Abstract

Introduction: Point-of-Injury (POI) care requires immediate specialized assistance, but delays and expertise lapses can lead to complications. In such scenarios, telementoring can benefit health practitioners by transmitting guidance from remote specialists. However, current telementoring systems are not appropriate for POI care. This paper clinically evaluates our System for Telementoring with Augmented Reality (STAR), a novel telementoring system based on an Augmented Reality Head-Mounted Display. The system is portable, self-contained, and displays virtual surgical guidance onto the operating field. These capabilities can facilitate telementoring in POI scenarios while mitigating limitations of conventional telementoring systems.

Methods: Twenty participants performed leg fasciotomies on cadaveric specimens under either one of two experimental conditions: telementoring using STAR; or without telementoring but reviewing the procedure beforehand. An expert surgeon evaluated the participants' performance in terms of completion time, number of errors, and procedure-related scores. Additional metrics included a self-reported confidence score and post-experiment questionnaires.

Results: System for Telementoring with Augmented Reality effectively delivered surgical guidance to non-specialist health practitioners: participants using STAR performed fewer errors and obtained higher procedure-related scores.

Conclusions: This work validates STAR as a viable surgical telementoring platform, which could be further explored to aid in scenarios where life-saving care must be delivered in a prehospital setting.

Introduction

When patient evacuation is infeasible, surgeons in the battlefield must administer critical lifesaving care to stabilize patients before transferring them to a Role 2 Care facility. Failure to provide adequate treatment in these situations can lead to risks such as limb amputations or even death.[1] Nonetheless, proper administration of critical care cannot always be guaranteed in austere scenarios and, thus, approaches to assist first responders at the Point of Injury (POI) have been explored, especially to provide assistance to non-specialists. Telementoring systems have been one of the most promising approaches for providing remote assistance effectively,[2-4] and their benefits has been previously demonstrated in contexts such as remote surgical assistance in trauma, emergency care and surgery.[5,6,7] Through this technique, non-specialist surgeons (mentees) can receive surgical guidance and coaching from remotely-located specialists (mentors).[4,8,9]

However, conventional telementoring systems are not designed to provide POI care, where undesired encumbrance and time delays must be avoided. Most conventional platforms rely on telestrators to convey surgical expertise.[10] In the telestrator approach, a remote mentor annotates a live view of the mentee's operating field with graphical surgical instructions, which a mentee then visualizes on a nearby display. Albeit effective, the approach requires the mentee to map the annotations from the telestrator to the actual operating field, which translates to extra cognitive loads that can lead to sub-optimal performance and even errors.[11,12] Augmented Reality (AR) technologies have been explored to address the issues introduced by telestrators by displaying the surgical guidance directly into the mentees' field of view (FOV), obtaining promising results.[13-15] Nonetheless, prior AR-based telementoring systems are not suited for POI scenarios. For example, some of these approaches propose the use of tablets in the operating field, limiting the surgeons' free selection of movements and degrading their depth perception due to loss of stereopsis.[15,16]

This paper presents a clinical validation of the System for Telementoring with Augmented Reality (STAR), a portable and self-contained telementoring platform based on an AR Head-Mounted Display (ARHMD).[17] This system can alleviate the aforementioned issues of conventional telementoring systems by displaying expert-authored surgical instructions directly into the FOV of mentees wearing the device. This clinical validation is considered a first step before evaluating our platform in an austere scenario, and addresses the question of whether this ARHMD-based platform can effectively be used to transmit surgical guidance remotely.

Methods

The study had ethical approval from Indiana University's Institutional Review Board (#1409037680), and written participant consent was acquired for each participant. Every participant was briefed on the logistics of the experiment before starting, and were able to withdraw from the experiment at any point. The experiment evaluated whether mentees receiving remote mentoring using the STAR platform can perform a surgical procedure better, faster and with fewer mistakes than participants who reviewed the procedure beforehand and then performed the procedure without mentoring. This evaluation was performed by doing leg fasciotomies on cadaveric specimens, a procedure relevant to combat zones, performed to restore blood flow to limbs affected by compartment syndrome.[14,18-20] This procedure is an ideal testbed for a telementoring scenario, as it requires specialized and immediate care but often has to be administered by non-expert trauma surgeons in the battlefield.

