Augmented Reality Future Step Visualization for Robust Surgical Telementoring

Daniel S. Andersen, MS;

Maria E. Cabrera, MSc;

Edgar J. Rojas-Muñoz, BS;

Voicu S. Popescu, PhD;

Glebys T. Gonzalez, BS;

Brian Mullis, MD;

Sherri Marley, RN;

Ben L. Zarzaur, MD;

Juan P. Wachs, PhD

Introduction: Surgical telementoring connects expert mentors with trainees performing urgent care in austere environments. However, such environments impose unreliable network quality, with significant latency and low bandwidth. We have developed an augmented reality telementoring system that includes future step visualization of the medical procedure. Pregenerated video instructions of the procedure are dynamically overlaid onto the trainee's view of the operating field when the network connection with a mentor is unreliable.
Methods: Our future step visualization uses a tablet suspended above the patient's body, through which the trainee views the operating field. Before trainee use, an expert records a future library of step-by-step video footage of the operation. Videos are displayed to the trainee as semitransparent graphical overlays. We conducted a study where participants completed a cricothyroidotomy under telementored guidance. Participants used one of two telementoring conditions: conventional telestrator or our system with future step visualization. During the operation, the connection between trainee and mentor was bandwidth throttled. Recorded metrics were idle time ratio, recall error, and task performance.
Results: Participants in the future step visualization condition had 48% smaller idle time ratio (14.5% vs. 27.9%, $P < 0.001$), 26% less recall error (119 vs. 161, $P = 0.042$), and 10% higher task performance scores (rater 1 = 90.83 vs. 81.88, $P = 0.008$; rater 2 = 88.54 vs. 79.17, $P = 0.042$) than participants in the telestrator condition.
Conclusions: Future step visualization in surgical telementoring is an important fallback mechanism when trainee/mentor network connection is poor, and it is a key step towards semiautonomous and then completely mentor-free medical assistance systems.
(*Sim Healthcare* 00:00–00, 2018)

Key Words: Augmented reality, telementoring, telemedicine, simulator training.

Surgical telementoring provides guidance to a trainee surgeon from a remote expert, enabling the trainee to deliver urgent, specialized care. Surgical telementoring is particularly useful in austere environments where distance limits the availability of necessary expertise. Examples include treating at forward operating base combat trauma injuries, providing immediate specialist care in rural environments, and connecting surgeons in developing countries with mentors from around the globe.[1] Telementoring can also benefit surgical training with simulators, especially when it is costly or difficult for an expert to be physically present and to interact with each of a large number of trainees.[2]

In recent years, telementoring has seen advances because of the use of augmented reality (AR), which can overlay graphical information authored by a remote mentor directly onto a trainee surgeon's view of the operating field. Such approaches have shown the important advantage of reducing the trainee's need to shift focus away from the operating field during a procedure.[3]

However, the network connection itself remains a pressing issue. The austere environments where telementoring is most valuable are also those where wireless Internet connections are often suboptimal. In such conditions, audiovisual transmission can be intermittent and vary greatly in latency and bandwidth during the course of a medical procedure. Mentor instructions may be garbled, delayed, or even completely lost at the moment when a trainee needs guidance. Overcoming this issue requires fallback mechanisms at the trainee site, where the trainee can still have some access to the necessary knowledge when the connection to the mentor is poor.[4]

In this article, we present a novel method for visualization in AR telementoring that allows the trainee to visualize future steps of a surgical procedure independently of the quality of the connection to the mentor (Fig. 1). This is in contrast to conventional AR interfaces, which only provide support for the current step of a procedure. This "future step visualization" illustrates to the trainee what the operating field will appear like after a future step of an operation has been completed, by superimposing prerecorded videos of future steps of the procedure directly onto the trainee's view of the operating field. The videos have been background subtracted and recorded from the trainee's viewpoint, which reduces visual misalignment between real and virtual content (although some minor

FIGURE 1. Prototype of our AR telementoring system, used here with a cricothyroidotomy simulator. The trainee looks at the patient simulator through the AR display that superimposes onto the simulator a visualization of future steps of the procedure, here the initial incision.

