ARTEMIS: A Collaborative Mixed-Reality System for Immersive Surgical Telementoring

Danilo Gasques
Janet G Johnson
Tommy Sharkey
UC San Diego
La Jolla, CA, USA
{gasques,janetjohnson,tsharkey}@ucsd.edu

Yuanyuan Feng
University of Maryland
Baltimore, MD, USA
fengy1@umbc.edu

Ru Wang
Zhuoqun Robin Xu
Enrique Zavala
UC San Diego
La Jolla, CA, USA
{ruw001,zhx068,enzavala}@ucsd.edu

Yifei Zhang
Wanze Xie
Xinming Zhang
UC San Diego
La Jolla, CA, USA
{yiz243,waxie,xiz015}@ucsd.edu

Konrad Davis
Naval Medical Center San Diego
San Diego, CA, USA
konrad.l.davis.mil@mail.mil

Michael Yip
Nadir Weibel
UC San Diego
La Jolla, CA, USA
{yip,weibel}@ucsd.edu
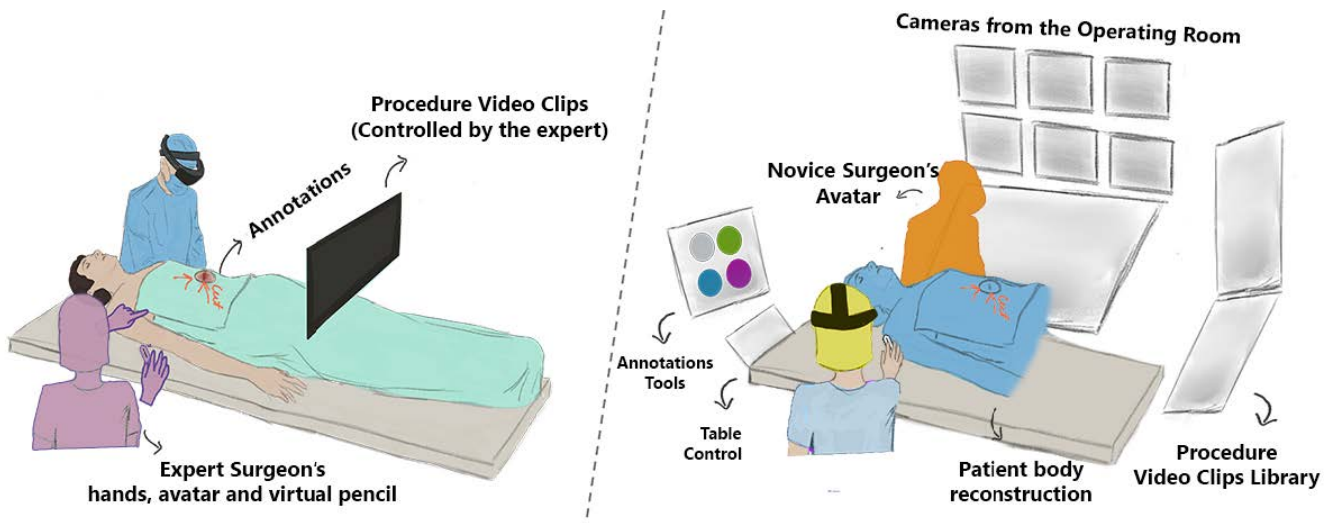
Figure 1: Artistic rendering of ARTEMIS and its features. Left: a Novice Surgeon in Augmented Reality receiving help from a remote expert. Right: a Remote Expert Surgeon in VR interacting with a 3D point-cloud of the patient, and engaging with the novice on a surgical procedure.

ABSTRACT

Traumatic injuries require timely intervention, but medical expertise is not always available at the patient's location. Despite recent advances in telecommunications, surgeons still have limited tools to remotely help inexperienced surgeons. Mixed Reality hints at a future where remote collaborators work side-by-side as if co-located; however, we still do not know how current technology can improve remote surgical collaboration. Through role-playing and iterative-prototyping, we identify collaboration practices used by expert surgeons to aid novice surgeons as well as technical requirements to facilitate these practices. We then introduce ARTEMIS, an AR-VR collaboration system that supports these key practices. Through an observational study with two expert surgeons and five novice surgeons operating on cadavers, we find that ARTEMIS supports remote surgical mentoring of novices through synchronous point, draw, and look affordances and asynchronous video clips. Most participants found that ARTEMIS facilitates collaboration despite existing technology limitations explored in this paper.

KEYWORDS

Mixed Reality, Augmented Reality, Virtual Reality, Surgery, Collaboration, Telementoring

1 INTRODUCTION

Whether from car accident, gunshot wound, or traumatic injury due to natural or man-made disasters, trauma often requires timely life-saving interventions; however, the expertise required to perform these actions is not ubiquitously present even within first world medical care systems. In situations where medical expertise or facilities are not available locally, surgeons rush to perform life-saving interventions that stabilize patient vitals before transferring the patient to a specialized surgical facility. When local surgeons lack the expertise to perform these life-saving interventions, they rely on remote guidance from expert surgeons through telementoring. Unfortunately, current telementoring systems limit how remote experts can guide and collaborate with local, inexperienced surgeons.

Telementoring is the use of information systems to provide real-time remote guidance to an inexperienced surgeon from an expert surgeon [4]. Today, telementoring systems typically require expert surgeons to instruct novice surgeons by annotating a live video feed of the surgical environment. Remote experts have to map actions they would normally express through gestures and actions into limited video-based interactions and verbalization. Novices operating on the patient have the extra burden to map instructions from a nearby screen to the operating field, with increased possibilities to make mistakes in the process [8].

Mixed Reality (MR) technology enables the creation of remote collaboration experiences where participants can interact with remote collaborators more naturally [6]. The space of remote collaboration in MR has seen a lot of promising works with a focus on training [11, 20, 58] and, more recently, also on telementoring projects [4, 52].

In this work, we build upon existing works and take a human-centered approach to designing a surgical telementoring system. Through role-playing sessions with US Navy surgeons and iterative prototyping, we gain insight into the features of MR that are useful in surgical and collaborative applications. Based on the communication needs of expert surgeons, and on novice surgeons' constraints, we designed ARTEMIS, a collaborative mixed-reality system for immersive surgical telementoring.

ARTEMIS is the first surgical telementoring system to give experts an immersive VR operating room where they can use gestures and 3D annotations on a 3D reconstruction of the patient's body to guide novice surgeons in Augmented Reality. ARTEMIS intuitive AR experience allows local surgeons to focus on the patient and

the tasks at hand, rather than having to interact with complicated AR interfaces.

Overall, this paper introduces three key contributions to the human-computer interaction, surgical innovation, and user interface software and technology research space:

- ARTEMIS's design process, including software artifacts that facilitated our technology exploration
- The development of ARTEMIS: a real-time, mixed reality collaborative system for telementoring surgical procedures
- An initial qualitative evaluation of ARTEMIS through mannequin and cadaveric subjects

2 RELATED WORK

ARTEMIS is inspired by previous works that use Mixed Reality as a medium for remote collaboration. This section summarizes recent advances in Mixed Reality and telementoring. Throughout this paper, we use the term Mixed Reality to represent the the spectrum of technologies ranging from Augmented Reality to Virtual Reality [39].

2.1 Collaboration in Mixed Reality

Studies show that participants collaborating on *physical tasks* over video usually under perform relative to participants that are collaborating side-by-side [28]. Unlike video systems, collocated collaborators can see and understand spatial relationships between each other, the task, and the environment. They use shared visual information to communicate more effectively through a combination of verbalization, gestures, and actions [21].

The continuous development of Mixed Reality input and display technology has been addressing the limitations of video communication by introducing support to key components that make side-by-side collaboration efficient: hand gestures [2, 49], sketches [23], annotations [11, 20, 38], real-time representations of a remote environment [12, 30, 46, 57], gaze [1, 12], shape of remote objects [30], collaborators as avatars [50], virtual replicas of task object [41], haptics [13], and more [32, 58].

These works provide a range of contributions, from enabling technology [46] to user evaluations of new interfaces [41] to new interaction modalities [50]. They approach remote collaboration from artificial domains, environments, tasks, and people in the environment to uncover general principles, but, as pointed out by Ens et al. [16], we need a more specific distillation of these principles in other to apply them to a specific domain (in this case, trauma surgery). For example, both remote laparoscopic procedures in practice [26] and studies like the ones by Fussell et al. [18] are based on a system where a mentor can annotate a live video of the mentee's task space; yet, communication in the robot-building task proposed by Fussel et al. [18] shares little resemblance to what is communicated in laparoscopic surgery [17]. Thus, while the interface in [18] leads to faster performance and better coordination between mentor and mentees, it can lead to mistakes and misunderstandings in a laparoscopic environment [53].

In our work, instead of adapting physical-collaboration systems such as Microsoft Remote Assist [38] and LOKI [58] to the surgical domain, we take a step back to better leverage the existing remote collaboration literature, and understand its limitations and

unique uses in surgery. We involve surgeons in the design process and explore the use of different enabling technologies (including Microsoft Remote Assist) through a design exploration grounded in participatory design techniques such as role-playing [56] and body-storming [47].

