A-RIFT: Visual Substitution of Force Feedback for a Zero-Cost Interface in Telemanipulation

Alexander Moortgat-Pick[1], Peter So[1], Michael J Sack[2], Emma G Cunningham[2], Benjamin P Hughes[2],
Anna Adamczyk[1], Andriy Sarabakha[1], Leila Takayama[2], Sami Haddadin[1]

***Abstract*—We present an accessible robot interface for telemanipulation (A-RIFT), which preserves the haptic channel partially in a zero-additional-cost interface by visual substitution of force feedback (VSFF). This work explores a gap in the literature, resulting from the focus on performance improvements in telerobotics at increasing interface costs. Unlike most telemanipulation interfaces for high-degree-of-freedom robotic systems, this one requires minimal training and can be run in a web browser under high latency conditions, using an Internet connected computer with the user's own mouse and keyboard. To evaluate the performance of the system, we ran a controlled user study (N=12) to test how different distances (local vs. remote) and VSFF (on vs. off) affect the system's usability. As expected, participants in remote conditions performed worse than those in closer proximity. Despite several participants claiming that the visual display of force feedback did not help them, our analysis of their task performance showed that operators in remote condition actually performed statistically significantly better with the visual force feedback display than without it. These results indicate a promising new interface design direction for low-cost telemanipulation.**

I. INTRODUCTION

Robotic telepresence has a rich history of advancing substantial areas such as medicine, space and underwater exploration. The creation of robots deployed in these telepresence scenarios was driven by the need to reach into places that are difficult and unsafe for people to access on their own. The corresponding control interfaces have been deliberately made with their remote highly trained human operators in mind and the associated costs were of secondary importance. Hence, historically, robotic manipulators have only been available in special domains like research, medicine, nuclear, and space robotics. As a result, state-of-the-art control instruments are developed for highly-trained, specialised experts (e.g., surgeons, pilots). Consequently, these professionals teleoperate robots at specific sites, like research centers, laboratories, headquarters, control and operating rooms, with high-investment equipment [1], [2], [3]. To date, limited attention has been paid to developing widely available and usable interface solutions for telepresence and telemanipulation.

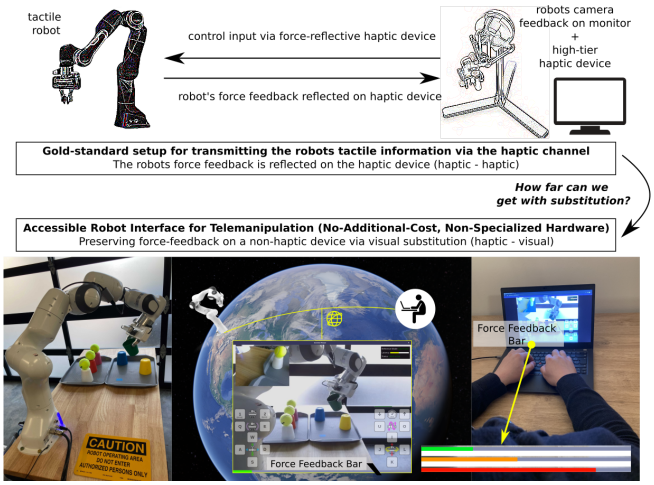


Fig. 1. Top: The gold-standard setup for haptic telemanipulation: force feedback from the robot reflected on a haptic device. Bottom: A web-based, internet-accessible user interface for telemanipulation connects users with robotic work cells across 9492 km using their personal computers. A web interface (bottom middle) allows users (right) to control a robotic arm to interact with objects (left). Force feedback is visually substituted to preserve the haptic information on the non-haptic, baseline device (here laptop) via a fillable bar overlaid on the remote operator's visual camera feed.
Bottom right corner: Three example states of the VSFF feedback bar are shown for low (green bar), medium (orange bar) and high (red bar) magnitudes of external forces $|\mathbf{F}_{ext}|$ at the robot's end-effector.

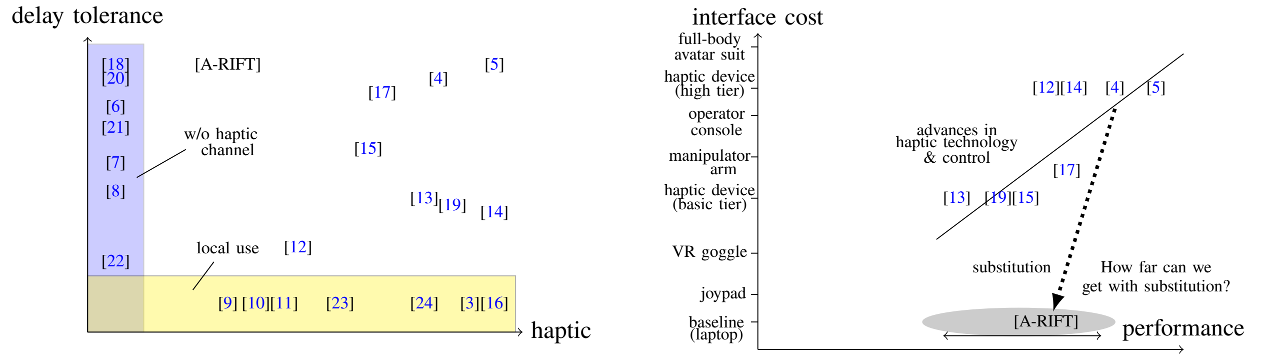
With the advancements of collaborative robots, a new generation of robots suitable for telepresence has come into focus [25]. Due to their force-sensitive and tactile abilities, these robots can interact with the environment in a delicate way without the need for special precautions, even aiming to reach human-inspired manipulation [26], [27]. Collaborative robots are becoming increasingly available and affordable, consequently shifting the bottleneck in telerobotics from the historical scarcity of suitable robots for remote operations to a scarcity of accessible, affordable, and usable control interfaces. Simultaneously, the Internet has revolutionized our everyday lives, from education, to work, to health care. Not only highly trained experts, but also minimally trained workers are beginning to use robotics in their everyday working lives in offices (e.g., [28]) and on factory floors (e.g., [29]). Moreover, networked connections of collaborative robots make it possible to upgrade to a tactile Internet [30]. Conveying tactile information in telemanipulation has been reserved for specialized, costly hardware. To bridge the gap between progress in telerobotic possibilities and non-expert users, novel concepts for low-barrier-to-entry telerobotic interfaces that can transmit tactile information without specialized hardware and handle global distances are needed.