misalignment does persist because of anatomical variation between the recordings and the real operating field). The trainee can view the videos when the connection to the mentor is lost or delayed. Previous work has demonstrated the value of standardized step-by-step video clips of procedures for surgical training and mentoring but has not integrated it into either a telementoring or AR context; visual information was still displayed to trainees on nearby monitors and not in context of the trainee's view of the operating field.[5]

We have implemented a prototype that visualizes the future steps of a cricothyroidotomy from stored knowledge, and we have validated it in a user study in which participants performed a cricothyroidotomy on a patient simulator under telementored guidance from a remote expert. Participants interacted with one of two systems: our telementoring system with visualization of future steps or a conventional telestrator. During the procedure, the network connection between trainee and mentor was randomly interrupted to replicate poor network connectivity. The idle time ratio, recall error, and task performance were recorded for participants in each condition and analyzed after the experiment.

METHODS

In this section, we first describe our method for adding future step surgical instruction to a telementoring system by using AR. We then describe the experimental design of a user study to validate our method in the context of a telementoring session under poor networking conditions.

Visualization of Future Steps

We have implemented future step visualization onto an existing telementoring platform called the System for Telementoring with Augmented Reality (STAR).[6] The platform consists of a trainee system, which transmits live imagery of the operating field to a mentor and which receives mentor annotations, and a mentor system, with which the mentor views the operating field remotely and authors annotations that are sent to the trainee. In this article, we focus solely on the details of the trainee system because they relate to the novel feature of future step visualization.

Figure 2 illustrates the addition of future step visualization at the trainee site. A tablet is held in a fixed position above the patient's body such that the trainee can see the screen when looking down at the operating field and can move their hands freely between the tablet and the patient's body. The tablet contains an onboard video camera; during operation, the tablet displays live video frames onto the tablet's screen. In this way, the trainee can view the patient's body and their own hands by looking "through" the tablet like a window and perform each step of the procedure.

During an operation, the trainee system is connected via a wireless network connection to a remote mentor system. The mentor system is a full-size interaction table, which displays the live video feed from the trainee's tablet; the mentor can view the trainee's operating field and draw graphical annotations (such as lines and icons of surgical instruments) using the interaction table's touch-screen interface. The annotations are transmitted back to the trainee tablet system, where they appear superimposed onto the trainee's view of the operating field.

During the operation, the tablet can display contextual visual instruction as AR overlay videos, which visualize future steps of the procedure. A set of videos are prerecorded and stored in a knowledge base before use, in which an expert user performs the same procedure on a patient simulator or on a
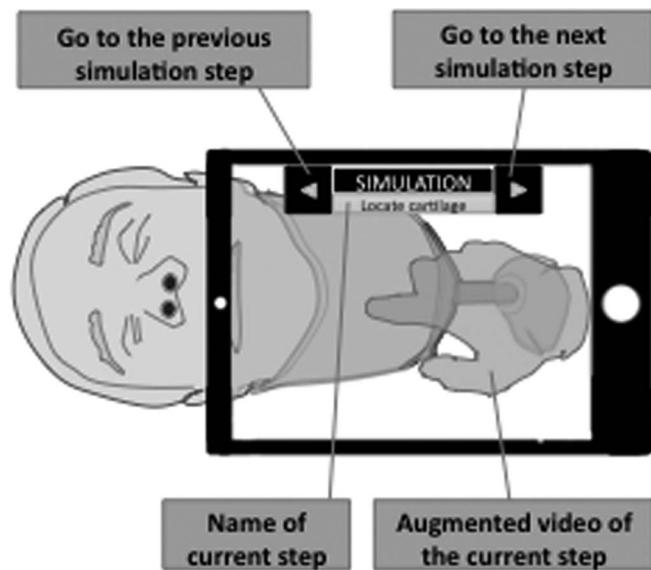
FIGURE 2. Diagram of the on-screen user interface of our AR future step visualization.

cadaver. These videos are recorded from the same relative viewpoint as the trainee's tablet camera. The videos show each stage of the operation, including the position of the expert's hands and any surgical instruments. The videos are segmented into individual video clips representing each step of the procedure, and any background imagery is subtracted using green screening. The videos are either preloaded onto the trainee's tablet before the operation (for standalone use) transmitted in the background from the mentor to the trainee during the telementoring call, so the trainee can access them without needing a live, robust Internet connection. Figure 3 illustrates frames from these videos, in the context of a cricothyroidotomy.