Twenty participants performed a leg fasciotomy under two experimental conditions: receiving remote guidance using the telementoring platform (*STAR* condition); or without external guidance beyond initial consultation of the Advanced Surgical Skills for Exposure in Trauma course manual (*Control* condition). The *Control* condition was selected because it can approximate a scenario in which a frontline surgeon must perform a procedure without any on-site help.

Materials

System for Telementoring with Augmented Reality is surgical telementoring platform that allows local mentees to visualize expert-authored surgical instructions directly into their FOV using AR.[15,17] The STAR platform is self-contained and portable, making it suitable for austere, pre-hospital scenarios. For this experimental setup, a top-down camera at the mentee site acquires a live view of the operating field, which is transmitted to the remote mentor. This video is displayed on a large interactive display, which the mentor can use to create two-dimensional (2D) surgical annotations via touch interactions. These annotations are sent to the ARHMD worn by the mentee, which displays the annotations as three-dimensional (3D) imagery superimposed onto the mentee's FOV of the patient's body. This mapping from mentor-authored 2D annotations to 3D annotations is possible thanks to the ARHMD's active depth acquisition of the environment, and a one-time calibration routine that maps the position of a top-down camera to the current position of the ARHMD. Figure 1 includes an augmented 3D annotation, visualized from the mentee's perspective. The ARHMD displays a different image for the left and right eyes, and this disparity provides the correct depth cue to the mentee, who sees the annotation as if it were actually drawn onto the patient.

The ARHMD platform was built based on four main design considerations: 1) overlaying mentor-authored graphical annotations directly into the mentee's FOV of the surgical field 2) showing the annotations to the mentee with correct binocular depth cues by rendering the annotations at their correct 3D location through stereo visualization; 3) avoiding workspace encumbrance that other state-of-the-art telementoring systems introduce;[15,21,22] and 4) providing the mentor with an adequate view of the operating field from an external top-down camera, to preserve the mentor's situational awareness.[23] Such visualization prevents potential disorientation and motion sickness that could arise from using the ARHMD's on-board camera visualization without proper image stabilization techniques.[24]

Procedures

The experiment took place between a Level 1 Trauma Center and a medical school campus, located approximately 1,650 ft. away from each other. The setup at the mentor site included a large interactive display and a conference speakerphone to communicate with the mentee via audio. The setup at the mentee site included two separate operating stations (divided by a mobile curtain) to perform the procedures under two mentoring conditions: *Control* vs. *STAR*. Figure 2 showcases the mentor (*right*) and mentee (*left*) sites, with an artist rendition of an augmented 3D annotation at the mentee site. Each station was equipped with one operating table, a set of surgical lights, a tripod-mounted camera, and a surgical tray with various surgical instruments. One nurse assistant (per operating station) stood by the mentees, assisting them during the procedure in aspects such as cleaning fluids and passing instruments. Additionally, the *STAR* condition included the ARHMD-based platform worn by the mentee, a phone for audio communication with the mentor, and a camera attached to the overhead surgical lights.

Participants

Residents and medical students were recruited as participants for the experiment. These populations were selected because of their relatively lower expertise levels when compared to expert surgeons: participants were expected to rely on the external sources of information rather than performing the procedure based on their own knowledge and experience. This setup has been considered as an adequate placeholder for a non-specialist attempting to do a procedure without having the proper expertise.[25-28] Additionally, a stratified sampling approach was adopted to recruit participants with various levels of expertise. Three expertise-based sub-groups were considered: medical students, residents with one year of post-graduate experience, and residents with two or more years of post-graduate experience. This stratified sampling was used to see the effect of the telementoring experience in a leg fasciotomy with respect to the participants' level of expertise.

Metrics

The participants' procedural performance was evaluated with the Individual Procedure Score (IPS),[29] adapted for fasciotomy procedures by expert surgeons of the research team. Individual Procedure Score is a metric that assesses whether a training course effectively improves the mentees' overall surgical expertise, covering both technical and non-technical aspects. In addition, IPS includes the Global Rating Score (GRS), a five-level Likert scale evaluating aspects such as management and readiness to complete the procedure independently; and the Evaluator Overall Rating (EOR), the evaluator's subjective rating (1 to 100) of how well the participant performed.

