Perception and Action Assistance for the Remote Control of Robotic Manipulation

by

Achyuthan Unni Krishnan

A Dissertation submitted to the Faculty of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Robotics Engineering

Committee in charge:

Prof. Jane Li. . . . . Robotics Engineering, WPI (Advisor)

Prof. Jing Xiao. . . . . Robotics Engineering, WPI

Prof. Berk Calli. . . . . Robotics Engineering, WPI

Prof. Carlo Pinciroli. . . . . Robotics Engineering, WPI

Prof. Karen Troy. . . . . Biomedical Engineering, WPI

July 2025

Worcester, Massachusetts

Perception and Action Assistance for the Remote Control of Robotic Manipulation

Achyuthan Unni Krishnan

ABSTRACT

Teleoperation systems offer tremendous potential in extending the reach of human abilities in domains such as healthcare, where remote manipulation can improve both access and efficiency. However, designing intuitive, efficient, and low-effort teleoperation interfaces remains a significant challenge, particularly for complex, high-precision tasks such as those in nursing. This dissertation explores the design and evaluation of control and perception assistances to enhance teleoperation performance, reduce physical and cognitive workload, and improve operator preference across a range of remote manipulation scenarios.

Through systematic user studies involving healthcare-relevant tasks and representative user populations, we first identify motion mapping as an intuitive and high-performing control interface. However, its benefits are offset by increased operator effort due to limited precision. To address this, we develop and evaluate action assistance strategies—including the separation of orientation and position control and environment-based motion scaling—which significantly reduce task completion time and body motions, while improving control precision and ergonomics. These approaches are extended to bi-manual teleoperation, demonstrating that both position and orientation supports yield workload and performance improvements. Additionally, assist-as-needed paradigms based on operator state (e.g., motion intent or physical activity) are shown to further improve performance and reduce frustration, highlighting the importance of adaptive interface design.

Recognizing that assistance may confuse the operator's understanding of robot autonomy, we further explore perception assistance through Augmented Reality visual cues to improve awareness and user experience. We show that the effectiveness and preference for these cues depend on the level of autonomy in the interface, and that training can shift user preferences to align with developer-recommended designs.

In conclusion, this dissertation offers design ideologies for building teleoperation systems tailored to healthcare contexts. These findings underscore the importance of adaptable, intuitive interfaces and highlight the value of workload- and intent-aware assistance. The work lays a foundation for future teleoperation systems that are accessible, efficient, and supportive of diverse healthcare professionals performing complex remote tasks.

To my family. Especially my parents Prema and Unnikrishnan. Thank you for your unconditional love, support and sacrifice to afford me the opportunities I am eternally grateful for.

Acknowledgments

I am deeply grateful to my advisor, Professor Jane Li, for her invaluable guidance, unwavering support, and mentorship throughout my PhD journey. Her dedication and encouragement through the highs and lows of this process have been instrumental in helping me achieve my goals and navigate this transformative phase of my life. I would also like to sincerely thank my dissertation committee—Professors Jing Xiao, Berk Calli, Carlo Pinciroli, and Karen Troy—for their insightful feedback and thoughtful suggestions, which have significantly enhanced the depth and quality of my research.

I would also like to thank Kristen Keane and the FORW-RD program for the support of my research This material is based upon work partially supported by the National Science Foundation (NSF) under Grant No DGE-1922761. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

A special note of thanks goes to all the user study participants, particularly the nursing students and healthcare workers. Their time, input, and feedback were critical to grounding this research in real-world needs and validating the applicability of teleoperation interfaces in healthcare settings.

I am thankful to all my past and present colleagues in the HiRo Lab—Kenechukwu, Zhaoyun, Nikita, Lorena, and Manav—for their camaraderie, collaboration, and the many meaningful conversations that shaped both my technical and personal growth. I am especially indebted to TC, who served not only as a colleague but as a mentor and confidant. Your unwavering support and wisdom guided me through some of the most challenging periods of my PhD journey, and I will always be grateful. Wishing you continued joy and success with ChiChi and Lilian. I extend my deepest thanks to my family, whose love, patience, and encouragement made this journey possible. I am equally grateful to my close friends—Ashay (the best roommate and test subject anyone could ever ask for), Abhijith, Abhinav, Zhaoyuan, Kailash, Joshua, and Shyam—for their support, understanding, and presence throughout these years. Aravind, thank you for everything you did for me. I hope, in whatever small way I could, I made you proud. You'll always be in my heart.

Finally, I would like to thank all those I may have inadvertently left unmentioned but who, in ways big or small, have contributed to my journey. I am excited and hopeful as I embark on the next chapter of my life and career.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Teleoperation and remote manipulation have the potential to address critical challenges such as workforce shortages and geographic isolation by extending the capabilities of teleoperators, especially in the healthcare domain. To enable widespread adoption, it is essential to develop control interfaces that enhance rather than hinder the work of healthcare professionals. Beyond control, the physical separation between the operator and the remote workspace introduces information gaps that must be addressed through effective perception interfaces. These interfaces should provide operators with sufficient situational awareness to complete their tasks reliably and efficiently. This underscores the importance of designing control and perception assistance systems that are efficient, intuitive, and informative—goals that are pursued through the *Design Objectives* outlined in the following section.

Design Objectives **—** Enable remote manipulation by identifying and validating action and perception assistance designs that improve control precision, reduce control effort in motion mapping control, improve robot and task state awareness and lower cognitive workload associated with remote manipulation for various human-robot control paradigms.

This research aims to identify specific design objectives for action and perception assistance in the human-robot control paradigm that improves operator performance, reduces physical and cognitive workload while improving the operators preference for teleoperation. In our prior work we compared and evaluated various teleoperation interfaces and identified free-form motion mapping control interface *easy to learn and use* and *improved task performance*. The free-form control interface designed using human motion capture reduced interface learning efforts and improved task performance thus making the interface more preferable and accessible to prospective users of teleoperated systems [4]. However, these motion mapping interfaces can lead to an increased physical workload due to the active nature of the interface, and reduced precision caused by limitations in motion capture techniques. To address these challenges, it is essential to identify action assistance mechanisms that enhance precision by reducing operational errors and lower control effort by minimizing operator motion and task completion time. Designing with a focus on reducing operator workload while maintaining or improving task performance requires a careful balance between automation and human input. Focusing the design to improve precision and reduce effort means lower operator workload while improving task performance via less task completion times and operational errors. Control interfaces should also be designed to provide assistance at correct moments, ensuring that such support does not hinder the operator or cause unnecessary confusion during operation. This highlights the need to investigate *assist-as-needed* interfaces that improve task performance and reduce operational workload without compromising intuitiveness or control. Perception assistance forms the other component of the human-robot control paradigm in addition to the teleoperation interface. The teleoperator receives their information about the workspace through information relayed by perception assistance(e.g visual interface/UI, haptic feedback, etc). *Reducing visual clutter* by providing information that the operator prefers according to the level of robot autonomy ensures the operator is not overwhelmed by information thus reducing cognitive workload. Cognitive workload can also be lowered by *reducing ambiguity of robot control* that occurs due to the introduction of automatic robot assistance. The introduction of robot assistance can lead to ambiguity regarding the level of control effort and operator engagement needed to teleoperate the robot [5, 6]. This ambiguity should be addressed by informing the user about the current robot state and helping the user better allocate the required mental effort needed for processing workspace conditions and robot operation. Finally, the information conveyed by the perception assistance should provide the operator necessary information about the workspace and thus *reduce operator control effort (reduced operator motion and task completion times*) to perform their task, especially when using low autonomy interfaces where the operator has greater control of robot operation. Attaining these desired features in perception assistance designs is crucial as it helps overcome the lack of sensory (haptic) or depth perception and mental workload of remote manipulation. These designs should be formulated through the evaluation of various visual cues for conveying information and by identifying users' preferred cues based on their level of engagement in robot control. In conclusion, implementation of these design objectives for both action and perception assistance can create a successful human-robot control paradigm that can improve adoption across various teleoperation tasks.

1.1 Design Objectives and Contributions

In order to identify the design philosophies that improve human-robot control paradigm, the following design objectives were investigated:

***Prior Work** – Design a control interface for a humanoid robot for nursing application that improves task performance and reduces learning effort.*

Limitations of Related Work — A tele-nursing robot should enable dextrous, precise and reliable motion control. Thus far, research efforts on tele-medicine interfaces are primarily focused on tele-surgical robots [7] where the teleoperation interfaces are focused on performing specific tasks. Control interfaces that span a spectrum of constrained to free-form control based on gamepad/joystick [8, 9], touchpad [10], Graphical User Interfaces (GUI) [11], stylus devices [12], exoskeletons [13], RGB-D cameras [14], customized cockpit [15], motion capture systems [16], etc have been used to control the motion of mobile manipulators and humanoid robots for nursing purposes. However, existing interface designs were implemented to work around specific limitations like:

Teleoperation Interfaces in Tele-nursing Robots		
Robot Platform	Type of Control Interface	Control Devices
Mobile manipulator, Remote manipulator	Constrained	**Gamepad [3],** Joystick[4], Touchpad [5]
Mobile Manipulator, Remote manipulator, Humanoid robot	Intermediate	Graphical User Interfaces (GUIs) [6], **Stylus Device [7]**
Mobile Manipulator, Remote manipulator, Humanoid robot	Free-form	Exoskeleton [8], RGB-D cameras [9], Customized cockpit [10], **Motion capture system [11]**

Prior Work: Design a control interface for a humanoid robot for nursing applications that improves task performance and reduces learning effort

	Task Completion times (Performance)	**Practice time (Learning effort)**	**Subjective Intuitiveness rating**	**Physical workload (NASA-TLX survey)**	**Control precision (subjective feedback)**
Constrained (Gamepad)	Slow	High	Low	Low	High
Intermediate (Stylus Device)	Slow	High	Low	Moderate	High
Free-form (Motion capture system)	Fast	Low	High	High	Low

Figure 1.1: Top: An overview of the contemporary robot platforms and types of control devices used for teleoperation of the robot system; Bottom: The features of 3 representative interfaces evaluated in [1]. This highlights the benefits of the intuitive and easy to use nature of free-form motion mapping control interfaces while resulting in higher physical workload and low control precision. In the work presented we aim to reduce this issue by developing action and perception assistance.

Figure 1.2: (From left-right) Representative control interfaces evaluated in user study across spectrum of constrained to free-form control [2]; Workspace clearing task; Muscle effort analysis using EMG sensors.

- Hardware constraints: The limited availability of hardware often results in poor visual perception, caused by reduced camera viewpoints and a narrow field of view [6, 17]. This leads to significant information loss [18] and diminished depth perception [19]. Additionally, motion control suffers due to the lack of precision and the unintuitive nature of robot control, which stems from the limited modalities in current user interfaces [20].

- Limited use cases: Many solutions are tailored to specific developer needs, such as constrained arm motions (e.g., handling a stethoscope [21]), robot navigation and guidance [22], or information gathering [23]. As a result, these designs are often less applicable to more complex, generalized motions.

However, for nursing applications which require more free-form teleoperation of complex motion coordination [8], interfaces should be easy to learn and facilitate improved performance for nurses who are not familiar with robotic interfaces to perform their duties. Thus, conducting a systematic evaluation of diverse control interfaces (representative interfaces ranging from constrained to free-form interfaces) in the context of general-purpose teleoperation tasks with active involvement from nursing workers is essential to identify human-robot control paradigm that enables easy to learn and easy to perform teleoperation interfaces.

Problem Statement — Teleoperation holds great promise in ensuring the safety of healthcare workers while improving their productivity. While several control interfaces within the spectrum of constrained to free-form control exist for robot teleoperation, we need to identify a control interface that is both easy for healthcare workers to learn—being subjectively preferable for teleoperators and requiring low learning time—while also enhancing task performance by reducing task completion times. To address this challenge, it's essential to conduct an interface evaluation that actively involves the intended end users, namely nurses and healthcare workers. The evaluation should identify the interface that is easy to learn and improves the performance for these individuals without teleoperation becoming an additional burden.

Proposed Methods — For identifying the operator preferred control interface three representative interfaces from across the spectrum of constrained to freeform control was evaluated (as described in Section 2.3). The three interfaces evaluated were a gamepad based interface (constrained control), a motion mapping interface based on motion capture technology (freeform control) and a stylus based device which lies between constrained and freeform control. The interfaces were used to control a mobile dual-armed humanoid robot to perform a "pseudo" nursing task of workspace clearing that involves aspects of teleoperation that are required for nursing tasks including: arm-hand coordination (object grasping), bi-manual coordination (handling deformable clothes) and navigation (moving the robot base in the workspace). In the user study (N = 8 nursing students), task completion time and and task errors (collisions with the table or objects, and inappropriate or failed grasping of objects) were recorded to evaluate the user's task performance. The learning effort for each interface was evaluated based on the practice time required to become familiar with the interface prior to experiment and post-study questionnaire. Electromyography (EMG) data of the participants active muscle groups was recorded to evaluate the participant's physical workload while performing the experiment. In this **prior work [24] where I am the second author**, I contributed by conducting the user study and post-study data analysis.

Key Findings and Novel Contributions — The comparison of the three representative interfaces identified that the motion mapping interface for bi-manual teleoperation took the least learning effort in terms of amount of practice time used. This was concluded based on the lesser practice time required by the participants during the training phase of the motion mapping interface compared to the gamepad and stylus based interfaces, which required significantly longer training times. The motion mapping interface also proved to facilitate the best task performance for the teleoperator as indicated by the lower task completion times compared to the gamepad and stylus based interfaces. The motion mapping interfaces resulted in improved task performance as this interface reduced task errors compared to the other two interfaces. Post-study survey corroborated the objective results as they chose the human motion mapping interface as the easiest one to learn and preferred using it as the robot teleoperation interface for future use.

However, subjective feedback from a post-study NASA-TLX survey and objective results from Electromyography (EMG) signals picked up from EMG sensors attached to important muscles on the operator's torso suggested increased physical workload due to the motion intensive nature of the motion mapping interface. Participants noted that fine motions required when precisely grasping objects and repeated motions due to operational errors were what caused the most physical workload. Thus while the motion mapping interface is easy to perform tasks with, the precision of the control interface needs to be improved with action assistance to make fine robot actions easier to execute. The participants also highlighted that limited information feedback from the camera video stream resulted in repeated trial and error motions to gauge depth and proximity to target objects resulting in increased control effort. Thus, it is also required to develop perception assistance designs that incorporate visual cues that enable operators to complete task objectives with increased workspace information while reducing required control effort.

Design Objective 1 (DO1) – *Design action assistance to improve control precision and reduce control effort of remote manipulation via motion mapping control interfaces [25, 26].*

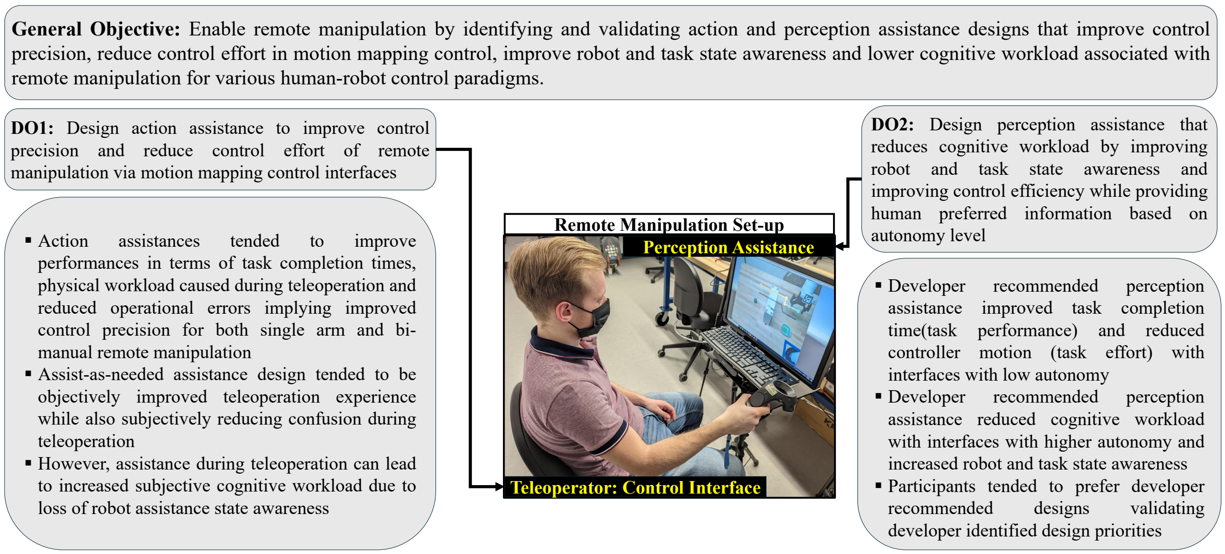


Figure 1.3: An overview of the design objectives for Action Assistance and Perception Assistance. **DO1:** Develop action assistance that helps reduce physical workload during operation while improving task performance; **DO2:** Develop perception assistance to reduce ambiguity due to the introduction of action assistance, reduce the cognitive workload and improve task performance during teleoperation.

Limitations of Related Work — Limitations in human motion tracking accuracy and the need to control multiple Degrees of Freedom (DOFs) can lead to interference between intended and unintended motions. For example, tremors in control input [27] or accidental orientation inputs affecting linear movements [28] can significantly reduce operator performance [29, 30]. Related work in 3D virtual object manipulation and teleoperation has shown that scaling motion (based on user's operating speed [31] or workspace conditions [32]) can improve control precision as required reducing operational errors. Motion constraints can be used to filter out unintended motions when teleoperating and this can be implemented by the separation of degrees of freedom in the design of interface mapping. In 3D virtual object manipulation this design has been achieved by on screen visual indicators like ring-and-arrow markers [33] or virtual handlebars [34] that enable the independent control of individual DOFs of the manipulated virtual object. Overall, our insight is that the control precision of remote manipulation can be improved by introducing appropriate motion *scaling* and *constraints*. However, contemporary literature highlights a limitation in current design ideologies—the inability to separate Degrees of Freedom (DOFs) in motion mapping interfaces and implement motion scaling for remote manipulation using motion mapping interfaces. Addressing this could enhance task performance by reducing task completion times, and lowering control effort by minimizing operator motions during remote manipulation. [35]. Thus, conducting an evaluation and validation of different designs of motion separation and motion scaling will help identify design ideologies that can be implemented to design future remote manipulation interfaces that improve task performance, operator effort and preference.

Problem Statement — While motion mapping interfaces are efficient and intuitive, they are often limited in control precision due to the inherent inaccuracies of human motion and the difficulty of simultaneously controlling multiple degrees of freedom, which can lead to unintended inputs [36]. These challenges, as discussed in the Limitations of Related Work section, contribute to increased workload and task errors. To address these shortcomings, the design of action assistance should aim to enhance task performance and control precision, reduce operator effort, and remain intuitive and preferable to use. Moreover, such assistance should be available in an adaptive and non-intrusive manner, ensuring that teleoperation performance is not hindered. This is especially crucial in bi-manual remote manipulation, where the demand for effective, intuitive, and efficient assistive strategies is significantly greater. Validating assist-as-needed interfaces in these complex scenarios is also needed to improve teleoperation experience.

Proposed Methods — Action assistance in the form of orientation and position motion separation and motion scaling can be used to improve the precision of remote manipulation when using motion mapping interfaces. Motion precision can be enhanced by separating position and orientation control such that unintended user motions do not interfere with intended control inputs [37]. Two design ideologies were implemented and evaluated to explore this approach: (1) a fully motion-based design where position and orientation were controlled using dominant and non-dominant hand controllers, respectively, and (2) a hybrid design where orientation control was implemented

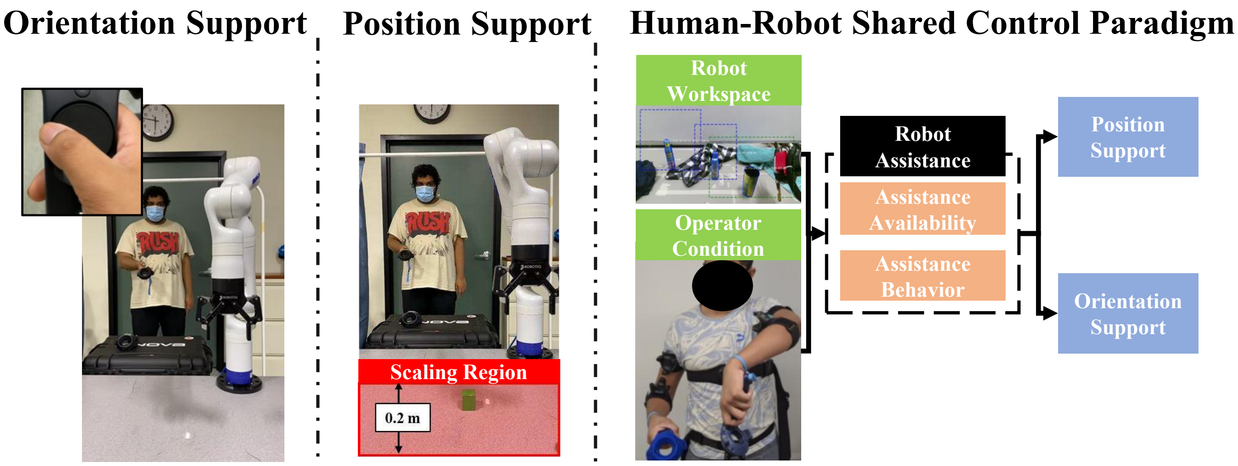


Figure 1.4: (From left-right) Orientation support in the form of orientation control using trackpad; Position support based on environment scaling; Human-Robot Shared Control Paradigm

via a discrete, constrained input (trackpad on the handheld controller). To further enhance precision, motion scaling techniques were evaluated. These included: (1) environment-based scaling triggered by constraints in the workspace and (2) user-based scaling triggered by the operator's motion speed. The experimental task involved pick-and-place of three small objects into a box, abstracting nursing-related tasks such as medication delivery and vital monitoring that require fine grasping and precise position and orientation control. These interfaces were first evaluated in a user study with $N = 8$ participants. The work was then extended to bi-manual remote manipulation, involving evaluation of seven control interfaces providing either position support (automatic, manual, or operator workload-based scaling) or orientation support (trackpad-based control or operator intent inference). In addition to quantifying the benefits of action assistance for bi-manual manipulation, this study also evaluated the benefits of assist-as-needed interfaces where assistance was conditionally triggered based on environmental or operator-centric factors. A second user study ($N = 16$ participants) was conducted, where participants performed two representative tasks: (1) opening a transparent pouch and (2) scanning and pouring the contents of a container. These tasks were designed to incorporate a variety of demands typical in bi-manual manipulation, including orientation control, precise actions, and gross movements. Across both studies, objective metrics such as operator workload and task completion time were recorded, alongside subjective evaluations via NASA-TLX, SUS questionnaires, and custom survey questions. All interface designs were finalized following a pilot study involving $N = 4$ experienced teleoperation users (each with over 100 hours of experience).

Key Findings and Novel Contributions— Separating orientation and position control can reduce operator workload and minimize unnecessary body motions by enabling precise control over individual degrees of freedom (DOFs), thereby avoiding interference from unintended motions. Among the evaluated interfaces, DOF separation implemented using two controllers (dominant hand for position, non-dominant hand for orientation) performed the best. This setup maintained the intuitive nature of motion mapping while improving control accuracy. In the pick-and-place user study, this interface reduced task completion time by 25% compared to the baseline, indicating improved efficiency. It also reduced body motion by 58% relative to the baseline and by 29% compared to the alternative DOF separation method using constrained orientation inputs via a trackpad. Since excessive body motion is associated with higher physical fatigue and operator frustration, these reductions are significant. The ability to individually control translation (e.g., object insertion, stacking) or orientation (e.g., adjusting the tilt of a telepresence camera) is critical for efficient manipulation in remote environments.

Regarding motion scaling, environment-based scaling was found to be both more effective and more preferred than operator-speed-based scaling. Environment-based scaling improved precision in constrained regions, reducing both control errors and unintended motions. Although it introduced slower robot motion during fine manipulation phases, task completion time still improved by 13% over the baseline interface. Post-study interviews revealed that users found environment-based scaling more predictable and controllable, whereas operator-based scaling was perceived as erratic and harder to manage due to unintended variations in scaling behavior.

These findings suggest that a balance of performance and manageable workload contributes to increased user acceptance of teleoperated systems. The combined benefits of DOF separation and environment-based motion scaling—both in terms of subjective preference and objective performance—offer valuable design insights for future remote manipulation interfaces. These strategies enable reduced control effort while enhancing task performance and usability.

When extended to bi-manual control, the benefits of action assistance were preserved. Interfaces providing position support significantly improved task completion times and reduced physical workload. Orientation support also lowered workload, although with minor increases in task duration. Notably, assist-as-needed strategies—triggering assistance based on operator activity levels or inferred intent—further reduced operational effort by up to 27.9% and 25%, respectively, in certain cases. These findings highlight that design principles established for single-arm control extend to bi-manual setups, reinforcing the value of assistance strategies tailored to operator state.

However, participants reported ambiguity about robot state and the activation of assistance, particularly in the bi-manual study where multiple assistance modalities (e.g., position and orientation support) were combined. This ambiguity negatively impacted subjective preferences and increased perceived workload. The issue was exacerbated in conditions involving higher autonomy levels, such as fully automated motions. To address this, perception assistance should be developed to enhance operator awareness of assistance activation and robot state, thereby minimizing cognitive burden and enabling users to adapt their control efforts effectively when assistance is active.

DO2 *– Design perception assistance that reduces cognitive workload by improving robot and task state awareness and improving control efficiency while providing human preferred information based on autonomy level [38].*

Limitations of Related Work — Recent related work in literature has provided a taxonomy for virtual, augmented and mixed reality for human-robot interaction [39]. Overall, the Augmented Reality (AR) visual cues can be augmented on the *robot*, *interface* and *environment*, in order to visualize the robot's internal and external states (e.g., internal reading and readiness, robot pose and location), as well as the robot's comprehension of *environment* (e.g., the purview, numerical readings, videos and images, 3D data from external sensors, robot-sensed/internal/user-defined spatial region,), *entities* (e.g., entity labels, attributes, locations and appearance), and *tasks* (e.g., heading, waypoint, call-out, trajectory, spatial preview, trajectory, alternation preview, command options, task statuses and instructions). Related work also recommends various effective integration of AR visual cues for manipulation tasks in the 3D environment (e.g., [40, 41, 42, 43, 44, 45]), these case-by-case designs did not provide insights into how AR visual cues should be structured for human-robot collaborative manipulation across different levels of autonomy. To improve the operator's remote perception, it is recommended to: (1) identify relevant information through data salience and clutter reduction (e.g., regions of interest), (2) detect anomalies in the workspace using statistical estimation, and (3) highlight risks by directing operator attention (e.g., identifying potential regions of failure in the workspace). Thus, it is crucial to evaluate the design of perception assistance tools used by operators in remote manipulation scenarios. By identifying Augmented Reality (AR) features that enhance operators' remote perception, we aim to improve their performance and reduce cognitive load. This involves streamlining visual information to minimize clutter while ensuring essential data remains accessible.

Problem Statement — The introduction of assistance to teleoperation interfaces has been shown to improve operator performance and increase their preference for teleoperated applications [25, 46]. However, action assistance can introduce ambiguity regarding the level of control the user has, leading to increased cognitive workload. Therefore, perception feedback must be adjusted to provide information that corresponds with the level of assistance. With varying levels of autonomy, users have different levels of engagement and participation in robot control. This means the amount of feedback desired by the operator, in terms of robot and task state awareness and operational guidance, varies accordingly. In interfaces with less robot assistance, perception assistance should

Table 1.1: Priority of AR features for each control mode.

Interface	Priority
Direct	Motion Guidance>Obstacle>Target
Assisted	Motion Guidance>Autonomy>Intent>Target>Obstacle
Shared	Autonomy>Motion Guidance>Intent>Target>Obstacle
Supervisory	Intent>Autonomy>Target>Obstacle

help reduce the control effort needed to complete the teleoperated task. Conversely, in interfaces with more robot assistance, perception assistance should focus on reducing cognitive workload by clarifying the robot's autonomy status and minimizing ambiguity, thus aiding the operator in performing task objectives effectively.

Proposed Methods — Different control interfaces, each corresponding to varying levels of autonomy, were used to identify preferred Augmented Reality visual cues for perception assistance. The control interfaces, arranged by increasing autonomy, included direct, assisted, shared, and supervisory control [47]. The participants were given several AR visual options for different aspects of teleoperation like motion guidance, intent of autonomy, assistance availability, environment and task objective based information. A systematic evaluation of the preferences for AR cues of the participants were recorded for a pick and place task where the operator also had to maneuver around an obstacle in the workspace. This task simplifies nursing tasks and involves aspects of navigating around workspace obstacles and object manipulation which contributes to elements of situational awareness and depth perception while teleoperating. To attain this design objective, pilot studies (N = 4, over 100 hours of teleoperation experience) were conducted. Based on the level of autonomy (direct, assisted, shared, and supervisory control), perception assistance recommended by developers was implemented for each control interface to complete the aforementioned pick-and-place task with priority for the information conveyed by the perception assistance for the different interfaces ordered as seen in Table 6.3. Information in the form of Augmented Reality features were provided for the various control interfaces with varying levels of priority given to information pertaining to motion guidance of the robot, robot autonomy state and intent, obstacle and target information. Depending on the level of control the operator had, which varies with level of robot autonomy, the salience with which this variety of information was modified. A user study (N = 18 participants) was conducted to evaluate the developer recommended designs and ideologies and identify preferred AR visual features among the general user population. In this study, the controller motion was recorded to estimate the control effort required to teleoperate the remote manipulator using the interface with no autonomy (direct control). The operator's variation in pupil diameter was recorded to estimate the operator's operational stress and mental workload while teleoperating the remote manipulator with different control interfaces and perception assistance.

Key Findings and Novel Contributions — As the level of autonomy transitions from direct to supervisory control, we found that: (1) humans' priority for AR visual cues shifts from guiding robot control to communicating autonomy activation and intention; (2) humans' preference for AR visual cues change from offering local information to offering global guidance in robot control as indicated by a large portion of participants (13 and 12 out of 18) selecting local information for direct control and global information for shared control. However, the use of global information to display autonomy's intention remains consistent across all interfaces with autonomy; (3) the efficacy of the AR features that share the same purpose as the robot autonomy is decreased as observed by the participants preferring not to have any AR cues for both obstacle indication and target hints in supervisory control mode. These design preferences of the users also aligned with the suggested perception assistance design recommended by the experienced users for the user study task identified during the pilot study phase (refer Table 6.3). With training and experience, the participants' final preferences for AR visual cues in perception assistance tended to align with the recommended design. In addition to being subjectively preferred by users, the recommended perception design ideologies also provided objective benefits. For the direct control interface, where users controlled the entire robot motion, participants who preferred the developer-recommended local information for motion guidance had significantly reduced controller motion (45% less controller motion) to complete the task compared to those who preferred alternative visual cues. This indicated that less control effort was required by the user to complete the task as they had more information of the workspace relayed to them with their preferred AR visual cues. Additionally, participants who used the experienced-user recommended perception assistance had lower cognitive workload compared to participants who used alternative perception assistance designs when using interfaces with robot assistance. During the post-study survey, participants indicated that obvious notifications for robot assistance activation in high-autonomy (shared control) interfaces ensured they did not have to exert effort when autonomy was active. These designs notified users of assistance activation in a user preferred manner clearing ambiguity in robot control in interfaces with robot assistance. This reduced their processing workload during these periods, resulting in significantly lower cognitive workload (25% less cognitive workload) compared to alternate perception assistance designs. There was also a tendency for reduced cognitive workload in assisted and supervisory control interfaces. The developer-recommended designs helped users focus their mental workload on processing visual information only when necessary. Additionally, perception assistance with reduced visual clutter and appropriate user preferred information feedback can assist with lowering cognitive workload as noted by the reduced pupil-based cognitive workload during the user studies. Thus, the design principles identified and validated in this user study can contribute to the development of perception assistance tailored to the level of action assistance provided. This approach enhances control efficiency, especially in scenarios with low autonomy levels, and reduces cognitive workload in interfaces with higher autonomy levels.

