A Multipurpose Interface for Close- and Far-Proximity Control of Mobile Collaborative Robots

Hamidreza Raei[1,2], Juan M. Gandarias[3], Elena De Momi[2], Pietro Balatti[1], and Arash Ajoudani[1]

Abstract— This letter introduces an innovative visuo-haptic interface to control Mobile Collaborative Robots (MCR). Thanks to a passive detachable mechanism, the interface can be attached/detached from a robot, offering two control modes: local control (attached) and teleoperation (detached). These modes are integrated with a robot whole-body controller and presented in a unified close- and far-proximity control framework for MCR. The earlier introduction of the haptic component in this interface enabled users to execute intricate loco-manipulation tasks via admittance-type control, effectively decoupling task dynamics and enhancing human capabilities. In contrast, this ongoing work proposes a novel design that integrates a visual component. This design utilizes Visual-Inertial Odometry (VIO) for teleoperation, estimating the interface's pose through stereo cameras and an Inertial Measurement Unit (IMU). The estimated pose serves as the reference for the robot's end-effector in teleoperation mode. Hence, the interface offers complete flexibility and adaptability, enabling any user to operate an MCR seamlessly without needing expert knowledge. In this letter, we primarily focus on the new visual feature of this interface, and first present a performance evaluation of different VIO-based methods for teleoperation. Next, the interface's usability is analyzed in a home-care application and compared to an alternative designed by a commercial MoCap system. Results show comparable performance in terms of accuracy, completion time, and usability. Nevertheless, the proposed interface is low-cost, poses minimal wearability constraints, and can be used anywhere and anytime without needing external devices or additional equipment, offering a versatile and accessible solution for teleoperation.

I. INTRODUCTION

MOBILE Collaborative Robots (MCR), often referred to as mobile cobots, are garnering substantial attention not limited to manufacturing and logistics but also expanding into the domains of healthcare, and in-home care [1]. Such widespread interest stems from the remarkable manipulation and stable locomotion capabilities inherently embedded in these robotic systems. However, despite their potential, the large-scale adoption of MCRs has been hindered by their lack of autonomy. This has been a major barrier to their integration into complex and dynamic environments. In response, significant efforts are devoted to increasing the MCRs' decisional and interaction autonomy [2].

In a parallel and synergistic effort, researchers are developing intuitive interfaces to enable human-in-the-loop

[1]HRI[2] Lab, Istituto Italiano di Tecnologia, Genoa, Italy. `hamidreza.raei@iit.it`
[2] Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy.
[3] Robotics and Mechatronics lab, Systems Engineering and Automation Department, University of Malaga, Malaga, Spain.
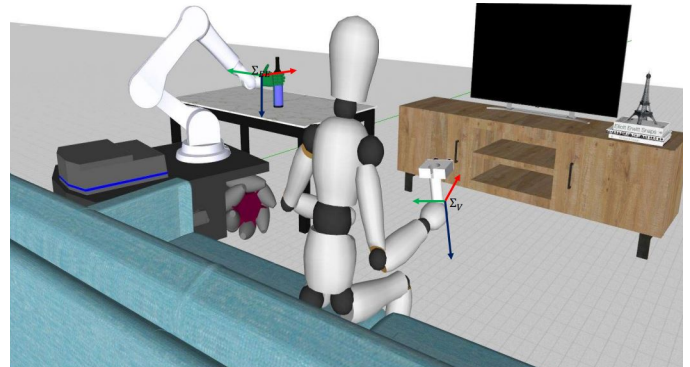
Fig. 1. The proposed VIO-based teleoperation interface enables individuals to control MCRs in both close and distant settings, facilitating the performance of Activities of Daily Living (ADLs) in home care environments.

control of MCRs, facilitating their use in close proximity and remote scenarios for effort-demanding or potentially hazardous tasks. In a step towards this goal, we have recently developed a collaborative framework that enables the MCR to function as a supernumerary mobile limb, leveraging its robust capabilities in tandem with human perceptual faculties [3]. Through a portable admittance-type interface and a priority-based whole-body controller, a user can move the arm, the mobile base, or a combination of both to perform complex loco-manipulation tasks [4].

With the new objective of enabling users to control the robot's locomotion and manipulation from a remote location (near or far from the MCR), especially for tasks such as assisting senior adults with limited mobility in their daily activities, this study introduces a novel design for an MCR teleoperation interface (see Fig. 1). To eliminate the need for any additional settings and external equipment, we propose to include a set of stereo cameras, and an inertial measurement unit (IMU) in the device for implementing visual-inertial odometry (VIO) to infer human commands.

In more details, this study proposes a novel visuo-haptic interface for MCR control, prioritizing critical factors essential for an efficient teleoperation interface, including portability, cost-effectiveness, independence from external sensors, and the absence of wearability requirements. In summary, the contributions of this work can be listed as follows:

- A passive mechanism that allows attaching and detaching the interface from the MCR without requiring additional actuation systems. This mechanism enhances adaptability and allows users to switch seamlessly from teleoperation to local operation.

- The integration of a VIO method into a low-cost wireless device that can be used for teleoperation. The method has been integrated into simple, inexpensive, and easily accessible hardware. It uses a stereo camera system and an IMU to estimate the position of the interface without the need for external sensors or complex and expensive tracking systems. This solution allows users to perform teleoperation tasks in an infinite workspace without limiting their movement.
- A performance evaluation of different hardware and VIO approaches to assess teleoperation accuracy according to several metrics is also presented.
- Integration of the VIO-based teleoperation approach into our previous haptic-based framework [5], both from hardware and software points of view. Thus, proposing a complete and unified framework that allows control both remotely, i.e., with the interface detached, thanks to the visual part of the interface, and locally, i.e., with the interface attached, thanks to the haptic part.
- Experimentation of the entire framework through a usability study and comparison with a state-of-the-art wireless teleoperation solution based on an IMU-based motion capture system.
- A detailed discussion on the advancements of our proposed solution incorporates both quantitative and qualitative metrics. These considerations are based on the results and performance revealed by the aforementioned experimental analyses.