When using the telementoring system, the trainee can select each pregenerated video clip from the knowledge base to be automatically overlaid onto the live video frames of the trainee's operating field. The video clips appear as semitransparent overlays on the tablet screen, so the trainee can see their own hands and surgical instruments and also those of the expert mentor as the expert performs that step of the procedure. Because the videos were captured from the same viewpoint as the trainee's tablet camera, the imagery appears overlaid onto the actual patient's body, as if the virtual expert was interacting with the patient. In this way, the trainee can view how an expert would perform each step of the
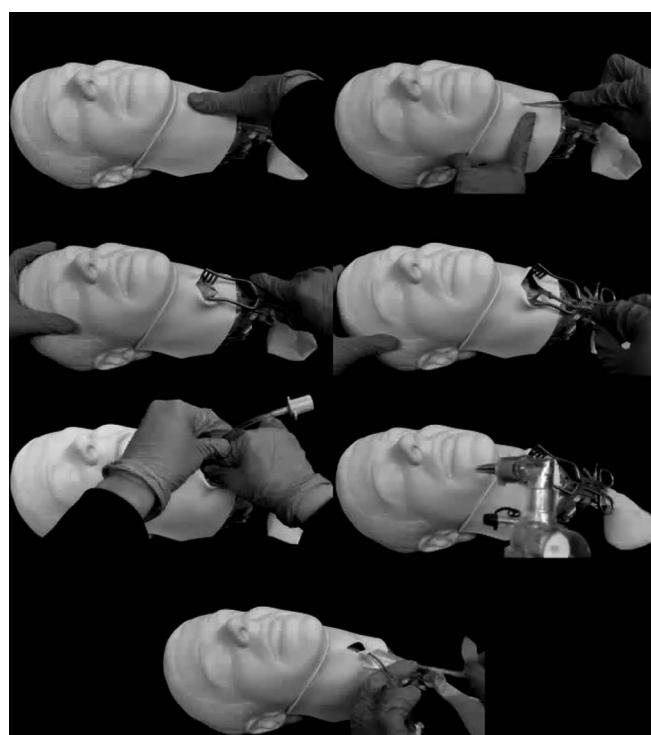
FIGURE 3. Prerecorded videos of an expert performing a simulated cricothyroidotomy used to provide a visualization of the steps of the procedure to the trainee: locating the cricoid cartilage (1), performing the skin incision to expose the cricoid (2), retracting the skin to expose the cricoid (3), performing the cricoid incision (4), inserting the breathing tube in the cricoid incision (5), and connecting the ventilation bag to the tube (6).

3

procedure and can follow along with their own hands and surgical instruments.

The trainee system's user interface allows the trainee to interact with the video instructions. The trainee can toggle the visibility of the visual instructions off (without virtual guidance) or on (to gain additional instruction). On-screen buttons allow the trainee to quickly navigate between each step of the procedure and visualize whichever stage is relevant to the trainee. The trainee can also adjust the overlaid videos' brightness and contrast to make the instructional videos more salient in different lighting conditions. The transparency level of the overlaid video is set to 50% by default, but the trainee can also adjust the transparency level with on-screen controls to see more or less of the instructional overlays. If the overlaid video appears misaligned on the patient's body, the trainee can use standard multitouch controls (pan/rotate/scale) to align the overlay.

User Study

To evaluate the effectiveness of a telementoring system with visualization of future steps in the context of a nonrobust network connection, we conducted a study in which participants performed tasks under telementored guidance with an unstable connection. Participants were tasked with completing a cricothyroidotomy on a patient simulator under one of two conditions: *conventional* telestrator-based telementoring (Telestrator) or telementoring using a visualization of future steps (*STAR*). The goal of the experiment was to compare the trainee's idle time ratio, recall error, and surgical performance, under the two conditions.

Twenty participants (10 female, 10 male) without previous medical training were recruited from Purdue University to act as trainees in our user study. Table 1 lists baseline characteristics of the participants. Participants had a mean age of 25.5 years, (SD = 2.2 years). The study was reviewed and approved by the Purdue University Institutional Review Board; written participant consent was acquired for each participant before the study.

