Mixed Reality Human Teleoperation

David Black*
University of British Columbia

Septimiu Salcudean†
University of British Columbia

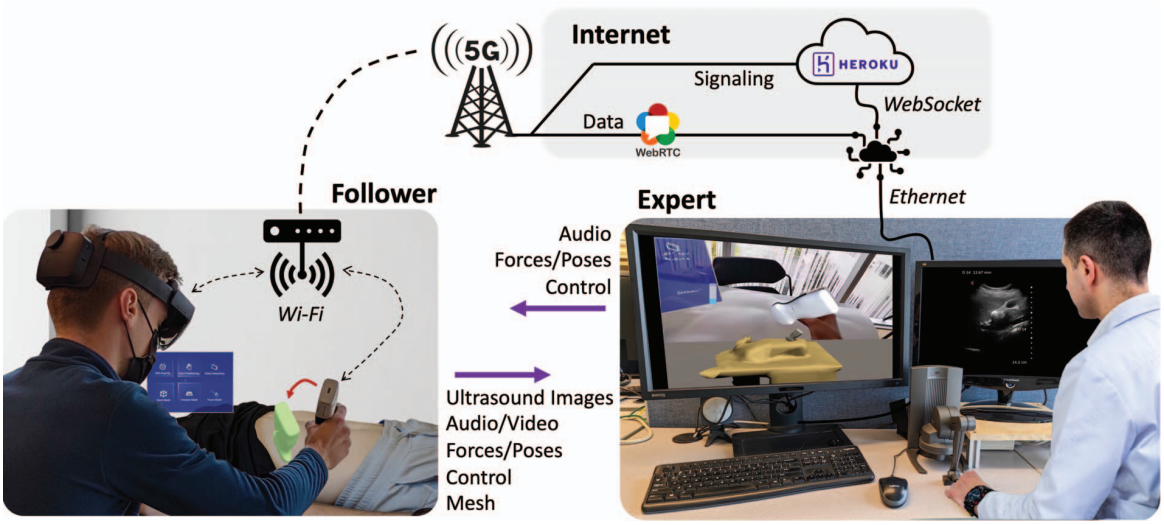

Figure 1: Overview of human teleoperation system.

Abstract

For many applications, remote guidance and telerobotics provide great advantages. For example, tele-ultrasound can bring much-needed expert healthcare to isolated communities. However, existing tele-guidance methods have serious limitations. A new concept called human teleoperation leverages mixed reality, haptics, and high-speed communication to provide tele-guidance that is more tightly coupled than existing methods yet more accessible than telerobotics. This paper provides an overview of the human teleoperation concept and its application to tele-ultrasound. The concept and its impact are discussed, the graphics, communications, controls, and haptics subsystems are explained, and results are presented that show the system's efficacy. These include tests of the communication architecture, of human performance in tracking mixed reality signals, and of human teleoperation in a limited clinical use-case. The results show good potential for teleultrasound, as well as possible other applications of human teleoperation including remote maintenance, inspection, and training.

Index Terms: Human-centered computing—Human Computer Interaction—Interaction Paradigms—Mixed / augmented reality; Human-centered computing—Human Computer Interaction—Interaction Devices—Haptic devices; Computer systems organization—Embedded and cyber-physical systems—Robotics—Robotic Control

*e-mail: dgblack@ece.ubc.ca
†e-mail: tims@ece.ubc.ca

1 Introduction

Remote guidance technologies can improve tele-medicine, inspection, maintenance, and teaching [16, 65]. For example, tele-ultrasound (tele-US) is useful for remote and under-resourced communities [25, 28]. Additionally, it is can be used in care homes for elderly patients with mobility issues [33], for COVID-19 [56, 60], for trauma assessment in ambulances [38], or for training of sonographers [15, 58]. Existing methods for tele-US consist of robotic teleoperation and audiovisual, video-conferencing-based guidance on a smartphone or tablet application.

Reviews of robotic US systems are found in [47] and more recently [53]. While one robotic tele-US system has been used in clinical trials [61], commercial success has been limited despite the robots' ability to provide precision, low latency, and haptic feedback [3, 14, 35, 51]. This is likely due to practical limitations including cost, restricted workspaces, time consuming set-up, and complex maintenance and operation. The cost is especially relevant when compared to otherwise inexpensive US devices, and makes it difficult to deploy such systems in small communities. Despite this, a large body of literature has studied autonomous robotic US [32, 63] using force-based positioning [26, 62], depth camera-based planning [24], and reinforcement learning [44]. However, guaranteed robustly safe human-robot interaction is an issue, particularly for regulatory bodies, and robotic tele-US remains relatively impractical.

On the other hand, there are several commercially available video conferencing-based mobile systems. Butterfly Network, Clarius Mobile Health Corp., and Philips use a point of care US (POCUS) device with live imaging and video conferencing available via a cloud interface on a mobile phone or tablet. Some visual guidance can be given by overlaying arrows or pointers on the US image. Though accessible and inexpensive, these systems are designed rather for quick expert review of a capable sonographer's captured images instead of teleoperation of an inexperienced novice. The resulting interaction is thus very inefficient for the latter case, leading

2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) | 979-8-3503-4839-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/VRW58643.2023.00083

to low precision and high latency.

Neither of the existing solutions is both flexible and accessible while being accurate and efficient. However, recent advances in extended reality (XR) research may solve this issue. Within the umbrella of XR, virtual reality (VR) immerses a user in a virtual environment, while augmented reality (AR) takes the real environment and adds visual information in the form of video overlays [11]. There are many definitions and classifications pertaining to the approach of augmenting the user's view, including Milgram and Kishino's "reality-virtuality continuum" [36], and more recent additions [55]. For clarity, we refer to our system as mixed reality (MR) according to the convention used by Microsoft, the manufacturer of the headset we use. Unlike in AR where overlays are applied to videos, in MR visual guides can be located within the real environment itself through the use of optically transparent headsets and waveguides [49].

The ability of MR to project 3D visual information seamlessly into the real world is the key enabling technology in a new concept we call "human teleoperation", introduced in [10], which leverages MR, haptics, and high-speed communication to bridge the gap in remote guidance techniques. In this system, a human follower is controlled as if they were a flexible, cognitive robot through an MR interface. In this way, both the input and the actuation are carried out by people, but with tight coupling, leading to latency and precision more similar to a tele-robotic system. This enables remote guidance that is more intuitive, accurate, and efficient than existing audiovisual systems, yet less expensive, more accessible, and more flexible than robotic teleoperation. The architecture and general function of human teleoperation is described in Sect. 2.

In developing this system, we follow several design objectives and requirements. To outperform existing methods, the system must be intuitive to use, efficient for procedures, and flexible for use in different procedures and locations. To achieve this, the communication system must be high throughput, low latency, and easily used in different networks and signal conditions. The visual control system should be user friendly for the follower and lead to good accuracy and little lag. Similarly, the expert should have a sensation as close as possible to carrying out the procedure in person, called teleoperation transparency, which involves visual, positional, and force feedback.