The remainder of the manuscript is organized as follows: Section II includes state-of-the-art analysis of teleoperation frameworks. The proposed framework is presented in section III. Section IV depicts the experimental setup. Then, the results of the experiments are presented and discussed in depth. Finally, the conclusions are presented in section VI.

II. Related Works

Teleoperation technology has found diverse applications in robotics, spanning domains such as search and rescue [6], industrial inspection [7], surgery [8], disaster relief [9], and space exploration [10].

Among all interfaces used for teleoperation, the joystick stands out as a prevalent teleoperation tool, extensively employed in various studies [11], [12]. Despite its widespread adoption, teleoperating a system with a high-DoF using a low-DoF interface poses a substantial challenge, as highlighted in the work of Garate et al. (2021) [13]. Furthermore, employing a joystick with a high-DoF could compromise the intuitiveness of the teleoperation interface [14]. An innovative approach to address this challenge involves embedding the robot's high-DoF dexterous behavior into low-DoF latent actions, facilitating human control, as illustrated in the study by Jeon et al. (2020) [15]. Despite its efficacy, this method induces task dependency and hinders adaptability to other applications.

Alternatively, Inertial and optical motion capture systems offer intuitive and accurate solutions for teleoperating various robots, ranging from robotic arms [16], and MCRs [17],

to humanoid [18] and legged robots [19]. However, It is important to point out, that the integration of motion capture systems as a teleoperation solution is usually expensive and may obstruct adaptability to diverse environments since it has wearability constraints and in the case of optical systems, necessitates the installation of external sensors within the operational environment.

Other solutions for teleoperation include employing single or multiple cameras to track specific markers [5] or to estimate the human-body pose in a markerless manner [20], [21]. Compared to motion capture systems, these alternatives offer more economical solutions at the cost of lower accuracy, whilst, their functionality remains dependent on the external sensors within the operational environment.

As previously mentioned, the primary objective of this work is to develop a non-wearable, portable, cost-effective, and user-friendly interface that does not rely on environmental sensors or external markers. Unlike previous studies, our approach aims to overcome the aforementioned limitations and broaden the applicability of the teleoperation interface to control MCRs when detached, augmenting the capabilities of our previous haptic interface [4].

III. Framework Overview

A. *The Open Multi-Purpose Interface (OMI)*

The proposed OMI serves as both a haptic and visual interface, allowing remote or local robot operation, as depicted in the system overall block diagram in Fig. 2. It facilitates local operation via a haptic interface when attached to the end-effector. When detached, it employs the visual interface for teleoperation, details of which are explicitly outlined in the following subsections.

1) Haptic Interface: This interface is built around a System on a Chip (SoC) solution, specifically the M5Core2 which utilizes an ESP-32 microcontroller and incorporated a touchscreen, and an F/T sensor. It excels in user-friendliness, enabling one-handed local operation that makes the OMI flexible and programmable. Users can efficiently oversee a range of robot functions and easily adapt to specific task needs. Notable features of this haptic interface include the capability to switch motion modes, manage the gripper, adjust task motion priorities between the arm and the base, or set the impedance. Specific details on the use of the haptic interface and its integration with the whole body controller are illustrated in our previous studies [3]–[5].

2) Visual Interface: Leveraging the VIO output, the visual interface interprets the motion of the interface to formulate a reference command for the robot. This capability is embedded in an all-in-one handheld device that is well-suited for teleoperation in any environment. The main idea is to perform VIO on the handheld device, the OMI, and eliminate the need for an external sensor in the environment to localize the interface with regard to its initial pose.

Several algorithms can be used to perform VIO, including RTAB-MAP [22], LSD-SLAM [23], and ORB-SLAM [24], which are typically compatible with RGB-D and stereo cameras. However, one of the specific differences of the
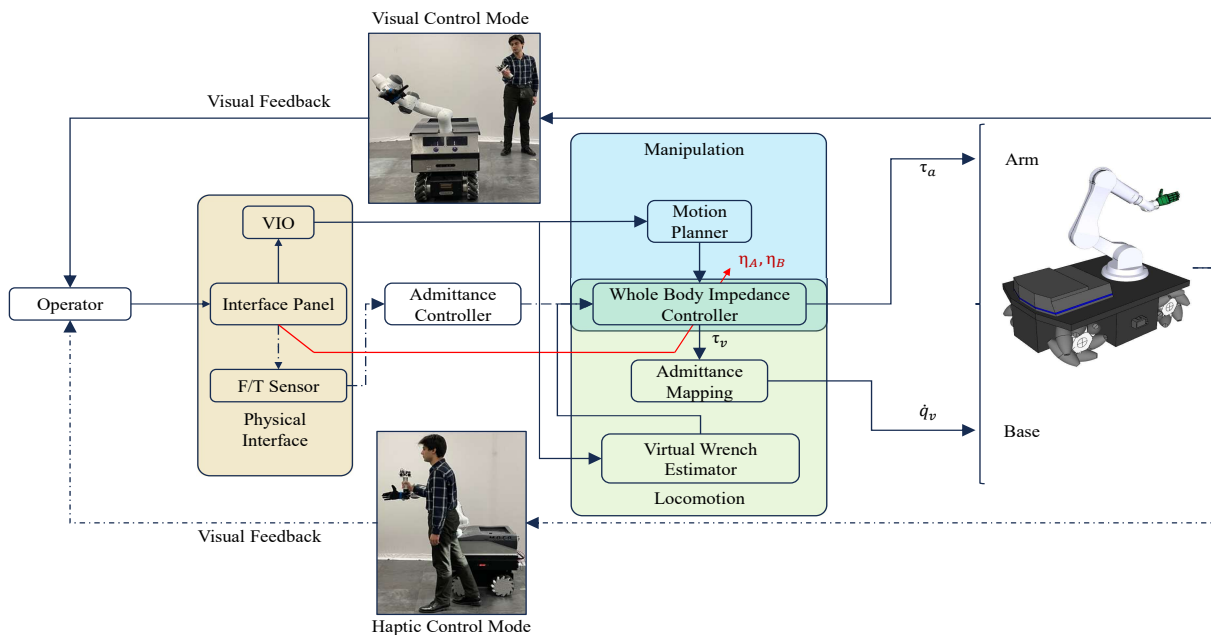
Fig. 2. Block diagram of the extended visuo-haptic framework, to be used for MCRs under a whole-body impedance controller. Dashed lines represent connections related to the haptic part whilst solid lines represent those related to the visual part.