TABLE 1. Baseline Characteristics of Participants in Each Condition

	STAR Condition	Telestrator Condition	Total
No. participants	10	10	20
Age, categorical (units: participants)			
≤18 yr	0	0	0
Between 18 and 65 yr	10	10	20
≥65 yr	0	0	0
Age, continuous, mean ± SD, yr	25.2 ± 2.04	25.8 ± 2.57	25.5 ± 2.28
Sex (units: participants)			
Female	4	6	10
Male	6	4	10
Region of enrollment (units: participants)			
United States	4	2	6
India	4	1	5
China	1	3	4
Latin America	1	2	3
Iran	0	1	1
Thailand	0	1	1

For both conditions (telestrator and STAR), the participant was located in a room with a patient cricothyroidotomy simulator on a table. In another building, a member of the research team who was trained in mentoring a cricothyroidotomy was connected with the trainee by using the mentor side of our telementoring system. An audio connection was set up so that the trainee participant and the remote mentor could speak with each other. In the *conventional* telestrator condition, participants received visual mentor instruction via a nearby monitor in front of them, which displayed imagery of the operating field overlaid with visual mentor instructions. In the *STAR* condition, participants had a tablet placed above the operating field, which displayed the same imagery of the operating field with mentor instructions.

Figure 4 illustrates the system architecture during the experiment. Participants were tasked with completing a cricothyroidotomy under remote telementored guidance. Participants were not given training in the steps of the procedure beforehand; only training on how to use the surgical instruments was given. To simulate intermittent network issues, bandwidth limiting software was introduced into the mentor/trainee network connection. At pseudo-random intervals, bandwidth was automatically limited, resulting in drops of audiovisual quality between the mentor and the trainee. Communications were interrupted using a normal random distribution around two means, where the connection switched between unimpaired (mean of 40 seconds) and impaired (mean of 15 seconds). The restricted connection made it almost impossible for the trainee to understand what the mentor was saying; visual information between mentor and trainee was similarly distorted. Participants were informed beforehand that network interruptions were a possibility but were not informed of the specific metrics being measured.

Telestrator Condition

Figure 5 shows the telestrator condition of the user study. In this condition, a screen was positioned in front of the trainee, which showed visual guidance from the remote mentor to the trainee. The guidance consisted of a set of annotated images, which the trainee could view via screen sharing. These images illustrated each step of the procedure and were augmented with animated line annotations.

STAR Condition

Figure 6 shows the STAR condition of the user study. In this condition, participants viewed the patient simulator by looking through a tablet preloaded with our future step visualization of a cricothyroidotomy procedure. The tablet system superimposed annotations (lines and icons of surgical instruments) that were drawn and transmitted in real time by the mentor. At any time, the trainee could switch to the future step instruction provided by the tablet.

Experimental Design and Data Collection

Each participant was video recorded while completing the procedure. After completion, participants also answered a questionnaire in which they described each of the steps of the procedure they had completed.

The number of times a participant in the STAR condition used the future step visualization feature was recorded. A team member observed the video recordings of the participant and
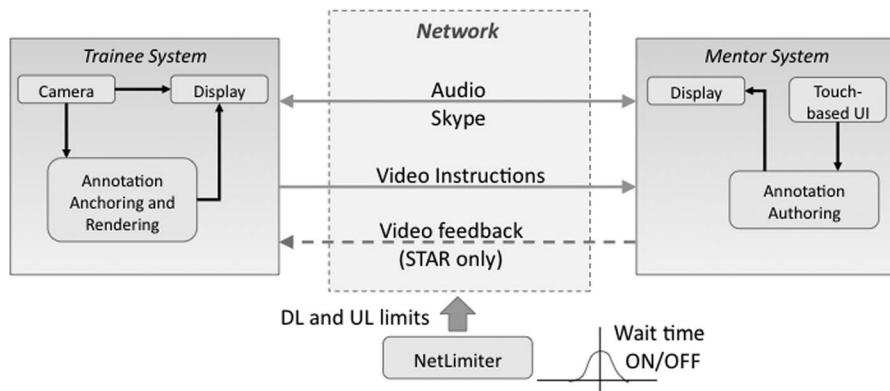
FIGURE 4. System architecture during the user study.

counted each time a participant used the tablet's touch screen either to start an animation or to change which step was currently visualized. In the event of misalignment between the video overlay and the real operating field, any user repositioning of the animation on screen was not counted as an interaction.