2.2 Telementoring in Mixed Reality

The domain of telementoring has greatly benefited from advances in telecommunication technologies [40]. In a conventional telementoring scenario, a senior or "expert" surgeon coaches a novice surgeon through a video stream. A common drawback for this conventional telementoring approach is that it requires novice surgeons to memorize the instructions (e.g., the position length and orientation of annotations) received on the display to perform them on the patient. This continuous focus in shift from patient to nearby tablet and back causes cognitive overload and potentially surgical errors [7].

Similar to the general remote collaboration literature, the introduction of Mixed Reality in telementoring aimed at overcoming drawbacks of conventional telementoring. STAR (2016) [3, 4] uses a tablet display between the novice surgeon and the patient body and prevents unnecessary focus shift by showing instructions directly over the surgical field. However, these tablet displays can only represent a flat image (no depth or 3D cues); thus, more recently, a new version of STAR (2018) [31, 51, 52] was introduced to provide novices with a head-mounted display so that they can have stereo view (depth cues) of annotations and instructions on the patient body. Situated instructions are known to increase procedural adherence as well as decrease execution times [25]. A key limitation of all STAR systems is that the expert surgeon sees and annotates the patient view on a flat monitor. These annotations then sit flat onto the patient as the expert surgeon is unable to give them a custom depth, required for example, to show the depth of an incision.

As we detail in the next section, surgeons annotating a 3D body need 3D annotations that they can shape freely with their hands. ARTEMIS builds on previous research to combine not only a novel 3D(expert)-3D(novice) annotation interface, but also to introduces non-verbal communication channels that bring telementoring closer to situated mentoring (e.g., gestures) and even improve upon it (e.g., procedure videos).

3 DESIGNING A MIXED REALITY TELEMENTORING SYSTEM

To design a collaborative mixed reality platform for surgery, we need to address two problems. First, we must understand how expert surgeons mentor inexperienced surgeons - their approach, their goals, and their unaddressed needs. Second, we need to understand how different MR interfaces can support their activities. Both problems are highly contextual and experiential. Moreover, the usefulness of an MR collaboration interface for a specific scenario might depend on various contextual and environmental factors.

To take into consideration the environment and the medical procedure, and to better understand user needs, we broke down our design process into two phases. (1) A role-playing phase in a mock-up surgical environment, and (2) An iterative prototyping

Figure 2: Our mock-up operating room (left) and office environments (right). In the left picture, we see a video frame from one of the role-playing sessions. In it, the expert enacts a *cricothyrotomy* on a medical mannequin while designers, developers, and other stakeholders engage in participatory design. In the right picture, we see the expert's office space with a number of paper and digital tools that we used to explore how experts would interact with the novice surgeons

phase where experts performed procedures on mannequins and cadavers.

3.1 Role-playing in a mock-up operating room

To better understand how experienced surgeons mentor inexperienced surgeons as well as to received feedback on the use of MR interactions in a trauma telementoring situation, we invited seven domain experts (4 surgeons and 3 operating room technology specialists) to participate in four role-playing [56] sessions. To contextualize the role-playing sessions, we simulated emergency procedures on a mannequin in a mock-up operating room (Fig. 2. left).

Each session was composed of two stages. First, we asked an expert surgeon to walk through an emergency procedure while they enacted it on the simulation mannequin. This allowed us to observe how mentors make sense of what they do – how they make decisions, how they communicate surgical steps to the novice surgeon, as well as what expectations they have from the novice surgeon. During this first stage, we asked questions to help us understand the specifics of the procedure as well as the specifics of their mentoring approach for that procedure. Then, in the second stage of the role-playing session, we invited experts to try existing MR applications and interfaces. We tested customized prototypes based on our observations as well as existing tools such as Microsoft Remote Assist [38]. This happened on an isolated part of the room where experts were unable to see the mannequin in the mock-up OR but were still able to talk to the designer who acted as a novice surgeon (Fig. 2. right). Finally, given our experts' backgrounds in battlefield care, we selected three commonly used procedures in military emergency scenarios [9, 10]: *needle decompression*, *leg fasciotomy* and *cricothyrotomy*. While we could not actually perform each procedure on our medical mannequin, we were able to walk through different stages of the procedure through printed images of a surgical video (i.e., each image showed a different stage of the procedure).

Role-playing is commonly used in the design of interactive system as it can help users, designers and developers communicate

needs and constraints. Hence, we also used these role-playing sessions as an opportunity to involve other team members as well as operating room technology specialists from the same hospital where we later deployed ARTEMIS. The most interesting results, however, were found in the sessions with four surgeons (S1-S4) .

In the remainder of this section, we report on our analysis of these role-playing sessions; our goal is to answer three key questions related to expert surgeons (tele)mentoring novice surgeons: **1)** *How do mentors guide mentees?* **2)** *How much information does the mentor need from the mentee side?* **3)** *What do mentors want to show to mentees?*

3.2 How do mentors guide mentees?

In a common, *co-located situation*, experts stand side-by-side with a novice surgeon as the novice surgeon operates on the patient (or cadaver). Novices are expected to have basic surgical skills, but often are unsure about the specifics of the procedure they are performing — *"where I put the needle, where do I make the incision, how do I put the tube in"* [S1]. Thus, experts chime in to explain to novices how to find the incision location, how to make the incision, and how to use other tools required by the procedure (e.g., tubes and clamps). To find the location of an incision, surgeons rely on anatomical landmarks - known body features - as well as touching and feeling the patient. According to the surgeons we interviewed, for many procedures, the hardest part of the procedure can be finding out where to make an incision. In a cricothyrotomy, for example, surgeons hold the patient's chin with their left hand and probe the neck with their thumb to find a soft spot where the incision will be made.

After finding the location of the procedure, expert surgeons then proceed by showing novices the landmarks and marking paths where novices should make incisions with a skin marker. Finally, if the novice needs to use a specialized tool, the expert surgeon demonstrates how to use the tool by mimicking the motion and maneuvering of the instrument above the patient. Mentoring styles vary as experts may stand side-by-side with novices [S4] or across the table from them [S3]. In the unlikely scenario a novice is unable to replicate instructions, some expert surgeons move the novice to the side, position themselves in the same location as the novice, and then enact the procedure from where the mentee should perform it (*"I move that person away to show them how to do it right (and not do it)"[S3]*).

3.3 How much information does the mentor need from the mentee side?

During role-playing sessions, we approached this question in two ways. First, as experts walked us through the different steps of a procedure, we asked them what information they used to make procedural decisions. We also asked them how they would be able to make a decision without that information. Second, in the technological exploration part of each role-playing session, we had them interact with existing AR and VR technology to help us understand how existing applications could support their efforts. In this short subsection, we summarize both findings.

First, as highlighted in the previous section, surgeons use a combination of touch and visual landmarks to locate the part of the

body where they will operate. Unfortunately, we still lack technology that can accurately sense and replicate pressure, temperature, and other haptics in a sterile environment. Hence, we focused our research on visual tools. This restriction to the visual field may seem limiting, but it allowed us to further investigate alternative ways of displaying patient/procedure specific information in MR.

In the technological exploration, part of our role-playing sessions, we presented experts with various ways of seeing what the novice sees:

(1) A first person view from the novice's augmented reality headset.
(2) Multiple wide-angle and close-up views of the operating room, including a top-down view of the patient (See monitor setup on the left side Figure (2).
(3) A static 3D model of the mannequin.

For the first person view, we used a Microsoft HoloLens 1 and its internal live-streaming application. Due to processing power limitations, the video quality lowers and becomes "blocky" [1] during sudden camera movements. Moreover, the resolution is considerably low (1216x684) if compared to a modern cellphone video camera (1920x1080). The other cameras consisted of Intel RealSense cameras and a Microsoft Kinect (v2) for the top-down view. For the static 3D model, we used a custom model of our mannequin in a Virtual Reality application on an HTC Vive headset.

First-Person View: *"All I want to see is what they're seeing."* [S3]. Unsurprisingly, all experts valued seeing exactly what the novice is seeing. S4 elaborates on possible reasons: For procedures such as leg fasciotomy, for example, after the first incision, novice surgeons get so close to the leg that any external camera looking from behind or from the top would be occluded by them. *"As a mentor, I would want a very good view of where they are retracting"* [S4] (and only a video stream from the novices' perspective can show that).

External cameras looking at the patient body: When away from the mock-up OR, experts relied mostly on the first-person view. However, S4, after observing the first-person view feed for a while, said that the video stream was *"choppy"* and that he would like the *"ability to see what is happening in the room, even around the learner."* As pointed out before, one of the technical limitations first-person view cameras is that as the novice surgeon moves around, the video stream both bounces with head movements and can decrease in quality (due to compression artifacts). When discussing possible locations for cameras, S4 mentioned the possibility of having a camera on wheels that an assisting nurse would be able to move around and re-position for per-procedure perspectives.

3D representation of the patient body: Most MR collaboration systems available at the time represented the novice's space on a flat screen (e.g., Microsoft Remote Assist [38, 52]). Before developing a real-time 3D reconstruction system, we wanted to understand whether or not surgeons would be interested in interacting with a 3D representation of the body in Virtual Reality. Overwhelming feedback on the reconstruction was positive, provided it was in real time and clear enough to show more than just gross anatomy (*"It would be important for the model of the patient to show the opened up*

---

[1]A visual artifact of the video codec used to compress the video stream

leg" [S4]). Crucially, the three-dimensional representation would also allow for three-dimensional annotations.