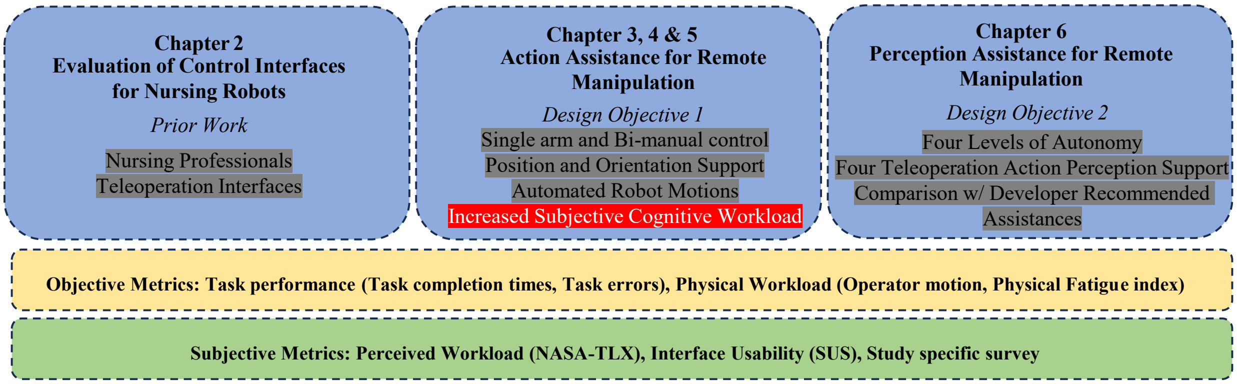


Figure 1.5: Dissertation Overview

1.2 Dissertation Structure

Chapter 1 — This chapter introduces the focus of the dissertation and an overview of the Design Objectives and corresponding findings and observations.

Chapter 2 — This chapter discusses prior work where different control interfaces for a bi-manual nursing robot were evaluated. Designs from more constrained to free-form interfaces were evaluated based on objective metrics and subjective feedback from nursing professionals. In this work where I am the second author, I contributed by helping implement control interfaces, aid conducting the user-study, and assist in data analysis. Findings in this chapter were published in the following work:

[1] T. C. Lin, **A. U. Krishnan**, and Z. Li, ”Intuitive, Efficient and Ergonomic Tele-Nursing Robot Interfaces: Design Evaluation and Evolution”, *ACM Transactions on Human-Robot Interaction (THRI), 2022.*

Chapter 3 — This chapter introduces the action assistance designs for a single-arm remote manipulation interface. A comprehensive evaluation of motion scaling and motion separation action assistance to improve control precision, performance and subjective preferences for remote manipulation was performed These findings were published in:

[1] **A. U. Krishnan**, T. C. Lin, and Z. Li, ”Improve the Control Precision of a free-form Telemanipulation Interface”, *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2022*.

Chapter 4 — This chapter extends the findings of Chapter 3 to bi-manual remote manipulation. Different position and orientation support designs along with different methodology for assistance availability were evaluated. These findings will be published in:

[1] **A. U. Krishnan**, and Z. Li, ”Towards Intuitive and Efficient Bi-manual Remote Manipulation: Adaptive Assistance to Improve Performance and Reduce Physical Workload”, *Submitted to IEEE Robotics and Automation Letters, 2025*.

Chapter 5 — This chapter describes the benefit of motions in a structured task being autonomously executed. However, these interfaces also resulted in increased ambiguity and uncertainty in robot state for the teleoperator requiring increased perception assistance. In this work where I am the second author I helped implement the shared autonomy and conduct the user study. These findings were published in:

[1] T. C. Lin, **A. U. Krishnan**, and Z. Li, ”Shared Autonomous Interface for Reducing Physical Effort in Robot Teleoperation via Human Motion Mapping”, *International Conference on Robotics and Automation (ICRA), 2020*.

Chapter 6 — This chapter presents the impact of autonomy on the operators preferred visual cues for information feedback in the form of perception assistance. The chapter expands on the design of the different control interfaces used in the evaluation process and the various AR visual cues available to the participant. The findings in this chapter are presented in the following paper:

[1] **A. U. Krishnan**, T. C. Lin, and Z. Li, ”Human Preferred Augmented Reality Visual Cues for Remote Robot Manipulation Assistance: from Direct to Supervisory Control”, *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2023*.

Chapter 7 — This chapter concludes the dissertation by highlighting the conclusions and key takeaways of all previous chapters while also examining the broader impacts of this work and the potential for future work.

Chapter 2

Prior Work: Evaluation of Control Interfaces for Tele-nursing Applications

2.1 Motivation

While tele-nursing robots provide a promising safe and cost-efficient approach for patient caring in quarantine areas [48], there has been limited research in identifying the desirable characteristics for control interfaces that allow healthcare workers to performing their nursing tasks remotely. Recent advances in human-robot interfaces offer various options for teleoperating advanced robots with navigation and manipulation capabilities. However, bridging the gap between interface development and practical application in the field requires identifying desirable interface characteristics. Designing teleoperation interfaces with direct involvement from end-users, particularly healthcare workers, is crucial to ensuring usability and effectiveness in real-world scenarios.

For the work in this chapter, we present a user study that involves healthcare workers and compare three control interfaces across the spectrum of types of teleoperation interfaces. With the feedback of the participants the different control interfaces were scored based on their intuitiveness, performance in the evaluation task, subjective workload and preference for the interface. Such an evaluation of control interfaces can help identify desirable design ideologies that can be adopted to develop tele-nursing systems that can be applied in nursing situations.

2.2 Literature Review

Tele-nursing robots provide a promising safe and cost-efficient approach for patient-caring in quarantine areas [48]. The recent outbreaks of infectious diseases like Ebola [49], Zika [50] and recently COVID-19 [51] necessitate the evolution of tele-nursing robots for the nursing workplace [52]. However, in order to effectively control the tele-nursing robot effectively, an efficient or robust control interface is required.

Table 2.1: An overview of commercial and prototype tele-nursing robots, with their Patient Assessment (A), Communication (C), Navigation (N) and Manipulation (M) capabilities.

Robot platform	Capability	Autonomy	Interfaces
Mobile telepresence [53, 54, 55, 56]	A/C/N	Self-navigation Obstacle avoidance Human-following	Touchpad, joystick
Mobile manipulator [8, 11, 57]	A/C/N/M	Self-navigation Obstacle avoidance Pick-and-place	Gamepad, stylus, GUI, touchpad, motion capture system
Humanoid [58]	A/C/N/M	NA	Exoskeleton

Thus far, research efforts on tele-medicine interfaces primarily focus on tele-surgical robots [7]. Since most of the surgical robots are focused on specific procedures (e.g., [59]), the interface design as well as the evaluation methods and metrics tend to be platform- and task-specific. The state-of-the-art interfaces and teleoperation assistance methods for complex robot platforms (e.g., mobile manipulators and humanoid robots) could be considered for tele-nursing robots of higher motion capabilities [16]. Teleoperation interfaces can be designed to range from point control to free-form control [60] where different levels of control modality and features. For the online control of motion coordination, these interfaces either: 1) discretize robot control across functions (like gamepad [8], keyboard and mouse [61], On-screen GUI [62], touchscreen motion [63], etc); 2) use motion of handheld objects like stylus devices [64], and joysticks [65]; 3) map the human motions to the robots (using customized cockpits [15], commercial virtual reality and gaming controllers [66, 67, 68, 69], soft/hard exoskeletons [70, 71] and data gloves [72, 73], marker-less or marker-based motion capturing device [70, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84]), or map human motor commands using myoelectric devices [68, 85, 86] and brain-computer interfaces [87, 88, 89] (refer to example representative interfaces in Table 2.1).

Limitations in Literature – An optimal control interface for tele-nursing robots should enable dexterous, precise, and reliable motion control. However, existing literature presents various control interfaces designed with specific constraints to meet developers' particular use cases. Conducting a systematic evaluation of diverse control interfaces in the context of general-purpose teleoperation tasks, with active involvement from nursing workers, is crucial. This evaluation aims to identify the ideal design features for tele-nursing systems.

2.3 Telerobotic System and Control Interfaces

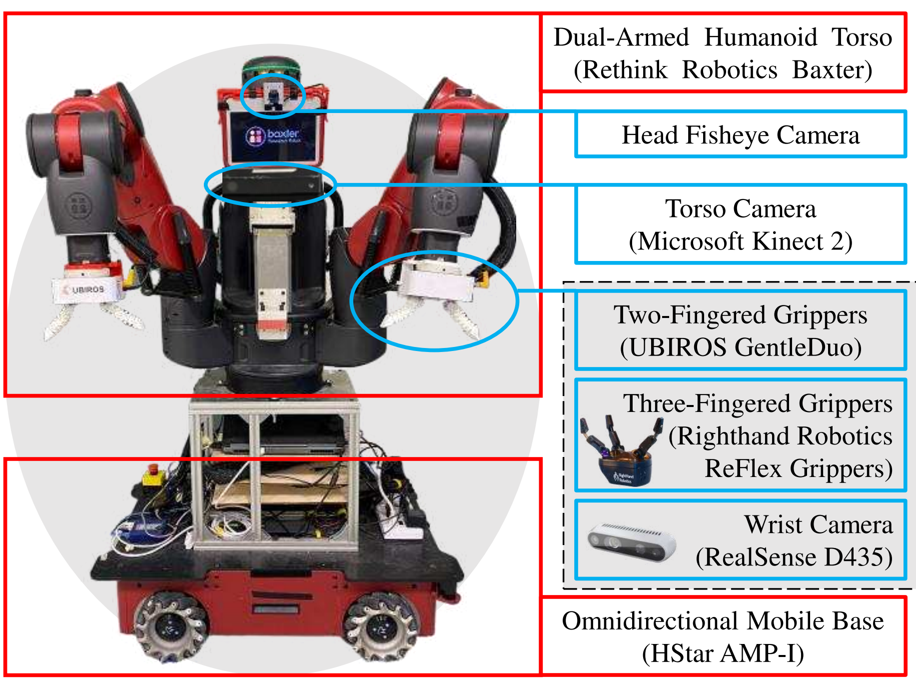


Figure 2.1: Tele-robotic Intelligent Nursing Assistant (TRINA) system.

Tele-Nursing Robot System – This section describes the robot platform. Shown in Figure 5.1, the Tele-Robotic Intelligent Nursing Assistant (TRINA) consists of a dual-armed humanoid torso (Rethink Robotics Baxter), and an omnidirectional mobile base (HStar AMP-I). For grasping objects, two Righthand Robotics Reflex grippers were used in the experiments evaluating physical fatigue indices and benefits of automating grasping while two two-fingered soft grippers (UBIROS GentleDuo) were used for evaluating teleoperation interfaces. The visual sensor suite consists of a ELP-USBFHD01M 180° fisheye camera mounted on the robot head and two Intel RealSense D435 cameras on the wrists.

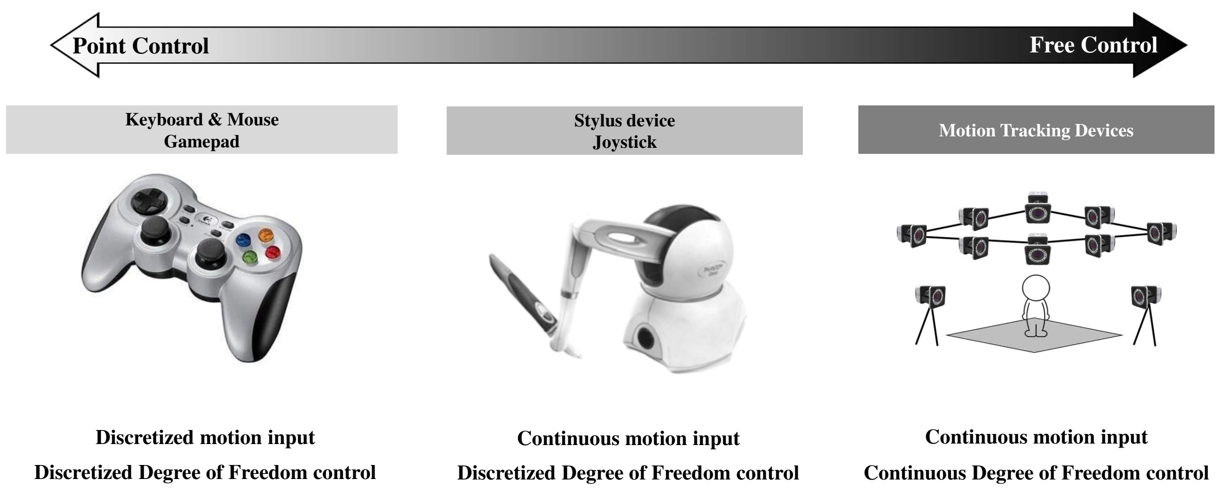


Figure 2.2: Examples of control devices across the spectrum of teleoperation control interfaces.

Teleoperation Interfaces – In the following section the different control interfaces used to represent different levels on the spectrum of teleoperation control as seen in Figure 2.2 will be presented.

Gamepad – The gamepad interface belongs to the point control type of teleoperation interfaces. Point control interfaces enable the operator to specify the point in the workspace to which the robot must move to. As seen in the interface developed for controlling TRINA (refer Figure 2.3, the robot arm and base motion is controlled one axis at a time with the motion of the joysticks and the control of different functions is spread across different buttons.

Stylus Device – The stylus device presents a point between point and free-form control. As seen

Figure 2.3: Gamepad controller configuration for teleoperation interface.

Figure 2.4: Stylus device (Geomagic Touch) configuration for teleoperation interface.

in Figure 2.4, the motion of the stylus interface controls the motion of the robot arm on multiple axes. In this design, each stylus device controls one of the robot arms, or the robot base motion depending upon what function is currently active.

Motion Mapping Interface – Motion mapping interfaces represent the free-form teleoperation type interface. For this user study, the interface was developed using VICON motion tracking and the interface could control all 6-degrees of freedom for both the robot arms and the motion of the robot base at the same time. The controls of the designed interface is seen in Table 2.2

Graphical User Interface — A Graphical User Interface (GUI) is used to provide the teleoperator with the video stream from the fisheye camera and the two wrist cameras. The video stream and the control GUI are presented to the user on a monitor at the teleoperation workstation. This video stream provides the teleoperator with a real-time view of the workspace. The GUI also tells

Table 2.2: Motion Mapping Teleoperation Interface. The arm posture is measured by the swivel angle, i.e., the rotation of the elbow position with respect to the axis connecting the shoulder and wrist positions [3].

Teleoperation Input	**Robot Function**
Robot's Upper Body	
Hand position and orientation	End-effector position and orientation
Arm posture and orientation	Manipulator arm posture
Rotate upper body	Rotate mobile base orientation
Hand open/close	Gripper opens/closes
Right shank flexion	Activate teleoperation assistance
Robot's Lower Body	
Squat	Engage/Disengage teleoperation
Leg steps forward/backward	Mobile base moves front/back
Left (right) leg steps left (right)	Mobile base moves left (right)
Lift right leg	Switch the camera view

which mode the operator is in while using the gamepad and stylus interface or which arm is being controlled when in the gamepad mode. The current robot state is also provided as a 3D model in a simulated environment.

2.4 User Study

The following user study tries to identify the desirable interface among the three control interfaces described in the previous section. The goal of this experiment is to find a suitable teleoperation interface that is intuitive and easy to learn for the primary users (nursing workers) who usually do not have an engineering background.

2.4.1 Participants

Our user study involves (N=8) nursing students (eight women, 19-21 years old) who represent future users of the teleoperation interface. We also recruited three registered nurses to get their feedback and attitude towards the use of tele-nursing assistive robots on a daily basis. All the participants have experience working in healthcare and are familiar with the hospital environment as either nursing students or experienced healthcare workers. They also do not have any robotics or engineering expertise and have almost zero gaming experience with the gamepad controller. These participants provide valuable insight into the preferences of future frontline healthcare workers who are likely to interact with teleoperation systems in clinical settings. Their feedback can help tailor the interface to better support usability, workflow integration, and user acceptance among similar end users. The experimental protocol was reviewed and approved by the Worcester Polytechnic Institute Institutional Review Board.

2.4.2 Experimental Procedure and Tasks

Figure 2.5: The tasks of the user study include collecting a single object (left), and cleaning and organizing a counter workspace (right).

The participants in the user study performed two tasks (see Figure 2.5). The *testing task* (left) is to collect an individual object on the counter workspace and is designed to examine the user's learning effort and the outcome of practice time for each interface. This task is performed both before and after the practice session to evaluate the learning outcome. The *evaluation task* (right) is to clean and organize several objects scattered in the workspace and is designed to evaluate the interface usability. Our prior study shows that most of the tele-nursing tasks require the users to perform: 1) free control of reaching-to-grasp rigid and deformable objects, 2) free control to move the mobile base to facilitate manipulation, and 3) point control to facilitate engaging/disengaging the interface or mode switching. Based on this finding, we set up a “pseudo-task” which incorporates these necessary teleoperation skills in the context of workspace cleaning and organization by a nurse. Specifically, the user will teleoperate the nursing robot to collect several rigid and deformable objects randomly placed on a counter workspace and sort them into two separate bins. This task integrates the free control of precise and gross manipulation in a cluttered environment, locomotion, and point control of robot states, all of which are necessary primitive robot control skills for tele-nursing tasks.

2.4.3 General Evaluation Metrics

In the user study, the intuitiveness and learning effort for each interface was identified by the time spent by the participant in the training phase. The time required to complete the task and the number of errors committed while performing the task were used to estimate operator performance with a control interface. These errors include: (1) dropping or knocking over objects, (2) inappropriate grasps and (3) collisions with the table. Participants were also required to answer simple arithmetic questions as they were performing the task. Lower cognitive workload while teleoperating results in lower response times for the arithmetic questions.

We adopted NASA-TLX as the workload assessment tool that helps record the user’s self-evaluation of Mental, Physical and Temporal Demands, Performance, Effort and Frustrations [90]. We further calculated the overall NASA-TLX score to identify the subjective workload by weighting each effort demand. The weighting coefficients were generated by choosing from a series of pairs of rating scale factors that were deemed to be important based on the official instructions. In addition to the NASA-TLX evaluation form, the comprehensive custom questionnaire is an integral part that captures the users' feedback and attitude toward the newly implemented methods and interfaces. Unlike the traditional customized questionnaire in human-robot interaction, we performed a post-study interview to identify the causing factors and features that will improve the interface usability.

2.5 Feasible Interface to Control Tele-nursing Robot

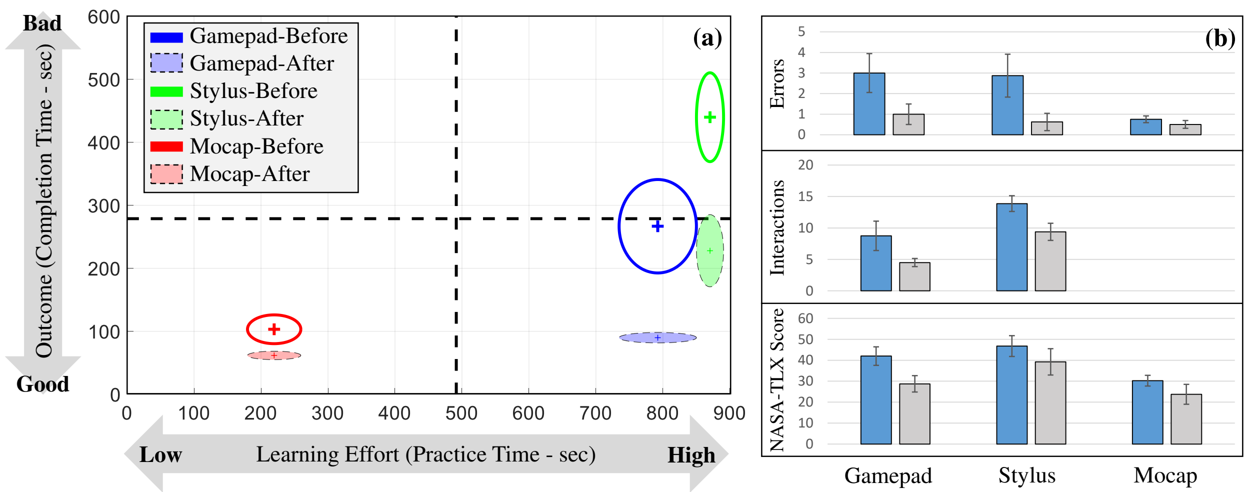


Figure 2.6: (a) Practice time vs Completion time for nursing students. (b) Comparison of number of errors, interactions, and subjective workload (NASA-TLX) across interfaces for nursing students.

Learning Effort – Figure 2.6(a) shows the comparison of interfaces in terms of learning effort. The dotted line indicates the mean of the maximum and minimum value from all participants for both learning effort and outcome. For N=8 nursing students, the motion mapping interface (Mocap) has a better learning outcome and lower learning effort, compared to the other interfaces. Each ellipse plots the mean and standard deviation of the testing task completion time with respect to the mean and standard deviation of the user's practice time. The red, green, and blue colors are for the motion mapping interface (Mocap), stylus and gamepad, respectively. The learning outcomes can be found by comparing the ellipses of the same color. On average, the nursing students spent less time (219±39 sec) learning the Mocap interface than the gamepad (792±57 sec) and stylus device (870±20 sec) interfaces. ANOVA analysis showed that: 1) The learning effort for Mocap was significantly lower than that for the gamepad ($F(2,21)=71.137$, $p<0.001$) and stylus interfaces ($F(2,21)=71.137$, $p<0.001$); 2) The Mocap interface also had the least completion time (61±6.7 sec) after practice for the testing task, followed by the gamepad (90±8.2 sec), and then the stylus (228±57 sec); 3) In the evaluation task, the Mocap interface also has a significantly faster completion time than the gamepad ($F(1,14)=6.979$, $p<0.05$) and stylus interfaces ($F(1,14)=8.296$, $p<0.05$). We also noticed that the completion time for the testing task before the practice was significantly slower than after practice, when using the gamepad ($F(1,14)=5.624$, $p<0.05$) and stylus device ($F(1,14)=5.442$, $p<0.05$). The significant effects of practice for these two interfaces were confirmed by the participants' reports in the post-study interview. The effect of practice is also more significant for the gamepad than for the stylus interface, which may be because the gamepad is a widely used gaming interface for the public. However, the participants also report that it is difficult to remember the many different functions associated with the gamepad buttons. On the other hand, the Mocap interface has a low learning effort and the trivial effect of practice indicates that this interface is the most intuitive one for nursing robot teleoperation.

In Figure 2.6(b), the Interactions field refers to the number of times a mode has to be switched while using the interface. A switch from controlling the base of the robot to the right arm of the robot is considered as an interaction when evaluating the gamepad and stylus interfaces. However, for the Mocap interface since the operator can control all aspects of the robot functionality at the same time, this part of the graph is empty as no interactions/mode switching is required.

Task Performance – Figure 2.7(a) compares the performance in the evaluation tasks among all the interfaces, using the following objective metrics: 1) the completion time of the evaluation task, and 2) the response time for each math question (i.e., the secondary task). The dotted line indicated the mean of the maximum and minimum value from all participants for both cognitive workload

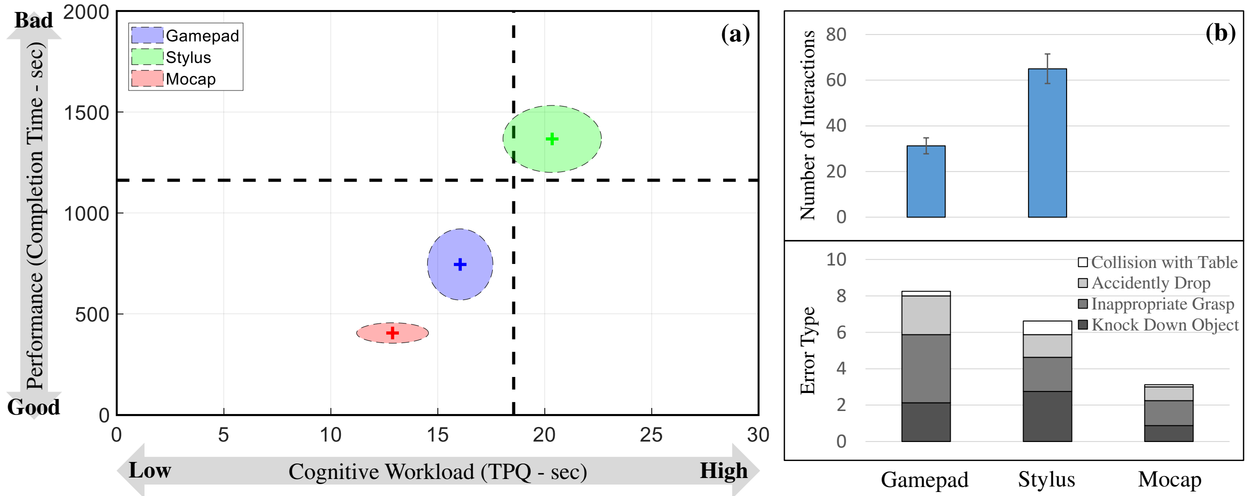
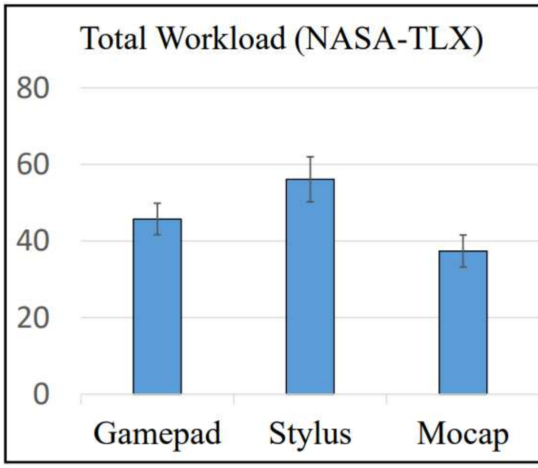


Figure 2.7: (a) Completion time vs Cognitive workload based on the time per question answered for nursing students. (b) Number of interactions and type of errors for nursing students.

and performance. For nursing students, the task completion time using the Mocap interface was less (404±50 sec) than the gamepad (745±176 sec) and stylus (1367±165 sec). ANOVA analysis shows the significant differences in completion time between Mocap and stylus ($F(2,21)=11.667$, $p<0.001$) and between gamepad and stylus device ($F(2,21)=11.667$, $p=0.015$). The results also indicates that the subjects took less time to solve the arithmetic questions while using the Mocap interface (12.8±1.6) than the gamepad (16±1.5) and stylus device (20.3±2.9).

Figure 2.7(b) further compares the interfaces by the number of errors and mode switches for the evaluating task. The nursing students have fewer total operation errors using Mocap interfaces (3.1±0.6), compared to using the gamepad (8.6±1.7) and stylus (6.6±1.7). The ANOVA analysis indicated significant differences between the Mocap interface and gamepad ($F(2,21)=3.513$, $p<0.05$). The breaking-down of error types shows that the Mocap and gamepad interfaces tend to cause more inappropriate grasps. Our post-study interviews show that this is due to the lack of depth perception in the visual feedback. Additionally, the stylus interface requires significantly more mode switches during operation than gamepad because the user has only one button to cycle between the hand, arm and base control. The Mocap interface does not need any mode switching as all robot components can be controlled simultaneously via whole-body motion mapping.

Subjective Operation Workload



	Gamepad	*Stylus*	*Mocap*
Mental Demand	64.4±6.9	75.6±6.9	36.2±5.4
Physical Demand	17.5±2.8	26.9±7.1	43.8±7.8
Temporal Demand	34.4±8.4	32.5±8	23.1±5.9
Performance	40±9.4	41.9±5.9	32.5±4
Effort	53.1±6.6	68.1±8.7	38.1±5.2
Frustration	47.5±6.3	55±7.5	28.8±6.1
Weighted Workload	**45.8±4.1**	**56.1±5.8**	**37.4±4.2**

Figure 2.8: Subjective workload (NASA-TLX) for nursing students.

While the motion mapping interface contributed to improvements in the mental demand required to teleoperate the robot, users also mentioned a significant physical workload (43.8 ± 7.8) incurred because of the physical nature of the teleoperation interface compared to the gamepad (17.5 ± 2.8) and stylus device (26.9 ± 7.1).

2.6 Summary and Outlook

We showed the advantage of using human motion mapping as the tele-nursing robot control interface by comparing it with a handheld gamepad and stylus-based device. The lower learning effort among nursing workers new to robot operation and the intuitive freeform motion control could also make the teleoperation experience more immersive. The post-study interviews also mentioned that participants appreciated being able to control the robot arms to perform complex orientations with ease. It is understandable that in a realistic patient-caring scenario, the nursing workers often need to take care of multiple tasks simultaneously. Nevertheless, we noticed that the control interface using a gamepad is suitable for more structured tasks since it can precisely and slowly control each motion for each degree of freedom. Interviews of the participants indicated that the ability to use different buttons to control different functions of the robot made teleoperation simpler. They felt this would prevent unintended motions of the arms or the base. The participants also stated that the ability to move the robot in small discrete increments via the joystick input gave them more confidence while performing delicate operations like picking up objects. This is unlike the motion control observed with the Vicon interface and the stylus-based interface where precise control is harder to achieve. On the other hand, when the workspace of the task is limited to a certain area, the stylus device will be a good fit to control the robot to perform tasks with small movements. Additionally, the buttons on the stylus hardware helps integrate discretized base and arm control while also enabling teleoperation through intuitive motion mapping of the stylus. Compared to the gamepad interface the motion mapping of the stylus interface was reported as being more intuitive to use according to some participant surveys. However, for nursing tasks that involve unstructured and large range of movement (e.g. laundry in the cluttered environment), freeform control (human motion mapping) is preferred.

However, there are still some issues with the motion mapping interfaces that need improvement. The ability to precisely control the robot is lost because of the free-form multi-modal operation facilitated by this interface. This results in more motions to correct the robot control which in turn results in increased body motion on the operator's side to provide this correctional input. Thus increasing the precision of the motion mapping interface and reducing the physical workload caused by the control interface will improve the acceptance of such teleoperated systems in the healthcare field.

