Noname manuscript No.
(will be inserted by the editor)

Virtual Annotations of the Surgical Field through an Augmented Reality Transparent Display

Daniel Andersen · Voicu Popescu · Maria Eugenia Cabrera · Aditya Shanghavi · Gerardo Gomez · Sherri Marley · Brian Mullis · Juan Wachs

Received: date / Accepted: date

Abstract Existing telestrator-based surgical telementoring systems require a trainee surgeon to shift focus frequently between the operating field and a nearby monitor to acquire and apply instructions from a remote mentor. We present a novel approach to surgical telementoring where annotations are superimposed directly onto the surgical field using an augmented reality (AR) simulated transparent display. We present our first steps towards realizing this vision, using two networked conventional tablets to allow a mentor to remotely annotate the operating field as seen by a trainee. Annotations are anchored to the surgical field as the trainee tablet moves and as the surgical field deforms or becomes occluded. The system is built exclusively from compact commodity-level components – all imaging and processing is performed on the two tablets.

Keywords Augmented reality · telementoring · telemedicine · annotation anchoring

1 Introduction

Telementoring has the potential to abstract away the geographic distance between a patient in need of expert surgical care and the surgeon with the required expertise. Consider the scenario of a patient urgently needing a complex procedure for which a rural hospital does not have a specialist. Telementoring could enable the rural surgeon to perform the procedure under the guidance of a remote expert, without the delays associated with transporting the patient to a major surgical center. Consider a second scenario where a surgeon is deployed to an overseas forward operating base with limited resources. The military surgeon could provide urgent specialized surgical care with the help of an expert surgeon located thousands of miles away. Further, consider the case of a life-saving innovative surgical technique not widely adopted yet. A surgeon could disseminate the novel procedure through telementoring. Finally, telementoring principles can be translated to telerobotics in the future where a single surgeon can participate in multiple surgeries simultaneously.

Current systems fall short of realizing the potential of surgical telementoring. In current systems, the remote mentor annotates a video feed of the surgery using a telestrator. The annotated video is sent back to the operating room where it is displayed on a nearby monitor. The trainee has to shift focus frequently between the operating field and the nearby monitor to acquire and apply the instructions from the mentor. The trainee first has to parse and understand the instructions on the monitor, then the trainee has to memorize the instructions, and, finally, after shifting focus back to the surgery, the trainee has to temporally and spatially project those instructions into the real-world context of the surgery. This indirect approach to acquiring and applying mentor instructions translates to a significant additional cognitive load for the trainee and interferes with natural hand-eye coordination, which can lead to surgery delays or even errors. Another shortcoming of current systems is that annotations are static and they can become disassociated from the surgical field elements for which they were defined. For example, an

D. Andersen (✉) · V. Popescu · M.E. Cabrera · A. Shanghavi · J. Wachs
Department of Computer Science,
Purdue University, West Lafayette, Indiana, USA
E-mail: andersed@purdue.edu

G. Gomez · S. Marley · B. Mullis
School of Medicine,
Indiana University, Indianapolis, Indiana, USA

Fig. 1 Concept illustration of AR transparent display telementoring approach: overall view of system at trainee surgeon site (left), and trainee view (right).

incision line drawn by the mentor can move away from its intended location as the surgical field changes.

We present a novel approach to surgical telementoring where annotations are superimposed directly onto the surgical field using an augmented reality (AR) transparent display. Fig. 1, left, gives the overall view of the system at the trainee site. The trainee surgeon sees the annotated surgical field through a display suspended into their field of view. The display is transparent except for the pixels where it displays the annotations created by the mentor. In Fig. 1, right, the annotation indicates the precise placement of an additional surgical clamp. The transparent display allows the trainee to see their hands, the surgical instruments, and the surgical field. The part of the surgical field seen by the trainee through the display is aligned with the surrounding region of the surgical field that the trainee sees directly. The annotations remain anchored to the surgical field elements for which they were defined as the display is repositioned, as the trainee head position changes, and as the surgical field changes over time.

The AR transparent display approach has the potential to bypass the shortcomings of the conventional telestrator-based approach. The transparent display integrates annotations into the surgical field, so the trainee can benefit from the annotations without shifting focus. The alignment between the annotated and the peripheral regions of the surgical field preserves the natural hand-eye coordination on which surgeons rely. The annotations are anchored to the surgical field and remain valid as the viewpoint and surgical field change. This reduces the need for the mentor to redraw annotations that have drifted out of place, improving the continuity of the visual guidance provided to the trainee.

In this paper we present our first steps towards realizing this vision. The transparent display at the trainee site (Fig. 2) is simulated using a conventional tablet that displays the video stream acquired by its back-facing camera. The video stream is sent wirelessly to the mentor site where it is displayed on the mentor's tablet. Using the tablet's touch-based user interface, the mentor adds graphical and textual annotations to a frame of the video stream with (Fig. 3). The annotations are sent back to the trainee site where they are overlaid on the tablet to provide guidance to the trainee.

Annotations are anchored to the surgical field as the trainee tablet moves and the surgical field deforms or becomes occluded. The anchoring algorithm positions, scales, and orients the annotations in each frame by computing a homography between the current frame and the initial frame where the annotations were defined. This is done by detecting features in the current frame and by matching them to features in the initial frame. The system is built exclusively from compact commodity-level components; all imaging and processing is performed on the two tablets. The annotation anchoring performance is on average 12fps.

We tested our system in three experiments. The first two experiments were conducted in our laboratory, once by approximating the patient using a flat anatomical poster and once using a surgical dummy. The third experiment was conducted in a teaching hospital where a faculty of trauma surgery instructed a surgery resident in conducting a cricothyrotomy on a porcine model. In the current system the video stream is displayed on the trainee tablet without adapting it to the trainee's viewpoint, so it only provides an approximate transparency effect. Annotation anchoring is robust to repositioning and occlusions but not to surgical field deformations.