3.4 What do mentors want to show to mentees?

Modern telementoring and remote collaboration systems augment the communication stream with deictic annotations. STAR [15, 31], for example, allows expert surgeons to annotate the novice's video stream or physical space. This allows them to highlight specific objects as well as describe actions [18]. Recent works, however, have shown that combining two or more visual cues such as gestures and annotations can have a significant positive impact on remote collaboration tasks [27]. In our role playing sessions, we wanted to understand which visual cues could lead to better surgical collaboration between mentors and mentees. From the expert surgeons' perspective, this section answers the following questions: What would help you best when guiding a novice surgeon? Do you need to bring in external content such as surgical videos or 3D anatomy? Should novices see virtual surgical tools? Where should novices see guidance information? This section summarizes this design exploration.

Pointing and sketching. *"I want to be able to do two things. I want to be able to point and say: 'this is your septum'. And then I want to be able to draw a line, like this, and say: 'make your incision here'. (...) [S3]*. Building upon their approach of showing anatomical landmarks and then marking incision lines, experts wanted the system to afford a similar approach. S4 envisioned pointing with his fingers while S3 suggested the use of virtual pointers, similar to a thin pencil.

Hands and gestures. Hands and gestures play a key role in how experts communicate what to do next in a surgical procedure. *"now just lift that up and just gently spread like this"* ([S3] while gesturing what to do) *"I would want them to be able to see this"*. Experts not only use hands to point at a specific location on the patient's body, but they also use gestures to enact different steps of the procedures. This behavior was consistently observed even when experts were not mocking the procedures. S2, for example, was enacting the procedure with his hands while narrating out loud.

Overall, a virtual representation of hands can provide great versatility to surgeons as they use their hands to point, to demonstrate hand positioning/angling, to mimick tool behaviors (e.g. clamps and scissors), and to quickly enact different ways of performing the same procedure (e.g., *"Being able to show hands would be very helpful with the dissection as well, because for blunt dissection there are a lot of ways of doing it"*, showing with their hands different ways of doing a blunt dissection).

As we spent more time with the surgeons we realized how critical these gestures were, not just because of their ability to show tool approaches and positioning techniques, but because surgery is fundamentally a physical act that cannot be easily verbalized. Many procedures rely on experts' tacit knowledge along side their strong understandings of both spatial layout and physical properties of anatomy. These understandings are difficult to convey verbally, but can be more easily communicated through physical gesturing.

Annotations in 3D. In the telementoring scenario, annotations allow experts to communicate the location and length of incisions. A important aspect of annotations for telementoring is the ability to communicate *depth* by drawing them in three-dimensions (*"It's hard to show you on a non-three-dimensional thing."* [S3, while explaining leg fasciotomy over a paper picture]). This became clearer when our experts telementored using existing MR collaboration technology such as Microsoft Remote Assist [38]. Similar to STAR [30], in Microsoft Remote Assist mentors sketch on a tablet and these annotations are then *projected* (ray cast) to the closest surface in the mentee environment. S4 was particularly concerned with that since procedures such as leg fasciotomy have *"a natural curve to it..."* and he wasn't able to sketch that.

Tools are not necessary, but could help in more complicated scenarios. Most emergency procedures we role-played required specialized tools such as clamps, scalpel, metzenbaum scissors, hooks, tracheostomy tubes, and others. Experts had mixed opinions on whether or not they should have access to equivalent virtual tools; they thought that these tools could be enacted with their hands (*"The mentee can see the mentor hands right? I think that I can show that with my hands. I don't think that the mentee need to physically see the tube"* [S4]) but only when the procedures were not "too complex" (*"for something more complex than this, I think that I would want to have the actual instrument.*[S3]).

Egocentric guidance. Easing the amount of cognitive workload on the novice was reinstated by experts several times during role-playing sessions. This impacts not only where information is displayed (*"(As a mentee) It would be beneficial to see information from my (point of) view"* [S4]) and how it is displayed (*"it would be great for the expert to use the same hand I am using"* [S4]) but also how much control novices should have over their MR interface (*"The learner is going to be fairly overwhelmed"* (because they are doing a new procedure alone) [S4]). Experts also agreed that it was a good idea to give novices as little control as possible: *"the idea of only getting an ON button is what most people would want"* [S4].

Educational content can support very inexperienced surgeons. Mixed Reality technologies creates an opportunity to bring 3D representations of human anatomy as well as other multi-media content novice's environment, something experts mentioned could be useful when novice has no familiarity with a procedure. While not their main goal while telementoring trauma surgery, experts expressed an interest in having access to a library of educational video clips that they could show novices. Interestingly, none of the experts expressed interest in showing 3D anatomy as — *"it varies a lot from person to person, and standard models won't be helpful"*[S4].

3.5 Summary

In summary, our analysis showed that when expert surgeons mentor novice surgeons, they focused on the following four expert goals (EG):

- EG1: Watch the procedure from the novice's perspective
- EG2: Show the location of anatomical landmarks
- EG3: Mark the location, length, and depth of incisions
- EG4: Enact the use of tools
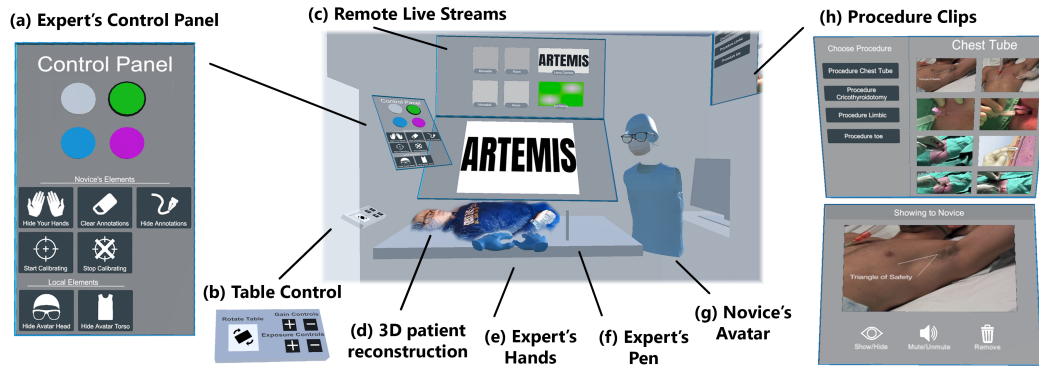
Figure 3: ARTEMIS VR - Expert surgeon's interface in Virtual Reality. At the center, the main view that the expert surgeon sees. (a) Expert's control panel with colors for 3D annotations and buttons to control novice's calibration and visibility of hands. (b) Table control to adjust orientation, camera exposure and gain. (c) Remote Live Streams with two *moveable* cameras (RealSense on wheels), two fixed cameras (RealSense), one patient-focused camera (Kinect) and a first-person view from the novice's HMD (HoloLens). (d) Real-time 3D Patient Visualization. (e) Expert's Hands as seen in VR. (f) Expert's Pen. (g) Novice's Avatar (when visible). (h) Expanded View of the Procedure Clips (the lower part is only visible when the expert is watching a procedure clip with the novice).

They need a system that enables them to fulfill all these four goals while also not overburden novice surgeons. Novice surgeons are already overwhelmed by the medical procedure at hand, and a new communication system shouldn't distract them from their main goal: keeping the patient alive.

4 THE ARTEMIS SYSTEM

After a 12-month long iterative design and development work that included rapid AR prototyping [19] and role playing in collaboration with expert surgeons, we created the **A**ugmented **R**eality **T**echnology-**E**nabled re**M**ote **I**ntegrated **S**urgery (ARTEMIS) system. ARTEMIS enables skilled surgeons and novices to work together in the same virtual space and approaches the problem of remote collaboration through a hybrid interface: expert surgeons in remote sites use Virtual Reality to access a 3D reconstruction of a patient's body and instruct novice surgeons on complex procedures; novice surgeons in the field focus on saving the patient's life while being guided through an intuitive Augmented Reality interface. From now on, we will refer to the expert interface as ARTEMIS VR and the novice interface as ARTEMIS AR. Also, to contextualize our implementation choices, we will refer back to the Expert's Goals (EG) when describing features of ARTEMIS AR or VR.

The novice bedside environment is equipped with a Microsoft HoloLens v1 [36] worn by the novice surgeon, 5x depth-cameras to capture the 3D scene (1x Microsoft Azure Kinect [35] placed on top of the bed and attached to a surgical lamp, 2x Intel RealSense cameras [24] in the corners of the room, and 2x Intel RealSense cameras on wheels movable in the OR), and an OptiTrack optical marker system [44] to track the movements of objects and people in the room (markers are attached to the HoloLens, the Kinect camera, and the surgical table).

The expert's remote environment is equipped with a video-see-trough Virtual Reality headset (HTC Vive Pro [22]) provided with a wireless adapter that makes it untethered, hand+finger tracking for the expert surgeon based on an IMU-equipped gloves [62], a

wireless mouse-pen used for annotations and VR-based interactions, and an OptiTrack optical marker system to track the movements of objects and people in the room (markers attached to the Vive Headset, surgical table proxy, gloves, and wireless pen).

Figure 5 shows ARTEMIS in action during one of our evaluation studies, and outlines the devices used for both ARTEMIS AR and ARTEMIS VR.

