Real-time Immersive Remote Telerobotics: Highlighting the Benefit of Humans in the Loop and Applying Machine Learning

Y. T. Tefera[1], J. V. Garcia[1], Kim Y.[1], Darwin G. Caldwell[1], S. Anastasi[2] and Deshpande N.[1]

[1]Istituto Italiano di Tecnologia (IIT), Genova, Italy

[2]Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro (INAIL), P.le Pastore 6, 00144 Rome, Italy

Abstract

Immersive remote teleoperation allows humans to interact with inaccessible or hazardous environments through robots using advanced displaying technologies like VR and AR. It ensures safe and efficient operations, facilitates knowledge transfer, and expands human involvement in challenging domains. At its core, successful remote robotic teleoperation hinges on intuitive interaction, and achieving accuracy and effectiveness relies on maintaining high fidelity in control actions and a deep perception of the remote environment. This paper introduces a Virtual Reality (VR)-based remote teleoperation interface, presented at the 1st International Workshop on Human-in-the-Loop Applied Machine Learning (HITLAML) held in Belval, Luxembourg. During the workshop, we demonstrated our work that aims to enhance real-time immersive interaction in Telerobotics, focusing on the vital role of human involvement. We explored the challenges associated with immersive remote teleoperation and presented potential solutions we have studied. In addition, We showcased a live demonstration illustrating the practical implementation of these solutions in real-time, overcoming geographical distances. Specifically, We presented a case study involving a teleoperation scenario from Luxembourg to Genova, Italy, spanning 863 km. Furthermore, we discussed specific areas where machine learning applications can enhance the user experience, improve efficiency, and optimize task execution.

Keywords

3D Reconstruction, Mixed Reality, Gaze Tracking, Telepresence, Teleoperation

1. Introduction

Remote telerobotics has gained a lot of attention, particularly after the COVID-19 pandemic. This significant interest is not without reason, as effective applications in this domain hold the

The 1st International Workshop on Human-in-the-Loop Applied Machine Learning (HITLAML)

yonas.tefera@iit.it (Y. T. Tefera); joaquin.vilagarcia@iit.it (J. V. Garcia); yaesol.kim@iit.it (K. Y.); Darwin.Caldwell@iit.it (D. G. Caldwell); s.anastasi@inail.it (S. Anastasi); nikhil.deshpande@iit.it (D. N.)

https://www.iit.it/it/people-details/-/people/yonas-tefera (Y. T. Tefera);
https://www.iit.it/it/web/guest/people-details/-/people/joaquin-vilagarcia (J. V. Garcia);
https://www.iit.it/it/people-details/-/people/yaesol-kim (K. Y.);
https://www.iit.it/it/people-details/-/people/darwin-caldwell (D. G. Caldwell);
https://www.iit.it/it/people-details/-/people/nikhil-deshpande (D. N.)

0000-0002-9938-4379 (Y. T. Tefera); 0009-0000-9078-6817 (J. V. Garcia); 0000-0003-0374-9570 (K. Y.); 0000-0002-6233-9961 (D. G. Caldwell); 0000-0003-2163-6684 (S. Anastasi); 0000-0003-4851-5655 (D. N.)

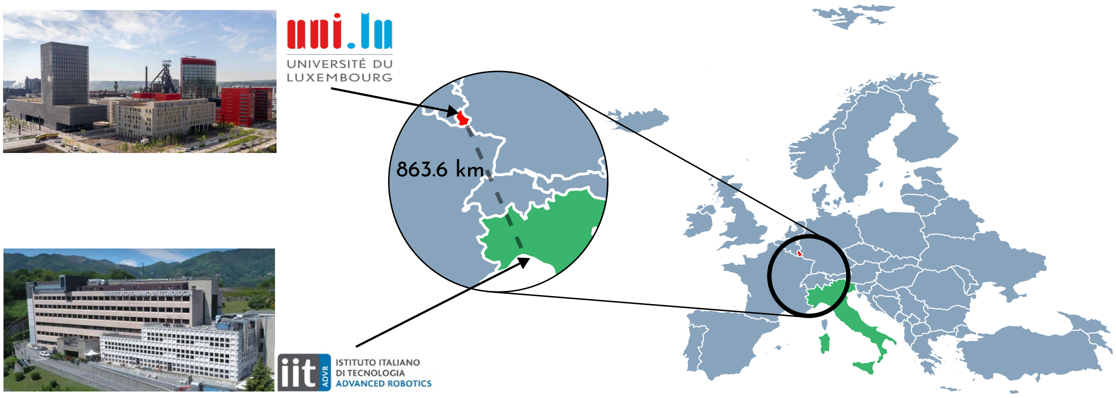

Figure 1: A map showing the geographical location of the remote teleoperation setup: The operator site was located in Luxembourg, and the remote environment (the robot) was at the IIT - Center for Robotics and Intelligent Systems hub (CRIS) (north-west of Genova, Italy) 863 km Kilometers away.