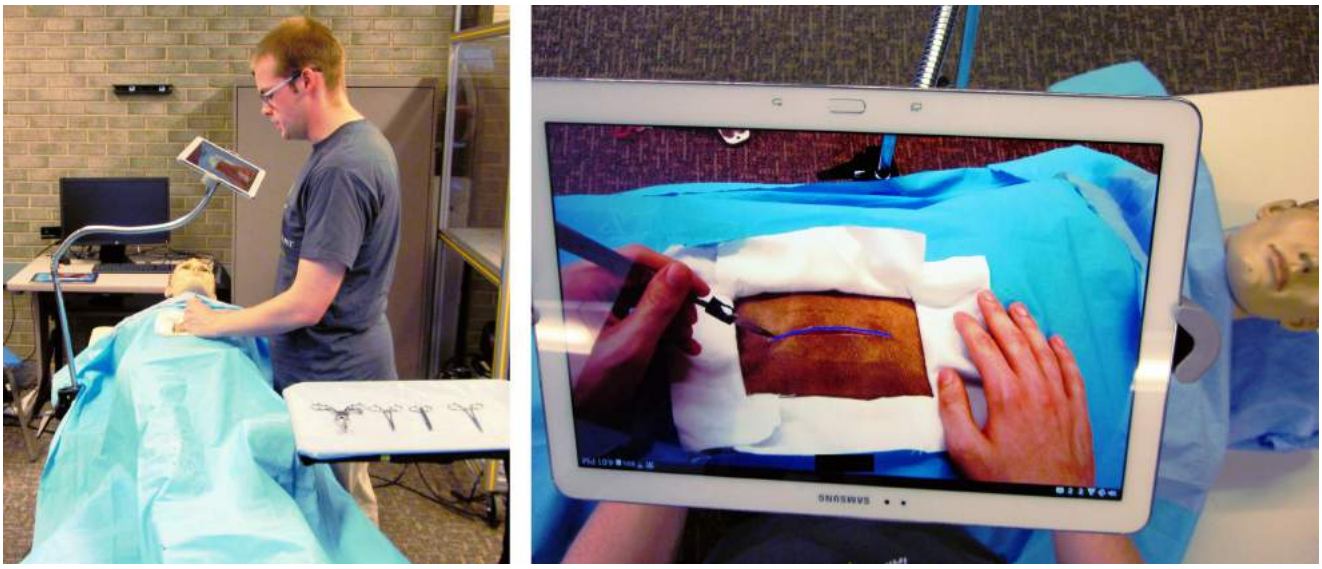

Fig. 2 Trainee system in our first implementation of the AR transparent display telementoring approach: overall view (left) and trainee view (right). The trainee surgeon sees the surgical field through the transparent display and performs an incision along the line suggested by the mentor.

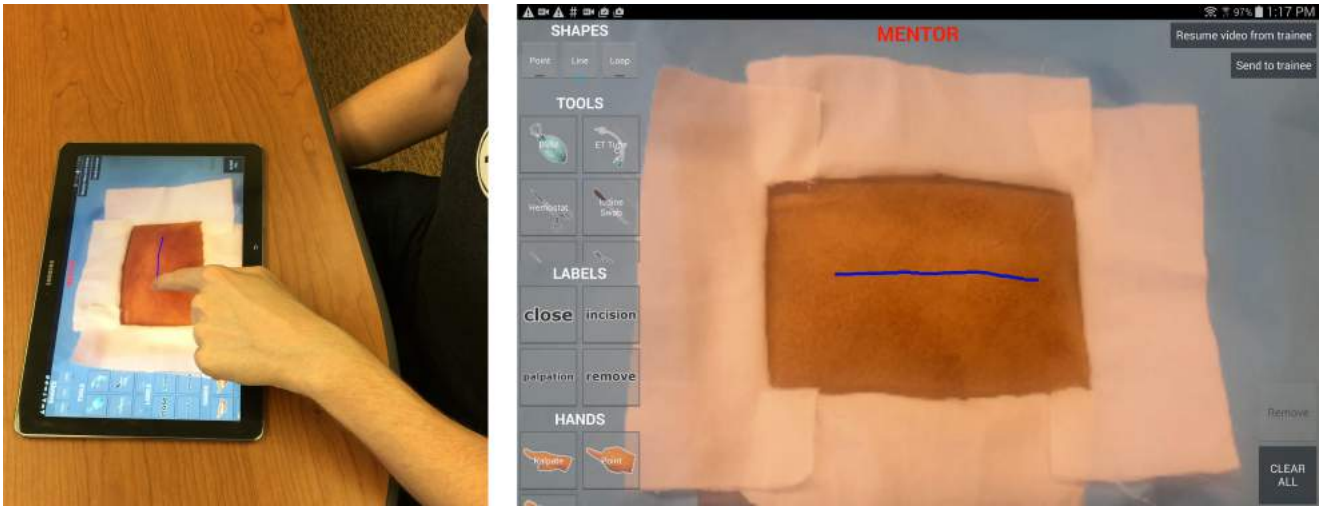

Fig. 3 Mentor system: overall view (left) and mentor touch-based user interface (right). The mentor suggests an incision line on the video stream received from trainee system.

2 Prior Work

Advances in telecommunications have impacted the medical field in the form of telemedicine, a new branch of medicine that focuses on the use of telecommunications technology to exchange medical information and provide medical services from a remote location [1,5]. One particular area of telemedicine – telementoring – allows for an experienced surgeon to provide relevant and immediate guidance. The potential of telementoring is that it promises to create a "virtual classroom" in which surgeons competent in general surgical techniques can gain additional, more sub-specialized experience from an expert surgeon, without needing to be physically co-located with the mentor [16,3].

There is a need for additional research on the effectiveness of telementoring in open surgery. Ereso et al. demonstrated the feasibility of telementoring for surgical consultation, by providing a mentor surgeon with a remote view of the operating field via a manipulable camera, and also providing the ability for the mentor to virtually gesture regions using a remote-controlled laser pointer. Performance of trainee surgeons benefited from the remote presence of a mentor when compared to unproctored performance. However, this study only compared unproctored experience to telementoring and did

not consider how effective telementoring is when compared to in-person mentoring [8].

Guo et al. integrated a commercial videoconferencing system in order to remotely mentor surgeons in laparoscopic surgery. This approach follows a more traditional form of telementoring [11]. Treter et al. also used video-conferencing for telementoring, this time using a multi-institutional effort focused on adrenalectomy procedures [26]. Limitations of these approaches involve inherent issues with traditional static telestrator-based approaches: as the mentor surgeon only draws static annotations, drawn lines do not remain rigidly anchored to the surgical field after movement, deformation, or structural change in the surgical field [6].

Telemedicine and telementoring applications rely on effective communication of medical expertise. AR can enhance telementoring, either as an interface or an environment [21]. In the first case, a virtualized interface can allow for more intuitive interaction between a surgeon and relevant medical information. For example, in laparoscopic surgery, the operating surgeon and the telementoring surgeon can share the same real-time laparoscopic video [8], so this video is displayed to the telementoring surgeon in conjunction with a view of the operating room [22]. Additional viewpoints can give greater visual context to trainee and mentor. In the second case, enhancing the trainee's perceived environment with imagery provided by a remote mentor can enhance the feeling of mentor-trainee co-presence.

Chou et al. proposed and successfully demonstrated the use of AR in preoperative planning for remote robot-assisted neurosurgery. This approach used physical markers placed on a patient's body, detected with a stereo camera, to calibrate the relative position of a robotic system to improve the safety of a remote-controlled surgical operation [7]. Shenai et al. created an AR surgical field, in which a remote mentor could make physical gestures with hands or surgical instruments, which would be overlaid onto the trainee's field of view. The virtual surgical field would also be augmented with relevant information, such as MRI volumetric renderings of the patient [23]. One issue with this approach is that the trainee must view the augmented surgical field with a binocular videoscope, which can be encumbering, bulky, or restrict the trainee's natural motion.