---

[1] Alexander Moortgat-Pick, Peter So, Anna Adamczyk, Andriy Sarabakha and Sami Haddadin are with Chair of Robotics and Systems Intelligence and MIRMI – Munich Institute of Robotics and Machine Intelligence (formerly MSRM), Technical University of Munich (TUM), 80797 Munich, Germany. `{alexander.moortgat-pick, peter.so, anna.adamczyk, andriy.sarabakha, haddadin}@tum.de`

[2] Michael J Sack, Emma G Cunningham, Benjamin P Hughes and Leila Takayama are with the Human-Robot Interaction Lab, University of California, Santa Cruz, California, USA. `{mjsack, emgcunni, bephughe, takayama}@ucsc.edu`

[*] S. Haddadin has a potential conflict of interest, being a shareholder of Franka Emika GmbH.



(a) Qualitative (and possibly subjective) display of teleoperation systems regarding tolerated delay and haptic feedback capabilities. Classification based on the available information in cited literature.

(b) Qualitative comparison of systems in terms of interface costs and a (subjective) performance metric that consists of tolerance to delay and simultaneous preservation of haptic feedback capabilities.

Fig. 2. Classification of A-RIFT with respect to the literature.

A. *Contribution*

We have identified a trend in advances in telerobotics, namely increased performance regarding task performance and tolerated roundtrip delay, but at the price of increasing interface costs (Fig. 2). From the extensive prior work in the field, we know the haptic channel is invaluable as a component in the quality of telepresence. While haptic interfaces provide great value in their respective areas like space or medicine, their prohibitive costs prevents access by the general population. As a step towards closing this gap, we ask the, beforehand possibly trivial sounding, following question: *Is it possible to sustain (some) performance advantages of force feedback in telemanipulation by means of visual substitution while using a no-additional-cost interface?*

To examine this we present an accessible robot interface for telemanipulation (A-RIFT) with a force sensitive robotic arm (Franka Emika Robot [31]). We define 'accessible interface' by: broad availability (connection via Internet, VPN) and using common computer hardware (laptop). A-RIFT preserves the haptic channel partially by visual substitution of force feedback (VSFF). Obviously, the performance is limited compared to the gold-standard of passive control with top-tier haptic devices.

Furthermore, A-RIFT follows a holistic interface design, which was developed by a close interdisciplinary collaboration between robotics and human-computer interaction team members. To date, most telerobotics systems have been developed from the perspective of control theory contributions. In contrast, A-RIFT was designed with a user-centered design process that included robotics control theory, human perception and performance, and end-user perspectives.

As the integration of VSFF in our system targets real-world use by non-specialized audience across the globe, an empirical user study investigates user performance and experience during transatlantic (9492 km beeline) telemanipulation in a 2x2 configuration setup: VSFF on/off and transatlantic/local robot-operator-distance, which induces high/low-varying-latency. We have found VSFF increased novice users' performance, especially in the case of higher roundtrip delays (hereafter referred to as "delay"). Which results in the, beforehand non-obvious observation, that the substitution of traditionally costly channels, like haptics, merged with above-defined 'accessible' interfaces forms a synergy into a distinct type of telemanipulation interface that induces no additional hardware nor cost, while preserving haptic information and a performance boost in telemanipulation due to VSFF. This type of accessible telemanipulation interfaces with channel substitution have the potential to democratize the access to telepresence and thus reduce the digital divide.

To the authors knowledge the current work presents the first visually rendered force feedback overlay in web-based telemanipulation under high latency conditions with a collaborative robot using a non-high tier device.

II. RELATED WORK

A. *Haptic feedback in Telerobotics*

In 1994, [20] reported the first successful teleoperation of a robot in space from the surface of the earth. The 6DoF robot was controlled with a 6DoF 'control ball' leader interface with a supervisory control scheme in simulation including robot feedback, i.e. a task was first solved in simulation, and the solution then replayed on the robot. Wave-variables were used for online control of a 2DoF robot in space [17] under varying, but lower time-delay during overflight of a space station. Several advancements in control, such as force control for teleoperation [19] or the time-domain passivity control [4] allowed to build systems that allow stable bilateral haptic teleoperation even under higher time-varying communication delays [5]. Other work explored stability via model-mediation in teleoperation [15] or shared autonomy for more efficient orbital teleoperation of space robots [18]. Simultaneously, telemanipulation systems for minimally invasive surgery were developed [16].

B. *Visual sensory substitution of force feedback*

Transparency in telerobotics can be increased by displaying the robot's sensory feedback to the operator. Force feedback can provide a tactile sense of the remote environment and has been shown to noticeably lead to improvement in telemanipulation [32], [33]. While the most immediate approach to display force feedback to the operator is through reflecting it on a leader [34], it requires force sensitive hardware and can introduce instabilities on the systems level

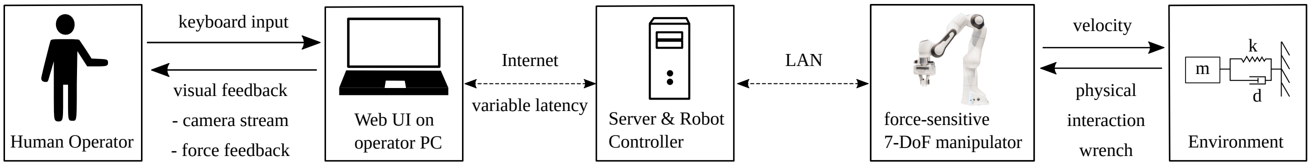


Fig. 3. The hardware setup of our web-based telemanipulation system. A human operator connects via the Internet and a web interface on a user computer to the server. The server provides access to the controlled robot and allows the operator to interact with the remote environment.

when used over a channel with time-varying delays, like an Internet connection [13], [35], [14]. In [12], [11] audio and vibrotactile rendering of forces has been presented as a viable alternative to reflective force feedback.

Visual rendering of forces for sensory substitution has been utilized in medical telemanipulation, where, during surgical applications, force feedback from a surgical tool tip is rendered visually into the camera stream showing the tool tip. For example, applied forces are rendered as a visual overlay during robot-assisted suture-tying, in [23] as a bar, in [10] as a dot and in a mock-up of teleoperated palpation as a bar [24]. All studies saw improved accuracy of inexperienced surgeons. Note, each of these previous setups used the high-tier da Vinci Surgical System (Intuitive Surgical, Inc.) and a local setup for the experiments. In [9], visual rendering of end-effector forces facilitated grasping with a construction robot via a joystick in a closed, local setup.