Participants were randomly assigned to one of the two telementoring methods. The independent variable was the telementoring method (*conventional* telestrator or *STAR* using visualization of future steps). The dependent variables were idle time ratio, recall error, and performance score.

Idle Time Ratio

We define idle time ratio as the ratio between the total time trainees remained idle (not doing any action) and the total time taken to complete the procedure. This metric indicates how much time was wasted because of faulty communications experienced by participants.

Recall Error

Recall error describes the error of participants when describing from memory each step of the procedure. Before the experiment, a ground truth text of each step of the procedure was defined. Each text description was a collection of important words relevant to that step (e.g., "open retractor," "connect bag pump check"). After completing the procedure, each participant wrote a description of each step of the procedure. In subsequent analysis, each description was interpreted as a vector of words and the Levenshtein vector distance between what each participant wrote and the ground truth was calculated.[7] This method, known as "bag-of-words similarity," determines the similarity between texts by quantifying the textual changes that must be performed to transform a given text into a target text.[8] In this vector space, lower distance means that participants' descriptions included more target keyboards for each step, indicating that participants could recall more of the relevant information than participants having a higher distance. The distance was taken to be that participant's recall error.

FIGURE 5. Experiment setup for the telestrator telementoring condition.

FIGURE 6. Experiment setup for the future step visualization telementoring condition.

Performance Score

This metric represents how well the participant performed each step of the procedure. Two team members, trained in performing cricothyroidotomy procedures, independently assessed each participant's performance as captured on video, and the Cronbach α measurement of internal consistency was recorded between the two raters.[9] Because video footage of the operating field shows the presence or absence of the tablet in the STAR condition, raters were unavoidably not blinded with respect to participant group. The assessment followed the United States Marine Corps Emergency Cricothyroidotomy Steps (FMST 1418), assigning a score (0–3, 0 being the lowest) depending on how well the instruction was performed.[10]

Statistical Analysis

A statistical analysis was conducted to compare two sampled populations: participants using the STAR condition and participants using the telestrator condition. The normality assumption of the data was confirmed using a Shapiro-Wilk test.[11] After this, a Levene test was run, which determined that variances between populations were not significantly different, and a t test using pooled variance was used to find statistically significant differences between the two sampled populations.[12] Continuous data were summarized as mean (micro) ± standard deviation (sigma$_x$). Outliers were detected based on distance from the first or third quartile exceeding 1.5 times the interquartile range.[13] For the recall error metric, one outlier was removed from the telestrator condition, giving nine for the telestrator condition and ten for the STAR condition. No outliers were removed for the rest of the metrics, giving ten per condition.

RESULTS

Figure 7 illustrates the results from our user study. On average, the number of interactions between a participant in the STAR condition and the future step visualization feature was 5.4 ± 4.62. The idle time ratio of participants using STAR was 48% less than those using the telestrator (14.5% ± 6.0% vs. 27.9% ± 6.4%, $P < 0.001$). Participants using STAR showed 26% less recall error than those using the telestrator (119 ± 35 vs. 161 ± 34, $P = 0.042$). The performance score for participants using STAR was approximately 10% higher than that achieved by participants using the telestrator (rater 1 = 90.83 ± 4.95 vs. 81.88 ± 7.57, $P = 0.008$; rater 2 = 88.54 ± 7.97 vs. 79.17 ± 10.08, $P = 0.042$; Cronbach α between raters = 0.8947). A post hoc power analysis was done (with α = 0.05), indicating a statistical power of 0.998 for the idle time ratio metric and 0.649 for the recall error metric. For the performance score metric, post hoc statistical power was

calculated independently for the results of the two raters (rater 1 = 0.879, rater 2 = 0.635).

DISCUSSION

The results indicate that surgical telementoring with future step instruction excelled when compared with conventional telestrator-based telementoring. Participants using future step visualization completed the operation with proportionally less idle time, with less recall error, and with improved task performance, compared with participants using conventional telementoring. A visualization of future steps is a valuable fallback solution to the question of interrupted connection during telementoring. Trainees using future step visualization are less idle or delayed by network interruptions than trainees depending on a robust connection with a remote mentor. Because delays in critical surgical care can greatly affect patient outcomes, our future step visualization could help save lives.