This paper summarizes conceptual, experimental, and implementation-based advances from several other papers to describe how human teleoperation meets these objectives and constitutes a promising and potentially impactful alternative to existing teleguidance methods. First, Sect. 2 explains how human teleoperation works before Sect. 3 delves into the communication architecture. Sect. 4 describes the spatial registration procedures, and Sect. 5 discusses the haptics aspects of this methodology. Experiments quantifying human MR tracking performance are outlined in Sect. 6, with some preliminary patient tests in Sect. 7. Qualitative experiences of the MR interface are discussed in Sect. 8, and several of the many interesting avenues of future work are explored in Sect. 10. Finally, a discussion of existing MR/AR guidance systems and how they differ from teleoperation is left to the end in Sect. 9, so the reader first gains a thorough understanding of human teleoperation.

2 System Architecture

In this section we describe the specific prototype system we built for tele-ultrasound. Other applications for human teleoperation are discussed in Sect. 10. A diagram of the whole system is found in Fig. 1.

In general teleoperation systems, there is a local agent, often a robot, which interacts with its environment, and a remote operator who receives feedback from the local agent and provides instructions on what actions to carry out. Traditionally, these are called the "master" and "slave" respectively, though we avoid this terminology. In human teleoperation, the local agent is a human, the "follower", while the remote operator is experienced in the task being performed, and is referred to as the "expert". For teleultrasound, the expert is a sonographer or radiologist with ultrasound experience, and the follower is an inexperienced novice, whose interactions with his/her environment constitute moving an US probe on a patient, as instructed by the expert.

The follower wears an MR headset, the HoloLens 2 (Microsoft, Redmond, WA), which projects a virtual US probe into their field of view. To perform the desired procedure, they align their real US probe with the virtual one and track it as it moves around on the patient. In this way, the desired position and orientation (pose) of the probe are achieved. The desired force is similarly reached through a visual control system, as explained in Sect. 5.

The desired pose and force of the probe are set in real time by the expert, who manipulates a haptic device (Touch X, 3D Systems, Rock Hill, SC). This is a small, 6-degree-of-freedom (DOF) serial manipulator that measures the 6-DOF pose of its pen-like handle and applies 3-DOF forces to the handle's tip for haptic feedback [34]. We replaced the Touch X conventional handle with a 3D-printed US probe-shaped end effector. As the expert grasps the end-effector and carries out his/her desired motion, the probe pose is transmitted to the virtual handle guiding the follower. The haptic device applies forces back to the expert. By pushing against them, the expert also inputs his/her desired force. This is described in detail in Sect. 5.

One of the primary objectives of the system is teleoperation transparency, or making the expert feel as if they are performing the procedure in person by matching the expert and follower positions and forces. To this end, a three or four-channel architecture is required, in which force and velocity are sent from expert to follower and vice versa at a high rate [22, 30]. Hence, the desired forces and poses are sent from expert to follower, along with an audio stream and some control packets. Conversely, the measured force and pose of the follower are returned to the expert, along with an audio stream, a video stream captured by the HoloLens which includes the virtual objects in place (called an MR capture), the live ultrasound images, control packets, and occasional mesh data. The communication system for transmitting this data is described in Sect. 3, and the meshes are discussed in Sect. 4.

In this way, the expert sees the patient, ultrasound probe, and ultrasound images live, and feels the applied forces. She/he consequently decides how to move and performs the motion on the haptic device, which updates the input signal to the follower through the visual control system, which the follower tracks. At the same time, the expert and follower are in verbal communication. Preliminary experiments with carrying out a remote ultrasound examination by a novice follower when teleoperated by an expert have shown promising results (Sect. 7).

3 Communication

As described in Sect. 2, a large amount of data must be transmitted at high speed. In particular, the throughput requirements on the follower side are outlined in Table 1. As the expert is stationary and attached via a wired connection, their throughput is of less concern.

To achieve the large throughputs required by such tight coupling between expert and follower, a high-performance communication system was developed and is described in [6, 8]. Several OSI model layers of the communication architecture are shown in Fig. 2 and Fig. 3. Since the follower side is mobile and could be installed anywhere, including a moving ambulance or a remote location, it is designed to work over mobile networks. A Sercomm 5G modem with a wired connection to a Wi-Fi router provides the interface between the mobile network and the user devices. This is equivalent to 5G CPE devices sold by Huawei, Yeastar, Oppo, and others, or to "MiFi" devices from Inseego. The HoloLens 2, US device, and force and pose sensors connect to a local Wi-Fi network, which in turn

Table 1: Worst-case uplink and downlink data throughput requirements on the follower side

Type	Uplink	Contents	Downlink	Contents
Timing	1.28 Kbps	64 bit long int × 20 Hz	3.84 Kbps	3 64 bit long ints × 20 Hz
Force	16 Kbps	3 32 bit floats + 1 64bit timestamp × 100 Hz	16 Kbps	Same as uplink
Pose	28.8 Kbps	7 32 bit floats + 1 64 bit timestamp × 100 Hz	28.8 Kbps	Same as uplink
Video	≈ 1-2 Mbps	960 × 540 px H.264 encoding, 25 Hz, Variable quality	0	No video
Audio	128 Kbps	Typical MP3 bitrate (Part of MPEG-4 stream)	128 Kbps	Same as uplink
US	4.64 Mbps	58 KB JPEG image (worst-case) × 10 Hz	0	No US
Mesh	2.3 Mbps	≈12k mesh triangles × 3 points and 3 indices × 32bit floats	0	No downlink mesh
Total	6.81 Mbps	Mesh sent rarely on demand. Peak throughput 9.11 Mbps	180 Kbps	Sum

connects to the Internet via mobile networks or directly. Conversely, the static expert side connects to the Internet via a wired connection.

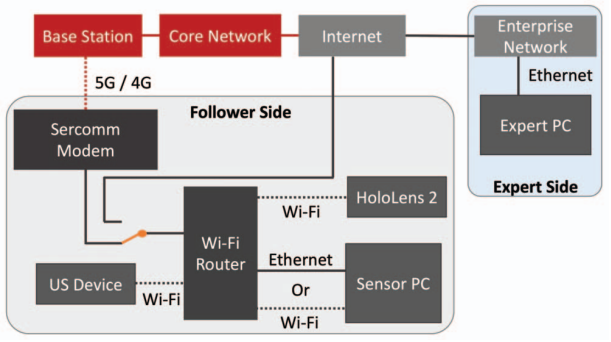

Figure 2: Physical, data link, and network layers of the communication architecture

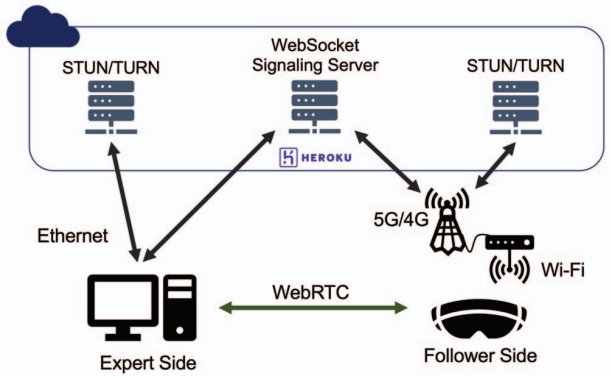

Figure 3: Session and application layers of the communication architecture. STUN: session traversal using NAT. This is a server that sends back a device's public address and connectivity information.