The previous work to investigate the effects of visual force feedback rendering in telemanipulation use a direct communication link between leader and follower, thus omitting effects of delay and packet loss. Unlike the previous work, we examine the effects of varying (low and high) latency coupled with visual force rendering on the systems usability.

C. *Web-based telemanipulation interfaces*

In [36], [37], the first web-based teleoperation interfaces were presented. A web-based robot control with buttons for interaction with moving objects was developed in [38]. A teleoperation interface was presented in [39] for a mobile robot via mouse/keyboard (M/KB) over the Internet, but without investigating the effect of time-delay. To control the point of view of a remote manipulator, [7] introduced a virtual reality interface with a VR headset and 3D hand-held controller input, compared against a M/KB interface, connecting to a remote location 64km away via the Internet. The proposed interface improves task performance and subjective workload, but introduces additional specialised devices. For teleoperation of a mobile manipulator used in CERN-facilities, [8] implemented an on-site local network based interface utilizing multiple input devices from M/KB to haptic devices and feedback from two cameras displayed in the UI. In [40] a web-based interface for robot telemanipulation in adult care scenarios was presented. The system is based on a 7-DoF robot arm controlled in local network via either M/KB or an kinesthetic device, using feedback from first and third person view cameras respectively. Participants of a user study were slower but more successful in task completion with M/KB compared to the kinesthetic device. In [21] a web interface for teleoperation of a service robot was shown, and later in [22] a VR-enhanced web interface was able to visually display contact of a service robot with it's environment. An unenhanced Internet connection with a large round trip delay was used in [6] to control a lightweight robot across the Atlantic for telemanipulation in an assembly task via a specialized leader console w/o force feedback.

From previous work on web-based telerobotics with a simple computer M/KB interface we can conclude that, although not as powerful as highly specialized interfaces, this setup could be sufficient for an operator to accomplish a variety of tasks. Furthermore, none of the web-based telerobotic systems utilize VSFF or aim to investigate how to provide accessibility to tactile telerobotics to a broad, non-expert audience without the need for special hardware.

D. *Classification of A-RIFT with respect to literature*

A qualitative and obviously potentially subjective comparison of the related work to A-RIFT regarding tolerance to delay and haptic capabilities is given in Fig. 2a based on the information available in the cited work. Three areas can be differentiated: Systems without a haptic channel (blue area), often web-based, exploring various input modalities such as joypads, joysticks, tablets or M/KB while mostly visual, vibro-tactile or auditory feedback. Secondly, systems with haptic display of the haptic channel via haptic joysticks or haptic devices without investigated tolerance to time-delay (yellow area). And lastly, time-delayed systems that use the haptic channel (uncolored area). The systems towards the top-right corner prove, that both stable haptic feedback and tolerance to high time-delays is possible, constituting the gold-standard in haptic teleoperation.

Focusing only on the systems capable of time-delayed haptic teleoperation in the uncolored area from Fig. 2a regarding the costs of the used interface, we can identify a trend, depicted as the line in Fig. 2b. With increasing performance (subjective metric consisting of haptic capabilities and delay tolerance), the cost of the interface also increases. A-RIFT takes a different approach via haptic-to-visual sensory mode substitution of the haptic channel, thus resetting the interface cost to the baseline (a computer) but obviously at the cost of performance compared to the most recent literature. The question is now, how far the performance gap towards the gold-standard can be closed by means of substitution alone.

III. SYSTEM DESIGN

A. *Interface Design*

Our system design goal was to create an accessible and usable interface for telemanipulation (Fig. 3), moving away from existing, high-end telemanipulation solutions and expanding it to a broader user base. To expand the accessibility of this system, we identified three leverage points: hardware availability and costs, locality, and training requirements.

The most pressing challenge is that we have a low-dimension input device on the user side (keyboard) but a high-dimension task space (6-DoF) on the robot side. Keyboard control is much harder for complex systems moving in 3D space, as opposed to driving the simpler 2D-motion of a mobile vehicle. In addition, the system has to operate via a high-delay, lossy channel (Internet). Apart from limited input options, the main display capabilities of computer peripherals are visual, thus the system will be limited to this. Obviously, this will limit the maximal performance of the system to be somewhat lower than the gold-standard of haptic devices.

We identified two distinct modes of teleoperation for a robot arm: *base frame control* makes it feel like operating the robot from a third-person perspective and *tool frame control*, making it feel like first-person flying. For the purpose of our study, we restricted participants to base frame control. During early user testing with professional remotely operated vehicle (ROV) pilots, we discovered that a picture-in-picture composited video feed of an arm-mounted and a fixed camera was beneficial, especially when the fixed camera was aligned with the base frame control axes.

Our application of *sensor fusion* [41] within the interface was influenced by literature in the field of psychology indicating factors such as cognitive load [42], attention, and useful field-of-view [43], [44], [42], [45] affect human performance. The sensory sources are the composite video feeds, a visual substitution of haptic feedback, a mapping of input controls, as well as a display of the robot's state. To decrease cognitive demand on users, we restricted the configuration of elements to a central viewport.

1) Visual substitution of force feedback: To preserve force feedback information on a non-haptic control device, we visually substituted the force feedback. Here, the VSFF has the form of a horizontal meter where the force feedback is associated with two characteristics in the visual substitution: color and length of the rendered bar (Fig. 1).

The VSFF is displayed as a bar b_{VSFF}, characterized by its RGB color c_{VSFF} and length l_{VSFF} in pixel where

$$\mathrm{b}_{VSFF}(\alpha) = f\left(c_{VSFF}(\alpha), l_{VSFF}(\alpha)\right) \tag{1}$$

$$c_{VSFF}(\alpha) = [255\alpha, 255(1-\alpha), 0]^T \tag{2}$$

$$l_{VSFF}(\alpha) = \alpha l_{viewport} \tag{3}$$

and α is the feedback gain (Eq. 12) and $l_{viewport}$ the width of the viewport in pixel, depending on the size of the browser window. This results in a continuously rendered colored amplitude inside the meter (Fig. 1 bottom right).

We made several assumptions during the development of our interface, the first of which being that teleoperation must involve visual camera feedback. Furthermore, we made the assumption that force feedback was not the primary objective of telemanipulation, but rather a means for supporting the completion of a given task quickly and effectively [12], [11]. If one were merely interested in accurately perceiving force exerted by the robot, the feedback could be depicted numerically. However, this mode of presenting haptic information might have a comparatively negative effect on task performance by simultaneously lowering the prominence of feedback while increasing an operator's cognitive load.