Vera et al. proposed and applied the use of AR in laparoscopic surgical training, using Chroma key technology to overlay live video of a mentor acting out a suturing task [28]. Ponce et al. successfully used the Google Glass wearable display to provide mentor guidance during the performance of a shoulder replacement. One reported issue was the divergent field of views between the Google Glass' on-board camera and the trainee surgeon's vision. In addition, the Google Glass display has low resolution and a very low field of view; mentor guidance appears on a small screen in the corner of the trainee's vision, not overlaid over the trainee's view of the surgical area [18]. Ponce et al. also developed a virtual interactive presence where the mentor surgeon's hands and other surgical tools are merged directly with the arthroscopic image and displayed on a sophisticated telestrator, which also allowed making annotations using a special pen-tool [18]. More recently, a new tool for surgical telementoring through haptic holograms, annotations and multi-model streaming has been suggested, though there is no evidence of such a tool evaluated in real surgeries at this point [24].

The fundamental challenge in using AR in surgical environments is integrating synthetic overlays seamlessly within a real-world scene. Many existing systems depend on the trainee surgeon looking at a screen that does not align with the trainee's view of the scene outside the screen. Systems that use AR head-mounted displays can interfere with the vision or the trainee's head motion and cause ocular fatigue. In addition, it is important for an augmented surgical field to avoid obscuring important real-world detail, while ensuring that the information provided by synthetic visuals is readily accessible to the trainee [15].

Loescher et al. described and developed a system that uses a tablet screen, held by a robotic arm between the trainee surgeon and the operating field, to overlay augmented annotations on the surgical scene. This approach surveyed a series of feature tracking and descriptor matching computer vision algorithms and compared their anchoring accuracy and performance. Limitations of the system include its reliance on processing video frames remotely and low processing frame rate [14].

Research in augmented reality is attempting to leverage the computational power and the compact form of tablets to simulate transparent displays. Tomioka et al. simulated a transparent display with a tablet, a camera for tracking the user, and a nearby workstation for warping the images acquired by the tablet to achieve continuity with the surrounding scene [25]. The warping is computed based on the assumption of a planar scene. Baričević et al. removed the planar scene assumption by acquiring depth passively, using stereo matching between the frames of two video cameras. The advantage of passive stereo acquisition is robustness with strong environment illumination, such as in the case of outdoor scenes. The classic disadvantage is the difficulty in establishing correspondences for scenes with little color variation [4]. Unuma et al. created a system that relies on active depth acquisition, which improves density, rate, and robustness. Compared to our work, these

systems have the advantage of attempting to create a better transparency illusion by reprojecting the tablet frames to the user viewpoint, which we will attempt in future work [27]. Our work has the advantage of processing the frames exclusively on the tablet, without the need of a nearby workstation. This is a crucial advantage for austere environments, which are targeted by our project. Furthermore, we are developing our system specifically for surgery telementoring, a demanding application of AR transparent displays, leveraging formative feedback from surgeons from day one.

3 System Overview

Fig. 4 gives an overview of our prototype system. The trainee system is implemented with a tablet whose camera acquires a video stream of the surgical field. Each frame is displayed (1), wirelessly sent to the remote mentor system (2), and processed for annotation anchoring that begins with Feature Detection (3).

The mentor system receives the current frame via a wireless network (4), the frame is displayed (5), and it is provided as an input to the Annotation Authoring module (6). Annotation authoring (Fig. 5 - 7) is described in Section 4. The mentor chooses a reference frame on which to define annotations using the Touch-Based UI (7). Fig. 8, left shows an incision line annotated by the mentor onto the reference frame. The annotations are displayed (8) and the reference frame is processed to prepare annotation anchoring. The first step is to detect salient features in the reference frame in the neighborhood of the annotations through Feature Detection (9) (see Fig. 9, left), and then to compute unique signatures for each feature through Descriptor Extraction (10) (see Fig. 9, right). Feature detection and descriptor extraction are described in detail in Section 5. Annotations, reference frame features, and associated descriptors are sent to the trainee system (11).

The trainee system receives the annotations and the reference frame data (12), and begins the process of anchoring the annotations to the current frame (3). Annotation anchoring is described in Section 5. The current frame's features are detected (13) and enhanced with descriptors (14) (Fig. 10, left). The current frame's descriptors (14) are matched with the reference frame's descriptors (15) where the annotations were defined (Fig. 10, right). The matched descriptors (16) are used to derive a homography for each annotation (Fig. 11, left). The homographies (17) transform the annotations from the reference frame (18) to the current frame (Fig. 11, right). The transformed annotations are rendered and overlaid onto the current frame (19), and appear anchored to the surgical field.

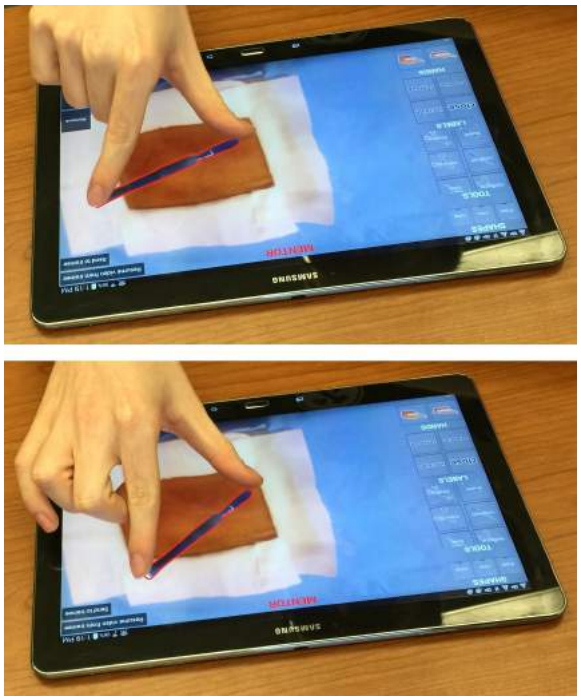

Fig. 5 Tool orientation using two-touch interaction.

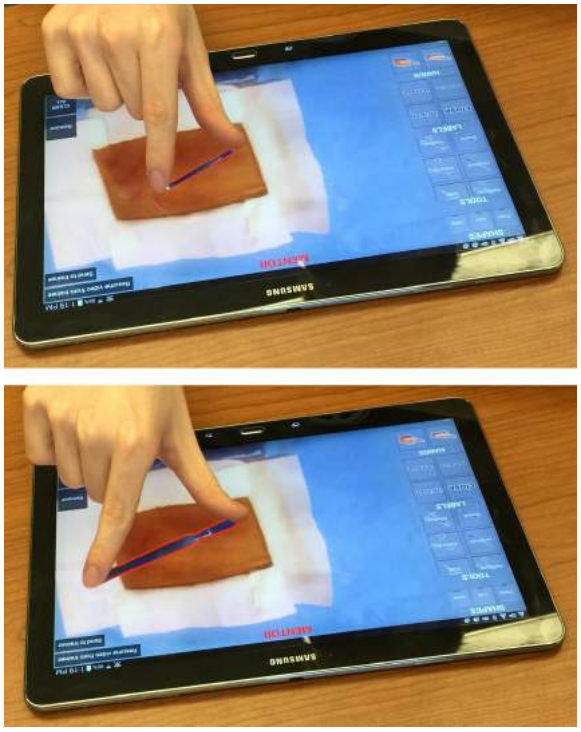

Fig. 6 Tool scaling using two-touch interaction.