potential to enhance the lives of frontline workers greatly. These applications could enable them to respond to specific emergencies without needing their physical presence [1]. The progress in this field can largely be attributed to the widespread availability of consumer-grade sensors, such as RGB-D cameras, like Microsoft Kinect, Intel RealSense, and ZED cameras, that offer high quality 3D visual data acquisition at a low cost. Simultaneously, the growth of immersive virtual reality (VR) devices has played a pivotal role, helping drive advanced rendering graphics at reducing cost. Their impact has been particularly notable in fostering the development of innovative algorithms for real-time video and point-cloud acquisition, streaming, and rendering. This influx of accessible yet powerful sensing and rendering technology has paved the way for novel solutions that have the potential to revolutionize the world of remote telerobotics.

Immersive Remote Telerobotics (IRT), i.e., the combination of VR and real-time 3D visual data from remote RGB-D cameras allows real-time immersive visualization and interaction for both individual users and multiple users, perceiving the colour and 3D profile of the remote scene and robotic agents simultaneously [2, 3, 4]. This combination is the key distinguishing factor against traditional teleoperation interfaces, which rely on mono- or stereo-video feedback and suffer from limitations in terms of fixed or non-adaptable camera viewpoints, occluded views of the remote space, etc. [5]. Nevertheless, the increased data footprint (3D vs 2D) in real-time IRT imposes constraints on resolution, latency, throughput, compression, acquisition, and the visualization of this information [2]. For instance, latency and low resolution negatively impact the sense of presence and provoke cybersickness [6, 7]. In unstructured environments, such as remote inspection and disaster response, these challenges become even more pronounced due to the inherently complex and hazardous nature of these environments. Teleoperators in such situations face at times extreme demands on their cognitive and physical capacities, even if they are experts in the particular applications. In addition, Teleoperating a robot over the local environment allows users to use wired connectivity between the robot and the operator site for a smooth robot control. However, teleoperation over remote or hazardous environments requires the operator to control a robot from many kilometers away: leading to wired connectivity limitations. Thus, we implemented a teleoperation system setup over

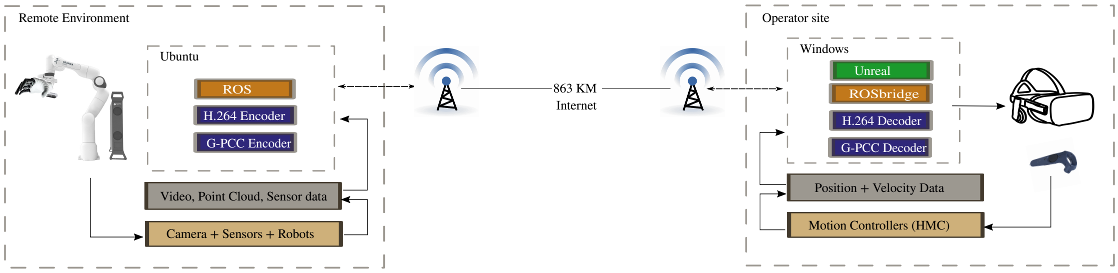

Figure 2: Overall software and hardware framework of internet teleoperation setup: The remote environment side components are shown on the left, and it includes the Franka Emika Panda robot for controlling the Hannes hand and a real-sene RGB-d camera for imaging the remote site. The operator side components are illustrated on the right. These include the HTC Vive Immersive visualization system with the HTC motion controllers (HMC). In the center, the public internet network is shown.

a public internet connection to understand how responsive the remote robot is to commands from the operator site and the overall acceptability of the system, as perceived by the operator. We showcased a live demonstration illustrating the practical implementation of these solutions in real-time, overcoming geographical distances. The demo was conducted during the 1st International Workshop on Human-in-the-Loop Applied Machine Learning (HITLAML) 2023 in Luxumberg. The operator site was located in Luxembourg, and the remote environment (the robot) was at the IIT - Center for Robotics and Intelligent Systems hub (CRIS) (north-west of Genova, Italy) 863 km Kilometers away as shown in Fig. 1.

