constants: *PROMINENCE*: 0.09 [mV],
  DISTANCE: 130 [samples],
  HEIGHT: [0.0, 1.0] [mV].

1: The *rawData* is sampled at 250 Hz. For the calculation, the last 60 seconds of the acquisition are taken.
2: The peaks are detected by the function find_peaks() from the SciPy.signal module, with *PROMINENCE*, *DISTANCE* and *HEIGHT* parameters.
3: The number of detected peaks is the *HR*.

The physiological state of the operator (R) is estimated as a weighted average of these triggers (Figure 11). The distinguished states are: relaxed (0-30%), standard (30-65%), and stressed (65-100%). The weights were selected experimentally according to the contribution of each physiological parameter to the state evaluation. The HR weight is 0.6 because it most responds to situational stress factors. The RR changes are slower and more long-term, and its indication combined with HR can further infer body unrest. Therefore, the chosen weight is 0.3. The EDA completes the

---

Algorithm 2 The RR Calculation From the Raw Respiration Belt PZT Sensor Acquisition Data

Inputs: *rawData*: sampled raw data from the acquisition unit with 250 Hz [V],

Output: *RR*: calculated respiration rate [brpm].

Constants: *PROMINENCE*: 0.1 [V],
  DISTANCE: 350 [samples],
  THRESHOLD: [0.0, 0.7] [V].

1: The *rawData* is sampled at 250 Hz. For the calculation, the last 60 seconds of the acquisition are taken.
2: The peaks are detected by the function find_peaks() from the SciPy.signal module, with *rawData*, with *PROMINENCE*, *DISTANCE* and *THRESHOLD* parameters.
3: The number of detected peaks is the *RR*.

Algorithm 3 The EDA Events Calculation From the Raw EDA Acquisition Data

Inputs: rawData: sampled raw data from the acquisition unit with 250 Hz [μS],

Output: *EDAEvents*: calculated EDA events [epm].

Temp: *SCR*: skin conductance response [μS].

Temp: *SCL*: skin conductance level [μS].

Temp: *EDA*: Electrodermal activity [μS].

Constants: *CUTOFFFREQ*: 0.1 [Hz],
  DISTANCE: 800 [samples],
  HEIGHT: 0.06 [μS].

1: The *rawData* is sampled at 250 Hz. For the calculation, the last 60 seconds of the acquisition are taken.
2: Use median filter from *SciPy.signal* module, with function *medfilt* to obtain EDA.
3: Use 2-nd order Butterworth filter with *CUTOFFFREQ* on rawData to obtain *SCL*. The filter is the function butter() from the SciPy.signal module.
4: Calculate *SCR = EDA - SCL*.
5: The peaks are detected by the function find_peaks() from the SciPy.signal module, with *SCR* data, with *rawData*, and with *DISTANCE* and *HEIGHT* parameters.
6: The number of detected peaks is the *EDAEvents*.

---

state estimation with a weight of 0.1. It offers the fastest response to stress but can be easily influenced by temperature and arousal [52].

III. EXPERIMENTAL METHODS

The multilayered organisation of the experiments is presented in Figure 12. The participants remotely controlled an operational robotic manipulator (Figure 3) located in the LHC. The operation was done from the Control Centre (Figure 13). The intervention scenario allowed testing the interface in actual conditions, characterised by collision risks, temporary network bandwidth and delays deterioration, and unexpected events (e.g. hardware or software failure, a tunnel light turning off, or other uncontrolled scenarios). The 9 DOF manipulator created further challenges for participants who had never
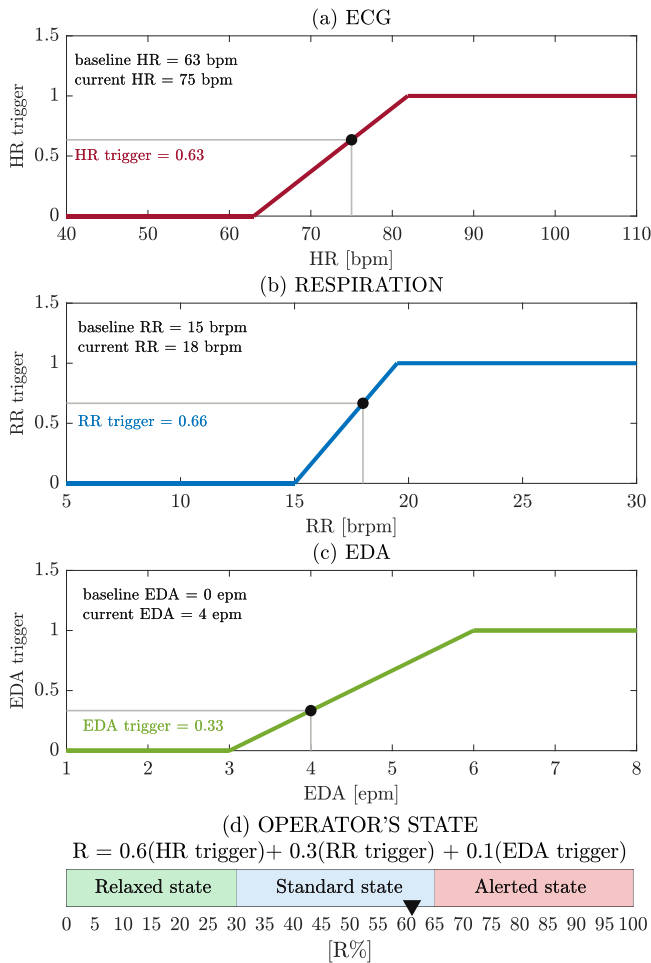
(a) ECG

baseline HR = 63 bpm
current HR = 75 bpm

HR trigger = 0.63

(b) RESPIRATION

baseline RR = 15 brpm
current RR = 18 brpm

RR trigger = 0.66

(c) EDA

baseline EDA = 0 epm
current EDA = 4 epm

EDA trigger = 0.33

(d) OPERATOR'S STATE
R = 0.6(HR trigger)+ 0.3(RR trigger) + 0.1(EDA trigger)

Relaxed state	Standard state	Alerted state

0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
[R%]

FIGURE 11. According to the current values of the physiological parameters, three different operator's states are identified: relaxed, standard and stressed state. The HR and RR triggers are calculated as ramps shown in the two upper graphs (a and b) and were calibrated individually accordingly to a participants' baselines (i.e. the start of the ramp is at the baseline value and it finishes at 130% of its value). The EDA trigger ramp (graph c) was fixed to start at 3 epm and finish at 6 epm. In graph d, the operator's state is evaluated according to the equation. In the presented example, the current HR, RR, and EDA values resulted in the standard state (R=61%).

operated a robot or had never operated a complex redundant manipulator. These conditions were sources of stress for participants.

A. TASKS AND INTERFACES USE

The tasks reflected real tasks that operators faced during BLM robotic measurements campaign [12]. There were two types of tasks:

1) An approach to a target with a specified distance from the end-effector tip, which holds a radioactive source, to the target (i.e. the BLM). The end-effector's orientation was not imposed because only the distance between the radioactive source to the device was essential.

2) An approach and then touch (a gentle push) of a target with the end-effector tip. The orientation of the end-effector was unrestricted. The only requirements
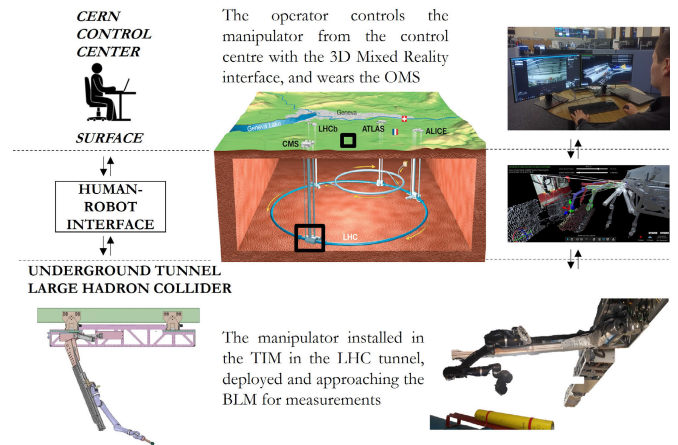

FIGURE 12. Remote robotic operations in the LHC at CERN. At the surface, the operator in the CERN Control Centre uses human-robot interfaces to control the robot in the underground tunnel of the LHC. The tunnel is, on average, 100 m below the surface and accessible by a few elevators. In the figure, on the left, there is a schematic representation of the control chain; on the right, the real operator, interface and robot pictures. In the case of the experiments in this publication, the robot was the TIM with a manipulator approaching the BLM for measurements. The operator wearing the OMS can be seen closer in Figure 13.
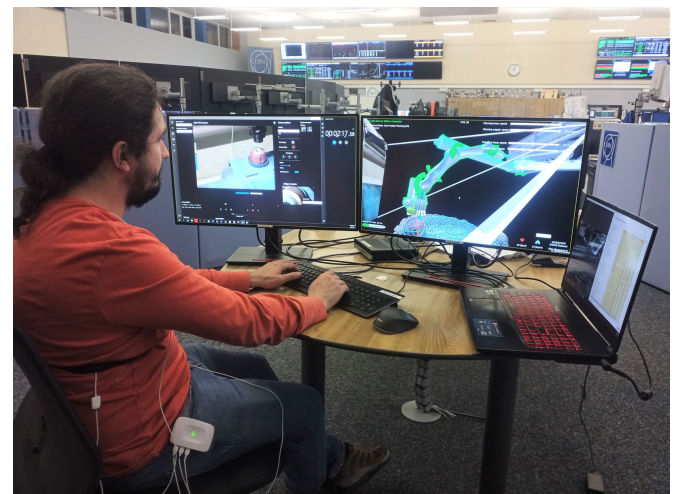
FIGURE 13. The picture shows an operator controlling the robot with the MR interface in the CERN Control Centre. The operator wore the OMS with attached ECG and EDA electrodes and an RR belt. The control station had two screens. The control commands were input with a keyboard and a mouse.

were the applied force direction and stopping when the contact was detected. This task reflected the manipulation task of the end-effector holding the radioactive source when it was remotely inserted into its protective container.

The operators alternately used the 2D GUI and the 3D MR interface for both tasks in the experiments. There were 16 attempts for each participant, 8 with the 2D GUI and 8 with the 3D interface. 4 attempts were made with each interface for each type of task. Table 1 presents the alternating sequence of the attempts. Before the actual attempts, each participant received a familiarisation session on using the interfaces,

with the possibility to try out functionalities later used to accomplish the tasks. During the attempts, communication with the experiment organisers was not allowed. Only in case of imminent danger could the experiment organiser react by advising to stop the movement or take over the robot's control. Such imminent dangers could be a potentially close distance from the equipment that should not be approached or a collision path that a participant overlooked.

During the use of each interface, the operator was free to choose any interface functionality to accomplish the task. This practically meant that in the 2D interface, the control mode could be real-time control joint by joint or with the inverse kinematics, and the cameras' quality, FPS orientation or zoom. For the 3D interface, additionally, the planning mode could be used, with previews of movements, detection and avoidance of collisions with the environment or self-collisions, as well as control of point cloud acquisition, display of distances from the arm to the environment and directing the arm to a selected point on a point cloud with an adjusted offset.

A portable computer and two external monitors were used to display and operate the robot with the interfaces. The inputs were a keyboard and a mouse for both interfaces.

B. QUESTIONNAIRES

The information gathered from participants before the experiments were:
- Experience in teleoperation (types and hours).
- Experience in video gaming (types of games and hours).

Based on their experience in teleoperation, the participants were categorized as beginners or experienced.

After finishing the tasks, the NASA TLX workload questionnaires were filled for 2D and 3D interfaces. Additional open questions about interfaces' use feedback, preferences or encountered problems were asked in a custom questionnaire.