Additional evaluation criteria considered the number of errors during the procedure (from 11 possible errors highlighted as part of IPS' evaluation), procedure completion time, and a self-reported confidence value describing the participants confidence to independently perform fasciotomies. These self-reported confidence scores were obtained both before and after the procedure. Finally, participants completed a post-experiment questionnaire evaluating their assigned mentoring condition.

Weighted IPS for fasciotomy procedure

The IPS metric used in this work included six sections: Anatomical Landmarks (A), to be highlighted before any incision is performed; Anterolateral Incision (B), where the participants needed to identify and release the fascia from the anterior and lateral compartments; Posteromedial Incision (C), where participants were required to identify and release the superficial posterior and the deep posterior compartments; and three general performance sections, namely Technique Points (D), Operative Field Maneuvers (E), and Instrument Use (F).

The metric, however, assumes that each of these six sections is equally important when performing a fasciotomy. To relax this assumption, this work introduces a weighted IPS (WIPS) to reflect the unequal importance of the different sections. Nine experienced Attending surgeons were surveyed to determine the relevance of the IPS sections. The survey revealed the following relationship: $A > B = C > E > D > F$; e.g. identification of Anatomical Landmarks was determined to be the most important section, while Instrument Use was determined to be the least important. This unequal importance is represented with numerical weights that are used to normalize the IPS values, leading to a higher fidelity score. The survey revealed the following section importance weights: A → 0.236, B → 0.217, C → 0.217, D → 0.106, E → 0.162, and F → 0.062.

Self-reported confidence

A self-reported confidence score evaluated how confident the participants felt towards performing a leg fasciotomy. This assessment contained 4 questions with 5-point Likert scales inquiring their confidence regarding anatomical landmarks, procedural steps, technique and handling instruments, as well as how confident they felt to perform the procedure on their own.

This confidence score was reported both before and after the experiment to obtain insights about how the mentoring experience affected the participants' confidence level.

Post-experiment questionnaire

The questionnaire included eight 5-point Likert scale questions regarding the participant's experience using the assigned mentoring method. The questions evaluated the mentoring method in terms of its effectiveness, of the frustration it generated, and of how similar it felt to in-person side-by-side mentoring. A section for comments was included at the end of the survey.

Statistical Analysis

A between subject design (with N = 10) statistical analysis was run to compare the mentoring conditions, treating the mentoring conditions as independent variables and using the evaluation metrics as dependent variables. The data's normality assumption was checked with the Shapiro-Wilk test.[30] Non-normal unpaired populations were compared with the nonparametric Mann-Whitney U test, whereas paired populations were compared with the nonparametric Wilcoxon signed-rank test.[31,32] Levene's test was run to assess normal data's equal variance assumption.[33] A 2-sample t-test was used over the data afterwards. Normally distributed data are presented as mean ± standard deviation, and non-normal distributed data as median ± interquartile range.

Findings

Fourteen residents and 6 medical students (12 male, 6 female, 2 N/A; age range of 28.5 ± 3.3) were assigned to the two conditions (seven residents and three medical students per condition). The medical students were in their first (two), second (one), and fourth (three) year of training. Residents varied between first (four), second (three), third (three), and fourth (four) year of general and orthopaedic surgery training. Given this distribution, the aforementioned expertise-based sub-groups were constructed as follows: only medical students (N = 6), only residents (N = 14), and medical students and first-year residents combined (N = 10). The latter sub-group was considered the one that would benefit the most from receiving expert guidance due to their lower level of expertise.

An expert Attending surgeon evaluated the participants on-site. Table 1 summarizes the quantitative metrics averaged over all participants in each condition: completion time (in seconds), IPS, WIPS, number of errors, GRS and EOR. Participants in the *STAR* condition performed significantly fewer errors ($p = 0.03$) and obtained significantly higher WIPS ($p = 0.04$). No statistically significant difference was found in term of procedure completion time. Likewise, the GRS and EOR metrics did not show statistical significance because of their large variances and were not consider for further analysis.