RTAB-MAP algorithm is its reliance on Appearance-Based loop-closure, as opposed to the dependence on geometric features or landmarks, used mostly in other algorithms. This special feature enhances the performance of the algorithm in environments with visual patterns but not many clear geometric landmarks. Implementing a technique known as the "bag of words", significantly improves loop closure detection while remaining efficient in terms of memory usage and computational load. Thereby mitigating drift by promptly recognizing previously visited scenes [22].

RGB-D and stereo cameras stand out as predominant choices for providing visual, and geometric data to VIO, owing to their widespread availability and compatibility with a majority of VIO algorithms. The type of visual sensors and the communication between the interface and the robot impact the accuracy and frame rate of the VIO system, influencing computational demands and data transmission requirements. Therefore, empirical testing is crucial to validate real-time teleoperation and precision of various setups.

B. Robot Whole-body Cartesian Impedance Controller

According to the formulation presented in [25], the MCR can perform two distinct behaviors: manipulation and locomotion. In the context of the Manipulation mode, when the mobile platform is deemed to be in a quasi-static state and a virtual inertial and damping representation is adopted for the entire robot, effectively replacing the inertial and damping parameters of both the arm and the mobile platform with their virtual counterparts, the resulting equation of overall dynamic of the MCR can be expressed as follows:

$$M_{adm}\ddot{q} + D_{adm}\dot{q} = \tau_c + \tau_{ext} \qquad (1)$$

which can be expanded as following:

$$\begin{pmatrix} M_v & 0 \\ 0 & M_r(q_r) \end{pmatrix} \ddot{q} + \begin{pmatrix} D_v & 0 \\ 0 & C_r(q_r, \dot{q}_r) \end{pmatrix} \dot{q} + \begin{pmatrix} 0 \\ g_r \end{pmatrix} =$$

$$\tau_c + \tau_{ext} \qquad (2)$$

Here, g_r is the gravity vector, $M_{adm} \in \mathbb{R}^{m \times m}$ and $D_{adm} \in \mathbb{R}^{m \times m}$ are the virtual inertial and virtual damping, and $q = (q_v, q_r)^T \in \mathbb{R}^{10}$ represents whole-body joint variables, and $\tau_c = [\tau_v^T \; \tau_r^T]^T$ and $\tau_{ext} = [0^T \; \tau_{r,ext}^T]^T$ are respectively the high-level commanded torque vector and the external torque vector that is passed for the mobile admittance controller (τ_v) and the arm low-level controller (τ_r).

The whole-body controller's advantage lies in its ability to execute operator-commanded motions by prioritizing specific joint movements. In other words, a weighting matrix is established to compute the high-level torque references(τ_c) as explained in our previous study [25]. This weighting matrix is determined as follows:

$$W(q) = H^T M^{-1}(q) H \qquad (3)$$

Here, the selection of H is based on the specific task at hand, whether it involves locomotion or manipulation, and it is constructed as a diagonal matrix:

$$H = \begin{pmatrix} \eta_B I_{n_b} & 0_{n_b \times n_a} \\ 0_{n_b \times n_a} & \eta_A I_{n_a} \end{pmatrix} \qquad (4)$$

where η_B and η_A penalize the motion of base or arm, respectively. These parameters can be adjusted either dynamically or manually to prioritize the movement of either the arm or the base, depending on the specific task at hand. Nonetheless, for the haptic interface, the study mentioned earlier [3] has clearly demonstrated its adaptability to both manipulation

and locomotion. In the following, the adaptability of the visual interface to these modes is illustrated.

1) Manipulation using VIO: During manipulation, the priority of the motion is given to the robotic arm and it works under the whole-body controller. Our objective is to achieve the same motion for the arm end-effector w.r.t its initial pose, as motion in the teleoperation interface w.r.t the interface initial pose. The world coordinate system is denoted as Σ_W and the transformation between the initial and the final pose of the end-effector is denoted as $T_{EE_i}^{EE_f}$, whilst, the transformation between the initial and the final pose of the teleoperation interface is denoted as $T_{V_i}^{V_f}$. Thus, our objective of the manipulation using VIO is to satisfy $T_{EE_i}^{EE_f} = T_{V_i}^{V_f}$.

The importance of a whole-body controller is emphasized when the operator extends the arm beyond its typical range. In such cases, the controller commands a coordinated motion involving both the manipulator and the mobile platform, ensuring that the end-effector attains the desired pose.

