

Robust High-Level Video Stabilization for Effective AR Telementoring

Chengyuan Lin^{*1}, Edgar Rojas-Muñoz¹, Maria Eugenia Cabrera¹, Natalia Sanchez-Tamayo¹, Daniel Andersen¹, Voicu Popescu¹, Juan Antonio Barragan Noguera¹, Ben Zarzaur², Pat Murphy², Kathryn Anderson², Thomas Douglas³, Clare Griffis³, and Juan Wachs¹

¹Purdue University, ²Indiana University School of Medicine, ³Naval Medical Center Portsmouth

ABSTRACT

This poster presents the design, implementation, and evaluation of a method for robust high-level stabilization of mentees first-person video in augmented reality (AR) telementoring. This video is captured by the front-facing built-in camera of an AR headset and stabilized by rendering from a stationary view a planar proxy of the workspace projectively texture mapped with the video feed. The result is stable, complete, up to date, continuous, distortion free, and rendered from the mentee's default viewpoint. The stabilization method was evaluated in two user studies, in the context of number matching and for cricothyroidotomy training, respectively. Both showed a significant advantage of our method compared with unstabilized visualization.

Index Terms: Human-centered computing—Visualization—Visualization design and evaluation methods Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Collaborative interaction Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality;

1 BACKGROUND

As science and technology specialize ever more deeply, it is more and more challenging to gather in the same place the many experts needed to perform a complex task. Telecollaboration is an important approach for transmitting expertise over large geographic distances promptly and effectively.

A special case of telecollaboration is telementoring, where a mentee performs a task under the guidance of a remote expert. One approach is to rely only on an audio channel for the communication between mentor and mentee. Telestrators enhance communication with a visual channel—the mentor annotates a video feed of the workspace, which is then shown to the mentee on a nearby display. The challenge with this approach is that the mentee has to switch focus repeatedly away from the workspace, and to remap the instructions from the nearby display to the actual workspace, which could lead to a high cognitive load on the mentee and ultimately to task performance delays and even errors [3]. Augmented Reality (AR) technology can solve this problem by directly integrating the annotations into the mentee's field of view. The mentee sees the annotations as if the mentor actually drew them on the 3D geometry of the workspace, eliminating focus shifts.

A problem less studied but nonetheless significant is the ability for the AR telementoring system to convey the workspace to the remote mentor effectively. One approach is to acquire the workspace with an auxiliary video camera, and to send its video feed to the mentor. In addition to the obvious shortcoming of needing additional hardware,

the auxiliary camera captures the workspace from a different view from that of the mentee. Effective telementoring requires the mentor to see what the mentee sees for the instructions to be as relevant and easy to understand as possible. For example, when using an auxiliary camera, the mentor might annotate parts of the workspace that are not visible to the mentee due to occlusions.

2 PROBLEM

With the advancement of AR technology, self-contained optical see-through head mounted displays (HMDs) are now available. Such HMDs typically incorporate a front-facing onboard camera, which can capture the workspace from a viewpoint close to the mentee's viewpoint. However, simply providing the mentee first-person video to the mentor is insufficient for effective telementoring. As the mentee frequently changes head position and view direction, the mentors visualization of the workspace changes frequently and substantially, which can lead to high cognitive load, low task performance, and simulator sickness on the mentor side. For example, when the mentee looks to the left, the workspace visualization shifts by hundreds of pixels to the right; when the mentee moves on the other side of the workspace as might be needed for best access during task performance, the visualization rolls 180 degrees, which results in an upside-down visualization that is frustratingly difficult to parse. What is needed is a robust stabilization of the mentee first-person video, such that it can provide an effective visualization of the workspace to the mentor. Considerable research efforts have been devoted to developing low-level stabilization solutions for videos acquired with a shaking camera, e.g. a hand-held or bicycle helmet mounted camera [1, 2]. However, what is needed in our context is a method for high-level stabilization, capable of removing the effects of viewpoint translations and view direction rotations that are substantial with respect to the distance to and the size of the workspace.

3 APPROACH

In this poster, we present the design, implementation, and evaluation of a method for the robust high-level stabilization of a video feed acquired from the mentee's first-person view, for it to serve as an effective visualization for a remote mentor. The output visualization has to be stable, i.e. to show the static parts of the scene at a constant image location, real-time, i.e. to keep up with the input feed, and of high quality, i.e. without distortions, tears and other artifacts. In addition to conveying the workspace to the mentor, the output visualization should also be a suitable canvas on which the mentor can author annotations to provide guidance. After investigating several approaches, we determined that the approach that best satisfies all these design requirements is projective texture mapping of the tracked video feed onto a proxy plane of the workspace, which is then rendered from a stable viewpoint.

4 FIRST USER STUDY: MATCHING NUMBERS

We evaluated the effectiveness of our stabilization method in two user studies. In one study ($n = 30$), the stabilized visualization is

*email: lin553@purdue.edu

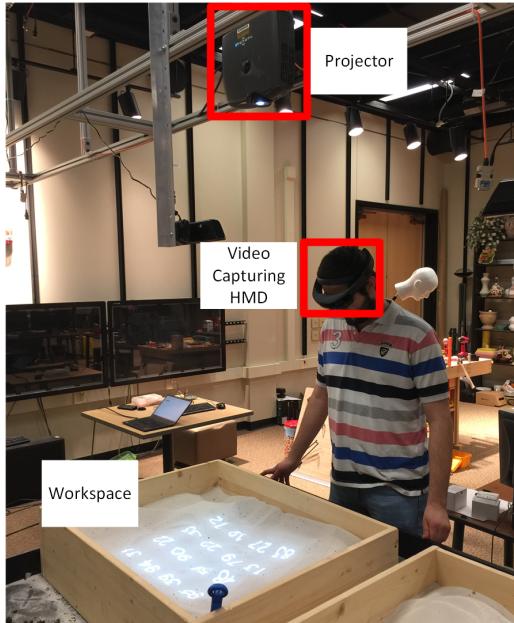


Figure 1: Acquisition of video used in number matching study.