5G provides ultra-reliable low latency communication and enhanced mobile broadband which are well suited to this application due to the large throughput, high reliability, and low latency requirements. The University of British Columbia was the first campus in North America equipped with a 5G network, so we partnered with Rogers Communications Inc., a Canadian telecommunications company, to utilize their non-standalone sub-6 GHz 5G network, shown in red in Fig. 2.

Built on this infrastructure, the communication system was designed to minimize latency using the WebRTC (Web Real Time Communication) standard. As shown in Fig. 3, WebRTC is a peer-to-peer architecture which eliminates server-related delays. It is built on stream control transmission protocol [2] and real time transport protocol [1], both of which are related to the user datagram protocol (UDP) which prioritizes speed over reliability. Instead of retransmission upon packet loss, which adds latency, packets are sent at a sufficient rate that lost ones are quickly replaced and local consistency checks are in place. To establish connectivity initially, WebRTC uses Session Description Protocol (SDP) and Interactive Connectivity Establishment (ICE) to find an optimal connection between two peers over any network and through most router NAT (Network Address Translation) schemes or firewalls. Both peers determine their own information from a STUN server, and then exchange this SDP over a signaling server which we implemented using Python WebSockets and host on a cloud platform called Heroku. Using this information, the peer-to-peer connection is established.

The signaling is encrypted locally, and the signaling server requires authentication. Similarly, WebRTC is built upon Datagram Transport Layer Security (DTLS), so all communication is encrypted and secure, which is important for medical applications.

In [6] we carried out tests of the communication performance over various networks, in different signal conditions, and with different throughput values to quantify the response to different conditions. We also experimented with configurations including retransmission, packet ordering guarantees, channel splitting, and queuing. Some of the key results are shown in Table 2. We found that the communication performed sufficiently well over 4G or 5G with good signal conditions up to about 5 Mbps continuous throughput, and well beyond this for WiFi or Ethernet. Typical round trip times (approximately two times latency) for 5G were 25 to 50 ms which is appropriate for teleoperation with direct force feedback in the case of a relatively soft contact environment like a patient.

4 Spatial Registration

The expert is presented with a live video stream from the follower's perspective, and the ultrasound images of the patient. Based on these, he/she decides how to move. If the expert sees the follower's US probe positioned too far left in the follower's view, it is natural for them to move their probe to the right. However, the follower coordinate frame is positioned by the HoloLens and varies every time the application is started, even in the same location. Therefore, a registration between the expert and follower spaces is required to ensure that directions in both frames correspond intuitively.

Similarly, the actual pose of the ultrasound device is measured using an electromagnetic tracking system (driveBAY, NDI, Waterloo, ON). A small sensing element embedded on the probe measures the field from an electromagnetic transmitter to determine the 6-DOF pose. To compare this to the expert's desired pose, the reading must be transformed from the transmitter frame to the HoloLens frame.

Furthermore, the HoloLens 2 continuously captures a 3D spatial mesh of its environment for SLAM purposes. This mesh can be accessed and sent to the expert side to provide 3D visual feedback as well as a haptic interaction, described in Sect. 5. This process is also discussed in detail in [9].

To extract the mesh of only the patient, not the rest of the room,

Table 2: Round trip time ($RTT \pm std$, ms) versus data throughput in good signal conditions for various networks.

Throughput	Ethernet	WiFi	4G LTE	5G NR
1.28 Kbps	1.07 ± 0.57	5.80 ± 3.30	38.41 ± 6.63	26.95 ± 7.72
46.08 Kbps	0.94 ± 0.61	5.90 ± 2.77	38.77 ± 7.92	27.67 ± 6.21
2.17 Mbps	0.93 ± 0.59	5.87 ± 1.75	43.49 ± 30.18	39.61 ± 6.14
4.97 Mbps	-	-	52.30 ± 59.30	47.64 ± 22.81
6.81 Mbps	1.07 ± 0.88	7.82 ± 9.90	66.58 ± 123.00	70.44 ± 78.29

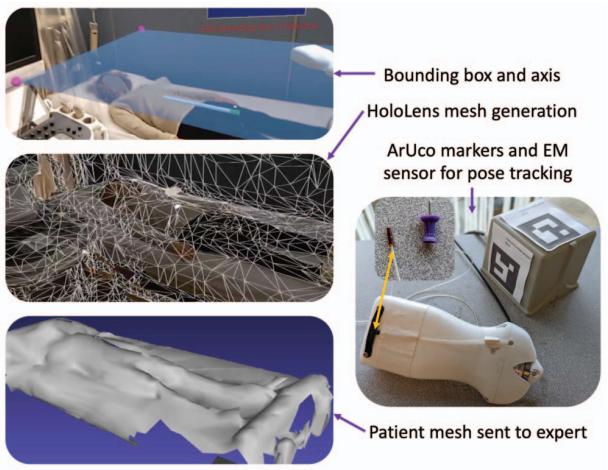

Figure 4: Aspects of expert-follower spatial registration. The US probe is a 3D printed dummy used for testing of the instrumentation. The electromagnetic (EM) transmitter is the box with two ArUco markers.

Figure 5: Color and error-bar force rendering schemes.

and to enable the spatial registration, the position and orientation of the patient and electromagnetic transmitter must be determined. To this end we have developed two methods. First, the follower is presented with a virtual bar which they can align with the axis of the patient, before moving three virtual markers to position a bounding box around the patient [9]. To speed up this procedure, we placed ArUco markers [17] in four corners of the test surface, as well as on the transmitter. The HoloLens automatically determines the poses of the markers and thus finds the patient bounding box, orientation, and the transmitter transform. These steps are shown in Fig. 4.

5 Haptics

We have so far discussed the visual pose control method on the follower side. However, force is almost equally important in ultrasound, determining what anatomical structures are visible and ensuring they are not too deformed. Moreover, sonographers usually look almost exclusively at the US images and rely on their sense of touch to move the probe on the patient. Hence, both controlling the applied force of the follower and feeding back that force to the expert are essential in human tele-operated ultrasound.