4 Annotation Authoring

The mentor creates, positions, orients, and sizes annotations via the tablet's multi-touch user interface.

Annotations are created by tapping icon-labeled buttons. There are four annotation categories: drawing shapes,

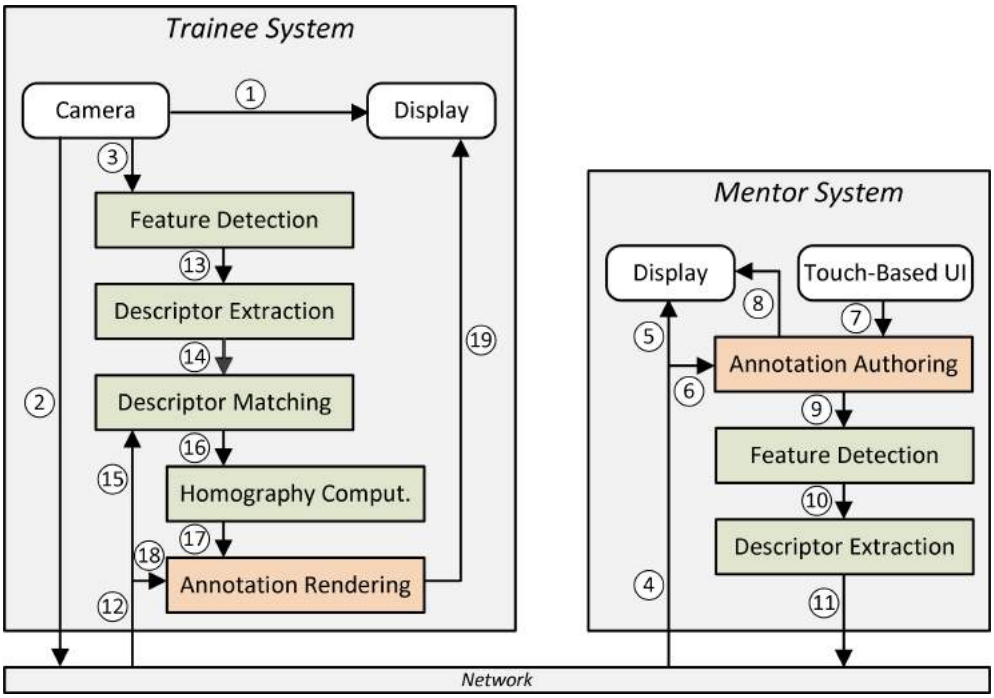

Fig. 4 Architecture of our first implementation of the proposed AR transparent display approach to surgery telementoring. The computer vision/computer graphics processing stages are highlighted in green/orange.

surgical tools, text labels, and hand gesture icons (Fig. 7). The types of drawing shapes are: points, lines, polygons. Each shape is defined with one or multiple points. The surgical tools include BVM, ET tube, hemostat, iodine swab, longhook, retractor, scalpel, scissors, stethoscope, surgical tape, syringe, and tweezers. The predefined text labels include "close," "incision," "palpation," "remove," and "stitch." The hand gesture annotations illustrate typical manual actions performed by the surgeon such as palpating, pointing, and stretching. Surgical tools, text labels, and hand gesture icons are positioned based on a reference point (e.g. the tip of the scalpel's blade); they are represented as an image with transparent background.

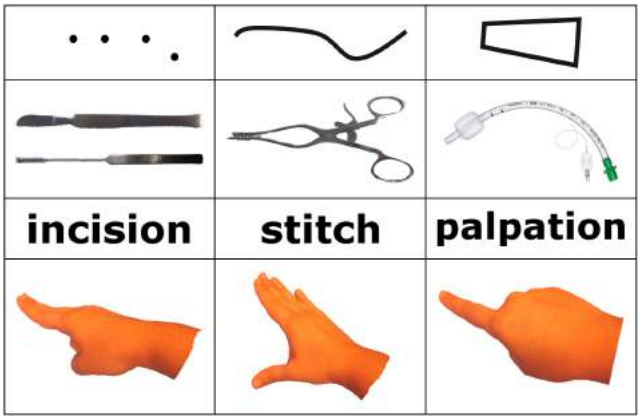

Fig. 7 Annotation examples: drawings, surgical tool icons, text labels, and hand gesture icons.

The annotations are positioned using single-touch drag and drop interaction. They are orientated using two-touch interaction: one touch for defining the center of rotation and one dragging motion for defining the rotation angle (Fig. 5). Scaling is done using two finger pinch-and-zoom interaction (Fig. 6).

The mentor system only needs to send to the trainee system the type of annotations and their position in the reference frame. This compact encoding of annotations saves bandwidth and is sufficient to recreate the annotations at the trainee system based on a local copy of the set of sprites.

5 Annotation Anchoring

As the tablet is repositioned, as the surgical field geometry changes, and as the surgical field becomes partially occluded due to the surgeon's hands and due to new instruments added to the surgical field, the annotations have to be repositioned to remain overlaid onto the surgical field elements that they describe. The process of computing the position of an annotation in the current video frame such that it remains in the same position relative to the surgical field as in the reference video frame where it was defined is called annotation anchoring. In Fig. 8, anchoring the annotation places it at

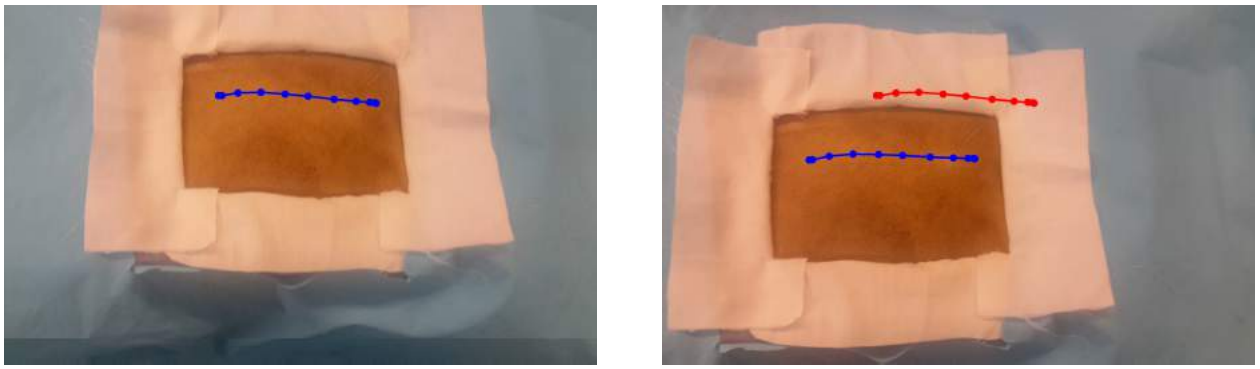

Fig. 8 Incision line annotation defined in reference frame with four segments (left, blue), obsolete annotation position in current frame (right, red), and correct position of annotation position in current frame (right, blue).

the correct location in the current frame, as the trainee tablet is repositioned.