4.1 Novice Surgeon's Interface

The novice surgeon's main goal is to operate on the patient. To avoid distractions, the ARTEMIS AR interface components are passive, and by design the novice surgeon is not able to *directly* interact with the interface. All the features described below are controlled remotely by the expert; for instance if the novices need to hide annotations or play a procedure clip, then they can request the expert surgeon to do so. This decision emerged directly from our role playing sessions where it became clear that novices were not able to directly interact with an AR application while operating on a patient. Our hypothesis is that this resulting interface will not overwhelm the novice surgeons nor distract them from their operating tasks.

The novice surgeons are able to see three main holographic representations in front of them (Fig. 4): (a) the expert's avatar and hands, (b) a remote pen and 3D annotations, and (c) procedural video clips. In addition, the novice surgeon is able to provide direct views of the patient and the surgery to the remote expert, by manipulating the position of the Kinect camera and by directly attending to regions of interest with the HoloLens device's camera.

Expert's Avatar and Hands – The novice surgeon can see both the expert surgeon's location and their hands (Fig. 4a). This enables the expert surgeon to communicate through gestures, for example, by pointing to a location on the patient body (EG2) or by showing how to handle a surgical tool (EG4). The expert surgeon's avatar automatically disappears if the novice surgeon walks into their virtual location. This interaction allows the novice surgeon to repeat
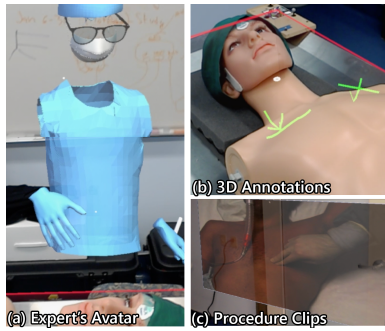
Figure 4: ARTEMIS AR Interface - Novice surgeon's interface in Augmented Reality. (a) Expert's avatar showing them holding a virtual pencil. (b) 3D Annotations as seen by the novice. (c) Procedure clips showing a video clip.

what the expert surgeon is gesturing in a more intuitive way as they are both doing it from the same point of view. In other words, the expert surgeon's hands can act as a second pair of hands that originate from the novice surgeon's body and guide the novice step-by-step.

3D Pen and Annotations – 3D annotations (Fig. 4b) allow the expert surgeons to instruct the novice by 3D sketching over the patient's body. Because these annotations are in 3D, they can directly communicate depth, length, and area (EG3), which are critical for surgical procedures such as incisions, tracheotomies, thoracotomies, etc. To allow for additional interaction space for the remote expert, 3D annotations can also happen in mid-air and are not limited to the patient's body. To facilitate the novice surgeon understanding of where the annotations will show up, if the remote surgeon is holding a pen, also ARTEMIS AR shows a 3D model of a pen. This is the same 3D pen that expert surgeons sees in their VR interface (see below), and it is shown in the hands of the expert surgeon's avatar when in use.

Procedure video clips – Procedure clips are an additional resource used to support guidance during remote telementoring. They are instructional, short video clips of different steps of specific surgical procedures that are available for the expert to show to the novice when needed (Fig. 4c). These video clips show up as a floating screen on top of surgical table (Fig. 1) and always face the novice surgeon. They keep repeating until disabled by the expert, and typically also contain audio instructions, that can be muted by the expert surgeon if needed.

Scanning the patient through the surgical lamp – ARTEMIS shows a 3D reconstruction of the patient to the expert surgeon in Virtual Reality (Fig. 3d), but the Microsoft Kinect depth camera that scans the patient is in the operating room (OR). To provide an intuitive interface that allow the novice surgeon to both know what the expert surgeon can see, and easily change that view as needed, we designed the system to take advantage of the OR layout, as well as the interaction that we observed during our role playing session; we therefore decided to attach the depth camera to the OR surgical lamp that is illuminating the patient (see Fig. 5, left). In this way, the novice surgeon is always aware of what the expert surgeon can see as the lamp illuminates that part of the patient.

4.2 Expert Surgeon's Interface

The expert surgeon's Virtual Reality interface provides surgeons with a virtual operating room. In this operating room, the 3D reconstruction of the patient is at the center of the stage, surrounded by tools and VR camera views (from the Kinect, RealSense and HoloLens cameras) designed to make the expert surgeons' interaction with ARTEMIS more efficient, and enable the expert to successfully explain procedural and conceptual steps of the surgery being performed. Figure 3 shows the entire VR interface and highlights its key components.

The expert surgeon interacts with the interface via a laser pointer, but unlike most VR experiences, this laser pointer is implemented through an actual physical pen, and not as a virtual tool that they "grab" with VR controllers. ARTEMIS VR does not use regular VR controllers to facilitate the use of gestures and tools by an expert surgeon that is most likely new to VR experiences. Thus, instead of having users learn how to map controller buttons to gestures and actions, the expert surgeons wear real gloves as if they were wearing surgical gloves (Fig. 6a) . Their only surgical tool is a pen that they can hold and feel with their hands, and can point, select, or draw (Fig. 6b).

Control Panel – The expert's control panel (Fig. 3a) provides four types of tools: (i) annotation controls and color palette, (ii) local space controls, (iii) novice surgeon's interface controls, and (iv) calibration controls. Through the annotation controls and color palette, the expert surgeons can change the color of the pen before making a 3D annotation. They can also erase annotations on both their side and the novice surgeon's side. The local space controls (Fig. 3a, bottom) allow experts to show and hide the novice's head or torso. The Novice surgeon's interface controls allow experts to change the visibility of their hands and annotations on the novice surgeon's space. Finally, the calibration controls allow experts to work with the novice surgeon to improve the alignment of annotations as seen from the novice surgeon's side. We will elaborate on the needs and the approach for alignments later in this and the next sections.

Table Control – This interface serves two purposes (Fig. 3b); the *Rotate Table* flips the orientation of the patient so that the expert surgeon can look at surgical field from two different points of view; the *Gain* and *Exposure* controls allow the expert surgeon to control settings of the remote Kinect camera, adjusting the visibility of the patient as needed.

Remote Live Streams – This interface (Fig. 3c) shows six different live video streams to the expert surgeon in the top part. Two displays show cameras attached to rolling wheels that the novice surgeon can move around the operating room. Two displays show cameras located at opposite corners of the operating room. One display shows the internal camera of the head-mounted display, and the last display shows the patient as seen by the depth camera attached to the surgical lamp. The expert surgeon can use the pen as a laser pointer to select and show any of these six videos streams in the bigger display at the bottom. The location and layout of these displays allow for the user to see both the patient reconstruction as well as the video displayed in the bigger screen without the need to switch focus or move their heads.
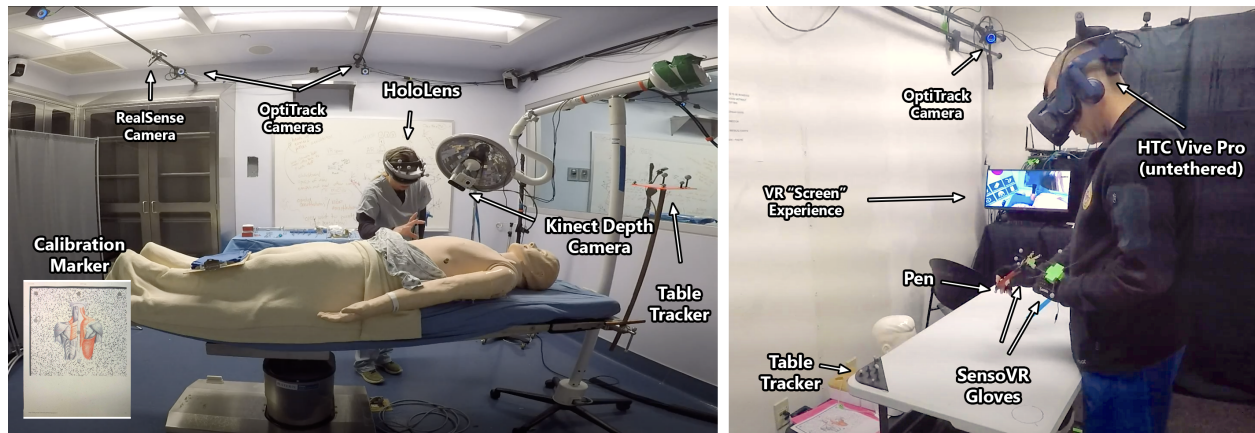
Figure 5: ARTEMIS in action. Left: ARTEMIS AR, with a novice wearing HoloLens and in the process of performing a surgical procedure on a mannequin; on the ceiling-mounted rods are visible the OptiTrack and RealSense cameras (on the left rode), while the Kinect is attached to the surgical lamp in the middle (the white zoomable cameras on the left and right of the image are not part of the system). Right: ARTEMIS VR, with an expert surgeon guiding the novice user from VR, wearing the IMU-equipped gloves, and holding the pen to annotate the 3D point-cloud in front of him; visible are the OptiTrack cameras on the ceiling-mounted rods, and a display showing the current VR view in the background.

3D Patient Reconstruction – This is the central interface available to the expert surgeon to guide the novice through specific surgical procedures (Fig. 3d). Through a point-cloud view, the expert surgeon can see the patient in a three-dimensional rendering that keeps real-world proportions. The point cloud view is a real-world live representation of the patient and it is placed on top of an actual table in the remote expert's environment. By looking at the point cloud the expert can see in real-time what is happening to the patient, and can interact with the patient representation by placing hands on particular parts of the body, and by annotating the body using 3D annotations. Both hand maneuvers and 3D annotations show up in real-time in the AR view of the novice.