C. SUBJECTS

In total, there were 12 participants in the experiments. The subjects were aged 23-34 years old. The categorization between gamers and non-gamers, or teleoperation beginners and experts, was based on the number of experience hours given by the participants before experiments. The participants were divided into groups based on their positioning with the median. In each group, there were 6 participants (i.e. 6 gamers vs 6 non-gamers, 6 beginners vs 6 experts). The declared gaming experience ranged from 150 to 21000 hours, with a median of 2000 hours. The teleoperation experience ranged from 0 hours to 3000 hours, with a median of 25 hours.

The 2D interface was used before experiments by 5 expert operators and 1 beginner operator. The 3D interface was used before experiments by 1 beginner operator and no expert operators. There were 3 expert operators with more gaming experience than the median, and 3 beginner operators had less gaming experience than the median. 4 gamers had used the 2D GUI before. None of the gamers had used the 3D GUI before. The described relations are presented in Table 2.

The study was performed with the Informed Consent of the participants. Personal and Sensitive Data were processed in the scope of scientific research in the context of CERN's specific activities (legal basis par. 28.5 and 29.5 of OC 11).

IV. RESULTS
A. VITAL PHYSIOLOGICAL PARAMETERS

An example of all vital parameters recording of one participant is shown in Figure 14. It is visible that the ECG, RR and EDA values varied throughout the experiment. Most of the time, the heartbeat rate goes up during task execution or decreases or stabilises during rest times. The subject collided during the touch task 2D-4, resulting in an elevated HR value. In the middle graph, it is noticeable that the RR values oscillated between ~8 until ~22 during operation and rest. In the lower chart, there are EDA events which did not present significant patterns correlated with tasks execution, collision moment or OMS validation period.

The system was adapted to the physiology of each participant by:
1) Calibration procedure: measuring baselines of HR and RR in rest condition before taking control of the robot. To present the diversity of these values, Figure 15 shows the distribution of participants' baselines.
2) The manufacturer's instructions for the respiration belt tightening were followed to obtain the best signal without discomfort. Different chest sizes were taken into account in the RR algorithm, based on the study in [48].
3) The parameters of signal processing algorithms were fine-tuned after being tested by more than 20 users.

The HR, RR, and EDA variations were calculated during executions of tasks for each participant, task type, and interface type, and then compared. The variations ΔHR, ΔRR, and ΔEDA were calculated individually for each participant as a difference between the mean HR, RR, and EDA values during the execution of tasks and the minimum recorded values. In Figure 16, it can be seen that there were no significant differences in the variations between the use of the 2D and 3D interfaces. The Pearson Correlation Coefficient (PCC) methodology was used to verify the linear correlation between all twelve participants' 2D and 3D data sets of HR, RR, and EDA. More description of the application of the PCC to compare such data can be found in Section II-C of [35], which explained how to translate the PCC value and p-value into correlation result. The analysis indicated their strong correlation with p-value < 0.0001, and PCC_{HR} of 0.936, PCC_{RR} of 0.969, and PCC_{EDA} of 0.961. Therefore, using the 3D interface resulted in a similar physiological response to the 2D interface. The variations were also compared between beginners and experts or gamers and non-gamers, and they did not show any distinct differences.

B. LEARNING CURVES

This section presents the comparison of execution times and learning curves between 2D and 3D interfaces (approach task in Figure 17, touch task in Figure 18, for

TABLE 1. Distribution and sequence of attempts per experiment participant. "App." = approach task; "Touch" = task of approaching and touching; "2D" = 2D GUI; "3D" = 3D MR interface.

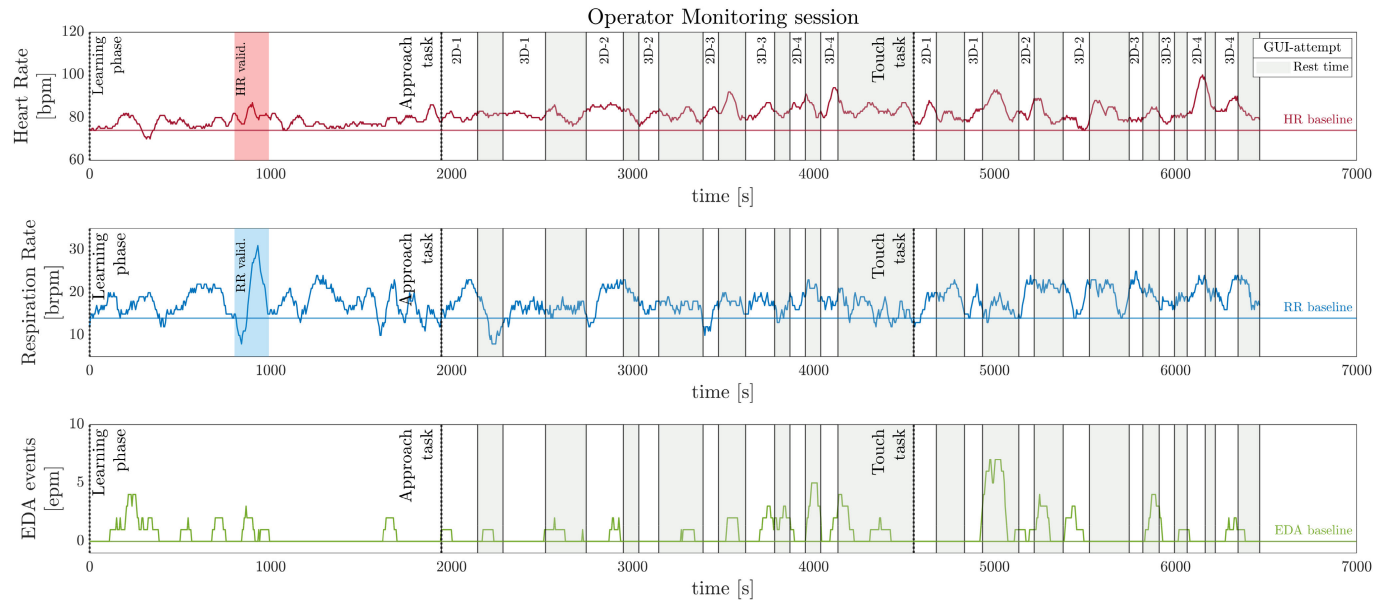| Attempt	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Interface + task	2D App	3D App	2D App	3D App	2D App	3D App	2D App	3D App	2D Touch	3D Touch	2D Touch	3D Touch	2D Touch	3D Touch	2D Touch	3D Touch

FIGURE 14. The graph presents recordings of HR, RR, and EDA values throughout the experiment of one participant. The horizontal timeline is divided into several phases of the experiment: learning, OMS testing as HR and RR validation, approach task and touch task repetitions. They are separated with annotated vertical lines. Each task is marked with its abbreviation on the upper graph (e.g. 2D-1, which means attempt 1 with 2D GUI). The HR and RR measurements were verified during the learning phase and are marked with red or blue backgrounds. The RR parameter was tested by having the participant hold their breath for up to 1 minute and then breathe with a specified rate for 1 minute to observe the increase and stabilization of the value. The HR parameter was verified with a reference device. The 16 task repetitions were done according to the tasks sequence in Table 1.

TABLE 2. The table shows the number of participants in the beginner or expert group. It also depicts how many beginners and experts were familiar with each interface and had gaming experience.

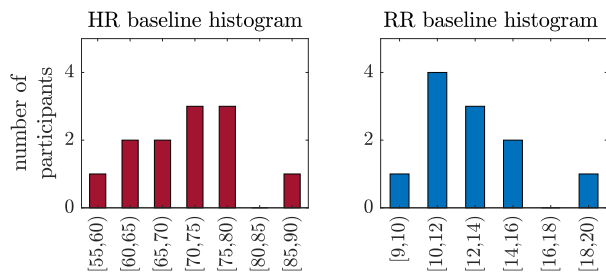|	Total	2D interface familiarity	3D MR interface familiarity	Gamers
Beginners	6	1	0	3
Experts	6	5	1	3

FIGURE 15. The distributions of HR and RR baselines among all participants are presented on the left and right, respectively. The range of HR baseline was between 55 and 90 and from 9 to 20 for RR.

gamers in Figure 19, for non-gamers in Figure 20, for beginner operators in Figure 21, for experts in Figure 22); and between gamers and non-gamers in Figure 23, experts and beginners in Figure 24; as well as combined gaming and remote control expertise in Figure 25.

The approach task was executed significantly faster with the 3D interface than the 2D (on the last attempt 42% faster), and the learning curve was steeper (Figure 17). The touch task was executed in the beginning 17% faster with the 2D interface, but in the end, the time was almost the same as the 3D, and the learning curve was steeper (Figure 18). The beginner operators were 23% faster with the 3D interface (Figure 21). The expert operators executed tasks 20% faster with the 2D interface on the first attempt, but on the last attempt, they were 14% faster with the 3D interface (Figure 22), which was due to the steeper learning curve. As expected, the expert operators executed the tasks 39% faster than beginner operators (Figure 24). Moreover, the expert operators with more gaming experience were 61% faster than beginner operators with less gaming experience (Figure 25).

It was observed in the group of experts that a more extensive video gaming experience resulted in faster learning of key bindings, understanding of controls and player movement, which resulted in faster goal achievement (expressed by an average time of all attempts).

C. NASA TLX QUESTIONNAIRES RESULTS

The results are split into 6 categories: 3 group users (beginner operators, experts, and all participants averaged) using the
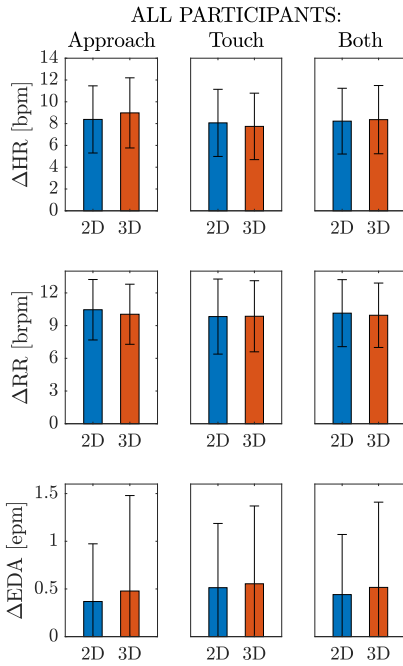
FIGURE 16. The HR, RR, and EDA variations compared for the approach, touch, and both combined tasks, for the 2D and 3D interface. All values are averaged for all participants.
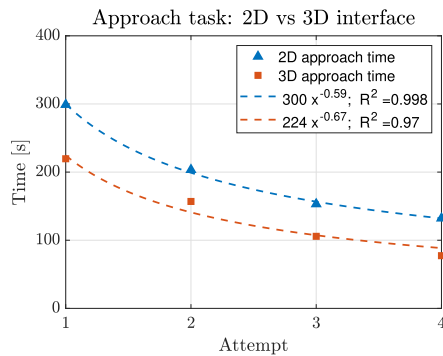

FIGURE 17. Learning curves comparison of all participants in the approach task with the 2D and 3D interface. It shows that, with the 3D interface, the approach task was executed 25% faster on the first attempt and 42% faster on the last attempt, and the learning curve was steeper.

2D interface; and, correspondingly, the 3D interface used by these 3 user groups. In Figure 26, there are four graphs of the raw rating, weights, adjusted rating and overall rating results, which compare the workload assessment. In the NASA TLX assessment methodology, the ratings range from 0% to 100%, and weights from 0 to 1. The raw rating analysis indicated the following characteristics:

- Mental Demand: No significant difference was found between 2D and 3D for all participants averaged. However, there was a substantial difference between beginners' and experts' demand values using the same interface. For 2D, experts had 39% less demand than beginners. For 3D, experts had 33% less demand than beginners.
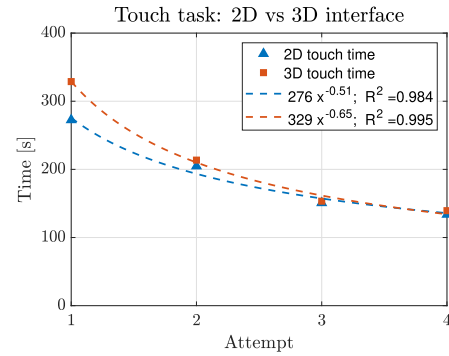- Physical Demand: The demand for all groups of participants was less than 6.