Necessary for both of these topics is force sensing at the ultrasound probe. This can be achieved by fabricating a shell around the US probe, connected to the probe using an off-the-shelf force sensor [18, 21, 40, 51, 54]. However, this is bulky and heavy. We are developing an alternative method, but for preliminary testing of our algorithms for force feedback we have 3D printed a dummy US probe with an ATI Nano25 force/torque sensor embedded in the tip.

To generate an error signal, the measured force is subtracted from the expert's desired force applied to the haptic device. This error signal is then displayed visually for the follower so they can move to minimize it and thus match the desired force. The visualization could involve an arrow that grows, shrinks and changes direction, or a second virtual probe, offset by an amount proportional to the force error and the patient tissue stiffness, in the direction of the force error. As the follower must track the virtual probe's pose, however, it is beneficial to have them focus exclusively on the probe. Additionally, the desired force is usually normal to the patient tissue, as the slippery ultrasound gel does not allow for large transverse forces. To avoid excessive cognitive load, therefore, we tested two different force rendering schemes that change only the virtual probe itself and do not contain direction information.

The two rendering schemes are color and error-bar and are shown in Fig. 5. In the former, the ultrasound probe color varies smoothly between blue (follower should apply more force), green (force error is small), and red (follower should apply less force). Similarly in the latter, the error bar grows toward the patient or away and changes color depending on the force error. In tests described in Sect. 6, we found that the error-bar rendering was superior, likely because it was easier to resolve errors close to zero using the error-bar than by differentiating slightly blueish green from slightly reddish green.

The force sensor can also be used for force feedback to the expert in one of several teleoperation configurations. These include a 2-channel force-position [13, 20] or dual-hybrid force/position [48] controller, a 3-channel system with local feedback [23], or a 4-channel bilateral parallel force/position [22] or matched impedance [52] controller. The problem with many of these approaches is that performance and stability are limited by time delays. In human teleoperation, delays stem not only from the communication system, which is relatively fast, but also from the reaction time of the follower person, which can be substantial. Wave variable-based schemes have been proposed to maintain stability in time-delayed force-reflecting teleoperation [43], but position tracking is sometimes sluggish and wave reflections can be disorienting [66].

Many papers explore robust time-delayed teleoperation, but our system presents an opportunity for a potentially simpler approach. As explained in [9], a 3D mesh of the patient, captured by the HoloLens 2 and transmitted to the expert, can be rendered as a virtual fixture on the haptic device. When the expert displaces the haptic device into the region delineated by the mesh, it is met with an opposing force proportional to configurable stiffness and damping

coefficients. Using the force and pose sensing of the US probe, it is possible to estimate the impedance parameters of the patient tissue [37]. The estimated parameters can be fed back along with the patient mesh at a much lower rate than forces would have to be, and as they are relatively constant, a temporary packet delay or loss in communication would have no great effect. This approach, with constant approximate parameters, was used in early tests of the system described in Sect. 7.

6 Human Performance

To evaluate the feasibility of human teleoperation, we carried out experiments of human pose and force tracking ability, and quantified the responses of the human and visual control system [7]. The tests involved 11 participants who ranged from 20 to 64 years of age (mean 32) and various backgrounds including law, medicine, engineering, and architecture. As with other control systems, we performed step and frequency response tests and arbitrary series of motions to determine the reaction time, overshoot, settling time, steady-state error, frequency dependence of relative phase and magnitude, RMS tracking error, and tracking lag for position, orientation, and force. Each test was performed with both force rendering methods and two different pose renderings.

To carry out these tests, the desired motion sequence was rendered using the visual control system, and the subject's response was measured using a force sensor (ATI Nano25) and electromagnetic pose sensor (NDI driveBAY) embedded on a 3D printed ultrasound probe dummy held by the subject. For the position step responses, the virtual probe jumped back and forth between positions 10 cm apart, holding each position for 5 seconds. The position changed each step so the user could not predict the motion. Similarly for force, the desired force suddenly increased or decreased every 5 seconds. For the frequency response, the subject tracked a periodic signal which increased in frequency every 5 oscillations. The recorded signals were afterward segmented into the constant frequency portions, and the relative gain and phase lag between the desired and measured signal were computed as a function of frequency.

The step and frequency response tests are shown in Fig. 6, and example force, position, and orientation tracking tests are found in Fig. 7. The key results of these tests are outlined in Table 3. While extensive discussion is found in [7], these results show that tracking both pose and force separately or simultaneously is feasible, with relatively small tracking error and lag of about 0.35 seconds. The worst case reaction time is to a step change in pose or force and varies based on the step amplitude, but remains well under 1 second. Similarly, the frequency response depends on the input signal's amplitude but for good performance the human force and pose tracking are limited to about 0.25 Hz and 1 Hz respectively. This is slower than the human hand's force bandwidth presented in [59], though in the same order of magnitude, showing that the actuations are limited by cognitive rather than motor factors.

All of these results point to excellent performance potential for human teleoperation. The motions in ultrasound, for example, are much less demanding than in the tests. Additionally, they indicate that careful consideration of time delays is essential for force feedback, as discussed in Section 5, because the time delays imposed by the human response time vary substantially and can be up to a large fraction of a second.

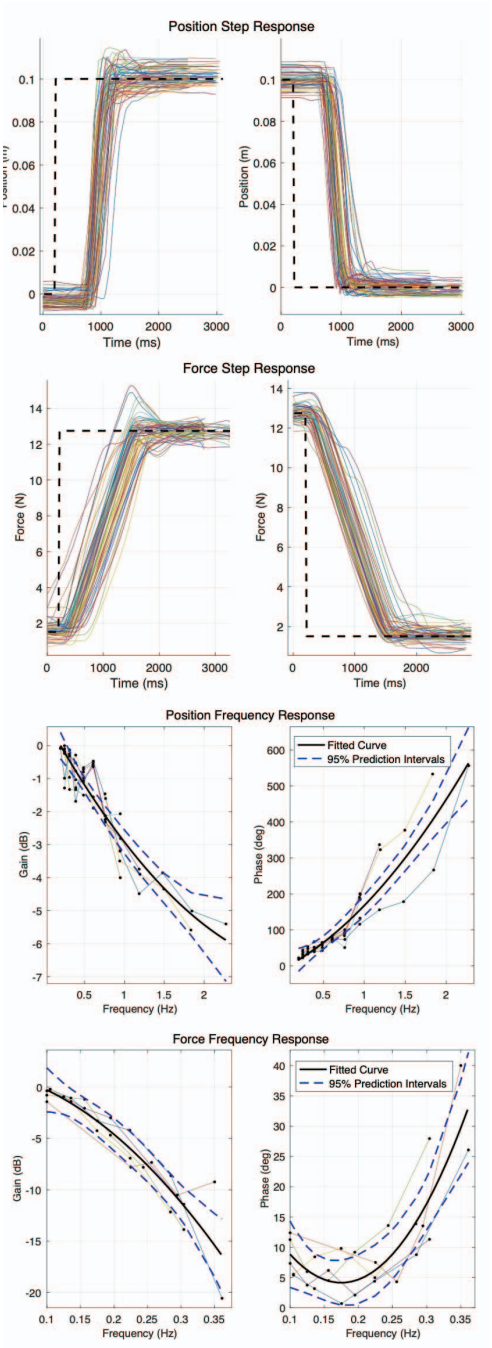

Figure 6: MR human tracking performance step and frequency response test results. The dotted step response is the desired signal. The other lines are the subjects' measured responses. The subject frequency responses are plotted, as is a fitted second degree polynomial (black solid line), and the 95% confidence interval of the fit (dashed lines). Numerical results from these tests are found in Table 3.