The mobile platform's odometry provides the transformation from the mobile platform's initial pose to its pose at any given moment, denoted as T_W^M. The T_M^R refers to the transformation from the mobile platform coordinate system to the Arm's base. $q_v \in \mathbb{R}^3$ represents mobile platform joint variables, which is 3 DoF shown as $q_v = (q_{vx}, q_{vy}, q_{vz})$ whilst $q_r \in \mathbb{R}^7$ represents joint variables of the arm with 7 DoF. The command to the MCR is the transformation from Σ_W to the final pose of the end-effector, which is supposed to replicate the motion from the interface while scaling the translational motion to accommodate the specific needs of the application. Hence, the transformation is obtained as follows:

$$T_W^{EE_f}(q) = T_W^M(q_v)\ T_M^R\ T_R^{EE_i}(q_r)\ T_{EE_i}^{EE_f}(q_r) \tag{5}$$

The scaling factor governing translation at the end-effector is dictated by the parameter denoted as α. This parameter is applied through element-wise multiplication, as expressed by the following:

$$T_{EE_i}^{EE_f} = \begin{pmatrix} 1 & 1 & 1 & \alpha \\ 1 & 1 & 1 & \alpha \\ 1 & 1 & 1 & \alpha \\ 1 & 1 & 1 & 1 \end{pmatrix} \odot T_{v_i}^{v_f} \tag{6}$$

2) Locomotion using VIO: During locomotion, the motion priority shifts to the mobile platform in the whole-body controller. The displacement of the teleoperation interface turns into a virtual force and Torque at the robot's end effector respectively denoted as $F_{vir} = \begin{pmatrix} F_x & F_y & F_z \end{pmatrix}^T$ and $\tau_{vir} = \begin{pmatrix} \tau_x & \tau_y & \tau_z \end{pmatrix}^T$, passing through the admittance controller, resulting in the mobile platform motion. Force and torque at the end effector are virtually obtained w.r.t to Σ_W coordinate system. In the following equations, linear stiffness and rotational stiffness are denoted as K and C respectively.

$$\begin{pmatrix} F_x \\ F_y \\ F_z \end{pmatrix} = \begin{pmatrix} K_x & 0 & 0 \\ 0 & K_y & 0 \\ 0 & 0 & K_z \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} \tag{7}$$

$$\begin{pmatrix} \tau_x \\ \tau_y \\ \tau_z \end{pmatrix} = \begin{pmatrix} C_x & 0 & 0 \\ 0 & C_y & 0 \\ 0 & 0 & C_z \end{pmatrix} \begin{pmatrix} \Delta\phi \\ \Delta\theta \\ \Delta\psi \end{pmatrix} \tag{8}$$

In locomotion, the mobile platform moves on the surface, involving 3 DoF which results in ($K_z = C_x = C_z = 0$). A circular virtual threshold is set for locomotion, requiring operators to cross it initially to map a virtual wrench onto the robot's end-effector. Additionally, a maximum radius is defined, denoted as R_m to limit the magnitude mapped wrench if the operator surpasses safety constraints $F_{vir} \in [-KR_m, -KR_{min}] \cup [KR_{min}, KR_m]$.

We observed that attempting diagonal movements with the robot caused several subjects to encounter various challenges, resulting in a modification that mandates robot motion in only one direction at a time, and changes in direction are only allowed after it comes to a stop. To optimize accuracy, the selected direction aligns with the predominant interface's motion axis, commanded by the operator.

This algorithm is implemented to enable long-distance locomotion without requiring the operator to physically move comparable to the mobile platform. Once the virtual wrench in this process is replaced by the output of the F/T sensor installed on the end-effector, the locomotion part is similar to the haptic interface [25]. The control block diagram of the teleoperation interface for both locomotion and manipulation mode is illustrated in Fig. 2.

IV. The Experimental Setup

This section outlines two types of experiments designed to assess the effectiveness of the VIO interface. In the first experiment, we compare the proposed VIO interface against alternative VIO configurations, evaluating their respective accuracies and update rates. The second experiment involves a comparative analysis of the proposed VIO interface against a well-established teleoperation method employed for mobile robots, in the application of home-care scenarios.

A. VIO Comparison

This comparison was made to compare various setups for VIO in terms of their accuracy, given a specific motion. The experiment involves translational and rotational motions along/around the x, y, and z axes, aimed to compare the performance of two VIO based setups. A structure was designed to link the VIO setups to the Franka Emika robot end-effector. This link facilitates precise motion control in different directions and orientations and enables accurate comparisons of various setups and configurations by providing ground truth data of the position and the orientation of the robot end-effector.

The wired stereo setup is a ZED2i, stereo camera, that streams images at 30 Frame Per Second (FPS) with a built-in IMU and a magnetometer. This camera features a Field of View (FoV) of 72 degrees horizontally and 44 degrees vertically. The RGB-D wired setup is an Intel Depth Camera D435i, it streams images at 30 FPS, and the FoV of 69 degrees horizontally and 42 degrees vertically.

For the wireless configuration of the RGB-D setup, the Intel D435i is connected to a Raspberry Pi 4, which streams images and depth frames via WiFi. The main difference between the wired RGB-D setup and the wireless one lies in its limited frame rate capability, with the wireless configuration achieving a maximum streaming rate of 15 FPS.
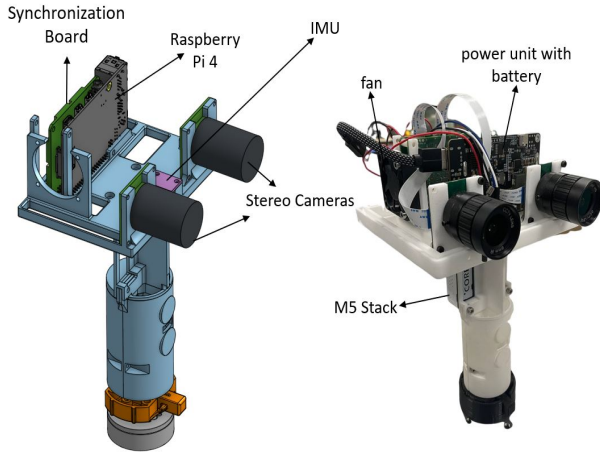

Fig. 3. The image showcases the designed model of the interface on the left, with the corresponding assembled version of the interface presented on the right.