Chapter 3

Action Assistance in Enhancing Motion Mapping Control for Tele-manipulation

3.1 Motivation

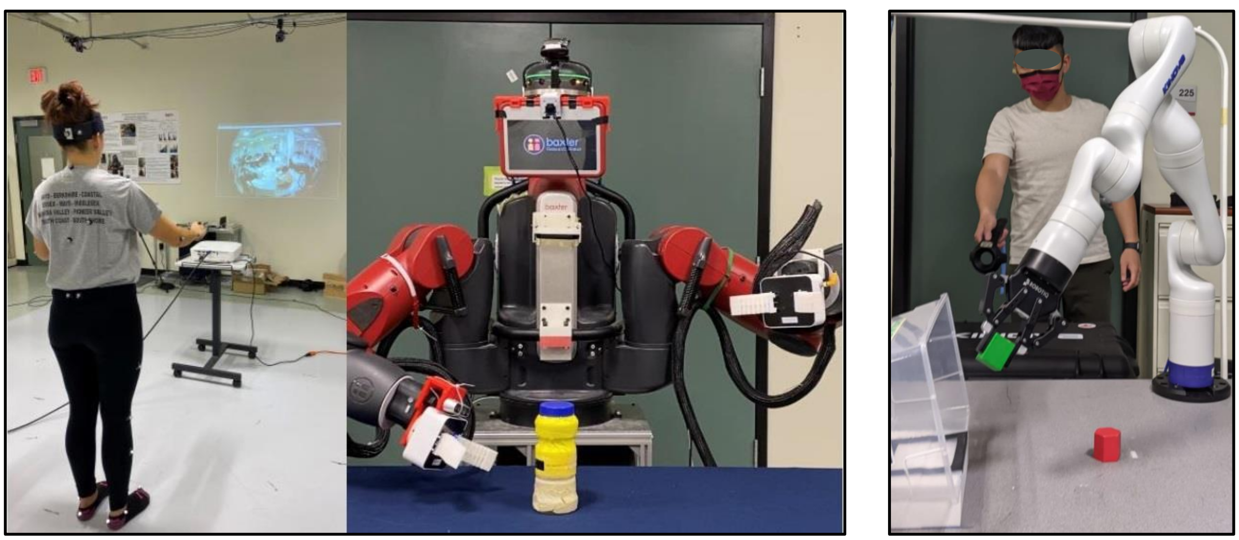


Figure 3.1: Robot teleoperation via motion tracking interfaces: Vicon motion capture system (Left); The HTC Vive hand-held controller (Right).

As seen in Section 2, motion tracking devices enable humans to use their natural motions to control remote robots. Such interfaces are efficient and intuitive for free-form robot teleoperation, because of how easily the operators can (learn to) control motion coordination of complex robot platforms [24]. However as seen in Section 2.6, controlling precise tele-manipulation using motion tracking interfaces is usually difficult, because of the simultaneous control of many degrees-of-freedom (DOFs). For instance, when the teleoperator needs to precisely adjust the pose of a remote object, the intended control of the object's position may cause unintended control of orientation, or vice versa. The control precision of tele-manipulation is also limited by the level of accuracy that human motion control can achieve. Since teleoperators use their own body or body parts for teleoperation control, their arms and hands may easily get tired, which prevents them from working on robot teleoperation for prolonged time and on a daily basis. The difficulty in precise tele-manipulation increases the task completion time, operation errors, additional cognitive and physical workload, and effort to learn interface controls, leading to reduced acceptance of these interfaces by novice users.

In this chapter we investigate two types of approaches to improve the control precision of tele-manipulation via motion tracking, without compromising on the control efficiency and intuitiveness. Related work in literature has proposed and validated various case-by-case interface designs that improve control precision by separating the controlled degrees-of-freedom (DOFs) [91], introducing motion constraints [92, 93] or using motion scaling methods [94, 95, 96]. However, prior research hasn't systematically evaluated and compared the many variants of these two types of approaches, and concluded generalizable interface design theories. To address this limitation, we compared several approaches to separation of controlled DOFs and motion scaling preferred by users in our pilot study, evaluated their performance and workload in the precise free-form tele-manipulation of 3D objects, and examined the factors that influence the users' preference. The main contribution in this chapter is to develop a framework of methods to improve the control precision of motion tracking tele-manipulation interfaces, and validate this framework in user studies.

3.2 Literature Review

Related work in robot teleoperation and 3D virtual object manipulation have provided many human motion tracking interfaces suitable for the efficient and intuitive control of free-form telemanipulation tasks. These interfaces, ranging from expensive motion capture systems [84] to touch screens and gamepads [8, 64, 96, 97], may capture multi-finger coordination [98, 99, 100, 101, 102], arm-hand coordination [64, 99, 103, 104], bi-manual coordination [105, 106] and whole-body coordination [46, 107, 108] They allow users to teleoperate mobile, manipulator, and humanoid robots or maneuver 3D objects in Virtual or Augmented Reality. While being efficient and intuitive, motion tracking interfaces are generally limited in their control precision, because human motions have limited accuracy, and because simultaneous control of many DOFs may cause the interference of intended and unintended motions [36]. The efficiency of controlled motions can be improved by various techniques like introducing constraints in terms of virtual fixtures [109, 110, 111] or autonomy for teleoperation assistance (e.g. collision avoidance [92, 93, 112, 113], motion guidance toward the intended goal [46, 106, 114, 115]).

Action Augmentation to Improve Control Efficiency — Interfaces that map gestures or point-and-click actions to autonomous robot motions or movement primitives [33, 60, 116, 117] can all be considered as some kind of constraints that limit the extent to which human operator can control the robot motion freely. Moreover, motion constraints can also be introduced by the separation of degrees of freedom (DOFs) in the design of the interface mapping. For instance, people may use *separate controllers to manipulate a 3D object's position and orientation*, to avoid the interference of intended and unintended motion control [91]. Some interface hardware, such as the trackpad of hand-held controllers, the joystick of gamepads, are naturally suitable for the separation of DOFs because they can clearly distinguish the control inputs for different motion directions. For screen- or projection-based interfaces, interactive avatars such as the ring-and-arrow markers [33, 60, 118, 119, 120, 121], the virtual handlebar [34, 94, 122, 123] enable the independent control of individual DOF(s) of the manipulated (virtual) object or robot end-effector. Alternatively, to improve the control precision, the motion tracking interfaces may *scale down the speeds and ranges of controlled motions*. The scaling ratios can be fixed (commonly used by tele-robotic laparoscopic or eye surgery interfaces [95, 124]), or vary with the user's operating speed (e.g., PRISM method [31, 125]) or regions of operation [64, 126].

Limitations and Research Questions — Overall, our insight is that the control precision of tele-manipulation can be improved by introducing appropriate motion *scaling* and *constraints*. This section will investigate two open research questions regarding the **design of interface mapping** for motion tracking interfaces: **RQ1**) How to appropriately separate the controlled DOFs, and **RQ2**) How to adjust the motion scaling ratio. For generalizable results, our proposed interface designs will be implemented on the hand-held controller of HTC Vive Virtual Reality system, which is a representative motion tracking interface that integrates free-form motion control (via hand motion tracking), directional motion control (using trackpad) and discrete state switching (using buttons). Our evaluation will use a general-purpose task that involves the free-form control of precise reaching, grasping and placing of small objects, which are the skills required in many tele-manipulation task scenarios.

3.3 Interface Design and Process

In this work, we used an interactive design process which leveraged findings from a pilot study conducted along with the implementation of interfaces to continuously improve the interface design.

3.3.1 Primitive Designs of Action Assistance

Our interactive design process starts with the baseline design and the seven primitive designs of interface mapping that address the control precision problem of motion tracking interfaces. The **baseline** design is to use a hand-held controller of the HTC Vive Virtual Reality (VR) system to control the 6-DOF pose of the end-effector of the teleoperated manipulator robot. To concisely describe these interface designs, we will use *ND* for non-dominant hand, *D* for dominant hand, *F* for the control using free-form motion tracking, *T* for the control using trackpad, *P* for position control; *O* for orientation control. The interfaces used in the final user study in addition to the baseline interface are in bold.

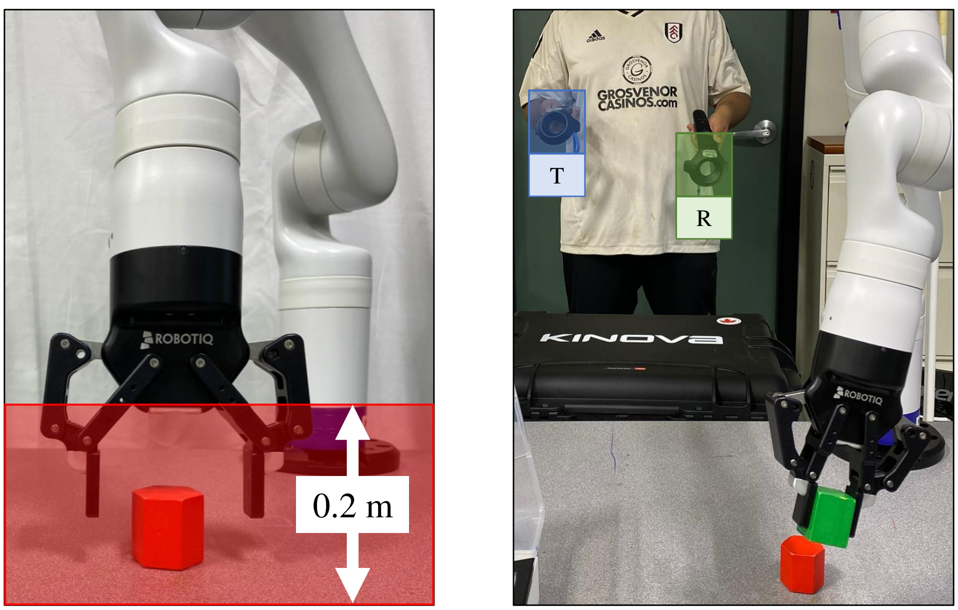


Figure 3.2: The $Scale_E$ interface (Left). The red region represents the distance between the robot end-effector and the table surface where the motion scaling is triggered. The P(D,F) + O(ND,F) interface (Right) uses two controllers, labelled “T” and “R” for the position and orientation control respectively.

The primitive interface designs for the separation of DOFs include:

- **P(D,F) + O(ND,F)**, where two hand-held controllers separate the free-form motion tracking control of the 3-DOF position and 3-DOF orientation of the robot end-effector(refer Figure 3.2);

- **P(D,F) + O(D,T)**, which uses the controller held in the dominant hand to control the 3-DOF position using free-form motion tracking and control the 3-DOF orientation using the trackpad.

- Mode switch, which uses one hand-held controller held in the dominant hand and allows switching between the three modes of Baseline or P(D,T) or O(D,T);

Meanwhile, the primitive interface designs for the motion scaling include:

- **Environment-based Scaling (**$Scale_E$**)** increases the scaling-down ratio of the robot end-effector motion as it approaches environment constraints or targeted objects. $Scale_E$ changes the scaling ratio between the operator's control speed and the controlled robot speed from 1:1 to 5:1 when the robot arm is close enough to (20 cm above; refer Figure 3.2) the table surface;

- **Operation-based Scaling (**$Scale_O$**)** increases the scaling-down ratio of the robot end-effector motion as the operator's control slows down. $Scale_O$ changes the controller-robot scaling ratio from 1:1 to 5:1 when the operator's control speed drops below 0.05 m/s. This method considers the operator's slow controller motion as the intent to perform precise manipulation. After some users couldn't perceive the scaling-ratio of 3:1, the final ratio 5:1 for a more noticeable scaling of motion for both $Scale_O$ and $Scale_E$ was finalized;

- Handle-Bar Scaling ($Scale_H$) defines a virtual handle bar by tracking the 3D position of two hand-held controllers, and allows the operator to adjust the scaling ratio by varying the distance between the two hand positions [122]. The $Scale_H$ interface changes the controller to robot scaling ratio from 1:1 to 3:1 when the distance between controllers goes below 10 cm. Thus, the position of robot end-effector can be controlled by the translation of the handle bar between two controllers, while the orientation can be adjusted by rotating the controller held in the dominant hand.

3.3.2 Interactive Design Process and Findings

For this pilot study process, 4 participants (3 males and 1 female) were recruited to continuously rest and improve the primitive interface designs. The participants were asked to teleoperate a 7-DOF manipulator robot (Kinova Gen3) with a two-fingered Robotiq gripper (2F-85) using the hand-held controller of a HTC Vive Virtual Reality (VR) system. As shown in Figure 3.3, the telemanipulation task is to pick up small wood blocks on a remote counter workspace and to place them in a target location inside of a plastic bin. Each of the interfaces was repeated for three trials. The interviews with the pilot study users helped us to identify the effective design features and factors that influence their decision. From the interactive design process, we find out that:

- **Separation of position and orientation control** is generally preferred, yet it is unclear whether the position and orientation control should use the same or separate controllers. Some users preferred the design of P(D,F)+O(ND,F) because the orientation control via motion tracking feels more natural, while others prefer to have the trackpad to control orientation because they can precisely adjust the object rotation in each DOF. Additionally, some users mentioned the ability to control one robot arm with one controller as preferable.
- Users also reported that they preferred performing **simultaneous rotation and translation** on P(D,F)+O(ND,F) compared to P(D,F) + O(D,T) or P(D,T) + O(D,F). They felt that being able to control both rotation and translation through motion mapping was intuitive and helped them do simultaneous rotation and translation.
- Users did not prefer **position control with the trackpad** as in the case of P(D,T) + O(D,F)

because it is slow and tedious, which makes it difficult to control large motions across the workspace.

- **Mode switching** should be avoided because users felt it tedious to switch between different control modes. They may even sacrifice performance to avoid switching modes as they felt it interrupted fluent operation.

- Regarding the design of **motion scaling**, scaling using handlebar ($Scale_H$) was difficult as it was hard to maintain the distance between two hand-held controllers required to maintain a steady scaling ratio. Pilot users generally agreed that the environment-based scaling, $Scale_E$, and the operation-based scaling, $Scale_O$, are more useful, but it is unclear which one is preferred. They also suggested that although the dynamic scaling is more convenient than manual adjustment of scaling ratio (e.g.,$Scale_H$), it is also important that they keep the scaling ratio to be steady for certain operations (e.g., steady low controller-robot scaling-down ratio for moving towards objects) or regions of the workspace (e.g., near the table constraints).

The feedback from the pilot study informs our final interface designs to be compared in our controlled user study.

3.4 Experiment

In the following user study (N=8) participants were recruited to evaluate the usability of the effective interface designs in the interactive design process. Specifically, we compared the users' performance and workload in dexterous manipulation using different DOF separation and motion scaling designs. We also examined the preferred combination of interface design features.

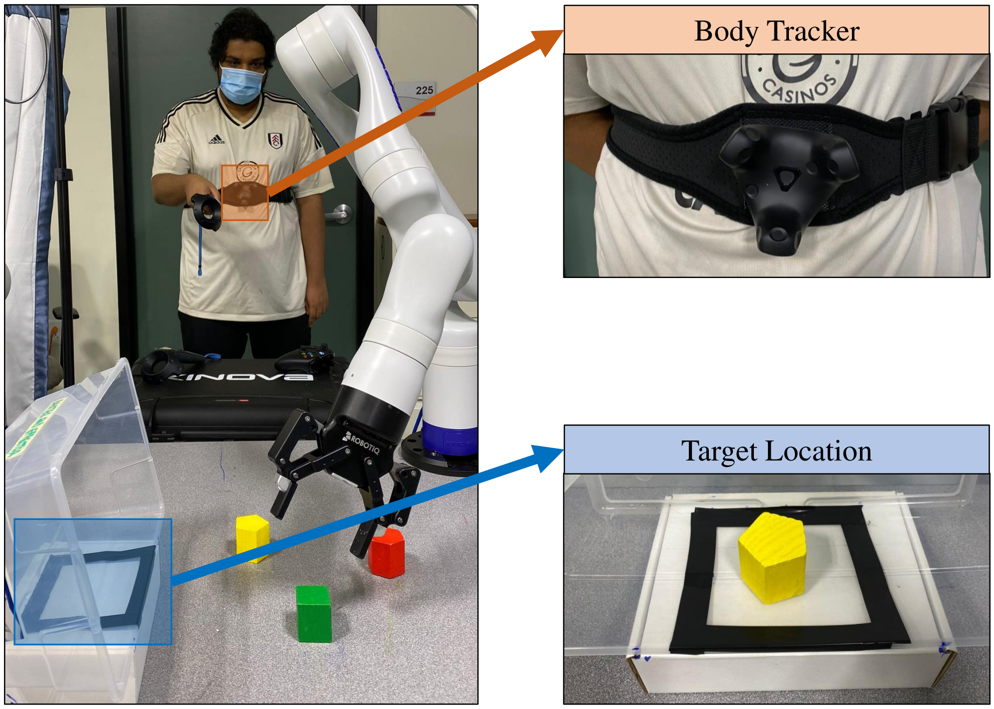


Figure 3.3: The task is to pick up the three small wooden blocks on the table and place them in the highlighted target location in a plastic bin. A Vive tracker is attached to the participant's torso to track the body motion.

3.4.1 Participants and Tasks

Our user study recruited (N=8) participants (7 male and 1 females, age = 26.4 $\pm$ 4.2 years) including student and general populations. Although these participants are not healthcare workers themselves, they are within the typical age range of nursing and medical students—future users of healthcare technologies. Their perspectives are valuable for evaluating usability and interface design, particularly for early-stage development where user experience and intuitive interaction are critical. The experiment protocol was approved by WPI's Institutional Review Board. The participants recruited for this final user study are different from the pilot user study. The experimental task is the same task as described in Section III-B and Figure 3.3. This task consists of general-purpose manipulation actions across the workspace (e.g., reaching, moving) as well as the precise grasping and placing of small objects. The plastic bin is placed sideways to force the operator to carefully maneuver the orientation of the robot end-effector which requires precise dexterous manipulation and orientation control.

3.4.2 Experimental Procedure

This experiment consists of six sections. Each section has a *training phase* to familiarize the participants with the interface design selected for that section, and a *task-performing phase* in which the participants officially perform manipulation tasks. The participants practiced and performed the tasks using the baseline interface design in Section 1, and using the four modes of interface designs (described in the Experiment Conditions) in Sections 2-5. The interface modes for Section 2-5 were randomized to avoid learning effects. In the last section, the participants combined their preferred mode for DOF separation and motion scaling, and used this combined interface. Total trials for each subject = 6 interfaces $\times$ 3 trials.

Experiment Conditions: Through the interactive design process P(D,F) + O(D,T) and P(D,F) + O(ND,F) for DOF separation, and the $Scale_E$ and the $Scale_O$ interfaces for motion scaling were selected for evaluation.

3.4.3 Data Collection and Analysis

During the experiment, we collected data to analyze the performance and workload to compare the interface design. For task performance, we measured the task completion time, and the occurrence and types of errors. The errors include: 1) collisions with the table and the object, and 2) misplacement of the object outside of the designated region. We also estimated the physical workload from the teleoperator's torso movement tracked by a Vive tracker attached to the operator's chest. Large torso motions indicate that the operator had to move their body to compensate for the limited range of arm motion due toto interface design limitations. In order to evaluate the user perceived workload the NASA-TLX questionnaire was used while a System Usability Scale (SUS) survey was used to evaluate the participants' rating of the interfaces' usability.

The Wilcox rank-sum test was used to identify significant differences between the different NASA-TLX and SUS responses while one-way ANOVA analysis was used to identify significant differences between the interfaces for average task completion times and body motion.

3.5 Impacts of Action Assistance

In the following sections, the "combination" mode refers to the mode where participants were allowed to use their preferred designs for scaling and for separation of DOFs. In the figures presenting the results, we address the interfaces as follows: A = Baseline Design; B = P(D,F) + O(ND,F); C = P(D,F) + O(D,T); D = $Scale_E$; E = $Scale_O$; F = Combination of preferred mode of DOF separation and motion scaling. Additionally, the bars with asterisk on top in Figure 3.4 and Figure 3.6 represents a significant difference between the operation modes at the line's starting and ending points.

3.5.1 Task Performance

Figure 3.4 shows the task completion time averaged across the 8 participants for each interface design. We found that based on the one-way ANOVA analysis using an additional controller for orientation control , i.e., the P(D,F) + O(ND,F) mode, significantly reduced the task completion time, compared to the baseline interface ($F(1,46)=11.63$, $p<0.05$), the interface using trackpad for orientation control, i.e., P(D,F) + O(D,T) ($F(1,46)=5.3$, $p<0.05$) and $Scale_O$ ($F(1,46)=6.3$, $p<0.05$). Regardless of what their preference was, the combination mode significantly reduced the task completion time compared to the baseline ($F(1,46)=13.06$, $p<0.05$), P(D,F) + O(D,T) ($F(1,46)=6.34$,

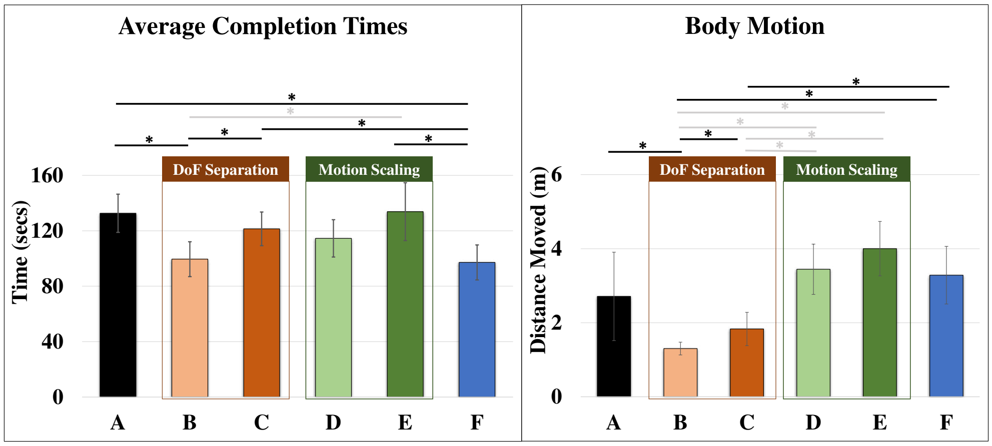


Figure 3.4: Comparison of the average task completion time (Left) and the induced body motions (Right) among six interface modes.

p<0.05) and the $Scale_O$ interfaces (F(1,46)=7.17, p<0.05) as well. Neither $Scale_E$ or $Scale_O$ significantly reduced the task completion time compared to the baseline mode.

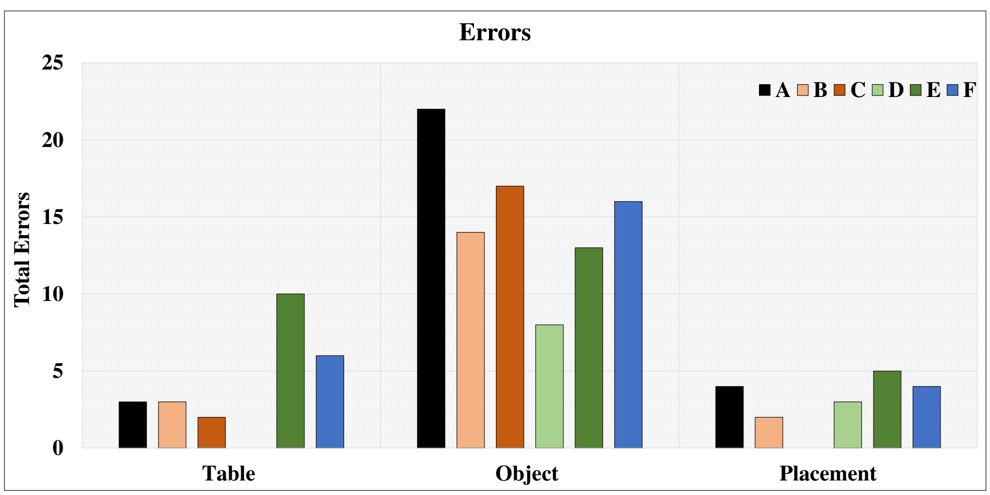


Figure 3.5: The total number of table collisions (Left), total number of collisions with objects (Middle) and total number of placement errors (Right) across all the trials and participants.

In Figure 3.5, we compared the total occurrence of each type of error for all the participants, because the mean and variance of each type of error was very small. The three types of errors we considered include collisions with the table, the objects, and misplacement of the object in the bin. It is shown that $Scale_O$ caused more collisions with the table compared to all the alternative designs with an average of 1.25 ± 0.33 times per trial, while $Scale_E$ had 0 instances of collision with the table across all the participants. $Scale_E$ also had the least number of collisions with the object with an average of 1 ± 0.61 times per trial, while the baseline had the most with an average of 2.75 ± 0.78 times per trial. The placement errors for all the interface modes were comparable. The combination of preferred interface designs were neither the most nor the least number of errors, for all types of errors.

We also analyzed the amount of body motion incurred when using each interface mode. The motion of the body tracker as attached to the operator in Figure 3.3 was used to monitor the operator's torso movements while teleoperating. More body motion during teleoperation implies a worse interface design, because the arm and hand motions for robot control have limited ranges and have to be compensated by (unconscious) movements of the body. As shown in Figure 3.4, the $P(D,F) + O(ND,F)$ mode, caused significantly less body motions compared to the baseline (F(1,46)=7.31, p<0.05), $P(D,F) + O(D,T)$ (F(1,46)=4.92, p<0.05), $Scale_E$ (F(1,46)=26.92, p<0.05), $Scale_O$ (F(1,46)=40.02, p<0.05) and combination modes (F(1,46)=30.96, p<0.05). The $P(D,F) + O(D,T)$ mode also caused significantly less body motions compared to the $Scale_E$ (F(1,46)=12.15, p<0.05), $Scale_O$ (F(1,46)=20.87, p<0.05) and combination modes (F(1,46)=12.1, p<0.05). Overall, the separation of DOFs were observed to significantly reduce body motion. However, although scaling down the robot control motions may limit the operator's workspace, we did not find a significant difference when compared with the baseline interface. The combination mode causes body motion comparable to the $Scale_E$ and $Scale_O$ modes, which implies motion scaling in general may induce more body motions in precise tele-manipulation.

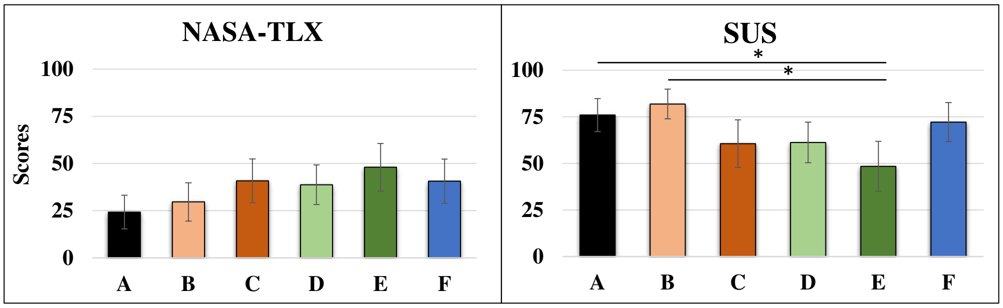


Figure 3.6: Comparison of the average weighted NASA-TLX scores (Left) and SUS scores (Right) among six interface modes.

3.5.2 Survey Feedback

Figure 3.6 shows the feedback for our NASA-TLX and SUS surveys. We first compared the weighted NASA-TLX scores to measure the subjective workload across interface modes. The weighting coefficients were selected according to the NASA-TLX guideline questions and are as follows: mental demand=5, physical demand=1, temporal demand=0, performance=4, effort=3, frustration=2. We found that none of the 6 interfaces had a significantly different score when compared against each other, indicating that all the interfaces have comparable perceived workload. The baseline interface scored the lowest perceived workload when averaged across the 8 participants with a total of 24.25 ± 8.9 while the $Scale_O$ interface scored the highest with a total of 48 ± 12.58. From the SUS survey, we found that the overall usability of the $Scale_O$ is significantly lower (p<0.05) than the baseline mode and the two controller DOF separation mode, i.e., the P(D,F) + O(ND,F). A more detailed usability analysis indicates that the $Scale_O$ mode is significantly worse (p$<$ 0.05) than the baseline and P(D,F) + O(ND,F) modes because of unnecessary complexity and significantly worse than the P(D,F) + O(ND,F) mode due to the inconsistencies in the interface. It is also significantly worse (p$<$ 0.05) than the P(D,F) + O(ND,F) and the combined preferred modes because the interface is perceived to be not as well-integrated.

Figure 5.7 shows the combination of preferred interface design features for each user, and

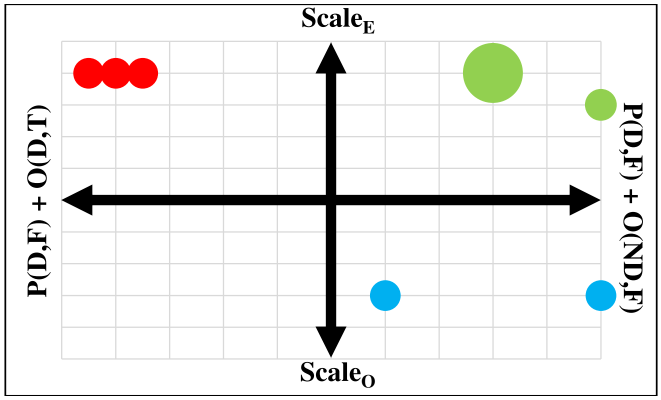


Figure 3.7: The preferred choices of DOF separation and motion scaling modes. The one larger dot represents two users with same preferences.

the extent of their preference. The preferred combinations of interface designs for scaling and DOF separation fell into three types, with a ratio of 3:3:2. Most participants (6 out of 8) preferred environment-based scaling ($Scale_E$) over the operation-based scaling ($Scale_O$), while 5 out of 8 participants preferred the P(D,F) + O(ND,F) interface when it came to DOF separation. No participant preferred the combination of operation-based scaling and the mode using trackpad for orientation control.

3.6 Summary and Outlook

3.6.1 Degree of Freedom Separation for Motion Mapping

Regarding Research Question **RQ1**, the separation of position and orientation control to two controllers is preferred over the alternative method. Overall, DOFs separation can improve control precision of a teleoperation interface because it enables the operator to precisely control which DOFs to be involved, and avoid the interference of unintended motions. The separation of position and orientation control to two controllers performs better because it balances the precision and efficiency in motion control. For the pick-and-place task in our user study, people do not need to precisely control the motion in each DOF, and controlling the 3-DOF for position or orientation leads to more efficient object manipulation. Using this interface design, the task completion time was significantly reduced by 18% compared to the alternative DOF separation method and by 25% compared to the baseline interface. However, being able to control individual DOF may be necessary for the precise control of translation (e.g., object inserting, stacking), or precise adjustment of orientation (e.g., adjusting the pan or tilt of an active telepresence camera). We also noticed that both the two DOF separation methods induced significantly less body motions compared to the two motion scaling methods and the combined method. Particularly, the body motions observed in the usage of the two-controller interface, P(D,F) + O(ND,F), is 29% less compared to mode using trackpad for orientation control, and 58% less compared to the baseline. Note that the induced body motion is not preferred because it increases a teleoperator's physical fatigue and frustration. From the survey feedback, we found that separating the position and orientation motion control did not significantly increase the perceived mental workload, and P(D,F) + O(ND,F) had the best usability score. Because using trackpad for orientation control is not as intuitive, 5 out of 8 participants preferred the two controller interface over the alternative for DOF separation.

3.6.2 Motion Scaling for Motion Mapping

Regarding the research question **RQ2**, we found that the environment-based scaling performed better and was preferred for our user study task. This is because the environment-based scaling improves the motion control precision in the neighborhood of environment constraints, or the local region where the precise manipulation is performed. In general, environment-based motion scaling may also benefit the dexterous manipulation or navigation in a cluttered environment, similar to methods for collision avoidance. Shown in Figure 3.5, the environment-based scaling, i.e., the $Scale_E$, has caused fewer errors than the operation-based scaling method ($Scale_O$) which adjusts the scaling ratio according to the operation speed. The environment-based scaling is particularly effective for reducing the collision with object when the participant attempted or failed to grasp. It also reduced the task completion time by 14% and 13% compared to $Scale_O$ and the baseline, respectively. We noticed that compared to baseline interface, $Scale_E$ has comparable perceived workload and usability score, while those of the $Scale_O$ are significantly worse. Our survey shows that 6 out of 8 participants preferring the environment-based scaling over the alternative method. The post-study interview shows that when using the environment-based scaling, participants can better predict when the scaling ratio will change, while the operation-based scaling, which is much less predictable and controllable, usually lead to unintended change of scaling ratio.