2. SYSTEM OVERVIEW

As illustrated in Figure 2, the teleoperation setup incorporates three main parts: the remote environment, the internet network, and the operator site systems.

2.1. The remote environment

The remote environment included the Franka Emkia panda Robot, attached with Hannes hand ([8], and a real-sense RGB-D camera. The robotic arm has 7 DOF with sensitive and agile features and each joint with torque sensors, allowing adjustable stiffness/compliance and advanced torque control. For visual feedback, two real-sense RGB-D cameras were mounted in the remote environment, acquiring real-time video and depth maps data to create point clouds. Acquired HD (1280x720 pixels) resolution video data are then compressed using the industry's standard H.264 codec with an adaptive video bit rate to ensure high video quality. Streaming were performed using the Real-time Transport Protocol (RTP) to maximize data transmission rates. Likewise, the point cloud data were compressed using the Motion Picture Experts Group (MPEG) geometry-based point cloud compression (G-PCC) techniques [9]. This technique encodes the point cloud content directly in 3D space using an octree that describes the point locations in 3D space, and transmission were performed using the Boost ASIO over a TCP socket for streaming to maximize data transmission rate.

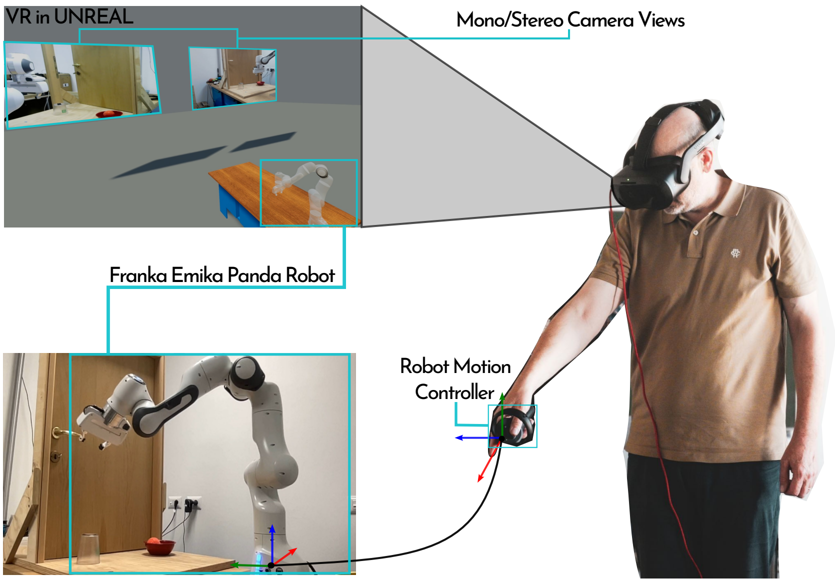

Figure 3: The teleoperation configuration spans both the user's location and the remote site. HTC Vive Motion Controllers and a head-mounted display are used at the user's end and the remote site includes the Franka Emika Panda Robot and RealSense cameras.

2.2. The Internet network

The Internet network provided a bidirectional data transmission channel to exchange robot command controls and visual feedback over the public internet network. The GlobalProtect VPN solution were used to deliver secure and reliable communication over public network. The robot command control communication between the operator site and the server side were established using a handshake protocol via ROSbridge. Before and during the teleoperation experiment, the download speed and upload speed were measured—the download speed at which data pockets need to reach from the computer to the internet was around 90 Mbps. Similarly, the upload speed, which is the data pockets, needs transfer from computer to the internet was around 158.4 Mbps.

2.3. The operator site

The operator site includes a range of interface devices to allow effective telepresence of the operator in the remote environment and intuitive remote control of the robotic systems. HTC Vive Pro head-mount display unit with a display resolution of 2880 x 1600, a 110-degree field of view, and a 90 Hz refresh rate was used to create immersive visualizations. Received video and point cloud packets were decoded using their respective decoders and were rendered using the Unreal graphics engine on Windows 10. Similarly, the robot state were received through ROSbridge is then used to display the virtual 3D models of the remote robot, rendering 1:1 the remote environment virtually. The HTC motion controllers (HMC) are then used to teleoperate the remote robots in real-time using an ungrounded motion controller interface.