The wireless stereo setup is designed, by prioritizing FPS, weight, and the system latency, in addition to constraints on power consumption and computational resources. The proposed wireless VIO interface consists of a Raspberry Pi 4, an Arducam synchronized stereo kit (two 12.3MP 477P camera modules), an M5Core2, an IMU (MPU-9250), and a power unit with batteries. The proposed stereo interface has a FOV of 65 degrees horizontally, and 40 degrees vertically and could stream up to 30 FPS. Additionally, the CAD (Computer-Aided Design) representation of the interface is displayed, along with the assembled interface depicted in Fig. 3. The CAD design is accessible through the following link: `https://cad.onshape.com/documents/a6ae72c`

Additionally, access to this study source code is available on our github using the following link:

`https://github.com/hrii-iit/hrii_vo.git`

B. Home-Care Application

This section outlines the scenario chosen to validate the proposed teleoperation interface against a widely used motion capture system, employed in diverse studies for teleoperating various robots such as mobile collaborative robots [17] and legged robots [19]. The motion capture system employed in this study is the Xsens system, which incorporates multiple IMUs affixed to wearable costumes. Several subjects were assigned to perform this scenario using teleoperation with both methods.

1) Experiment Scenario: The determined scenario for this experiment entails the subject executing three sequential tasks as follows:

i. *Grasping a relatively small ball in manipulation mode.*

ii. *Locomoting the mobile platform from its initial position to the vicinity of the drawer.*

iii. *Depositing the ball into the drawer and subsequently closing the drawer.*

In Fig. 4 the experiment setup and distances between the starting point and the drawer are illustrated.

2) Study Protocol: During the comparative study, subjects were asked to sit on a chair and perform remotely while looking at the robot on their side. Participants were asked to perform the sequential tasks two times, unaware of the interface type assigned for each trial, and they wore the Xsens suit and grasped the VIO interface throughout both trials. The decision aimed to prevent participants from altering their behavior and performance based on prior knowledge of the teleoperation interface. The allocation of each interface was randomized for each participant. This study sought to evaluate these teleoperation systems based on both quantitative and qualitative parameters. The specified quantitative parameters included in this study are listed as follows:

- The success rate, which is determined by dividing the number of successfully completed subtasks by the total trials across all subtasks for each participant.
- The corresponding time duration (T_c) of overall performance for each subject which is the sum of the time duration for all three successful subtasks.
- Average preparation time for each subject to use the teleoperation system denoted as T_p.

Additionally, to assess the effectiveness of the proposed teleoperation interface in terms of qualitative parameters, the NASA Task Load Index (TLX) questionnaire scores were obtained from each subject upon completion of all three sequential tasks using each teleoperation system.
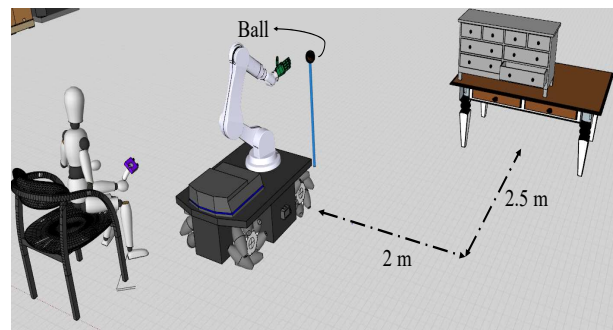

Fig. 4. A sketch of the experimental setup, used for home-care scenario application.

V. RESULTS AND DISCUSSION

A. VIO Comparison

The translational and orientational motion given to the VIO setups attached to the end-effector have been plotted in Fig. 5 as input motion, which is considered as ground truth since it is measured through inverse kinematics of the robotic arm.
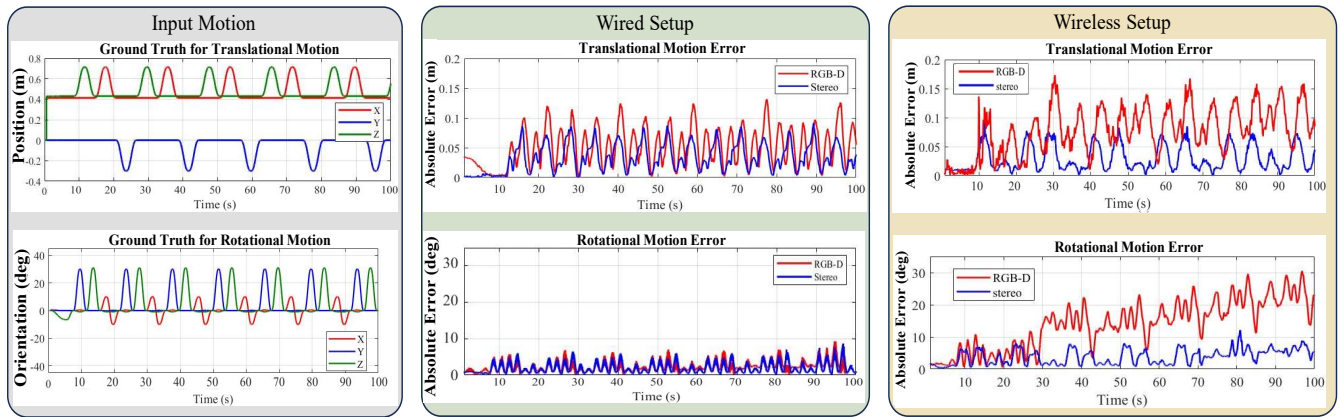
Fig. 5. This figure compares VIO performance using Stereo and RGB-D setups. The left column shows the translational and rotational inputs. The middle column displays the error for wired configurations, and the right column shows the error for wireless configurations. The top plots in both columns illustrate the absolute position estimation error, with red for RGB-D and blue for Stereo and the bottom ones indicate the rotational error.

