==

Human Teleoperation - A Haptically Enabled Mixed Reality System for Teleultrasound

ARTICLE HISTORY
Compiled January 25, 2025

ABSTRACT

Current teleultrasound methods include audiovisual guidance and robotic teleoperation, which constitute tradeoffs between precision and latency versus flexibility and cost. We present a novel concept of "human teleoperation" which bridges the gap between these two methods. In the concept, an expert remotely teloperates a person (the follower) wearing a mixed-reality headset by controlling a virtual ultrasound probe projected into the person's scene. The follower matches the pose and force of the virtual device with a real probe. The pose, force, video, ultrasound images, and 3-dimensional mesh of the scene are fed back to the expert. This control framework, where the actuation is carried out by people, allows more precision and speed than verbal guidance, yet is more flexible and inexpensive than robotic teleoperation. The purpose of this paper is to introduce this concept as well as a prototype teleultrasound system with limited haptics and local communication. The system was tested to show its potential, including mean teleoperation latencies of 0.32±0.05 seconds and steady-state errors of 4.4±2.8 mm and 5.4±2.8° in position and orientation tracking respectively. A preliminary test with an ultrasonographer and four patients was completed, showing lower measurement error and a completion time of 1:36±0:23 minutes using human teleoperation compared to 4:13±3:58 using audiovisual teleguidance.

1. Introduction

In remote areas, access to expert care and diagnosis by sonographers is often severely lacking or infrequent (Ferreira, O'Mahony, Oliani, Araujo Júnior, & da Silva Costa, 2015). By enabling expert sonographers to remotely guide or teleoperate ultrasound (US) procedures in these communities, teleultrasound has immense potential to improve the quality of care of patients, both in rural regions and in ambulances. Teleultrasound also decreases costs associated with transporting the patients or medical workers, and increase safety in a pandemic such as COVID-19 (Wu, Li, et al., 2020).

Ultrasound teleguidance systems have been implemented by numerous groups. For trauma patients, verbal guidance via radio while viewing a stream of the ultrasound images was explored by Boniface et al. (Boniface, Shokoohi, Smith, & Scantlebury, 2011). More modern systems sold by Clarius Mobile Health Corp. and Butterfly Network combine a mobile phone application with a wireless ultrasound transducer and remote access to the images and video conferencing via a cloud interface (Strumia, Costa, Pascarella, Del Buono, & Agrò, 2020). However, in all these solutions the instructions for probe positioning, orientation, and

force are given verbally or with limited augmented reality overlays of arrows or pointers, which is very inefficient, leading to high latency and low precision.

Conversely, robotic teleultrasound systems have also been developed which provide low latencies and high precision, as well as haptic feedback (Delgorge et al., 2005)(Kontaxakis, Walter, & Sakas, 2000)(Mathiassen, Fjellin, Glette, Hol, & Elle, 2016)(Adriana et al., 2001). These involve a robotic arm with ultrasound probe end effector which is teleoperated by a remote expert sonographer. Salcudean et al. presented a robot whose control was shared between the expert and a visual servoing system to maintain correct positioning on the carotid artery (Abolmaesumi, Salcudean, Zhu, Sirouspour, & DiMaio, 2002). Much interesting work has been carried out on autonomous robotic ultrasound to optimize ultrasound image quality (Chatelain, Krupa, & Navab, 2017). Another system, named OTELO (Courreges, Vieyres, & Istepanian, 2004)(Vieyres et al., 2006), has demonstrated clinical utility in trials (Courreges, Vieyres, Istepanian, Arbeille, & Bru, 2005). Recent work has even investigated the control of such systems over 5G and in the context of COVID-19 (Wu, Wu, et al., 2020).

However, there are many drawbacks with robotic systems. While some are designed to be inherently backdriveable and lightweight (Salcudean et al., 1999), the issues of safe human-robot interaction and predictable and consistent autonomy remain unsolved (Lasota, Fong, Shah, et al., 2017). As a result, a human follower is usually still needed on-site to monitor the robot (Victorova, Navarro-Alarcon, & Zheng, 2019), and potentially check and approve planned motion trajectories. This limits the efficiency of such systems. Furthermore, such robots have restricted workspaces, are time consuming to set up, too large to store on ambulances, and incongruously expensive compared to ultrasound systems. While ultrasound is usually an inexpensive procedure and is thus well suited to being a standard of care in remote communities, installing an expensive robot in every small town is infeasible.

In this paper, we introduce the concept of "human teleoperation" to bridge the gap between teleguidance and robotic systems. In human teleoperation the follower, or person carrying out the procedure on site, is guided by a remote expert through a real-time, mixed reality (MR) interface on a Microsoft HoloLens 2. A 3-dimensional (3D) virtual ultrasound transducer controlled by the expert is projected into the follower's environment for the follower to match. In terms of classical teleoperation concepts (Aliaga, Rubio, & Sanchez, 2004), the "remote" robot acting on the environment is replaced by a human "follower" (Fig. 1). In the concept, the follower copies the desired position (P_o) and force (F_o) of the "master" or "expert" by aligning the tool to its MR projection on a HoloLens 2 worn by the follower. In turn, the expert is presented visually with the end-effector pose (P') via an MR capture of the follower's environment with the virtual tool in place, as well as the forces (F'), if sensed, returned through a haptic device. The implemented prototype contains a slightly limited version of this haptic interaction concept, as outlined in Table 1, which is nonetheless functional and sufficient for the proof-of-concept. See Sections 2.4 and 4 for details and future improvements.

The key enabling technology for this system is mixed reality. While augmented reality (AR) captures the real environment and renders it on a screen, for example on a smartphone or tablet, where virtual cues can be embedded into the scene, MR projects the 3D virtual objects into the real environment using a partially-transparent headset. This allows the follower wearing the MR headset to interact

seamlessly with both the real environment and the virtual objects. The idea of using augmented and mixed reality to aid in medical procedures has been explored extensively, from providing guidance for tissue biopsies by overlaying medical images and guiding pointers (Park, Hunt, Nadolski, & Gade, 2020)(Bettati et al., 2020), to training and simulation (Ni et al., 2009)(Escobar-Castillejos, Noguez, Neri, Magana, & Benes, 2016)(Wang et al., 2017). In teleultrasound, several patents for using augmented reality interfaces to guide ultrasound procedures have been filed by Butterfly Networks, Inc. (Rothberg et al., U.S. Patent 10702242, December 21, 2017) (Rothberg et al., U.S. Patent 20190261957) and others (Dalvin & Alkaitis, U.S. Patent 20190239850, August 8, 2019) (Buras, Russel, & Nguyen, U.S. Patent 10636323, April 28, 2020).

The use of AR and MR to provide remote assistance has been used not only in telemedicine, but in countless industries including manufacturing and remote maintenance. Masoni et al. created an augmented reality system that places helpful labels and 2D text in the follower's scene to assist them in their task (Masoni et al., 2017). Conversely, Mourtzis et al. developed a framework to obtain information about a scene and create an AR application off-line which contains visual instructions that can be overlaid onto the scene (Mourtzis, Zogopoulos, & Vlachou, 2017).

All the AR/MR tele-assistance solutions mentioned above are static or preplanned, are applied only to predefined, known environments, or include only simplistic labels and arrows for guidance. Thus, our human teleoperation concept provides several contributions, which we frame here in terms of a teleultrasound system, but which are trivially extended to other applications. Our system:

(1) Allows the expert to dynamically control a 3D virtual object such as a virtual ultrasound probe in the follower's scene in real time, so the follower can follow its pose with their real probe.
(2) Captures the 3-dimensional follower-side scene on demand and relays it to the expert so the expert can interact with it visually and haptically.
(3) Allows the expert to provide input by directly manipulating a dummy ultrasound probe.
(4) Includes haptic feedback so the expert has the sensation of touching the actual patient, and can guide the follower's input force.

These contributions form the basis of the human teleoperation system proposed in this paper. They allow teleguidance that is more precise, intuitive, and with lower latency than verbal guidance, yet more flexible, inexpensive, accessible, and more feasible than robotic teleultrasound. By providing a control framework where both the input and the actuation are carried out by people, this system can be deployed in any new, unfamiliar environment, and faces none of the regulatory problems related to unpredictable and potentially unsafe behaviour of robotic systems.

In the following sections, the human teleoperation teleultrasound system will be introduced. First, the application-specific requirements and design objectives are discussed (Section 2.1). In Section 2.2 and those following, the implementation of a prototype system is shown. Finally, preliminary tests were carried out to validate the effectiveness of the system. The results are in Section 3, and the system's limitations and extension to other fields are discussed in Section 4.

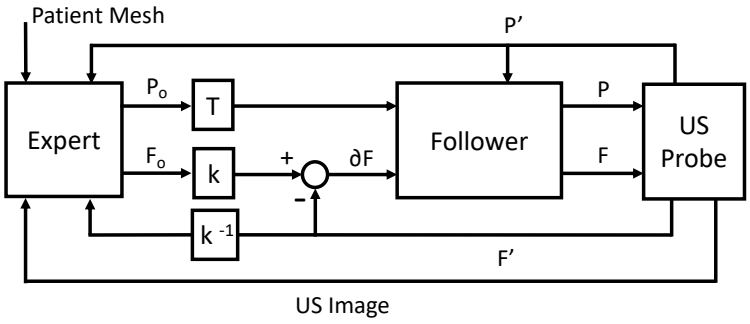

Figure 1. Block Diagram of the Human Teleoperation System Concept: The desired position, P_o, is transformed to the follower coordinate system using transform T. The forces are scaled at the expert side by a factor of k^{-1}. Instead of a controller and actuators at the "slave" side, there is a human "follower".