Individual Procedure Score related metrics

Table 2 summarizes the IPS, WIPS and number of errors of the expertise-based sub-groups. Medical students in the *STAR* condition performed significantly fewer errors ($p = 0.02$) and obtained a significantly higher WIPS ($p = 0.05$). Conversely, the sub-group of only residents did not report statistically significant differences between the conditions for any of the metrics. Finally, participants in the *STAR* condition from the medical students plus first year residents sub-group performed significantly fewer errors ($p < 0.001$) and obtained significantly higher IPS and WIPS values ($p = 0.04$ and $p < 0.01$ respectively).

Self-reported confidence assessment

Participants in the *STAR* condition had an average confidence increase of 1.28 (on a 5-point scale), whereas participants in the *Control* condition reported an average increase of 0.675.

This difference was statistically significant (p = 0.017). The results of the increases in the participants' confidence scores are summarized in Table 3. The table presents the average difference between the pre and post-experiment scores, and compares whether the increases in confidence reported by the participants were significantly different. Overall, participants in the *STAR* condition reported lower confidence scores than participants in the *Control* condition at the beginning of the experiment. This was reported in their scores of anatomical landmarks identification (3.00 ± 1.25 for *STAR*; 3.50 ± 1.00 for *Control*), instrument handling (3.0 ± 2.0 for *STAR*; 4.0 ± 1.5 for *Control*), and capacity to perform the procedure alone (2.50 ± 1.25 for *STAR*; 3.00 ± 1.25 for *Control*). These initial differences between the conditions were not statistically significant. Nonetheless, participants in the *STAR* condition reported higher or equal confidence scores than participants in the *Control* condition after to the experiment.

Post-experiment questionnaire

Participants significantly preferred STAR in terms of the amount of information it provided and the ease of following instructions. Participants also considered that STAR generated significantly less frustration. Additionally, although participants considered both mentoring techniques worse than side-by-side mentoring, participants considered the *Control* condition to be worse than the *STAR* condition when compared to side-by-side mentoring. Finally, some participants in the *STAR* condition commented that the weight of the ARHMD causes strain in the neck and shoulders, which interfered with their normal surgical posture and comfort. Other participants also mentioned that the augmented 3D annotations appeared to be drifting in space (approximately 1cm) from the mentee's perspective relative to the operating field, which they found confusing. Supplemental Table 1 presents the average results for each question of the survey.

Discussion

Participants in the *STAR* condition performed significantly fewer errors when performing the procedure. This is a critical outcome: not only identifying and avoiding surgical errors is a predictor for technical skill and performance, but it is critical for ensuring patient safety.[34,35] Additionally, participants in the *STAR* condition obtained higher technical and non-technical skill scores, as reported by the IPS metric. With respect to the WIPS metric, the technical and non-technical skill scores obtained by participants in the *STAR* condition were significantly higher. Overall, the results obtained demonstrate that participants using our novel telementoring platform increased their overall procedure performance.

The self-reported confidence scores indicate that participants in the STAR condition had a significantly higher increase in their confidence about performing fasciotomies. A breakdown of the differences revealed that, before to the experiment, participants in the *Control* condition reported to be more confident than participants in the *STAR* condition. However, the scores revealed the opposite trend after the experiment: participants in the *STAR* condition reported equal or higher confidences than participants in the *Control* condition. Although studies have shown that the confidence of health practitioners' in their surgical skills is correlated to their competence,[36,37] surveys have reported that some populations are dissatisfied with their confidence level.[38] Our results demonstrate that a telementoring experience with STAR's ARHMD can have a positive impact on the surgical skills confidence of health practitioners. These results can be extrapolated to indicate that integrating STAR with current coaching and training programs can be beneficial to reinforce the confidence, and therefore competence, of health practitioners.

The post-experiment questionnaires also revealed that participants preferred the *STAR* condition in all the evaluated criteria. Participants in the *STAR* condition reported significantly better results in aspects related to the information provided, to the ease of following this information, and to the overall frustration generated throughout the experiment. These aspects indicate that participants considered STAR as an effective method to receive external surgical guidance without negatively impacting their overall performance. These results also indicate that STAR could be a useful technique to implement as part of training and coaching programs of health practitioners, as mentees could use the platform to receive extra guidance when learning how to perform new surgical procedure. Furthermore, STAR could alleviate the mentor availability bottleneck, by connecting one mentor to multiple mentees. However, further studies should be performed to validate STAR as a training tool.