Furthermore, in the interest of minimizing the delay between input and feedback, we determined that this rendering should be coupled to the refresh rate of the operator's monitor [13], [35], [14]. The implication here being that force feedback must be perceived concurrently with video feedback, and must not impede task execution.

B. Control Scheme

The control scheme resulting from the general requirements and the design choices made for the interface is shown in Fig. 4. Via a keyboard an operator issues a desired twist $\dot{\boldsymbol{x}}_d \in \mathbb{R}^6$ of the robots end effector in cartesian space

$$\dot{\boldsymbol{x}}_d = v_u \ \boldsymbol{A} \ \dot{\boldsymbol{x}}_{max}, \tag{4}$$

where $v_u \in [-1, 1]$ is the velocity scaling, $\dot{\boldsymbol{x}}_{max} \in \mathbb{R}^6$ is the maximal velocity and $\boldsymbol{A} \in \mathbb{R}^{6\times 6}$ a diagonal matrix denoting the keyboard state (for ANSI or ISO keyboard layouts)

$$\boldsymbol{A} = diag(a_1, a_2, a_3, a_4, a_5, a_6), \tag{5}$$

$$with \begin{cases} a_1 = 1, \text{if 'W' held;} & -1, \text{if 'S' held; } 0, \text{else} \\ a_2 = 1, \text{if 'A' held;} & -1, \text{if 'D' held; } 0, \text{else} \\ a_3 = 1, \text{if 'I' held;} & -1, \text{if 'K' held; } 0, \text{else} \\ a_4 = 1, \text{if 'Q' held;} & -1, \text{if 'E' held; } 0, \text{else} \\ a_5 = 1, \text{if 'U' held;} & -1, \text{if 'O' held; } 0, \text{else} \\ a_6 = 1, \text{if 'J' held;} & -1, \text{if 'L' held; } 0, \text{else.} \end{cases}$$

The Cartesian policy in Fig. 4 running at guaranteed sample time is expressed as

$$\dot{\boldsymbol{x}}_d = \begin{bmatrix} \dot{\boldsymbol{p}}_d \\ \boldsymbol{\omega}_d \end{bmatrix} \tag{6}$$

$$\mathbf{R}_d = e^{\int \boldsymbol{\omega}_d \mathrm{d}t} \tag{7}$$

$$\mathbf{T}_{EE,d} = \begin{bmatrix} \mathbf{R}_d & \int \dot{\boldsymbol{p}}_d \mathrm{d}t \\ 0\,0\,0 & 1 \end{bmatrix}, \tag{8}$$

where $\mathbf{R}_d \in \mathbb{R}^{3x3}$ is the synchronous desired rotation matrix computed from the asynchronous desired translational and rotational end-effector velocity $\dot{\boldsymbol{p}}_d$ and $\boldsymbol{\omega}_d \in \mathbb{R}^{3x1}$, respectively. The desired pose $\mathbf{T}_{EE,d} \in \mathbb{R}^{4x4}$ with respect to local base frame is then fed to a Cartesian impedance controller.

The flexible joint robot is modelled as [46]

$$\mathbf{M}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \mathbf{C}(\boldsymbol{q}, \dot{\boldsymbol{q}})\dot{\boldsymbol{q}} + \boldsymbol{g}(\boldsymbol{q}) = \mathbf{K}(\boldsymbol{\theta} - \boldsymbol{q}) + \boldsymbol{\tau}_{ext} \tag{9}$$

$$\mathbf{B}\ddot{\boldsymbol{\theta}} + \mathbf{K}(\boldsymbol{\theta} - \boldsymbol{q}) = \boldsymbol{\tau}_m, \tag{10}$$

where $\boldsymbol{q} \in \mathbb{R}^7$ are the link and $\boldsymbol{\theta} \in \mathbb{R}^7$ the motor positions, $\mathbf{M}(\boldsymbol{q}) \in \mathbb{R}^{7x7}$ is the inertia matrix, $\mathbf{C}(\boldsymbol{q}, \dot{\boldsymbol{q}})\dot{\boldsymbol{q}} \in \mathbb{R}^{7x7}$ the centripetal and Coriolis vector, $\boldsymbol{g}(\boldsymbol{q}) \in \mathbb{R}^7$ the gravity vector, $\mathbf{K}$ and $\mathbf{B}$ are the joint stiffness and motor inertia matrices, $\boldsymbol{\tau}_m$ is the motor torque vector and $\boldsymbol{\tau}_{ext}$ the vector of external forces acting on the end-effector by the environment. Based on above model, we use a state of the art impedance

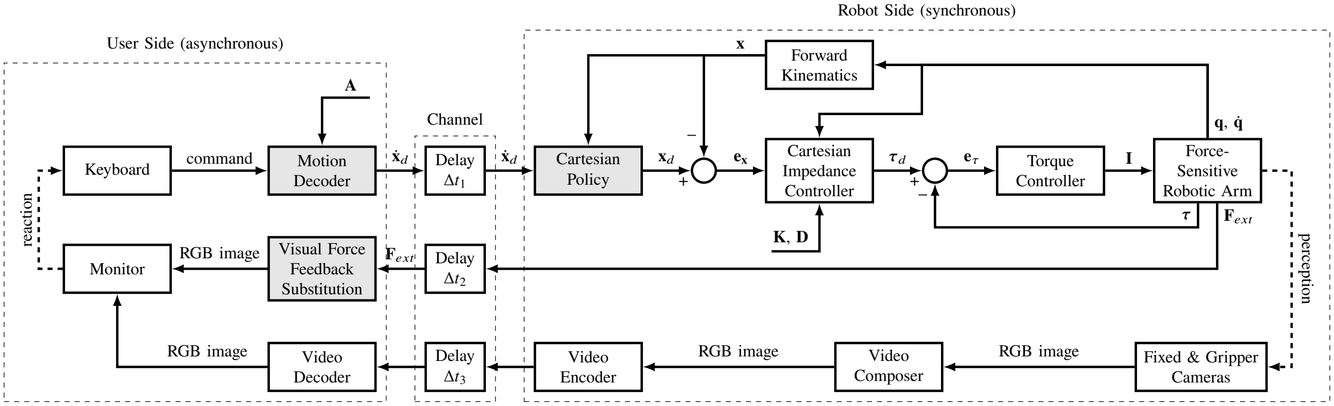


Fig. 4. Overview of the control scheme. A user sends the desired velocity twist $\dot{\mathbf{x}}_d$ via a keyboard and motion decoder over a channel as input to a local robot control loop. A Cartesian impedance controller derives the desired end-effector torque $\boldsymbol{\tau}_d$. The estimated external force $\mathbf{F}_{ext}$ is fed back to the operator console, where it is rendered as a visual UI element that is displayed together with the composed video stream on a screen.