FIGURE 18. Learning curves comparison of all participants in the touch task with the 2D and 3D interface. It shows that, with the 3D interface, the approach task was executed 17% slower on the first attempt, but the times were similar on the last attempt, and that is because the learning curve of the 3D interface was steeper.

- Temporal Demand: No significant difference between 2D and 3D was found. For experts, 3D had 16% more, while for beginners, 3D had 5% less.
- Performance: For all participants averaged, 3D had 11% more demand than 2D to achieve the performance. Also, there was a difference between experts and beginners. Beginners had 18% more demand than experts.
- Effort: No significant difference was found between 2D and 3D. However, there was a significant difference between experts and beginners. Experts had 23% Effort less than beginners.
- Frustration: There was a significant difference between beginners using 2D and other groups using 2D or 3D. The value was 92% higher than, for example, for beginners using 3D.

The weights analysis demonstrated the following points:
- Mental Demand: Average weights for all groups varied between 0.18 and 0.24.
- Physical Demand: For all groups, the average weights were lower than 0.01.
- Temporal Demand: Average weights for all groups varied between 0.18 and 0.23.
- Performance: Average weights for all groups varied between 0.24 and 0.30
- Effort: Average weights for all groups varied between 0.14 and 0.21. The value was much higher for beginners using 3D than for other groups using 2D or 3D.
- Frustration: Average weights for all groups varied between 0.10 and 0.17. The value was the highest for beginners using 2D.

The adjusted rating analysis exhibited the following attributes:
- Mental Demand: For all participants averaged, the demand was 11% lower for 3D. There was a substantial difference between beginners' and experts' demand values. For 2D, experts had 45% less demand than beginners. For 3D, experts had 48% less demand.
- Physical Demand: For all groups of participants, the demand was negligible.
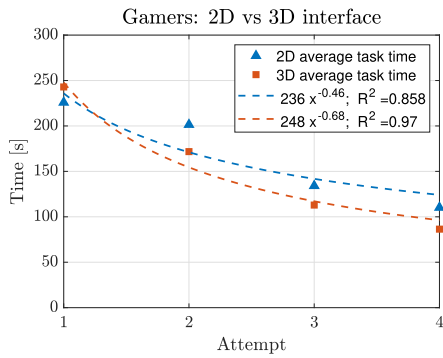
FIGURE 19. Learning curves comparison of participants with gaming experience using the 2D and 3D interface. It shows that participants with gaming experience had a steeper learning curve. Initially, they executed tasks 8% slower with the 3D interface, but in the end, they were 22% faster than with the 2D interface.
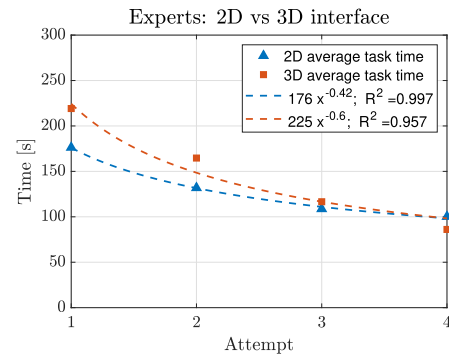

FIGURE 20. Learning curves comparison of non-gamers using the 2D and 3D interface. It shows that tasks were executed, on average, 11% faster with the 3D interface.

FIGURE 21. Learning curves comparison of beginner operators using the 2D and 3D interface. The tasks were executed on average 23% faster with the 3D interface.

FIGURE 22. Learning curves comparison of expert operators using the 2D and 3D interface. On the first attempt, the tasks were executed 20% faster with the 2D interface, but because of a steeper learning curve, in the last attempt, the tasks were done 14% faster with the 3D interface.
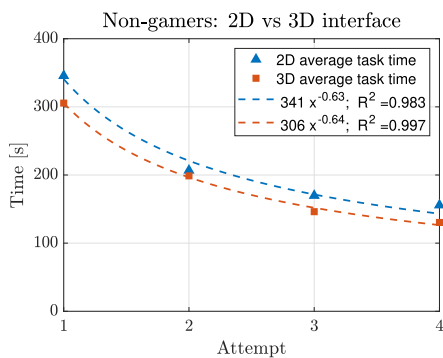

FIGURE 23. Learning curves comparison of participants concerning their gaming experience. The time was averaged for both interfaces and tasks. It shows that participants with gaming experience executed tasks 28% faster on the first attempt and 31% faster on the last attempt.

FIGURE 24. Learning curves comparison of participants with and without teleoperation experience averaged for both interfaces and tasks. It shows that expert operators were, on average, 39% faster than beginner operators. However, the learning curve for beginners was steeper.

- Temporal Demand: For 3D, overall, the value was 12% higher than for 2D.
- Performance: For all participants averaged, 3D had 11% more demand than 2D to achieve the performance, which was similar between all groups.
- Effort: 3D had, on average, 21% more demand than 2D. The situation is different from raw ratings due to beginners' weight responses.
- Frustration: The most significant difference was registered between beginners using 2D and 3D. The value

was 184% higher for 2D. There was no difference in Frustration for experts between using 2D and 3D.

The overall rating analysis indicated no substantial differences between the 2D and 3D interfaces in each beginner, expert, and all averaged participants. However, the analysis showed a few interesting findings in contributing factors'

FIGURE 25. Learning curves comparison of expert operators with more gaming experience (gamers) and beginner operators with less gaming experience (non-gamers) using the 2D and 3D interface. The tasks were executed on average 61% faster by the group of expert operators and gamers.
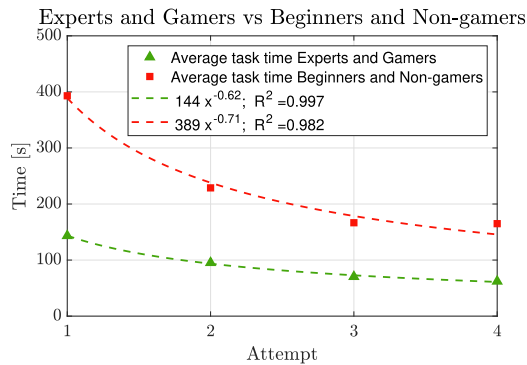
weights and differences between beginner and expert operators. These observations were made:

1) The Mental Demand in adjusted rating was much lower (47%) for experts than for beginners.
2) Performance contributed to the workload (in terms of weight) the most (27%), then Temporal Demand (22%), Mental Demand (21%), Effort (17%) and Frustration (13%). The Physical Demand did not contribute at all to the workload assessment.
3) Frustration while using the 2D interface was much higher (184%) for beginners than other groups using 2D and 3D interfaces.
4) It must be noted that variations of answers were significant between participants for most parameters as indicated by the standard deviation bars in Figure 26, which broadly overlapped for both interfaces. However, the Pearson Correlation Coefficient (PCC) methodology was used to verify the statistical relationship between the 2D and 3D responses of all participants for each adjusted rating. The analysis indicated their correlations:
   - PCC_{Mental} of 0.703 and p-value < 0.01;
   - $PCC_{Physical}$ of 1 and p-value < 0.0001;
   - $PCC_{Temporal}$ of 0.927 and p-value < 0.0001;
   - $PCC_{Performance}$ of 0.702 and p-value < 0.01;
   - PCC_{Effort} of 0.683 and p-value < 0.015;
   - $PCC_{Frustration}$ of 0.778 and p-value < 0.003;

All ratings showed higher moderate or strong correlations. Therefore, the ratings for an individual participant were correlated between the 2D and 3D interfaces, although there were significant inter-subject differences.

D. CUSTOM QUESTIONNAIRES

After the experiments, the participants filled in a questionnaire with several questions about the 2D and 3D interfaces and their comparisons. These detailed questions allowed to gather valuable insights into the reasons for the operator's higher workload and the interfaces' potential improvements. The listed questions with grouped detailed responses are in Appendix.

In summary, based on the questionnaires filled by participants, the following advantages were identified in the 3D interface in comparison to the 2D interface:

1) Better immersion in the scene. The 3D interface generally gave more confidence, especially to beginner operators. All participants highlighted that the 3D representation of the arm in the 3D environment was one of the best advantages of the 3D interface.
2) The point clouds complemented the video feedback information. It was confirmed that the point cloud helped approach a target accurately and avoid or detect collisions.
3) Several subtasks were more manageable in 3D than in 2D control: visualizing commands, trajectories and movement execution. The 3D feedback and advanced functionalities, such as a normal point selection, made the approach with a precise distance easier.
4) Safety was more assured by higher environmental awareness, collision detection and avoidance with the preview function.
5) The estimation of distances was easier with the 3D model, point cloud, and the numeric indicators of distances.
6) The 3D representation was fully confirmed to be beneficial for understanding the arm pose better.
7) Most participants (apart from expert operators who already operated the arm with the 2D interface) agreed that after working with the 3D interface, they better understood the joints' movement with the 2D interface. Therefore, the 3D interface could be used as a training tool before using the 2D interface, which requires more advanced knowledge.
8) If participants were asked to select the interface depending on the task, the 3D interface was preferred by all participants for the task of approaching a target behind obstacles with a higher risk of collision. For a task of simple approach, the preference was split (5 for 3D, 7 for 2D) due to not much use of advanced 3D functionalities. For the touch task, the choice was also split (7 for 3D and 5 for 2D); some participants used the advanced functionalities to detect the touch, while others preferred to depend only on the visual camera feedback. Figure 27 presents the results of the polls.

The following future improvements in the 3D interface were identified:

1) Simplification of complex controls (key bindings, input device, control modes). Using a different input device (space mouse, joystick or gamepad) for a player or arm movement.
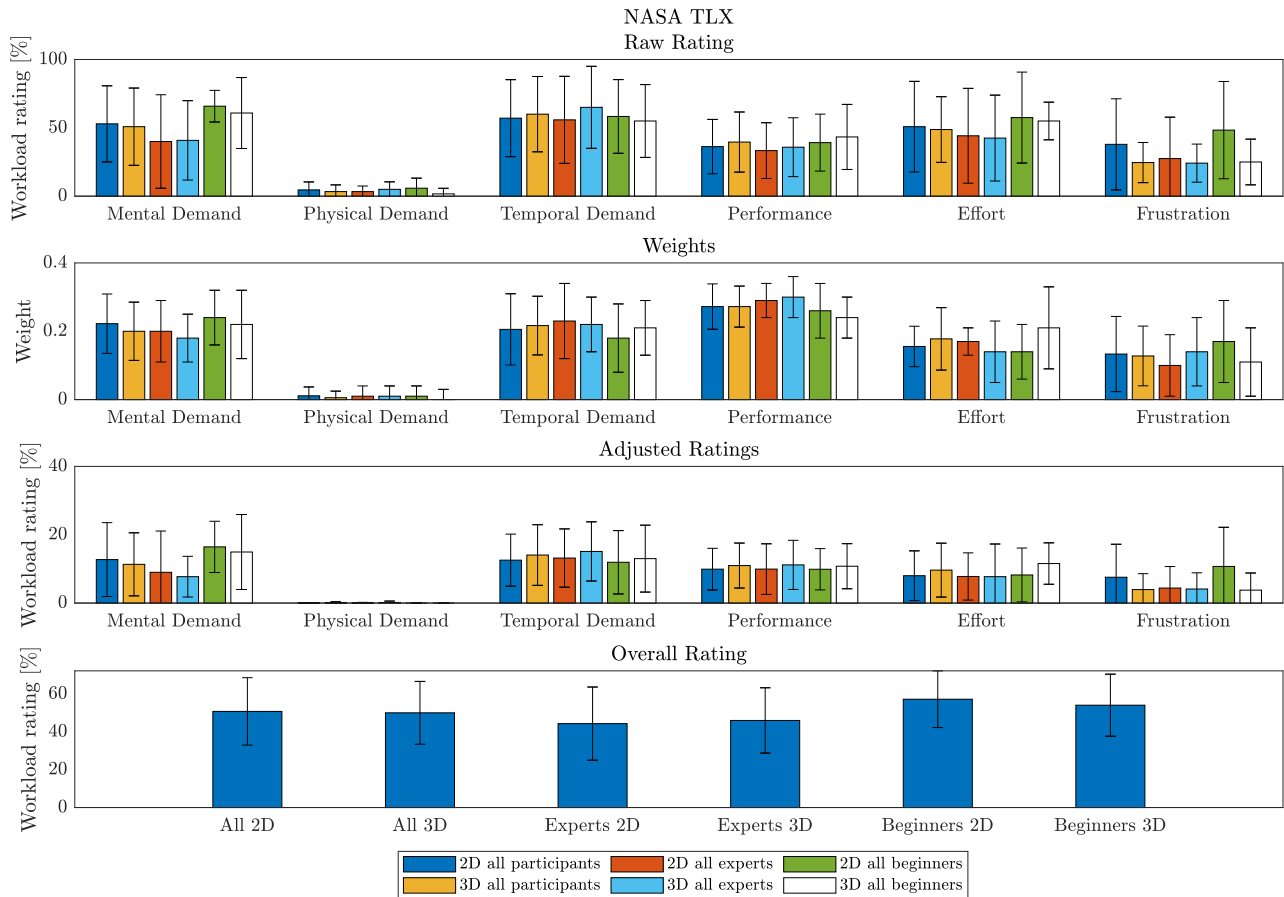2) Adding a possibility to control multiple joints simultaneously in joint control mode.

FIGURE 26. NASA TLX results overview for 2D and 3D interfaces, for all participants, experts, and beginners groups. The upper left chart presents the raw ratings. The upper right chart shows the weights attributed to the ratings. The lower left chart presents the adjusted ratings. The lower right chart shows the overall ratings. The scale for ratings was 0-100%, and for the weight, 0-1.
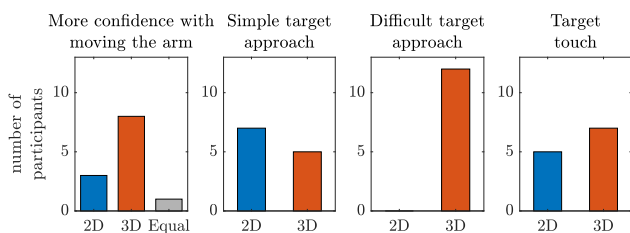
FIGURE 27. Operators' confidence comparison and choice of interface for three tasks with the 2D and 3D interfaces. The 3D interface received more votes in terms of confidence during arm movement and for a difficult approach task.

3) Making the transition to control like with the 2D interface easier (smooth switching between 2D and 3D interface depending on the task).
4) Improving point cloud visualization with surface reconstruction.
5) The operator's attention should be drawn to the physiological state display in the Head-Up Display more when there is stress.

Regarding the OMS, in the opinion of most participants, the system correctly measured their heartbeat and respiration signals. Only sometimes, the respiration value was prone to noise from movements or speaking. The participants observed no clear correlation between EDA and stress.

V. DISCUSSION

The discussions in this section focus on NASA TLX assessment and questionnaires methodology considerations, the concluded need for metrics in the training tool for operator training, and the OMS.

A. NASA TLX ASSESSMENT AND QUESTIONNAIRES METHODOLOGY

The NASA TLX methodology showed some differences between the interfaces. However, it did not fully reflect the complexity and factors contributing to the workload in the telerobotic context and could only serve as a general indicator. The specific contributing factors in NASA TLX subscales must have been deducted from participants' additional questionnaires and their descriptive feedback during interviews after the experiments. For example, the Mental Demand in the 3D interface was decreased by a higher spatial awareness and 3D representation of the robot, which helped to understand the pose. But the workload increased by a multiplied number of keyboard bindings necessary to remember or look up.

Using the NASA TLX standard protocol gave too much room for participants' interpretation of the Effort or Performance subscales. For example, Performance was understood as the minimum time needed to finish a task, the safety of the operation, or the smoother control and sureness of all movements.

Therefore, the next attempt would be to create a more precise human-robot interface workload estimation with interface-specific calibrated questions reflecting the areas of potential improvement. The additional criteria could be focused on the interface inputs' workload (to check if using a joystick, space mouse, gestures, or hand tracking is easier for a particular task). In a hazardous environment, safety is a higher priority than the time of execution. Therefore, an estimation of how much safety assurance and what usability tools (such as collision avoidance or detection) the interface provides is needed. The operator's reactions and an interface's adaptation to unexpected situations should be measured. In limited network scenarios, communication and feedback delays are a source of frustration and stress. Hence, an estimator of how the interface was coping with such a problem is crucial. For example, the interface could automatically adjust speed to the delay or modify feedback quality to minimise the delay and avoid the network collapse; or have more supervisory control, such as position-based trajectories, which are more resistant to unstable communication. All these techniques can minimise the operator's workload by explicitly addressing the delay problem that contributed to Mental and Temporal Demand, Effort and Frustration in the NASA TLX assessment.

B. OPERATOR TRAINING AND NEED FOR METRICS IN THE TRAINING TOOL

A person must undergo training before becoming an independent robot operator. During this process, usually, there is an expert teaching and giving feedback on the trainee's actions. However, when the process is long and complex, additional training tools such as training simulators are needed for skill perfection and for trying out different control techniques or scenarios by the newly trained operator.

In the performed experiments, some operators reacted in a time-based competitive manner, which could have provoked collisions at high velocities, or the manipulator was moved in close vicinity of the environment instead of keeping a safe distance and making the approach in the last phase. Hence, in the training tool, there is the need to create metrics that give feedback on how safe the teleoperation was, taking into account the distances to collisions, near-misses, or safe adjustment of speed in the vicinity of objects. These metrics can be easily calculated in simulation with a virtual robot and objects. When a real robot is used for training, the robot's modules' distances to the captured point cloud environment can be calculated. The number of keyboard key presses or mouse clicks could be counted for the performance quantification of the input systems. It was noticed in the experiments that the number of presses on a joint movement key was higher for beginner operators. If the feedback delay

became high, operators changed to the control strategy of short pressing and waiting for video feedback to check the effect. This significantly increased the task time. There are also other methods for performance calculation in human-computer interaction. For example, Fitt's law was used to assess the operator performance using virtual fixtures [53], or to analyse perceptual-motor performance within teleoperated or telepresence systems [54], [55].

C. THE OPERATOR MONITORING SYSTEM

The considerations regarding the OMS design, its future advancements and improvements were drawn from the practical use during experiments by multiple participants and are the following:

1) The ECG acquisition provided a good signal/noise ratio, and the R-peaks had large amplitudes compared to the surrounding data points. The ECG acquisition was reliable when the operator focused on the manipulation.

2) The RR measurements provided good feedback when the person had a stable breathing pattern, for example, when focused on the task. However, speaking or movements could be a source of chest movements that interfered with breaths. Therefore this parameter should not be used when the operator must use speech or move during the remote control. Or a specific algorithm recognising these interferences is needed to disable the measurement. Also, the physiological differences between operators required some tuning of thresholds when a breath is recognised and adjusting this parameter to each person or finding a typical value for a group of operators in the team.

3) The EDA measurement was the most experimental, but its feedback fulfilled its role in most cases, showing more activity when the task became more intense. However, the skin activity varied among participants, and threshold tuning was necessary and difficult. Similarly to other signals, it was susceptible to movements and friction between the electrodes or wires and clothing.

4) It was observed that sometimes the OMS's fuzzy logic calculated state did not precisely reflect the operator's feeling. It may have been caused by signal noise caused by movement, changing a position on a chair, contact of electrodes or cables with clothes, or simply talking.

5) Measurements of the baseline of the heartbeat, respiration, and EDA signals varied depending on the person's recent activity. Therefore, this measurement must be standardised, with a longer resting period, after a fixed-time low-activity work at a computer without stressors.

6) All used sensors require an attachment of wired electrodes to the body or wearing a respiration belt, which can generally interfere with tasks such as driving, walking or gestures. During the experiments, a few participants gave feedback that although the system did

not deteriorate the robot's remote control capabilities, it was not entirely comfortable to wear due to wires.

7) During the experiments, there were no simulated stressful situations (such as increased delay, robot not responding, disconnection, alert message, turning off the light, or collision sound) due to the complexity and dangers while using an operational robot in the LHC. However, for future system testing, these simulations would be an indispensable method for validation, especially for the EDA measurements.

VI. CONCLUSION

The telerobotic tasks in hazardous, radioactive, underground and semi-structured environments are a source of increased operator stress and workload compared to non-hazardous environments. Therefore, the design, development and evaluation of adequate human-centred robot interfaces play an essential role in the success of reliable and safe missions. In this publication, two contributions aimed to improve this evaluation process and steer the subsequent developments of human-robot interfaces: the Operator Monitoring System (OMS) and the assessment methodology.

Currently, in most human-machine systems, the state of the human is neglected (e.g., by using only a dead man's switch to check if the robot's or train's operator is conscious). The proposed approach of the OMS can significantly improve human inclusion in the system. The OMS and its experimental findings can be practically applied in any mission-critical human-robot interface, in which the operator's state should be monitored and the control adapted to the attention and stress. For example, it could be applied to surgical robots operated by doctors, or unmanned public and commercial transport vehicles controlled by remote operators (metro, drones, aeroplanes, semi-autonomous cars or machinery).

Often, developments assume that new or changed functionalities in an interface will improve the user experience (UX). However, a well-prepared iterative investigation with proper methods can steer this development better. The showcased customised methodology of assessing and comparing human-robot interfaces can be used directly or inspire applications where one interface must be evaluated or compared with a previous or a new novel interface.

The experimental application of the methodology and the OMS was done with 12 participants operating a real robot with a 9 DOF redundant manipulator in the scenario of the world's biggest and most complex machine, the LHC. The 3D MR interface was compared with the 2D interface, which has been widely used during the last 9 years in more than 160 real interventions, 500 performed tasks and 500 hours of operation. The results showed that the operations with the 3D MR on screens increased the safety and performance indicators while maintaining a similar workload and physiological response of operators. The experiments also studied intrinsic human factors, such as experience in remotely controlling robots or gaming, and the results can be used in operator selection and training processes.

Hypothesis 1 was confirmed. Improved understanding of the robotic manipulator thanks to the 3D representation and better perception of the environment with spatial point clouds were confirmed by detailed feedback from the operators. The NASA TLX assessment method did not reveal significant differences in the assessed workload comparison between interfaces.

Hypothesis 2 was confirmed. The 3D interface resulted in a similar physiological response to the 2D interface. Therefore, it can be concluded that the 3D interface did not cause higher physiological demand for operators.

Hypothesis 3 was partially confirmed. With the MR 3D interface, the execution of the approach task and learning process were 25-42% faster, especially for beginner operators. However, the touch task required more time for expert operators with the 3D interface, although the time difference became small (3%) in the last attempt. The difference in the previous familiarity with the 2D interface of 5 participants compared to only 1 participant who operated the 3D interface before may have been the main factor in favour of the 2D interface.

Hypothesis 4 was confirmed. Operators with more gaming experience executed tasks around 30% faster. Moreover, they had steeper learning curves with the 3D interface than with the 2D.