3.6.3 Combining Two Types of Action Assistance

We also found that the combination of a user's preferred choices for DOF separation and motion scaling may lead to a balanced performance and workload. The combination of preferred methods, regardless of individual difference in the choices has comparable task completion time compared to the most preferred DOF separation method, i.e, P(D,F) + O(ND,F), even if for some part of the manipulation the controlled motion became very slow due to the motion scaling. The combined method also induced less body motions compared to the two scaling methods. The survey feedback shows that the usability score of the combined method is comparable to the baseline interface and higher than the DOF separation method using trackpad. For the task in this user study, a balanced performance and acceptable workload may contribute to improved user's acceptance, yet the choices of preference may be different for individuals and other tasks and types of precise manipulation.

Chapter 4

Adaptive Action Assistance for Bi-manual Remote Manipulation

4.1 Motivation

Teleoperated robotic systems offer a powerful solution for enabling remote expert interaction across a wide range of applications — from healthcare delivery to operations in hazardous environments [127, 128]. In the context of healthcare, such systems can bridge geographical barriers, allowing clinicians to treat patients in remote or underserved locations [129]. They can also protect healthcare workers from exposure in high-risk scenarios, such as quarantine wards, by enabling them to provide care while remaining safely outside the affected area [130]. To maximize usability and facilitate effective translation of expert motor skills to robotic actions, motion capture-based free-form teleoperation interfaces have been preferred [46]. These interfaces provide intuitive and efficient control by directly mapping human motions to corresponding robot movements, supporting complex, multi-axis robot arm motions required in bi-manual remote manipulation tasks [84].

While intuitive, free-form interfaces can also be physically demanding for users, especially when controlling multiple degrees of freedom over extended periods. These physical demands are further exacerbated in bi-manual remote manipulation, where both arms are used in coordinated tasks such as opening containers, pouring, or dual-arm object manipulation. Additionally, teleoperation systems often suffer from limited feedback, such as the absence of haptic information or

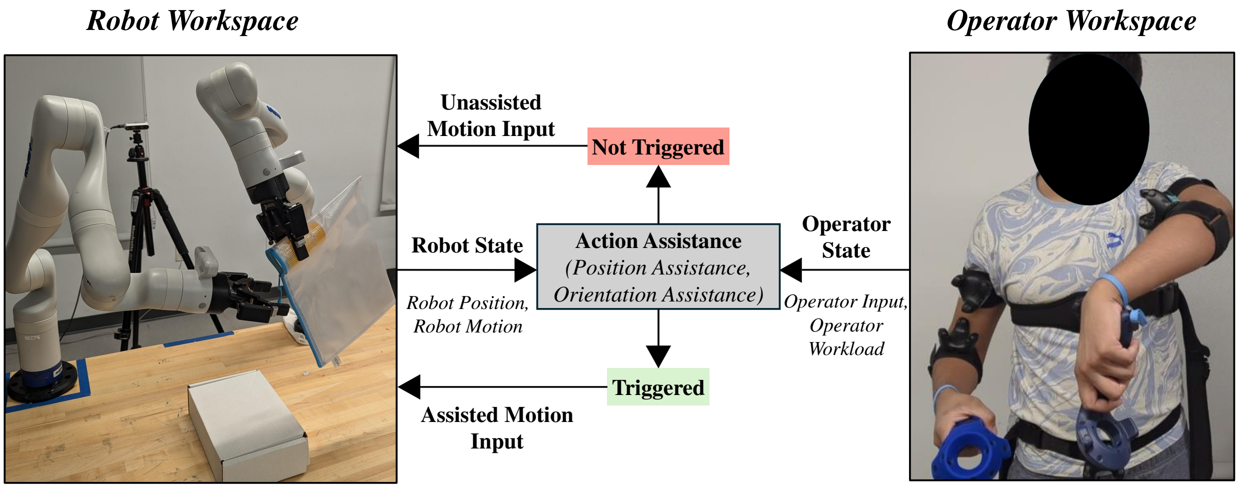


Figure 4.1: Workflow during action-assisted bi-manual remote manipulation.

incomplete visual perception of the robot workspace, making it difficult for users to execute precise actions [131]. To address this, various forms of action support or assistance have been proposed, aimed at compensating for perception gaps or guiding motion [25].

Motion tracking systems can introduce noise due to unintentional user motions or reduced precision, making it difficult to carry out high-precision tasks [31]. Techniques such as position support in the form of motion scaling or orientation support by eliminating unintended motions during robot rotation have shown promise in improving operational accuracy. However, these support mechanisms, if applied indiscriminately, may interfere with user intention or cause frustration, especially when they override desired control inputs or increase required operational efforts [132, 133]. Thus there is a need to develop assist-as-needed control interfaces that activate support only when appropriate—leveraging the benefits of assistance while minimizing its drawbacks.

Although prior research has proposed various action assistance techniques for bi-manual remote manipulation, there has been no systematic evaluation of how these designs influence operator performance, workload, or subjective preferences. Moreover, there is limited exploration into incorporating real-time operator state—such as physical activity levels or inferred control intent—

to govern when and how assistance should be delivered. This paper aims to address these gaps by conducting a comprehensive evaluation of multiple interface designs, each incorporating different forms of assistance. These designs are tested across two bi-manual remote manipulation tasks that involve object handling, information gathering, and high-precision actions — core components of many teleoperated procedures. The main contribution of this work is to identify design principles that enhance the effectiveness of bi-manual remote manipulation assistance for the proposed action assistance while avoiding unnecessary or intrusive activation.

4.2 Literature Review

Action Assistances for Bi-manual Remote Manipulation: Several remote manipulation interfaces have been designed using motion capture technologies, including RGB-D cameras [134], exoskeletons [135], infrared sensor-based systems [24], and Virtual Reality (VR) setups [136]. Our prior research has shown that while motion-mapped remote manipulation interfaces support intuitive and efficient teleoperation [25], they continue to face key limitations:

- **Reduced precision** in fine manipulation tasks, due to inherent limitations in motion tracking fidelity and the natural variability in human motor inputs.
- **Unintended motion coupling** across multiple degrees of freedom (DoFs), where simultaneous multi-axis control can result in interference between intended and unintended motions.

To address reduced precision, a commonly adopted solution is ***position support*** through motion scaling, wherein user input is scaled down to allow for finer control of robotic movement. Multiple interface designs have explored motion scaling strategies for both single-arm and bi-manual teleoperation, often employing task-based or robot state-based scaling schemes. For instance, fixed-ratio scaling methods—commonly used in robotic telesurgery [137]—apply region-based or task-state-based scaling with limited levels of variability during operation. These have also been successfully extended to general bi-manual manipulation scenarios [32, 138, 139]. Other approaches employ dynamic scaling, which continuously varies based on contextual factors such as operator speed (e.g., PRISM scaling [31]) or proximity to task objectives (e.g., HOMER scaling [140]).

Another major challenge in bi-manual remote manipulation is managing interference caused by unintended DoF activations. For example, an operator attempting to rotate an object may unintentionally produce linear motion due to the coupled nature of human motion control [30]. To reduce such interference, several ***orientation support*** techniques have been proposed. These include decoupling position and orientation control or constraining specific DoFs to isolate only the intended motions. Such support has been realized through control interfaces that allow discrete control (e.g., trackpads or joysticks [141, 142]), virtual fixtures [143], motion intent inference [144], and control interface designs using multiple controllers [145].

Overall, prior work suggests that incorporating such action assistance mechanisms can improve operator performance, reduce workload, and enhance subjective preference. However, while these techniques have been well studied for virtual object manipulation [32] or single-arm remote control [25], there remains a significant gap in effectively integrating and evaluating such support in bi-manual teleoperation systems and identifying user preferred position and orientation support for bi-manual tasks remains an open challenge.

Design of Assist-as-Needed Interfaces: A critical consideration in the design of action assistance interfaces is determining when such support should be provided. Assistance that is triggered inappropriately—when the user does not require it—can lead to unnecessary frustration, added cognitive and physical workload, and unintended disruptions in task execution [146]. Thus, the notion of assist-as-needed control can provide desirable behavior, where assistance is applied selectively, rather than continuously or indiscriminately.

To effectively implement assist-as-needed paradigms, it is essential to identify suitable triggers for support activation. These triggers can be derived from three main sources:

- Workspace-based triggers: These triggers rely on characteristics such as proximity to task-critical elements [147] or environmental constraints like collisions [25].
- Robot-based triggers: These triggers include factors such as avoiding kinematic instability [28], operational latency [148] or self-collision [149].
- Operator-based triggers: These triggers are typically based on real-time estimates of physical or cognitive workload [150], or rely on more explicit operator-initiated requests for assistance, where the user actively activates or deactivates support as needed [151].

While each strategy has demonstrated utility in specific contexts, their effectiveness can vary based on the teleoperation scenario.

The challenge, therefore, lies in selecting the trigger mechanism that best enhances task performance without compromising user control or introducing confusion. Ideally, an assist-as-needed interface should balance the benefits of support with user preference, offering help only when it aligns with user intent or operational demand. This requires robust detection of assistance needs so that users remain aware of when and why assistance is being activated.

4.3 Realtime Physical Workload Estimation Based on Human Motion

The physical workload was estimated using a methodology developed in our prior work [150, 152], where arm and torso configurations are mapped to a muscle effort index (scaled 1–100). This mapping was built from EMG data collected from one male and one female participant performing a

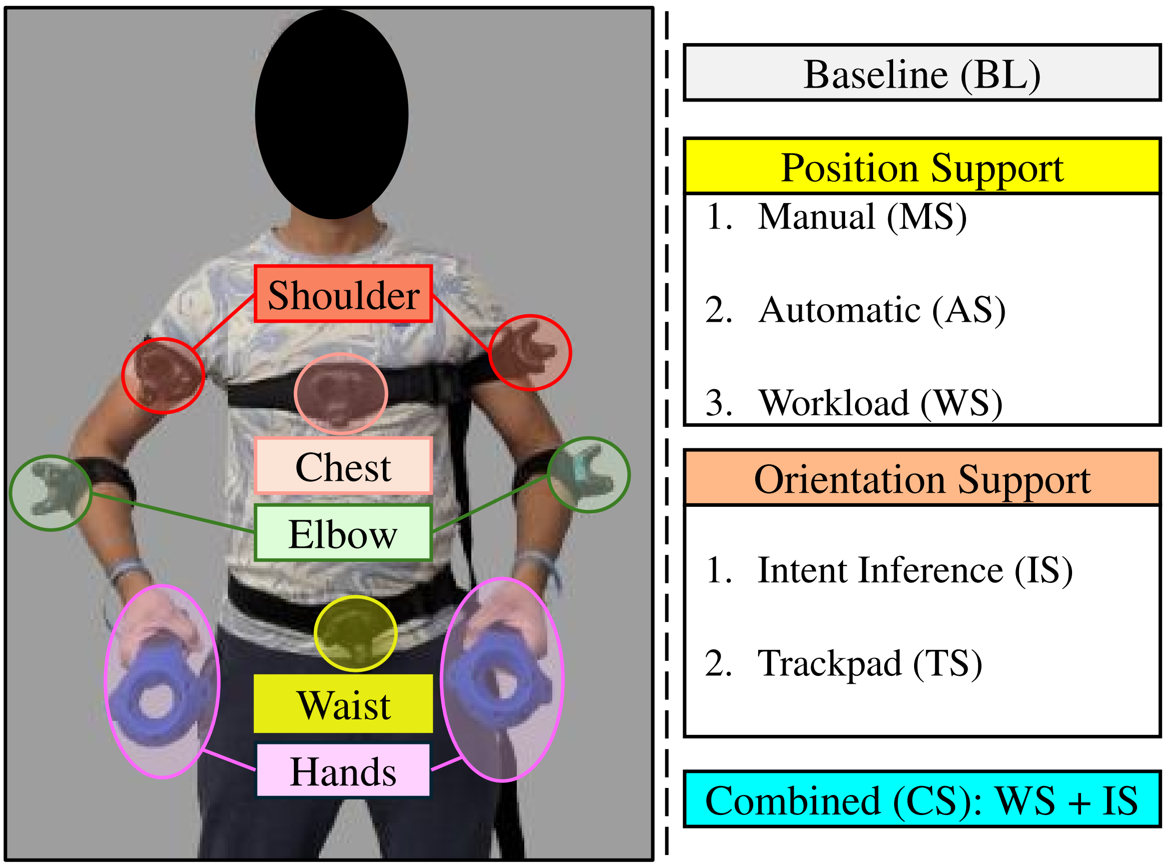


Figure 4.2: **Left**: Arrangement of VR trackers and controllers used for estimating physical workload; **Right**: Overview of the seven interfaces evaluated in the user study: the Baseline interface, three position support interfaces, two orientation support interfaces, and a combined interface using workload-based scaling and intent inference.

range of upper-limb motions. The index reflects the estimated muscular demand of various configurations based on joint angles and muscle group activations. Shoulder effort is computed as a weighted sum of the anterior and lateral deltoid activations, using a 3:4 ratio to reflect their relative force-generation capacities [153], while elbow effort is inferred primarily from biceps flexion. The resulting effort index, scaled from 1 to 100, reflects the physical strain associated with a specific arm posture. To account for the effect of prolonged static postures on fatigue, we incorporate the dwell time spent in a given configuration. The overall physical workload is computed as $W_{phy} = E \times T_{dwell}$, where W_{phy} denotes the physical workload corresponding to the arm configuration, E is the configuration specific effort index and T_{dwell} refers to the time (in seconds) the operator dwells in a particular configuration. Dwell time accumulates when the operator's motion falls below 0.125 m/s (linear) or 0.4 rad/s (angular) and resets when movement exceeds these thresholds. These velocity thresholds were determined through pilot trials with five experienced participants (four male, one female), each with over 20 hours of teleoperation experience. The selected thresholds reflect the participants' feedback on the motion speeds at which they perceived themselves to be holding stationary and effortful poses. The value of W_{phy} is maxed out at 100. To reflect the higher fatigue associated with static arm postures, workload is designed to increase proportionally with dwell time [84, 154], and W_{phy} is capped at a maximum value of 100. During the experiment, we captured operator torso and arm motion using six VR trackers and two handheld controllers, as shown in Figure 4.2 (Left).

4.4 Interface Designs for Bi-manual Remote Manipulation Tasks

The **Baseline (BL)** control interface is the default design where the motion of the handheld controllers is mapped to the robot arm in a 1:1 ratio, controlling all six Degrees-of-Freedom (DoF) simultaneously.

Position Support: Refers to control interfaces that provide the operator support with positioning and scaling down motion as required. In these designs the operator to robot motion scaling ratio varies as 1:1 when the robot end-effector is greater than 0.4 m from the table or the other arm, 1:0.75 when the robot arm is between 0.2 m and 0.4 m from the table or the other arm and 1:0.5 when the robot arm is less than 0.2 m from the table or the other arm.

- **Manual Support (MS):** This interface uses the baseline design with the scaling ratios described above. The motion scaling is manually triggered by the operator by pressing a button on the handheld controller.
- **Auto Support (AS):** In this interface, position support is automatically activated based on the distance between the robot end-effector and the table or another arm, applying the appropriate motion scaling ratio.
- **Workload Support (WS):** Position support is triggered when the physical workload estimated on the operator is reached, using the methodology described in Section 4.3.

Orientation Support: Refers to support interfaces that assist the operator with controlling rotational motion, ensuring that only the desired rotation occurs without interference from other rotations.

- **Intent Inference Support (IS):** This interface detects the operator's intended direction of rotation and filters out other directions. It only allows rotations along directions where the angular velocity exceeds 0.65 $radians/second$.
- **Trackpad Support (TS):** When activated by pressing a button on the controller, the operator can use the trackpad to control rotations along the x and y-axes more discreetly.

The **Combined Support (CS)** is a combination of Workload Support and Intent Inference Sup-

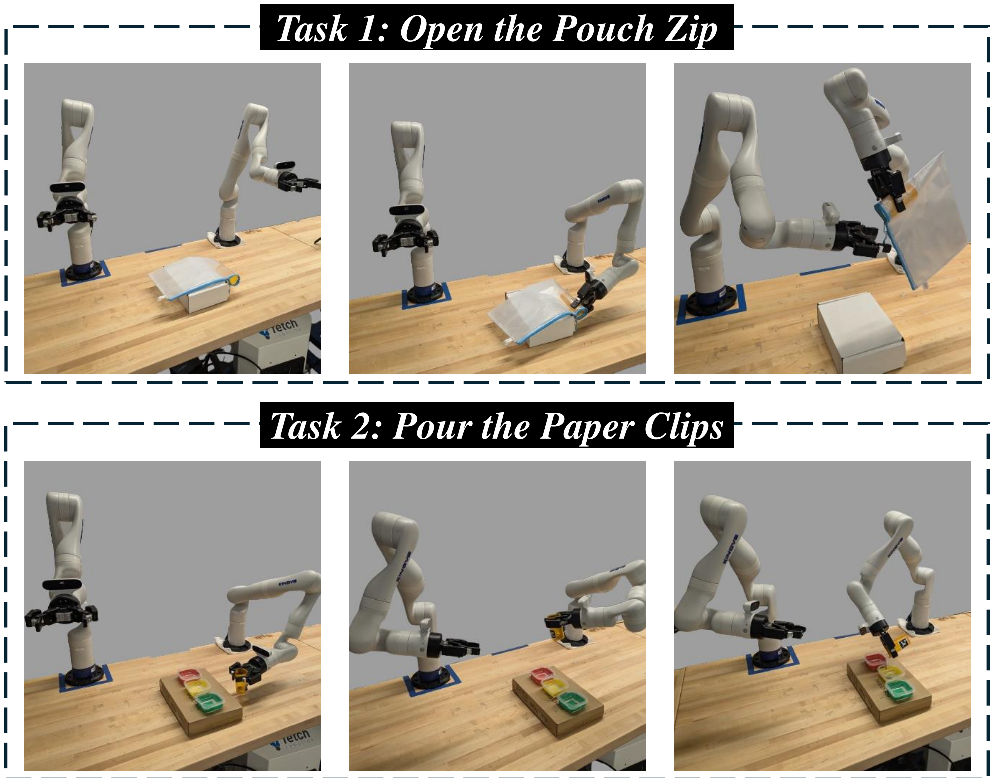


Figure 4.3: **Top**: The participants pick up the pouch and open the zip; **Bottom**: Participants pick up the container with paper clips, scan the label on the container and pour paper clips in the target box.

port. This combination was chosen based on pilot trials and was found to be the preferred combination for both position and orientation support.

During operation, if the operator wants to pause an unused robot arm or reconfigure their own arm's position, they can do so by pressing the engage/disengage button on the controller that corresponds to the arm they wish to control.

4.5 Experiment

Participants and Tasks: We conducted user studies (N = 16 participants, 13 male, 3 female) to investigate the effectiveness of the the implemented assistance designs for bi-manual tele-manipulation.

Similar to the previous chapter, these participants fall within the typical age range of nursing students, and it is assumed that their perspectives can closely resemble those of users with comparable demographic profiles but more healthcare experience. The evaluation focused on performance (task completion times), workload (physical workload) and subjective preferences. The experimental protocol was approved by WPI's Institutional Review Board (IRB-21-0004). As shown in Figure 4.3), participants performed two object manipulation tasks involving the control of two 7-Degree-of-Freedom (DoF) robot arms (Kinova Gen3):

- **Task 1:** Participants used the right robotic arm to pick up a transparent pouch placed on an elevated platform (refer Figure 4.3 (a)) and opened the pouch's zip using the left robotic arm. This task was divided into the following sub-actions: 1) picking up the pouch from the table and 2) opening the pouch's zip.
- **Task 2:** Participants used the right robotic arm to pick up a container holding 20 paper clips, then scanned a label on the container using a camera mounted on the left robotic arm. After scanning, participants poured the paper clips into one of the color-coded containers in the workspace (refer Figure 4.3 (b)) corresponding to the color of the label observed during scanning. This task comprised the following sub-actions: 1) picking the container; 2) scanning the label and 3) pouring the paper clips in the box.

To ensure uniformity across participants and trials, the positions of the color-coded containers and the observed label color remained consistent throughout the experiment.

While performing the tasks, participants viewed the workspace through a visual interface that streamed the workspace camera view, secondary camera view, and a status bar. The status bar displays the state of position and orientation support and pause/unpause status of each robot arm (refer Figure 4.4). The secondary camera view switched between the side camera view and the robot arm camera view based on the state of the task. All camera streams are captured using Intel

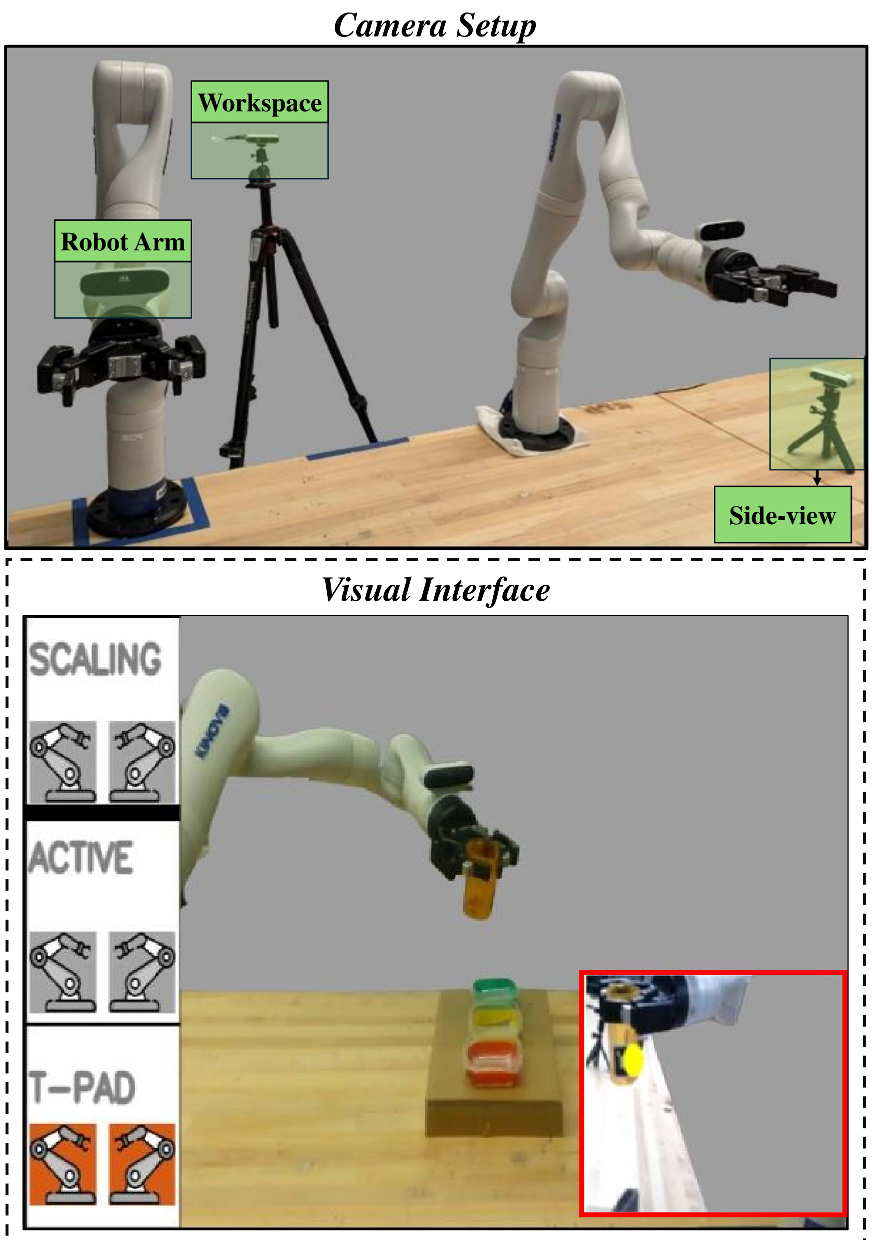


Figure 4.4: **Top**: The camera set-up highlighting the locations of the workspace camera, side-view camera and left robot arm camera; **Bottom**: The visual interface with the status bar on the left, the workspace camera view in the middle and secondary camera view in the bottom right. The secondary view above displays the robot arm camera view scanning the label on the container.

Realsense d435i cameras.

Experimental Procedure: The experiment required participants to perform the tasks using the seven interfaces described in Section 4.4. Before starting the user study, participants completed a training phase to familiarize themselves with the interfaces. The training phase consisted of two tasks: 1) Picking and placing a container in a box; 2) Searching for and scanning a label on a container using the robot arm camera. These training tasks introduced participants to the position and orientation support mechanisms described in Section 4.4 and ensured they became comfortable with bi-manual tele-manipulation. Participants proceeded to the *task-performing phase* only after completing the training tasks without errors and reporting subjective comfort with all the interfaces. In the task-performing phase the participants perform the two user study tasks with each of the seven interfaces and their performances and subjective preferences are recorded. Each participant completed a total of: 2 tasks $\times$ 7 interfaces $\times$ 2 trials $=$ 28 trials To mitigate learning effects, the order of interface usage was randomized across participants during the user study.

Evaluation Metrics: During the experiment, we collected data to analyze performance and workload, enabling a comparison of the different interface designs. Task performance was assessed using task completion time, which serves as an indicator of the ease of operation for performing the user study tasks. Physical workload was estimated using the method described in Section 4.3. The total physical workload for an entire task was calculated as a weighted average of the workload observed in the left and right arms, with a ratio of 7:3. This ratio reflects the greater motion requirements for the left arm during the tasks. For specific sub-tasks, the workload ratios were adjusted as follows:

- 9:1 (left:right): Applied to the pouch pickup sub-action in Task 1 and the container gathering and pouring sub-actions in Task 2, where extensive left-arm motion was required.

- 1:9 (left:right): Applied to the zip opening sub-action in Task 1 and the scanning sub-action in Task 2, where right-arm motion was predominant.

The NASA Task Load Index (NASA-TLX) questionnaire was used to capture subjective workload perceptions during the post-study survey. Participant preferences for interface usability were evaluated using the System Usability Scale (SUS) survey.

A one-way ANOVA followed by pairwise t-tests was used to identify significant differences between interfaces for average task completion times and physical workload. The Wilcoxon rank-sum test was employed to determine significant differences in the NASA-TLX and SUS survey responses.

Design Objectives: This experiment is designed to investigate the following Design Objectives (DOs) for bi-manual remote manipulation: **DO1**) How to design position **DO1** and orientation **DO2** support to improve operator performance and reduce workload while improving operator preference for bi-manual remote manipulation?

4.6 Results

Table 4.1: The Average NASA-TLX and SUS Scores of Baseline (BL), Auto (AS), Manual (MS), Workload (WS), Intent Inference (IS), Trackpad (TS) and Combined (CS) Support Interfaces

Interface	**BL**	**AS**	**MS**	**WS**	**IS**	**TS**	**CS**
NASA-TLX Scores	50.2 ± 9.6	48.4 ± 9.6	55.6 ± 9.8	44.7 ± 8.3	48.3 ± 10.4	52.6 ± 10.5	45.8 ± 10.2
System Usability Scores (SUS)	71.7 ± 8.4	73.7 ± 10.1	64.4 ± 10.8	75.1 ± 10.1	67.9 ± 13.1	60.4 ± 11.9	69.5 ± 12.7

In this section we present the results of our experiment and the impact of the various assistance designs on the operator performance (task completion time) and physical workload (refer Section 4.3). In the presented figures significant differences between the two data points is represented by a line with an asterisk above it.

4.6.1 Impact of Position Support for Bi-manual Tasks

Figure 4.5 represents the performance of the position support interfaces (Auto, Manual and Workload support) compared to the Baseline (BL) interface for both Tasks. The top half of Figure 4.5 represents the average physical workload for Tasks 1 and 2. The average physical workloads for the various interfaces for Task 1 are 32.3 $\pm$ 7.8, 31.9 $\pm$ 9.5, 37.2 $\pm$ 10, and 23.3 $\pm$ 6.7 for the Baseline (BL), Auto (AS), Manual (MS), and Workload (WS) support interfaces respectively. Similarly, for Task 2, the incurred physical workloads are 34.2 $\pm$ 10.1, 34.6 $\pm$ 9.8, 35.3 $\pm$ 9.7, and 28.3 $\pm$ 7.5 for the BL, AS, MS, and WS interfaces respectively. The bottom half of Figure 4.5 reports the task completion times for both tasks. For Task 1, the average task completion times (in seconds) are 103.3 $\pm$ 13.6, 121.1 $\pm$ 14, 116.8 $\pm$ 17.5, and 92 $\pm$ 10.5 for the Baseline, Auto, Manual, and Workload-based supports respectively. For Task 2, the average task completion times for the BL, AS, MS, and WS interfaces are 106.6 $\pm$ 17.6, 109.8 $\pm$ 14.6, 116.4 $\pm$ 21, and 99.3 $\pm$ 10.4 respectively. Post hoc comparisons show that the workload-based scalings yielded significantly lower ($p < 0.05$) physical workload and task completion times compared to all other interfaces for Task 1 (Opening the pouch zip). In terms of subjective perceived workload, the average weighted NASA-TLX scores (scale: 1–100) for the BL, AS, MS, and WS interfaces were 50.2 $\pm$ 9.6, 48.4 $\pm$ 9.6, 55.6 $\pm$ 9.8, and 44.7 $\pm$ 8.4 respectively. The average System Usability Scores (SUS) for the Baseline, Auto, Manual, and Workload scaling interfaces were 71.7 $\pm$ 8.4, 73.7 $\pm$ 10.1, 64.4 $\pm$ 10.8, and 75 $\pm$ 9.5 respectively.