Figure 2: Study participant trying to find matching numbers in video.

compared to unstabilized visualization and to ground truth (perfectly stable) visualization, in the context of a number matching task. The workspace is a sandbox onto which an overhead projector displays a 4 by 4 matrix of two digit numbers. The workspace video is acquired by an experimenter who walks around the sandbox, wearing an AR HMD that has a built-in camera (Fig. 1). The ground truth video is acquired with the AR HMD on a tripod. The participant sits in a chair and watches the video on a monitor (Fig. 2) and tries to find matching numbers.

Our method showed significant advantages in terms of task performance, quantified as the number of pairs of matching numbers, in terms of participant workload, quantified using the NASA-TLX questionnaire, and in terms of simulator sickness, quantified using the SSQ. The perfectly stable visualization showed no significant advantages over our stabilization method.

5 SECOND USER STUDY: AR TELEMENTORED PRACTICE CRICOHYOIDOTOMY

In a second study, our stabilization was tested in the context of surgical telementoring, where it was used in cricothyroidotomy training in an austere setting. The austere setting was simulated with an empty room with the patient on the floor, with low visibility achieved with a fog machine, and with loud, combat-like noises (Fig. 3a and Fig. 3b). Each participant ($n = 20$) performed a cricothyroidotomy in two conditions (i.e. a within-subjects design).

In the experimental condition, the mentee wore an AR HMD (Fig. 3a). The AR HMD has a built-in video camera that acquires

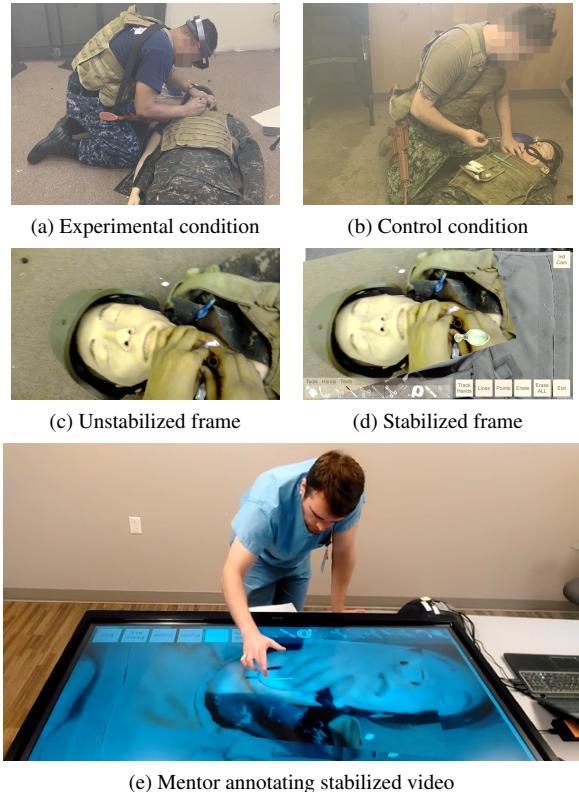


Figure 3: Second user study: practice cricothyroidotomy on synthetic patient simulator in an austere setting.

the patient from the mentee's vantage point. The video feed is transmitted to the remote mentor, where it is stabilized (Fig. 3c, Fig. 3d). The mentor annotates the feed through a touch-based interface to provide visual guidance to the mentee (Fig. 3e). The annotations are sent back to the mentee, where they are overlaid on the mentee's view of the patient using the AR HMD. The mentor and mentee could also talk to each other, through an audio connection. In the control condition, the mentor only guided the mentee verbally, through audio communication (Fig. 3b).

The cricothyroidotomy procedure performance of each mentee was evaluated by an expert surgeon, who assigned a score for each one of the ten steps of the procedure. The experimental condition had an advantage over the control condition for each of the ten sub-procedure scores. The advantage was significant for two sub-procedure scores, as well as for the overall procedure score, obtained by averaging the ten sub-procedure scores.

ACKNOWLEDGMENTS

We thank our Augmented Reality Tea group for insightful comments and suggestions. This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award No. W81XWH-14-1-0042, and by the NSF under Grant DGE-1333468.

REFERENCES

- [1] J. Kopf, M. F. Cohen, and R. Szeliski. First-person hyper-lapse videos. *ACM Transactions on Graphics (TOG)*, 33(4):78, 2014.
- [2] T. Lee and T. Hollerer. Viewpoint stabilization for live collaborative video augmentations. In *Mixed and Augmented Reality, 2006. ISMAR 2006. IEEE/ACM International Symposium on*, pp. 241–242. IEEE, 2006.
- [3] E. Rojas-Muñoz, M. E. Cabrera, D. Andersen, V. Popescu, S. Marley, B. Mullis, B. Zarzaur, and J. Wachs. Surgical telementoring without encumbrance: A comparative study of see-through augmented reality-based approaches. *Annals of surgery*, 2018.