Analyzing the expertise-based sub-groups revealed peculiar trends. First, medical students in the *STAR* condition reported significantly fewer errors and a significantly higher WIPS. While the sample size was low, this trend suggests that our ARHMD-based platform led to better overall surgical performance on fasciotomies. The third sub-group, that enlarged the student group with first-year residents, further confirmed the benefit of STAR for mentees with a relatively lower expertise. Contrariwise, no significant difference between the conditions was found for the residents only sub-group. This suggests that receiving telementoring for a relatively simple procedure like a leg fasciotomy was irrelevant for experienced surgery residents, particularly the ones with more than two years of post-graduate experience at a Level-1 trauma center. For example, one experienced surgical resident in the *Control* condition was able to complete the procedure by heart, obtaining a near-perfect score. We hypothesize that these trends were observed due to the relative low difficulty of the procedure, and that evaluating our system with more complex procedures could reveal benefits even to health practitioners with high expertise levels.

The experimental apparatus introduced some limitations in the study. First, a within-subjects design could have been used to evaluate the effect of each mentoring condition in each participant. Another limitation is the sample size and distribution, as it was not possible to recruit enough participants from each stratum to perform a complete statistical design. Further studies should be performed to corroborate the findings reported from each of this different sub-groups. Moreover, although prior work in the STAR platform compared the ARHMD device against another common AR telementoring method (telestrator),[17] further studies should be performed to compare STAR against another telementoring technique used in POI scenarios such as audio-only communication. Furthermore, some technical limitations need to be addressed before adopting ARHMD technologies to provide remote assistance during Role 1 Care scenarios, namely the weight of the ARHMD and the phenomenon of annotations appearing to drift in space. Additionally, more general limitations for POI scenarios include aspects such as the bulkiness of the equipment, mostly introduced in our setup via the external top-down camera. Currently, we are working on replacing this top-down camera with a stabilized first-person view acquired with the ARHMD's on-board camera. This aspect will enhance STAR's portability, a valuable aspect in POI settings. Finally, the internet connection quality in POI scenarios can be low and fluctuating, which can impact the transmission of video and annotations of our platform. To tackle this limitation, the system is currently being enhanced with autonomous mentoring capabilities which should complement or even take over mentoring when the connection to the human mentor degrades.

One of the intended uses of our system is to assist in austere environments such as conflict zones and rural areas. In these scenarios, mentees would not have any other resource to refresh their skills other than watching videos of procedures, or resort to a surgical course manual. As a first order approximation of these scenarios, our experiment indicates that STAR can excel by transmitting real-time assistance through visual annotations from a remote expert surgeon. Further studies in less controlled settings are required before transferring this technology to POI settings, which should include aspects as anesthesia application processes and maintaining hemostasis. Nonetheless, the current results are promising in indicating that the STAR platform can effectively communicate mentees with remote specialists to receive assistance while performing a procedure in which they are not proficient.

Conclusion

This paper presented a clinical validation of STAR in the context of leg fasciotomies on cadaveric specimens. System for Telementoring with Augmented Reality leverages an ARHMD that allows health practitioners to visualize 3D surgical guidance from remote specialists directly in their FOV, and without obstructing their workspace. The portable and self-contained design of the system can be beneficial in POI scenarios, where encumbrance and delays cannot be tolerated. The performed clinical validation compared participants performing the procedure while receiving remote guidance through STAR against participants that received no guidance other than reviewing the procedure steps on a surgical course manual. Participants in the *STAR* condition completed the fasciotomies with significantly fewer errors, and with significantly higher procedural scores. In addition, they reported higher confidence in their abilities to perform fasciotomies after the experiment. Finally, a usability questionnaire revealed that participants preferred receiving remote guidance with our telementoring system. These results validate

STAR's ARMHD potential to guide unexperienced mentees through a procedure that is frequently performed in combat zones by surgeons without trauma expertise. Our portable and self-contained solution for surgical telementoring can be a key contribution to enhance surgical performance in austere settings, which should be explored in subsequent studies conducted in realistic POI scenarios.