VII. FUTURE WORK

Future work regarding the methodologies of interface evaluation and metrics:

- For interface comparisons and human-robot interface workload assessment, a specific assessment methodology must be designed with interface-specific calibrated questions for hazardous scenarios and remotely operated robots, which would assess the workload due to more specific sources, such as inputs, safety assurance, network delays and uncertainty of robot's status.
- In the training process without an expert trainer or during the perfecting phase, there is a need for a tool that can give feedback and metrics on teleoperation safety. The boundaries of safe speeds and control techniques depending on situation should be also communicated by the system to the trainee.
- In the used setup for the experiments, the operator's eyes' focus was not used. However, it can be precisely measured with head-mount devices (HMDs) such as those for Virtual Reality (VR) or Augmented Reality (AR) stereoscopic displays. This information can be used to check if the operator is currently looking at the operated robot and the scene and adjust the behaviour of the interface. This will be implemented in the MR human-robot interface that uses the AR HMD, which has been recently commissioned at CERN [56], [57].

Future work regarding the Operator Monitoring System:

- Machine learning techniques and Artificial Intelligence (AI) could be applied to better recognize non-standard situations, which are not easily filtered by

analytic signal processing. Also, stress can manifest itself in different patterns under different circumstances and external conditions, which should be deeply studied. Further study must focus on a detailed estimation and quantification of particular sensors' stress recognition trust levels and testing algorithms with more participants.

- Initial tests of the OMS have been performed with the interfaces in spatial VR and AR human-robot interfaces using HMDs, where the operators must walk and use hand gestures or voice control. In such applications, the wired connections are troublesome and cause signal noise. Therefore a non-invasive heartbeat and respiration measurement should be used. Standalone contactless monitoring using radar and cardiac activity estimation with a pan-tilt-zoom camera has already been developed and successfully tested at CERN, as presented in [34] and [35], and must be integrated with the OMS.
- Other signals estimating stress and workload, such as eye tracking or electrical brain activity with electroencephalography (EEG), will be tested.

APPENDIX
CUSTOM QUESTIONNAIRE - RESPONSES

1) *Which tasks or actions were easier with the 2D interface compared to the 3D interface?*

- The 2D interface allowed the movement of multiple joints simultaneously in the real-time direct joint by joint control mode. Therefore, in this mode, it was faster to move the arm by usually combining two joints simultaneously, especially for experts and for repetitive tasks or simple trajectories that needed only a few movements with joints in a known environment.
- In the 2D interface, the video camera canvases were always visible, while in 3D, it was sometimes necessary to change the player's position to see the video canvas better. Therefore, the final part of the touch task, which had to be based more on video camera feedback and multiple movements with small increments, was faster with the 2D interface. In this task phase, there was less risk of collisions with the other arm elements, so the 3D collision avoidance was not much in use.
- The 2D interface was more intuitive because it had simpler controls. On the other hand, it lacked the advanced functionalities of the 3D interface.

2) *Which tasks or actions were easier with the 3D interface compared to the 2D interface?*

- For beginner operators, the 3D interface allowed straightforward learning of each joint movement and interpretation of the arm's pose. With the 3D representation of the arm, there was no need to look at the video camera to understand the arm's pose and prevent self-collisions.

- The estimation of distances and orientation of the end-effector was easier thanks to the additional information based on the point cloud. It benefited the approach task to keep a specific distance from the target, and the control was more precise.
- It was easier to visualize commands with a preview showing the trajectory, planning functions, and position-based control functionalities. The 3D interface allowed to ''jump'' to the target location with the point cloud normal point functionality and the planning arm. It did not require going joint by joint. With the preview, safety was assured, and then the arm could be moved directly to the target with one button and under supervision.
- It was manageable to avoid a collision. The preview functionality increased trust. Collision detection helped to know when the arm was in contact with itself or the target, especially when controlling the arm in real-time.
- The 3D interface gave better immersion in the entire space and the definition of obstacles. The combination of point clouds and video streams allowed a complete view of the area. The operator could better feel where the robot was in the environment (better prioperception). With the 2D interface, operators spent more time planning and checking movements with a restricted view.

3) *2D interface: What problems did you encounter?*

- The only feedback available was the video stream, and when the communication delays were high, it was difficult to operate with significantly delayed video feedback. With high delays, the control felt unresponsive.
- Due to a camera perspective, estimating the distance to the target was difficult. Indirect visual clues, such as shadows in the environment, must have been used.
- The joint and Cartesian movement had no indicators of joint direction. Initially, the trial and error method had to be used to understand the movement directions.
- For beginner operators, no understanding of the arm's pose and environment and how each joint moves was a source of stress to use the system safely. There was always a collision possibility, which could not have been noticed.

4) *3D interface: What problems did you encounter?*

- The 3D interface had advanced functionalities, but the input system and key bindings were difficult to remember, although there was a screen to look them up. It had too many modes of control. However, closer to the end of the experiment (after ~2 hours of practice), most participants remembered the key bindings fluently without external help.

- In the beginning, the locomotion around the scene was not intuitive for operators without gaming experience (the QWEASD keys, SHIFT to accelerate, and mouse to look around or interact). Therefore, there was an additional workload of moving the viewer (player) in the scene in addition to controlling the robot. Another input device, such as a space mouse used often in the 3D design software, could be better for player movement.
- With multiple functionalities used simultaneously (normal point, planning arm, trajectories, FABRIK end-effector control, collisions, torque arrows), the display became too cluttered, and it wasn't easy to see the environment.
- Representation of the point cloud as points but not surfaces was not natural, although it was possible to see the shapes. Sometimes, using the normal point required much time to define the target. A more intuitive method of changing its position and orientation would be better.
- The planning with the inverse kinematics and FABRIK mode allowed for swift deployment, but an inappropriate use could place the arm in a strange configuration. It required more practice before it could be used efficiently.
- For a few participants, the use of the point cloud for the touch determination was not always reliable due to point cloud resolution (for example, 10 mm) and system mechanical elasticity - the wagon with the arm could tilt, and the previous point cloud reading became shifted in comparison to the actual position. The video feedback for that task was more reliable in seeing the arm-environment contact.

5) *Did the vital parameters displayed in the interface correspond to your stress, workload or feelings?*
   - Most participants answered: "The heartbeat matched the real feeling. It corresponded to my attempts to improve the time and performance."
   - "The respiration was usually correct. But when I was moving, it was overestimated."
   - Four participants did not pay much attention to the indicators. They were focused more on the tasks.
   - Participants could not tell if the EDA measurement was correct.

6) *If you could choose which interface, 2D or 3D, for different tasks, which one would you choose? With which interface were you more confident about the movement of the arm?*
   The 3D interface received higher confidence primarily due to positioning with the 3D model showing the arm position, especially when the network delay was longer. Operators who voted for more confidence with the 2D interface explained that it was more straightforward with less complex control modes or they had already been familiar with it. The 3D interface was selected

by everybody for a difficult approach task, while 2D and 3D received a similar number of votes for a simple approach or the touch task.

7) *After operating with the 3D interface, did you understand the movement better in the 2D interface?* Most participants (7) answered "definitely yes", apart from the operators who already knew and operated the arm with the 2D interface (5 participants).

REFERENCES

[1] J. Y. C. Chen and M. J. Barnes, "Human–agent teaming for multirobot control: A review of human factors issues," *IEEE Trans. Human-Machine Syst.*, vol. 44, no. 1, pp. 13–29, Feb. 2014.

[2] N. J. Currie and B. Peacock, "International space station robotic systems operations—A human factors perspective," *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, vol. 46, no. 1, pp. 26–30, Sep. 2002, doi: 10.1177/154193120204600106.

[3] J. K. Portello, M. Rosenfield, Y. Bababekova, J. M. Estrada, and A. Leon, "Computer-related visual symptoms in office workers," *Ophthalmic Physiol. Opt.*, vol. 32, no. 5, pp. 375–382, Sep. 2012. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-1313.2012.00925.x

[4] M. M. Bragatto, D. Bevilaqua-Grossi, S. C. H. Regalo, J. D. Sousa, and T. C. Chaves, "Associations among temporomandibular disorders, chronic neck pain and neck pain disability in computer office workers: A pilot study," *J. Oral Rehabil.*, vol. 43, no. 5, pp. 321–332, May 2016. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/joor.12377

[5] C. L. Cooper and J. Marshall, "Occupational sources of stress: A review of the literature relating to coronary heart disease and mental ill health," *From Stress Wellbeing*, vol. 1, pp. 3–23, Jan. 2013.

[6] 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," in *Proc. 15th Int. Conf. Informat. Control, Autom. Robot. (ICINCO)*, vol. 2, 2018, pp. 283–290.

[7] L. B. Rosenberg, "Virtual fixtures: Perceptual tools for telerobotic manipulation," in *Proc. IEEE Virtual Reality Annu. Int. Symp.*, Sep. 1993, pp. 76–82.

[8] D. Markov-Vetter, J. Millberg, and O. Staadt, "Mobile augmented reality for space operation procedures: A generic approach of authoring and guiding on-board payload activities," in *Proc. Int. Astron. Congr. (IAC)*, vol. 6, Jan. 2013, pp. 1–14.

[9] A. Martín-Barrio, J. J. Roldán, S. Terrile, J. del Cerro, and A. Barrientos, "Application of immersive technologies and natural language to hyper-redundant robot teleoperation," *Virtual Reality*, vol. 24, no. 3, pp. 541–555, Sep. 2020, doi: 10.1007/s10055-019-00414-9.

[10] 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.

[11] M. Ostanin, S. Mikhel, A. Evlampiev, V. Skvortsova, and A. Klimchik, "Human-robot interaction for robotic manipulator programming in mixed reality," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 2805–2811.

[12] K. Szczurek, R. Prades, E. Matheson, H. Perier, L. Buonocore, and M. Di Castro, "From 2D to 3D mixed reality human-robot interface in hazardous robotic interventions with the use of redundant mobile manipulator," in *Proc. 18th Int. Conf. Informat. Control, Autom. Robot. (ICINCO)*, 2021, pp. 388–395.

[13] J. A. Waxenbaum, V. Reddy, and M. Varacallo, *Anatomy, Autonomic Nervous System*. Treasure Island, FL, USA: StatPearls Publishing, 2022. [Online]. Available: http://europepmc.org/books/NBK539845

[14] M. Mestanik, A. Mestanikova, Z. Visnovcova, A. Calkovska, and I. Tonhajzerova, "Cardiovascular sympathetic arousal in response to different mental stressors," *Physiol. Res.*, vol. 64, pp. S585–S594, Dec. 2015.

[15] M. N. Alshak and J. M. Das, *Neuroanatomy, Sympathetic Nervous System*. Treasure Island, FL, USA: StatPearls Publishing, 2022. [Online]. Available: http://europepmc.org/books/NBK542195

[16] M. N. Jarczok, M. Jarczok, and J. F. Thayer, "Work stress and autonomic nervous system activity," in *Handbook of Socioeconomic Determinants of Occupational Health: From Macro-Level to Micro-Level Evidence*. Cham, Switzerland: Springer, 2020, pp. 1–33.

[17] C. M. A. Panigrahi, "Managing stress at workplace," *J. Manage. Res. Anal.*, vol. 3, no. 4, pp. 154–160, 2016.

[18] P. Arpaia, N. Moccaldi, R. Prevete, I. Sannino, and A. Tedesco, "A wearable EEG instrument for real-time frontal asymmetry monitoring in worker stress analysis," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 10, pp. 8335–8343, Oct. 2020.

[19] T. Maurya, K. Karena, H. Vardhan, M. Aruna, and M. G. Raj, "Effect of heat on underground mine workers," *Proc. Earth Planet. Sci.*, vol. 11, pp. 491–498, Jan. 2015.

