# AR HMD for Remote Instruction in Healthcare

Helena Mentis, Ignacio Avellino, Jwawon Seo

# AR HMD for Remote Instruction in Healthcare

Helena M. Mentis*
University of Maryland Baltimore County

Ignacio Avellino†
Sorbonne Université, CNRS, ISIR

Jwawon Seo‡
University of Maryland
Baltimore County

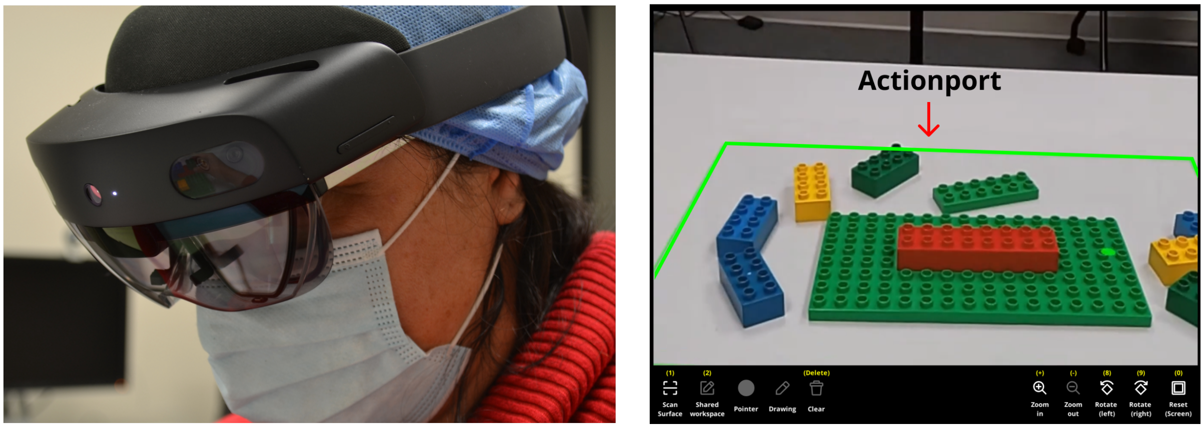


Figure 1: Left: Local healthcare worker receiving instructions in AR HMD. Right: Remote helper's system with Actionport.

## Abstract

In the following position paper we introduce the use of AR HMD for remote instruction in healthcare and present the challenges our team has faced in achieving this application in two contexts: surgical telementoring and paramedic teleconsulting. After the presentation of how these challenges come to be and indications on how to address them, we argue that those who wish to pursue this area of research must be grounded in best practices from the field of CSCW integrated with technical innovations in AR interaction development. This is a truly interdisciplinary research and development area that has many challenging topics to tackle through collaborative efforts.

**Index Terms:** Human-centered computing—Visualization—

## 1 Introduction

Collaborative systems that provide shared visual information through shared workspaces have been studied for decades in the field of computer supported cooperative work (CSCW), and their benefits when remote experts guide local novices during physical tasks are well known [5,7,8]. For example, shared workspaces improve situation awareness and are resources during conversational grounding [7] improving communication between collaborators [8], and virtual pointing in shared workspaces improves performance by reducing movement quantity for physical tasks [5] as implicit guidance becomes explicit visual cues [4].

These benefits have been studied for shared workspaces composed of the live video feed from a fixed camera pointing at the local worker's physical space, and an overlaid layer where the remote helper can point and annotate. The local worker sees this shared digital workspace through a separate monitor, which means that they need to switch their attention between the real world and the virtual. Now, when novices wear an Augment Reality (AR) Head Mounted Display (HMD), they can see the annotations experts produce overlaid directly onto the real world - seemingly a better orientation and setup for conveying instructions during hands-on tasks.

Thus, industry developers and academic researchers alike have begun designing for remote instruction with AR HMDs. For instance, Microsoft touted the benefit of the Hololens for remote instruction [9] and researchers have been studying the benefits of augmenting the local worker's view with virtual arms for pointing by a remote helper [10, 11].

In the following position paper, we will present the challenges our team has faced in achieving the application of AR HMDs in two telemedicine contexts: surgical telementoring and paramedicine teleconsulting. We argue that these challenges are unique in both AR and CSCW research, and require innovative solutions beyond what has been considered to date.