controller with redundancy resolution [47] with closed loop dynamics

$$\boldsymbol{\Lambda}(\boldsymbol{q})\ddot{\boldsymbol{e}}_x + \mathbf{D}_d\dot{\boldsymbol{e}} + \mathbf{K}_d\boldsymbol{e}_x + \bar{\mathbf{C}}(\boldsymbol{q},\dot{\boldsymbol{q}})\dot{\boldsymbol{e}}_x = \mathbf{F}_{ext}. \tag{11}$$

Here, $\boldsymbol{\Lambda}(\boldsymbol{q}) = (\mathbf{J}(\boldsymbol{q})\bar{\mathbf{M}}(\boldsymbol{q})^{-1}\mathbf{J}(\boldsymbol{q})^T)^{-1}$ denotes the equivalent Cartesian mass matrix with $\mathbf{J}(\boldsymbol{q})$ being Jacobian. The desired damping and stiffness matrices are $\mathbf{D}_d$ and $\mathbf{K}_d$, respectively. $\bar{\mathbf{C}}(\boldsymbol{q},\dot{\boldsymbol{q}})$ is an arbitrary skew symmetric matrix and $\boldsymbol{e}_x = \boldsymbol{x}_d - \boldsymbol{x}$ is the control error.

The estimated external joint torque $\boldsymbol{\tau}_{ext}$ is then used to estimate the generalized external force on the end-effector $\boldsymbol{F}_{ext} \in \mathbb{R}^6$. This is sent back to the user side over a lossy channel with delay Δt_2.

On the user side, the feedback gain α is

$$\alpha = \begin{cases} \frac{|\mathbf{F}_{ext}|}{\mathbf{F}_{ext,max}}, & \mathbf{F}_{ext} \le \mathbf{F}_{ext,max} \\ 1, & \mathbf{F}_{ext} > \mathbf{F}_{ext,max}, \end{cases} \tag{12}$$

encoding the expected maximal force $\mathbf{F}_{ext,max}$ issued by the environment on the end-effector (here: $\mathbf{F}_{ext,max} = 40\ N$). Visual substitution of the haptic channel is done by rendering an image of a bar, which size and green-yellow-red color scheme is chosen to be directly proportional to α (2),(3).

Additionally, feedback from two RGB cameras is used: An external fixed camera for third-person view and a camera mounted onto the gripper for first-person view. Both video streams are composed into a single stream, compressed and transmitted to the user side over a lossy channel with delay Δt_3. On the user side, the stream is decoded by the video decoder and displayed on the monitor along with the force feedback. Finally, the human operator closes the loop by reacting to the feedback and providing a new command.

C. System Architecture and Networking

The networked system architecture is depicted in Fig. 5 and consists of three main components: The web interface runs within a browser on the user's computer (top left), the application server hosting A-RIFT (bottom middle) and connecting it to the robot runs on the main physical server, and the robot controller (top right) runs on the robot host computer. We chose a web-based implementation for the user interface to maximize the availability and accessibility, successfully tested during the study on Windows, Mac OS and Ubuntu Linux. Minimizing latency is extremely important in teleoperation as it affects its transparency. Thus we chose a peer-to-peer (p2p) architecture for the communication between all components, as opposed to e.g. broker-based. The Web UI is first loaded via HTTP from the server, then it creates p2p websocket connections to exchange commands, using also the robot operating system (ROS) [48] to stream input, force feedback and video. This implementation can be improved by using webRTC for all web streaming, as it transports via UDP instead of TCP compared to websockets.

The application server handles the camera input and connects the Web UI to the robot controller via custom UDP and websocket protocols. On the robot host, the outer control loop including the impedance controller is executed at 1 kHz, calculating desired torques that are transmitted to the motor controller via the Franka Control Interface API.

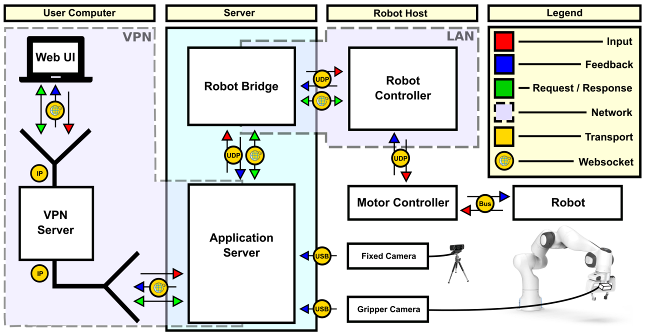


Fig. 5. Overview of the system architecture. The user computer is connected with the application server through a VPN. The server connects with the robot controller in a LAN via a bridge. Only the robot controller has direct access to the motor controller of the robot.

IV. EXPERIMENTS

A. Experiment design

We ran an initial user study to evaluate our systems performance. 12 participants in Bavaria, 7 male and 5 female, were asked to accomplish a complex object manipulation task with a robotic arm located in California via the Internet, and as a comparison with a robot located in Bavaria, using the A-RIFT from their homes or offices in the greater Munich area. The task was to flip three cups using the robot's gripper and

(a) Task: start state. (b) Task: goal state.

Fig. 6. The task for the experiments in the user study: Participants were asked to flip cups and insert one tennis ball into each cup using A-RIFT.

placing a tennis ball into each within three minutes (Fig. 6). During this, participants were either aided by VSFF or not. It was almost impossible to complete all three in the allotted time, allowing us to hold the experiment duration constant. We ran our user study as a 2 x 2 Latin squares within-subjects experiment design, such that each participant experienced all possible conditions in a counterbalanced order to account for potential learning effects. A within-subjects design was chosen to account for large individual differences in teleoperation performance. The two independent variables were VSFF (on vs. off) and distance between the teleoperator and the robotic arm (local vs. remote).