4.6.2 Impact of Orientation Support for Bi-manual Tasks

The top half of Figure 4.6 reports the physical workload index for Tasks 1 and 2 of the Baseline interface along with the two orientation support (Intent Inference (IS) and Trackpad (TS)) interfaces. The average physical workload incurred by the operator while using the BL, IS, and TS

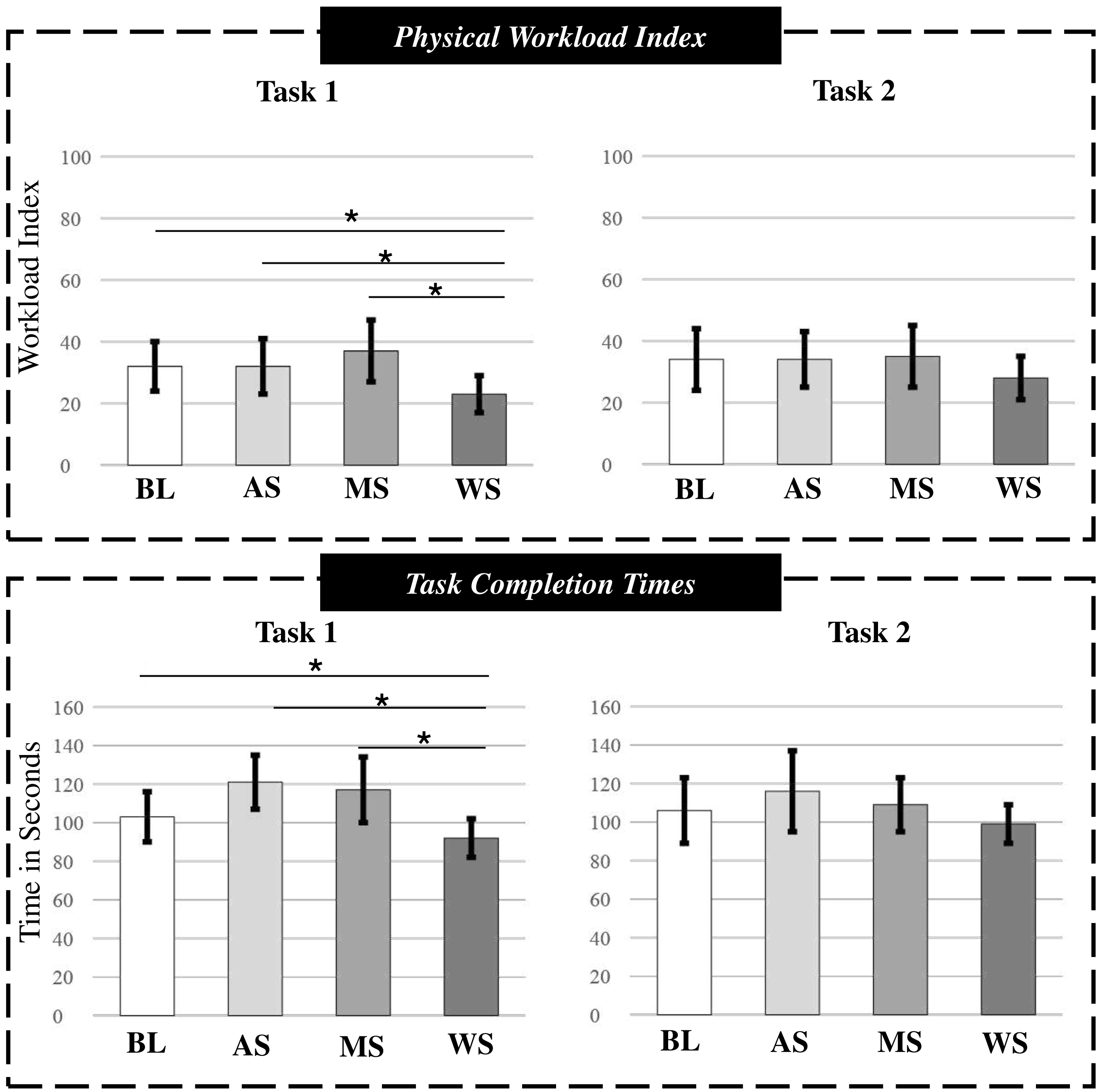


Figure 4.5: The baseline interface is represented as BL, while the position support interfaces are denoted as AS, MS, and WS for Auto, Manual, and Workload Scaling, respectively. **Top**: Comparison of physical workload indices for Tasks 1 and 2; **Bottom**: Task completion times for Tasks 1 and 2.

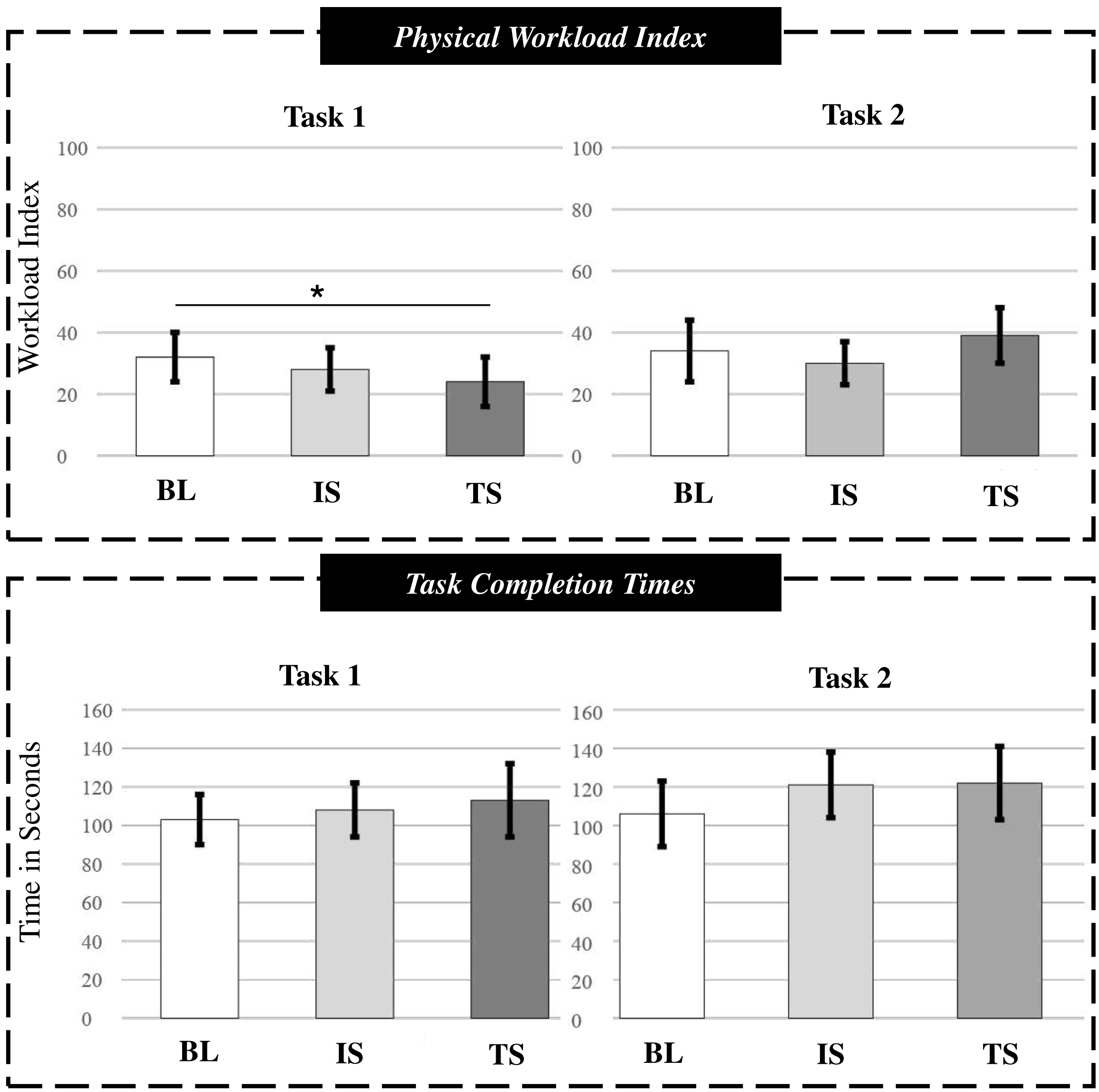


Figure 4.6: The baseline interface is represented as BL, while the orientation support interfaces are denoted as IS, and TS for Intent Inference, and Trackpad support, respectively. **Top**: Comparison of physical workload indices for Tasks 1 and 2; **Bottom**: Task completion times for Tasks 1 and 2.

interfaces were as follows: Task 1 (opening the pouch zip), the values were 32.3 ± 7.8, 28.4 ± 7.4, and 23.9 ± 7.6 respectively; for Task 2 (pouring the paper clips), the values were 34.3 ± 10.1, 30.1 ± 7.4, and 39.5 ± 9.1 respectively. Post hoc comparison shows for Task 1, the Trackpad support causes significantly lower ($p < 0.05$) physical workload than the Baseline interface. Similarly the average task completion times for the Baseline, Intent Inference and Trackpad supports are: Task 1, the time in seconds were 103.3 ± 13.6, 108 ± 14.6 and 113.2 ± 19 respectively; for Task 2, the times were 106.6 ± 17.6, 121 ± 17 and 121.9 ± 19.2 respectively. Perceived workload reported through the NASA-TLX survey suggests average scores of 50.2 ± 9.6, 48.3 ± 10.4 and 52.6 ± 10.5 for the Baseline, Intent Inference and Trackpad interfaces. The reported average System Usability Scores were 72 ± 8.4, 67.9 ± 13.1 and 60.4 ± 11.9 for BL, IS and TS respectively.

4.6.3 Impact of Combining Position and Orientation Support for Bi-manual Tasks

As seen in Figure 4.7, the combination (CS) of Workload scaling and Intent Inference for orientation support resulted in an average of 30.6 ± 7.9 and 36.1 ± 9.4 compared to the 32.3 ± 7.9 and 34.3 ± 10.1 caused by the Baseline interface for Tasks 1 and 2 respectively. The average task completion times in seconds for BL and CS interfaces for Task 1 was 103.3 ± 14 and 126.6 ± 20.2 respectively. Similarly, for Task 2 the task completion times were 106.6 ± 16.6 and 130.6 ± 19.7 for the Baseline and Combined interfaces. Post hoc comparisons show that the Combined interface exhibited significantly higher ($p < 0.05$) task completion times compared to the Baseline interface for both the pouch opening and paper clip pouring tasks. The average NASA-TLX and SUS scores for the Combined interface were 45.8 ± 10.2 and 69.5 ± 12.7, respectively, compared to 50.2 ± 9.6 and 72 ± 8.4 for the Baseline interface.

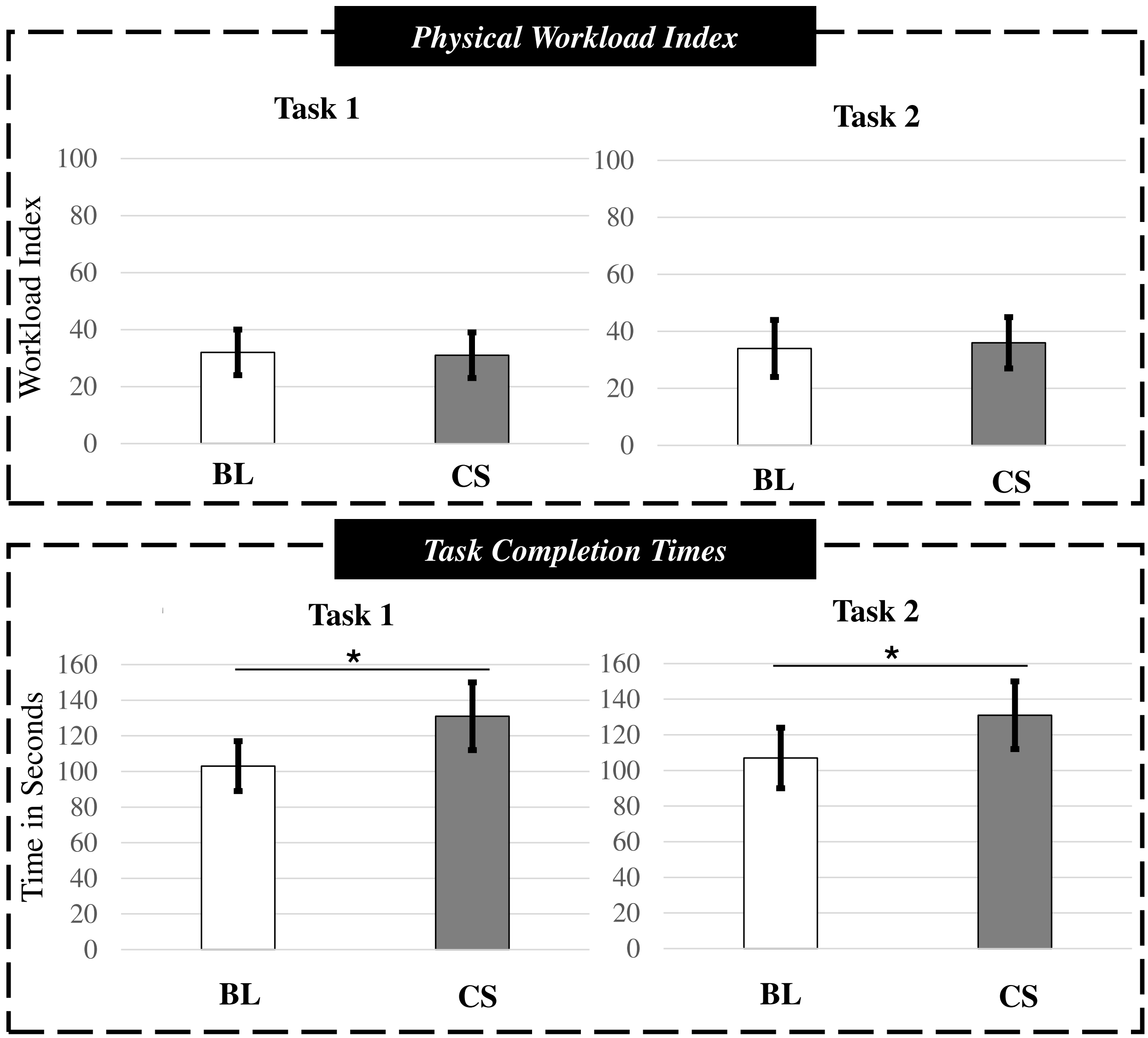


Figure 4.7: The baseline interface is represented as BL, while the combination of recommended position support (WS) and orientation support (IS) is referred to as CS. **Top**: Comparison of physical workload indices for Tasks 1 and 2; **Bottom**: Task completion times for Tasks 1 and 2.

4.7 Discussion

4.7.1 Design of Position Support for Bi-manual Telemanipulation

Regarding **DO1**, we observed that the Workload-based scaling interface significantly reduced both physical workload and task completion times compared to the Baseline interface for both tasks. Specifically, it led to a 27.9% and 17.3% reduction in physical workload for Tasks 1 and 2, respectively, indicating lower physical effort required from the operator. Task completion times were also reduced by 10.9% for the pouch opening task and 6.9% for the pouring task, highlighting improved task efficiency. These results underscore the effectiveness of motion scaling in enhancing precision during object interactions—such as picking up the pouch, grasping the zip, and manipulating the paper clip container—by mitigating the impact of operator noise. The Workload-based scaling interface was also the most preferred among the seven interfaces, selected by 9 out of 16 participants. It received the lowest perceived workload in the NASA-TLX survey and the highest SUS scores, indicating strong subjective performance. In contrast, other motion scaling approaches like Auto and Manual scaling often reduced operator performance and, in some cases, increased physical workload. For instance, during Task 1, the Manual scaling interface resulted in a 15.2% increase in physical workload and a 13.3% increase in task completion time compared to the Baseline. Similarly, the Auto scaling interface increased task completion times by 17.2% and 3% for Tasks 1 and 2, respectively. One possible reason for the reduced performance of the Auto scaling interface is the inefficient design of triggering assistance. While Auto scaling provides continuous access to action assistance—making it the ”safest” design—it lacks explicit operator control, which can cause the interface to feel unresponsive or unpredictable. This may disrupt the operator's motor planning and lead to over- or under-compensation in control inputs, increasing effort and task duration. Notably, no participant selected Auto scaling as their preferred interface, and many highlighted the increased effort and frustration associated with its use. Manual scaling, on the other hand, requires users to explicitly toggle between scaled and unscaled motion, potentially disrupting task flow and adding cognitive load. Only 1 out of 16 participants selected it as their preferred interface. Participants reported that the need to manually manage scaling increased interface complexity and sometimes led to confusion about whether assistance was active. Several participants avoided using Manual scaling altogether to keep the interaction simple. These findings support the benefit of motion scaling under an assist-as-needed paradigm—triggered by operator workload—which improves task performance, reduces physical workload, and enhances the remote manipulation experience in object interaction tasks like pouch opening and object pouring. However, if assistance is triggered incorrectly or unpredictably, it can negatively impact performance and experience, as also suggested in [146].

4.7.2 Design of Orientation Support for Bi-manual Telemanipulation

For **DO2**, both orientation support interfaces—Intent Inference (IS) and Trackpad (TS)—were effective in reducing physical workload during the pouch opening task. Specifically, IS and TS resulted in 12.5% and 25% lower physical workload, respectively, compared to the Baseline interface. The ability of the participants to operate the rotational motions from a comfortable arm posture could have resulted in the decrease in physical workload while using the TS interface for this task. However, this came at the cost of increased task completion times: 4.5% for IS and 9.6% for TS relative to the Baseline. Notably, during the pouch opening sub-action—the task segment requiring the most orientation control—both interfaces actually reduced completion time by 8.2% (IS) and 9.4% (TS), demonstrating the benefit of orientation support for rotational motions. In Task 2, the IS interface remained the lowest fatigue-inducing orientation support design, causing 12% less physical workload than the Baseline. Within Task 2, the scanning and pouring sub-actions saw physical workload reductions of 11% and 13%, respectively, with IS, further supporting its efficacy in rotational assistance. This benefit was observed despite a 14.1% increase in overall task completion time for the pouring task across both IS and TS interfaces. Conversely, the TS interface led to a 13% increase in physical workload, suggesting usability challenges. This may stem from the unintuitive mapping between the user's trackpad inputs and the robot's rotational behavior, often leading to control mistakes and the need for corrective actions. These issues were also reflected in subjective feedback: the Trackpad interface received the lowest System Usability Scale (SUS) score, averaging 60.4 ± 11.9. In contrast, the IS interface outperformed the Baseline in terms of physical workload and improved rotational motion handling. Participants reported it to be more intuitive, and it was the second most preferred interface overall, chosen by 4 out of 16 participants for general-purpose bi-manual remote manipulation. While some users noted that minor intended rotations could be mistakenly filtered out by the IS logic (described in Section 4.4), this was generally considered a minor issue that did not outweigh the benefits of reduced effort and increased control precision thus being the preferable orientation support design for **DO2**.

4.7.3 Drawback of Combining Multiple Action Support

The Combination of Workload scaling and Intent Inference for orientation support performed considerably worse than the Baseline interface in terms of task completion times and physical workload compared to all other interfaces. Despite being the combination of the two most successful action assistance in this study, participants suggested that the ambiguity caused by multiple different assistance caused confusion and frustration that resulted in decreased performance.

4.8 Conclusions

The user study presented in this paper evaluated different position and orientation support mechanisms and their impact on improving operator performance and reducing workload in remote manipulation tasks. Our results demonstrate that motion scaling triggered based on operator workload effectively reduced physical workload and task completion time, while intent-based orientation support improved performance during rotation-heavy subtasks. These interfaces also received favorable subjective feedback, suggesting that assistance triggered based on user state or intent enhances usability and efficiency. In contrast, interfaces that provided constant or manually controlled assistance—such as Auto scaling, Manual scaling, and Trackpad—often increased task complexity, led to frustrating corrective actions, and were rated poorly by participants. These findings indicate that excessive or poorly timed assistance may negatively impact the remote manipulation experience, emphasizing the importance of designing assist-as-needed interfaces. This work is limited by its relatively small participant pool and task simplicity. Future work will expand this evaluation to more complex and generalizable bi-manual remote manipulation tasks and a larger, more diverse user population. Additionally, a more systematic investigation of how position and orientation supports can be combined is warranted. The assistance should ideally work in conjunction with robust perception systems and Augmented Reality visual cues to enhance operator awareness of robot state and assistance activation while maximizing the benefits of multi-modal action assistance that have shown promise when evaluated independently.

Chapter 5

Action Assistance for Motion Mapping Interface to Reduce Physical Workload

5.1 Motivation

Motion mapping interfaces are suitable for freeform teleoperation which can perform unstructured tasks, like collecting a mixture of deformable and rigid objects in a cluttered workspace. Such tasks are challenging for fully automated systems. However the physical fatigue caused due to robot teleoperation via a motion mapping interface is not trivial, particularly when teleoperation lasts for extended durations. Such physical workload not only influences the teleoperator's task performance, but may also negatively influence the worker's perception and attitude towards the usage of teleoperation interface as well as nursing robot technologies.

This chapter will investigate how shared autonomy reduces the physical fatigue incurred while using the motion mapping teleoperation interface. Shared autonomy for teleoperation assistance has been used to enhance the functionality of the slave platform [155] and improve the accuracy of robot teleoperation [156]. Our prior research as seen in Section 2 identified the fatigue causing actions by assessing the muscle effort while using the whole-body motion mapping teleoperation interface to perform general assistive tasks. Precise manipulation and steady posture maintenance were identified as actions that caused physical fatigue. Based on our findings, we hypothesize that automating the fatigue-causing task components will reduce the physical effort of teleoperation via motion mapping.

5.2 Literature Review

Tele-action Assistance for Motion Mapping Interface — Within the spectrum of automation ranging from fully manual to fully automated, action support and shared control are often used to assist freeform teleoperation using motion mapping interfaces (for a review of levels of robot autonomy refer [47]). Action support like Degree of Freedom separation and motion scaling (as seen in Section 3) have been used to improve upon the performance of motion mapping control interfaces. Shared control is mostly used to assist the operator in actions towards a goal or generating motion along certain trajectories. Rakita et. al. have recently developed teleoperation assistance for a motion mapping interface which uses a predict-then-act strategy where the implementation infers an action based on a bimanual action library and engages an appropriate assistance mode to enhance efficiency [157]. This implementation blends the suboptimal user translational and rotation control inputs with known translation and rotation paths in space in which the user can easily guide the robot. Laghi et al[158] combined arm motion tracking, impedance control and hand gesture recognition for using a single arm of the operator to perform bimanual manipulation.

As workload is an essential factor in robot teleoperation, both subjective and objective methods have been proposed to comprehensively evaluate the mental and physical workload. Most of the subjective measurement for both mental and physical workload relies on user surveys [159] like NASA Task Load Index [160] and customized questionnaires. The commonly used objective metrics for mental workload is heart rate measurement which increases with increase in cognitive workload [161]. Limited work has used quantitative and objective metrics to assess and monitor the physical workload in robot teleoperation via motion mapping.

Limitations in Literature – Limited work has been done to design shared autonomy which considers the effects of physical fatigue in teleoperation and thus improve the "ergonomics of teleoperation assistance". To fill this gap, we will explore how to use shared autonomy to manage the physical workload in freeform teleoperation. In this chapter we propose to use a manually-triggered autonomous function to assist precise grasping such that the teleoperators will not be constrained to follow a specific task structure or trajectory and identify the actions that can be automated to reduce physical demand. Additionally, both subjective metrics like user surveys in addition to objective metrics for physical workload monitoring muscle activity will be used to comprehensively analyze and identify the impact of the developed human-robot control paradigm on physical workload.

5.3 Platform And Teleoperation Assistance

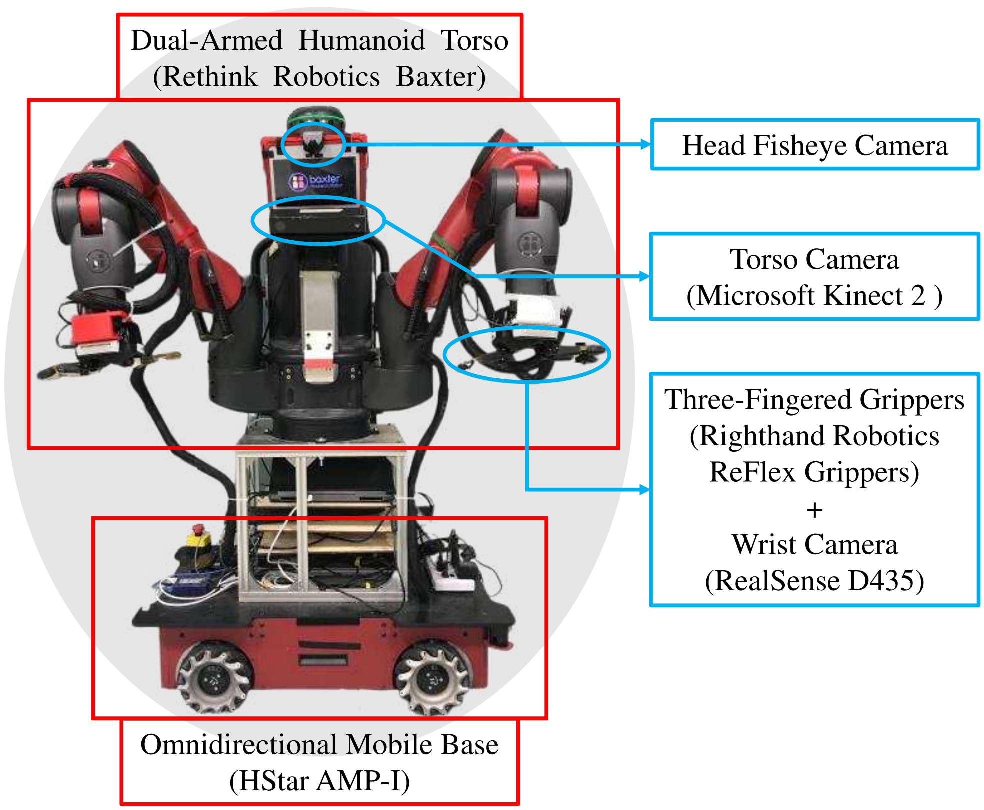


Figure 5.1: Tele-robotic Intelligent Nursing Assistant (TRINA) system.

Here we describe the robot platform, the design of the whole-body motion mapping interface and the autonomous function for teleoperation assistance. The Tele-Robotic Intelligent Nursing Assistant (TRINA) (Figure 5.1) consists of a dual-armed humanoid torso (Rethink Robotics Baxter), an omnidirectional mobile base (HStar AMP-I) and two three-fingered grippers (Righthand Robotics ReFlex grippers). The visual sensor suite of this nursing robot includes a $180°$ fisheye camera on the head, a Microsoft Kinect 2 attached to the robot's chest and two Intel RealSense D435 depth cameras attached to the robot's wrists. Table 5.1 defines the controls for the motion mapping interface. Motion capture of the teleoperator is done using the Vicon motion capture system (10 Vero cameras). This system captures passive reflective markers attached to the human torso, arms and legs. The users thus control actions like reaching, grasping, locomotion of the mobile humanoid robot, the selection of cameras and the movement of these cameras. Human motion was captured at 100 Hz and streamed at 50 Hz for robot control. The proportions of the subjects (height, limb lengths, etc) do not affect the end-effector positions of the robot. This is because the position and orientation of the wrist and the swivel angle of the teloperator is mapped to the robot during teleoperation. The swivel angle is defined as the rotation of the position of the elbow of the operator with respect to the axis connecting the centers of the shoulder and wrist joints [3], which indicates the operator's arm posture.

The flowchart in Figure 5.2 describes the design of the autonomous grasping function for teleoperation assistance. The Kinect camera detected all the objects to grasp in the workspace using Mask-RCNN [162, 163]. As the teleoperator controls the robot hand to reach into the Teleoperation Assistance Zone (TAZ) — a bounding box of $(2 \times height) \times (3 \times thickness) \times (5 \times width)$ (cm^3) around the center of an object, the object will be locked as the "target", and an autonomous reaching-to-grasp motion will be planned for this object. Based on the bounding box created by the computer vision module, points on the mid-point of the left and right sides of the box were identified as the target grasping points. According to which robot arm is in the TAZ, the corresponding

Teleoperation Input	**Robot Function**
Robot's Upper Body	
Hand position & orientation	End-effector position & orientation
Arm posture & orientation	Manipulator arm posture
Rotate upper body	Rotate mobile base orientation
Hand open/close	Gripper opens/closes
Angle between feet $\geq 60^o$	Gripper preshape pinch grasp
Angle between feet $< 60^o$	Gripper preshape power grasp
Right shank flexion	Activate teleoperation assistance
Robot's Lower Body	
Squat	Engage/Disengage teleoperation
Lift left leg	Switch primary camera view
Lift right leg	Switch secondary camera view
Leg steps forward/backward	Mobile base moves front/back
Left (right) leg steps left (right)	Mobile base moves left (right)

Table 5.1: Motion Mapping Teleoperation Interface.

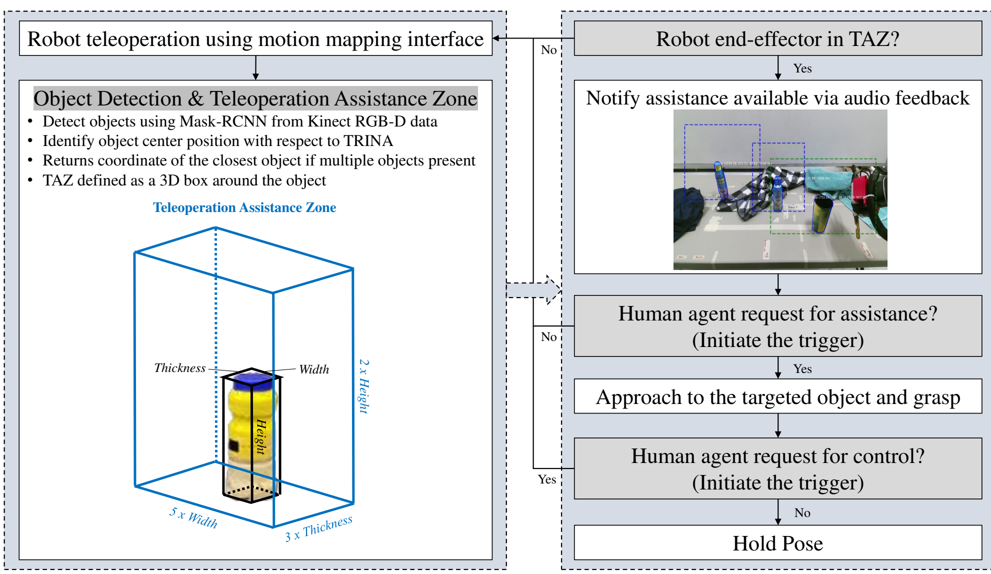


Figure 5.2: Autonomous Grasping Function for Teleoperation Assistance.

nearest target point is selected. If both the arms are in the TAZ, the right hand is selected by default to move to the target point on the right side of the bounding box. The inverse kinematics for this target location is solved and the joints of the selected robot arm are moved to these desired joint angles. The user is informed that the autonomous grasp function is ready to be triggered based on auditory and visual cues (Figure 5.3).

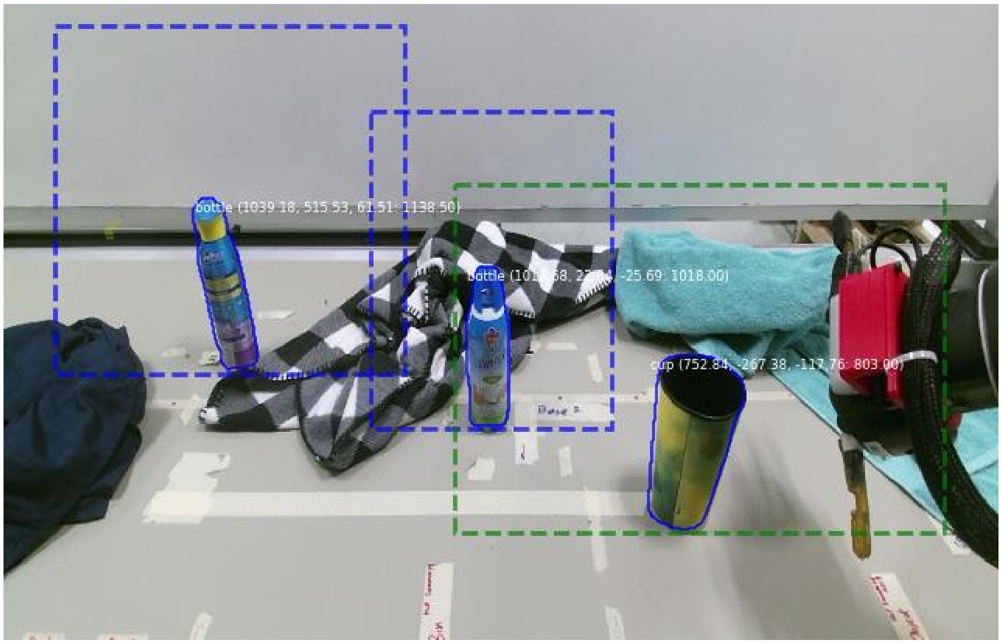


Figure 5.3: Demonstration of object detection in cluttered environment.