Novice Surgeon's Avatar – The novice's avatar (Fig. 3f) shows the location of the novice surgeon with respect to the surgical table

at all times. Experts use the avatar as a communication and interaction anchor when guiding the novice through their procedures.

Procedure Clips Control – This interface provides a video library containing a number of surgical procedure video clips for different procedures (Fig. 3h). By selecting one of the options, a series of video clips pop up on the right side of the interface (for instance the chest tube procedure); when the expert surgeon selects one of these video clips, it displays on the larger screen at the bottom of the interface, and it plays synchronously on both the expert side in VR, and as holographic representations on the novice surgeon's side (Fig. 4c). The expert surgeon can show, hide, pause, mute, and remove this interface from both the expert's and the novice's side.

4.3 Calibrating 3D annotations

Because novice surgeons may *only* see the 3D annotations floating in front of them (Fig. 7, right), annotations displayed at the wrong location can lead them to perform an action (i.e. an incision on the patient's body) at the wrong place. To avoid this, before starting a procedure, expert surgeons can use the calibration buttons in the Control Panel (Fig. 3a) to calibrate the novice surgeon's head-mounted display. To calibrate, the novice surgeon brings a *Calibration Marker* (See Fig. 9) under the surgical lamp view. The expert surgeon then selects "Start Calibrating" in the Control Panel (Fig. 3a); when the novice surgeon looks at the calibration marker through the HoloLens camera, the system performs a calibration between the different tracking devices. The expert surgeon can confirm alignment accuracy by annotating over the calibration marker.

5 IMPLEMENTATION

As explained earlier in this paper, ARTEMIS consists of two separate spaces: ARTEMIS AR and ARTEMIS VR. ARTEMIS AR encompasses the novice surgeon's AR head-mounted display and the
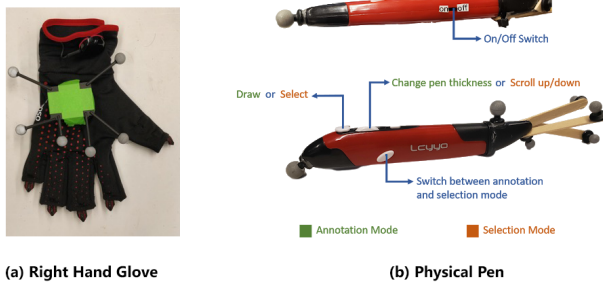
Figure 6: ARTEMIS VR Input Interfaces. Expert surgeon's wear gloves and interact with the interface through a physical pen. (a) Right-Hand glove showing OptiTrack markers (small silver spheres) installed on the top side of the glove. (b) Physical pen, implemented using a wireless mouse, and showing OptiTrack markers; highlighted are its two modes: selection mode and annotation mode.

server that connects to the cameras and trackers in the operating room. ARTEMIS VR encompasses all the tracking hardware used for the expert surgeon. Each computing device (HoloLens v1, Server Computer, and VR computer) run an application developed in Unity 2018.3 [60]. Figure 8 summarizes the ARTEMIS' hardware and streams of data.

ARTEMIS AR uses HoloLens v1 [36] as the AR headset and a standalone computer (*Server Computer*). HoloLens sends audio and video directly to ARTEMIS VR through WebRTC [37]. It also receives tracking data and commands directly from the *VR computer*. We use a separate computer, the *Server Computer*, to encode and stream data from the many cameras installed in the operating room. The Azure Kinect depth camera [35] is used to create the patient's point-cloud at a resolution of 1280x720, 30fps. The *Server Computer* also encodes and streams frames from four Intel RealSense color cameras [24] (2 on rolling wheels, 2 in the corners of the room) at a resolution of 800x600. Finally, we use five OptiTrack Prime 13 motion capture cameras [44] to cover the entire surgical space. These cameras track HoloLens, the surgical table, the surgical lamp, and the calibration marker used to realign annotations. Tracking data from these cameras flow from OptiTrack's Motive [42] to the *Server Computer* and then to both the HoloLens and the VR Computer through a custom protocol (Fig. 8, left).

ARTEMIS VR uses an untethered HTC Vive Pro [22] as the VR headset. Similar to the AR space, it also uses OptiTrack motion capture cameras to track the VR headset, the pen, the gloves, and a physical table that serves as a proxy to the surgical table (Fig. 6 shows the spherical passive infrared OptiTrack reflectors attached to the gloves and the pen). Both the pen and the gloves [62] are commodity hardware that connect to the VR computer through bluetooth. We use OptiTrack's active Vive tracker [43] to track the VR headset instead of using Vive's default tracking technology.

In developing ARTEMIS, we faced four key technical challenges: (1) HoloLens limitations, (2) a constant change of input and interaction, (3) rendering the point-cloud without distortions, and (4) calibrating the different coordinate systems. The remainder of this section describes the specific software and hardware strategies that we used to address these challenges.

5.1 Overcoming HoloLens Limitations

Released to the public in 2016, HoloLens v1 was the first commercial, untethered optical see-through head-mounted display to use inside-out global sensor fusion for head-tracking [29]. However, its capability to track itself in space, makes HoloLens integration

Figure 7: The expert surgeon annotates on the reconstruction of the patient body in VR (left) while the novice surgeon sees it overlaid onto the real patient in AR (right)

.

with external tracking systems such as OptiTrack quite challenging, especially when HoloLens' internal tracking system drifts over time [61]. To check that devices are properly integrated and therefore ensure smooth user experience, as part of ARTEMIS we created an interface for the expert surgeon and the novice surgeon to verify that OptiTrack-HoloLens calibration is accurate.

As an untethered device, HoloLens v1 runs on a battery that lasts around 2-3 hours during continuous use [36], but our surgeons were (correctly) worried that some procedures could last longer than that. To allow surgeons to quickly switch to a new device, each device connects to the *Server Computer* that acts as a hub and advertises that a new AR/VR display is available. This allows a newly connected HoloLens to quickly take over an existing session when a device is about to run out of battery.

As an embedded computer, HoloLens has also limited processing power. For example, our procedure clips (Fig. 3h) would not play while WebRTC was encoding and streaming the front-facing camera. To resolve this problem, we disable the first-person view whenever the expert surgeon plays a procedure clip. In these situations, we expect the expert surgeon to rely on the other cameras to gather more visual information of what the novice surgeon is doing.

5.2 Rapid Technology integration

ARTEMIS implementation's is the result of several iterations over a period of 12 months. During this time, we tested different commodity technologies to address the interaction needs we uncovered during our role-playing sessions with experts. AR and VR technologies are still maturing and off-the-shelf products have limitations that only show up through extensive testing. For example, we initially integrated Leap Motion [59] on the VR headset to track the expert hands, but our tests showed that Leap Motion fails with different gestures when fingers overlap each other. As a result, we integrated VR gloves as a replacement for Leap Motion, and later upgraded again to a computer vision hand tracking system.

Often, switching from one commercial technology to another means migrating to a new Software Development Kit (SDK) with different programming language or environment support. To facilitate our rapid-technology integration, we implemented abstraction layers so that the expected data from each side were independent of the hardware and software used. For example, the point-cloud is encoded as JPEG and Raw 16 bits depth. We then used a dedicated networking library to connect and synchronize data streams across systems.

Networking Library – Unity has poor support for custom, high-throughput networking protocols. With all relevant data going and coming through the network, we implemented a networking library for Unity with support for Python, C++, and other platforms such as NodeJS. The Unity counterpart of the library provides a high-level interface to TCP clients and TCP servers as well as UDP sockets. It also receives and decodes network packets in a external thread to avoid impacting rendering performance. With the exception of the WebRTC audio and video streams, we implemented all the network streams described in Figure 8 with this
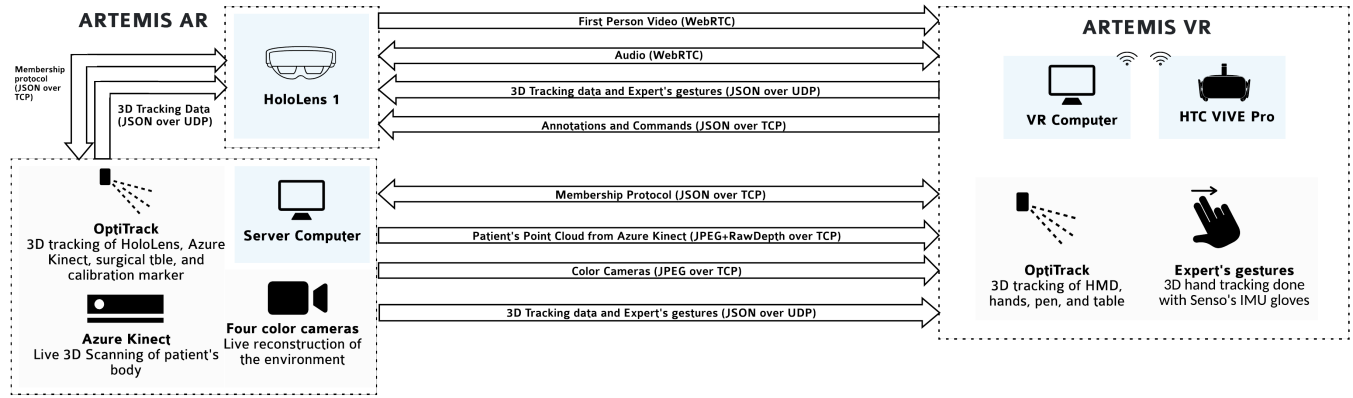
Figure 8: ARTEMIS is composed of two separate environments: ARTEMIS AR (left) and ARTEMIS VR (right). This diagram shows the origin and direction of data flowing from each environment.

library. While it is possible to replicate ARTEMIS without our networking library, we have it available as an open source project at https://github.com/WeibelLab/Comms-Unity/.