B. Study Procedure and Metrics

After consenting to participate, participants engaged in a 15 minute training period to familiarize themselves with the interface and task. Participants then completed the task four times, once for each of the within-subjects conditions. After each trial they completed a brief survey measuring task load, remote presence, and embodiment. They completed a short demographic survey after the final trial. We measured observable behaviours during each of the study's four conditions, including *task performance*, defined as the number balls placed in upturned cups, and *errors*, defined as unintentional impacts to any of the cups or balls that resulted in them changing position.

We measured features of cognitive load including temporal demand using the NASA TLX questionnaire [49]; remote presence in the virtual environment, using items from the Slater-Usoh-Steed (SUS) [50]; and embodiment of the robot arm, using a modified question from [51].

V. RESULTS

To statistically test the effects of distance (local vs. remote) and visual displays of force feedback (displayed vs. not), we ran repeated measures ANOVAs, using a statistical significance cut-off of p=.05.

A. Network delays

Fig. 7a shows that users experienced an average network delay of 50.28 msec (SE=4.30) in the local and 219.63 msec (SE=5.86) in the remote location. This was a statistically significant main effect of location, $F(1,11)=3371.59$, $p<.001$, partial η^2=.99. We did not find a main effect of force feedback nor did we find an interaction effect between the two independent variables. When participants operated the robot in the remote location, the robot experienced higher forces (M=4.17 Newtons, SE=0.22) than when it was in the local location (M=3.30 Newtons, SE=0.34), $F(1,11)=11.47$, p=.006, partial η^2=.51. See Fig. 7b.

B. Task performance

Participants performed best in the local conditions (Fig. 7c), placing an average of 0.54 (SE=0.13) balls in cups, compared to an average of 0.17 (SE=0.07) in the remote condition. This was a statistically significant main effect of location (local vs. remote) upon the number of balls placed in cups $F(1,11)=11.05$, p=.005, partial η^2=.52. We did not find a significant main effect of force feedback or an interaction effect between the two independent variables.

In terms of errors made while performing the teleoperation task (Fig. 7d), we found a statistically significant interaction effect between location (local vs. remote) and force feedback (displayed vs. not), $F(1,11)=5.50$, p=.04, partial η^2=.33. Participants made fewer errors the remote location when they had force feedback displayed (M=0.08, SE=0.08) than when they did not have force feedback displayed (M=0.42, SE=0.15), pair-wise comparison, p=.04.

C. User experience

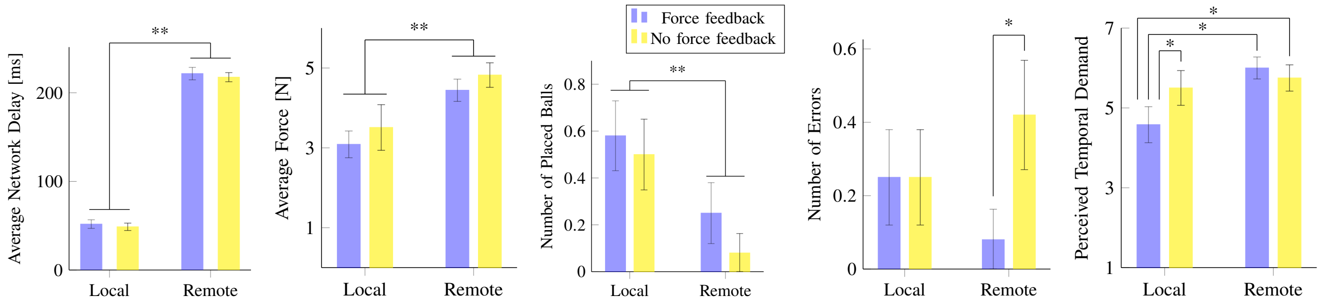
In terms of how much temporal demand people experienced, we found a statistically significant interaction effect between location (local vs. remote) and force feedback (displayed vs. not), $F(1,11)=4.59$, p=.02, partial η^2=.41. People experienced the least temporal demand in the local condition with force feedback displayed (Fig. 7e).

D. Cup configurations

During the last session of our study, we noticed that the cups at the local site were arranged inconsistently (left, right, left) compared to the cups at the remote site (right, left, right). The inconsistency at the local site occurred just after the first two participants had participated in the study. To see if our inconsistent cup configurations affected task performance or user experience in this study, we ran follow-up statistical tests. To account for potential ordering effects, we compared the first two participants against the other participants, who experienced the conditions in the same orders as them, n=6. We found no statistically significant effects of local site cup configuration upon the number of balls placed in cups at the local site – local site with force feedback (χ^2=1.5, p=.22) and local site without force feedback (χ^2=1.5, p=.22). We found no statistically significant differences for any of the task performance or user experience measures reported above. These analyses gave us reason to believe that this oversight did not significantly affect the performance and user experience measures.

E. Implications for the design of telemanipulation systems

The results of our user study demonstrate how users can telemanipulate objects, even while operating at a physical distance of 9,490 kilometers with relatively high latency.



(a) The average network delay (in ms) experienced by participants was influenced by the location of the participant relative to the robot.

(b) The robot arm experienced higher forces (in N) when participants were operating it from the remote location.

(c) Participants were able to place the most balls in cups when they were close to the robot and could see the force feedback information.

(d) Participants made the most errors when they were far from the robot and could not see the force feedback information.

(e) The temporal demand (min 1, max 7) experienced by participants was influenced by the combination of location and force feedback.

Fig. 7. Experimental results.

A-RIFT has the potential to allow people to manipulate a robot over the Internet with nothing more than their personal computers and keyboards, even with minimal training. The addition of visually rendered force feedback to the UI aided in the prevention of mistakes during the task. The force feedback meter affords users information about the contact made between the robot and its environment, even when the end effector is obscured from view of the camera. It is therefore possible that this allowed participants to avoid errors that would have otherwise resulted from partial obstruction of the visual field. This suggests that visually rendered force feedback may have allowed users to adjust their strategy based on relevant information from the perspective of the robot. Furthermore, force feedback reduced feelings of temporal demand in the local condition. These results underscore the importance of visually rendered feedback for robotic telemanipulation systems.

VI. Conclusion

This paper presents A-RIFT, a zero-additional-cost interface with visual substitution of force feedback (VSFF) of the haptic channel for telemanipulation with a force sensitive robotic arm. With A-RIFT we investigate a gap in the current literature: low cost interfaces for telemanipulation with haptic information, available to the broader population. In our user study, we found that substituted haptic feedback increased users task performance under high latency conditions, such as teleoperation across the Atlantic. Furthermore, we found that visually rendered force feedback reduced the perceived temporal demand experienced by users.