5.4 Physical Workload Estimation

We used Wireless sEMG sensors (Trigno^{TM} from Delsys Inc.) to record the EMG signals at 1,000 Hz for 10 individual muscles, namely the Anterior and Middle fibers of the Deltoid, the Biceps, the Brachioradialis and the Trapezius of the left and right sides of the body. These muscles were selected as they are the most involved in controlling human upper body motion.

Our analysis of the sEMG data aims to evaluate individual muscle effort during teleoperation using motion mapping with and without the assistance feature. Fig. 5.4 illustrates our data analysis process where we have determined individual muscle contraction levels and contraction duration. In the graph on the bottom right the black line represents the muscle contraction and the green bars collectively represent the contraction duration.

The recorded EMG signals are within the 40 Hz-700 Hz range in the spectrum domain. In the muscle effort analysis, the raw EMG data was pre-processed using a high pass filter (cutoff

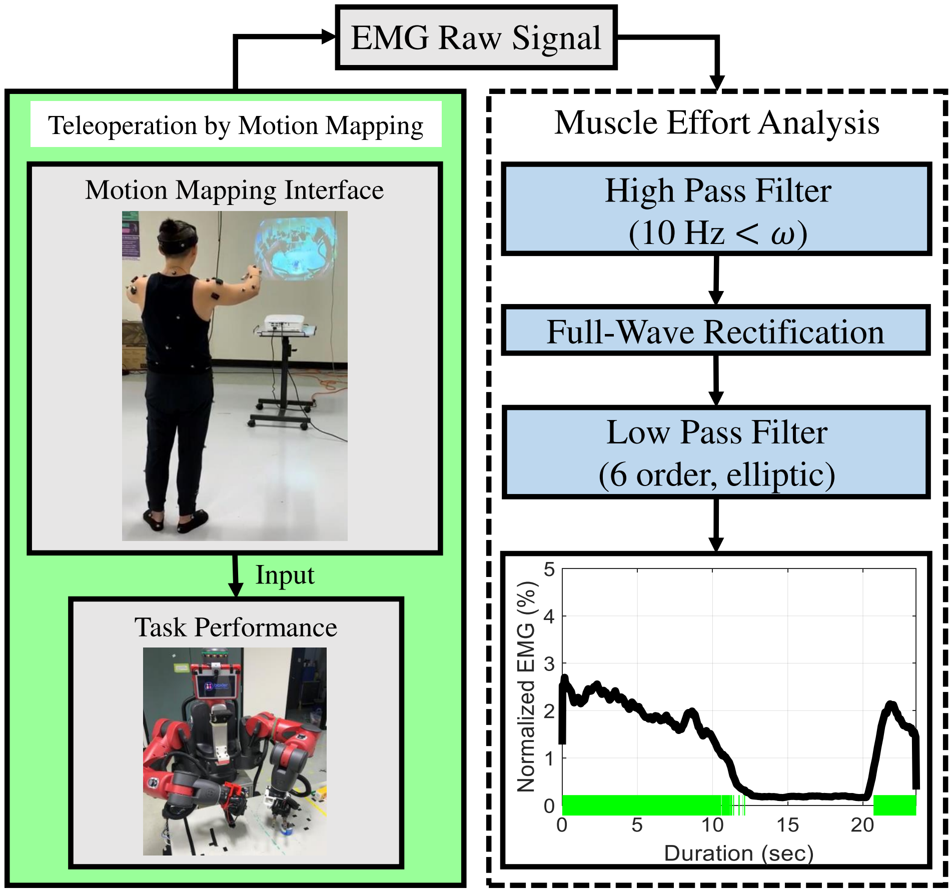


Figure 5.4: Muscle efforts analysis process.

frequency 10 Hz), to remove the soft tissue artifact and offset the frequency baseline. The processed signal further went through a full wave rectification and then a sixth order elliptical low pass filter (cutoff frequency 50 Hz), to remove noise and transients and develop a linear envelope of the EMG signal [164]. Using the method in [164], we determine the appropriate threshold for muscle contraction to be the signal baseline offset by thrice the standard deviation of the muscle static contraction obtained from the first 200 frames of the EMG signal in the MVC test.

5.5 User Study

Our user study aims to evaluate if the teleoperation assistance will reduce the physical workload and improve the task performance of robot teleoperation via whole-body motion mapping interface. The hypotheses we evaluate include:

- **Hypothesis 1:** The proposed teleoperation assistance will reduce the teleoperator's physical workload in terms of muscle effort.
- **Hypothesis 2:** Teleoperators will prefer to use the teleoperation assistance based on their experience performing the tasks with and without teleoperation assistance. With teleoperation assistance, users will prefer to work more with the teleoperated robots in the future.

Participant and Task –

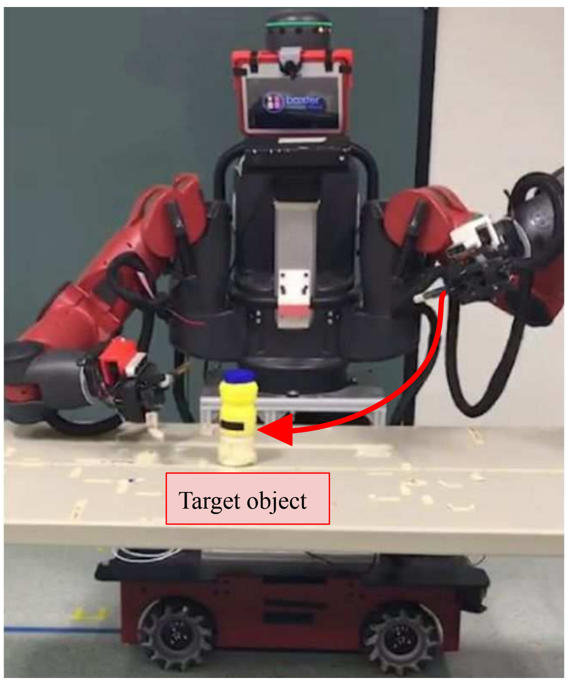


Figure 5.5: The operator is tasked with picking up a single object in the robot workspace similar to the object shown above with either their dominant or non-dominant arm.

We recruited N=8 participants (6 male and 2 female, all right-handed) to use the teleoperated robot system described in Section 5.3 to perform the following task: reaching to grasp an individual object with the dominant and non-dominant arm (as seen in Figure 5.5). These participants share a similar age demographic with future nursing professionals and likely possess comparable physical capabilities and interface preferences, making their input valuable for the early-stage development of telerobotic healthcare interfaces. We choose this task because precise orientation control in reaching-to-grasp has been demonstrated to be challenging for teleoperation and requires careful design of teleoperation interface assistance (e.g., [60]). Our prior work has indicated the teleoperation of precise manipulation is one of the most fatigue-causing factors.

Experiment Procedure –

Preparation: Our experiment uses EMG-measured muscle effort to assess the physical workload. Before the experiment, each participant performs a maximum voluntary contraction (MVC) test for each muscle. The collected data is used for normalizing the EMG signal with respect to the maximum force generated by each muscle [165]. Each participant undergoes a training session to get familiar with the teleoperation interface, the autonomous grasping function and the robot. The training task is to pick up a bottle on the counter and place it in a basket. The participants are allowed to practice in this training session until they feel confident and comfortable to use the teleoperation interface and assistive function.

Task Performance: In this session, a participant was instructed to reach and grab a bottle placed on the counter (Figure 5.5). The participants were asked to grab the objects for five repetitions, using their dominant and non-dominant arms, with and without the teleoperation assistance (Total number of trials = 5 repetitions $\times$ 2 arms $\times$ 2 modes). The order of arms and modes were randomized. All the repetitions of the object grasping task were set to have the same initial robot arm configuration, initial and final location of the object. The participants were required to pick up and place the object in a stable manner. During each trial, we record the time for completing the task, the number of times the object was knocked down and the EMG signal of the muscle groups for muscle effort analysis (described in Section 5.4.). The participants also answered survey questions about their teleoperation experience, in the NASA Task Load Index (NASA-TLX) format on a 1–-7 Likert scale.

5.6 Impact of Assistive Autonomy

We compared the muscle efforts, task completion time and numbers of errors in the object grasping task, to objectively and quantitatively assess the teleoperators' physical workload reduction when using teleoperation assistance. We further use the results from the NASA-TLX survey and customized questionnaires to assess their perception of workload, preference of teleoperation assistance and their change of attitude toward teleoperated robot technologies.

Reduced Muscle Effort with Assistive Autonomy –

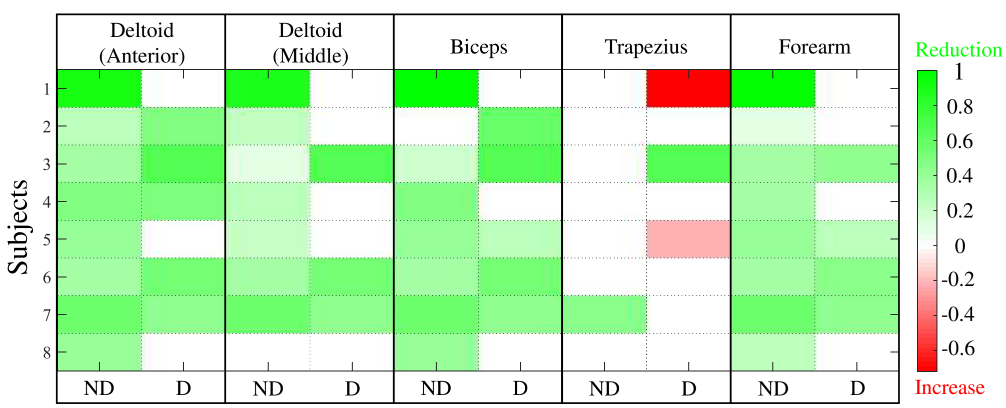


Figure 5.6: Comparison of physical effort across all muscles with dominant (D) and non-dominant (ND) hand.

We compared the muscle efforts between teleoperation with and without the assistance across all muscle groups for each participant. As shown in Figure 5.6, most of the muscles had a significant reduction in physical effort (marked as green) with a higher level of relaxation for the deltoids and biceps of the dominant/non-dominant hand. The different levels of muscle effort was calculated using the Kullback-Leibler (KL) divergence measurement and all the results were normalized by the maximum value. It is noted that the Trapezius muscle however has reduced reduction (marked as white) or increased physical effort as shown by the red marks for 2 subjects. Overall, the assistance function performed equally effectively on both arms for all subjects thus validating **Hypothesis 1**: Assistive autonomy reduces the teleoeprator's physical workload in terms of muscle effort.

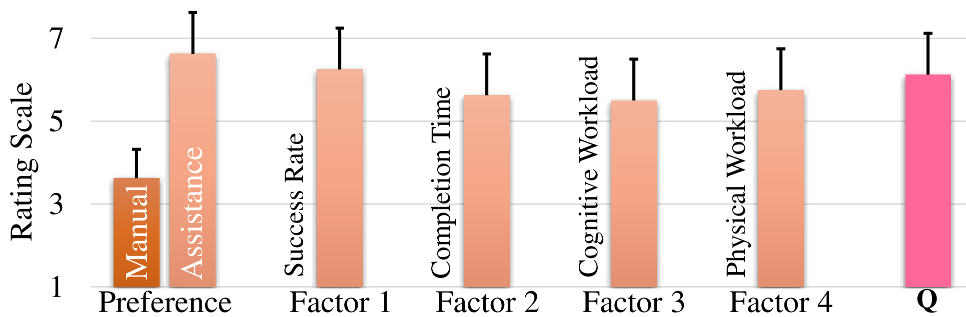


Figure 5.7: Rating of preference.

Subjective Preference for Teleoperation Assistance – After the experiment, participants rated in hindsight their preference for teleoperation assistance and manual control during robot teleoperation. As there was a greater preference for the assistive function, the users were questioned on what factors made them favor teleoperation assistance more. They were asked to state to what extent the teleoperation assistance can (1) increase the success rate; (2) reduce the task completion time; (3) reduce the cognitive workload; and (4) reduce the physical workload based on their experience on a 1-7 Likert scale with 1 being the least and 7 being the most in terms of agreement. Finally, we evaluate their acceptance of using teleoperated robot technologies by asking the question (Q): "With the teleoperation assistance, do you prefer to work more with teleoperated robots?". The results represented in Figure 5.7 highlight the participant's belief that teleoperation assistance improves performance. This further supports **Hypothesis 2**: Teleoperators will prefer to use the teleoperation assistance based on their experience performing the tasks with and without teleoperation assistance. With teleoperation assistance, users will prefer to work more with the teleoperated robots in the future.

5.7 Summary and Outlook

Motion mapping as a teleoperation interface proves to be the most intuitive and preferred means of teleoperation. However, the physical fatigue developed because of using this interface cannot be ignored and it can result in the rejection of this interface as a means of everyday sustained use.

We used a shared autonomous control interface to tackle the issue of physical fatigue. Nursing tasks require a lot of decision-making skills and occur in an unstructured environment. As a result, the entire task cannot be automated as current automation techniques do not capture the nuances of operator controlled teleoperation. In this chapter, we have proposed how automating reach-to-grasp reduces physical effort and fatigue in the operator and improves their perception towards teleoperation. Aspects of teleoperation like locomotion and gross manipulation is left to the operator while the finer manipulation involved with object grasping is automated and can be triggered on and off base on the operator's needs.

We have demonstrated that augmenting the direct, freeform interfaces for robot control with a little bit of robot autonomy will lead to flexible and reliable robot control for complex tasks. Our proposed robot autonomy effectively reduced the operator's physical workload in the control of reaching-to-grasp motions. Our future work will further develop a variety of robot autonomy to other fine motor skills necessary for quarantine patient care (see the fine manipulation tasks listed in [166]). We will also explore how to design robot autonomy to be fatigue-adaptive, such as triggering the autonomous function based on task context, inferred user intents, and estimated physical fatigue level.

Chapter 6

Perception Assistance for Remote Manipulation Autonomy

6.1 Motivation

The work in this chapter aims to investigate what augmented reality (AR) visual cues humans prefer to use when controlling or supervising remote robot manipulation. To enhance remote perception, AR visual cues are overlaid on the video stream from remote cameras to indicate the robot and task states, and the spatial relationship in a 3D environment, which may be difficult for human operators to estimate precisely. They are also used to indicate the motion, action, path and task that robot autonomy plans to perform, in order to enhance human's understanding of robot's intent, behavior and capabilities. As seen in Section 4, action assistance improves user performance and their preference for remote manipulation systems. Perception assistance in the form of Augmented Reality cues is needed to reflect the change in requirements based on the type of assistance available to the user. Thus far, effective AR visual cues are mostly developed case-by-case for remote robot manipulation under direct to supervisory control. It is still not clear what AR visual cues humans need or prefer to use to control a remote manipulator robot with various levels of autonomy.

To this end, we proposed systematic AR visual cues for remote robot manipulation assistance. These AR cues can guide human's control of robot motion toward the target and around the obstacles in a 3D workspace, and indicate whether the robot autonomy is activated,and its planned

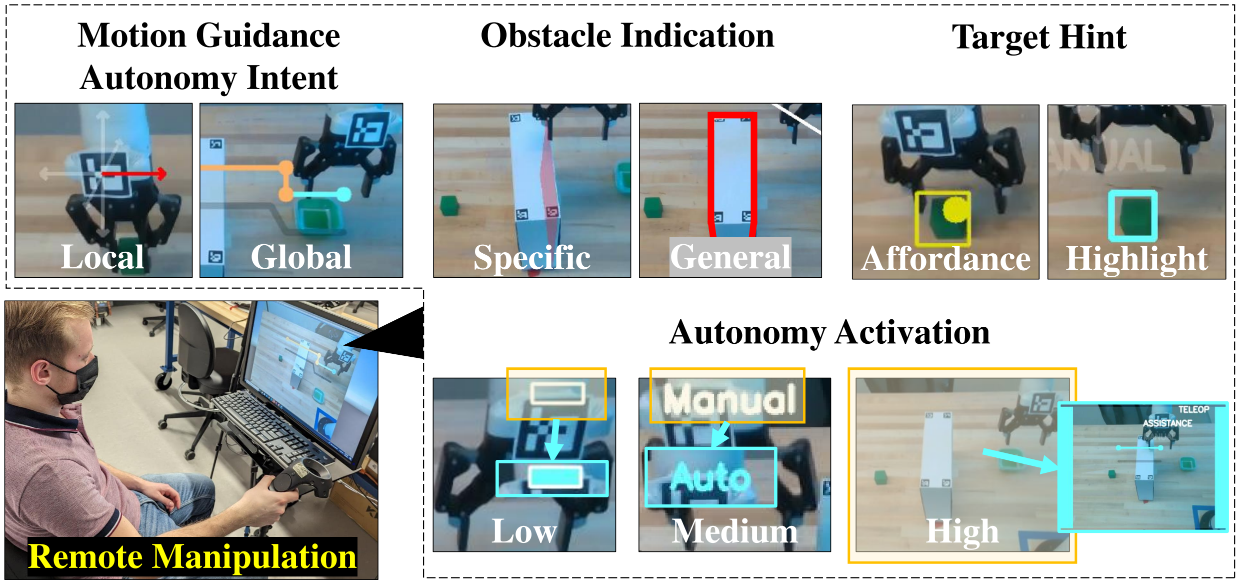


Figure 6.1: Proposed AR visual cues to assist humans to control or supervise remote robot manipulation, and to communicate the robot autonomy's activation, capabilities, and intents.

motion or action. We conducted a user study where the participants (N=18) controlled a remote manipulator robot with adjustable autonomy to move around an obstacle to pick and place an object on the counter space. The robot can operate under direct human control, or assist humans with autonomy to avoid obstacles and environmental constraints, to perform error-prone precise manipulation actions, or plan and execute the entire robot motion to pick and place an object under human supervision. We compared the AR cues the participants chose for each level of autonomy, to the choices of experienced interface users. We also compared the participants' initial choices based on video demos of AR features used to understand the interface to their choices after hands-on practice with the interface. Our results show that the preference for AR visual cues varied with the level of autonomy used to control the robot. From direct to supervisory control, the preference shifted from AR visual cues that guide their control of robot motion to cues that communicate the robot's intended action and path plan. Additionally, participants tended to change their initial preference for the AR visual cues after hands-on practice tending to agree with the recommendation of experienced users.

6.2 Literature Review

Our proposed AR visual interface aims to enable humans to effectively assist the robot with limited autonomy for structured pick-and-place object manipulation to handle a wide range of unstructured, complex manipulation tasks in a cluttered environment. Although the robot autonomy cannot handle the entire task, it can reliably enforce motion constraints, and plan and execute autonomous actions. Effective human-robot collaboration for such tasks enables humans and robots to contribute complementary skills to improve the overall task performance, reduce the human workload, and task complexity for robot autonomy [167, 168, 169]. Related work shows that humans can operate the robot to separate the cluttered and entangled objects based on their task knowledge and experience, rearrange the objects and workspace into more organized and predictable positions, in order to reduce the robot sensor occlusion and facilitate the robot's autonomous semantic segmentation of workspace [167]. To reduce the complexity of autonomous motion planning, humans can guide a robot towards target locations or objects across cluttered workspace [46], control nonprehensile object manipulation [170], select the grasping points to handle (deformable) objects [118]. In shared and supervisory control, humans can select or confirm the robot's actions [171], review and revise robot action sequence or task plan [172], and supervise the autonomy's execution [173].

Recent related work in literature has provided a taxonomy for the virtual, augmented, and mixed reality for human-robot interaction [39]. Overall, the AR visual cues can be augmented on the *robot*, *interface* and *environment*, in order to visualize the robot's internal and external states (e.g., internal reading and readiness, robot pose and location), as well as the robot's comprehension of *environment* (e.g., the purview, numerical readings, videos and images, 3D data from external sensors, robot-sensed/internal/user-defined spatial region,), *entities* (e.g., entity labels, attributes, locations and appearance), and *tasks* (e.g., heading, waypoint, call-out, trajectory, spatial preview, trajectory, alternation preview, command options, task statuses and instructions). Related work also recommends various effective integration of AR visual cues for manipulation tasks in the 3D environment (e.g., [40, 41, 42, 43, 44, 45]), yet these case-by-case designs did not reveal how the AR visual cues should be designed for a human-robot collaborative manipulation with various levels of autonomy. Another recent work revealed similarity in the design principles for data visualization and AR for assisting remote manipulation control [174].

Limitations in literature – To enhance the operator's remote perception, it is recommended to *find relevant information* by data salience and clutter, *synthesize data across sources* by comparison and multiple views, *identify anomalies* by data provenance and statistical estimation, *make predictions* by uncertainty and temporal data, *assess risks* by direct attention and value estimation. While the design principles are valid for manipulation tasks, it is unclear how to apply them to the AR visual designed for various human-robot collaborations for remote robot manipulation.

6.3 Proposed Method

Here we introduce the remote manipulation system and the human-robot task division in each control mode, in order to provide context for AR visual cue design, and present our systematic AR visual cue design for the human-robot collaborative control at four different levels of autonomy. In this work, Augmented Reality cues were focused on as our prior work [44] identified Augmented Reality cues preferable to other forms of notification like haptic cues for information pertaining to continuous monitoring while teleoperating.

Remote Manipulation System: Our prior work [25] had integrated a robotic system with an adjustable level of autonomy for remote, unstructured manipulation. This system uses a HTC Vive handheld controller to control a 7 degrees of freedom Kinova Gen 3 robotic manipulator with a two-fingered Robotiq gripper. A desktop monitor displayed a graphical user interface (GUI) on a Unity 3D window (1920 $\times$ 1150 pixel) to stream the video from the workspace camera (RealSense

Table 6.1: Human (**H**) and Robot (**R**) task division in each control mode.

Control Mode	**Gross Manipulation**	**Obstacle Avoidance**	**Precise Manipulation**	**Autonomy Activation**
Direct	H	H	H	N/A
Assisted	H/R	H/R	H/R	Auto
Shared	H/R	R	R	Auto
Supervisory	R	R	R	Auto

D435f) at 30 Hz.

Human-Robot Control Paradigm: Table 6.1 illustrates the human-robot task division in each control mode. In the **Direct Control** mode, a human manually controls the robot through the entire task. In the **Assisted Control** mode, robot autonomy uses a virtual fixture to constrain human-controlled robot motion in directions that avoid collision with environmental constraints and obstacles as well as move to grasp/place an object. In the **Shared Control** mode, the robot assists human-controlled gross manipulation motions (e.g., approaching or moving an object) with automated motions for collision avoidance, and for precise manipulation (e.g., grasping or placing) when the robot end-effector is close enough to the target object or location. In the **Supervisory Control** mode, robot autonomy plans the motion for the entire pick-and-place task, while humans supervise its execution. The robot's autonomous perception using AruCo markers [175] are available for all the control modes, in order to locate the target object, container, obstacle, and track the robot end-effector.

AR Visual Cues: Shown in Figure 6.1, we propose five types of AR visual cues and representation options:

- **Motion Guidance** indicates the robot's instantaneous motion direction using a *3-axis arrows* overlaid on the robot end-effector or display the robot's suggested path [176].

Table 6.2: AR visual cue choices recommended by experienced users.

Control Mode	Motion Guidance	Obstacle Indication	Target Hint	Autonomy Activation	Autonomy Intention
Direct	Arrow	Planes	Affordance	–	–
Assisted	Arrow	Box	None	Low	Path
Shared	Path	None	Highlight	High	Path
Supervisory	–	None	None	Medium	Path

- **Obstacle Indication** has the options to highlight close-to-collision features on the obstacle (e.g., plane, edge or vertex, as in [171]), and to display the obstacle's 3D bounding box (as in [177]).

- **Target Hint** provides the grasp/place *affordance* [44] by changing the color of the square around the target and the dot overlaid on the robot end-effector. Alternatively, the target can be *highlighted* [46] to provide an intent inference to the user.

- **Autonomy Activation** indicates if the robot autonomy has been activated. We provide three options for the representation of different visual salience, including a blue light (low salience, as in [178]), text with "AUTO" (medium salience, as in [179]), and blue bars on both sides of the camera view (high salience, as in [171]).

- **Autonomy Intention** has the option to indicate the robot's motion, action, and path plan (as in [178, 180]). The autonomy intention is displayed similar to the visual cues used for motion guidance.

Our user study allows participants to choose the types and options of AR visual cues for each control mode. Table 6.2 shows the choices recommended by experienced users (users (N=5) with at least 100 hours of experience in remote manipulation.)

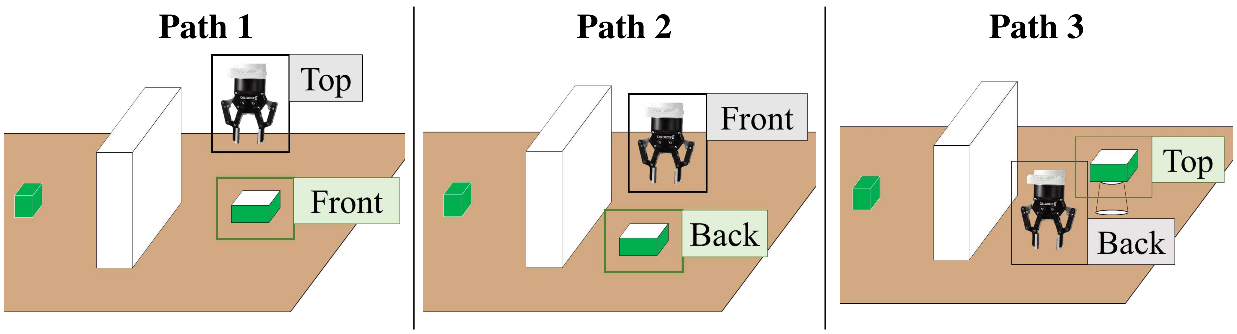


Figure 6.2: Workspace configurations for the pick and place experimental task. The robot end-effector starts towards the top, front and back of the container for Paths 1, 2 and 3 respectively. The target container is towards the front, back and top of the obstacle for Paths 1, 2 and 3 respectively.

6.4 User Study

In the user study presented in this chapter we explored the following Research Questions: We conduct a user study to investigate: **(RQ1)** what AR visual cues humans prefer when controlling the robot with various levels of autonomy, and **(RQ2)** whether this preference can be influenced by the way humans learn to use the interface.

Participants and Tasks — We recruited 18 participants (13 male and 5 female, 25.4 ± 6.9 years) from the WPI campus to perform a single object pick-and-place task with an obstacle between the object and the box. These participants not only share a similar age demographic with future nursing professionals but are also likely to have comparable exposure to video games and interactive media, making them relevant for evaluating AR-based interfaces where familiarity with spatial interaction and immersive technologies can influence usability and design feedback. The experimental protocol was approved by WPI's Institutional Review Board. Shown in Figure 6.2, participants controlled the robot to pick up an object on the other side of an obstacle (box of size 200 mm $\times$ 80 mm $\times$ 200 mm), and bring it back and place it in a small container. Participants performed the task under three different initial task states: the object location remained the same, while the container and robot were placed in three different ways such that the robot planned path would move around different sides of the obstacle to reach the object and container.

Experimental Procedure — Participants were trained on the baseline control interface and three levels of assistive autonomy before the experiment. The experimenter then provided video instructions to demonstrate each AR feature and gather their preferences (Preference 1) for each level of autonomy. Participants performed a single trial of the task in Path 1 using their selected AR visual cues for all control modes. Following the completion of the task in Path 1, participants were given the opportunity to switch their preferences (Preference 2) based on their hands-on experience and then perform the same task again with Paths 2 and 3 for one trial each. Lastly, participants were required to perform the same task using the AR feature proposed by the experienced user for Paths 2 and 3 for one trial each and made their final selection of AR preferences (Preference 3). Note that a trial was skipped if the preferred AR combination was the same between Preference 1, 2, and 3.

Data Collection and Analysis — To assess control efficiency, we recorded the complete length of the trajectory covered by the handheld controller. Previous research has shown that the size of the pupil increases as the level of stress rises [181]. We utilized pupil diameter as a measure for estimating subject-specific cognitive workload. To establish a baseline for each participant, we instructed them to look at a blank screen for 30 seconds, assuming a stress-free state. The cognitive workload was calculated by finding the average difference between real-time pupil diameter and baseline value, which was then normalized by the maximum average distance between real-time pupil diameter and baseline value across all trials for each participant.

6.5 AR Preferences and Impact of Training

A post-hoc power analysis with (1-β) = 0.8 and α = 0.05 found an observed power of 0.85 (d = -1.05) for the participant size (N=18). For the comparisons in control efficiency and cognitive workload, we first used F-test to compare the variances of the two sample groups, and then used Student's t-test (equal variance) and Welch's t-test (unequal variance) with $p < .05$. Note that

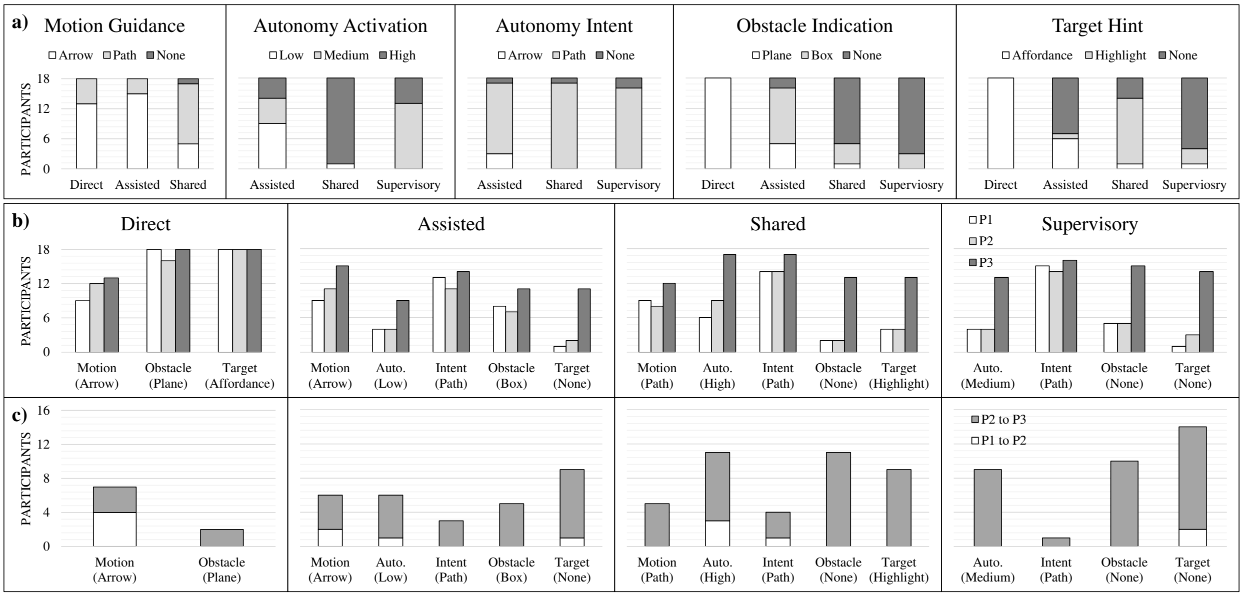


Figure 6.3: a) The number of participants who selected the recommended AR feature for all the control interfaces; b) The number of participants who changed their preferences from other AR features to the recommended AR feature when moving from Pref 1-2 and Pref 2-3; c) The final selections of the participants for the different AR features for all the control interfaces.

preferences 1, 2, and 3 are referred to as P1, P2, and P3 in this section.