2.4. Robot Control

The motion controllers of the HTC Vive Pro Eye allow users to send control commands in real-time using a wireless motion controller interface. These HTC Motion Controllers (HMC) are tracked in space, allowing users to freely explore the virtual reality (VR) environment and interact naturally with the remote robots. The HMC has two buttons: To engage/disengage motion commands between the HMC and remote robot, overcoming range limitations, and to open/close the remote robot's end-effector for precise object manipulation and control.

Since the operator directly command the motion of the remote robot using a HMC controller, our experimental setup follows a human-in-the-loop design paradigm without assuming any autonomy or semi-autonomy of the robots [8, 10]. The remote scene point-cloud is sampled, streamed, and rendered for the human operator in real-time. Based on the remote scene, the operator plans the next step and the action commands are sent to the remote robot. The remote robots are considered passive and do not act autonomously. The chosen control algorithm (e.g., Jacobian, Inverse Kinematics) calculates joint angles, velocities, and accelerations, which are then transmitted from the HMC interface to the remote robots. The joint states of the remote robots are relayed back to the interface to update the poses of the virtual robot models. This integration allows the operator to see and command the motion of the virtual robot, with the motion of the real robot being mapped 1:1 to the virtual robot motion. Due to the non-homothetic nature of the kinematics between the remote robot and the operator controllers, a velocity-control approach is used to command the 7-degrees-of-freedom (DOF) pose of the robot. The inverse Jacobian method is employed to calculate the velocity of the HMC, which is then mapped to the robot. To command the robot pose, the position error $\mathbf{e} \in SE(3)$ is determined by comparing the current pose with the desired pose. The damped least-squares solution, as shown in Eq. (1), is iteratively used to find the change in joint angles $\Delta\mathbf{q}$ that minimizes the error $\mathbf{e}$ [11].

$$\Delta\mathbf{q} = (\mathbf{J}^T\mathbf{J} + \lambda\mathbf{I})^{-1}\mathbf{J}^T\mathbf{e}, \tag{1}$$

where $\mathbf{J}$ is the manipulator Jacobian, $\lambda \in R$ is a non-zero damping constant, and $\mathbf{I}$ is the identity matrix. Finally, a proportional controller $\dot{\mathbf{q}} = K_p \cdot \Delta\mathbf{q}$ sets the joint velocities to achieve the desired pose. The values for λ and K_p were determined empirically as 0.001 and 0.6, respectively.

2.5. Remote Teleoperation Experiment

Attendees of the workshop actively engaged in the Remote Teleoperation Experiment. Each participant used HTC Vive Pro Eye headsets to view both the robot and its distant surroundings. Simultaneously, the remote station used the Intel Realsense RGB-D cameras, capturing the remote environment in real-time, and providing an immersive experience for the participants. The remote environment featured a variety of objects, including strategically placed red tennis balls and a designated target location containing a bowl. Additionally, there was a small door that needed to be opened as part of the scenario. A sample scene is seen in Fig. 3. The first task was to pick-up the ball and drop it off in the bowl and the second task was to open the door after drop off. At the "Go" signal from the experimenter, with the robot starting from a home location, the participant picked up the tennis ball and placed it inside the bowl. Participants released the grasped ball, based on their judgement of the end-effector location at the target location and then open the door.

3. Conclusions and Future Work

This study introduced a VR interface as the foundation for an immersive and user-friendly remote robotic teleoperation system from Luxembourg to Genova. The interface incorporates several key elements: (i) utilization of the Unreal graphics engine for superior VR rendering, enabling seamless manipulation of robotic platforms in distant settings, viewed through the HTC system; (ii) real-time video and point cloud streaming facilitated by RGB-D cameras, ensuring instantaneous perception and updates of the remote environment; and (iii) implementation of teleportation within VR for user viewpoint adjustments, overcoming obstacles and occlusions. The features integrated in the interface allow users to navigate the VR environment effortlessly, enhancing their understanding, visibility, positioning, and interaction within remote settings. This adaptability makes it particularly valuable for challenging telerobotic applications such as disaster response, nuclear decommissioning, and telesurgery, where precise control and remote interaction are essential. Nevertheless, operators are responsible for maintaining constant attention and focus on their tasks in real time. Challenges such as latency and reduced quality further increase the workload on perceptual and cognitive abilities while carrying out tasks. Semi-autonomous robots perform repetitive or time-consuming tasks with remarkable efficiency, exceeding human capabilities in these areas. The synergy between human intelligence and robotic precision proves to be more effective than either entity operating independently. For instance, a surgeon utilizing a robotic surgical system achieves unprecedented precision compared to manual methods, although the surgeon's expertise and judgment remain indispensable. Furthermore, even with sophisticated AI, robots can struggle in unforeseen situations. With their creativity and problem-solving abilities, humans can approach novel challenges that machines might find dificult. This collaboration between human ingenuity and robotic efficiency presents a powerful solution in diverse contexts.