Annotation anchoring is done in two major stages. The first stage preprocesses the reference frame where annotations are defined to prepare for annotation anchoring in future frames. The second stage uses the preprocessed reference frame and processes the current frame to anchor the annotation.

5.1 Reference frame preprocessing

The reference frame is preprocessed with an annotation anchoring preprocessing algorithm shown in Algorithm 1.

input : Reference frame F_0, annotation A defined in F_0
output: ORB features and descriptors of A region in F_0

Compute region R of A in F_0
Detect features f_{0i} in R using ORB
foreach f_{0i} **do**
  compute a descriptor d_{0i} using ORB
end
Return f_{0i} and d_{0i}

Algorithm 1: Annotation anchoring preprocessing of reference frame

The region R of the annotation is defined with an axis aligned rectangle that is obtained by enlarging the 2D axis aligned bounding box of the annotation points (step 1 of Algorithm 1). R is the black rectangle in Fig. 9. Feature points are identified in the region using the ORB feature detection algorithm, which uses FAST feature detection along with image pyramids to find multiscale features (step 2, Fig. 9 left) [19]. A descriptor is computed for each feature point using the ORB descriptor extraction algorithm (step 3, Fig. 9 right) [20]. The descriptor is a bit string that describes the pixel intensities at each pixel in an image patch surrounding the keypoint. This allows comparing the descriptors from the reference frame to descriptors of future frames. The annotation with its set of descriptors is sent to the trainee system where the annotation is tracked and displayed.

5.2 Actual annotation anchoring in current frame

input : Annotation A defined in reference frame F_0, ORB features f_{0i} and descriptors d_{0i} of A region in F_0, current frame F
output: Frame F with A overlaid at correct position

Detect features f_j in F using ORB
foreach f_j **do**
  compute a descriptor d_i using ORB
end
foreach d_{0i} **do**
  $d_{0i}.matchIndex = 0$
  $d_{0i}.matchDist = HammingDist(d0i, d0)$
  foreach d_j **do**
    if $d_{0i}.matchDist > HammingDist(d_{0i}, d_j)$ **then**
      $d_{0i}.matchIndex = j$
      $d_{0i}.matchDist = HammingDist(d_{0i}, d_j)$
    end
  end
end
$H = RANSACHomography(d_{0i}, d_j)$
foreach *point* p_i *of* A **do**
  $p'_i = Hp_i$
end
Render A with points p'_i in F
Return F

Algorithm 2: Annotation anchoring in current frame

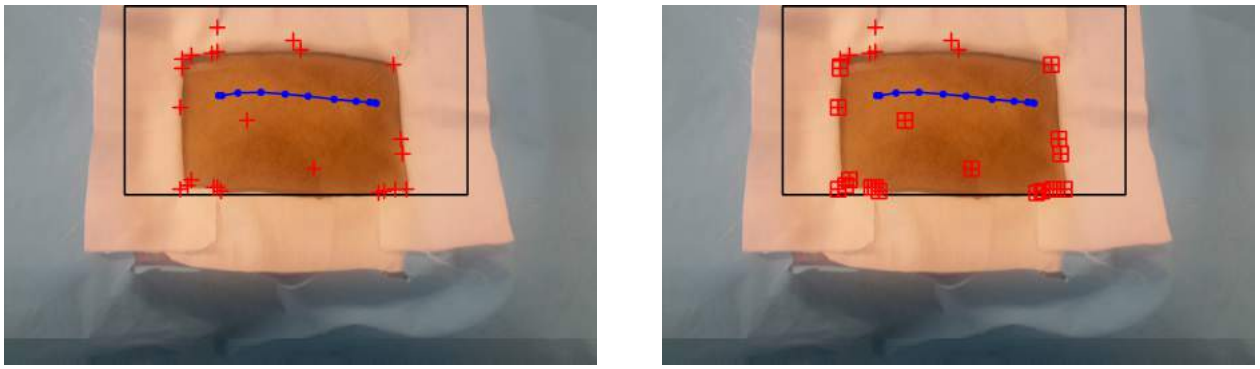

Fig. 9 Left: features (red crosses) detected in the reference frame in the region (black rectangle) of the incision line annotation (blue line). Right: descriptors (small red rectangles) computed for features to enable comparison and matching to descriptors in new frames.

The current frame is first processed similarly to the reference frame: features are detected and then enhanced with descriptor data (steps 1 and 2 of Algorithm 2, and Fig. 10 left). For some features near the edges of the frame, descriptor computation fails. This is because descriptor extraction involves reading the intensities of pixels in a ring surrounding the feature; if that ring extends beyond the edges of the image, there is insufficient information to complete the descriptor extraction.

Next, the reference frame's descriptors are matched to the current frame's descriptors using an all-pairs brute-force matching algorithm (step 3 of Algorithm 2). Each reference frame descriptor d_{0i} is matched against each current frame descriptor dj, selecting the match with the lowest Hamming distance between the descriptors. The matched descriptors are used to define a homography H from the reference frame to the current frame (Fig. 11, left) using a RANSAC-based algorithm (step 4) [9]. It should be noted that this homography computation method takes as one of its parameters a reprojection threshold, which determines whether a match is considered to be an inlier or an outlier. This threshold value is scaled based on the downsample factor of the input frame; otherwise, a smaller image with a relatively larger reprojection threshold would allow too many outliers to find a good homography. H maps a reference frame point to a current frame point. The homography is applied to each annotation point p_i, positioning the annotation in the current frame (step 5 and Fig. 11 right). Finally, the annotation is rendered with F as background at the position defined by the transformed points p'_i (step 6).

6 Results and Discussion

In this section we briefly describe the implementation of our first prototype system (Section 6.1), we report the results of our performance measuring experiments (Section 6.2), we summarize the feedback provided by the surgeons on our team after first trying the system (Section 6.3), we describe a user study we conducted to test the user experience and task efficiency of the system (Section 6.4), and we enumerate the limitations of this first prototype (Section 6.5).

6.1 Implementation overview

We have implemented the first system prototype using two Samsung Galaxy Tab Pro 12.2-inch tablets (each running Android 4.4.2), one for the trainee system, and one for the mentor system. Each tablet has a 1.9GHz and a 1.3GHz Quadcore processor, 3GB of RAM, 1,920 x 1,080 video camera, and a 2,560 x 1,600 display. All processing was performed exclusively on the two tablets. The system does not rely on additional workstations and it is therefore suitable for use in austere, resource-limited environments. The trainee tablet was suspended above the surgical field, into the trainee surgeon's field of view using a mechanical arm with interlocking joints.