2. Methods

2.1. Design Objectives

Our research goal was to design and build a system that has the high precision and low latency of robotic teleultrasound without all the disadvantages listed in Section 1. In particular, we aimed to achieve a small error between the desired and actual pose and force, and low latency between issuing a command and achieving the desired state. It has also been shown that haptic feedback for the expert improves teleoperation task performance (Westebring-van der Putten, Goossens, Jakimowicz, & Dankelman, 2008) and is more intuitive for the expert, so teleoperation transparency was also an objective in this system. The expert should have the sensation of touching the actual patient and should be able to guide the follower's force without distracting the follower from following the pose. While these objectives can be achieved in a robotic system, we additionally aimed to make the patient-side interface wireless and portable. The system should be fast to set up, accessible, inexpensive (compared to a robot), and intuitive to use for both the expert and the follower. Furthermore, through meetings with expert sonographers of the British Columbia Ultrasonographers' Society, it was established that high quality ultrasound image transmission and a video conferencing interface are essential. The intent of our system is to keep the ultrasound image quality the same in terms of resolution and frame rate; this is a priority with probe position and orientation feedback to the expert as ancillary information (Stember, 2018).

2.2. System Overview

This section describes the overall system structure while implementation details are found in the following sections. While this is a description of both the human teleoperation concept and the implemented prototype, the prototype has three primary limitations which are found in Table 1. The teleultrasound system consists of two distinct halves, the follower side and the expert side, which communicate wirelessly. A conceptual overview of the system is seen in Fig. 2, and a demo video is linked in Section 6.

The follower wears a Microsoft HoloLens 2 which projects a virtual ultrasound

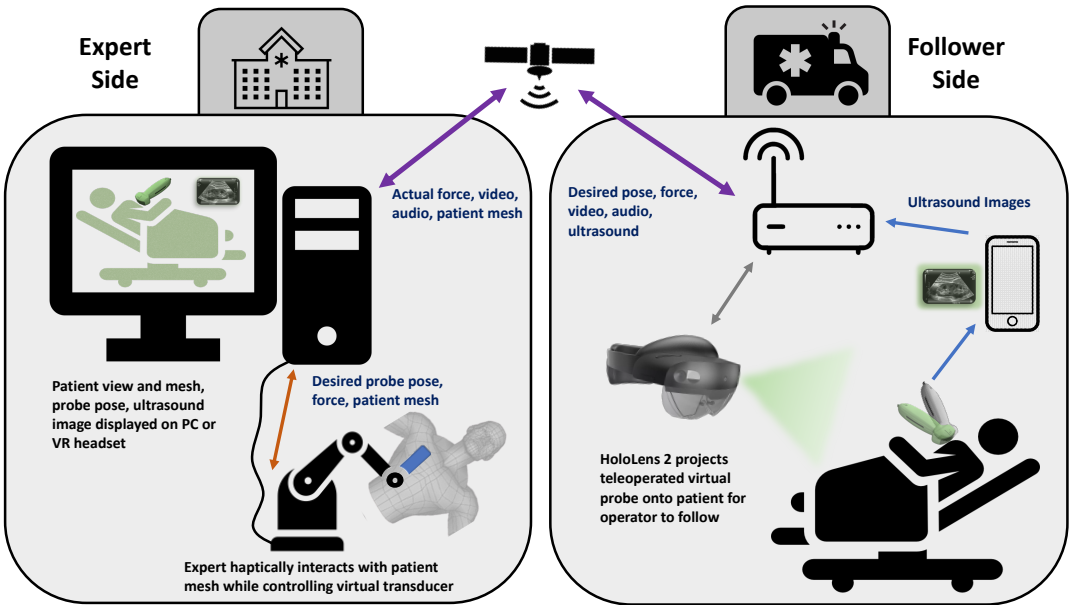

Figure 2. Conceptual System Level Diagram: The follower wears a Microsoft HoloLens 2 which projects a virtual transducer into the follower's scene. The expert controls this virtual probe using a haptic controller while observing the ultrasound images obtained by the follower. The expert and follower communicate via a (Mixed Reality WebRTC) video call interface and the probe pose, force, and patient mesh are sent by a WebSocket.

transducer into the follower's scene. The expert remotely controls this virtual probe using a haptic controller (Phantom Omni, 3D Systems, Inc - now sold as the Touch) to input the desired pose (position and orientation) and force. The follower follows the virtual probe with the real probe, thus achieving the human teleoperation. The follower-side interface is seen in Fig. 3, with a few frames showing the teleoperation. At the same time, the live ultrasound images are transmitted wirelessly from a handheld ultrasound device (C3HD, Clarius Mobile Health, Vancouver, BC) to the follower's smartphone and the expert PC. The HoloLens 2 also captures an MR video of the scene with the MR overlays in position (known as an MR capture) and shares these live with the expert via a WebRTC interface for positional feedback. In this way, the expert receives the high quality ultrasound images in real time, can see the actual patient with the virtual and real probes, and is in verbal communication with the follower.

Additionally, the follower sends a spatial mesh of the patient, generated automatically by the HoloLens 2, to the expert on demand (Section 2.5). This mesh is rendered haptically as a virtual fixture for the Phantom Omni, giving the expert the sensation that they are physically touching the patient (Section 2.4). Finally, the mesh is shown on the expert PC along with the virtual transducer in position for further pose feedback. This also allows the virtual transducer pose to be registered to the real patient, as explained in Section 2.6.

While the haptic device is used to control fine pose, the rough positioning can be changed on the expert side using the PC's arrow keys, and on the follower side by pinching and dragging the virtual probe. When the follower changes the probe position, the input from the haptic device is ignored to avoid conflicting pose commands. The haptic controller is also used to input the desired force, which is displayed on the follower side by changing the color of the virtual transducer.

Figure 3. Follower-Side Interface and Teleoperation: The follower sees the virtual transducer and a control menu. In frames 1-2 the follower starts in a random position and matches the virtual probe pose precisely within < 0.5 sec, before the probe pose changes and is immediately matched by the follower again in frame 3. Frame 4 shows the follower during this test. An image of the expert side during the creation of this sequence is shown in Fig. 4. The full test is available as a video - see Section 6.

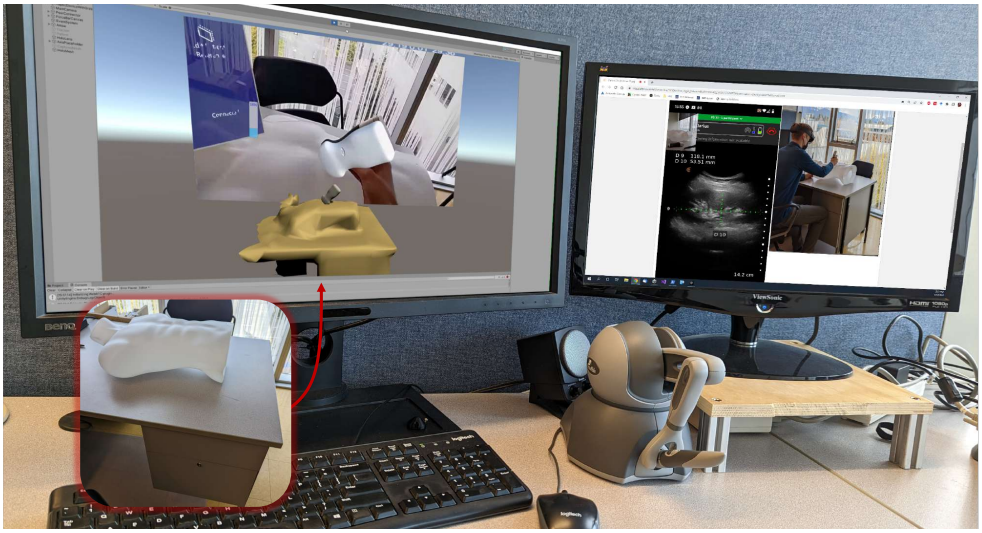

Figure 4. Expert Workstation: the Phantom Omni haptic controller (front center) is used to input pose and force and provide haptic feedback. The virtual probe relative to the patient mesh is visualized on the left monitor along with the live MR capture. A cutaway on the bottom left shows the real environment which is recreated closely by the mesh. Fig. 3 shows further MR captures. The Clarius live ultrasound images and video call are shown on the right monitor. Note, no actual ultrasound image was captured of the dummy, so the displayed image is a screenshot from a patient test, pasted here to portray the complete expert side.

In this way, the follower receives feedback on the applied force without being distracted from the pose control. The force applied to the ultrasound probe by the follower is an important part of obtaining a quality ultrasound image. Finally, the expert views the ultrasound images, MR capture, and patient mesh with the virtual transducer in position on the monitor of the expert PC, as shown in Fig. 4. The expert PC application can be viewed immersively on a virtual reality

	Concept	Prototype
Communication Section 2.3	Over Internet via Wi-Fi or mobile networks	Within local Wi-Fi networks
Force Feedforward Section 2.4	Expert force measured by haptic controller	Discrete expert force input by buttons
Force Feedback Section 2.4	Follower force measured by sensor and fed back to expert	Patient-specific mesh used as haptic interface for expert

Table 1. In a few aspects, the implemented first prototype deviates from the concept. These are listed here and described in detail in their respective sections.

headset, if desired. This further increases the immersive and realistic nature of the expert side teleoperation interface, and allows more intuitive visualization of the virtual probe on the patient mesh in 3D.

With this overview in mind, the following subsections explain the system design in more detail.

2.3. System Architecture and Communication

This section explores in detail the implementation of each component and how they all communicate. Fig. 5 shows the different communication layers and what data is sent through which interface. This mirrors Fig. 1, but shows how each connection is implemented. The required bandwidths are listed in Table 2.