In future work, we intend to investigate how we can further improve the benefits of VSFF and compare it against haptic interfaces. Instead of the distance condition, latency can be artificially induced to investigate its effects on task performance with VSFF more directly. Other substitution of the haptic channel such as auditory will be investigated, as well as different options for visual substitution such as logarithmic instead of linear feedback display.

Acknowledgements

The authors would like to thank the Dobeneck Technologie Stiftung for financial support as part of the SVAN: Synchronous Team-Robot Van project. This work was supported by the Federal Ministry of Education and Research of the Federal Republic of Germany (BMBF) within the project VERITAS under the Project Number 01IS18073A. Funded by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) as part of Germany's Excellence Strategy – EXC 2050/1 – Project ID 390696704 – Cluster of Excellence "Centre for Tactile Internet with Human-in-the-Loop" (CeTI) of Technische Universität Dresden. We gratefully acknowledge the funding of the Lighthouse Initiative Geriatronics by StMWi Bayern (Project X, grant no. IUK-1807-0007// IUK582/001) and LongLeif GaPa gGmbH (Project Y). The authors would like to gratefully acknowledge the general support by the Technical University of Munich's Institute for Advanced Study.

References

[1] G. S. Guthart and J. K. Salisbury, "The intuitive/sup tm/telesurgery system: overview and application," in *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*, vol. 1. IEEE, 2000, pp. 618–621.

[2] S. Martin and N. Hillier, "Characterisation of the novint falcon haptic device for application as a robot manipulator," in *Australasian Conference on Robotics and Automation (ACRA)*. Citeseer, 2009, pp. 291–292.

[3] A. Tobergte, P. Helmer, U. Hagn, P. Rouiller, S. Thielmann, S. Grange, A. Albu-Schäffer, F. Conti, and G. Hirzinger, "The sigma. 7 haptic interface for mirosurge: A new bi-manual surgical console," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2011, pp. 3023–3030.

[4] J.-H. Ryu, D.-S. Kwon, and B. Hannaford, "Stable teleoperation with time-domain passivity control," *IEEE Transactions on robotics and automation*, vol. 20, no. 2, pp. 365–373, 2004.

[5] J.-H. Ryu, J. Artigas, and C. Preusche, "A passive bilateral control scheme for a teleoperator with time-varying communication delay," *Mechatronics*, vol. 20, no. 7, pp. 812–823, 2010.

[6] J. Bohren, C. Papazov, D. Burschka, K. Krieger, S. Parusel, S. Haddadin, W. L. Shepherdson, G. D. Hager, and L. L. Whitcomb, "A pilot study in vision-based augmented telemanipulation for remote assembly over high-latency networks," in *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 3631–3638.

[7] D. Whitney, E. Rosen, E. Phillips, G. Konidaris, and S. Tellex, "Comparing robot grasping teleoperation across desktop and virtual reality with ros reality," in *Robotics Research*. Springer, 2020, pp. 335–350.

[8] G. Lunghi, R. M. Prades, and M. Di Castro, "An advanced, adaptive and multimodal graphical user interface for human-robot teleoperation in radioactive scenarios." in *ICINCO (2)*, 2016, pp. 224–231.

[9] A. A. Yusof, T. Kawamura, and H. Yamada, “Evaluation of construction robot telegrasping force perception using visual, auditory and force feedback integration.” *J. Robotics Mechatronics*, vol. 24, no. 6, pp. 949–957, 2012.

[10] C. E. Reiley, T. Akinbiyi, D. Burschka, D. C. Chang, A. M. Okamura, and D. D. Yuh, “Effects of visual force feedback on robot-assisted surgical task performance,” *The Journal of thoracic and cardiovascular surgery*, vol. 135, no. 1, pp. 196–202, 2008.

[11] M. J. Massimino, “Improved force perception through sensory substitution,” *Control Engineering Practice*, vol. 3, no. 2, pp. 215–222, 1995.

[12] M. J. Massimino and T. B. Sheridan, “Sensory substitution for force feedback in teleoperation,” *Presence: Teleoperators & Virtual Environments*, vol. 2, no. 4, pp. 344–352, 1993.

[13] G. Niemeyer and J.-J. E. Slotine, “Towards force-reflecting teleoperation over the internet,” in *Proceedings. 1998 IEEE International Conference on Robotics and Automation (Cat. No. 98CH36146)*, vol. 3. IEEE, 1998, pp. 1909–1915.

[14] J. J. Gil, E. Sánchez, T. Hulin, C. Preusche, and G. Hirzinger, “Stability boundary for haptic rendering: Influence of damping and delay,” *Journal of Computing and Information Science in Engineering*, vol. 9, no. 1, 2009.

[15] B. Willaert, H. V. Brussel, and G. Niemeyer, “Stability of model-mediated teleoperation: Discussion and experiments,” in *International conference on human haptic sensing and touch enabled computer applications*. Springer, 2012, pp. 625–636.

[16] A. Tobergte, R. Konietschke, and G. Hirzinger, “Planning and control of a teleoperation system for research in minimally invasive robotic surgery,” in *2009 IEEE International Conference on Robotics and Automation*. IEEE, 2009, pp. 4225–4232.

[17] G. Hirzinger, K. Landzettel, D. Reintsema, C. Preusche, A. Albu-Schäffer, B. Rebele, and M. Turk, “Rokviss-robotics component verification on iss,” in *Proc. 8th Int. Symp. Artif. Intell. Robot. Autom. Space (iSAIRAS)(Munich 2005) p. Session2B*, 2005.

[18] P. Schmaus, D. Leidner, T. Krüger, R. Bayer, B. Pleintinger, A. Schiele, and N. Y. Lii, “Knowledge driven orbit-to-ground teleoperation of a robot coworker,” *IEEE Robotics and Automation Letters*, vol. 5, no. 1, pp. 143–150, 2019.

[19] J. Park and O. Khatib, “A haptic teleoperation approach based on contact force control,” *The International Journal of Robotics Research*, vol. 25, no. 5-6, pp. 575–591, 2006.

[20] G. Hirzinger, B. Brunner, J. Dietrich, and J. Heindl, “Rotex-the first remotely controlled robot in space,” in *Proceedings of the 1994 IEEE international conference on robotics and automation*. IEEE, 1994, pp. 2604–2611.

[21] B. Pitzer, S. Osentoski, G. Jay, C. Crick, and O. C. Jenkins, “Pr2 remote lab: An environment for remote development and experimentation,” in *2012 IEEE International Conference on Robotics and Automation*. IEEE, 2012, pp. 3200–3205.