In our experiments the mentor was located in a room adjacent to the trainees room. The distance was sufficiently short for the tablets to communicate via an ad-hoc Wi-Fi Direct network. For scenarios where the mentor is separated from the trainee by considerable geographic distance, the communication would be implemented via Wi-Fi and the Internet, with only minor modification to the systems software implementation.

Annotation anchoring was implemented relying on OpenCVs implementation of the ORB feature detection and descriptor extraction algorithms, and of a brute-force algorithm for estimating a homography from matched descriptors [12]. The annotations are overlayed onto video frames are drawn using OpenGL ES [13].

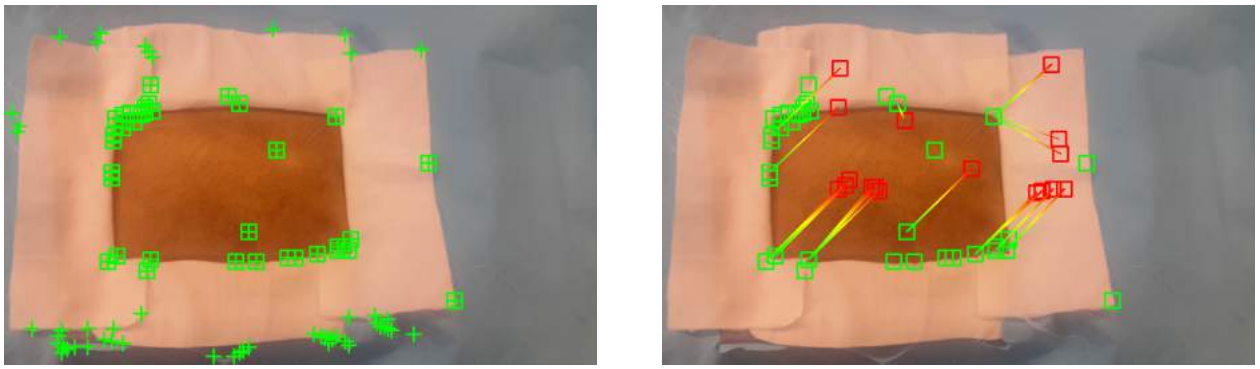

Fig. 10 Left: features (crosses) and descriptors (rectangles) in the current frame. Right: reference frame descriptors (red rectangles) matched to current frame descriptors (green rectangles).

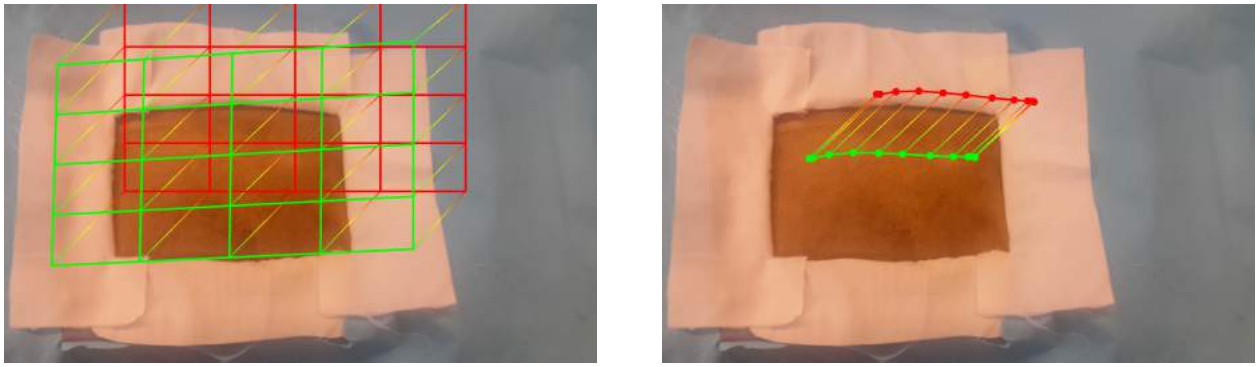

Fig. 11 Left: homography linking reference frame to current frame, visualized for a regular grid defined in the reference frame (red) that is mapped to the current frame (green). Right: annotation is anchored by mapping the annotation points from the reference to the current frame.

6.2 Performance

We quantify the system's performance through the *annotation anchoring error*, the *trainee system frame rate*, and the *mentor system frame rate*.

Annotation anchoring error The annotation anchoring error in one frame is measured in display pixels and it is defined as the Euclidean distance between the anchored annotation's location and the ground truth location of the annotation. The ground truth location of an annotation was defined with the following process. First, given the initial reference frame F_0, we inscribed each point of the annotation r_0 in a reference frame triangle whose vertices (r_{01}, r_{02}, and r_{03}) are defined by salient point features (Fig. 12, left).

The location of the annotation point within the triangle is defined by the point's barycentric coordinates, which are computed by solving a linear equation that inverts the barycentric interpolation:

$$r_0 = \lambda_1 r_{01} + \lambda_2 r_{02} + \lambda_3 r_{03} \quad (1)$$

$$\lambda_1 + \lambda_2 + \lambda_3 = 1 \quad (2)$$

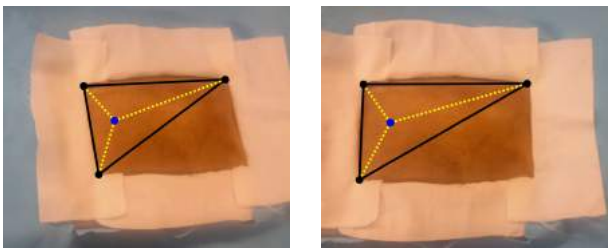

Fig. 12 Definition of ground truth annotation position. The reference frame (left) is used to compute the barycentric coordinates of the annotation point (blue) with respect to the reference triangle (black lines). The barycentric coordinates define the ground truth position of the annotation point in subsequent frames (right).

Then, in each subsequent frame F_i (Fig. 12, right), the salient point features of each triangle (r_{i1}, r_{i2}, and r_{i3}) are marked using a graphical user interface. Finally, the ground truth location r_i of the annotation point is derived by interpolation of the triangle vertices using the barycentric coordinates defined in the reference frame:

$$r_i = \lambda_1 r_{i1} + \lambda_2 r_{i2} + \lambda_3 r_{i3} \quad (3)$$

This process allows computing ground truth for an annotation point precisely even when the annotation point is in the middle of an area where features are scarce. It is most accurate for rigid, planar surfaces but is sufficiently accurate for ground truth acquisition in the experimental cases here. For an annotation defined with multiple points, e.g. the incision line annotation in Fig. 8 left, the annotation anchoring error is defined as the average of the anchoring errors at the individual points.