7 Patient Tests

Though the human performance tests in Sect. 6 and the communication experiments in Sect. 3 show good potential for human teleoperation, neither evaluates the performance of the system in a practical clinical setting. The human teleoperation system should improve precision, efficiency, and completion time of US procedures compared to existing audiovisual guidance methods, and improve set-up time, ease-of-use, and accessibility compared to robotic systems. While larger-scale trials with patients as well as comparative studies between human and robot teleoperation are planned future work, we have completed a preliminary, small-scale study of four healthy volunteers and a physician with extensive US experience on two example procedures in [9]. The volunteer followers had no

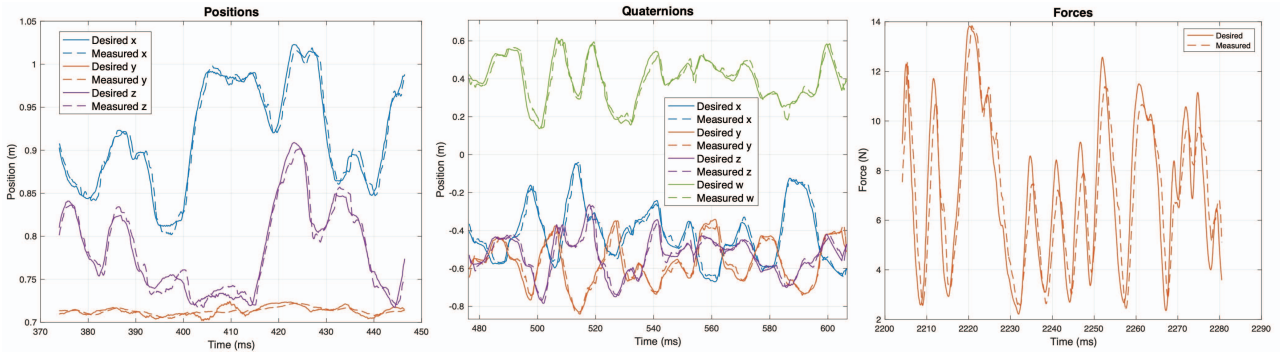

Figure 7: Results from one example human performance test. We see good tracking in all axes of position and orientation. The force with error-bar rendering also exhibits low tracking error. Solid lines are desired values while dotted lines are the measured response.

Table 3: Human performance test key results from [7]. This shows that tracking pose or force is cognitively easier than both at once, but that when tracking both simultaneously the speed and accuracy are sufficient for relatively high-performing teleoperation.

	Force	Pose
Single Parameter (pose or force):		
Continuous tracking lag	255 ± 118.88 ms	346.23 ± 118.15 ms
Continuous tracking rms error	0.99 ± 0.29 N	6.24 ± 1.93 mm, $5.93 \pm 1.85°$
Dual Parameter (pose and force simultaneously):		
Continuous tracking lag	345.5 ± 87.60 ms	345.5 ± 87.60 ms
Continuous tracking rms error	1.25 ± 0.33 N	8.5 ± 1.4 mm, $7.27 \pm 2.25°$
Step Responses:	*(10 N step)*	*(10 cm step)*
Reaction time to step changes	171.5 ± 85.9 ms	628.3 ± 102.3 ms
Steady state error	0.26 ± 0.16 N	2.8 ± 2.1 mm
Frequency Responses:	*(10 N magnitude)*	*(10 cm magnitude)*
Max. frequency for good tracking	0.25 Hz	1 Hz
For smaller input magnitudes, max. frequency	decreases	increases

ultrasound experience.

Both procedures involved quantitative endpoints and were carried out three times: once directly by the physician, once using the Clarius video conferencing interface, and once using human teleoperation. Each volunteer follower carried out one procedure using human teleoperation, and the other procedure using audiovisual guidance, to avoid bias from learning effects if they performed the same procedure twice. It is unlikely that the expert learned or improved much between trials as he was already very experienced. The direct US measurements were taken as a reference, and the measurement error and completion time of the different methods were compared. The results are outlined in Table 4. Despite the small sample size, it was found that human teleoperation was faster and more accurate than audiovisual guidance ($p = 0.052$) and not much slower than direct teleoperation.

Table 4: Results from preliminary volunteer study with expert sonographer. Each procedure and method has a completion time and measurement percent error. Kidney values are displayed as transverse $\times$ craniocaudal dimension error. This shows that human teleoperation is faster and more precise than existing audiovisual guidance ($p = 0.052$).

	Kidney	Vena Cava
Direct	$1{:}28 \pm 0{:}21$	$0{:}42 \pm 0{:}04$
	-	-
Audiovisual	$4{:}13 \pm 3{:}58$	$3{:}55 \pm 0{:}25$
	$3.5\% \times 14\%$	12.8%
Human Teleop.	$1{:}36 \pm 0{:}23$	$0{:}49 \pm 0{:}02$
	$3.5\% \times 4.6\%$	8.3%

8 Follower MR Experience

In the patient tests as well as numerous demos given to technical, non-technical, medical, and non-medical people of all ages, feedback has uniformly been that tracking the virtual tool is very easy and intuitive. All participants were able to track motions within seconds of putting on the HoloLens 2, sometimes with no specific instruction. The sole difficulty was an occasional initial incorrect depth perception, leading to the follower person holding the real probe closer to the headset than the virtual probe was. After a single prompting telling the user the approximate location of the virtual probe, this was rectified in all cases and there were no recurrences.

In the human performance tests [7], users filled out a questionnaire regarding their experience. The results are shown in Table 5. It was found that again tracking both force and pose is very intuitive, though requires some focus to perform well. In dimmer lighting conditions, the holographic US probe can occlude the real one, thus blocking visual positioning feedback for the follower. In this case, using a virtual probe with transparent sections or decreasing its overall opacity was sometimes useful, depending on user preference.