References

1. Eastridge BJ, Mabry RL, Seguin P, et al.: Death on the battlefield (2001–2011): implications for the future of combat casualty care. J Trauma Acute Care Surg 2012; 73(6): S431–7.
2. Garcia P: Telemedicine for the battlefield: present and future technologies. In: Surgical Robotics, Boston, MA, Springer, 2011: 33–68.
3. Glenn IC, Bruns NE, Hayek D, Hughes T, Ponsky TA: Rural surgeons would embrace surgical telementoring for help with difficult cases and acquisition of new skills. Surg Endosc 2017; 31(3): 1264–8.
4. Boedeker D, Numanoglu A, Hall T, Boedeker B: Description of a novel telemedicine imaging system which could be used to teach surgical burn management techniques by telementoring in remote areas. J Int Soc Telemed EHealth 2017; 5: e12.
5. Doarn CR, Latifi R: Telementoring and Teleproctoring in Trauma and Emergency Care. Curr Trauma Rep 2016; 2(3): 138–43.
6. Latifi R, Hadeed GJ, Rhee P, et al.: Initial experiences and outcomes of telepresence in the management of trauma and emergency surgical patients. Am J Surg 2009; 198(6): 905–10.

7. Kirkpatrick AW, McKee JL, McBeth PB, et al.: The Damage Control Surgery in Austere Environments Research Group (DCSAERG): a dynamic program to facilitate real-time telementoring/telediagnosis to address exsanguination in extreme and austere environments. J Trauma Acute Care Surg 2017; 83(1): S156–63.

8. Greenberg CC, Ghousseini HN, Quamme SRP, et al.: A statewide surgical coaching program provides opportunity for continuous professional development. Ann Surg 2018; 267(5): 868–73.

9. Bogen EM, Augestad KM, Patel HR, Lindsetmo R-O: Telementoring in education of laparoscopic surgeons: an emerging technology. World J Gastrointest Endosc 2014; 6(5): 148-55.

10. Budrionis A, Augestad KM, Patel HR, Bellika JG: An evaluation framework for defining the contributions of telestration in surgical telementoring. Interact J Med Res 2013; 2(2): e14.

11. Patrick E, Cosgrove D, Slavkovic A, Rode JA, Verratti T, Chiselko G: Using a large projection screen as an alternative to head-mounted displays for virtual environments. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2000: 478–85.

12. Downs RM, Stea D: Image and Environment: Cognitive Mapping and Spatial Behavior. Transaction Publishers, 1974.

13. Shuhaiber JH: Augmented reality in surgery. Arch Surg 2004; 139(2): 170–4.

14. Talbot M, Harvey E, Berry G, et al.: A pilot study of surgical telementoring for leg fasciotomy. J R Army Med Corps. 2018;164(2):83–6.

15. Andersen D, Popescu V, Cabrera ME, et al.: Medical telementoring using an augmented reality transparent display. Surgery 2016;159(6):1646–53.

16. Luursema J-M, Verwey WB, Kommers PA, Annema J-H: The role of stereopsis in virtual anatomical learning. Interact Comput 2008; 20(4–5): 455–60.

17. Rojas-Muñoz E, Cabrera ME, Andersen D, et al.: Surgical telementoring without encumbrance: a comparative study of see-through augmented reality-based approaches. Ann Surg 2018; in-press.

18. Kragh JF Jr, Wade CE, Baer DG, et al.: Fasciotomy rates in Operations Enduring Freedom and Iraqi Freedom: association with injury severity and tourniquet use. J Orthop Trauma 2011; 25(3): 134–9.

19. Kragh JF Jr, Dubick MA, Aden JK III, et al.: US Military experience from 2001 to 2010 with extremity fasciotomy in war surgery. J Mil Med 2016; 181(5): 463–8.

20. Shackelford SA, Butler Jr FK, Kragh JF Jr, et al.: Optimizing the use of limb tourniquets in tactical combat casualty care: TCCC guidelines change 14-02. Army Inst Of Surgical Research, Fort Sam, Houston TX, 2014.

21. Vera AM, Russo M, Mohsin A, Tsuda S: Augmented reality telementoring (ART) platform: a randomized controlled trial to assess the efficacy of a new surgical education technology. Surg Endosc 2014; 28(12): 3467–72.