## 2 Healthcare Environments and AR Applications

### 2.1 Surgical Telementoring

Much of our work has been in the realm of laparoscopic surgery. Laparoscopy adds specific complexities on top of those of open surgery, as now perception and action are decoupled: surgeons operate by looking at a 2D monitor that displays the live video of a laparoscope (camera), and perform actions through elongated instruments, each of which are inserted into the patient through small incisions on the body. However, expert surgeons that can teach new skills and techniques are not always locally available. The Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) has promoted telementoring as an advanced educational approach to improve laparoscopic training and wider patient care [16]. Through this approach, an expert surgeon can guide another surgeon remotely on new techniques, procedures, and tools. This enables a broader transfer of knowledge and improves patient care in remote regions that do not have access to specialized knowledge.

*e-mail: mentis@umbc.edu
†e-mail: ignacio.avellino@sorbonne-universite.fr
‡e-mail: jwawon.seo@umbc.edu

## 2.2 Paramedic Teleconsulting

A more recent healthcare context for remote instruction that we have been working in is paramedicine. Access to emergency medical services (EMS) for both rural and urban patients is hampered by an overstressed healthcare workforce [13]. With the recent opioid crisis, COVID-19 pandemic, and growing aging population, many people are best served through "hospital at home" medical care as they do not need the resources of an emergency room or hospital [2]. EMS Agenda 2050: A People-Centered Vision released in 2019 created a bold vision for EMS in the United States at the culmination of a three year task force [1]. Among other futuristic ideals, the concept of telemedics was put forth – a large regional network of entry-level medical personnel who can provide immediate care in the field by tele-consulting with highly-trained medical personnel [17]. These consultants would provide everything, from a second opinion to specific instructions on how to provide care for a significant health crisis. We imagined the use of AR to provide a shared workspace that is hands-free for the telemedic to simultaneously provide care and receive instruction from a remote expert.

# 3 Challenges Across both Healthcare Contexts

## 3.1 Challenge 1: Multiple Local Workers

One of the challenges that exists in both of the healthcare contexts described above is that the care is not usually provided by a single clinician. Usually there are two or more surgeons or medics who are part of the care team – working together to achieve the desired health outcomes and, thus, consistently monitoring and maintaining awareness of one another's activities and intentions.

In laparoscopic surgery, for instance, there are two surgeons typically on either side of the patient table – coordinating their maneuvers with the instruments and scope and each focusing their attention on the actions displayed in their own monitors. This is in addition to the nurses and other techs who are monitoring the surgeons' actions in order to provide assistance or equipment at the right time. Likewise, in paramedicine, two or more medics arrive on a scene to provide care. If a patient has multiple injuries, for instance, from a car accident, one medic may be splinting the patient's leg while another is bandaging the patient's head or attending to another passenger in the car.

When a trainer or expert is collocated, all who are in attendance hear what the trainer/expert says. Likewise, as telementoring systems have typically been applied, there is a broadcast video and audio channel from the remote helper that everyone in the vicinity can hear/see and incorporate the conveyed information into their own work patterns. The problem arises when an individual AR HMD system is used - in essence, the HMD creates a private space. This is because of the 1-on-1 conversation between the remote helper and one of the local workers. The rest of the team only understands half the story when it comes to the conversation between the two, as they can only hear one side. But, in team-based healthcare, good situation awareness is key, knowing what others are doing and talking about reduces errors, and improves anticipation of others' actions.

A seemingly simple solution to this social problem would be to introduce additional AR HMD systems for the other local workers to wear. However, this, in turn, brings up the problem of referencing and annotating different perspectives or orientations.

These two AR HMD situations provide both technical and social problems. In the case of surgical telemedicine, where the local workers are working on the same physical space, the orientation of one's body to that space may be different. Thus, if the remote surgeon is annotating a line over the live video of the patient's abdomen that is provided by the head-mounted camera of the surgeon standing on the left side of the table, a line in the proper orientation needs to be displayed for the surgeon standing on the right side of the table.