Chan et al. showed that data speeds of at least 1Mbps are needed for high quality transmission of ultrasound images (Chan et al., 1999). However, with more modern imaging systems and higher expectations for quality and frame rate, this may be substantially higher. In addition, the sonographers stressed the importance of an audio/video conferencing system, which adds another several Mbps. The transducer pose and force have to be transmitted with low latency for haptic feedback, and finally a spatial mesh of the patient measured by the HoloLens 2 is sent as well (See Section 2.4). The bandwidth accounting is shown in Table 2. In total, the data being communicated may amount to up to 10Mbps peak. Given these large bandwidths, a 5G system would be ideal for the remote operation. However, this proof-of-concept prototype was developed to run on local networks only, and extension to 5G is a future improvement.

Data	Size	Rate	Bandwidth
Force	48 bits	100Hz	4.8kbps
Position	48 bits	100Hz	4.8kbps
Orientation	64 bits	100Hz	6.4kbps
Video	-	30Hz	4Mbps
Audio	-	-	96kbps
Ultrasound	-	-	3Mbps

Table 2. Bandwidth Accounting: approximation of required bandwidth for teleultrasound system. The total is about 7-10Mbps. The video bandwidth is for a 720p, 30fps video, which was deemed sufficient. Pose and force are 16-bit floating-point numbers, and orientation is sent as a quaternion.

Starting on the right side of Fig. 5, the HoloLens 2 provides the main interface for the follower through a Unity application built with the Microsoft Mixed Reality Toolkit (MRTK). It receives the desired pose and force from the expert and the actual force from the follower side, if measured, and send the patient

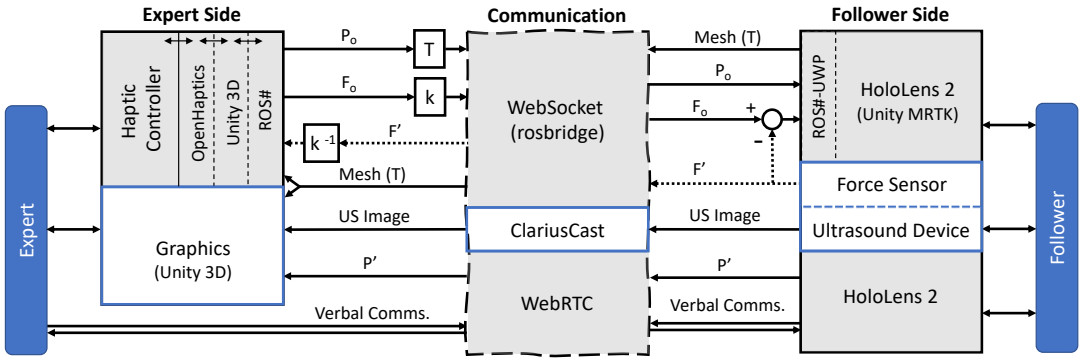

Figure 5. Communication Architecture: The wireless communication (centre) utilizes rosbridge, Mixed Reality WebRTC, and ClariusCast API. The rosbridge core runs on the expert side PC, on a Windows Subsystem for Linux (WSL) Ubuntu 18.04 machine. Everything else on the expert PC requires Windows. The expert and follower user interfaces are both implemented in Unity, using OpenHaptics SDK and Mixed Reality Toolkit (MRTK) respectively to interface with the Phantom Omni and the HoloLens 2. Both sides access rosbridge via ROS#. The patient mesh captured by the HoloLens 2 provides the coordinate transform, T, as explained in Section 2.6. Dotted lines indicate parts of the concept that are not yet fully implemented in the prototype.

mesh as well as MR captures of the scene. All communication between the expert PC and the HoloLens 2 is achieved via the rosbridge suite (Crick, Jay, Osentoski, Pitzer, & Jenkins, 2017) except the MR capture and audio communication which are sent using Microsoft's Mixed Reality WebRTC API.

Rosbridge is an API which allows Robot Operating System (ROS) communication networks to be extended from a single device to a distributed set of devices on a local wireless network. These remote devices each run one of the rosbridge client libraries (ROS# for C#, roslibpy for Python, roslibjs for Javascript) through which they can publish and subscribe to ROS topics, actions, and services. The ROS messages are first serialized into JSON (JavaScript Object Notation) before being sent to the rosbridge server on the expert PC via a WebSocket interface, which facilitates the high-speed, persistent connection needed for this application.

In the teleultrasound system, the rosbridge server is set up on a Windows Subsystem for Linux (WSL) running Ubuntu 18.04 on the expert PC. This allows for seamless integration with the expert's Unity application and Phantom Omni drivers, which require Windows. Both the expert and follower user interfaces are 3D graphics applications built in Unity (Unity Technologies, Inc) using C#. The expert and follower interfaces therefore communicate with rosbridge via ROS#, an open source rosbridge client library from Siemens. The HoloLens runs a different build of the library called ROS#-UWP, which is compatible with the Universal Windows Platform (UWP) architecture of the device. In order to minimize latency, the orientation of the probe is encoded as a quaternion. The mesh is also preprocessed to decrease the required data transfer. This is discussed in the following section.

The expert Unity application uses OpenHaptics SDK to drive the Phantom Omni and the haptic interactions, as well as OpenVR SDK to provide an optional immersive view on an Oculus Rift DK2 VR headset. As shown in Fig. 4, the live ultrasound images and MR capture are shown in the Unity application along with the patient mesh and virtual transducer. This gives the expert multiple channels of information to work with and make clinical and diagnostic decisions.

Clarius Cast API by Clarius Mobile Health Corp. allows real time streaming of the ultrasound images from the wireless transducer to devices on the local network. The audio/video call uses the HoloLens 2's microphones and front-facing cameras to stream an MR capture, as described before.

2.4. Haptics

The control of pose and force, as well as force feedback to the expert are achieved using a Phantom Omni haptic device. The Phantom Omni is a 6 degree of freedom serial arm with three actuated arm joints that can provide haptic feedback, a passive spherical wrist, and a stylus-like end effector with two buttons.

The expert determines whether more/less force is needed based on the quality of the ultrasound image, the video feed of the patient, and verbal communication with the follower. They then indicate the desired force through the haptic controller. Though the Phantom Omni used in this prototype can apply forces precisely, it is limited to 3.3N. In the 2-10 N force range, the human hand's just noticeable difference (JND) in force is about 10% (Allin, Matsuoka, & Klatzky, 2002), so for ultrasonographers accustomed to working in the 5-20 N range (Smith-Guerin et al., 2003), a 10% JND is comparable in magnitude to the entire force range of the haptic device. Thus, in practice it was found to be very difficult to precisely modulate the applied force without saturating the device, making it impractical for the expert to directly input a force by pressing harder. Instead, however, for the proof-of-concept, the two buttons on the stylus end-effector are used to indicate "more force", "less force", or "good force". On the follower side this is shown by changing the color of the transducer. "More force" makes the probe red, "less force" turns it blue, and "good force" is green. In this way, the follower can remain completely focused on following the desired pose, and does not have to look away to determine the desired force.

For force feedback to the expert, a three dimensional spatial mesh of the patient, extracted from the HoloLens 2's Simultaneous Localization and Mapping (SLAM) system, is sent to the expert side. This is rendered as a virtual fixture for the expert using the OpenHaptics SDK; i.e. whenever the haptic device's end effector is moved such that it would enter the volume inside the mesh, the haptic device applies a force normal to the surface, according to a configurable spring constant. This allows the expert to feel the virtual patient with the ultrasound probe and to rest the probe against it. The normal force can be a function of penetration depth and velocity. Additionally, friction and damping forces can be simulated for motion along the surface. This gives the expert the sensation of interacting with the actual patient.

In future work, methods for force sensing at the follower's ultrasound device will be investigated, as discussed in Section 4. For testing purposes in this work, a Raspberry Pi was set up to simulate force data and connect to rosbridge using Python's roslibpy library. From discussions with sonographers and since the ultrasound gel makes the patient surface very slippery, it is assumed that torques play a very small role and can be ignored. Furthermore, the haptic device can apply only point forces, not moments. Hence, only forces would be fed back to the expert.

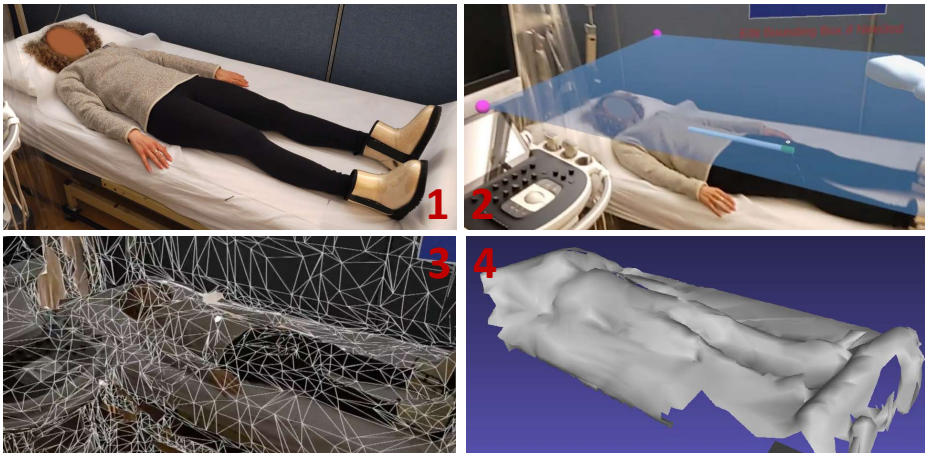

Figure 6. Example Bounding Box Definition, Scanning, and Mesh Transfer: Frame 1 shows the patient before an ultrasound. Frames 2 and 3 are an MR capture from the HoloLens 2. A bounding box is delineated by the follower using virtual markers in frame 2. In frame 3, the mesh created by the HoloLens is previewed in place to check for quality. When the "Send Mesh" button is pressed, only the mesh from under the bounding box is sent. This is seen in frame 4. The entire process is seen in a demo video (Section 6).