In upcoming iterations, we will leverage machine learning and artificial intelligence methods to enhance the robot's capacity for making informed decisions. An illustrative example is reinforcement learning, which can progressively train robots to carry out designated tasks with greater autonomy. Furthermore, we intend to integrate foveated point-cloud rendering, as outlined in the study by [12], aligning it with the user's gaze while applying downsampling techniques to the peripheral vision. This strategic approach aims to reduce latency and bandwidth consumption significantly. Our ongoing research will also assess the impact on user immersion, task execution, and situational awareness, particularly in scenarios involving multiple users.

Acknowledgments

This research is supported by and in collaboration with the Italian National Institute for Insurance against Accidents at Work (INAIL), under the project "Sistemi Cibernetici Collaborativi - Robot Teleoperativo 2".

References

[1] G. Yang, H. Lv, Z. Zhang, L. Yang, J. Deng, S. You, J. Du, H. Yang, Keep Healthcare Workers Safe: Application of Teleoperated Robot in Isolation Ward for COVID-19 Prevention and Control, Chinese Journal of Mechanical Engineering 33 (2020). doi:https://doi.org/10.1186/s10033-020-00464-0.

[2] P. Stotko, S. Krumpen, M. Schwarz, C. Lenz, S. Behnke, R. Klein, M. Weinmann, A VR System for Immersive Teleoperation and Live Exploration with a Mobile Robot, in: IEEE/RSJ IROS, 2019, pp. 3630–3637. doi:10.1109/IROS.2012.6386012.

[3] Y. T. Tefera, D. Mazzanti, S. Anastasi, D. G. Caldwell, P. Fiorini, N. Deshpande, Towards Foveated Rendering For Immersive Remote Telerobotics, in: 5^{th} Intl. VAM-HRI Workshop at IEEE HRI, 2022.

[4] K. A. Szczurek, R. M. Prades, E. Matheson, J. Rodriguez-Nogueira, M. D. Castro, Multimodal multi-user mixed reality human–robot interface for remote operations in hazardous environments, IEEE Access 11 (2023) 17305–17333. doi:10.1109/ACCESS.2023.3245833.

[5] J. Y. C. Chen, E. C. Haas, M. J. Barnes, Human Performance Issues and User Interface Design for Teleoperated Robots, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 37 (2007) 1231–1245.

[6] M. Meehan, S. Razzaque, M. Whitton, F. Brooks, Effect of Latency on Presence in Stressful Virtual Environments, in: IEEE Virtual Reality, 2003. Proceedings., 2003, pp. 141–148.

[7] J.-P. Stauffert, F. Niebling, M. E. Latoschik, Latency and Cybersickness: Impact, Causes, and Measures. A Review, Frontiers in Virtual Reality 1 (2020) 31.

[8] A. Naceri, D. Mazzanti, J. Bimbo, Y. T. Tefera, D. Prattichizzo, D. G. Caldwell, L. S. Mattos, N. Deshpande, The *Vicarios* Virtual Reality Interface for Remote Robotic Teleoperation, Journal of Intelligent & Robotic Systems 101 (2021).

[9] R. Mekuria, K. Blom, P. Cesar, Design, implementation, and evaluation of a point cloud codec for tele-immersive video, IEEE Transactions on Circuits and Systems for Video Technology 27 (2016) 828–842.

[10] J. Calzado, A. Lindsay, C. Chen, G. Samuels, J. I. Olszewska, Sami: Interactive, multi-sense robot architecture, in: 2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES), 2018, pp. 000317–000322. doi:10.1109/INES.2018.8523933.

[11] S. Buss, Introduction to inverse kinematics with jacobian transpose, pseudoinverse and damped least squares methods, IEEE Transactions in Robotics and Automation 17 (2004).

[12] Y. T. Tefera, D. Mazzanti, S. Anastasi, D. G. Caldwell, P. Fiorini, N. Deshpande, FoReCast: Real-time Foveated Rendering and Unicasting for Immersive Remote Telepresence, in: H. Uchiyama, J.-M. Normand (Eds.), ICAT-EGVE 2022 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments, The Eurographics Association, 2022. doi:10.2312/egve.20221278.