Table 1 gives the average anchoring error and the anchoring success rate for an incision line annotation over a sequence of frames for various experimental conditions and scenes. The success rate is computed as the number of frames where anchoring succeeds over the total number of frames. Anchoring succeeds when annotation anchoring error is below a threshold (we use 20 pixels). It should be emphasized that all errors are here recorded in terms of screen space pixels for the 2,560x1,600 display of our tablet; because our input frames are downsampled by a factor of 4 during processing, a 20-pixel error on screen is equivalent to a 5-pixel error on the frame as it is processed. The average error is computed over the frames where anchoring succeeds. In one scene the patient is approximated with a color anatomical poster printed at real world scale (top four frames in Fig. 13). In a second scene, the patient is approximated with a surgical dummy (bottom four frames in Fig. 13). The tablet repositioning conditions include lateral tablet translation (row 1 in Fig. 13), tablet rotation (left in row 2), and forward tablet translation that achieves a zoom effect (right in row 2). For minor and major occlusion conditions the tablet is fixed in the reference frame position and orientation and the frame is partially occluded by the trainee surgeon's hands (row 3). The surgical field deformation conditions were only applied to the surgical dummy scene. In the small deformation condition, the skin deforms as pressure is applied to the scalpel to perform the incision, and the incision becomes apparent (left in row 4). The large deformation condition corresponds to placing a retractor that opens up the wound, substantially changing the surgical field's appearance (right in row 4).

Annotation anchoring is more robust to tablet repositioning (78%-90% success rate in the surgical dummy scenario) and occlusion compared to deformation. In the case of tablet repositioning, anchoring succeeds as long as a sufficient set of reference frame features are still captured by the current frame. Anchoring is more robust with tablet translation since it only displaces the features, without changing their scale or orientation, leading to a high anchoring success rate of 89%-98% for tablet translation.

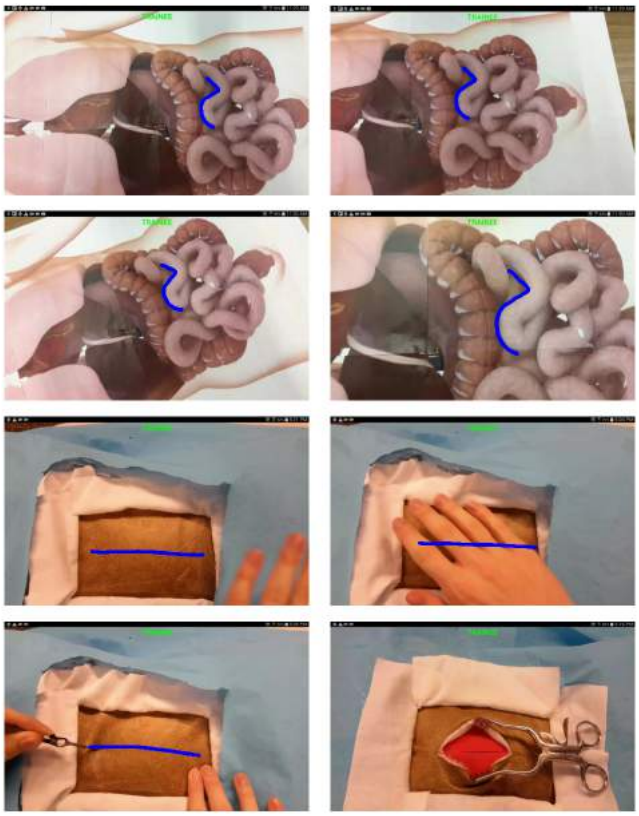

Fig. 13 Frames from the trainee tablet during our experiments with the anatomical poster (rows 1 and 2) and the surgical dummy (rows 3 and 4) scenes.

Anchoring is also robust with occlusions (60-100% success rate in the anatomical poster scenario, and 74%-96% in the surgical dummy scenario) because the tablet does not move with respect to the surgical field and the changes in the frame are confined to the occluded areas. The features that are not occluded have the same position and appearance as in the reference frame. Anchoring is the least robust with deformation.

Deformations and the addition of surgical instruments (Fig. 13, row 4, right) change the appearance of the surgical field substantially. Many of the original features are lost, new features are added, and even for the original features that persist, the homography model of the transformation is not sufficiently powerful. This leads to low anchoring success rates (15%-63%).

Fig.14 and 15 give the anchoring error for individual frames for the sequences used in Table 1. For each graph, the red curve gives the annotation error, the blue curve gives a measure of the difference between the current frame and the reference frame, and the black line shows the error threshold for successful anchoring.

For the tablet repositioning conditions (Fig. 14), the difference is measured by how much the annotation moved from the reference frame to its ground truth po-

Table 1 Average anchoring error in display pixels and annotation anchoring success rate.

		Experimental condition						
		Tablet repositioning			Surgical field occlusion		Surgical field deformation	
		Trans	Rot	Zoom	Minor	Major	Small	Large
Scene	Anatomical poster	2.66 98%	15.17 55%	7.27 80%	1.7 100%	1.27 60%	n/a	n/a
	Surgical dummy	3.65 89%	8.79 90%	6.41 78%	1.48 96%	3.65 74%	2.90 63%	2.73 15%

sition in the current frame. For translation, the difference is measured as the average translation of the annotation points. For rotation, the difference is measured as the angle between the incision line in the reference and current frames. For zoom, the difference is measured as the percentage ratio of the length of the incision line in the current frame over its length in the reference frame. For occlusion and deformation, the difference is measured as the percentage of the current frame that has changed compared to the reference frame.

For the tablet repositioning conditions, the graphs for both scenes show that: (1) anchoring is robust with translation even for large translation amplitudes; (2) anchoring fails for translation intermittently, for individual frames, but is regained on the following frame; (3) anchoring is less robust with rotation and zoom, being lost consistently for large rotation and zoom-in amplitudes; and (4) anchoring is more robust with zooming out compared to zooming in. Anchoring robustness decreases when only a few reference frame features are still visible in the current frame, as in frames with large amounts of translation, rotation, and zoom-in.

For the occlusion and deformation conditions (Fig. 15) the difference between the reference and current frames is measured as the percentage of pixels that changed, due to occlusions or to deformations. The number of changed pixels was computed automatically using background subtraction. For the occlusion conditions, anchoring recovers once the occlusion is removed. For the surgical dummy scene, anchoring is less robust to occlusions as most features are concentrated at the surgical field which represents a small fraction of the total frame, so even a small occlusion perturbs the detection of a large percentage of features. The minor deformation condition corresponds to performing the incision. The major deformation condition corresponds to several attempts to place the retractor. Anchoring recovers as the amount of deformation goes down. Once the retractor is placed and the deformation becomes permanent, annotation anchoring does not recover.

Trainee system frame rate As noted earlier, all computation is performed on the two tablets, without any help from auxiliary workstations. The overall and the individual stage running times of the annotation anchoring pipeline are given in Table 2. The figures were measured for the anatomical poster / tablet translation scene sequence. The running times are similar for other scenes and conditions. The running times were measured as averages over the frames of the sequence. The running times are given for various frame resolutions, starting with the full resolution and ending with a frame that was downsampled by a factor of 8.