5.3 Rendering the Point-cloud

We implemented a custom application to read, encode and stream color and depth frames from Kinect Azure. Color frames are encoded into JPEGs and depth frames are sent raw through a C++ interface part of our networking library. On the VR application, we use a custom shader to render the color and depth frames as a point cloud (Fig. 7, left). Our shader uses a pre-computed look-up table to calculate the point-cloud from the depth map similar to Azure Kinect Fast Point Cloud example [34].

The AR space tracks both the surgical table and the surgical lamp equipped with the ARTEMIS depth camera. The novice surgeon can move the camera around the patient, and the expert surgeon will see the different parts of the patient at different locations on their virtual surgical table. Moving the surgical table and the lamp together will not move the virtual surgical table nor the patient point-cloud in VR.

5.4 Calibrating Different Coordinate Systems

Showing annotations on the patient's body requires us to know their location with respect to the HoloLens coordinate system. Unfortunately, HoloLens is unable to directly track the body of the patient. We rely on OptiTrack to track both the patient and HoloLens so that HoloLens can translate incoming annotations from the patient's body coordinate system to its internal coordinate system. We summarize the different coordinate systems and required transformations in Fig. 9. Overall, there are two main transformations needed.

The first step is to establish a coordinate system for the patient's body. This requires a single, offline calibration between Azure Kinect and OptiTrack. We perform this calibration through a variation of Chiodini et al. [14] using one camera instead of using a stereo setup. This calibration gives us $\overrightarrow{K_O}$, the location of Azure Kinect with respect to the OptiTrack coordinate system ($\overrightarrow{O}$). Given

that Kinect is not directly tracked by OptiTrack, during calibration, we use the surgical lamp location ($\overrightarrow{S_O}$) to find a transformation between itself and Kinect ($\overrightarrow{T_{SK}}$). $\overrightarrow{T_{SK}}$ allows us to know the location of Kinect at all times, even after a surgeon moves the surgical lamp.

The second step is to find $\overrightarrow{H_O}$–HoloLens' location with respect to OptiTrack ($\overrightarrow{O}$). Unfortunately, HoloLens' internal tracking system resets its origin every time ARTEMIS AR is started. This requires us to find $\overrightarrow{T_{HO}}$ every time we start the system. To facilitate calculating $\overrightarrow{M_H}$, we created a Calibration Marker that can be tracked both by OptiTrack and HoloLens. We use Vuforia [63] on HoloLens to track the marker through its front-facing camera. With known equivalent points $\overrightarrow{M_H}$ and $\overrightarrow{M_O}$ we can find a transformation between HoloLens and Optitrack ($\overrightarrow{T_{HO}}$).
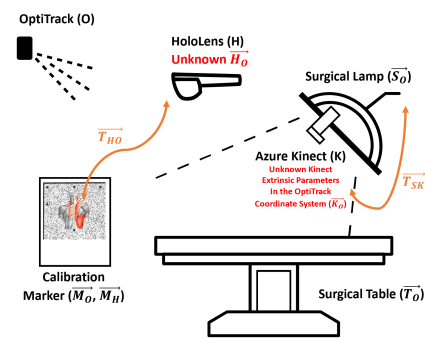
Figure 9: Diagram representing all tracked objects in the Augmented Reality space. We represent the three independent coordinate systems HoloLens, OptiTrack, and Kinect by H, O, and K, respectively. OptiTrack tracks all physical objects of interest: the calibration marker ($\overrightarrow{M_O}$), the surgical table ($\overrightarrow{S_O}$), the surgical lamp($\overrightarrow{L_O}$). The goal of the calibration is to find the transformations $\overrightarrow{T_{HO}}$ and $\overrightarrow{T_{SK}}$.

6 SYSTEM EVALUATION

ARTEMIS is being prospectively evaluated within a cadaveric trauma model. Novices are randomly assigned to ARTEMIS support or the Control group (audio only consultation), and evaluation is performed retrospectively by experts blinded to the study participants group (as all participants wear a headset platform) using standardized graduate medical education methodology. Key outcome measures are surgical technique (evaluated using the OSATS tool [33]), critical decision making, and timeliness of life saving interventions.

It is outside of the scope of this paper to further discuss the large cadaver-based randomized trial or to understand if ARTEMIS could be used for training and for the provision of emergency care. We, however, performed a qualitative evaluation of the use of ARTEMIS and we report here on initial results involving 5 experts, 6 novices and 22 procedures on mannequins and cadavers.

6.1 Procedure

As part of the initial deployment of ARTEMIS at Naval Medical Center San Diego, we enrolled 6 novices (N1-N6) and invited 5 experts (E1-E5) to work with the novices. Together they participated in pilot studies with mannequins, as well as actual studies with cadavers. Our research team engaged in direct observation, and all participants were video recorded. Direct observations from cadaver studies were used in conjunction with videos from mannequin studies. Approval for participation in the study was obtained following local human subject protection guidelines.

Of the five experts, two were Senior Critical Care Intensivists, and three expert were Staff Surgeons. The six novices were Surgical Technicians, Medics, and Junior Surgical Residents. One of the experts mentored two different novices.

The five experts engaged in a total of 22 procedures across all of the sessions with novices, with two novices only performing one procedure (cricothyroidotomy), and the other four performing five back-to-back procedures each (cricothyroidotomy, dual-incision leg fasciotomy, femoral artery exposure, axillary artery exposure, and resuscitative thoracotomy). Procedures on mannequins did not entail any actual incision, while cadaver studies performed the procedure as if it was on a real patient. All but one experts (E1) and all novices were trained on the system, but neither experts nor novices knew in advance the procedure to perform. Sessions covering one to five procedures spanned 15-60min. All experts and novices had the VR/AR headsets (HTC Vive Pro and HoloLens 1) calibrated for their inter-pupillary distance (IPDs)

The research team was composed by at least three researchers (sometimes four) who observed all the sessions, took notes, and analyzed the videos of the procedures to uncover usability issues, usage patterns, communication and coordination processes specific to immersive remote telementoring. Experts and novices were also informally interviewed at the end of their session to gather their feedback on the use of the system. We summarize results of this initial qualitative evaluation in the next section. To structure the results as presented below, we conducted a qualitative synthesis through a thematic analysis that aggregated observations from videos, interviews, and researcher notes.

6.2 Results and Discussion

In general ARTEMIS worked well, and we witnessed both novices and experts being able to communicate with increased precision, accuracy, and clarity. Novices were able to successfully complete the procedures assigned to their sessions, including those that they had never performed previously. The overall feedback from experts when asked about the system is nicely summarized by E2's comment *"To do those 5 procedures in 40 minutes, especially 2 of which he's never done before ... is pretty great"* [E2].

After reviewing the data collected during our exploratory evaluation, we organized our findings into seven major themes. The first three themes cover technology limitations and interaction limitations. The last four show us how ARTEMIS enabled experts to overcome those limitations, and how the system accommodated for different mentoring styles.

(1) The 3D point cloud doesn't represent fine details – 3D point clouds are as reliable as the depth information associated with each pixel. In situations where experts needed a closer look of the body, novices moved the surgical lamp supporting the depth camera to cover distant parts of the body (e.g. the legs during a fasciotomy), as well as to have better views (e.g. the side of the chest during a lateral puncture). Unfortunately, modern depth cameras still have a huge gap in resolution between the depth camera and the color camera. Overall, we found out that point-clouds are unable to to represent finer details [45, 57] such as veins and soft tissue such as the fascia.

(2) Annotations' depth and alignment are difficult to perceive – One of the most difficult tasks for both experts and novices was to understand how to make annotations at the right depth (expert in VR) and at what depth were annotations done (novice in AR). Experts using VR for the first time had some difficulty making annotations at the right depth. This difficulty with the interface is partially due to depth perception in VR. VR headsets are still limited in how they model and create depth perception [48]. In our studies we had to explain to one expert that they were writing annotations far above the actual 3D body because they were looking at the body from the top: *"here, bend low / crouch there to annotate the body"* - *"Oh"* [E4] – referring to stereoscopic understanding of depth. While we tackled this limitation by training our expert surgeons, a future venue of exploration is the use of visual aids and sketching constraints [5].

(3) Not everything is being shared, but this is not always clear – During our first tests, E1 tried to guide the novice surgeon by referring to one of the many video clips visible to him on the procedure clips library interface (Fig. 3h). It took him some time to realize that the novice was only able to see one selected video, and not the entire video library. As mentioned before, E1 was not trained on ARTEMIS as other experts, and did not realize that the AR and VR user interfaces are quite different. In addition, given that E1 was already quite experienced in collaborative VR interfaces, he was expecting similar interfaces as the ones he experienced in the past. A related problem would happen if the expert surgeon points to the position in space where the video library is located in VR, to refer to the currently playing video in AR. The location of the video player on the AR side is not necessarily the same as the location on the VR side, and therefore the pointing gestures would fail. In

situations like these, it is important to not take the user interface for granted, and train even the most experienced AR/VR users. In addition, to solve unbalanced interfaces, we could intervene and "warp deixis" [55] by changing pointing gestures so that they align with what users see in their environments.