In addition, the different perspectives of the local workers demand high cognitive load on the remote helper. In the telemedic example, when receiving information from two or more local medics using AR HMD, the remote helper would then be managing two simultaneous scenes from the local workers. Even if the medics are working together on the same patient concern, they may have two different orientations and foci on the concern that requires orientation specific instructions (e.g. "to the right"). Requiring the remote helper to constantly shift focus between two scenes and know whose scene is displayed can lead to high cognitive load and errors [15]. Therefore, it is necessary to understand how the remote helper manages different perspectives to give precise instructions or annotations.

## 3.2 Challenge 2: Mobile Local Workspace Views

In our current work for designing AR HMD for remote instruction in healthcare, we have identified an interaction challenge that is due to the head-mounted camera: the view sent to the remote instructor is not stable. Thus, as the expert uses a mouse to click on the view and draw an annotation, the result not only depends on their actions but now also on the actions of the local worker. As it stands now, Microsoft Hololens with Dynamics 365 Remote Assist requires a snapshot of the view to be taken in order to then remotely gesture or annotate the view.

Ideally for a remote helper to augment an AR HMD's view, they would click and draw with their mouse on the live video. The system would then make a projection of the point into the 3D world and stabilize the annotation. And so, when the local person wearing the HMD shifts their view, the point stays overlaid in the correct location with respect to the real world. The problem that we have identified, is that, if the local worker is looking to the right when the remote expert is drawing, and then moves their head to the left, the intended annotation will be adversely affected by the head movement. Thus, there is a gap between the 2D screen where the remote instructor interacts and the 3D environment where the remote instructor perceives.

In addition, there is an added complexity when having to translate 2D interactions to a 3D view. Referencing and drawing of annotation on a 2D visual display are well known as very simple and suitable means to convey information in collaboration. However, a remote helper's annotations on a 2D screen are not delivered as intended due to perspective problems in AR environments [14]. Lin et al. [12] proposed a system that can translate 2D annotations drawn on a tablet to be optimized for a 3D environment using a fixed overhead camera to calibrate 3D geometry. Although this system addressed a solution for translating 2D interactions to a 3D view, they did not consider a mobile view in the annotation system as described above. Thus, solutions that have been found for fixed AR views need to be reconsidered and innovated upon for a true AR HMD application as we have in many healthcare contexts.

# 4 Interdisciplinary Research to Address These Challenges

When applying AR HMD to a healthcare context, the specificities and interactions that exist "in the wild" quickly provide unique challenges not encountered in more controlled conditions. What we are finding with the use of AR HMD in remote instruction are both socio and technical problems. These are problems that may either have socio or technical solutions. But more importantly, one solution may impart the other's problems – i.e. a socio-solution may give rise to a technical need not realized before or vice versa.

We argue that, in applying AR to the very real world problem of healthcare, we are delving into a very interdisciplinary research space that can benefit from those researchers who focus on socio-technical work (i.e. CSCW) and those who have the necessary expertise in techno-mathematical solutions. Although collaborative research like this has been regularly called upon, we use this opportunity to highlight, yet again, such an important problem area that truly can benefit from interdisciplinary collaboration and solutions.

### 4.1 Example Collaborative Exploration

We can provide a concrete example of how this played out on a recent research project. Specifically, the following story outlines how our collaborative team had to negotiate the socio-technical problem of remote instruction over live mobile views.

When we first were considering the use of the Microsoft HoloLens in Paramedic Consulting, we were surprised to find that Dynamics 365 Remote Assist only enables a remote helper to add annotations (e.g. arrows) on a *snapshot* of the local worker's view to support the local worker in performing the task. When a user selects a pen icon and clicks the live view to start drawing, the view is frozen automatically. After drawing, the user clicks "stop editing" and then the drawing is transferred to the local worker's view [1]. In addition, Remote Assist does not provide a pointer that can represent the remote helper's dynamic cursor movements over that snapshot or the live view. From our prior work in surgical telementoring as well as other prior work in the field of CSCW, we knew that deictic referencing was an important part of collaboration (e.g. [3,6]). Thus, the team immediately agreed that we needed to find a mechanism for pointing and annotating a live mobile view as one would naturally do on a static video image.

One of the first solutions we came to was the use of a world-stabilized segment of the local worker's physical environment where a dynamic pointer can accurately reference and draw over a live mobile view – we termed these *actionports*. However, the real challenge in meeting this interaction need came from finding an input mechanism to accurately and intuitively interact within these ports.