[20] C. Zhao, M. Zhao, J. Liu, and C. Zheng, "Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator," *Accident Anal. Prevention*, vol. 45, pp. 83–90, Mar. 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0001457511003241

[21] R. Buendia, F. Forcolin, J. Karlsson, B. A. Sjöqvist, A. Anund, and S. Candefjord, "Deriving heart rate variability indices from cardiac monitoring—An indicator of driver sleepiness," *Traffic Injury Prevention*, vol. 20, no. 3, pp. 249–254, Mar. 2019, doi: 10.1080/15389588.2018.1548766.

[22] N. Maeda, Y. Hirabe, Y. Arakawa, and K. Yasumoto, "COSMS: Unconscious stress monitoring system for office worker," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput., Adjunct*, Sep. 2016, pp. 329–332.

[23] S. Sarkar, A. Ghosh, and S. S. Ghosh, "Study of cardiorespiratory and sweat monitoring wearable architecture for coal mine workers," in *Proc. IEEE REGION Conf. (TENCON)*, Nov. 2020, pp. 355–360.

[24] C. S. S. Tan, J. Schöning, K. Luyten, and K. Coninx, "Investigating the effects of using biofeedback as visual stress indicator during video-mediated collaboration," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 2014, pp. 71–80.

[25] I. K. Waterhouse and I. L. Child, "Frustration and the quality of performance," *J. Personality*, vol. 21, p. 298, Mar. 1953. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-6494.1953.tb01773.x

[26] J. Lazar, A. Jones, and B. Shneiderman, "Workplace user frustration with computers: An exploratory investigation of the causes and severity," *Behav. Inf. Technol.*, vol. 25, no. 3, pp. 239–251, May 2006, doi: 10.1080/01449290500196963.

[27] L. Antao, R. Pinto, J. Reis, G. Gonçalves, and F. L. Pereira, "Cooperative human-machine interaction in industrial environments," in *Proc. 13th APCA Int. Conf. Control Soft Comput. (CONTROLO)*, Jun. 2018, pp. 430–435.

[28] P. Rani, J. Sims, R. Brackin, and N. Sarkar, "Online stress detection using psychophysiological signals for implicit human-robot cooperation," *Robotica*, vol. 20, no. 6, pp. 673–685, Nov. 2002.

[29] F. J. Badesa, J. A. Diez, J. M. Catalan, E. Trigili, F. Cordella, M. Nann, S. Crea, S. R. Soekadar, L. Zollo, N. Vitiello, and N. Garcia-Aracil, "Physiological responses during hybrid BNCI control of an upper-limb exoskeleton," *Sensors*, vol. 19, no. 22, p. 4931, Nov. 2019. [Online]. Available: https://www.mdpi.com/1424-8220/19/22/4931

[30] E. Yang and M. C. Dorneich, "The emotional, cognitive, physiological, and performance effects of variable time delay in robotic teleoperation," *Int. J. Social Robot.*, vol. 9, no. 4, pp. 491–508, Sep. 2017.

[31] C. Lauretti, F. Cordella, C. Tamantini, C. Gentile, F. S. D. Luzio, and L. Zollo, "A surgeon-robot shared control for ergonomic pedicle screw fixation," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 2554–2561, Apr. 2020.

[32] L. Evans, "Particle accelerators at CERN: From the early days to the LHC and beyond," *Technological Forecasting Social Change*, vol. 112, pp. 4–12, Nov. 2016.

[33] M. F. Alam, E. Adamidi, and S. Hadjiefthymiades, "Wireless personnel safety system (WPSS), a baseline towards advance system architecture," in *Proc. 18th Panhellenic Conf. Informat.*, Oct. 2014, pp. 1–6, doi: 10.1145/2645791.2645855.

[34] A. Ivanovs, A. Nikitenko, M. Di Castro, T. Torims, A. Masi, and M. Ferre, "Multisensor low-cost system for real time human detection and remote respiration monitoring," in *Proc. 3rd IEEE Int. Conf. Robotic Comput. (IRC)*, Feb. 2019, pp. 254–257.

[35] R. Cittadini, L. R. Buonocore, E. Matheson, M. Di Castro, and L. Zollo, "Robot-aided contactless monitoring of workers' cardiac activity in hazardous environment," *IEEE Access*, vol. 10, pp. 133427–133438, 2022.

[36] 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.

[37] 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.

[38] M. Di 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," in *Proc. IEEE Int. Symp. Saf., Secur., Rescue Robot. (SSRR)*, Aug. 2018, pp. 1–6.

[39] J. Y. C. Chen and M. J. Barnes, "Supervisory control of multiple robots: Effects of imperfect automation and individual differences," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 54, no. 2, pp. 157–174, Apr. 2012, doi: 10.1177/0018720811435843.

[40] C. Green and D. Bavelier, "Enumeration versus multiple object tracking: The case of action video game players," *Cognition*, vol. 101, no. 1, pp. 217–245, Aug. 2006.

[41] B. Hubert-Wallander, C. S. Green, and D. Bavelier, "Stretching the limits of visual attention: The case of action video games," *Wiley Interdiscipl. Rev., Cogn. Sci.*, vol. 2, no. 2, pp. 222–230, Mar. 2011.

[42] I. Spence and J. Feng, "Video games and spatial cognition," *Rev. Gen. Psychol.*, vol. 14, no. 2, pp. 92–104, Jun. 2010.

[43] 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.

[44] K. A. Szczurek and R. Cittadini. *Cern Robot Operator Physiological Parameters Monitoring System—Project Code Source*. Accessed: Mar. 24, 2023. [Online]. Available: https://gitlab.cern.ch/kszczure/cern-robot-operator-physiological-parameters-monitoring-system

[45] B. E. Jin, H. Wulff, J. H. Widdicombe, J. Zheng, D. M. Bers, and J. L. Puglisi, "A simple device to illustrate the Einthoven triangle," *Adv. Physiol. Educ.*, vol. 36, no. 4, pp. 319–324, Dec. 2012.

[46] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biol. Psychol.*, vol. 84, no. 3, pp. 394–421, 2010.

[47] R. E. De Meersman, "Heart rate variability and aerobic fitness," *Amer. Heart J.*, vol. 125, no. 3, pp. 726–731, Mar. 1993. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0002870393901645

[48] H. Kaneko and J. Horie, "Breathing movements of the chest and abdominal wall in healthy subjects," *Respirat. Care*, vol. 57, no. 9, pp. 1442–1451, Sep. 2012.

[49] A. Joshua, L. Shetty, and V. Pare, "Variations in dimensions and shape of thoracic cage with aging: an anatomical review," *Anatomy J. Afr.*, vol. 3, pp. 346–355, Jul. 2014.

[50] E. Lutin, R. Hashimoto, W. De Raedt, and C. Van Hoof, "Feature extraction for stress detection in electrodermal activity," in *Proc. 14th Int. Joint Conf. Biomed. Eng. Syst. Technol.*, vol. 2, 2021, pp. 177–185.

[51] F. J. Badesa, R. Morales, N. M. Garcia-Aracil, J. M. Sabater, L. Zollo, E. Papaleo, and E. Guglielmelli, "Dynamic adaptive system for robot-assisted motion rehabilitation," *IEEE Syst. J.*, vol. 10, no. 3, pp. 984–991, Sep. 2016.

[52] M. Egger, M. Ley, and S. Hanke, "Emotion recognition from physiological signal analysis: A review," *Electron. Notes Theor. Comput. Sci.*, vol. 343, pp. 35–55, May 2019.

[53] L. B. Rosenberg. (Mar. 1993). *The Use of Virtual Fixtures to Enhance Operator Performance in Time Delayed Teleoperation*. [Online]. Available: https://apps.dtic.mil/sti/citations/ADA296363

[54] J. W. Hill. (Mar. 1979). *Study of Modeling and Evaluation of Remote Manipulation Tasks With Force Feedback*. [Online]. Available: https://ntrs.nasa.gov/citations/19790018614

[55] R. L. Pepper and P. K. Kaomea, "Teleoperation: Telepresence and performance assessment," in *Proc. Annu. Int. Ergonom. Soc. Meeting, Teleoperation Control*, 1988, pp. 227–234.

[56] K. A. Szczurek, R. M. Prades, E. Matheson, J. Rodriguez-Nogueira, and M. D. Castro, "Multimodal multi-user mixed reality human–robot interface for remote operations in hazardous environments," *IEEE Access*, vol. 11, pp. 17305–17333, 2023.

[57] 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

KRZYSZTOF ADAM SZCZUREK received the M.Sc. 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 Jaume I University, Spain. From 2013 to 2014, he was with American Axle & Manufacturing on industrial controls for the automotive sector. In 2017, he was 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 and space robotics, and specializes in designing and implementing control systems consisting of complex and state-of-the-art solutions.

ROBERTO CITTADINI received the B.Sc. degree in medical engineering from the University of Rome Tor Vergata, in 2017, and the M.Sc. degree in biomedical engineering and robotics from the Campus Bio-Medico University of Rome, Italy, in 2020. He is currently pursuing the Ph.D. degree with the CERN and the Advanced Robotics and Human-Centered Technologies Research Unit, Campus Bio-Medico University of Rome. In 2020, he joined CERN's Mechatronics, Robotics and Operations Section, as a Cooperation Associate (COAS) Engineer, working on health contactless monitoring of physiological parameters using cameras, radars, and distance sensors. His research interests include worker monitoring in harsh environments, physiological signal analysis, and human-centered technologies.

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

ELOISE MATHESON (Member, IEEE) received the B.Sc./B.Eng. degree in mechatronics (space) engineering from The University of Sydney, Australia, in 2010, the M.Sc. degree in advanced robotics from 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 her Ph.D. degree 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 with the Telerobotics and Haptics Laboratory, European Space Agency, The Netherlands, largely working on telerobotic activities under the METERON Project. In 2020, she joined CERN's Mechatronics, Robotics and Operations Section 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 teleoperation, supervisory control and autonomous operations, haptics, and human–machine interfaces.

MARIO DI CASTRO received 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 with CERN in charge of advanced magnetic measurements and studies for LHC superconducting magnets. From 2007 to 2011, he was with EMBL c/o DESY in charge of advanced mechatronics solutions for synchrotron beamlines controls. Since 2011, he has been with 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 targets) in harsh environments. 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, telerobotics, human–robot interfaces, machine learning, enhanced reality, automatic controls, mechatronics, precise motion control in harsh environments, and advanced robotics for search and rescue scenarios.

● ● ●

Chapter 5

Summary

This chapter summarises the results achieved in the thesis and published in the three journal papers. Figure 5.1 depicts the research and operational challenges of mobile telerobotics in CERN accelerators and experimentation locations and their respective solutions and strategies formulated in this thesis.

The first paper presented a framework optimising human-robot interfaces for hazardous teleoperation scenarios with Mixed Reality techniques. The redundancy and complexity of the mobile manipulator increase the risk of collisions, which can be mitigated with spacial models and point cloud-based collision avoidance. The proposed framework enabled the automatic setting of video and point cloud settings to optimise available bandwidth, reduce latency, and improve the operator's environmental awareness. The presented application context and algorithms were thoroughly characterized and tested to evaluate their specific use case, advantages, and how they reduced the workload. The presented solutions can be applied to multiple industrial fields where network parameters are limited and dynamically changing, and Mixed-Reality control using 3D environmental feedback is needed. The positive results obtained with the 3D MR interface on screens motivated the further implementation of this interface for the holographic AR HMD with hand, voice, and user motion control. These achievements are presented as the next journal paper.

The second paper presented the findings of extensive tests and measurements of the multi-user Mixed Reality human-robot interface using the AR HMD in remote teleoperation scenarios. Despite challenging conditions, including multifunctional user interface capabilities, network infrastructure limitations, and the interface's high reliability, operators performed stable AR remote teleoperations. The solution provided multimodal communication links via 4G mobile, Wi-Fi and cabled networks available at CERN. The network architecture was proposed to optimise its usage depending on connection bandwidth, delays, and executed robotic tasks. Improvements to the multi-user collaboration network architecture and other technical implementations, such as foveated rendering and focus tracking adapting acquisition parameters, were proposed to optimise the communication load further. A high-level autonomous behaviour applied to different interactions with the robot was recommended to lower the operator's workload.