Despite these drawbacks, we found that ARTEMIS' different communication features served as a way to overcome technical limitations.

(4) Expert and novice surgeons overlap each other – ARTEMIS allows experts to (virtually) move around the operating room, but they mostly stayed at same remote location as the novice surgeon. This resonates with works that investigated the impact of giving view independence to remote mentors. Amores et al. [2], for example, touch upon a similar observation where experts commented on the ease of guiding someone by offering a first-person perspective with their own hands. This not only allowed them to benefit from seeing the 3D reconstruction from the same perspective as the novice [27], but it also facilitated enacting the procedure from the novice's perspective, something that experts can't do even when working with novices side-by-side.

(5) Expert surgeons alternate between watching the novice's first-person view and sketching on 3D reconstruction – Most experts were initially expecting the 3D reconstruction to be as detailed as a video feed or as a real-patient. Unfortunately, as pointed out before, point cloud renders are unable to represent small details such as veins inside cadavers. Nonetheless, by providing more than a single of view of the body, ARTEMIS allowed experts to overcome this limitation while still benefiting from the 3D annotation interface. After the initial incision was made, experts would alternate between watching the first-person view of the novice to gather context of what they were doing, then looking at the 3D reconstruction whenever they needed to show a maneuver with their hands or make more markings.

(6) Verbalization and context can help overcome visual alignment issues in AR – Similar to other AR systems such as STAR [31, 51], annotations as seen by the novices were often not perfectly aligned with the cadaver, sometimes being off by up to 1-2cm. Interestingly the expert and novice were often able to smoothly account for this — using verbal communication, hand gestures and body landmarks to resolve alignment problems. For instance, in one of of the sessions where annotations were not perfectly aligned, E2 directly asked N3 for her feedback: *"Does the 3D writing help spatially?" (despite its inaccuracies)* [E2], *"Yeah, definitely. It's pretty neat actually"* [N3]. Another expert-novice dyad [E5-N6] learned that by standing over the body from a similar point-of-view, they could align annotations better. This helped them deal with the lack of depth cues in virtual spaces such as the lack of occlusion of the annotation by the patient body. Despite these difficulties, as mentioned above, users were always able to resolve possible annotation positioning problems, by talking through them and using hands movements to better explain.

(7) A Digital Whiteboard is needed ... but not really – In most telementoring systems, experts annotate a video interface to contextualize the instructions they want to convey - for example, by marking the location of an incision. In fact, during our role-playing sessions, experts annotated the body to show the location

of anatomical landmarks (EG2) as well as mark the location, length, and depth of incisions (EG3). Naturally, in addressing these two goals, ARTEMIS' interface is centered around 3D annotations on the patient body. In previous telementoring systems, however, experts could use the video interface to make annotations that are unrelated to the patient body. For example, they might want to explain a technical concept or describe future steps of the procedure [54]. During the exploratory study, a couple of experts asked for additional shared writing supports; E1, for instance, mentioned how he *"would like to have a whiteboard so that I can make pauses and explain steps"* [E1]. While this could inspire a new feature for ARTEMIS, we noticed how most experts relied on annotations over the virtual body (in accordance to EG1 and EG3). We also noticed that the lack of a digital whiteboard did not constrain experts in any way. E3, for example, used the space above the patient to create 3D mid-air handwritten notes that the novices were able to read and act upon.

7 CONCLUSION AND FUTURE WORK

In this paper we introduced ARTEMIS, a Mixed-Reality system for immersive surgical telementoring. Through a participatory design process with expert surgeons, we explored the use of Mixed Reality technology in a collaborative surgical task and defined four design goals that systems should address to better support expert surgeons' needs. Our system, ARTEMIS, addresses these goals to create a high-fidelity remote collaboration environment for time-critical environments. Through a qualitative evaluation, we showed that ARTEMIS allows untrained medical personnel to respond to an emergency, and to perform complex surgeries on critical patients under direct guidance from remote experts. The qualitative evaluation of ARTEMIS in a real-world surgery scenario outlined a number of important aspects that will be key for the further development of immersive collaborative environments for time-critical applications.

While telementoring for trauma care was the lead use of ARTEMIS, concepts explored here can be generalized to remote physical collaboration that requires a high-degree of fidelity and accuracy (down to the centimeter scale). Unlike previous systems that support physical task mentoring through a single, complex interface for both mentors and mentees (e.g., LOKI [58]), ARTEMIS provides specialized interfaces that allow mentors in VR to control what mentees immersed in AR can see so that mentees can focus on the task at hand.

While much can be achieved in the future with a system like ARTEMIS, we believe that a critical role of this work will be to allow researchers to understand the unique needs of surgical collaboration in trauma settings as well as the impact of new mixed-reality technology for collaborative environments, specifically investigating the development of common ground, collaboration, and communication.

In the near future, along with conducting a formal evaluation comparing ARTEMIS with other telementoring systems, we plan on refining this novel technology for mobile immersive AR and VR to enable 3D real-time telementoring in two ways. First, to scale and support time-critical tasks that require co-presence of experts alongside novices. Second, by studying ARTEMIS from the

perspective of the novice surgeon. Currently, a limitation of this work is that it was designed through expert surgeons' perspective. We expect our clinical evaluation to help us uncover unknown mentees' needs that would further improve our understanding of how a system such as ARTEMIS can provide remote guidance to novice surgeons.

ACKNOWLEDGMENTS

REFERENCES

[1] Deepak Akkil, Jobin Mathew James, Poika Isokoski, and Jari Kangas. 2016. Gaze-Torch: Enabling gaze awareness in collaborative physical tasks. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 1151–1158.

[2] Judith Amores, Xavier Benavides, and Pattie Maes. 2015. Showme: A remote collaboration system that supports immersive gestural communication. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. 1343–1348.

[3] Daniel Andersen, Voicu Popescu, Maria Eugenia Cabrera, Aditya Shanghavi, Gerardo Gomez, Sherri Marley, Brian Mullis, and Juan Wachs. 2016. Virtual annotations of the surgical field through an augmented reality transparent display. *The Visual Computer* 32, 11 (2016), 1481–1498.

[4] Daniel Andersen, Voicu Popescu, Maria Eugenia Cabrera, Aditya Shanghavi, Gerardo Gomez, Sherri Marley, Brian Mullis, and Juan P Wachs. 2016. Medical telementoring using an augmented reality transparent display. *Surgery* 159, 6 (2016), 1646–1653.

[5] Rahul Arora, Rubaiat Habib Kazi, Fraser Anderson, Tovi Grossman, Karan Singh, and George W Fitzmaurice. 2017. Experimental Evaluation of Sketching on Surfaces in VR.. In *CHI*, Vol. 17. 5643–5654.

[6] Mark Billinghurst and Hirokazu Kato. 2002. Collaborative augmented reality. *Commun. ACM* 45, 7 (2002), 64–70.

[7] Etai M Bogen, Knut M Augestad, Hiten RH Patel, and Rolv-Ole Lindsetmo. 2014. Telementoring in education of laparoscopic surgeons: An emerging technology. *World journal of gastrointestinal endoscopy* 6, 5 (2014), 148.

[8] Andrius Budrionis, Knut Magne Augestad, Hiten RH Patel, and Johan Gustav Bellika. 2013. An evaluation framework for defining the contributions of telestration in surgical telementoring. *Interactive journal of medical research* 2, 2 (2013), e14.

[9] Frank K Butler. 2010. Tactical combat casualty care: update 2009. *Journal of Trauma and Acute Care Surgery* 69, 1 (2010), S10–S13.

[10] Frank K Butler Jr, John Hagmann, and E George Butler. 1996. Tactical combat casualty care in special operations. *Military medicine* 161, suppl_1 (1996), 3–16.

[11] Henry Chen, Austin S Lee, Mark Swift, and John C Tang. 2015. 3D collaboration method over HoloLens™ and Skype™ end points. In *Proceedings of the 3rd International Workshop on Immersive Media Experiences*. 27–30.

[12] Long Chen, Thomas W Day, Wen Tang, and Nigel W John. 2017. Recent Developments and Future Challenges in Medical Mixed Reality. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 123–135.

[13] Lung-Pan Cheng, Thijs Roumen, Hannes Rantzsch, Sven Köhler, Patrick Schmidt, Robert Kovacs, Johannes Jasper, Jonas Kemper, and Patrick Baudisch. 2015. TurkDeck: Physical Virtual Reality Based on People. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. 417–426.

[14] Sebastiano Chiodini, Marco Pertile, Riccardo Giubilato, Federico Salvioli, Marco Barrera, Paola Franceschetti, and Stefano Debei. 2018. Camera rig extrinsic calibration using a motion capture system. In *2018 5th IEEE International Workshop on Metrology for AeroSpace (MetroAeroSpace)*. IEEE, 590–595.

[15] Charles R Doarn and Rifat Latifi. 2016. Telementoring and teleproctoring in trauma and emergency care. *Current Trauma Reports* 2, 3 (2016), 138–143.

[16] Barrett Ens, Joel Lanir, Anthony Tang, Scott Bateman, Gun Lee, Thammathip Piumsomboon, and Mark Billinghurst. 2019. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies* 131 (2019), 81–98.