2.5. Mesh Management

The mesh described above is sent via the WebSocket and rosbridge, as explained in Section 2.3, after some data preprocessing. The HoloLens constantly captures a spatial mesh of as much of the environment as it sees. However, for the teleultrasound system, only the patient's mesh is desired. Thus, a bounding box is defined which delineates from which region of space the mesh vertices for the patient should be extracted. This is shown in Fig. 6 and is achieved as follows: when starting up the application, the follower is presented with four spherical markers in a rectangle configuration, with a semi-transparent plane spanning them, and a bar pointing along the major axis of the rectangle. The follower can pinch and drag the markers to resize the box, and rotate the bar to align it with the patient's bed. The height of the bounding box is also important to eliminate mesh points from the ceiling. The markers and plane are hidden a few seconds after hitting the "Finished" button, and can be recalled by pressing a button on the control menu to edit the bounding box.

When the follower presses the "Send Mesh" button on their menu, for example because the patient's position has changed, the follower is first encouraged to scan the patient with the HoloLens for 5 seconds to capture the required details. During this process, the mesh edges are projected visually onto the real world to give an idea of the mesh quality and which areas should be improved by scanning over them. Each vertex of the mesh is then iterated through to check if it is within the bounding box. To do so, the point is first projected down into the plane of the defined rectangle. Each edge of the rectangle represents a half-space partition $\boldsymbol{a}_i^\top \boldsymbol{x} \leq b_i$, so in total the rectangle is a convex set of points defined by the intersection of the four half-spaces. By placing the four $\boldsymbol{a}_i^\top$ vectors as the rows of a matrix, A, a mesh point's inclusion in the rectangle can easily be determined by checking if $A\boldsymbol{x} \leq \boldsymbol{b}$ (component-wise) and the vertical component is less than the bounding box height.

Any mesh triangles with only one vertex left within the bounding box are ignored, while mesh triangles with two vertices in the bounding box are completed

by replacing the third vertex with a its projection onto the box's boundary. This smoothes the edges of the cropped patient mesh, which is then expressed as a list of vertex points (3-vectors) and a list of indices defining which points form triangles together. These are sent via ROS as a simple message containing float and int arrays, and are converted back to a Unity mesh on the expert side.

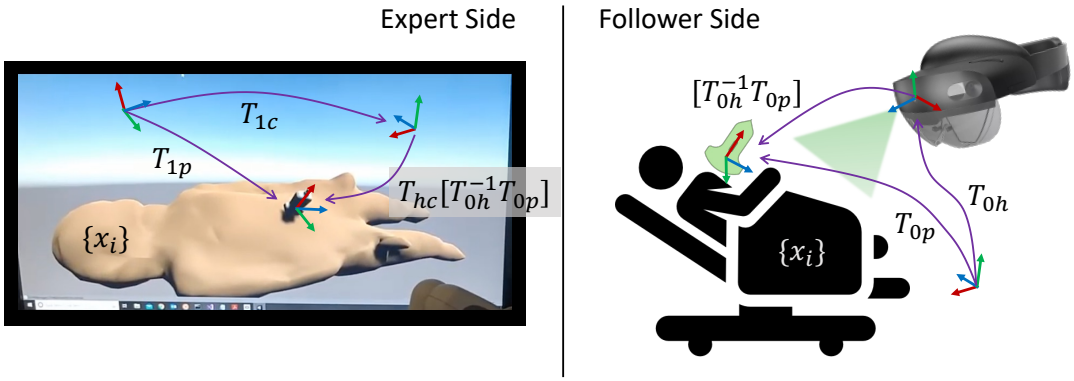

Figure 7. Coordinate transforms on expert and follower sides, giving the registration of the virtual probe to the real patient and patient mesh on the follower and expert sides respectively.

2.6. Pose Registration

In addition to facilitating haptic interaction, the mesh provides visual feedback for the expert regarding transducer positioning, and facilitates the pose registration between the expert-side virtual probe, the follower-side virtual probe, and the real patient as mentioned in Section 2.2.

In the following, $T_{ij} \in \mathrm{SE}(3)$ is the 4×4 homogeneous transformation matrix transforming frame i to frame j.

The patient mesh is measured by the HoloLens 2 as a set of points in space, $\{x_i\}$, in the HoloLens's head-attached coordinate frame. When the mesh is sent, it is placed in the expert's scene in the centre of the screen, at a comfortable distance from the camera. It is oriented such that the expert's x-axis (left-right on the expert's monitor) aligns with the major axis of the bounding box described in the previous section, so that the expert observes the patient from the side on, and the vertical axis is kept constant. This sequence uniquely defines a transformation, T_{hc}, that transforms from the HoloLens head frame to the expert's Unity camera frame. The camera pose in Unity, T_{1c}, is known. The HoloLens provides accurate SLAM through its spatial awareness interface, so the transform from the HoloLens base frame to the head frame, T_{0h}, is also known. Finally, the virtual ultrasound probe is roughly positioned by the follower relative to the patient, as explained before. This sets the pose of the probe on the follower side, in the HoloLens base coordinate frame: T_{0p}. Thus, we can define a chain of transformations to find the virtual probe pose in the expert scene:

$$T_{1p} = \left[T_{1c} T_{hc} T_{0h}^{-1}\right] T_{0p} \tag{1}$$

Thus the registration is achieved. This gives the transform $T = T_{1c}T_{hc}T_{0h}^{-1}$ in Figs. 1 and 5. The coordinate transforms are visualized in Fig. 7. The transform T includes a scaling as well, which scales expert motions up by a factor of three. This allows the expert to reach more of the patient without having to reset the virtual probe position and redo the registration. Though the haptic device can measure sub-millimeter positions, this is likely unachievable for the follower, so teleoperation precision is probably not sacrificed by the motion scaling. However, tracking accuracy by a person is still to be studied in detail.

3. Testing and Validation

In the design objectives, a number of goals were outlined involving latency, precision in position, orientation, and force, and fast and easy setup. In addition, the system aimed to be intuitive and easy to use for both the follower and expert. To verify that the human teleoperation concept implemented in the described prototype achieves these objectives and can be effective in improving teleultrasound procedures, preliminary patient tests with an expert sonographer were carried out as a proof-of-concept. A number of latency and precision tests were also carried out, all with approval from the University of British Columbia Clinical Research Ethics Board (CREB) (No. H22-01195). Rigorous evaluation of the system will follow in a separate publication.

3.1. Data Latency:

To determine the latency of the rosbridge system for sending forces and poses, the time taken to receive 100 messages from the HoloLens and Raspberry Pi was measured for both types of data. The times were then divided by 100, to find the mean latencies for the rosbridge communication channels. The test was repeated 100 times, and the resultant latency histogram is shown in Fig. 8. The latency for the video conferencing was determined by making an obvious, sharp sound which was picked up by the HoloLens 2, transmitted to the expert PC, and replayed loudly. A microphone recorded both sounds, and the delay time was determined in MATLAB. This test was repeated 20 times. The results are summarized in Table 3. The force and pose latencies are similar even though the pose involves more data, showing that they are limited by Unity rather than the communication system. Both meet the design objectives, and the WebRTC video conferencing system is sufficiently fast for efficient communication.

Data Channel	Mean Latency	Standard Deviation	Maximum
Force	11.2 ms	4.5 ms	16 ms
Pose	11.6 ms	4.6 ms	32 ms
MR Capture	160 ms	3.2 ms	166 ms

Table 3. Latency of the Communication System: Measured latencies of the 3 main communication channels, transmitting force (here just randomly generated vectors since the force sensing is not yet implemented) from Raspberry Pi to expert PC, pose from expert PC to HoloLens 2, and images to expert PC. Note that these results will change depending on network utilization and other conditions. See histogram in Fig. 8. The latency of the Clarius Live interface is difficult to measure, but Clarius states that the lag is not noticeable with a good Internet connection. Image lag due to the Internet would be approximately equal to the lag of the MR capture. Thus our system does not significantly alter the latency of the ultrasound image display - i.e. if there are image delays or dropped frames, they would be there regardless of whether we have a local exam with display over Wi-Fi or a remote one with display over Wi-Fi and transmission to the expert.

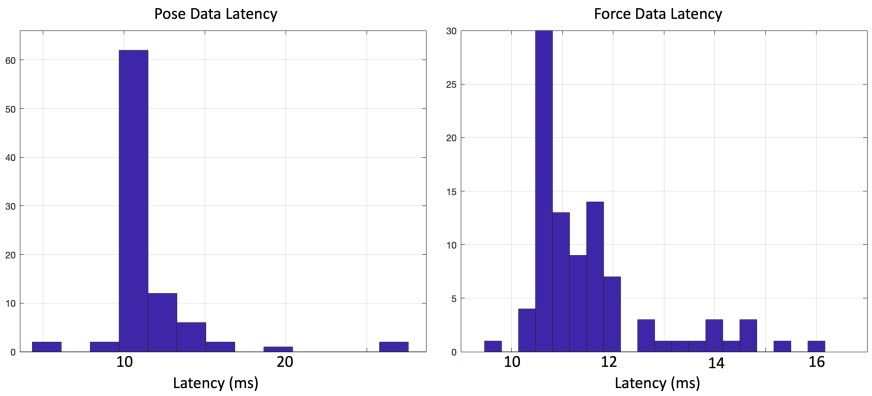

Figure 8. Histograms of Communication Latencies: The time taken to send 100 messages in the pose and force channels was measured, divided by 100 to find the average, and repeated 100 times.

3.2. Teleoperation Latency and Precision:

Up to this point, all reported latencies have been implementation-dependent and not a reflection of how well the follower can track the pose. In good network conditions, however, the actual teleoperation is likely to be limited by the reaction times of the follower in following the virtual probe pose. To test the resulting latency of the system as a whole, as well as the precision of the teleoperation, two series of arbitrary motions were recorded using the haptic device. A Python script subscribed to the expert ROS topics and saved the time-series of poses in a CSV file. Trial 1 consisted of smooth, continuous motions while trial 2 consisted of sharp motions followed by holding the pose for a few seconds (See Fig. 11). The latter series is much like a sequence of step response tests. Both series lasted 150 seconds and together contained hundreds of different poses. An end-effector similar to the shell of the ultrasound device was mounted on the haptic controller as seen in Fig. 9, and each series was played back on the HoloLens by replacing the expert side with another Python script, publishing the poses from the recorded CSV file.