[22] P. M. Grice and C. C. Kemp, “In-home and remote use of robotic body surrogates by people with profound motor deficits,” *PloS one*, vol. 14, no. 3, p. e0212904, 2019.

[23] M. Kitagawa, D. Dokko, A. M. Okamura, and D. D. Yuh, “Effect of sensory substitution on suture-manipulation forces for robotic surgical systems,” *The Journal of thoracic and cardiovascular surgery*, vol. 129, no. 1, pp. 151–158, 2005.

[24] J. C. Gwilliam, M. Mahvash, B. Vagvolgyi, A. Vacharat, D. D. Yuh, and A. M. Okamura, “Effects of haptic and graphical force feedback on teleoperated palpation,” in *2009 IEEE international conference on robotics and automation*. IEEE, 2009, pp. 677–682.

[25] A. Bauer, D. Wollherr, and M. Buss, “Human–robot collaboration: a survey,” *International Journal of Humanoid Robotics*, vol. 5, no. 01, pp. 47–66, 2008.

[26] S. Haddadin, A. De Luca, and A. Albu-Schäffer, “Robot collisions: A survey on detection, isolation, and identification,” *IEEE Transactions on Robotics*, vol. 33, no. 6, pp. 1292–1312, 2017.

[27] L. Johannsmeier, M. Gerchow, and S. Haddadin, “A framework for robot manipulation: Skill formalism, meta learning and adaptive control,” in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 5844–5850.

[28] M. K. Lee and L. Takayama, “" now, i have a body" uses and social norms for mobile remote presence in the workplace,” in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2011, pp. 33–42.

[29] A. Sauppé and B. Mutlu, “The social impact of a robot co-worker in industrial settings,” in *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 2015, pp. 3613–3622.

[30] S. Haddadin, L. Johannsmeier, and F. D. Ledezma, “Tactile robots as a central embodiment of the tactile internet,” *Proceedings of the IEEE*, vol. 107, no. 2, pp. 471–487, 2018.

[31] Franka Emika Robot, https://www.franka.de/, Accessed: 2021/04/07.

[32] T. B. Sheridan, *Telerobotics, automation, and human supervisory control*. MIT press, 1992.

[33] W. R. Ferrell, “Delayed force feedback,” *Human factors*, vol. 8, no. 5, pp. 449–455, 1966.

[34] M. J. Massimino and T. B. Sheridan, “Variable force and visual feedback effects on teleoperator man/machine performance,” in *Proceedings of the NASA Conference on Space Telerobotic*, 1989, pp. 1751–1756.

[35] N. Chopra, M. W. Spong, S. Hirche, and M. Buss, “Bilateral teleoperation over the internet: the time varying delay,” in *Proceedings of the American Control Conference (ACC)*, 2003.

[36] K. Goldberg, M. Mascha, S. Gentner, N. Rothenberg, C. Sutter, and J. Wiegley, “Desktop teleoperation via the world wide web,” in *Proceedings of 1995 IEEE International Conference on Robotics and Automation*, vol. 1. IEEE, 1995, pp. 654–659.

[37] K. Taylor and J. Trevelyan, “Australia’s telerobot on the web,” in *Proceedings of the International Symposium on Industrial Robots*, vol. 26. INTERNATIONAL FEDERATION OF ROBOTICS, & ROBOTIC INDUSTRIES, 1995, pp. 39–44.

[38] I. Belousov, S. Chebukov, and V. Sazonov, “Web-based teleoperation of the robot interacting with fast moving objects,” in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*. IEEE, 2005, pp. 673–678.

[39] R. L. Alves, R. F. Vassallo, E. O. Freire, and T. F. Bastos-Filho, “Teleoperation of a mobile robot through the internet,” in *Proceedings of the 43rd IEEE Midwest Symposium on Circuits and Systems (Cat. No. CH37144)*, vol. 2. IEEE, 2000, pp. 930–933.

[40] 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. 172, 2020.

[41] N. Xiong and P. Svensson, “Multi-sensor management for information fusion: issues and approaches,” *Information Fusion*, vol. 3, no. 2, pp. 163–186, 2002.

[42] J. G. Gaspar, N. Ward, M. B. Neider, J. Crowell, R. Carbonari, H. Kaczmarski, R. V. Ringer, A. P. Johnson, A. F. Kramer, and L. C. Loschky, “Measuring the useful field of view during simulated driving with gaze-contingent displays.” *Human Factors*, vol. 58, no. 4, pp. 630 – 641, 2016.

[43] R. Sekuler and K. Ball, “Visual localization: age and practice,” *J. Opt. Soc. Am. A*, vol. 3, no. 6, pp. 864–867, Jun 1986.

[44] D. L. Roenker, G. M. Cissell, K. K. Ball, V. G. Wadley, and J. D. Edwards, “Speed-of-processing and driving simulator training result in improved driving performance,” *Human factors*, vol. 45, no. 2, pp. 218–233, 2003.

[45] A. O’Shea and A. J. Woods, “Useful field of view (ufov),” in *Encyclopedia of Clinical Neuropsychology*, J. Kreutzer, J. DeLuca, and B. Caplan, Eds. Springer, 2018.

[46] M. W. Spong, “Modeling and control of elastic joint robots,” 1987.

[47] A. Albu-Schaffer, C. Ott, U. Frese, and G. Hirzinger, “Cartesian impedance control of redundant robots: Recent results with the dlr-light-weight-arms,” in *2003 IEEE International conference on robotics and automation (Cat. No. 03CH37422)*, vol. 3. IEEE, 2003, pp. 3704–3709.

[48] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, “Ros: an open-source robot operating system,” in *ICRA workshop on open source software*, vol. 3, no. 3.2. Kobe, Japan, 2009, p. 5.

[49] S. G. Hart, “Nasa-task load index (nasa-tlx); 20 years later,” in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, no. 9. Sage publications Sage CA: Los Angeles, CA, 2006, pp. 904–908.

[50] V. Schwind, P. Knierim, N. Haas, and N. Henze, “Using presence questionnaires in virtual reality,” in *Proceedings of the 2019 CHI conference on human factors in computing systems*, 2019, pp. 1–12.

[51] M. R. Longo, F. Schüür, M. P. Kammers, M. Tsakiris, and P. Haggard, “What is embodiment? a psychometric approach,” *Cognition*, vol. 107, no. 3, pp. 978–998, 2008.