Table 5: User questionnaire scores out of 5 for the MR follower interface [7] (5 = a lot / demanding; 1 = a little / easy / intuitive)

Participants who became dizzy	0
Preference for error-bar over color force	4.55 ± 0.52
Mental demand of force tracking	3.36 ± 0.5
Mental demand of pose tracking	2.18 ± 0.6
Intuitiveness of tracking interface	4.82 ± 0.4
Physical demand	2.36 ± 0.8

9 Comparison to Existing AR/MR Guidance

Human teleoperation bridges the gap between robotic teleoperation and existing video conference-based systems, as described in the Introduction. These systems involve some use of AR overlays, including arrows and pointers drawn on the images. MR has also been used extensively in other fields to guide tasks. However, these uses differ fundamentally from the human teleoperation concept.

The idea of using mixed reality to overlay ultrasound images into the radiologist's field of view was introduced in 1992 [5]. The same group later performed a randomized trial and found that MR guidance improved accuracy in reaching a target during needle biopsies, decreasing mean deviation from 2.48 mm to 1.62 mm [50]. Similar work has been used for ultrasound-guided needle biopsies on more modern MR headsets such as the Microsoft HoloLens [19]. In these systems, the US images are projected into the imaging plane of the US device, and the needle is extrapolated linearly to show its predicted trajectory. This effectively allows the radiologist to "see inside" the patient and better aim their needle. In [4], the overlay was displayed on a monitor, which is less immersive and intuitive than displaying it in an MR headset or 3D display as found in the surgeon console of the da Vinci surgical system (Intuitive Surgical, Sunnyvale, CA).

The concept of overlaying medical images and 3D volumetric models of anatomy in position on a patient has been applied to other branches of medicine including laparoscopic surgery [29, 41], robot-assisted surgery [27], and even treatment of depression through transcranial magnetic stimulation [31]. A very attractive implementation for open resection of liver metastases is presented in [45]. In [12], the images and models are not overlaid directly, but available for the surgeon to drag into whichever position is convenient for them to look at. This is due in part to the difficulty of registering pre-operative images to intra-operative anatomy, which tends to move and deform. Some papers attempt to deal with this by registering and deforming the pre-operative images according to US images captured intra-operatively [57]. A different approach is to use fluorescent markers [64].

In addition to image overlays which mostly extend a physician's view into otherwise obscured anatomy, some systems for manufacturing include static labels or pointers [42]. The authors of [39] developed a framework to gather information from a scene and create an MR application offline which contains visual instructions that can be overlaid onto the scene.

What every one of these applications has in common is a relatively static overlay of images or pointers intended to extend a user's vision or indicate a target to reach. To our knowledge, no other system exists in which a virtual guiding tool, in our case an US probe, is controlled in real time by a remote person to guide the user in a hand-over-hand manner with tight coupling. Even recent advances in "holoportation" focus on creating a natural social interaction rather than providing dynamic guidance [46]. By approaching the MR guidance from a controls and tele-robotics perspective, human teleoperation enables very flexible remote control of procedures such as US exams. In this way, unlike most other MR guidance research, human teleoperation does not necessarily enhance the user's sensory capabilities by overlaying images and guides, but rather transports an expert's knowledge and skill into a remote location where it is needed.

10 Future Work

In addition to the further patient tests and comparison to robotics mentioned in Sect. 7, this project presents many avenues for further research. For example, the human teleoperation concept can be applied to other fields beyond ultrasound, including any application where tightly coupled hand-over-hand guidance is needed. For example, this could include maintenance, inspection, and training. Furthermore, because the actuations are ultimately carried out by a human, not a computer or robot, this presents an interesting application for artificial intelligence. The human expert could be replaced by an AI agent, trained through learning from demonstration or reinforcement learning, for example, to guide procedures autonomously on demand. This is not limited by the concerns of robustly safe human-robot interaction that affect autonomous robotics.

In terms of communications, it is possible to leverage further aspects of 5G, including exporting costly computations used for computer vision or AI to proximal edge servers with minimal latency. In addition, we plan to test the communication over mm-wave band 5G, which promises the greatest increase in throughput.

Furthermore, the spatial registration process could be made faster and easier using computer vision to segment the patient automatically and track the US probe visually using the HoloLens. For improved haptics, local convex meshes could be approximated and streamed in real time to remove artefacts such as splinters and to take into account a moving patient. Different visual control architectures have the potential to reduce cognitive load and thus remove the decreased performance in Table 3 when tracking pose and force simultaneously compared to one at a time. Moreover, delay-robust force feedback should be implemented and tested, potentially in a shared architecture with mesh and impedance feedback.

Finally, future work will evaluate alternative MR headsets such as the Nreal Light, Magic Leap 2, HoloBoard, or Creal LightField display.

11 Conclusion

This paper has described a novel system for teleoperating a human with precision and latency similar to that of a robot through a mixed reality interface. The work of several papers is summarized and key results are shown which demonstrate the performance and feasibility of human teleoperation. Not only does this system bring up interesting basic science questions about mixed reality, teleoperation, and human computer interaction, but it has the potential to impact many remote communities that otherwise have difficult access to healthcare. It may additionally be of use in myriad industrial applications.

Acknowledgments

We gratefully acknowledge scholarship support from the Vanier Canada Graduate Scholarships program, infrastructure support from CFI and funding support from NSERC and the Charles Laszlo Chair in Biomedical Engineering, as well as infrastructure, technical, and funding support for the communication system from Rogers Communications and MITACS.

References

[1] Rfc 1889: Rtp: A transport protocol for real-time applications, Jan 1996.

[2] Rfc 4960: Stream control transmission protocol, Sept 2007.

[3] M. Akbari, J. Carriere, T. Meyer, R. Sloboda, S. Husain, N. Usmani, and M. Tavakoli. Robotic ultrasound scanning with real-time image-based force adjustment: Quick response for enabling physical distancing during the covid-19 pandemic. *Frontiers in Robotics and AI*, 8, 2021. doi: 10.3389/frobt.2021.645424

[4] G. Ameri, J. S. Baxter, D. Bainbridge, T. M. Peters, and E. Chen. Mixed reality ultrasound guidance system: a case study in system development and a cautionary tale. *International journal of computer assisted radiology and surgery*, 13(4):495–505, 2018.

[5] M. Bajura, H. Fuchs, and R. Ohbuchi. Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. *ACM SIGGRAPH Computer Graphics*, 26(2):203–210, 1992.

[6] D. Black, D. Andjelic, and S. Salcudean. Evaluation of communication and human response latency for (human) teleoperation. *TechRxiv*, Oct 2022. doi: 10.36227/techrxiv.21432009.v1

[7] D. Black, H. Moradi, and S. Salcudean. Human-as-a-robot performance in augmented reality teleultrasound. *TechRxiv*, Oct 2022. doi: 10.36227/techrxiv.21432012.v1

[8] D. Black and S. Salcudean. A mixed reality system for human teleoperation in tele-ultrasound. In *Hamlyn symposium for medical robotics*, pp. 91–92, 2022.

[9] D. Black, Y. O. Yazdi, A. H. Hadi Hosseinabadi, and S. Salcudean. Human teleoperation-a haptically enabled mixed reality system for teleultrasound. *Human Computer Interaction (Accepted)*, 2022.