To obtain the initial correct orientation, the haptic device was positioned in the recorded starting position, and the HoloLens was aligned carefully to match

Figure 9. Ultrasound probe-shaped end effector mounted on haptic device for pose following tests.

the pose. This was repeated in three different poses. The follower's head was supported to avoid any positional drift. The follower then followed the virtual probe pose with the real "probe" mounted on the haptic controller, and this was again recorded by a Python script. In this way, the expert and follower signals could be compared precisely, as seen in Fig. 11. No position or orientation drift was observed, and error in the initial alignment only increases the recorded teleoperation error, so the presented results are conservative estimates.

Using these measurements, it is possible to approximate the average latency of the teleoperation by determining the time delay between the leader and follower position signals. This is calculated as shown in Fig. 10 by applying a varying time delay to the follower signal and maximizing the absolute value of the resulting normalized cross-correlation between the signals as a function of time delay. The approximate teleoperation latencies in the three positional axes are given in Table 4. On average, the total teleoperation latency is 0.32 ± 0.05 seconds.

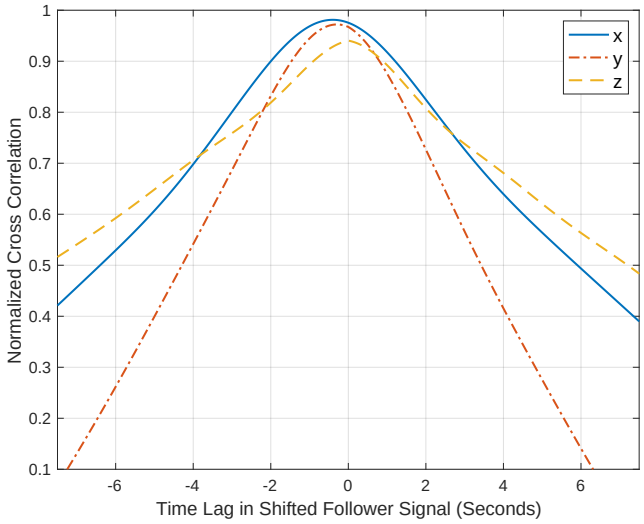

Figure 10. Normalized cross-correlation between leader and follower signals as a function of time delay from trial 2. The three values are maximized at the values shown in Table 4, showing a mean latency of 0.32 seconds, and high peak correlation of 0.96 on average.

	X	Y	Z	Mean
Trial 1 Latency (sec)	0.34	0.27	0.31	0.31
Trial 2 Latency (sec)	0.40	0.30	0.07	0.35

Table 4. Mean teleoperation latencies determined by finding the time delay in the leader position signal that maximized the absolute normalized cross-correlation between the signals. The very small z-axis latency in trial 2 is likely in part an artifact of the delay approximation algorithm. For example, rapid follower motion in the x or y axis could lead to the follower inadvertently moving the z axis before the leader moves, thus apparently leading the expert signal in that axis. This would not happen on a patient because the z motion is constrained to the patient surface. This value has been excluded from the overall average latency.

The precision was characterized separately for the position and orientation of the probe. For position, each axis was compared individually and an error signal was obtained by subtracting the leader and follower position elements. The signals for the series of sharp motions are plotted in Fig. 11. The RMS positional error of each axis and the resulting Euclidean displacement for both

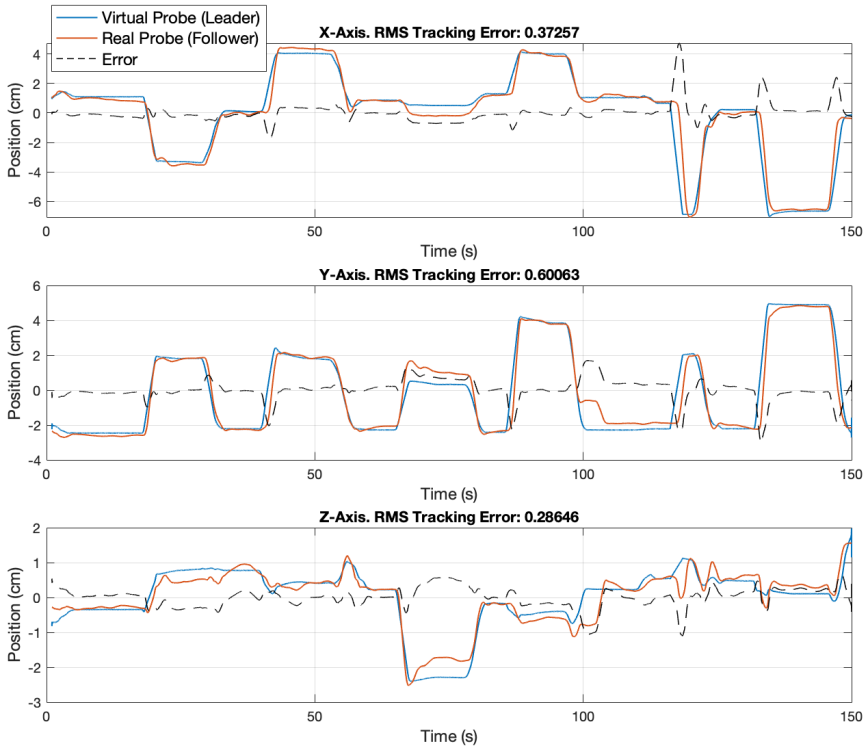

Figure 11. Positional tracking of follower with error signal. The position error is shown in Table 5, and the lag is in Table 4. Orientation error is shown in Fig. 12.

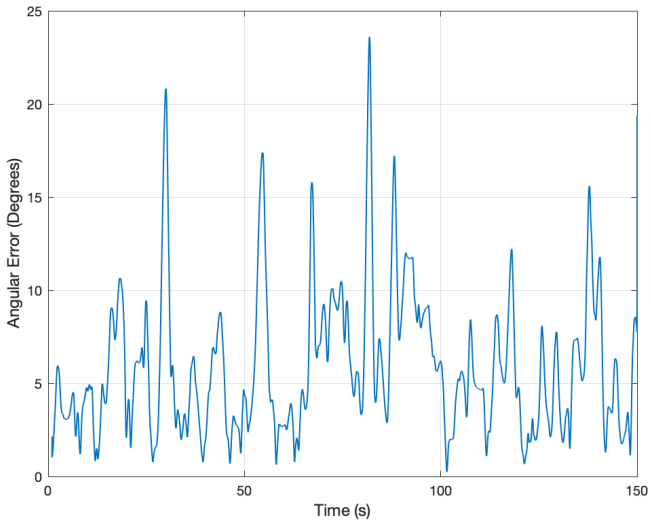

Figure 12. Angular displacement between follower and leader vs. time, representing the orientation error. The sharp peaks are where the orientation was suddenly changed and the follower had not yet reacted.

trials are found in Table 5. Both trials show very similar positional results despite the different character of the motion. Both average values are slightly inflated because they include the initial large position error at the start of a motion. The sharper motions in trial 2 are likely the reason why the mean offset in that trial

is larger. To address this and determine the steady-state error, a third trial much like trial 2 was carried out containing 12 step motions over 150 seconds, and only the steady state error was recorded. This was defined as the rms tracking error starting 1 second after a step motion, until the next motion, i.e. just the flat plateaus seen in Fig. 11. The mean Euclidean steady state error is 22% of the width of the transducer head, which was 2 cm in these tests. To quantify the orientation error, the rotation quaternion from leader to follower was calculated at every time step and converted to its axis-angle representation to find the error as a single angular value in degrees. This is plotted for trial 1 in Fig. 12. The mean angular displacements between leader and follower were $5.87 \pm 3.81^\circ$ and $6.89 \pm 4.11^\circ$ for trial 1 and 2 respectively. If we consider the steady state error again, most of the higher peaks above approximately 12° where the orientation was suddenly changed and the follower had not yet reacted are eliminated and these errors are reduced to $5.2 \pm 2.78^\circ$ and $5.50 \pm 2.71^\circ$ respectively. As expected, the mean non-steady-state error in trial 2 is larger because the motions were sharper.

Axis	X	Y	Z	Eucl.
Trial 1 Error (mm)	4.6 ± 2.9	3.1 ± 2.1	3.7 ± 2.6	6.7 ± 4.4
Trial 2 Error (mm)	3.7 ± 2.4	6.0 ± 3.1	2.9 ± 2.0	7.6 ± 4.4
Steady State Error (mm)	3.0 ± 1.9	2.5 ± 1.7	2.1 ± 1.3	4.4 ± 2.8

Table 5. RMS tracking error ($\pm$ standard deviation) of each axis and the resulting Euclidean displacement in teleoperation. We see similar errors in both trials despite the different character of the motions. The steady state error was determined in a separate trial of step-like motions.

In summary, the average tracking error was measured to be 7.1 ± 4.4 mm and $6.3 \pm 4.0^\circ$ for general teleoperation, and smaller in smoother, slower motions as in ultrasonography. The steady state position error was 4.4 ± 2.8 mm, and steady state orientation error was $5.4 \pm 2.8^\circ$.