All mentioned setups in wired and wireless configurations were tested with the same motions and the VIO estimation of the motion is compared to the ground truth. Thus, the absolute error of the RGB-D and the stereo setup have been plotted for both wireless and wired configurations in Fig. 5.

For the case of wired configurations, although in translational motion higher error for the RGB-D setup has been observed, the orientational motion does not exhibit a considerable difference between stereo and RGB-D setup.

In the wireless configuration, the RGB-D camera exhibits significant errors exceeding those in the stereo setup, with an apparent drift over time affecting both translational and orientational accuracy. Translational errors differ notably between wired and wireless setups for the RGB-D camera, while the stereo camera shows similar errors in both configurations. Average errors for the RGB-D camera are 0.0541 (wired) and 0.0833 (wireless), and for the stereo camera, 0.0365 (wired) and 0.0384 (wireless). Thus, the stereo setup showed lower average error compared to the RGB-D setup in both configurations, respectively 32% and 53% less than the RGB-D setup in wired and wireless configuration.

It is worth mentioning that regarding the update rate of the VIO output achieved by the teleoperation interface, for the wired configuration, the odometry update rate is consistent at approximately 30 Hz for both setups. However, wireless setups face several challenges ranging from computation limitations to bandwidth of wireless communication. Considering a trade-off between the update rate and the image resolution resulting in the accuracy of the odometry, we could reach 30 Hz for the stereo setup and 15 Hz for the RGB-D camera.

B. Home-Care Application

Eight participants performed the Home-Care application scenario utilizing the proposed VIO interface and the Xsens system. In Fig. 6, the depicted success rates for each interface highlight a higher success rate with the VIO interface compared to the Xsens system in the designated scenario.
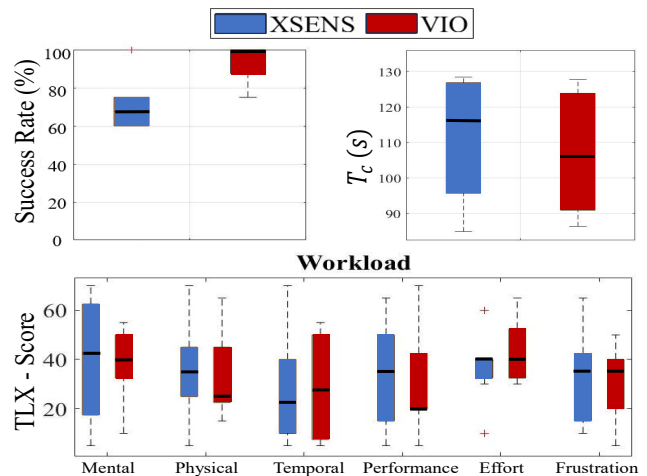

Fig. 6. Illustrates the comparison between qualitative and quantitative parameters of the proposed VIO interface and Xsens system performance, used for teleoperation in the home-care application scenario

The average success rate for the Xsens system is 70.625% and the average for the proposed VIO interface is 93.75%.

Regarding the corresponding time duration (T_c) for each teleoperation system, both systems had similar time lengths. However, the median T_c for the proposed VIO interface is slightly lower (106 seconds) compared to the Xsens system's median T_c of 116 seconds, indicating a marginally more efficient task completion with the VIO interface which is depicted in Fig. 6 at the top right side of the figure. Regarding the teleoperation system's preparation time, the Xsens system necessitates about 3 minutes for donning, plus an additional approximately 1 minute for calibration, a requisite for every inertial motion capture system. Conversely, the VIO interface, being handheld and not wearable, demands no calibration, enabling the subject to initiate usage within 30 seconds.

The NASA-TLX questionnaire result is depicted at the bottom of Fig. 6, and a statistical analysis of TLX pa-
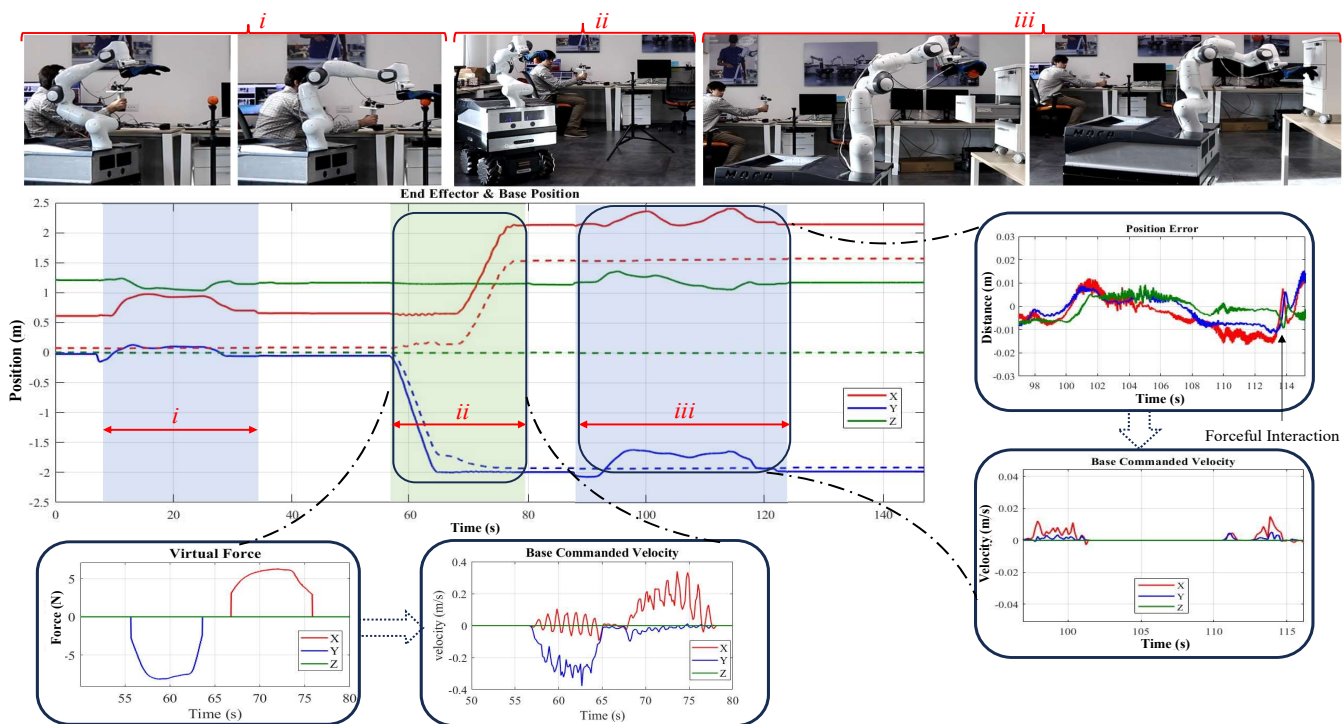
Fig. 7. The figure illustrates VIO interface performance in the Home-Care Scenario, with dashed lines for the mobile platform position and solid lines for the End-Effector Position. The blue highlight indicates manipulation mode, and the green one signifies locomotion mode.