[10] D. Black, Y. O. Yazdi, A. H. H. Hosseinabadi, and S. Salcudean. Human teleoperation - a haptically enabled mixed reality system for teleultrasound. *Human Computer Interaction*, 2022 (In press).

[11] J. Carmigniani and B. Furht. Augmented reality: an overview. *Handbook of augmented reality*, pp. 3–46, 2011.

[12] J. Cartucho, D. Shapira, H. Ashrafian, and S. Giannarou. Multimodal mixed reality visualisation for intraoperative surgical guidance. *International journal of computer assisted radiology and surgery*, 15(5):819–826, 2020.

[13] R. Daniel and P. R. McAree. Fundamental limits of performance for force reflecting teleoperation. *The international journal of robotics research*, 17(8):811–830, 1998.

[14] C. Delgorge, F. Courrèges, L. A. Bassit, C. Novales, C. Rosenberger, N. Smith-Guerin, C. Brù, R. Gilabert, M. Vannoni, G. Poisson, et al. A tele-operated mobile ultrasound scanner using a light-weight robot. *IEEE transactions on information technology in biomedicine*, 9(1):50–58, 2005.

[15] A. E. Drake, J. Hy, G. A. MacDougall, B. Holmes, L. Icken, J. W. Schrock, and R. A. Jones. Innovations with tele-ultrasound in education sonography: the use of tele-ultrasound to train novice scanners. *The ultrasound journal*, 13(1):1–8, 2021.

[16] S. N. Gajarawala and J. N. Pelkowski. Telehealth benefits and barriers. *The Journal for Nurse Practitioners*, 17(2):218–221, 2021.

[17] 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*, 47(6):2280–2292, 2014.

[18] M. W. Gilbertson and B. W. Anthony. An ergonomic, instrumented ultrasound probe for 6-axis force/torque measurement. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 140–143. IEEE, 2013.

[19] L. Groves, N. Li, T. M. Peters, and E. Chen. Towards a first-person perspective mixed reality guidance system for needle interventions. *Journal of Imaging*, 8(1):7, 2022.

[20] B. Hannaford and R. Anderson. Experimental and simulation studies of hard contact in force reflecting teleoperation. In *Proceedings. 1988 IEEE International Conference on Robotics and Automation*, pp. 584–589. IEEE, 1988.

[21] M. O. Harris-Love, C. Ismail, R. Monfaredi, H. J. Hernandez, D. Pennington, P. Woletz, V. McIntosh, B. Adams, and M. R. Blackman. Interrater reliability of quantitative ultrasound using force feedback among examiners with varied levels of experience. *PeerJ*, 4:e2146, 2016.

[22] K. Hashtrudi-Zaad and S. E. Salcudean. Transparency in time-delayed systems and the effect of local force feedback for transparent teleoperation. *IEEE Transactions on Robotics and Automation*, 18(1):108–114, 2002.

[23] K. Hastrudi-Zaad and S. E. Salcudean. On the use of local force feedback for transparent teleoperation. In *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No. 99CH36288C)*, vol. 3, pp. 1863–1869. IEEE, 1999.

[24] Q. Huang, J. Lan, and X. Li. Robotic arm based automatic ultrasound scanning for three-dimensional imaging. *IEEE Transactions on Industrial Informatics*, 15(2):1173–1182, 2018.

[25] K. Jemal, D. Ayana, F. Tadesse, M. Adefris, M. Awol, M. Tesema, B. Dagne, S. Abeje, A. Bantie, M. Butler, et al. Implementation and evaluation of a pilot antenatal ultrasound imaging programme using tele-ultrasound in ethiopia. *Journal of Telemedicine and Telecare*, p. 1357633X221115746, 2022.

[26] Z. Jiang, M. Grimm, M. Zhou, Y. Hu, J. Esteban, and N. Navab. Automatic force-based probe positioning for precise robotic ultrasound acquisition. *IEEE Transactions on Industrial Electronics*, 68(11):11200–11211, 2020.

[27] M. Kalia, A. Avinash, N. Navab, and S. Salcudean. Preclinical evaluation of a markerless, real-time, augmented reality guidance system for robot-assisted radical prostatectomy. *International Journal of Computer Assisted Radiology and Surgery*, 16(7):1181–1188, 2021.

[28] T. Kaneko, N. Kagiyama, Y. Nakamura, T. Hirasawa, A. Murata, R. Morimoto, S. Miyazaki, and T. Minamino. Effectiveness of real-time teleultrasound for echocardiography in resource-limited medical teams. *Journal of Echocardiography*, 20(1):16–23, 2022.

[29] A. Kolagunda, S. Sorensen, S. Mehralivand, P. Saponaro, W. Treible, B. Turkbey, P. Pinto, P. Choyke, and C. Kambhamettu. A mixed reality guidance system for robot assisted laparoscopic radical prostatectomy. In *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*, pp. 164–174. Springer, 2018.

[30] D. A. Lawrence. Stability and transparency in bilateral teleoperation. *IEEE transactions on robotics and automation*, 9(5):624–637, 1993.

[31] C. Leuze, G. Yang, B. Hargreaves, B. Daniel, and J. A. McNab. Mixed-reality guidance for brain stimulation treatment of depression. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 377–380. IEEE, 2018.

[32] K. Li, Y. Xu, and M. Q.-H. Meng. An overview of systems and techniques for autonomous robotic ultrasound acquisitions. *IEEE Transactions on Medical Robotics and Bionics*, 3(2):510–524, 2021.

[33] P. J. Mariani and J. A. Setla. Palliative ultrasound for home care hospice patients. *Academic Emergency Medicine*, 17(3):293–296, 2010.

[34] T. H. Massie, J. K. Salisbury, et al. The phantom haptic interface: A device for probing virtual objects. In *Proceedings of the ASME winter annual meeting, symposium on haptic interfaces for virtual environment and teleoperator systems*, vol. 55, pp. 295–300. Chicago, IL, 1994.

[35] K. Mathiassen, J. E. Fjellin, K. Glette, P. K. Hol, and O. J. Elle. An ultrasound robotic system using the commercial robot ur5. *Frontiers in Robotics and AI*, 3:1, 2016.

[36] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994.

[37] S. Misra and A. M. Okamura. Environment parameter estimation during bilateral telemanipulation. In *2006 14th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pp. 301–307. IEEE, 2006.

[38] J. Montoya, S. Stawicki, D. C. Evans, D. Bahner, S. Sparks, R. Sharpe, and J. Cipolla. From fast to e-fast: an overview of the evolution of ultrasound-based traumatic injury assessment. *European journal of trauma and emergency surgery*, 42(2):119–126, 2016.

[39] D. Mourtzis, V. Zogopoulos, and E. Vlachou. Augmented reality application to support remote maintenance as a service in the robotics industry. *Procedia Cirp*, 63:46–51, 2017.