AR Preferences for Each Level of Autonomy – Figure 6.3.a presents participants' final selection (after using the recommendation of experienced users, refer to as P3) of each AR visual cue for different levels of robot autonomy.

Direct Control — Most of the participants (13 out of 18) selected local information (arrow) as a preferred way to continuously guide their motion to perform the task and avoid an obstacle. All the participants (18 out of 18) preferred having detailed (planes of the obstacle) information indicating possible collisions and specific (target affordance) methods to indicate if the position of the robot end-effector is good to perform the precise manipulation.

Assisted Control — Most of the participants (15 out of 18) still selected local information (arrow) as a preferred way to guide their motion, especially in the aspect of explicitly showing the required control direction while the assisted autonomy is activated. Half of the participants preferred notification of the activation of the autonomy with a low salience (light-up a small square) method while using global information (path) to be informed of the autonomy intention had been selected by most of the participants (14 out of 18). Most of the participants (11 out of 18) preferred having a general obstacle indication (highlight with a red boundary) that potentially provides a reason for the activation of assisted autonomy to avoid an obstacle. Most of the participants (11 out of 18), however, preferred not to have AR visual cues for precise manipulation with grasping and placing an object because they feel assisted autonomy will handle it and like to have minimal visual clutter on the screen.

Shared Control — In contrast to the direct and assisted control, most of the participants (12 out of 18) selected global information (path) as a preferred AR visual cue to guide their motion to help approach the autonomy zone around the targets and obstacle. Almost all the participants (17 out of 18) preferred the most salience (highlight the entire user interface with color and two thick bars) method to indicate the activation of autonomy and global information (path) to indicate the intent of autonomy. This way the participants could get a better sense of the timing of resuming control when the autonomy was completed. Most of the participants (13 out of 18) preferred to have no AR visual cue for obstacle indication as they feel the autonomy will handle it automatically, and use general information (highlight the current target) to ensure the autonomy is being applied to the correct target.

Supervisory Control — Most of the participants (13 and 16 out of 18 participants respectively) preferred having a medium salience (pop-up text) to indicate autonomy activation and global information (path) to demonstrate the autonomy intention. No AR visual cues for obstacle indication and target hint were selected as they were deemed not necessary by most of the participants (15 and 14 out of 18) as these will be handled by robot autonomy.

Influence of Interface Learning Method — Figure 6.3.b shows the users' preference changes for each AR feature suggested by the experienced user from P1 to P3.

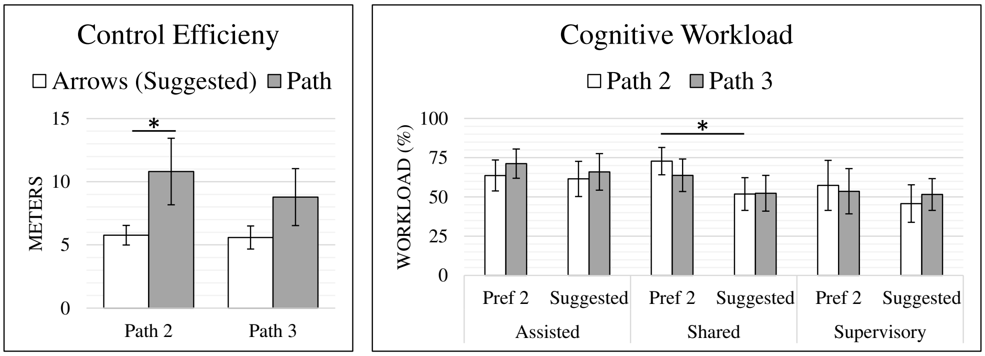


Figure 6.4: Control efficiency evaluated by the handheld controller's trajectory length in direct control and cognitive workload.

Video Instruction-Based (Initial Selection) — In **direct control** mode, half the participants selected local and half the participants selected global information as their preferred motion guidance indication. All the participants already preferred having detailed information (planes of possible collision with the obstacle) for avoiding the obstacle and grasp/place affordance to assist precise manipulation. In **assisted control** mode, half the participants selected local and half the participants selected global information as their preferred motion guidance indication. Few participants selected the low salience (light-up small square) method to be informed of the autonomy activation while most of the participants already chose the global information (planned path) to show the autonomy intention. Similar to direct control, most of the participants still preferred having detailed information (planes) over the general method (highlight obstacle boundary) to avoid the obstacle. Most of the participants chose grasp/place affordance to assist precise manipulation even if assisted autonomy was provided. In **shared control** mode, half the participants selected local and half the participants selected global information as their preferred motion guidance indication. Few of the participants selected the high salience (highlight the entire screen with bars) method to show the autonomy activation while most of the participants already chose the global information (planned path) to show the autonomy intention. Most of the participants selected the general information for obstacle avoidance, over having no AR visual cues, and grasp/place affordance to assist precise manipulation. In **supervisory control** mode, most of the participants preferred having the most salience method to indicate the activation of the autonomy while most of the participants already chose the global information (planned path) to show the autonomy intention. Similar to direct control, most of the participants selected the general information for obstacle avoidance over having no AR visual cues and grasp/place affordance to assist precise manipulation even the autonomy will handle both obstacle avoidance and precise manipulation.

Hands-On Engagement-Based (Intermediate Selection) — In **direct control** mode, a few participants changed their preferences for motion guidance from global information to local information in the form of arrows. The preferences for obstacle collision, and target grasping and placing largely remained the same. For **assisted control** and **shared control** mode, the preferences for the AR features for motion guidance, obstacle collision formation, target grasping/placing, and autonomy indication and intent generally remained the same. This trend continued for **supervisory control** mode as well with minimal changes in the selections for the AR features for obstacle collision information and autonomy indication and intent from the previous selections.

Expert Recommendation-Based (Final Selection) — For **direct control** mode, minimal changes happened across all the AR features between the intermediate selection and the final selection with most of the participants selecting the local information for motion guidance and all the participants wanting detailed information for obstacle collision (planes) and target grasp/place (affordance). For the **assisted control** mode, most of the participants selected local information in the form of arrows for motion guidance with some participants changing their preferences from global information in the form of path AR. Half the participants selected the low salience light notification for autonomy indication with a few participants changing their choices from high salience to low salience when moving from intermediate to final selection. Most participants continued to select the path AR to provide global information about the intent of autonomy. For information about a collision with the obstacle, most participants selected general information in the form of box AR. Finally, for the target hint, most participants chose to have no hint to help with target grasp/place with nearly half the participants changing their intermediate selections. Most of the participants selected global information for motion guidance while making their final selection for **shared control** mode, with some participants changing their selections from local information for motion guidance. Nearly all the participants selected the high salience autonomy indication with nearly half the participants changing their selections from the intermediate selection. Similar to assisted control, the selections for autonomy intent remained largely unchanged, with nearly all the participants preferring to learn about robot autonomy intent through the global information provided by path AR. More than half the participants changed their preference for obstacle information, with most of the participants now preferring no information about the obstacle collision. Most of the participants prefer to highlight the target while grasping or placing with half the participants changing their preference from the previous intermediate selection. For **supervisory control** mode, most of the participants selected the medium salience text indication for autonomy indication with half the participants changing their selection to medium salience. Similar to both assisted and shared control, most participants selected the global information via path AR for autonomy intent with minimal participants changing their preferences. Most of the participants preferred no information about collisions with the obstacle with nearly half the participants changing their preferences from other AR features. Finally, most of the participants now preferred no information regarding the target with most of the participants changing their preferences from the intermediate selection.

6.6 Summary and Outlook

Table 6.3: Usefulness of AR features for each control mode.

Interface	**Priority**
Direct	Motion Guidance>Obstacle>Target
Assisted	Motion Guidance>Autonomy>Intent>Target>Obstacle
Shared	Autonomy>Motion Guidance>Intent>Target>Obstacle
Supervisory	Intent>Autonomy>Target>Obstacle

Impact of Autonomy Level on Perception Assistance Preference — To address **RQ1**, in addition to the participants' final selection of AR features, Table 6.3 shows their subjective feedback on the usefulness rankings of AR features for each control mode. As the level of autonomy transitions from direct to supervisory control, we found that: (1) humans' priority for AR visual cues shifts ***from guiding robot control to communicating autonomy activation and intention*** based on their subjective feedback in Table 6.3; (2) humans' preference for AR visual cues changes ***from providing local information to offering global guidance in robot control*** as indicated by a large portion of participants (13 and 12 out of 18) selecting arrow for direct control and the path for shared control. However, the use of ***global information to display the autonomy's intention*** remains consistent across all interfaces with autonomy, as more than 14 participants preferred having a planned path to display it; (3) the ***efficacy of the AR features that share the same purpose as the robot autonomy is decreased*** as observed by most of the participants preferring not to have any AR visual cues for both obstacle indication and target hint in supervisory control mode.

Impact of Interface Training and Experience on Assistance Preference — Regarding the influences of how participants learn to use the AR visual cues (**RQ2**), we found that ***participants' preference for AR visual cues converged to the recommendation of experienced users***, which is observed by a large change after hands-on robot control using suggested AR features (Figure 6.3.c) and most of the participants selecting the suggested AR features as their final choice across all control modes. Additionally, the participants commented: *"...would like to have clear guidance on how I should move the robot when not much robot autonomy is available."* which was supported by the total trajectory lengths of the handheld controller being significantly shorter ($p < .05$) while using the suggested AR visual cue (arrows) in direct control mode (Figure 6.4). The participants also commented: *"...the most obvious way to inform the activation of the autonomy will be preferred for shared control so that I do not need to put too much effort when the autonomy is on."* and this was supported by the significantly lower cognitive workload ($p < .05$) while using the suggested

AR visual cues in shared control (Figure 6.4). We also found that ***video instruction and hands-on practice tend to provide sufficient information for the selection of AR visual cues in direct control without autonomy*** while experience and proficiency play a role in selecting suitable AR features when various autonomy is available. This is supported by the observation that participants' final preference for AR visual cues remained consistent with their initial selection when using direct control, but there were notable differences when using assisted, shared, and supervisory control.

Chapter 7

Conclusion and Future Scope

Building on the findings presented in Chapters 3, 4, 5, and 6 the research highlights the substantial advantages of action and perception assistance in enhancing both user preference and performance.

7.1 Conclusions

This dissertation presents systematically identifying and evaluating design ideologies for action and perception assistance that help improve remote manipulation experience for the teleoperator. The primary contributions presented in this document are:

Preferred teleoperation interface for nursing applications – In Chapter 2, we systematically compared and evaluated three representative interfaces across the spectrum of teleoperation control—from point-based to free-form control. These interfaces were assessed through a general-purpose workspace cleaning task conducted with nursing participants. Our findings indicate that motion mapping interfaces were the most intuitive and yielded the best performance. However, these interfaces also led to increased physical workload and reduced control efficiency due to limitations in precision.

Design ideologies for action assistance – Chapter 3 explores various forms of action assistance aimed at improving the precision and ergonomics of motion mapping interfaces. We found that separating orientation and position control significantly enhanced teleoperation performance and reduced physical exertion, with a purely motion-mapped interface for DOF separation performing best. Furthermore, motion scaling based on environmental constraints improved control precision and task performance in an intuitive and predictable manner. These findings were extended to bi-manual manipulation in Chapter 4, where position support (via motion scaling) and orientation support (for streamlining rotational control) were shown to improve performance and reduce operator workload when implemented correctly. Improperly triggered assistance, however, led to user frustration, which was mitigated through an assist-as-needed strategy. Assistance based on operator-specific metrics such as physical activity or inferred motion intent yielded the most effective results. These design principles can be generalized to inform the development of action assistance for both single-arm and bi-manual teleoperation tasks.

Design ideologies for perception assistance – In Chapter 5, we introduced a shared human-robot control paradigm that automates physically demanding actions. This interface reduced workload and improved user preference. However, user feedback revealed that increased autonomy introduced ambiguity about the robot's actions and the current state of autonomy. To address this, we studied the role of perception assistance using Augmented Reality (AR) cues, which were tailored to the operator's level of involvement in teleoperation. Chapter 6 presents a study examining operator preferences for AR cues under different control autonomy conditions, as well as how training influence operator preferences. Notably, preferences tended to converge toward developer-recommended designs with increased exposure. These findings offer critical insights for designing perception assistance that aligns with both user needs and developer goals.

7.2 Limitations

This research presents valuable insights into the design of teleoperation and AR-based healthcare interfaces; however, there are important limitations to consider. The studies involved a relatively small sample size, with fewer than 30 participants, which limits the ability to draw statistically conclusive results. Despite this, the trends and observations gathered throughout the research offer a meaningful foundation for making informed design decisions in future work and can guide the iterative development of user interfaces in similar contexts.

The design guidelines and interface evaluations were developed based on a specific set of teleoperation tasks. These tasks focused primarily on core aspects such as rotation and orientation control, information gathering, and basic manipulation. While representative of essential teleoperation capabilities, this scope does not extend to more complex scenarios, including in-hand manipulation, work with highly deformable or delicate objects, or tasks requiring fine-grained dexterity. These limitations highlight the need for continued research to address more intricate task requirements and to validate the applicability of the proposed design strategies in broader contexts.

Additionally, the AR interface studies were conducted under the assumption that participants had prior experience with video games and interactive media, which likely contributed to a higher level of spatial understanding and familiarity with augmented reality features. While this provided useful insight into AR usability among digitally experienced users, it may not accurately reflect the preferences and performance of individuals with less spatial awareness or limited exposure to such media. As a result, further studies are needed to understand how users with different levels of spatial awareness and technological proficiency interact with AR interfaces and whether alternative forms of information, like haptic or auditory feedback might be preferred in those cases.

7.3 Broader Impacts of Research

This research contributes meaningfully to the development of intuitive, efficient, and user-centered teleoperation systems, with particular relevance to nursing and broader healthcare contexts. The design and evaluation of novel control interfaces, action assistance strategies, and perception aids have the potential to significantly improve the accessibility and effectiveness of remote robotic systems in real-world clinical and care environments.

One of the central societal impacts of this work lies in enhancing the capabilities and reach of healthcare providers, especially nurses. By enabling remote physical interaction through teleoperated robotic systems, this research supports the growing need for decentralized and scalable healthcare delivery. For example, in rural or underserved areas where trained personnel may be limited, nurses could remotely perform essential tasks such as workspace setup, equipment handling, or patient assistance through intuitive robotic interfaces without being physically present. Currently, only 12% of the rural population get hospitalized treatment in rural hospitals [182]. This ability to extend the healthcare workers' expertise via teleoperation not only improves access to care but also allows for faster response times and better resource utilization.

Moreover, the developed techniques—such as environment-aware motion scaling, intuitive DOF separation, and assist-as-needed strategies—help reduce the physical and cognitive workload associated with prolonged robot teleoperation. This is particularly crucial in healthcare, where professionals are often required to manage high-stress environments and multitask continuously [183]. By lowering the fatigue and learning curve associated with robotic interfaces, the solutions proposed in this work can help reduce burnout and improve overall job satisfaction and safety.

Importantly, this research incorporates feedback and involvement from real healthcare workers, ensuring that the proposed systems are grounded in practical needs and constraints. This user-driven development approach enhances the likelihood of real-world adoption and impact. Additionally, the insights gained on user training, preference convergence, and perception support in autonomous systems help lay the foundation for broader human-robot collaboration in complex, safety-critical tasks.

In summary, this work aims not only to advance the field of human-robot interaction and teleoperation but also to make tangible contributions to improving healthcare delivery, workforce sustainability, and patient outcomes through thoughtful, human-centered robotic interface design.

7.4 Future Scope

Online cognitive workload estimation – We hypothesize that for perception assistance utilization of gaze based workload estimations can identify the users' intended operation and utilization on on screen perception assistance. We also hypothesize that removing perception assistance not being used by the operator helps reduce visual clutter and thus improve cognitive workload while teleoperating. Workload induced by stress can be estimated by monitoring changes in pupil dilation during teleoperation. An increase in pupil dilation correlates with heightened cognitive workload, influenced by physical and visual factors such as visual clutter in perception feedback. Online cognitive workload is quantified as the fraction of time pupil dilation increases relative to the total task completion time (see Figure 7.1) [184]. This workload can serve as an objective way of identifying cognitive workload while using a visual interface for teleoperation. Concentration workload, representing the cognitive effort needed for task processing, is defined as the ratio of time spent fixated in an area of interest to the total task operation time. This is grounded in the observation that individuals tend to focus their gaze on task-relevant areas [185]. Interface complexity is gauged by gaze motion, calculated as the ratio of the major axis of the ellipse formed by recent gaze points to the diagonal length of the screen displaying the perception assistance. The size and location of this ellipse help estimate the utilization of specific augmented reality features, allowing for their omission from the screen when not in use. This online realtime cognitive workload estimation can address the availability of perception assistance and alleviate mental fatigue as described in the following section.

Gaze based assist-as-needed interface – Elevated concentration workload, in tandem with the

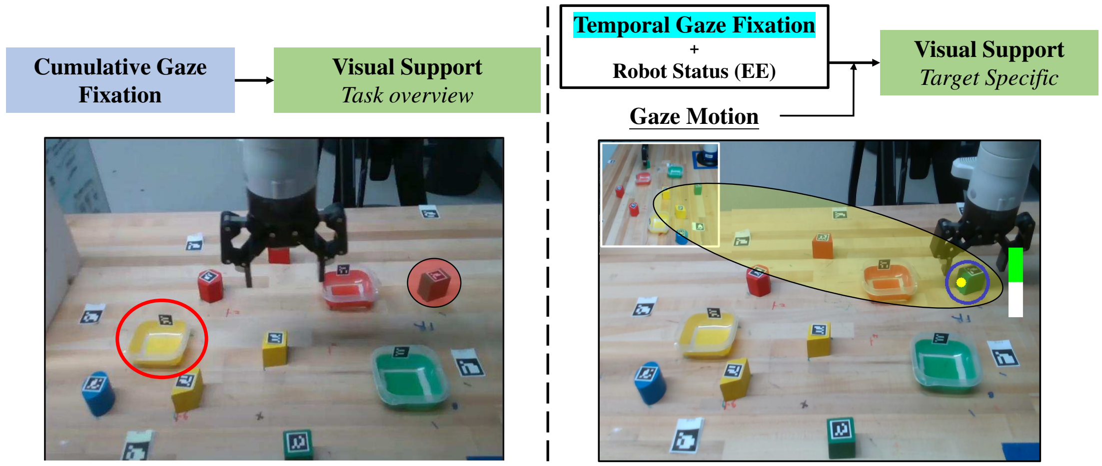


Figure 7.1: **Left:** Cumulative gaze fixation over the task duration reveals the Augmented Reality (AR) view providing information about the task overview, including the target object and its location. **Right:** Temporal fixation (current area of fixation) in conjunction with the proximity of the robot end-effector provides task-specific information, such as grasping support and a secondary view for depth perception. The yellow ellipse highlights the region of the most recent gaze fixations, serving as an indicator of whether the secondary view is in use and, by extension, whether it is required.

robot state, serves as a valuable indicator for determining the specific visual assistance required. For instance, when there is a high concentration workload related to a specific object and the robot end-effector is distant from that object, it suggests that the operator may prefer a task overview displaying the sequence of actions associated with that object. Conversely, if the concentration workload is high and the robot end-effector is in proximity to the same object, it indicates a preference for performing object-specific actions such as picking or placing. Furthermore, interface complexity workload aids in identifying visual features within the perception feedback. For example as seen in Figure 7.1, if the operator exhibits minimal gaze motion and their gaze does not fixate on a distant feature (e.g., a secondary view for grasping assistance), it suggests that this feature can be removed to alleviate visual overload. These interfaces can thus complement the observations from the study presented in Chapter 4 where assist-as-needed interfaces helped improve performance, reduce workload and enhanced the quality of teleoperation.

User Study Population and Generalizable Design Ideologies – While this dissertation presents comprehensive evaluations of teleoperation interfaces and assistance across various tasks, the user studies conducted were limited in scale and demographic diversity. Future work should focus on expanding the participant pool—particularly including a larger and more representative sample of healthcare professionals. This includes involving older healthcare workers, who make up a significant portion of the nursing workforce and may have different ergonomic needs, preferences, and interactions with robotic systems compared to younger or more tech-savvy users. Their inclusion is critical to ensure that the developed systems are accessible, intuitive, and effective across a broader spectrum of end-users.

Additionally, while the current studies focused on general-purpose manipulation and common nursing-related tasks, future evaluations should explore more complex, multi-step tasks that better represent the real-world demands of clinical environments. This will help validate and refine the proposed control and assistance strategies under conditions that more closely mirror actual nursing workflows. Testing under increased task complexity and varied environments will further ensure the robustness and generalization of these interface design ideologies. Broadening both the participant demographics and task scope is essential for developing truly deployable and impactful teleoperation systems that address the nuanced challenges of modern healthcare delivery.

Bibliography

[1] T.-C. Lin, A. U. Krishnan, and Z. Li, "Intuitive, efficient and ergonomic tele-nursing robot interfaces: Design evaluation and evolution," *Accepted by the Transactions on Human-Robot Interaction*, 2021.

[2] Y. Wang, P. Praveena, and M. Gleicher, "A design space of control coordinate systems in telemanipulation," *IEEE Access*, 2024.

[3] D. Tolani and N. I. Badler, "Real-time inverse kinematics of the human arm," *Presence: Teleoperators & Virtual Environments*, vol. 5, no. 4, pp. 393–401, 1996.

[4] I. B. Wijayasinghe, M. N. Saadatzi, S. Peetha, D. O. Popa, and S. Cremer, "Adaptive interface for robot teleoperation using a genetic algorithm," in *2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)*, pp. 50–56, IEEE, 2018.

[5] S. Li, M. Bowman, H. Nobarani, and X. Zhang, "Inference of manipulation intent in teleoperation for robotic assistance," *Journal of Intelligent & Robotic Systems*, vol. 99, no. 1, pp. 29–43, 2020.

[6] J. Y. Chen, E. C. Haas, and 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)*, vol. 37, no. 6, pp. 1231–1245, 2007.

[7] B. S. Peters, P. R. Armijo, C. Krause, S. A. Choudhury, and D. Oleynikov, "Review of emerging surgical robotic technology," *Surgical endoscopy*, vol. 32, no. 4, pp. 1636–1655, 2018.

[8] Z. Li, P. Moran, Q. Dong, R. J. Shaw, and K. Hauser, "Development of a tele-nursing mobile manipulator for remote care-giving in quarantine areas," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3581–3586, IEEE, 2017.

[9] M. Wrock and S. B. Nokleby, "Decoupled teleoperation of a holonomic mobile-manipulator system using automatic switching," in *2011 24th Canadian Conference on Electrical and Computer Engineering (CCECE)*, pp. 001164–001168, IEEE, 2011.

[10] D. Rakita, B. Mutlu, and M. Gleicher, "An autonomous dynamic camera method for effective remote teleoperation," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 325–333, 2018.

[11] K. Hashimoto, F. Saito, T. Yamamoto, and K. Ikeda, "A field study of the human support robot in the home environment," in *2013 IEEE Workshop on Advanced Robotics and its Social Impacts*, pp. 143–150, IEEE, 2013.

[12] F. Abi-Farrajl, B. Henze, A. Werner, M. Panzirsch, C. Ott, and M. A. Roa, "Humanoid teleoperation using task-relevant haptic feedback," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5010–5017, IEEE, 2018.

[13] J. Oh, O. Sim, H. Jeong, and J.-H. Oh, "Humanoid whole-body remote-control framework with delayed reference generator for imitating human motion," *Mechatronics*, vol. 62, p. 102253, 2019.

[14] E.-J. Rolley-Parnell, D. Kanoulas, A. Laurenzi, B. Delhaisse, L. Rozo, D. G. Caldwell, and N. G. Tsagarakis, "Bi-manual articulated robot teleoperation using an external rgb-d range sensor," in *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pp. 298–304, IEEE, 2018.

[15] K. Yokoi, K. Nakashima, M. Kobayashi, H. Mihune, H. Hasunuma, Y. Yanagihara, T. Ueno,

T. Gokyuu, and K. Endou, “A tele-operated humanoid operator,” *The International Journal of Robotics Research*, vol. 25, no. 5-6, pp. 593–602, 2006.

[16] M. A. Goodrich, J. W. Crandall, and E. Barakova, “Teleoperation and beyond for assistive humanoid robots,” *Reviews of Human factors and ergonomics*, vol. 9, no. 1, pp. 175–226, 2013.

[17] J. B. Van Erp and P. Padmos, “Image parameters for driving with indirect viewing systems,” *Ergonomics*, vol. 46, no. 15, pp. 1471–1499, 2003.

[18] M. Voshell, D. D. Woods, and F. Phillips, “Overcoming the keyhole in human-robot coordination: simulation and evaluation,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 49, pp. 442–446, Sage Publications Sage CA: Los Angeles, CA, 2005.

[19] A. K. Kanduri, G. Thomas, N. Cabrol, E. Grin, and R. C. Anderson, “The (in) accuracy of novice rover operators’ perception of obstacle height from monoscopic images,” *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 35, no. 4, pp. 505–512, 2005.

[20] E. Triantafyllidis, C. Mcgreavy, J. Gu, and Z. Li, “Study of multimodal interfaces and the improvements on teleoperation,” *IEEE Access*, vol. 8, pp. 78213–78227, 2020.

[21] C. Cao, E. Danahy, and J. Noble, “Educational robotics for teleoperated general exam,” in *2012 IEEE International Conference on Technologies for Practical Robot Applications (TePRA)*, pp. 59–66, IEEE, 2012.

[22] T. L. Chen and C. C. Kemp, “Lead me by the hand: Evaluation of a direct physical interface for nursing assistant robots,” in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 367–374, IEEE, 2010.

[23] G. Randelli, M. Venanzi, and D. Nardi, “Evaluating tangible paradigms for ground robot teleoperation,” in *2011 RO-MAN*, pp. 389–394, IEEE, 2011.

[24] T.-C. Lin, A. U. Krishnan, and Z. Li, “Intuitive, efficient and ergonomic tele-nursing robot interfaces: Design evaluation and evolution,” *ACM Transactions on Human-Robot Interaction*, 2022.

[25] A. U. Krishnan, T.-C. Lin, and Z. Li, “Design interface mapping for efficient free-form tele-manipulation,” in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6221–6226, IEEE, 2022.

[26] A. Unni Krishnan and Z. Li, “Towards intuitive and efficient bi-manual remote manipulation: Adaptive assistance to improve performance and reduce physical workload,” *Submitted to IEEE Robotics and Automation Letters*, Apr. 2025.

[27] Y. Zhu, A. Smith, and K. Hauser, “Automated heart and lung auscultation in robotic physical examinations,” *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4204–4211, 2022.

[28] W.-H. Zhu and S. E. Salcudean, “Stability guaranteed teleoperation: an adaptive motion/-force control approach,” *IEEE transactions on automatic control*, vol. 45, no. 11, pp. 1951–1969, 2000.

[29] Y. Fan, C. Yang, and X. Wu, “Improved teleoperation of an industrial robot arm system using leap motion and myo armband,” in *2019 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 1670–1675, IEEE, 2019.

[30] T. Zhou, M. E. Cabrera, J. P. Wachs, T. Low, and C. Sundaram, “A comparative study for telerobotic surgery using free hand gestures,” *Journal of Human-Robot Interaction*, vol. 5, no. 2, pp. 1–28, 2016.

[31] S. Frees, G. D. Kessler, and E. Kay, "Prism interaction for enhancing control in immersive virtual environments," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 14, no. 1, pp. 2–es, 2007.

[32] D. Mendes, F. M. Caputo, A. Giachetti, A. Ferreira, and J. Jorge, "A survey on 3d virtual object manipulation: From the desktop to immersive virtual environments," in *Computer graphics forum*, vol. 38, pp. 21–45, Wiley Online Library, 2019.

[33] W. Pryor, B. P. Vagvolgyi, A. Deguet, S. Leonard, L. L. Whitcomb, and P. Kazanzides, "Interactive planning and supervised execution for high-risk, high-latency teleoperation," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1857–1864, IEEE, 2020.

[34] P. C. Gloumeau, W. Stuerzlinger, and J. Han, "Pinnpivot: Object manipulation using pins in immersive virtual environments," *IEEE transactions on visualization and computer graphics*, vol. 27, no. 4, pp. 2488–2494, 2020.

[35] F. Argelaguet and C. Andujar, "A survey of 3d object selection techniques for virtual environments," *Computers & Graphics*, vol. 37, no. 3, pp. 121–136, 2013.

[36] Y. Huang, E. Burdet, L. Cao, P. T. Phan, A. M. H. Tiong, and S. J. Phee, "A subject-specific four-degree-of-freedom foot interface to control a surgical robot," *IEEE/ASME Transactions on Mechatronics*, vol. 25, no. 2, pp. 951–963, 2020.

[37] J. Funda, R. H. Taylor, B. Eldridge, S. Gomory, and K. G. Gruben, "Constrained cartesian motion control for teleoperated surgical robots," *IEEE Transactions on Robotics and Automation*, vol. 12, no. 3, pp. 453–465, 1996.

[38] A. U. Krishnan, T.-C. Lin, and Z. Li, "Human preferred augmented reality visual cues for remote robot manipulation assistance: from direct to supervisory control," in *2023 IEEE/RSJ*

International Conference on Intelligent Robots and Systems (IROS), pp. 7034–7039, IEEE, 2023.

[39] M. Walker, T. Phung, T. Chakraborti, T. Williams, and D. Szafir, "Virtual, augmented, and mixed reality for human-robot interaction: A survey and virtual design element taxonomy," *arXiv preprint arXiv:2202.11249*, 2022.

[40] C. P. Quintero, O. Ramirez, and M. Jägersand, "Vibi: Assistive vision-based interface for robot manipulation," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4458–4463, IEEE, 2015.

[41] D. Krupke, L. Einig, E. Langbehn, J. Zhang, and F. Steinicke, "Immersive remote grasping: realtime gripper control by a heterogenous robot control system," in *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pp. 337–338, 2016.

[42] J. I. Lipton, A. J. Fay, and D. Rus, "Baxter's homunculus: Virtual reality spaces for teleoperation in manufacturing," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 179–186, 2017.

[43] E. Rosen, D. Whitney, E. Phillips, G. Chien, J. Tompkin, G. Konidaris, and S. Tellex, "Communicating robot arm motion intent through mixed reality head-mounted displays," in *Robotics Research: The 18th International Symposium ISRR*, pp. 301–316, Springer, 2020.

[44] T.-C. Lin, A. U. Krishnan, and Z. Li, "Comparison of haptic and augmented reality visual cues for assisting tele-manipulation," in *2022 International Conference on Robotics and Automation (ICRA)*, pp. 9309–9316, IEEE, 2022.

[45] W. P. Chan, C. P. Quintero, M. K. Pan, M. Sakr, H. M. Van der Loos, and E. Croft, "A multimodal system using augmented reality, gestures, and tactile feedback for robot trajectory programming and execution," in *Virtual Reality*, pp. 142–158, River Publishers, 2022.

[46] T.-C. Lin, A. U. Krishnan, and Z. Li, "Shared autonomous interface for reducing physical effort in robot teleoperation via human motion mapping," in *to appear in the 2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020.