rameters is performed. A Wilcoxon signed-rank test was conducted on NASA-TLX data, employing a significance level of 0.05. Findings indicate no statistical significance in all NASA-TLX parameters between the two teleoperation systems under consideration. The p-values for the Wilcoxon signed-rank test are as follows: mental demand (0.8984), physical demand (0.5), temporal demand (0.9062), performance (0.6719), effort (0.3281), and frustration (0.8906). These results suggest comparable task load indices for both teleoperation systems.

While both systems exhibit reasonable and relative accuracy, we claim that providing subjects with awareness of the specific points used for manipulation or locomotion in the teleoperation paradigm enhances task performance. In contrast, the Xsens system lacks clarity regarding the designated segment on the subject's arm for mimicking motions using teleoperation, potentially leading to errors as subjects may not be aware of the precise manipulation points. That could possibly be the reason for a slightly higher success rate, in the VIO teleoperation interface.

The comparative analysis of the VIO teleoperation system and the Xsens system reveals their similar performance across quantitative and qualitative parameters. Notably, the VIO interface exhibits slightly superior results in terms of success rate, in the aforementioned home-care scenario. The VIO interface being relatively cost-effective in addition to the extended preparation time, calibration requirements, and the wearability constraint of the Xsens system would suggest that the VIO interface could be more adaptable to various

environments and applications.

An example of this home-care application scenario has been performed and visualized in Fig. 7. In this figure, the motion of the end effector is depicted with solid lines, and the motion of mobile platform is depicted with dashed lines. Additionally, segments, where the operator executed tasks in manipulation mode, are highlighted in blue, while locomotion mode is represented in green. Throughout the locomotion mode, the virtual force acting on the end effector is graphed at the bottom left of Fig. 7, influencing the commanded velocity for the mobile platform depicted on the right side of the virtual force plot.

Another parameter that influences the commanded velocity for the mobile platform is the extension of the arm beyond its regular range, leading to an error between the commanded and actual arm position, which results in the motion of the mobile platform under the whole-body impedance controller. The integration of an impedance controller offers several advantages, such as demonstrating compliance—allowing slight yielding or deformation in response to external forces. On the right side of Fig. 7, the position error resulting from arm extension and forceful interactions is indicated. Where exerting force is required to close the drawer, the error between the commanded and actual position of the arm increases which leads to small motion in the mobile platform evident in the commanded velocity depicted on the bottom right side of Fig. 7. The video showcasing the performance of the proposed VIO interface is available at the following link: https://youtu.be/cZpI6Cp2Atc.

While the VIO interface showed slightly better performance, it is worth mentioning that factors such as illumination levels and a highly dynamic teleoperation environment can affect VIO pose estimation and its performance in teleoperation.

VI. CONCLUSIONS

This study introduced a novel and low-cost visuo-haptic interface for MCR featuring a passive detachable mechanism, allowing local loco-manipulation when attached and teleoperation when detached. The control framework included a whole-body impedance controller to enhance the interaction of the robot with external objects in the environment.

Additionally, a home-care scenario was determined to validate and compare the performance of the proposed VIO interface with one of the widely known means of teleoperation. Results showed that both systems performed similarly in terms of qualitative parameters concerning the human workload, whilst in terms of quantitative parameters, the VIO interface slightly outperformed the Xsens system. The evaluation result obtained in this experiment indicates that the primary objective of this study, which is the development of a teleoperation interface that operates without dependency on environmental sensors while ensuring precision, and wireless connectivity, has been achieved. It is worth mentioning that the proposed teleoperation interface is not limited to being used for mobile cobots only and can be seamlessly integrated with various other robots.

In the future, considering the accuracy of the proposed visual interface validated in this study, it can be integrated with learning from demonstrations (LfD) applications. Additionally, this system can be integrated with shared autonomy, through perception algorithms that can detect and identify objects in the close vicinity of the robot, and potentially be used to predict the operator's intention of a given command to reduce the workload on the operator.

REFERENCES

[1] A. Abou Allaban, M. Wang, and T. Padır, "A systematic review of robotics research in support of in-home care for older adults," *Information*, vol. 11, no. 2, p. 75, 2020.