The first option was the desktop application's mouse. However, this proved to not be ideal. Due to the potential movement of the head-mounted camera, we could not use the mouse as a *direct* input mechanism to the area of the actionport - i.e. hovering the mouse's cursor directly over the camera video to indicate where in the local worker's physical space to place the pointer hologram. As a remote helper's pointer would be placed in an area of the actionport, the result of the camera's movement would either offset the mouse cursor's location from the pointer's position on the camera view or relatively move the pointer away from the intended location. This is also the reason why a touchscreen could not be used - as a remote helper is touching an XY coordinate on the screen, any camera movement under their finger would offset the view and thus offset the pointer from its intended target.

In order to then use the desktop application's mouse to manipulate a holographic pointer in the actionport, we needed to translate the coordinates of interaction on a 2D display to a 3D plane that may lie in a different orientation. As you can see in Fig. 1 on the right, the borders of the actionport represented by the green box lies on the table (fundamentally the Y plane) while the angle of the camera and, thus, the video which is shown on the remote helper's display, is at a 45°offset. We transformed the coordinates of the mouse cursor's position in the window displaying the camera video to the corresponding coordinates in the actionport, assuming border alignment. Thus, when a remote helper would move their mouse towards the upper right corner of the camera view display window, the pointer hologram would correspondingly move to the upper right corner of the actionport. Finally, in order to not confuse the remote helper with two cursors/pointers on their display reacting to their mouse movements, we made the mouse cursor invisible. However, this led to even more confusion for the user as they did not know how far to move the mouse to get to a corner of the view without having that cursor feedback. In addition, there was a discrepancy in the degree of movement with the mouse when over the camera view display and the distance the pointer moved in the actionport. This was due to the size of the camera view window the remote instructor was actually interacting with and the size of perceiving the actionport that the pointer moves are different.

The final reason that the desktop application's mouse was not ideal is because the remote helper's desktop application consisted of a menu bar below a window encompassing the camera view. That menu bar offered functionality for the remote helper such as "scan surface", "set actionport" and "clear" to remove annotations. To select those buttons, the remote instructor used the desktop computer's mouse as well - i.e. the same mouse as used for interacting in the actionport. This meant that when the remote helper moved their mouse from the actionport to the button menu, they would cross through the camera view 'void' between the edge of the actionport and the start of the application (refer to Fig. 1 on the right to see this gap between the green outline of the actionport and the black outline of the camera view). This 'void' was because the position of the pointer can only be visible in the actionport, and so it was difficult for a user to know how much to move it out of the camera view in order to appear over the application's menu. Therefore, we identified the need to separate the interaction with the application frame and its menus from the interaction in the actionport by the remote helper.

Thus, we had to rely on an *indirect* input mechanism. In addition, we determined that an absolute position input device that supports the one-to-one mapping from the position of one's finger to the cursor location would be ideal. We, thus, devised **actionpad** for remote instructors to *act* on a tablet through touch and *perceive* the effect on the computer's screen displaying the actionport overlaid on the view of the local worker's physical space.

We have since tested this system in tele-instruction with a building blocks task. Our initial study has shown the ease of use of the system in addition to its ability to support the desired collaborative behaviors of pointing and annotating to effectively and efficiently convey instructions to a local worker completing a physical task.

This example shows a deep understanding of the research that comes from CSCW – in this case, the need to support deictic references which translated into the need for a dynamic pointer on a live view of the mobile camera – and how that motivated a process of determining a suitable interaction method for achieving that vision and seamlessly interacting with it.

## 5 Conclusion

In this position paper, we argue that the opportunity to capture CSCW best practices for remote instruction through an AR environment comes with some very unique challenges. These challenges demand innovative solutions that are quite different than needs from other application contexts. By outlining these challenges we hope to provide a call to other researchers to join us in devising innovative solutions and ensure that AR systems can be the next successful technological solution in collaborative healthcare.

## Acknowledgments

The authors wish to thank Azin Semsar who has collaborated with us on surgical telementoring and Anita Komlodi who has collaborated with us on paramedic teleconsulting. This work was supported in part by NSF IIS-1552837 and NSF BCS-2026510.