[40] G. P. Mylonas, P. Giataganas, M. Chaudery, V. Vitiello, A. Darzi, and G.-Z. Yang. Autonomous efast ultrasound scanning by a robotic manipulator using learning from demonstrations. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3251–3256. IEEE, 2013.

[41] N. Navab, J. Traub, T. Sielhorst, M. Feuerstein, and C. Bichlmeier. Action-and workflow-driven augmented reality for computer-aided medical procedures. *IEEE computer graphics and applications*, 27(5):10–14, 2007.

[42] A. Y. Nee, S. Ong, G. Chryssolouris, and D. Mourtzis. Augmented reality applications in design and manufacturing. *CIRP annals*, 61(2):657–679, 2012.

[43] G. Niemeyer and J.-J. Slotine. Stable adaptive teleoperation. *IEEE Journal of oceanic engineering*, 16(1):152–162, 1991.

[44] G. Ning, X. Zhang, and H. Liao. Autonomic robotic ultrasound imaging system based on reinforcement learning. *IEEE Transactions on Biomedical Engineering*, 68(9):2787–2797, 2021.

[45] D. Ntourakis, R. Memeo, L. Soler, J. Marescaux, D. Mutter, and P. Pessaux. Augmented reality guidance for the resection of missing colorectal liver metastases: an initial experience. *World journal of surgery*, 40(2):419–426, 2016.

[46] S. Orts-Escolano, C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Degtyarev, D. Kim, P. L. Davidson, S. Khamis, M. Dou, et al. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th annual symposium on user interface software and technology*, pp. 741–754, 2016.

[47] A. M. Priester, S. Natarajan, and M. O. Culjat. Robotic ultrasound systems in medicine. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 60(3):507–523, 2013.

[48] C. Reboulet, Y. Plihon, and Y. Briere. Interest of the dual hybrid control scheme for teleoperation with time delays for proceeding of iser'95. In *Experimental Robotics IV*, pp. 498–506. Springer, 1997.

[49] S. Rokhsaritalemi, A. Sadeghi-Niaraki, and S.-M. Choi. A review on mixed reality: Current trends, challenges and prospects. *Applied Sciences*, 10(2):636, 2020.

[50] M. Rosenthal, A. State, J. Lee, G. Hirota, J. Ackerman, K. Keller, E. D. Pisano, M. Jiroutek, K. Muller, and H. Fuchs. Augmented reality guidance for needle biopsies: an initial randomized, controlled trial in phantoms. *Medical Image Analysis*, 6(3):313–320, 2002.

[51] S. E. Salcudean, G. Bell, S. Bachmann, W.-H. Zhu, P. Abolmaesumi, and P. D. Lawrence. Robot-assisted diagnostic ultrasound–design and feasibility experiments. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 1062–1071. Springer, 1999.

[52] S. E. Salcudean, K. Hashtrudi-Zaad, S. Tafazoli, S. P. DiMaio, and C. Reboulet. Bilateral matched impedance teleoperation with application to excavator control. *IEEE Control Systems Magazine*, 19(6):29–37, 1999.

[53] S. E. Salcudean, H. Moradi, D. G. Black, and N. Navab. Robot-assisted medical imaging: A review. *Proceedings of the IEEE*, 2022.

[54] T. Schimmoeller, R. Colbrunn, T. Nagle, M. Lobosky, E. E. Neumann, T. M. Owings, B. Landis, J. E. Jelovsek, and A. Erdemir. Instrumentation of off-the-shelf ultrasound system for measurement of probe forces during freehand imaging. *Journal of biomechanics*, 83:117–124, 2019.

[55] R. Skarbez, M. Smith, and M. Whitton. Revisiting milgram and kishino's reality-virtuality continuum. front. *Presence and Beyond: Evaluating User Experience in AR/MR/VR*, p. 8, 2022.

[56] N. Smallwood, A. Walden, P. Parulekar, and M. Dachsel. Should point-of-care ultrasound become part of healthcare worker testing for covid? *Clinical Medicine*, 20(5):486, 2020.

[57] H. Song, H. Moradi, B. Jiang, K. Xu, Y. Wu, R. H. Taylor, A. Deguet, J. U. Kang, S. E. Salcudean, and E. M. Boctor. Real-time intraoperative surgical guidance system in the da vinci surgical robot based on transrectal ultrasound/photoacoustic imaging with photoacoustic markers: an ex vivo demonstration. *IEEE Robotics and Automation Letters*, 2022.

[58] N. J. Soni, J. S. Boyd, G. Mints, K. C. Proud, T. P. Jensen, G. Liu, B. K. Mathews, C. K. Schott, L. Kurian, C. M. LoPresti, et al. Comparison of in-person versus tele-ultrasound point-of-care ultrasound training during the covid-19 pandemic. *The ultrasound journal*, 13(1):1–7, 2021.

[59] L. Stocco and S. E. Salcudean. A coarse-fine approach to force-reflecting hand controller design. In *Proceedings of IEEE International Conference on Robotics and Automation*, vol. 1, pp. 404–410. IEEE, 1996.

[60] C. Uschnig, F. Recker, M. Blaivas, Y. Dong, and C. F. Dietrich. Tele-ultrasound in the era of covid-19: A practical guide. *Ultrasound in Medicine & Biology*, 2022.

[61] P. Vieyres, G. Poisson, and et al. A tele-operated robotic system for mobile tele-echography: The otelo project. In *M-health*, pp. 461–473. Springer, 2006.

[62] S. Virga, O. Zettinig, M. Esposito, K. Pfister, B. Frisch, T. Neff, N. Navab, and C. Hennersperger. Automatic force-compliant robotic ultrasound screening of abdominal aortic aneurysms. In *2016 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp. 508–513. IEEE, 2016.

[63] S. Wang, J. Housden, Y. Noh, D. Singh, and A. Singh. Robotic-assisted ultrasound for fetal imaging: Evolution from single-arm to dual-arm system. In *Annual Conference Towards Autonomous Robotic Systems*, vol. 11650, pp. 27–38, July 2019.

[64] E. Wild, D. Teber, D. Schmid, T. Simpfendörfer, M. Müller, A.-C. Baranski, H. Kenngott, K. Kopka, and L. Maier-Hein. Robust augmented reality guidance with fluorescent markers in laparoscopic surgery. *International journal of computer assisted radiology and surgery*, 11(6):899–907, 2016.

[65] T. Yamamoto, M. Otsuki, H. Kuzuoka, and Y. Suzuki. Tele-guidance system to support anticipation during communication. *Multimodal Technologies and Interaction*, 2(3):55, 2018.

[66] Y. Ye and P. X. Liu. Improving trajectory tracking in wave-variable-based teleoperation. *IEEE/ASME Transactions on Mechatronics*, 15(2):321–326, 2009.