3.3. Procedure Efficiency:

While the previous tests establish the efficacy of the human teleoperation architecture in general, it remains to be shown that this concept is useful in practice and in teleultrasound specifically. One of the primary benefits of this control scheme is that it should make remote ultrasound procedures faster and more precise by improving the efficiency of the communication through direct teleoperation. While future work will perform rigorous evaluation of the system with various sonographers, procedures, network conditions, and locations, we present here a preliminary test performed during COVID-19 which demonstrates the system's feasibility. Two procedures were carried out on two patients each, first directly by a physician with extensive sonography experience, then by inexperienced subjects guided verbally by the expert, and finally by different inexperienced subjects guided through human teleoperation by the expert. The subjects were also the patients, and consisted of two males, aged 19 and 22, and two females, aged 25 and 54. All four subjects were healthy, with normal anatomy and no prior sonography experience whatsoever.

The first test establishes the baseline for the measured values and the time taken to complete the procedure. The second and third tests form a comparison

between currently commercially available tele-guidance systems such as Clarius versus human teleoperation. The two procedures involved specific, quantitative endpoints so the effectiveness of the method could be quantified by comparison of the measured values, and the time taken to complete the procedure was well defined. The procedures were (1) measurement of the right kidney size (craniocaudal and transverse dimensions) and (2) measurement of the inferior vena cava (IVC) diameter in the infrarenal segment. An example capture of the IVC during one of the tests is shown in Fig. 13. Each subject was teleoperated on one procedure and verbally guided on the other to avoid learning the procedure and thus introducing bias into the experiment. Each person was also subjected to one of each procedure, with a screen so they could not see what was being done. Procedure times and values differ between the patients due to differences in anatomy. However, these differences should cancel out when studying the percent changes in the metrics between tests on a given patient. Additionally, though one follower may be a faster learner than another, each follower participates in one test of each teleguidance method, so again no bias is introduced. The results are outlined in Table 6. Though preliminary and with a small sample size, there is apparent improvement in the results in both speed and precision using human teleoperation over existing systems ($p = 0.052$).

Procedure	**Control**	**Verbal**	**Teleoperation**
	Time Value	Time Value	Time Value
Kidney 1	1:13 113×49mm	7:01 110×59mm	1:20 111×54mm
Kidney 2	1:43 118×50mm	1:25 123×46.5mm	1:52 112×46.9mm
Vena Cava 1	0:45 18.2mm	4:20 17.3mm	0:50 16.8mm
Vena Cava 2	0:39 17.4mm	3:30 21mm	0:47 15.9mm
Averages	Time	Time Error	Time Error
Kidney	1:28 ± 0:21	4:13 ± 3:58 4×12mm	1:36 ± 0:23 4×4mm
Vena Cava	0:42 ± 0:04	3:55 ± 0:25 2.3mm	0:49 ± 0:02 1.5mm

Table 6. Results from testing with four patients, four inexperienced followers, and one expert. Each procedure was carried out directly by the expert, then using verbal teleoperation on a Clarius system, and finally using human teleoperation. Setup time was not considered. This took less than 1 minute for the teleoperation. On average, the teleoperation is only slightly slower than the control, and substantially faster than the verbal method. The accuracy, which shows good performance, is similar but also slightly better in the teleoperation. The variation in the verbal time results is discussed in the Discussion section. The p-values for elapsed time are 0.049 between Control and Verbal, 0.052 between Verbal and Teleoperation, and 0.84 between Control and Teleoperation, though this is not reliable given the sample size.

It was also found that the audio communication between expert and follower during teleoperation was very useful, especially during setup and when the expert requested more ultrasound gel or adjustment of ultrasound device parameters (e.g. gain, depth). Additionally, the binary, color-based force controller could in theory induce oscillations with an overzealous follower (i.e. large gain). However, it was found that a natural hesitancy in the followers to apply larger forces on patients led to an over-damped force control, complemented by audio commands, giving acceptable performance.

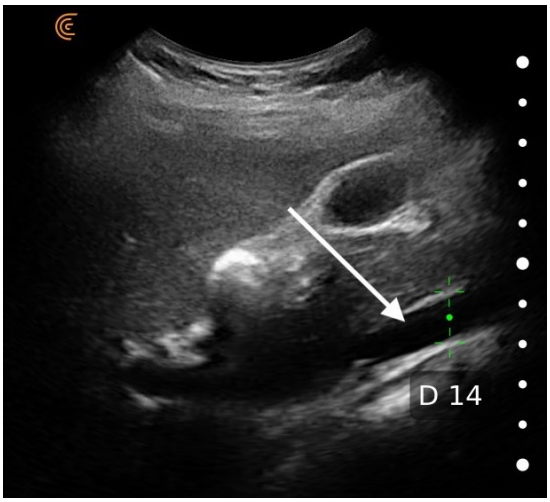

Figure 13. Infrarenal segment of inferior vena cava during human-teleoperated ultrasound exam. The ultrasound was performed using a Clarius C3HD, and the measurement was made by dragging the two green markers on the Clarius interface. The IVC is indicated by the arrow.

4. Discussion

This paper introduces the concept of human teleoperation for the broad range of applications where the control system in Fig. 1 could be applied. To better understand the implementation challenges, performance, limitations, and efficacy of the concept, a prototype system was developed for teleultrasound. Through various tests it was shown that the teleoperation error is small: approximately 7.1 ± 4.4mm and $6.3 \pm 4.0°$. While a human hand supported at the forearm can achieve accuracy up to 0.34 ± 0.16 mm, it is expected that an unsupported arm on a slippery surface like in ultrasonography has much lower accuracy (Riviere & Khosla, 1996). Hence, the precision of the teleoperation system is approximately on the same order of magnitude as that of the human hand itself, which shows good performance. The latency is 0.32 ± 0.05 sec on average, and the concept is likely to outperform existing, commercially available teleguidance methods in both precision and speed.

4.1. Comparison to Robotics

While it has not been compared directly to robotic systems, the measured teleoperation precision and latency can be contrasted with the literature. For example, the robotic teleultrasound system described in (Gilbertson & Anthony, 2011) had a rise time of about 0.08 seconds. Stable teleoperation under time delays in various conditions has been studied in detail (Arcara & Melchiorri, 2002)(Niemeyer & Slotine, 1991), though the delay can degrade performance. The most realistic model for the network-induced communication delays in this system is one of asymmetric, time-varying delays, which as shown in (Hua & Liu, 2010), can be teleoperated stably if the delays are less than 1 second. Thus, the 0.32 second latency of our system is well within the safe bounds and can enable a performant control system.

Beyond these performance characteristics are important practical factors such

as cost, portability, and setup time. Many existing robotic teleultrasound systems have used large industrial manipulators (Fang, Zhang, Finocchi, Taylor, & Boctor, 2017)(Carriere et al., 2019)(De Cunha et al., 1998)(Degoulange et al., 1998), which are expensive and not portable. The follower first has to move the robot into position, prepare it for use, and home it on the patient, making for a slow and potentially challenging setup. Custom, lighter-weight robotic ultrasound systems have also been developed (Abbasi Moshaii & Najafi, 2019)(Najafi & Sepehri, 2011) which are smaller but more complex to operate and likely very expensive. Conversely, in our system the follower simply puts on the HoloLens 2 and drags the virtual bounding box into position as shown in Fig. 3.

The primary expenses for human teleoperation are the HoloLens 2 and Phantom Omni, which together cost a fraction of an industrial robot. The HoloLens 2 was purchased for $3500 USD, and the haptic device was obtained used for $1600 USD, though a new device is more expensive. Given the new devices on the market (for example: Magic Leap One, ThirdEye X2), HoloLens prices are likely to drop. Beyond this, a normal Windows PC and a fast wireless connection are required, and usually available in a hospital. Both the PC and the haptic device can be used in the hospital for many other tasks such as teaching and simulation, and only a single set is required to cover a wide geographical area. Only the HoloLens and instrumented ultrasound probe are needed at every patient site, so their cost can be compared to that of an ultrasound robot and transducer, also instrumented with force sensing.

Further, having a human follower rather than a robot is safer as human hand actuations inherently resemble a passive system (Hogan, 1989). Thus, the human teleoperation concept has multiple advantages over existing robotic systems as well as teleguidance methods.

4.2. Implementation Limitations

Though the results are promising, the implemented system also has certain limitations, which are discussed here. First, the tele-ultrasound system was implemented on local networks to allow rapid prototyping and development. However, to be truly useful in the real world, it would have to be expanded to run on external networks. With the advent of 5G, the required bandwidths outlined in Table 2 can easily be supported. Current work is porting the communication system to WebRTC, which can support secure teleoperation over the Internet, through almost any firewall and router NAT (Network Address Translation). This is enabled by the Interactive Connectivity Establishment (ICE) protocol. Though the Internet adds some latency over local networks, WebRTC is a peer-to-peer, UDP-based architecture which is inherently faster than the system used in the presented prototype. We therefore expect the latency to continue to be limited by the human response time rather than the communication delays, so the tests and discussion about control architectures presented here would still be equally applicable. This improvement to the system is now discussed in Black & Salcudean (2022).

In addition, our system relies partly on the patient mesh to provide real-time, 3D positional and force feedback to the expert. However, though the mesh captured by the HoloLens 2 is sufficiently accurate to create a haptic surface of larger anatomies for the expert to interact with, it captures only the broad

shape, not the fine details of the patient as would be required, for example, to feel the spaces between ribs or work on smaller anatomies such as the neck. It would therefore be of interest to improve the resolution of the 3D perception to better capture the details of the patient, which could potentially be achieved using Microsoft's Research Mode APIs (Ungureanu et al., 2020). This could also be useful in other fields such as manufacturing where a more precise mesh might be required. However, it will also be addressed by adding force sensing at the tool. Additionally, the mesh does not capture the patient's color or texture (as seen in Fig. 6). It could improve the expert's experience to overlay the existing mesh with a registered and deformed MR capture of the patient to create an avatar using known methods for deformed registration and overlay (Thivierge-Gaulin et al., 2012)(Hayashi, De Sorbier, & Saito, 2012).