## References

[1] E. Agenda et al. Ems agenda 2050: A people-centered vision for the future of emergency medical services: Report no. dot hs 812 664. 2019.

[2] V. J. Dzau, M. B. McClellan, J. M. McGinnis, S. P. Burke, M. J. Coye, A. Diaz, T. A. Daschle, W. H. Frist, M. Gaines, M. A. Hamburg, et al. Vital directions for health and health care: priorities from a national academy of medicine initiative. *Jama*, 317(14):1461–1470, 2017.

[3] Y. Feng, H. McGowan, A. Semsar, H. R. Zahiri, I. M. George, A. Park, A. Kleinsmith, and H. Mentis. Virtual pointer for gaze guidance in laparoscopic surgery. *Surgical Endoscopy*, 34:3533–3539, Oct. 2019. doi: 10.1007/s00464-019-07141-x

[1] https://youtu.be/lYcYWlghfQQ?t=84

[4] Y. Feng, H. McGowan, A. Semsar, H. R. Zahiri, I. M. George, A. Park, A. Kleinsmith, and H. Mentis. Virtual pointer for gaze guidance in laparoscopic surgery. *Surgical endoscopy*, 34(8):3533–3539, 2020.

[5] Y. Feng, H. McGowan, A. Semsar, H. R. Zahiri, I. M. George, T. Turner, A. Park, A. Kleinsmith, and H. M. Mentis. A virtual pointer to support the adoption of professional vision in laparoscopic training. *International journal of computer assisted radiology and surgery*, 13(9):1463–1472, 2018.

[6] Y. Feng, H. McGowan, A. Semsar, H. R. Zahiri, I. M. George, T. Turner, A. Park, A. Kleinsmith, and H. M. Mentis. A virtual pointer to support the adoption of professional vision in laparoscopic training. *International Journal of Computer Assisted Radiology and Surgery*, 13(9):1463–1472, Sept. 2018. doi: 10.1007/s11548-018-1792-9

[7] D. Gergle, R. E. Kraut, and S. R. Fussell. Using visual information for grounding and awareness in collaborative tasks. *Human–Computer Interaction*, 28(1):1–39, 2013.

[8] C. Gutwin and S. Greenberg. A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work (CSCW)*, 11(3):411–446, 2002.

[9] M. HoloLens. Introducing dynamics 365 remote assist for hololens 2 and mobile devices, Apr 2020.

[10] S. Kim, G. Lee, M. Billinghurst, and W. Huang. The combination of visual communication cues in mixed reality remote collaboration. *Journal on Multimodal User Interfaces*, 14(4):321–335, 2020.

[11] M. Le Chénéchal, T. Duval, V. Gouranton, J. Royan, and B. Arnaldi. Help! i need a remote guide in my mixed reality collaborative environment. *Frontiers in Robotics and AI*, 6:106, 2019.

[12] C. Lin, D. Andersen, V. Popescu, E. Rojas-Munoz, M. E. Cabrera, B. Mullis, B. Zarzaur, K. Anderson, S. Marley, and J. Wachs. 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)*, pp. 3–8. IEEE, 2018.

[13] K. K. McGinnis. *Rural and frontier emergency medical services: Agenda for the future*. National Rural Health Association, 2004.

[14] B. Nuernberger, K.-C. Lien, T. Höllerer, and M. Turk. Interpreting 2d gesture annotations in 3d augmented reality. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 149–158. IEEE, 2016.

[15] E. Rojas-Muñoz, C. Lin, N. Sanchez-Tamayo, M. E. Cabrera, D. Andersen, V. Popescu, J. A. Barragan, B. Zarzaur, P. Murphy, K. Anderson, et al. Evaluation of an augmented reality platform for austere surgical telementoring: a randomized controlled crossover study in cricothyroidotomies. *NPJ digital medicine*, 3(1):1–9, 2020.

[16] C. M. Schlachta, N. T. Nguyen, T. Ponsky, and B. Dunkin. Project 6 summit: Sages telementoring initiative. *Surgical endoscopy*, 30(9):3665–3672, 2016.

[17] B. Schooley and T. Horan. Emerging digital technologies in emergency medical services: Considerations and strategies to strengthen the continuum of care. *National Highway Traffic Safety Administration, DOT HS 811 999c*, 2015.