The results for recall error and performance score show the value of video clips to visualize future steps. Videos provide a good indication of the steps because they were acquired from a viewpoint similar to the trainee's and were background subtracted to only illustrate relevant details. Videos are rich content that show more than a mentor could sketch through a telestrator. We foresee that future telementoring systems could benefit from introducing a library of pregenerated instructions even during synchronous telementoring, where a mentor could initiate playback of videos for illustration purposes or to facilitate recall during training. Because the trainee can go step-by-step through the visualization, the steps themselves become more memorable, as indicated in the recall error metrics. The number of times STAR participants used the future step visualization feature (5.4 times on average) was high, considering that the operation steps were illustrated using a total of seven videos. Typically, participants would begin using future step visualization after the first time the network connection was interrupted and would continue to use it for the rest of the session.

Several limitations remain in our current implementation. First, in our user interface, the trainee presses buttons on the tablet screen to move between visualizations of each future step. Although not an issue for the patient simulator used in our study, such interactions raise questions of sterility during an actual operation. Future work will allow the trainee to progress through each step using voice commands or, alternatively, a fully autonomous prediction system.

Because the visualization's videos in the library must be pregenerated, the proportions of the patient simulator and the
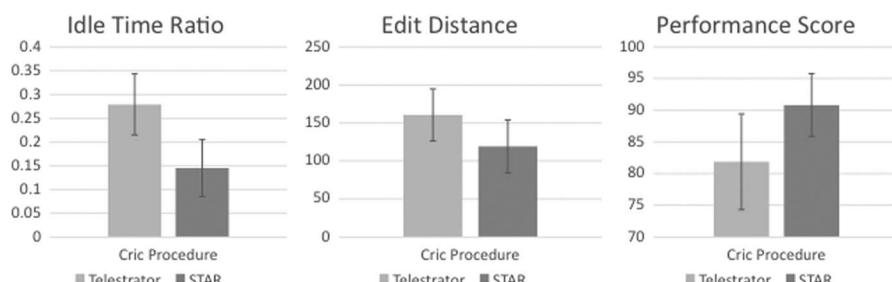

FIGURE 7. Results from the future step visualization user study.

actual patient's body will be necessarily misaligned. It should be noted, however, that in the case of simulator training, alignment is expected to be good because video footage can identically match a patient simulator. Future work will investigate future step visualization on an actual patient, where the task of adapting stock footage is more challenging because of patient variability and would require automatically morphing and warping the captured mentor video imagery to fit different anatomy.

Beyond this potential variability, some mismatch is unavoidable between the view of the operating field as seen by the tablet camera, and the view as would be seen directly by the trainee, because the tablet camera is not co-located with the trainee's eyes. The result is that the position and scale of the patient may appear incorrect, which could impact trainee hand-eye coordination. This is known as the "dual-view problem" in AR.[14] Recent research into user-perspective rendering and simulated transparent displays can help overcome this by tracking the three-dimensional geometry of the operating field and the position of the trainee's eyes to reproject imagery to the correct perspective.[15,16]

Conventional tablets possess some camera latency (approximately 100 milliseconds), so the trainee will experience a minor delay between performing an action on the patient and seeing it on the tablet screen. Although we have not investigated the effect of local video latency in our AR system, previous research has indicated that video latency can adversely impact surgical performance during laparoscopic surgery.[17,18] However, because all videos are displayed and rendered locally, the latency is less than in telesurgery, where a surgeon is operating remotely. In addition, local video latency is far lower than what the trainee's idle time would be if awaiting instruction from a remote mentor without future step visualization. Emerging advances in see-through AR displays such as the Microsoft HoloLens would sidestep this latency.

Our system allows a user to adjust the transparency level of the future step overlays by accessing a menu in the user interface. However, for the sake of simplicity, this ability was not indicated to study participants and the default transparency value of 50% remained unchanged. Therefore, it would be interesting to formally evaluate different transparency levels to determine an ideal default that balances visibility of the overlaid videos with the visibility of the background operating field. Such an evaluation should also determine whether participants find value in an ability to adjust transparency on-the-fly or whether such fine level of control would be a distraction.