[2] H. Engemann, S. Du, S. Kallweit, P. Cönen, and H. Dawar, "Omnivil—an autonomous mobile manipulator for flexible production," *Sensors*, vol. 20, no. 24, p. 7249, 2020.

[3] A. Giammarino, J. M. Gandarias, P. Balatti, M. Leonori, M. Lorenzini, and A. Ajoudani, "Super-man: Supernumerary robotic bodies for physical assistance in human-robot conjoined actions," *arXiv preprint arXiv:2201.06365*, 2022.

[4] J. M. Gandarias, P. Balatti, E. Lamon, M. Lorenzini, and A. Ajoudani, "Enhancing flexibility and adaptability in conjoined human-robot industrial tasks with a minimalist physical interface," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2022, pp. 8061–8067.

[5] A. Giammarino, J. M. Gandarias, and A. Ajoudani, "An open tele-impedance framework to generate data for contact-rich tasks in robotic manipulation," in *2023 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO)*. IEEE, 2023, pp. 140–146.

[6] F. Perez-Grau, R. Ragel, F. Caballero, A. Viguria, and A. Ollero, "Semi-autonomous teleoperation of uavs in search and rescue scenarios," in *2017 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2017, pp. 1066–1074.

[7] F. Rocha, G. Garcia, R. F. Pereira, H. D. Faria, T. H. Silva, R. H. Andrade, E. S. Barbosa, A. Almeida, E. Cruz, W. Andrade *et al.*, "Rosi: A robotic system for harsh outdoor industrial inspection-system design and applications," *Journal of Intelligent & Robotic Systems*, vol. 103, pp. 1–22, 2021.

[8] J. Leven, "A telerobotic surgical system with integrated robot-assisted laparoscopic ultrasound capability," Ph.D. dissertation, Thesis for Master of Science in Engineering in Computer Science, The Johns . . . , 2005.

[9] T. Klamt, D. Rodriguez, M. Schwarz, C. Lenz, D. Pavlichenko, D. Droeschel, and S. Behnke, "Supervised autonomous locomotion and manipulation for disaster response with a centaur-like robot," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 1–8.

[10] W. Bluethmann, R. Ambrose, M. Diftler, S. Askew, E. Huber, M. Goza, F. Rehnmark, C. Lovchik, and D. Magruder, "Robonaut: A robot designed to work with humans in space," *Autonomous robots*, vol. 14, pp. 179–197, 2003.

[11] N. Mavridis, G. Pierris, P. Gallina, N. Moustakas, and A. Astaras, "Subjective difficulty and indicators of performance of joystick-based robot arm teleoperation with auditory feedback," in *2015 International Conference on Advanced Robotics (ICAR)*. IEEE, 2015, pp. 91–98.

[12] J. E. Solanes, A. Munoz, L. Gracia, A. Martí, V. Girbés-Juan, and J. Tornero, "Teleoperation of industrial robot manipulators based on augmented reality," *The International Journal of Advanced Manufacturing Technology*, vol. 111, pp. 1077–1097, 2020.

[13] V. R. Garate, S. Gholami, and A. Ajoudani, "A scalable framework for multi-robot tele-impedance control," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 2052–2066, 2021.

[14] I. Dekker, K. Kellens, and E. Demeester, "Design and evaluation of an intuitive haptic teleoperation control system for 6-dof industrial manipulators," *Robotics*, vol. 12, no. 2, p. 54, 2023.

[15] H. J. Jeon, D. P. Losey, and D. Sadigh, "Shared autonomy with learned latent actions," *arXiv preprint arXiv:2005.03210*, 2020.

[16] N. Mavridis, N. Giakoumidis, and E. L. Machado, "A novel evaluation framework for teleoperation and a case study on natural Human-Arm-Imitation through motion capture," *International Journal of Social Robotics*, vol. 4, no. S1, pp. 5–18, 11 2011. [Online]. Available: https://doi.org/10.1007/s12369-011-0117-8

[17] Y. Wu, P. Balatti, M. Lorenzini, F. Zhao, W. Kim, and A. Ajoudani, "A teleoperation interface for loco-manipulation control of mobile collaborative robotic assistant," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3593–3600, 2019.

[18] D. Dajles, F. Siles *et al.*, "Teleoperation of a humanoid robot using an optical motion capture system," in *2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI)*. IEEE, 2018, pp. 1–8.

[19] C. Zhou, C. Peers, Y. Wan, R. Richardson, and D. Kanoulas, "Teleman: Teleoperation for legged robot loco-manipulation using wearable imu-based motion capture," *arXiv preprint arXiv:2209.10314*, 2022.

[20] M. Hirschmanner, C. Tsiourti, T. Patten, and M. Vincze, "Virtual reality teleoperation of a humanoid robot using markerless human upper body pose imitation," in *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2019, pp. 259–265.

[21] L. Fortini, M. Leonori, J. M. Gandarias, E. de Momi, and A. Ajoudani, "Markerless 3d human pose tracking through multiple cameras and ai: Enabling high accuracy, robustness, and real-time performance," *arXiv preprint arXiv:2303.18119*, 2023.

[22] M. Labbé and F. Michaud, "Rtab-map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation," *Journal of field robotics*, vol. 36, no. 2, pp. 416–446, 2019.

[23] J. Engel, T. Schöps, and D. Cremers, "Lsd-slam: Large-scale direct monocular slam," in *European conference on computer vision*. Springer, 2014, pp. 834–849.

[24] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE transactions on robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[25] W. Kim, P. Balatti, E. Lamon, and A. Ajoudani, "Moca-man: A mobile and reconfigurable collaborative robot assistant for conjoined human-robot actions," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 10191–10197.