22. Shenai MB, Dillavou M, Shum C, et al.: Virtual interactive presence and augmented reality (VIPAR) for remote surgical assistance. Oper Neurosurg 2011; 68(suppl_1): ons200–7.

23. Morin A: Levels of consciousness and self-awareness: a comparison and integration of various neurocognitive views. Conscious Cogn 2006; 15(2): 358–71.

24. LaViola JJ Jr: A discussion of cybersickness in virtual environments. ACM SIGCHI Bull 2000; 32(1): 47–56.

25. Snyderman CH, Gardner PA, Lanisnik B, Ravnik J: Surgical telementoring: A new model for surgical training. The Laryngoscope 2016; 126(6): 1334–8.

26. Rafiq A, Moore JA, Zhao X, Doarn CR, Merrell RC: Digital video capture and synchronous consultation in open surgery. Ann Surg 2004; 239(4): 567-73.

27. Panait L, Rafiq A, Tomulescu V, et al.: Telementoring versus on-site mentoring in virtual reality-based surgical training. Surg Endosc Interv Tech 2006; 20(1): 113–8.

28. Sereno S, Mutter D, Dallemagne B, Smith C, Marescaux J: Telementoring for minimally invasive surgical training by wireless robot. Surg Innov 2007; 14(3): 184–91.

29. Mackenzie CF, Garofalo E, Shackelford S, et al.: Using an individual procedure score before and after the advanced surgical skills exposure for trauma course training to benchmark a hemorrhage-control performance metric. J Surg Educ 2015; 72(6): 1278–89.

30. Shapiro SS, Wilk MB: An analysis of variance test for normality (complete samples). Biometrika 1965; 52(3/4): 591–611.

31. Mann HB, Whitney DR: On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat 1947; 50–60.

32. Wilcoxon F: Individual comparisons by ranking methods. Biom Bull 1945; 1(6): 80–3.

33. Levene H: Robust tests for equality of variances. Contrib Probab Stat Essays Honor Harold Hotel 1961; 279–92.

34. Bann S, Khan M, Datta V, Darzi A: Surgical skill is predicted by the ability to detect errors. Am J Surg 2005; 189(4): 412–5.

35. Wiegmann DA, ElBardissi AW, Dearani JA, Daly RC, Sundt III TM: Disruptions in surgical flow and their relationship to surgical errors: an exploratory investigation. Surgery 2007; 142(5): 658–65.

36. Leopold SS, Morgan HD, Kadel NJ, Gardner GC, Schaad DC, Wolf FM: Impact of educational intervention on confidence and competence in the performance of a simple surgical task. JBJS 2005; 87(5): 1031–7.

37. Clanton J, Gardner A, Cheung M, Mellert L, Evancho-Chapman M, George RL: The relationship between confidence and competence in the development of surgical skills. J Surg Educ 2014; 71(3): 405–12.

38. Bucholz EM, Sue GR, Yeo H, Roman SA, Bell RH, Sosa JA: Our trainees' confidence: results from a national survey of 4136 US general surgery residents. Arch Surg 2011; 146(8): 907–14.

Figure 1. Example of an augmented 3D annotation visualized from the perspective of a mentee wearing the ARHMD. The annotation appears at the correct position and depth, projected onto the patient's body.