[17] Yuanyuan Feng, Christopher Wong, Adrian Park, and Helena Mentis. 2016. Taxonomy of instructions given to residents in laparoscopic cholecystectomy. *Surgical endoscopy* 30, 3 (2016), 1073–1077.

[18] Susan R Fussell, Leslie D Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer, and Adam DI Kramer. 2004. Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction* 19, 3 (2004), 273–309.

[19] Danilo Gasques, Janet Johnson, Thomas Sharkey, and Nadir Weibel. 2019. What You Sketch Is What You Get: Quick and Easy Augmented Reality Prototyping with PintAR. In *Proc. CHI 2019 (CHI EA '19)*.

[20] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. World-stabilized annotations and virtual scene navigation for remote collaboration. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. 449–459.

[21] Darren Gergle, Robert E Kraut, and Susan R Fussell. 2004. Action as language in a shared visual space. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. 487–496.

[22] HTC. 2020. VIVE Pro Starter Kit | The professional-grade VR headset. https://www.vive.com/us/product/vive-pro-starter-kit/

[23] Weidong Huang, Seungwon Kim, Mark Billinghurst, and Leila Alem. 2019. Sharing hand gesture and sketch cues in remote collaboration. *Journal of Visual Communication and Image Representation* 58 (2019), 428–438.

[24] Intel. 2020. Intel RealSense Depth Camera D435. https://www.intelrealsense.com/

[25] Janet G Johnson, Danilo Gasques Rodrigues, Madhuri Gubbala, and Nadir Weibel. 2018. Holocpr: Designing and evaluating a mixed reality interface for time-critical emergencies. In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 67–76.

[26] Rohana Abdul Karim, Nor Farizan Zakaria, Mohd Asyraf Zulkifley, Mohd Marzuki Mustafa, Ismail Sagap, and Nani Harlina Md Latar. 2013. Telepointer technology in telemedicine: a review. *Biomedical engineering online* 12, 1 (2013), 21.

[27] Seungwon Kim, Gun Lee, Mark Billinghurst, and Weidong Huang. 2020. The combination of visual communication cues in mixed reality remote collaboration. *Journal on Multimodal User Interfaces* 14, 4 (2020), 321–335.

[28] Robert E Kraut, Darren Gergle, and Susan R Fussell. 2002. The use of visual information in shared visual spaces: Informing the development of virtual co-presence. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work*. 31–40.

[29] Bernard C. Kress and William J. Cummings. 2017. Invited Paper: Towards the Ultimate Mixed Reality Experience: HoloLens Display Architecture Choices. *SID Symposium Digest of Technical Papers* 48, 1 (2017), 127–131.

[30] Gun A. Lee, Theophilus Teo, Seungwon Kim, and Mark Billinghurst. 2018. A User Study on MR Remote Collaboration Using Live 360 Video. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 153–164.

[31] Chengyuan Lin, Daniel Andersen, Voicu Popescu, Edgar Rojas-Munoz, Maria Eugenia Cabrera, Brian Mullis, Ben Zarzaur, Kathryn Anderson, Sherri Marley, and Juan Wachs. 2018. A first-person mentee second-person mentor AR interface for surgical telementoring. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 3–8.

[32] Stephan Lukosch, Mark Billinghurst, Leila Alem, and Kiyoshi Kiyokawa. 2015. Collaboration in augmented reality. *Computer Supported Cooperative Work (CSCW)* 24, 6 (2015), 515–525.

[33] JA Martin, Glenn Regehr, Richard Reznick, Helen Macrae, John Murnaghan, Carol Hutchison, and M Brown. 1997. Objective structured assessment of technical skill (OSATS) for surgical residents. *British journal of surgery* 84, 2 (1997), 273–278.

[34] Microsoft. 2020. Azure Kinect - Fast Point Cloud. https://github.com/microsoft/Azure-Kinect-Sensor-SDK/tree/develop/examples/fastpointcloud

[35] Microsoft. 2020. Azure Kinect DK hardware specifications. https://docs.microsoft.com/en-us/azure/kinect-dk/hardware-specification

[36] Microsoft. 2020. HoloLens (1st gen) hardware. https://docs.microsoft.com/en-us/hololens/hololens1-hardware

[37] Microsoft. 2020. MixedReality-WebRTC. https://github.com/microsoft/MixedReality-WebRTC

[38] Microsoft. 2020. Overview of Dynamics 365 Remote Assist on HoloLens and HoloLens 2. https://docs.microsoft.com/en-us/dynamics365/mixed-reality/remote-assist/overview-hololens

[39] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. 1995. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, Vol. 2351. International Society for Optics and Photonics, 282–292.

[40] Bonnie A Nardi, Heinrich Schwarz, Allan Kuchinsky, Robert Leichner, Steve Whittaker, and Robert Sclabassi. 1993. Turning away from talking heads: The use

---

of video-as-data in neurosurgery. In *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*. 327–334.

[41] Ohan Oda, Carmine Elvezio, Mengu Sukan, Steven Feiner, and Barbara Tversky. 2015. Virtual replicas for remote assistance in virtual and augmented reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. 405–415.

[42] OptiTrack. 2020. Motive - Optical motion capture software. https://optitrack.com/products/motive/

[43] OptiTrack. 2020. OptiTrack - Active Components. https://optitrack.com/products/active-components/

[44] OptiTrack. 2020. OptiTrack - Hardware. https://optitrack.com/hardware/

[45] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L Davidson, Sameh Khamis, Mingsong Dou, et al. 2016. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 741–754.

[46] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchyn, Cem Keskin, and Shahram Izadi. 2016. Holoportation: Virtual 3D Teleportation in Real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 741–754.

[47] Antti Oulasvirta, Esko Kurvinen, and Tomi Kankainen. 2003. Understanding contexts by being there: case studies in bodystorming. *Personal and ubiquitous computing* 7, 2 (2003), 125–134.

[48] Nitish Padmanaban, Robert Konrad, Tal Stramer, Emily A Cooper, and Gordon Wetzstein. 2017. Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays. *Proceedings of the National Academy of Sciences* 114, 9 (2017), 2183–2188.

[49] Thammathip Piumsomboon, Arindam Day, Barrett Ens, Youngho Lee, Gun Lee, and Mark Billinghurst. 2017. Exploring enhancements for remote mixed reality collaboration. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*. 1–5.

[50] Thammathip Piumsomboon, Gun A Lee, Jonathon D Hart, Barrett Ens, Robert W Lindeman, Bruce H Thomas, and Mark Billinghurst. 2018. Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

[51] Edgar Rojas-Muñoz, Maria E Cabrera, Chengyuan Lin, Daniel Andersen, Voicu Popescu, Kathryn Anderson, Ben L Zarzaur, Brian Mullis, and Juan P Wachs. 2020. The System for Telementoring with Augmented Reality (STAR): A head-mounted display to improve surgical coaching and confidence in remote areas. *Surgery* (2020).

[52] Edgar Rojas-Muñoz, Maria Eugenia Cabrera, Chengyuan Lin, Natalia Sánchez-Tamayo, Dan Andersen, Voicu Popescu, Kathryn Anderson, Ben Zarzaur, Brian Mullis, and Juan P Wachs. 2020. Telementoring in Leg Fasciotomies via Mixed-Reality: Clinical Evaluation of the STAR Platform. *Military Medicine* 185, Supplement_1 (2020), 513–520.

[53] Azin Semsar, Hannah McGowan, Yuanyuan Feng, H Reza Zahiri, Adrian Park, Andrea Kleinsmith, and Helena Mentis. 2019. How Trainees Use the Information from Telepointers in Remote Instruction. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–20.

[54] Dave Snowdon, Elizabeth F Churchill, and Alan J Munro. 2001. Collaborative virtual environments: Digital spaces and places for CSCW: An introduction. In *Collaborative virtual environments*. Springer, 3–17.

[55] Maurício Sousa, Rafael Kufner dos Anjos, Daniel Mendes, Mark Billinghurst, and Joaquim Jorge. 2019. WARPING DEIXIS: Distorting Gestures to Enhance Collaboration. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[56] Dag Svanaes and Gry Seland. 2004. Putting the users center stage: role playing and low-fi prototyping enable end users to design mobile systems. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 479–486.

[57] Theophilus Teo, Louise Lawrence, Gun A Lee, Mark Billinghurst, and Matt Adcock. 2019. Mixed Reality Remote Collaboration Combining 360 Video and 3D Reconstruction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.

[58] Balasaravanan Thoravi Kumaravel, Fraser Anderson, George Fitzmaurice, Bjoern Hartmann, and Tovi Grossman. 2019. Loki: Facilitating Remote Instruction of Physical Tasks Using Bi-Directional Mixed-Reality Telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 161–174.

[59] Ultraleap. 2020. Tracking | Ultraleap. https://www.ultraleap.com/tracking/

[60] Unity. 2020. Unity Real-Time Development Platform | 3D, 2D VR & AR Visualizations. https://unity.com/

[61] Reid Vassallo, Adam Rankin, Elvis CS Chen, and Terry M Peters. 2017. Hologram stability evaluation for Microsoft HoloLens. In *Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment*, Vol. 10136. International Society for Optics and Photonics, 1013614.

[62] Senso VR. 2020. Senso VR | Interactive virtual and augmented reality. https://senso.me/

[63] Vuforia. 2020. Vuforia Developer Portal. https://www.vuforia.com