We plan to investigate the potential of future step visualization for patient simulator training outside of telementoring. Local availability of instructions can also benefit completely mentorless scenarios such as medical training with simulators. Semiautonomous training systems could enable independent practice (alone or in groups) without needing direct continuous access to experts who may be in short supply. Simulator-only future step visualization is also particularly promising because the structure of a patient simulator is known in advance, so videos can always be well aligned and matched. We anticipate that future step visualization can become part of adaptive, responsive, and automated systems for trauma care and for simulator training.

We also plan to perform additional validations of future step visualization features in user studies that quantify trainee performance to situations in which no network interruptions are present. Our current investigations assumed a setting with constant network interruptions in both telestrator and STAR conditions and determined that STAR improved participant performance in such a setting. However, it would also be useful to compare against the "gold standard" in which participants had no network interruptions at all, and to determine how completely future step visualization can overcome the problem of network interruptions.

In conclusion, we have developed a method of future step visualization suitable for trainees in telementoring contexts with austere network conditions. Our system provides mentor guidance to a trainee user in context on the operating field. We conducted a user study that demonstrated that telementoring systems are more effective if trainees can keep following surgical instructions when network conditions are austere or even when a mentor is not available. Further studies should be conducted to explore the potential of future step visualization in a wider range of austere environments and in surgical training with simulators.

REFERENCES

1. 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–S163.

2. Augestad K, Han H, Paige J, et al. Educational implications for surgical telementoring: a current review with recommendations for future practice, policy, and research. *Surg Endosc* 2017;31:1–11.

3. Wachs JP, Gomez G. Telementoring systems in the operating room: a new approach in medical training. *Medicina (B Aires)* 2013;73(6):539–542.

4. St Julien J, Perrier ND. Video telementoring to accelerate learning of new surgical techniques. *JAMA Surg* 2016;151(7):671–672.

5. Lee JH, Tanaka E, Woo Y, et al. Advanced real-time multi-display educational system (ARMES): an innovative real-time audiovisual mentoring tool for complex robotic surgery. *J Surg Oncol* 2017; 116(7):894–897.

6. Andersen D, Popescu V, Cabrera ME, et al. An augmented reality-based approach for surgical telementoring in austere environments. *Mil Med* 2017;182(S1):310–315.

7. Levenshtein VI. Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady* 1966;10(8):707–710.

8. Klinkmüller C, Weber I, Mendling J, et al. Increasing recall of process model matching by improved activity label matching. *In Business Process Management.* Berlin: 2013:211–218.

9. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16(3).

10. United States Marine Corps Training & Education Command. FMST 1418 Perform Emergency Cricothyroidotomy. Available at: http://www.tecom. marines.mil/Portals/120/Docs/Student%20Materials/FMST%20Manual/ CricothyroidotomyFMST1418.doc. Accessed October 1, 2017.

11. Shapiro S. An analysis of variance test for normality (complete samples). *Biometrika* 1965;52(3):591.

12. Brown MB, Forsythe AB. Robust tests for the equality of variances. *J Am Stat Assoc* 1974;69(346):364–367.

13. Rousseeuw PJ, Hubert M. Robust statistics for outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2011: 73–79.

14. Čopič Pucihar K, Coulton P, Alexander J. Evaluating dual-view perceptual issues in handheld augmented reality: device vs. user perspective rendering. *Proceedings of the 15th ACM on International Conference on Multimodal Interaction* 2013: 381–388.

15. Unuma Y, Komuro T. 3D interaction with virtual objects in a precisely-aligned view using a see-through mobile AR system. *ITE Transactions on Media Technology and Applications* 2017;5(2): 49–56.

16. Andersen D, Popescu V, Lin C, et al. *A Hand-Held, Self-Contained Simulated Transparent Display*. Merida, Mexico: IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct); 2016:96–101.

17. Kumcu A, Vermeulen L, Elprama SA, et al. Effect of video lag on laparoscopic surgery: correlation between performance and usability at low latencies. *Int J Med Robot Comput Assisted Surg* 2016.

18. Kim T, Zimmerman PM, Wade MJ, et al. The effect of delayed visual feedback on telerobotic surgery. *Surg Endosc* 2005;19:683–686.