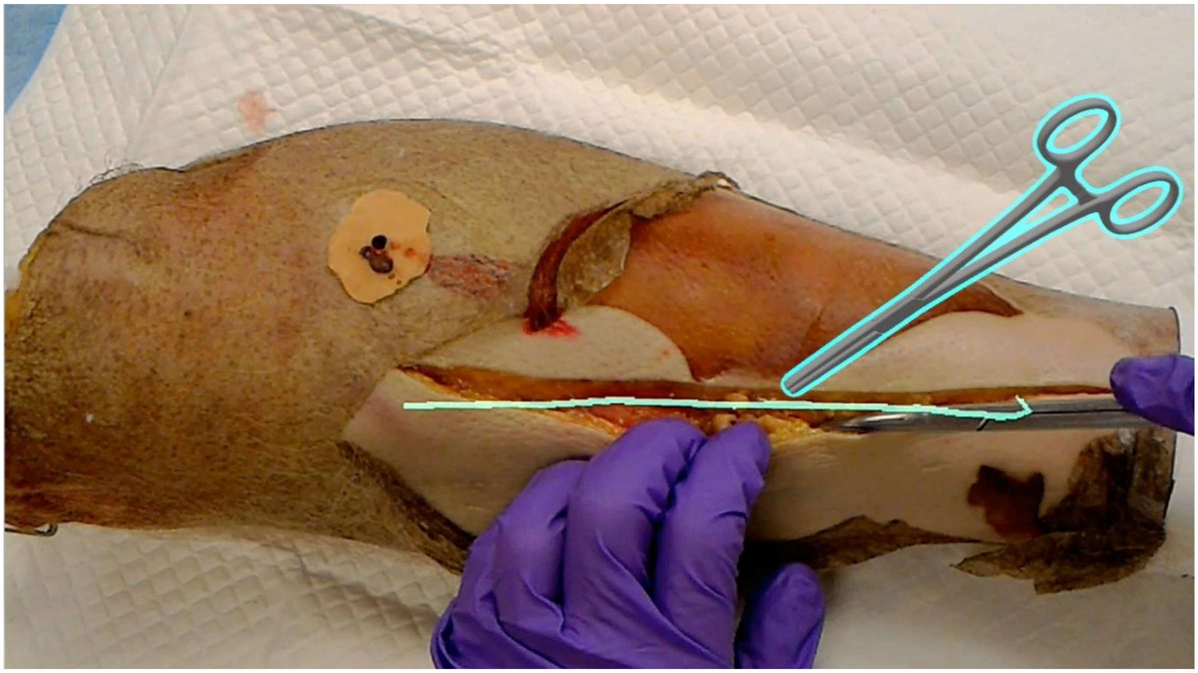

Figure 2. Example use of the STAR ARHMD platform. A remotely located mentor can use the platform to transmit surgical guidance to a local mentee. This guidance will be visualized in 3D by the mentee wearing the ARHMD. A nurse assistant, member of the research team, stood by the mentees during our clinical validation to assist the mentees in aspects such as cleaning fluids

Table 1. Quantitative metrics averaged over participants in each condition. Metrics with * do not satisfy the normality assumption. Only p-values with an obelisk (†) represent a significant difference between the metrics.

Metric	Alone	System for Telementoring with Augmented Reality	p-value
Individual Performance Score	437.4 ± 96.8	483.1 ± 65.3	0.1
Weighted Individual Performance Score	74.6 ± 14.8	82.7 ± 7.6	0.03†
Errors	1.8 ± 1.4	0.6 ± 0.7	0.04†
Global Rating Scale*	60.00 ± 46.67	56.65 ± 29.95	0.3
Evaluator’s Overall Rating*	76.5 ± 27.5	75.0 ± 12.0	0.3
Completion Time*	1444.0 ± 685.0	1379.0 ± 380.5	0.6

Table 2. Participants' performance according to the expertise-based sub-groups.

Metric	Control			System for Telementoring with Augmented Reality		
	Med Students	Residents	Med Students and 1st Year Residents	Med Students	Residents	Med Students and 1st Year Residents
Individual Performance Score	366.5 ± 58.8	467.8 ± 96.5	376.4 ± 46.2	462.8 ± 88.6	491.8 ± 59.0	451.6 ± 69.7
Weighted Individual Performance Score	64.3 ± 10.9	79.00 ± 14.7	54.9 ± 7.7	82.0 ± 10.3	82.9 ± 7.2	80.5 ± 7.9
Errors	2.7 ± 1.2	1.4 ± 1.4	2.6 ± 0.9	0.3 ± 0.6	0.7 ± 0.8	0.2 ± 0.4

Table 3. Difference in participants' self-reported confidence scores. P-values with an asterisk (*) represent that the increases in confidence generated by using the mentoring methods were significantly different.

Confidence Assessment Aspect	Control	System for Telementoring with Augmented Reality	p-value
Identify anatomical landmarks	0.50 ± 1.00	1.00 ± 1.25	0.036*
Knowledge of procedural steps	1.00 ± 2.00	1.00 ± 1.00	0.225
Instrument handling technique	0.00 ± 1.00	1.00 ± 1.25	0.014*
Perform procedure alone	0.05 ± 1.25	1.00 ± 1.00	0.006*