A final limitation and area for further research is the haptics aspect of the system. As explained in Section 2.4, the force control is currently almost entirely open-loop, with no force sensing at the ultrasound probe. Though the expert can feel the mesh, this is limited by the accuracy of the mesh, and uses a fixed spring constant that does not necessarily reflect changes to tissue impedance in different procedures. The choice of a force sensing method is very application dependent and may differ widely between teleultrasound and other applications of the human teleoperation concept such as manufacturing.

However, to improve the reliability, accuracy, and transparency of the control system, the forces applied by the ultrasound probe should be determined (Lawrence, 1993). This can be achieved by instrumenting the probe itself with a force sensor (Gilbertson & Anthony, 2013)(Abolmaesumi, Salcudean, Zhu, DiMaio, & Sirouspour, 2001), or by estimating the forces visually using the HoloLens through recurrent neural networks (Aviles, Marban, Sobrevilla, Fernandez, & Casals, 2014) or with a model-based approach, looking at tissue deformation (Giannarou et al., 2016). In this way, more complex force teleoperation architectures can also be implemented, including 4-channel teleoperation for optimal transparency (Hashtrudi-Zaad & Salcudean, 2002). Here the expert would not have a virtual fixture to interact with, but rather would have the exact forces applied by the follower on the patient reflected through the haptic controller. In addition, the forces could be scaled down at the expert side to reduce fatigue and stress-related injuries common in ultrasonographers (Mirk, Magnavita, Masini, Bazzocchi, & Fileni, 1999).

In order to realize these improvements, a more capable haptic device is required, as explained in the Section 2.4. With the ability for the expert to input a precise force vector rather than a binary more/less, the rendering of the haptic feedback at the follower must be adapted as well. A continuous spectrum of colors can be used to indicate force magnitude, and an arrow for direction. Alternatively, a second virtual transducer could be positioned with a slight offset from the original, where the direction of the offset indicates the direction of desired force and the magnitude of the offset conveys the magnitude of the commanded force, proportional to some stiffness parameter. For example, to increase the pressure, the second virtual probe could be positioned further into the patient. Then the follower would push their probe harder into the patient to reach the second probe, thus increasing the force in that direction to equal the desired force.

4.3. Human Teleoperation Limitations

While these are all implementation details and not fundamental limitations of the human teleoperation concept, the reaction time latency is the primary systematic limitation that affects the concept itself. The latencies presented in Section 3 represent close to the minimum possible response times because they are limited by the reaction time of the follower. Thus, this system can never achieve robot-level latency. However, this was clear from the start, and as the results show, the 0.32 second latency is relatively small, is much faster than alternative teleguidance techniques, and is well below the cutoff time delay for stable teleoperation given in (Hua & Liu, 2010). Furthermore, the tests of pose error and latency studied unconstrained motion in three dimensions while in an ultrasound procedure the transducer is approximately constrained into two dimensions on the surface of the patient, so the error would likely be lower. On the other hand, the measured 0.32 second latency value will likely vary between followers and can be affected by external influences such as stress, fatigue, and distractions, which is unlike a robotic system.

In the expert ultrasound tests, the standard deviations in timing were large, and in one case the verbal communication was faster than the direct measurement. This instance was an outlier where the follower coincidentally set the inital pose so that little adjustment was necessary to obtain the image. This outlier, however, does not affect the outcome which shows improved precision and speed in human teleoperation compared to existing methods. Indeed, while the tested procedures were very simple, it is expected that the teleoperation will prove even more beneficial when used in longer and more involved procedures, for example with multiple measurements or with a qualitative aspect where the expert's judgement is needed. This is because the teleultrasound system provides the ability for them to consistently have the ultrasound exactly where they want it over an extended period of time, and it offsets the initial setup time.

5. Conclusion and Future Work

In this paper, we presented a novel concept of "human teleoperation" through haptically-enabled mixed reality which bridges the gap between robotic and verbal methods of teleguidance. In this control framework, both the input and the actuation are carried out by people, but with near robot-like latency and precision. This allows teleguidance that is far more precise, intuitive, and low latency than verbal guidance, yet it is more flexible, inexpensive, and accessible than robotic teleoperation. A prototype system was implemented in the context of teleultrasound which shows the efficacy of the concept for a variety of potential applications including telemedicine, remote manufacturing, maintenance, and teaching. The system was subjected to a number of tests that show its effectiveness, including teleoperation latencies of 0.32 ± 0.05 seconds on average, and steady-state error in the pose tracking of 4.4 ± 2.8mm and $5.4 \pm 2.8°$.

A range of additional research is possible for the human teleoperation concept, including instrumenting a probe with force sensing and exploring stable and transparent force reflection in bilateral teleoperation under time delays. This has been studied extensively in the context of robotics, for example using passivity and scattering theory (Anderson & Spong, 1988), wave variables

(Niemeyer & Slotine, 2004)(Aziminejad, Tavakoli, Patel, & Moallem, 2008), μ-synthesis (Leung, Francis, & Apkarian, 1995), and an input-to-output stability small gain approach (Polushin, Liu, & Lung, 2007). However, in this system the communication delays are imposed by the human response time in the actuations, so this would constitute an interesting bridge between control theory and human teleoperation. Further work can involve extension of the human teleoperation concept to other applications, which is achievable simply by replacing the virtual ultrasound probe with a different tool, and the ultrasound images with information relevant to the application.

6. Supplementary Material

A video capture of the test from which Figs. 3 and 4 were created is available here: https://youtu.be/gEhq0ZQiyrw.

References

Abbasi Moshaii, A., & Najafi, F. (2019). Design, evaluation and prototyping of a new robotic mechanism for ultrasound imaging. *Journal of Computational Applied Mechanics*, *50*(1), 108–117.

Abolmaesumi, P., Salcudean, S., Zhu, W., DiMaio, S., & Sirouspour, M. (2001). A user interface for robot-assisted diagnostic ultrasound. In *Proceedings 2001 icra. ieee international conference on robotics and automation (cat. no.01ch37164)* (Vol. 2, p. 1549-1554 vol.2).

Abolmaesumi, P., Salcudean, S. E., Zhu, W.-H., Sirouspour, M. R., & DiMaio, S. P. (2002). Image-guided control of a robot for medical ultrasound. *IEEE Transactions on Robotics and Automation*, *18*(1), 11–23.

Adriana, V. G., Troccaz, J., Cinquin, P., Courreges, F., Poisson, G., & Tondu, B. (2001). Robotic tele-ultrasound system (ter): slave robot control. *IFAC Proceedings Volumes*, *34*(9), 425–430.

Aliaga, I., Rubio, A., & Sanchez, E. (2004). Experimental quantitative comparison of different control architectures for master-slave teleoperation. *IEEE Transactions on Control Systems Technology*, *12*(1), 2-11.

Allin, S., Matsuoka, Y., & Klatzky, R. (2002). Measuring just noticeable differences for haptic force feedback: implications for rehabilitation. In *Proceedings 10th symposium on haptic interfaces for virtual environment and teleoperator systems. haptics 2002* (p. 299-302).

Anderson, R. J., & Spong, M. W. (1988). Bilateral control of teleoperators with time delay. , *1*, 131–138.

Arcara, P., & Melchiorri, C. (2002). Control schemes for teleoperation with time delay: A comparative study. *Robotics and Autonomous systems*, *38*(1), 49–64.

Aviles, A. I., Marban, A., Sobrevilla, P., Fernandez, J., & Casals, A. (2014). A recurrent neural network approach for 3d vision-based force estimation. In *2014 4th international conference on image processing theory, tools and applications (ipta)* (p. 1-6).

Aziminejad, A., Tavakoli, M., Patel, R. V., & Moallem, M. (2008). Transparent time-delayed bilateral teleoperation using wave variables. *IEEE Transactions on Control Systems Technology*, *16*(3), 548–555.

Bettati, P., Chalian, M., Huang, J., Dormer, J. D., Shahedi, M., & Fei, B. (2020). Augmented reality-assisted biopsy of soft tissue lesions. In *Medical imaging 2020: Image-guided procedures, robotic interventions, and modeling* (Vol. 11315, p. 113150W).

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

Boniface, K. S., Shokoohi, H., Smith, E. R., & Scantlebury, K. (2011). Tele-ultrasound and paramedics: real-time remote physician guidance of the focused assessment with sonography for trauma examination. *The American journal of emergency medicine*, *29*(5), 477–481.

Buras, W. R., Russel, C. S., & Nguyen, K. Q. (U.S. Patent 10636323, April 28, 2020). *System and method for three-dimensional augmented reality guidance for use of medical equipment.* Tietronix Software Inc.

Carriere, J., Fong, J., Meyer, T., Sloboda, R., Husain, S., Usmani, N., & Tavakoli, M. (2019). An admittance-controlled robotic assistant for semi-autonomous breast ultrasound scanning. In *2019 international symposium on medical robotics (ismr)* (pp. 1–7).

Chan, F., Whitehall, J., Hayes, L., Taylor, A., Soong, B., Lessing, K., ... others (1999). Minimum requirements for remote realtime fetal tele-ultrasound consultation. *Journal of Telemedicine and Telecare*, *5*(3), 171–176.

Chatelain, P., Krupa, A., & Navab, N. (2017). Confidence-driven control of an ultrasound probe. *IEEE Transactions on Robotics*, *33*(6), 1410–1424.

Courreges, F., Vieyres, P., & Istepanian, R. (2004). Advances in robotic tele-echography services-the otelo system. In *The 26th annual international conference of the ieee engineering in medicine and biology society* (Vol. 2, pp. 5371–5374).

Courreges, F., Vieyres, P., Istepanian, R. S., Arbeille, P., & Bru, C. (2005). Clinical trials and evaluation of a mobile, robotic tele-ultrasound system. *Journal of telemedicine and telecare*, *11*(1_suppl), 46–49.