[47] J. M. Beer, A. D. Fisk, and W. A. Rogers, "Toward a framework for levels of robot autonomy in human-robot interaction," *Journal of human-robot interaction*, vol. 3, no. 2, pp. 74–99, 2014.

[48] R. Ebad and K. Jazan, "Telemedicine: Current and future perspectives telemedicine: Current and future perspectives," 2013.

[49] A. J. Kucharski, A. Camacho, S. Flasche, R. E. Glover, W. J. Edmunds, and S. Funk, "Measuring the impact of ebola control measures in sierra leone," *Proceedings of the National Academy of Sciences*, vol. 112, no. 46, pp. 14366–14371, 2015.

[50] I. Rodriguez-Barraquer, F. Costa, E. J. Nascimento, N. Nery, P. M. Castanha, G. A. Sacramento, J. Cruz, M. Carvalho, D. De Olivera, J. E. Hagan, *et al.*, "Impact of preexisting dengue immunity on zika virus emergence in a dengue endemic region," *Science*, vol. 363, no. 6427, pp. 607–610, 2019.

[51] V. J. Munster, M. Koopmans, N. van Doremalen, D. van Riel, and E. de Wit, "A novel coronavirus emerging in china—key questions for impact assessment," *New England Journal of Medicine*, 2020.

[52] G.-Z. Yang, B. J. Nelson, R. R. Murphy, H. Choset, H. Christensen, S. H. Collins, P. Dario, K. Goldberg, K. Ikuta, N. Jacobstein, *et al.*, "Combating covid-19—the role of robotics in managing public health and infectious diseases," 2020.

[53] E. Ackerman, "irobot and intouch health announce rp-vita telemedecine robot," *IEEE Spectrum*, 2012.

[54] E. Ackerman, "Ava robotics introduces autonomous telepresence robot," *IEEE Spectrum*, 2018.

[55] E. Ackerman, "Suitable tech introduces beampro 2 telepresence platform," *IEEE Spectrum*, 2018.

[56] K. Tsui, A. Norton, D. Brooks, H. Yanco, and D. Kontak, "Designing telepresence robot systems for use by people with special needs," in *Int. Symposium on Quality of Life Technologies: Intelligent Systems for Better Living*, 2011.

[57] E. Ackerman, "Moxi prototype from diligent robotics starts helping out in hospitals," *IEEE Spectrum. https://spectrum. ieee. org/automaton/robotics/industrial-robots/moxi-prototype-fro m-diligent-robotics-starts-helping-out-in-hospitals*, 2018.

[58] E. Ackerman, "Toyota gets back into humanoid robots with new t-hr3," *IEEE Spectrum*, 2017.

[59] J. Bodner, H. Wykypiel, G. Wetscher, and T. Schmid, "First experiences with the da vinci™ operating robot in thoracic surgery," *European Journal of Cardio-thoracic surgery*, vol. 25, no. 5, pp. 844–851, 2004.

[60] D. Kent, C. Saldanha, and S. Chernova, "A comparison of remote robot teleoperation interfaces for general object manipulation," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 371–379, 2017.

[61] P. Gliesche, T. Krick, M. Pfingsthorn, S. Drolshagen, C. Kowalski, and A. Hein, "Kinesthetic device vs. keyboard/mouse: a comparison in home care telemanipulation," *Frontiers in Robotics and AI*, vol. 7, p. 561015, 2020.

[62] K. Yaovaja and J. Klunngien, "Teleoperation of an industrial robot using a non-standalone

5g mobile network," in *2019 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE)*, pp. 320–323, IEEE, 2019.

[63] S. Bier, R. Li, and W. Wang, "A full-dimensional robot teleoperation platform," in *2020 11th International Conference on Mechanical and Aerospace Engineering (ICMAE)*, pp. 186–191, IEEE, 2020.

[64] D. Rakita, B. Mutlu, and M. Gleicher, "A motion retargeting method for effective mimicry-based teleoperation of robot arms," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 361–370, 2017.

[65] D. Q. Truong, B. N. M. Truong, N. T. Trung, S. A. Nahian, and K. K. Ahn, "Force reflecting joystick control for applications to bilateral teleoperation in construction machinery," *International Journal of Precision Engineering and Manufacturing*, vol. 18, pp. 301–315, 2017.

[66] L. Fritsche, F. Unverzag, J. Peters, and R. Calandra, "First-person tele-operation of a humanoid robot," in *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*, pp. 997–1002, IEEE, 2015.

[67] G. Du, P. Zhang, and X. Liu, "Markerless human-manipulator interface using leap motion with interval kalman filter and improved particle filter," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 2, pp. 694–704, 2016.

[68] O. Porges, M. Connan, B. Henze, A. Gigli, C. Castellini, and M. A. Roa, "A wearable, ultralight interface for bimanual teleoperation of a compliant, whole-body-controlled humanoid robot," in *Proceedings of ICRA-International Conference on Robotics and Automation*, 2019.

[69] E. Rosen, D. Whitney, E. Phillips, G. Chien, J. Tompkin, G. Konidaris, and S. Tellex, "Communicating and controlling robot arm motion intent through mixed-reality head-mounted displays," *The International Journal of Robotics Research*, vol. 38, no. 12-13, pp. 1513–1526, 2019.

[70] A. Cela, J. J. Yebes, R. Arroyo, L. M. Bergasa, R. Barea, and E. López, "Complete low-cost implementation of a teleoperated control system for a humanoid robot," *Sensors*, vol. 13, no. 2, pp. 1385–1401, 2013.

[71] J. Oh, I. Lee, H. Jeong, and J.-H. Oh, "Real-time humanoid whole-body remote control framework for imitating human motion based on kinematic mapping and motion constraints," *Advanced Robotics*, vol. 33, no. 6, pp. 293–305, 2019.

[72] J. Liu and Y. Zhang, "Mapping human hand motion to dexterous robotic hand," in *2007 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 829–834, IEEE, 2007.

[73] K. Fischer, F. Kirstein, L. C. Jensen, N. Krüger, K. Kukliński, M. V. aus der Wieschen, and T. R. Savarimuthu, "A comparison of types of robot control for programming by demonstration," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 213–220, IEEE, 2016.

[74] N. Miller, O. C. Jenkins, M. Kallmann, and M. J. Mataric, "Motion capture from inertial sensing for untethered humanoid teleoperation," in *4th IEEE/RAS International Conference on Humanoid Robots, 2004.*, vol. 2, pp. 547–565, IEEE, 2004.

[75] K. Yamane and J. Hodgins, "Simultaneous tracking and balancing of humanoid robots for imitating human motion capture data," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2510–2517, IEEE, 2009.

[76] B. Dariush, M. Gienger, A. Arumbakkam, Y. Zhu, B. Jian, K. Fujimura, and C. Goerick, “Online transfer of human motion to humanoids,” *International Journal of Humanoid Robotics*, vol. 6, no. 02, pp. 265–289, 2009.

[77] J. Koenemann and M. Bennewitz, “Whole-body imitation of human motions with a nao humanoid,” in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pp. 425–426, ACM, 2012.

[78] G. Du, P. Zhang, J. Mai, and Z. Li, “Markerless kinect-based hand tracking for robot teleoperation,” *International Journal of Advanced Robotic Systems*, vol. 9, no. 2, p. 36, 2012.

[79] J. Koenemann, F. Burget, and M. Bennewitz, “Real-time imitation of human whole-body motions by humanoids,” in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2806–2812, IEEE, 2014.

[80] L. Penco, B. Clément, V. Modugno, E. M. Hoffman, G. Nava, D. Pucci, N. G. Tsagarakis, J.-B. Mouret, and S. Ivaldi, “Robust real-time whole-body motion retargeting from human to humanoid,” in *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*, pp. 425–432, IEEE, 2018.

[81] R. Elbasiony and W. Gomaa, “Humanoids skill learning based on real-time human motion imitation using kinect,” *Intelligent Service Robotics*, vol. 11, no. 2, pp. 149–169, 2018.

[82] C.-L. Hwang and G.-H. Liao, “Real-time pose imitation by mid-size humanoid robot with servo-cradle-head rgb-d vision system,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 1, pp. 181–191, 2018.

[83] B. Fang, F. Sun, H. Liu, D. Guo, W. Chen, and G. Yao, “Robotic teleoperation systems using a wearable multimodal fusion device,” *International Journal of advanced robotic systems*, vol. 14, no. 4, p. 1729881417717057, 2017.

[84] T.-C. Lin, A. U. Krishnan, and Z. Li, "Physical fatigue analysis of assistive robot teleoperation via whole-body motion mapping," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2240–2245, IEEE, 2019.

[85] M. K. Kim, K. Ryu, Y. Oh, S.-R. Oh, and K. Kim, "Implementation of real-time motion and force capturing system for tele-manipulation based on semg signals and imu motion data," pp. 5658–5664, IEEE, 2014.

[86] S. Tortora, M. Moro, and E. Menegatti, "Dual-myo real-time control of a humanoid arm for teleoperation," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 624–625, IEEE, 2019.

[87] Y. Chae, J. Jeong, and S. Jo, "Toward brain-actuated humanoid robots: asynchronous direct control using an eeg-based bci," *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1131–1144, 2012.

[88] K. Muelling, A. Venkatraman, J.-S. Valois, J. E. Downey, J. Weiss, S. Javdani, M. Hebert, A. B. Schwartz, J. L. Collinger, and J. A. Bagnell, "Autonomy infused teleoperation with application to brain computer interface controlled manipulation," *Autonomous Robots*, vol. 41, no. 6, pp. 1401–1422, 2017.

[89] N. K. N. Aznan, J. D. Connolly, N. Al Moubayed, and T. P. Breckon, "Using variable natural environment brain-computer interface stimuli for real-time humanoid robot navigation," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 4889–4895, IEEE, 2019.

[90] S. Chernova and A. L. Thomaz, "Robot learning from human teachers," vol. 8, ch. 10, pp. 1–121, Morgan & Claypool Publishers, 2014.

[91] D. Mendes, F. Relvas, A. Ferreira, and J. Jorge, "The benefits of dof separation in mid-air 3d

object manipulation," in *Proceedings of the 22nd ACM conference on virtual reality software and technology*, pp. 261–268, 2016.

[92] N. Pedemonte, F. Abi-Farraj, and P. R. Giordano, "Visual-based shared control for remote telemanipulation with integral haptic feedback," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5342–5349, IEEE, 2017.

[93] M. A. R. Garcia, R. A. Rojas, and F. Pirri, "Object-centered teleoperation of mobile manipulators with remote center of motion constraint," *IEEE robotics and automation letters*, vol. 4, no. 2, pp. 1745–1752, 2019.

[94] I. Cho and Z. Wartell, "Evaluation of a bimanual simultaneous 7dof interaction technique in virtual environments," in *2015 IEEE symposium on 3D User Interfaces (3DUI)*, pp. 133–136, IEEE, 2015.

[95] F. Richter, R. K. Orosco, and M. C. Yip, "Motion scaling solutions for improved performance in high delay surgical teleoperation," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 1590–1595, IEEE, 2019.

[96] Y. Cho and F. L. Hammond, "Improving efficiency and safety in teleoperated robotic manipulators using motion scaling and force feedback," in *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 1236–1242, IEEE, 2020.

[97] J. E. Solanes, A. Muñoz, 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, no. 3, pp. 1077–1097, 2020.

[98] A. Handa, K. Van Wyk, W. Yang, J. Liang, Y.-W. Chao, Q. Wan, S. Birchfield, N. Ratliff, and D. Fox, "Dexpilot: Vision-based teleoperation of dexterous robotic hand-arm system,"

in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 9164–9170, IEEE, 2020.

[99] S. Li, J. Jiang, P. Ruppel, H. Liang, X. Ma, N. Hendrich, F. Sun, and J. Zhang, "A mobile robot hand-arm teleoperation system by vision and imu," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 10900–10906, IEEE, 2020.

[100] C. Meeker, T. Rasmussen, and M. Ciocarlie, "Intuitive hand teleoperation by novice operators using a continuous teleoperation subspace," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5821–5827, IEEE, 2018.

[101] M. Möllers, P. Zimmer, and J. Borchers, "Direct manipulation and the third dimension: coplanar dragging on 3d displays," in *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*, pp. 11–20, 2012.

[102] F. Daiber, E. Falk, and A. Krüger, "Balloon selection revisited: multi-touch selection techniques for stereoscopic data," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pp. 441–444, 2012.

[103] B. Fang, F. Sun, H. Liu, and D. Guo, "A novel data glove using inertial and magnetic sensors for motion capture and robotic arm-hand teleoperation," *Industrial Robot: An International Journal*, 2017.

[104] H. Zhang, Z. Zhao, Y. Yu, K. Gui, X. Sheng, and X. Zhu, "A feasibility study on an intuitive teleoperation system combining imu with semg sensors," in *International Conference on Intelligent Robotics and Applications*, pp. 465–474, Springer, 2018.

[105] T. Zhang, Z. McCarthy, O. Jow, D. Lee, X. Chen, K. Goldberg, and P. Abbeel, "Deep imitation learning for complex manipulation tasks from virtual reality teleoperation," in *2018*

IEEE International Conference on Robotics and Automation (ICRA), pp. 5628–5635, IEEE, 2018.

[106] G. Gorjup, A. Dwivedi, N. Elangovan, and M. Liarokapis, "An intuitive, affordances oriented telemanipulation framework for a dual robot arm hand system: On the execution of bimanual tasks," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3611–3616, IEEE, 2019.

[107] Z. Zhang, Y. Niu, Z. Yan, and S. Lin, "Real-time whole-body imitation by humanoid robots and task-oriented teleoperation using an analytical mapping method and quantitative evaluation," *Applied Sciences*, vol. 8, no. 10, p. 2005, 2018.

[108] K. Darvish, Y. Tirupachuri, G. Romualdi, L. Rapetti, D. Ferigo, F. J. A. Chavez, and D. Pucci, "Whole-body geometric retargeting for humanoid robots," in *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*, pp. 679–686, IEEE, 2019.

[109] C. P. Quintero, M. Dehghan, O. Ramirez, M. H. Ang, and M. Jagersand, "Flexible virtual fixture interface for path specification in tele-manipulation," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5363–5368, IEEE, 2017.

[110] D. Lee and Y. S. Park, "Implementation of augmented teleoperation system based on robot operating system (ros)," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5497–5502, IEEE, 2018.

[111] M. M. Marinho, B. V. Adorno, K. Harada, K. Deie, A. Deguet, P. Kazanzides, R. H. Taylor, and M. Mitsuishi, "A unified framework for the teleoperation of surgical robots in constrained workspaces," in *2019 international conference on robotics and automation (ICRA)*, pp. 2721–2727, IEEE, 2019.

[112] M. Minelli, F. Ferraguti, N. Piccinelli, R. Muradore, and C. Secchi, “An energy-shared two-layer approach for multi-master-multi-slave bilateral teleoperation systems,” in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 423–429, IEEE, 2019.

[113] M. Rubagotti, T. Taunyazov, B. Omarali, and A. Shintemirov, “Semi-autonomous robot teleoperation with obstacle avoidance via model predictive control,” *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2746–2753, 2019.

[114] S. Javdani, H. Admoni, S. Pellegrinelli, S. S. Srinivasa, and J. A. Bagnell, “Shared autonomy via hindsight optimization for teleoperation and teaming,” *The International Journal of Robotics Research*, vol. 37, no. 7, pp. 717–742, 2018.

[115] C. Schultz, S. Gaurav, M. Monfort, L. Zhang, and B. D. Ziebart, “Goal-predictive robotic teleoperation from noisy sensors,” in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5377–5383, IEEE, 2017.

[116] D. Krupke, F. Steinicke, P. Lubos, Y. Jonetzko, M. Görner, and J. Zhang, “Comparison of multimodal heading and pointing gestures for co-located mixed reality human-robot interaction,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1–9, IEEE, 2018.

[117] W. Zhang, H. Cheng, L. Zhao, L. Hao, M. Tao, and C. Xiang, “A gesture-based teleoperation system for compliant robot motion,” *Applied Sciences*, vol. 9, no. 24, p. 5290, 2019.

[118] A. E. Leeper, K. Hsiao, M. Ciocarlie, L. Takayama, and D. Gossow, “Strategies for human-in-the-loop robotic grasping,” in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pp. 1–8, 2012.

[119] D. Kent, M. Behrooz, and S. Chernova, “Construction of a 3d object recognition and manip-

ulation database from grasp demonstrations," *Autonomous Robots*, vol. 40, no. 1, pp. 175–192, 2016.

[120] R. Toris, J. Kammerl, D. V. Lu, J. Lee, O. C. Jenkins, S. Osentoski, M. Wills, and S. Chernova, "Robot web tools: Efficient messaging for cloud robotics," in *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp. 4530–4537, IEEE, 2015.

[121] V. Annem, P. Rajendran, S. Thakar, and S. K. Gupta, "Towards remote teleoperation of a semi-autonomous mobile manipulator system in machine tending tasks," in *International Manufacturing Science and Engineering Conference*, vol. 58745, p. V001T02A027, American Society of Mechanical Engineers, 2019.

[122] P. Song, W. B. Goh, W. Hutama, C.-W. Fu, and X. Liu, "A handle bar metaphor for virtual object manipulation with mid-air interaction," in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1297–1306, 2012.

[123] B. Bossavit, A. Marzo, O. Ardaiz, L. D. De Cerio, and A. Pina, "Design choices and their implications for 3d mid-air manipulation techniques," *Presence: Teleoperators and Virtual Environments*, vol. 23, no. 4, pp. 377–392, 2014.

[124] J. Guo, C. Liu, and P. Poignet, "A scaled bilateral teleoperation system for robotic-assisted surgery with time delay," *Journal of Intelligent & Robotic Systems*, vol. 95, no. 1, pp. 165–192, 2019.

[125] N. Osawa, "Two-handed and one-handed techniques for precise and efficient manipulation in immersive virtual environments," in *International Symposium on Visual Computing*, pp. 987–997, Springer, 2008.

[126] R. V. Dubey, S. Everett, N. Pernalete, and K. A. Manocha, "Teleoperation assistance through

variable velocity mapping," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 5, pp. 761–766, 2001.

[127] D. G. Black, D. Andjelic, and S. E. Salcudean, "Evaluation of communication and human response latency for (human) teleoperation," *IEEE Transactions on Medical Robotics and Bionics*, vol. 6, no. 1, pp. 53–63, 2024.

[128] W. Tang, Y. Tian, L. Tan, S. Xue, T. Jiang, S. Zhou, and C. Wang, "Human factors design and evaluation of china's space manipulator teleoperation system," *International Journal of Human–Computer Interaction*, vol. 40, no. 8, pp. 1943–1959, 2024.

[129] P. Barba, J. Stramiello, E. K. Funk, F. Richter, M. C. Yip, and R. K. Orosco, "Remote telesurgery in humans: a systematic review," *Surgical endoscopy*, vol. 36, no. 5, pp. 2771–2777, 2022.

[130] G. Yang, H. Lv, Z. Zhang, L. Yang, J. Deng, S. You, J. Du, and H. Yang, "Keep healthcare workers safe: application of teleoperated robot in isolation ward for covid-19 prevention and control," *Chinese Journal of Mechanical Engineering*, vol. 33, pp. 1–4, 2020.

[131] J. Du, W. Vann, T. Zhou, Y. Ye, and Q. Zhu, "Sensory manipulation as a countermeasure to robot teleoperation delays: system and evidence," *Scientific Reports*, vol. 14, no. 1, p. 4333, 2024.

[132] H. Igarashi, A. Takeya, F. Harashima, and M. Kakikura, "Human adaptive assist planning for teleoperation," in *IECON 2006-32nd Annual Conference on IEEE Industrial Electronics*, pp. 4522–4527, IEEE, 2006.

[133] N. Small, K. Lee, and G. Mann, "An assigned responsibility system for robotic teleoperation control," *International journal of intelligent robotics and applications*, vol. 2, no. 1, pp. 81–97, 2018.

[134] Q. Gao, Z. Ju, Y. Chen, Q. Wang, and C. Chi, "An efficient rgb-d hand gesture detection framework for dexterous robot hand-arm teleoperation system," *IEEE Transactions on Human-Machine Systems*, vol. 53, no. 1, pp. 13–23, 2022.

[135] A. Toedtheide, X. Chen, H. Sadeghian, A. Naceri, and S. Haddadin, "A force-sensitive exoskeleton for teleoperation: An application in elderly care robotics," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 12624–12630, IEEE, 2023.

[136] B. R. Galarza, P. Ayala, S. Manzano, and M. V. Garcia, "Virtual reality teleoperation system for mobile robot manipulation," *Robotics*, vol. 12, no. 6, p. 163, 2023.

[137] F. Richter, E. K. Funk, W. S. Park, R. K. Orosco, and M. C. Yip, "From bench to bedside: The first live robotic surgery on the dvrk to enable remote telesurgery with motion scaling," in *2021 International Symposium on Medical Robotics (ISMR)*, pp. 1–7, IEEE, 2021.

[138] J. R. Boehm, N. P. Fey, and A. Majewicz, "Inherent kinematic features of dynamic bimanual path following tasks," *IEEE transactions on human-machine systems*, vol. 50, no. 6, pp. 613–622, 2020.

[139] M. Zhang, C. Sun, Y. Liu, and X. Wu, "A robotic system to deliver multiple physically bimanual tasks via varying force fields," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 688–698, 2022.

[140] C. Wilkes and D. A. Bowman, "Advantages of velocity-based scaling for distant 3d manipulation," in *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, pp. 23–29, 2008.

[141] A. García, J. E. Solanes, A. Muñoz, L. Gracia, and J. Tornero, "Augmented reality-based interface for bimanual robot teleoperation," *Applied Sciences*, vol. 12, no. 9, p. 4379, 2022.

[142] M. Mine, A. Yoganandan, and D. Coffey, “Principles, interactions and devices for real-world immersive modeling,” *Computers & Graphics*, vol. 48, pp. 84–98, 2015.

[143] A. Kapoor and R. H. Taylor, “A constrained optimization approach to virtual fixtures for multi-handed tasks,” in *2008 IEEE International Conference on Robotics and Automation*, pp. 3401–3406, IEEE, 2008.

[144] Y. Oh, T. Schäfer, B. Rüther, M. Toussaint, and J. Mainprice, “A system for traded control teleoperation of manipulation tasks using intent prediction from hand gestures,” in *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, pp. 503–508, IEEE, 2021.

[145] D. Sun, Q. Liao, and A. Loutfi, “Single master bimanual teleoperation system with efficient regulation,” *IEEE Transactions on Robotics*, vol. 36, no. 4, pp. 1022–1037, 2020.

[146] Y. Li, R. Cui, W. Yan, C. Feng, and S. Zhang, “Mitigating over-assistance in teleoperated mobile robots via human-centered shared autonomy: Leveraging suboptimal rationality insights,” *IEEE Robotics and Automation Letters*, 2024.

[147] S. Javdani, S. S. Srinivasa, and J. A. Bagnell, “Shared autonomy via hindsight optimization,” *Robotics science and systems: online proceedings*, vol. 2015, 2015.

[148] S. Lu, M. Y. Zhang, T. Ersal, and X. J. Yang, “Workload management in teleoperation of unmanned ground vehicles: Effects of a delay compensation aid on human operators’ workload and teleoperation performance,” *International Journal of Human–Computer Interaction*, vol. 35, no. 19, pp. 1820–1830, 2019.

[149] M. Selvaggio, S. Grazioso, G. Notomista, and F. Chen, “Towards a self-collision aware teleoperation framework for compound robots,” in *2017 IEEE World Haptics Conference (WHC)*, pp. 460–465, IEEE, 2017.

[150] T.-C. Lin, A. U. Krishnan, and Z. Li, “The impacts of unreliable autonomy in human-robot collaboration on shared and supervisory control for remote manipulation,” *IEEE Robotics and Automation Letters*, vol. 8, no. 8, pp. 4641–4648, 2023.

[151] S. Parsa, H. A. Maior, A. R. E. Thumwood, M. L. Wilson, M. Hanheide, and A. G. Esfahani, “The impact of motion scaling and haptic guidance on operators’ workload and performance in teleoperation,” in *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pp. 1–7, 2022.

[152] T.-C. Lin, A. U. Krishnan, and Z. Li, “Perception and action augmentation for teleoperation assistance in freeform telemanipulation,” *ACM Transactions on Human-Robot Interaction*, vol. 13, no. 1, pp. 1–40, 2024.

[153] G. M. Kontakis, K. Steriopoulos, J. Damilakis, and E. Michalodimitrakis, “The position of the axillary nerve in the deltoid muscle: A cadaveric study,” *Acta orthopaedica Scandinavica*, vol. 70, no. 1, pp. 9–11, 1999.

[154] V. Balasubramanian, K. Adalarasu, and R. Regulapati, “Comparing dynamic and stationary standing postures in an assembly task,” *International Journal of Industrial Ergonomics*, vol. 39, no. 5, pp. 649–654, 2009.

[155] H. Admoni and S. Srinivasa, “Predicting user intent through eye gaze for shared autonomy,” in *2016 AAAI Fall Symposium Series*, 2016.

[156] A. D. Dragan, S. S. Srinivasa, and K. C. Lee, “Teleoperation with intelligent and customizable interfaces,” *Journal of Human-Robot Interaction*, vol. 2, no. 2, pp. 33–57, 2013.

[157] D. Rakita, B. Mutlu, M. Gleicher, and L. M. Hiatt, “Shared control–based bimanual robot manipulation,” *Science Robotics*, vol. 4, no. 30, p. eaaw0955, 2019.

[158] M. Laghi, M. Maimeri, M. Marchand, C. Leparoux, M. Catalano, A. Ajoudani, and A. Bicchi, "Shared-autonomy control for intuitive bimanual tele-manipulation," in *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*, pp. 1–9, IEEE, 2018.

[159] C. E. Harriott, G. L. Buford, J. A. Adams, and T. Zhang, "Mental workload and task performance in peer-based human-robot teams," *Journal of Human-Robot Interaction*, vol. 4, no. 2, pp. 61–96, 2015.

[160] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Advances in psychology*, vol. 52, pp. 139–183, Elsevier, 1988.

[161] M. Castor, E. Hanson, E. Svensson, S. Nählinder, P. LeBlaye, I. MacLeod, N. Wright, J. Alfredson, L. Ågren, P. Berggren, *et al.*, "Garteur handbook of mental workload measurement," *GARTEUR, Group for Aeronautical Research and Technology in Europe, Flight Mechanics Action Group FM AG13*, vol. 164, 2003.

[162] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.

[163] W. Abdulla, "Mask r-cnn for object detection and instance segmentation on keras and tensorflow." `https://github.com/matterport/Mask_RCNN`, 2017.

[164] P. W. Hodges and B. H. Bui, "A comparison of computer-based methods for the determination of onset of muscle contraction using electromyography," *Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control*, vol. 101, no. 6, pp. 511–519, 1996.

[165] C. E. Boettcher, K. A. Ginn, and I. Cathers, "Standard maximum isometric voluntary

contraction tests for normalizing shoulder muscle emg," *Journal of orthopaedic research*, vol. 26, no. 12, pp. 1591–1597, 2008.

[166] K. Hauser and R. Shaw, "How medical robots will help treat patients in future outbreaks," *IEEE Spectrum*, 2020.

[167] M. T. Mason, "Toward robotic manipulation," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, pp. 1–28, 2018.

[168] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, p. eaat8414, 2019.

[169] J. Zhu, A. Cherubini, C. Dune, D. Navarro-Alarcon, F. Alambeigi, D. Berenson, F. Ficuciello, K. Harada, J. Kober, X. Li, *et al.*, "Challenges and outlook in robotic manipulation of deformable objects," *IEEE Robotics & Automation Magazine*, vol. 29, no. 3, pp. 67–77, 2022.

[170] M. Selvaggio, J. Cacace, C. Pacchierotti, F. Ruggiero, and P. R. Giordano, "A shared-control teleoperation architecture for nonprehensile object transportation," *IEEE Transactions on Robotics*, vol. 38, no. 1, pp. 569–583, 2021.

[171] S. Arevalo Arboleda, F. Rücker, T. Dierks, and J. Gerken, "Assisting manipulation and grasping in robot teleoperation with augmented reality visual cues," in *Proceedings of the 2021 CHI conference on human factors in computing systems*, pp. 1–14, 2021.

[172] S. S. White, K. W. Bisland, M. C. Collins, and Z. Li, "Design of a high-level teleoperation interface resilient to the effects of unreliable robot autonomy," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 11519–11524, IEEE, 2020.

[173] D. Schitz, S. Bao, D. Rieth, and H. Aschemann, "Shared autonomy for teleoperated driving:

A real-time interactive path planning approach," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 999–1004, IEEE, 2021.

[174] D. Szafir and D. A. Szafir, "Connecting human-robot interaction and data visualization," in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 281–292, 2021.

[175] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.

[176] M. D. Coovert, T. Lee, I. Shindev, and Y. Sun, "Spatial augmented reality as a method for a mobile robot to communicate intended movement," *Computers in Human Behavior*, vol. 34, pp. 241–248, 2014.

[177] C. Piyavichayanon, M. Koga, E. Hayashi, and S. Chumkamon, "Collision-aware ar telemanipulation using depth mesh," in *2022 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, pp. 386–392, IEEE, 2022.

[178] J. Larsson, M. Broxvall, and A. Saffiotti, "An evaluation of local autonomy applied to teleoperated vehicles in underground mines," in *2010 IEEE International Conference on Robotics and Automation*, pp. 1745–1752, IEEE, 2010.

[179] J. F. Mullen, J. Mosier, S. Chakrabarti, A. Chen, T. White, and D. P. Losey, "Communicating inferred goals with passive augmented reality and active haptic feedback," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 8522–8529, 2021.

[180] K. Chintamani, A. Cao, R. D. Ellis, and A. K. Pandya, "Improved telemanipulator navigation during display-control misalignments using augmented reality cues," *IEEE Transactions on*

Systems, Man, and Cybernetics-Part A: Systems and Humans, vol. 40, no. 1, pp. 29–39, 2009.

[181] A. D. Souchet, S. Philippe, D. Lourdeaux, and L. Leroy, "Measuring visual fatigue and cognitive load via eye tracking while learning with virtual reality head-mounted displays: A review," *International Journal of Human–Computer Interaction*, vol. 38, no. 9, pp. 801–824, 2022.

[182] N. Douthit, S. Kiv, T. Dwolatzky, and S. Biswas, "Exposing some important barriers to health care access in the rural usa," *Public health*, vol. 129, no. 6, pp. 611–620, 2015.

[183] D. Saparniene, B. Strukcinskiene, G. Mineviciute, A. Cizauskaite, L. Rapoliene, R. Grigoliene, I. Paciauskaite, and A. Genowska, "Working environment of health care professionals–focus on occupational stress," *Annals of Agricultural and Environmental Medicine*, vol. 30, no. 4, pp. 721–728, 2023.

[184] R. Naik, A. Kogkas, H. Ashrafian, G. Mylonas, and A. Darzi, "The measurement of cognitive workload in surgery using pupil metrics: a systematic review and narrative analysis," *Journal of Surgical Research*, vol. 280, pp. 258–272, 2022.

[185] Y.-F. Tsai, E. Viirre, C. Strychacz, B. Chase, and T.-P. Jung, "Task performance and eye activity: predicting behavior relating to cognitive workload," *Aviation, space, and environmental medicine*, vol. 78, no. 5, pp. B176–B185, 2007.