The work done and presented in the previous two journal papers showed the need for a structured assessment methodology and comparisons of human-robot interfaces. Such assessments and comparisons are foundations of a good understanding of problematic areas and functionalities that should be improved in the interfaces. The novel MR techniques were never used in operational scenarios at CERN. Therefore, such feedback was necessary to check if the teleoperators could practically utilise the enhanced spatial awareness, supervisory control and functionalities to improve safety and if the user's workload and stress were not increased. In the
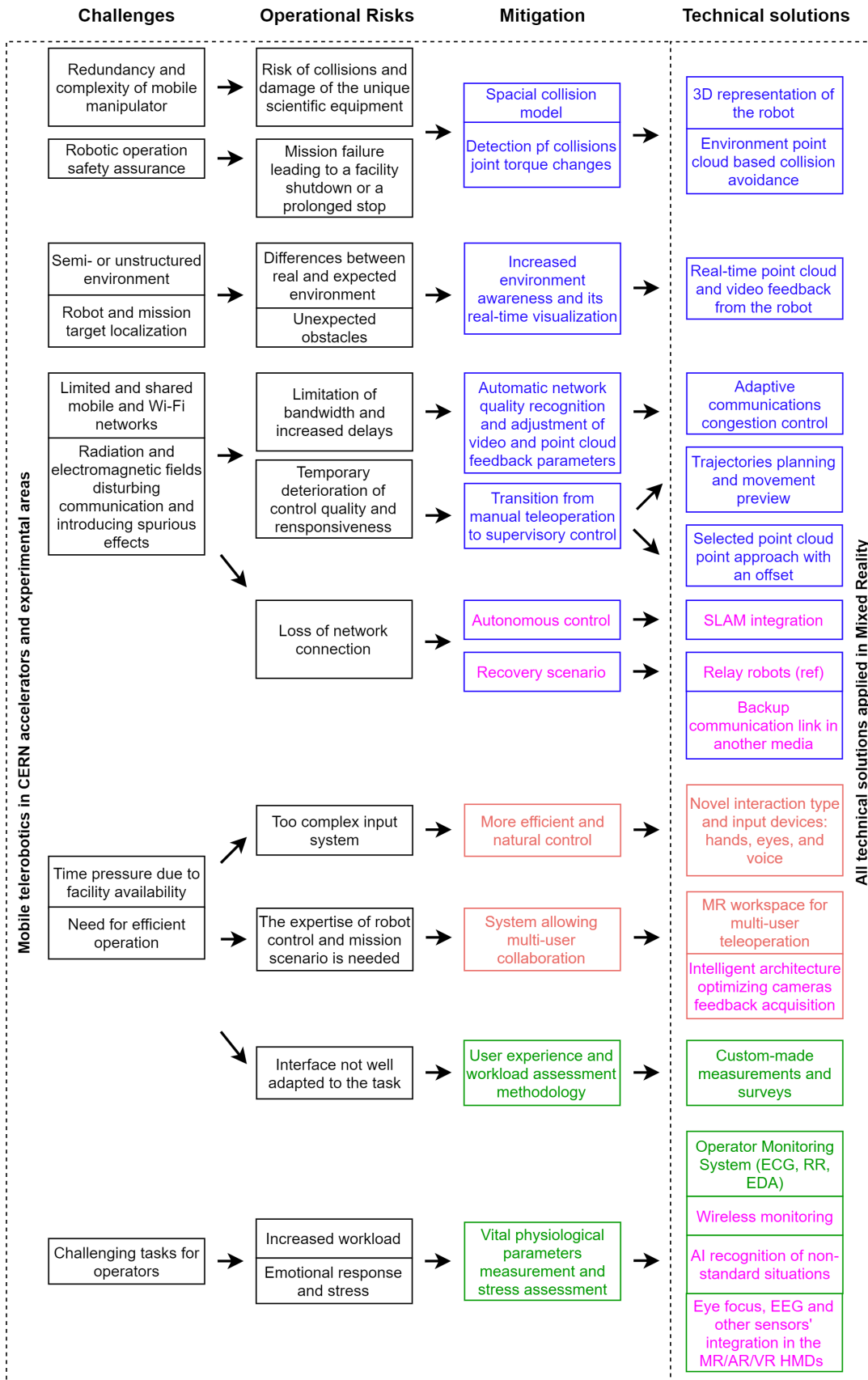
FIGURE 5.1: The diagram presents the challenges and operational risks, alongside their mitigation and technical solutions developed within this thesis, of mobile telerobotics in CERN accelerators and experimental areas. The colours in mitigation and technical solutions indicate the corresponding focus and novelties in the three presented before journal papers (Chapter 2, Chapter 3, Chapter 4. The main future work is also indicated, and described in Chapter 6.

third paper, the CERN 3D MR and the operational 2D interfaces were compared using the NASA TLX assessment method, custom questionnaires, task execution time curves, and the Operator Monitoring System (OMS) for measuring heart rate (HR), respiration rate (RR), and skin electrodermal activity (EDA). NASA TLX found no significant differences in workload assessment between the two interfaces. Expert operators had 45% less Mental Demand than beginners. The execution of precise approach tasks was faster (up to 42%) with the 3D MR interface. Operators with more gaming experience completed tasks 30% quicker and had a steeper learning curve. Custom questionnaires and interviews were the most helpful in understanding workload sources. OMS experiments showed that HR was the most reliable for tracking operator state, while RR and EDA were more noise-prone. The 3D interface produced a similar physiological response as the 2D. This method and comparative study identified critical improvements for both interfaces. Further advances were proposed, such as safety metrics and calibrated questions. The need for applying machine learning techniques in signal interpretation to detect non-standard situations and contactless monitoring technology was identified.

These three complementing publications resulted in research-based, novel and human-centred human-machine interaction advancements that took into account the challenges in hazardous environments. Specific safety requirements, such as unique scientific equipment protection from collisions, were addressed through increased spatial perception and automated collision detection and avoidance. Limiting communication network conditions were mitigated with adaptive interface behaviours. And the physiological state of the operator, who is an very important factor contributing to a mission's success, was included in the system.

Chapter 6

Conclusions

The research has been done to enhance the efficiency and safety of robotic operations in hazardous environments, with direct application in CERN particle accelerators and experimental underground areas. The novel 3D MR techniques for human-robot interfaces have been chosen to improve spatial awareness, user experience and safety. In the state of the art, this approach had not been applied in hazardous industrial environments for remote robot control, especially with MR HMDs. Therefore, this solution for CERN mobile robotics had to be evaluated. A methodology of interface workload and operator vital physiological parameters assessment was also developed, which proved to be a source of a good understanding of interfaces' problems and needs for improvements.

All the objectives listed in Section 1.2 have been achieved. The risks and challenges of telerobotics in hazardous environments were evaluated. The risks of collisions and damage to the unique scientific equipment were addressed by providing spacial collision detection and avoidance mechanisms based on the robot's geometry, joint torques and environmental feedback. Environmental awareness was increased by real-time 3D point cloud and video feedback. The volatility of network bandwidth limitation and increased delays were mitigated by automatically recognising network quality and adjusting video and point cloud feedback parameters. The interaction level was increased by transiting from manual teleoperation to the supervisory control of a complex redundant manipulator. Available MR HMDs and their applicability to human-robot interfaces were evaluated. The problem of a too complex and not intuitive input system was addressed by applying novel control with hand, locomotion, eye tracking, and voice inputs. A multi-user system in the MR workspace for the collaboration of multiple experts was proposed. A methodical assessment of user experience, workload, and vital physiological parameters and its comparison between the standard 2D interface and the developed 3D MR interface was performed. And notably, the research in the mentioned areas has brought multiple advances in the 2D, 3D and MR human-robot interfaces applied in real operational scenarios at CERN.

However, further research and optimisation must be done to mitigate some disadvantages, such as the increased complexity of the system compared to standard 2D interfaces. Based on the results, further developments bringing valuable gains in the multi-user MR human-robot interface technology with operator physiological parameters monitoring are proposed below (and marked in Figure 5.1).

- The operational risk of totally losing network connection must be addressed. Two solution groups are available: autonomous control or external recovery. In autonomous control, the next milestone is to achieve SLAM integration with the human-robot interface, and the operator has to specify the robot's actions if the connection is lost. In the recovery scenario, other robots must be used, or a backup communication link must be established. The communication link

should preferably use another medium, such as light or sound waves, in case of too much electromagnetic disturbance. Moreover, having another robot in the scenario would allow better spatial awareness thanks to an external viewpoint of the primary robot executing the task.

- Further optimisation of communication load must be pursued, especially in the multi-user architecture. In Chapter 3, an architecture was presented where each user created a separate connection to the robot. This led to an increase in throughput proportional to the number of users. Although it was observed that with two connections, the total bandwidth is only marginally greater than with one, this may not be enough in cases with more than two users and a low-bandwidth network. A single link from the robot to an interface multi-server should be implemented to improve the feedback quality. The multi-server would then further distribute data to each user already in the high-bandwidth layer of the network.

- In terms of the operator's physiological parameters monitoring system (OMS), two further improvements are proposed. Firstly, machine learning and AI methods should be applied to recognise non-conventional situations that are not easily identified through analytical signal processing. It can allow the use of the system for operators who must walk, talk and interact with the control interface in the MR workspace with hands and voice. Different types of stress can reveal themselves in various patterns under various conditions and external environments, which should be carefully investigated and addressed with AI methods. Secondly, non-invasive heart rate and respiration measurements should be used for the operator wearing a VR or AR headset. Contactless respiration monitoring by radar and cardiac activity estimation with a pan-tilt-zoom camera could be used, or an AR/VR HMD with integrated biosensors is required.

- The MR interface with an AR HMD was developed with the HoloLens 2 device. Evaluations of other devices providing Virtual, Augmented or Mixed realities for robotic operations must be done due to the fast evolution of HMD technology. New devices will provide a wider field of view, more rendering capabilities, better hologram quality, foveated rendering, or light-field displays to lower the eyes' accommodation and depth-matching discomfort.