Crick, C., Jay, G., Osentoski, S., Pitzer, B., & Jenkins, O. C. (2017). Rosbridge: Ros for non-ros users. In *Robotics research* (pp. 493–504). Springer.

Dalvin, S. P., & Alkaitis, M. S. (U.S. Patent 20190239850, August 8, 2019). *Augmented/mixed reality system and method for guidance of a medical exam.*

De Cunha, D., Gravez, P., Leroy, C., Maillard, E., Jouan, J., Varley, P., ... others (1998). The midstep system for ultrasound guided remote telesurgery. In *Proceedings of the 20th annual international conference of the ieee engineering in medicine and biology society. vol. 20 biomedical engineering towards the year 2000 and beyond (cat. no. 98ch36286)* (Vol. 3, pp. 1266–1269).

Degoulange, E., Urbain, L., Caron, P., Boudet, S., Gariépy, J., Megnien, J.-L., ... Dombre, E. (1998). Hippocrate: an intrinsically safe robot for medical applications. In *Proceedings. 1998 ieee/rsj international conference on intelligent robots and systems. innovations in theory, practice and applications (cat. no. 98ch36190)* (Vol. 2, pp. 959–964).

Delgorge, C., Courrèges, F., Bassit, L. A., Novales, C., Rosenberger, C., Smith-Guerin, N., ... others (2005). A tele-operated mobile ultrasound scanner using a light-weight robot. *IEEE transactions on information technology in biomedicine*, *9*(1), 50–58.

Escobar-Castillejos, D., Noguez, J., Neri, L., Magana, A., & Benes, B. (2016). A review of simulators with haptic devices for medical training. *Journal of medical systems*, *40*(4), 104.

Fang, T.-Y., Zhang, H. K., Finocchi, R., Taylor, R. H., & Boctor, E. M. (2017). Force-assisted ultrasound imaging system through dual force sensing and admittance robot control. *International journal of computer assisted radiology and surgery*, *12*(6), 983–991.

Ferreira, A. C., O'Mahony, E., Oliani, A. H., Araujo Júnior, E., & da Silva Costa, F. (2015). Teleultrasound: historical perspective and clinical application. *International journal of telemedicine and applications*, *2015*.

Giannarou, S., Ye, M., Gras, G., Leibrandt, K., Marcus, H. J., & Yang, G.-Z. (2016). Vision-based deformation recovery for intraoperative force estimation of tool–tissue interaction for neurosurgery. *International journal of computer assisted radiology and surgery*, *11*(6), 929–936.

Gilbertson, M. W., & Anthony, B. W. (2011). Impedance-controlled ultrasound probe. , *7968*, 796816.

Gilbertson, M. W., & Anthony, B. W. (2013). An ergonomic, instrumented ultrasound probe for 6-axis force/torque measurement. , 140–143.

Hashtrudi-Zaad, K., & Salcudean, S. E. (2002). Bilateral parallel force/position teleoperation control. *Journal of Robotic Systems*, *19*(4), 155–167.

Hayashi, T., De Sorbier, F., & Saito, H. (2012). Texture overlay onto non-rigid surface using commodity depth camera. In *Visapp (2)* (pp. 66–71).

Hogan, N. (1989). Controlling impedance at the man/machine interface. In *Proceedings, 1989 international conference on robotics and automation* (p. 1626-1631 vol.3).

Hua, C.-C., & Liu, X. P. (2010). Delay-dependent stability criteria of teleoperation systems with asymmetric time-varying delays. *IEEE Transactions on Robotics*, *26*(5), 925-932.

Kontaxakis, G., Walter, S., & Sakas, G. (2000). Eu-teleinvivo: an integrated portable telemedicine workstation featuring acquisition, processing and transmission over low-bandwidth lines of 3d ultrasound volume images. In *Proceedings 2000 ieee embs international conference on information technology applications in biomedicine. itab-itis 2000. joint meeting third ieee embs international conference on information technol* (pp. 158–163).

Lasota, P. A., Fong, T., Shah, J. A., et al. (2017). *A survey of methods for safe human-robot interaction*. Now Publishers.

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

Leung, G. M., Francis, B. A., & Apkarian, J. (1995). Bilateral controller for teleoperators with time delay via mu-synthesis. *IEEE Transactions on Robotics and Automation*, *11*(1), 105–116.

Masoni, R., Ferrise, F., Bordegoni, M., Gattullo, M., Uva, A. E., Fiorentino, M., ... Di Donato, M. (2017). Supporting remote maintenance in industry 4.0 through augmented reality. *Procedia manufacturing*, *11*, 1296–1302.

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

Mirk, P., Magnavita, N., Masini, L., Bazzocchi, M., & Fileni, A. (1999). Frequency of musculoskeletal symptoms in diagnostic medical sonographers. results of a pilot survey. *La Radiologia Medica*, *98*(4), 236–241.

Mourtzis, D., Zogopoulos, V., & Vlachou, E. (2017). Augmented reality application to support remote maintenance as a service in the robotics industry. *Procedia Cirp*, *63*, 46–51.

Najafi, F., & Sepehri, N. (2011). A robotic wrist for remote ultrasound imaging. *Mechanism and machine theory*, *46*(8), 1153–1170.

Ni, D., Chan, W. Y., Qin, J., Chui, Y.-P., Qu, I., Ho, S. S., & Heng, P.-A. (2009). A virtual reality simulator for ultrasound-guided biopsy training. *IEEE computer graphics and applications*, *31*(2), 36–48.

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

Niemeyer, G., & Slotine, J.-J. E. (2004). Telemanipulation with time delays. *The International Journal of Robotics Research*, *23*(9), 873–890.

Park, B. J., Hunt, S. J., Nadolski, G. J., & Gade, T. P. (2020). Augmented reality improves procedural efficiency and reduces radiation dose for ct-guided lesion targeting: a phantom study using hololens 2. *Scientific Reports*, *10*(1), 1–8.

Polushin, I. G., Liu, P. X., & Lung, C.-H. (2007). A force-reflection algorithm for improved transparency in bilateral teleoperation with communication delay. *IEEE/ASME Transactions on Mechatronics*, *12*(3), 361-374.

Riviere, C., & Khosla, P. (1996). Accuracy in positioning of handheld instruments. In *Proceedings of 18th annual international conference of the ieee engineering in medicine and biology society* (Vol. 1, p. 212-213 vol.1).

Rothberg, A., De Jonge, M., Jia, J., Nouri, D., Rothberg, J. M., & Sofka, M. (U.S. Patent 10702242, December 21, 2017). *Augmented reality interface for assisting a user to operate an ultrasound device.* Butterfly Networks Inc.

Rothberg, A., De Jonge, M., Jia, J., Nouri, D., Rothberg, J. M., & Sofka, M. (U.S. Patent 20190261957). *Methods and apparatus for tele-medicine.* Butterfly Networks Inc.

Salcudean, S. E., Bell, G., Bachmann, S., Zhu, W.-H., Abolmaesumi, P., & Lawrence, P. D. (1999). Robot-assisted diagnostic ultrasound–design and feasibility experiments. In *International conference on medical image computing and computer-assisted intervention* (pp.

1062–1071).

Smith-Guerin, N., Al Bassit, L., Poisson, G., Delgorge, C., Arbeille, P., & Vieyres, P. (2003). Clinical validation of a mobile patient-expert tele-echography system using isdn lines. In *4th international ieee embs special topic conference on information technology applications in biomedicine, 2003.* (pp. 23–26).

Stember, J. N. (2018). Three-dimensional surface point cloud ultrasound for better understanding and transmission of ultrasound scan information. *Journal of Digital Imaging*, *31*(6), 904–911.

Strumia, A., Costa, F., Pascarella, G., Del Buono, R., & Agrò, F. E. (2020). U smart: ultrasound in your pocket. *Journal of clinical monitoring and computing*, 1–3.

Thivierge-Gaulin, D., Chou, C.-R., Kiraly, A. P., Chefd'Hotel, C., Strobel, N., & Cheriet, F. (2012). 3d-2d registration based on mesh-derived image bisection. In *International workshop on biomedical image registration* (pp. 70–78).

Ungureanu, D., Bogo, F., Galliani, S., Sama, P., Duan, X., Meekhof, C., ... Pollefeys, M. (2020). Hololens 2 research mode as a tool for computer vision research. *arXiv:2008.11239*.

Victorova, M., Navarro-Alarcon, D., & Zheng, Y.-P. (2019). 3d ultrasound imaging of scoliosis with force-sensitive robotic scanning. In *2019 third ieee international conference on robotic computing (irc)* (p. 262-265).

Vieyres, P., Poisson, G., Courrèges, F., Smith-Guerin, N., Novales, C., & Arbeille, P. (2006). A tele-operated robotic system for mobile tele-echography: The otelo project. In *M-health* (pp. 461–473). Springer.

Wang, S., Parsons, M., Stone-McLean, J., Rogers, P., Boyd, S., Hoover, K., ... Smith, A. (2017). Augmented reality as a telemedicine platform for remote procedural training. *Sensors*, *17*(10), 2294.

Westebring-van der Putten, E. P., Goossens, R. H., Jakimowicz, J. J., & Dankelman, J. (2008). Haptics in minimally invasive surgery–a review. *Minimally Invasive Therapy & Allied Technologies*, *17*(1), 3–16.

Wu, S., Li, K., Ye, R., Lu, Y., Xu, J., Xiong, L., ... Lv, F. (2020). Robot-assisted teleultrasound assessment of cardiopulmonary function on a patient with confirmed covid-19 in a cabin hospital. *Advanced Ultrasound in Diagnosis and Therapy*, *4*(2), 128–130.

Wu, S., Wu, D., Ye, R., Li, K., Lu, Y., Xu, J., ... others (2020). Pilot study of robot-assisted teleultrasound based on 5g network: a new feasible strategy for early imaging assessment during covid-19 pandemic. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, *67*(11), 2241–2248.