6.1   Other achievements

6.1.1   Mixed Reality astronaut training

The PhD has been mainly conducted at CERN on MR human-robot interfaces to control robots in radioactive areas remotely. However, the research of MR techniques was also extended to astronaut training in robotics. Further experience was gained as the author's 5-month internship in the XR Lab of the European Space Agency (ESA) European Astronaut Centre (EAC) in Cologne, Germany. The work improved the Joint Investigation into VR for Education (JIVE) software (Figure 6.1 and assessed the benefits of MR by implementing the techniques with the state-of-the-art MR technology of the Varjo XR-3 MR HMD. The JIVE is a VR application with tools to aid robotics instructors in delivering astronaut lessons. It has been a collaborative project between EAC's XR Lab and the National Aeronautics and Space Administration (NASA) Johnson Space Center. The software has been provided to NASA
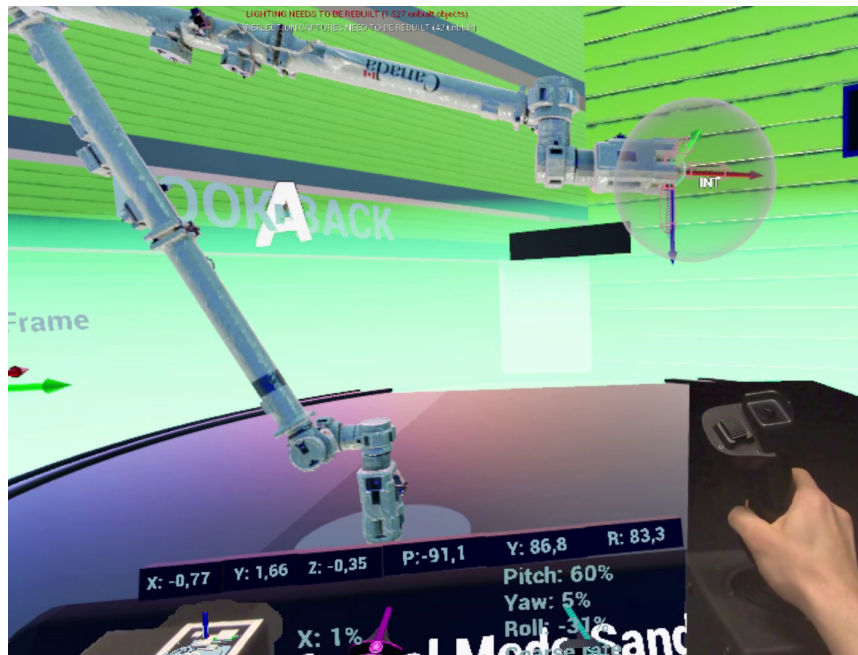
FIGURE 6.1: The JIVE teaches fundamentals of robotic arm control and allows astronauts to control a simulated arm in manual joint-by-joint modes or with inverse kinematics. In the picture, Mixed Reality shows real controllers and the user's hands blended with the Virtual Reality content.

and ESA and used for astronaut training in robotics. The work was a part of the improvement process using haptics and enhancing the overall immersive experience. The view in the Virtual Reality from the International Space Station Cupola observatory module, where the astronau operate the Canada robotic ar from, is shown in Figure 6.2. Another example of a training simulator developed for moonwalk is shown in Figure 6.3.

6.1.2 Knowledge transfer and entrepreneurial follow-up

During the R&D of the human-robot interface systems, the knowledge and expertise exchange took place with the following institutions:

- The Interactive and Robotic Systems Lab (IRSLab) from the Research Center for Robotics and Underwater Technology (CIRTESU) at the Jaume I University of Castellón. The experience was shared from the CERN underground remote robotics operations, and vice versa, the inspiration was taken from novel human interfacing and control methods for underwater robots;

- The XR-Lab at the European Astronaut Centre (EAC) of the European Space Agency (ESA) within the project described above;

- The European Space Operations Centre (ESOC) of the ESA to exchange know-how in space robotics and operations enhanced with the XR technology.

- The Lusospace company in Portugal in terms of multi-user interactions with AR technologies for concurrent mechanical design and assembly procedures guidance;

FIGURE 6.2: The International Space Station (ISS) robotic arm simulator from the Cupola observatory module. The work done by the author was focused on integrating the real robotic arm controllers in Mixed Reality.
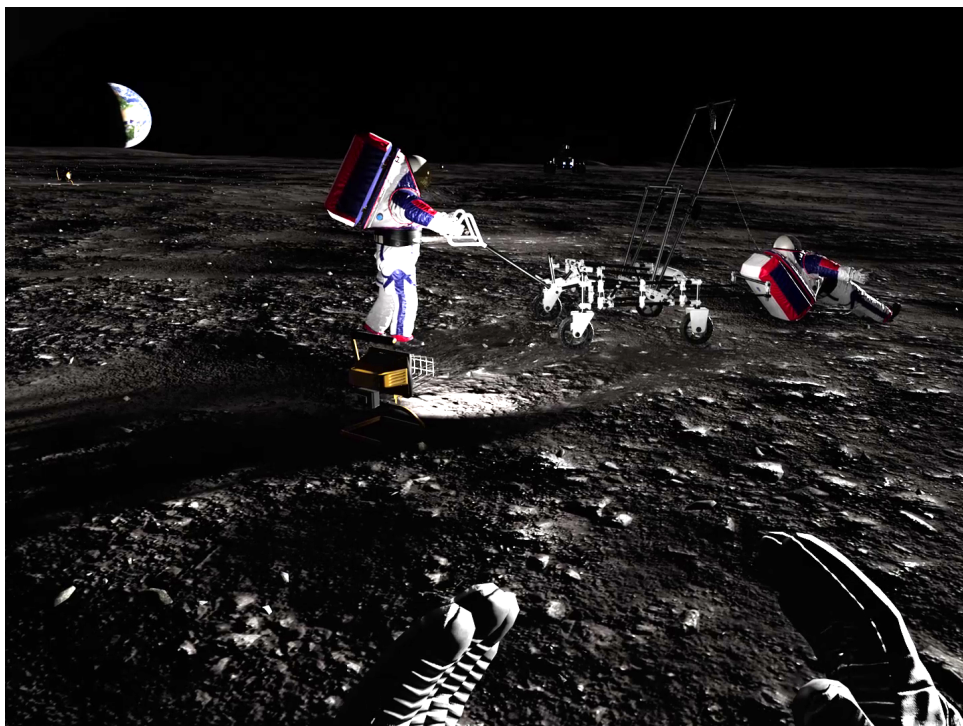

FIGURE 6.3: The moonwalk simulator developed in the XR Lab in the European Astronaut Centre. The Unreal game engine created realistic effects and physics with partial gravity and no atmosphere on the Moon. The astronaut candidates can interact with objects, moon rocks, spacesuits, and other equipment to train before the planned Artemis missions.

- The CREAL company in Switzerland in terms of Light-Field Display technology for VR and AR HMDs to understand how genuine focal depth can improve user experience in MR, especially while working with objects within arm's reach;

- Human-robot interfacing technology discussion was held with the European Spallation Source (ESS) in Sweden and the Joint European Torus (JET) operated by Culham Centre for Fusion Energy in the United Kingdom. The practical use of the VR4Robots commercial solutions or custom-made frameworks was considered.

Moreover, as an entrepreneurial initiative, the process was launched with the CERN Knowledge Transfer department to create a product or a startup utilizing the Mixed Reality robotics experience and toolset developed at CERN in potential commercial and industrial applications.

Bibliography

[1] *Telerob teodor specifications.* [Online]. Available: https://www.telerob.com/en/products/teodor-evo.

[2] *Telerob telemax specifications.* [Online]. Available: https://www.telerob.com/en/products/telemax-family.

[3] N. J. Currie and B. Peacock, "International space station robotic systems operations - a human factors perspective," vol. 46, 2002, pp. 26–30. DOI: 10.1177/154193120204600106.

[4] M. Bajracharya, M. W. Maimone, and D. Helmick, "Autonomy for mars rovers: Past, present, and future," *Computer*, vol. 41, pp. 44–50, 12 Dec. 2008, ISSN: 1558-0814. DOI: 10.1109/MC.2008.479.

[5] R. Skilton, N. Hamilton, R. Howell, C. Lamb, and J. Rodriguez, "Mascot 6: Achieving high dexterity tele-manipulation with a modern architectural design for fusion remote maintenance," vol. 136, 2018, pp. 575–578. DOI: https://doi.org/10.1016/j.fusengdes.2018.03.026.

[6] T. Sakaue, S. Yoshino, K. Nishizawa, and K. Takeda, "Survey in fukushima daiichi nps by combination of human and remotely-controlled robot," Oct. 2017, pp. 7–12. DOI: 10.1109/SSRR.2017.8088132.

[7] S. Kawatsuma, R. Mimura, and H. Asama, "Unitization for portability of emergency response surveillance robot system: Experiences and lessons learned from the deployment of the jaea-3 emergency response robot at the fukushima daiichi nuclear power plants," *ROBOMECH Journal*, vol. 4, p. 6, 1 2017, ISSN: 2197-4225. DOI: 10.1186/s40648-017-0073-7.

[8] R. Buckingham and A. Loving, "Remote-handling challenges in fusion research and beyond," *Nature Physics*, vol. 12, pp. 391–393, 5 2016, ISSN: 1745-2481. DOI: 10.1038/nphys3755.

[9] J. Adams, "Critical considerations for human-robot interface development," *AAAI Fall Symposium: Human Robot Interaction Technical Report*, pp. 1–8, Feb. 2002.

[10] S. H. Alsamhi, O. Ma, and M. S. Ansari, "Survey on artificial intelligence based techniques for emerging robotic communication," *Telecommunication Systems*, vol. 72, pp. 483–503, 3 2019, ISSN: 1572-9451. DOI: 10.1007/s11235-019-00561-z.

[11] T. B. Sheridan and W. L. Verplank, "Human and computer control of undersea teleoperators," *NASA. Ames Res. Center The 14th Ann. Conf. on Manual Control*, Jul. 1978.

[12] A. Seefried, A. Pollok, R. Kuchar, *et al.*, "Multi-domain flight simulation with the dlr robotic motion simulator," Feb. 2019, pp. 1–12. DOI: 10.23919/SpringSim.2019.8732921.

[13] H. Dol, M. Colin, P. van Walree, and R. Otnes, "Field experiments with a dual-frequency-band underwater acoustic network," 2018, pp. 1–5. DOI: 10.1109/UComms.2018.8493194.

[14] T. Sawa, N. Nishimura, and S. Ito, "Wireless optical ethernet modem for underwater vehicles," 2018, pp. 1–4. DOI: 10.1109/CCNC.2018.8319252.

[15] B. Kelley and K. Naishadham, "Rf multicarrier signaling and antenna systems for low snr broadband underwater communications," 2013, pp. 340–342. DOI: 10.1109/RWS.2013.6486734.

[16] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, "Augmented reality: A class of displays on the reality-virtuality continuum," *Telemanipulator and Telepresence Technologies*, vol. 2351, Oct. 1994. DOI: 10.1117/12.197321.

[17] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE Transactions on Information and Systems*, vol. 77, pp. 1321–1329, 1994.

[18] P. Milgram and J. H. W. Colquhoun, "A framework for relating head-mounted displays to mixed reality displays," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 43, pp. 1177–1181, 22 1999. DOI: 10.1177/154193129904302202.

[19] R. Wirz, R. Marín, J. M. Claver, M. Ferre, R. Aracil, and J. Fernández, "End-to-end congestion control protocols for remote programming of robots, using heterogeneous networks: A comparative analysis," *Robotics and Autonomous Systems*, vol. 56, pp. 865–874, 10 2008, ISSN: 0921-8890. DOI: https://doi.org/10.1016/j.robot.2008.06.005.

[20] M. E. Walker, H. Hedayati, and D. Szafir, "Robot teleoperation with augmented reality virtual surrogates," Mar. 2019, pp. 202–210. DOI: 10.1109/HRI.2019.8673306.

[21] 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," Feb. 2013, pp. 23–26. DOI: 10.1109/VR.2013.6549352.

[22] L. Antão, R. Pinto, J. Reis, G. Gonçalves, and F. L. Pereira, "Cooperative human-machine interaction in industrial environments," Jun. 2018, pp. 430–435. DOI: 10.1109/CONTROLO.2018.8514549.

[23] M. D. 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. 37 506–37 522, 2018, ISSN: 21693536. DOI: 10.1109/ACCESS.2018.2849572.

[24] G Lunghi, R Marin, M. D. Castro, A Masi, and P. J. Sanz, "Multimodal human-robot interface for accessible remote robotic interventions in hazardous environments," *IEEE Access*, vol. 7, pp. 127 290–127 319, 2019, ISSN: 2169-3536 VO - 7. DOI: 10.1109/ACCESS.2019.2939493.

[25] C. V. Almagro, G. Lunghi, M. D. Castro, *et al.*, "Cooperative and multimodal capabilities enhancement in the cerntauro human–robot interface for hazardous and underwater scenarios," *Applied Sciences*, vol. 10, 17 2020, ISSN: 2076-3417. DOI: 10.3390/app10176144.

[26] 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," SciTePress, 2021, pp. 388–395, ISBN: 978-989-758-522-7. DOI: 10.5220/0010528503880395.