As expected, overall and individual stage performance is strongly dependent on resolution (compare a processing time of 956 ms for a 1:1 scale image, and 153 ms for a 1:8 scale image). Higher resolution frames imply more pixels to examine when finding features, more features for which to find descriptors, more descriptors to match, and more matched features from where to compute the homographies. Descriptor extraction is usually a more laborious stage of the pipeline than feature detection (e.g., 585 ms versus 326 ms in the 1:1 scale image). Descriptor matching is a very fast process; even though the approach we use involves a brute-force method to find the most similar descriptors, the number of descriptors to match is usually low (about 50-100 descriptors per frame). The processing time for homography computation increases as the resolution decreases (ranging from 42 ms in the 1:1 image to 131 ms in the 1:8 image). This result is due to the rescaling of the RANSAC reprojection threshold with the change in resolution. To get accurate homographies, the error threshold for an match outlier must scale with the resolution; as there are fewer and more error-prone features in smaller images, the number of iterations in the RANSAC homography computation increases as it searches for acceptable inlier matches.

Annotation anchoring accuracy also depends on frame resolution. Table 3 gives the annotation anchoring error and the success rate as a function of the video frame resolution on which the anchoring algorithm is run. The

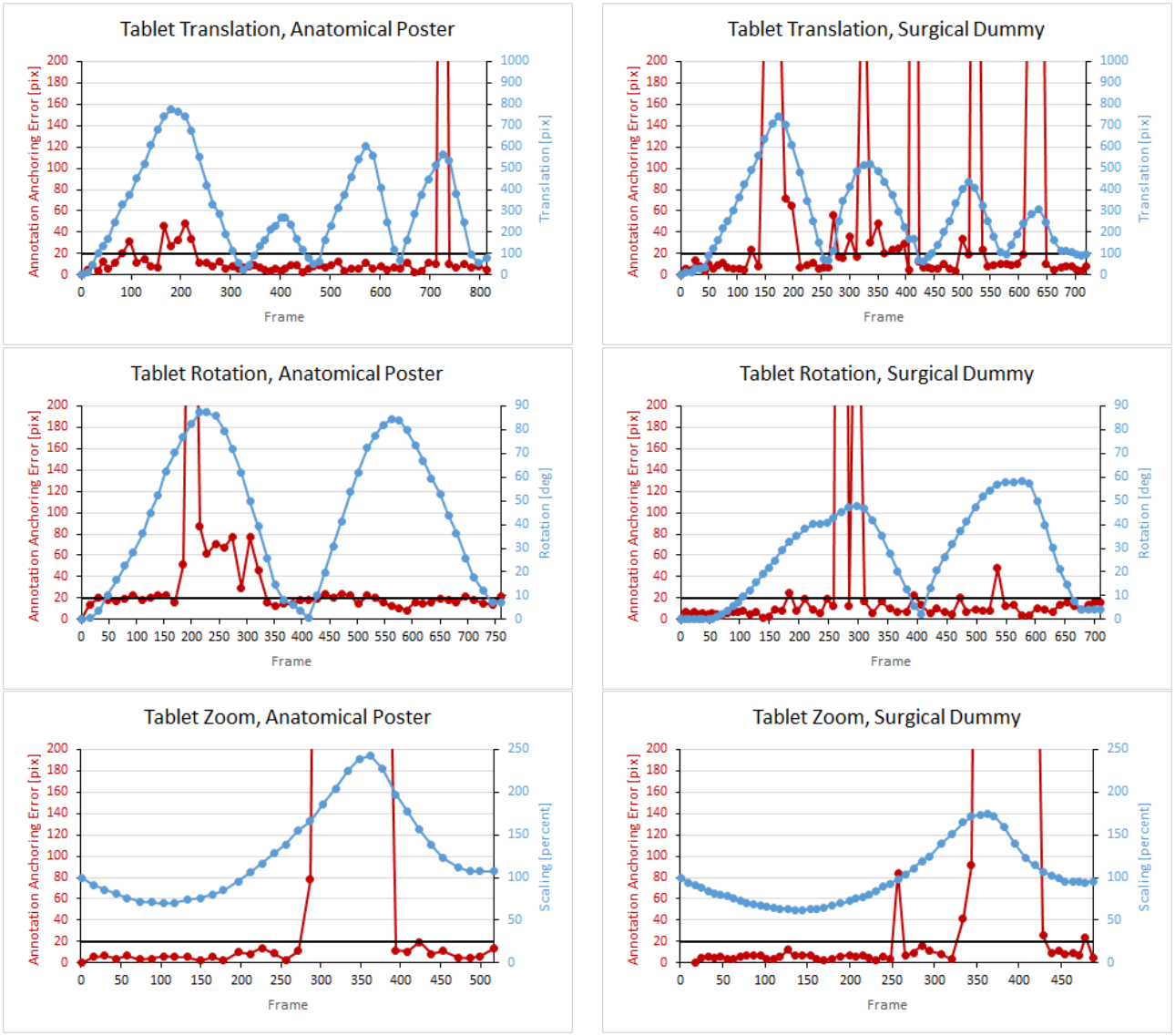

Fig. 14 Anchoring error graphs for the tablet repositioning conditions for the sequences from Table 1. The blue lines graph the change in tablet pose, the red lines graph the error values, and the black lines show the error threshold below which tracking was considered succesful.

Table 2 Running times for the annotation anchoring stages for various input image resolutions.

	Total Frame Time [ms]	**Feature Detection [ms]**	**Descriptor Extraction [ms]**	**Descriptor Matching [ms]**	**Homography Computation [ms]**
1920 x 1080 1:1	956	326 34.1%	585 61.2%	2 0.2%	42 4.4%
960 x 540 1:2	312	82 26.4%	162 52.1%	2 0.4%	65 20.9%
480 x 270 1:4	198	24 12.3%	44 22.1%	1 0.5%	128 65.0%
240 x 135 1:8	153	11 7.3%	10 6.6%	1 0.6%	131 85.7%

Fig. 15 Anchoring error graphs for the occlusion and deformation conditions for the sequences from Table 1. The blue lines graph the amount of occlusion or deformation, the red lines graph the error, and the black lines graph the error threshold.

scene corresponds to the surgical dummy, and the condition corresponds to the tablet translation. The error is given in output image pixels. The success rate increases (60% to 78%) when switching from full resolution to half resolution, which we attribute to the noise filtering benefit of downsampling. Downsampling the frame aggressively with factors of 1:8 and beyond drastically reduces the success rate (38% success rate in the 1:8 scale image). In practice we use a downsampling factor of 1:4 which achieves a good tradeoff between annotation anchoring robustness, accuracy, and performance, resulting in an 89% success rate in this particular scenario.

The systems provides two modes of displaying the annotations on the trainee tablet. In a first mode, the display frame rate is decoupled from the annotation anchoring frame rate. The display is updated at the video acquisition frame rate of 30Hz and the annotations' positions are updated at annotation anchoring frame rate. This decoupled mode has the advantage of a fluid, real-time display of the surgical field. The disadvantage of the decoupled mode is that, during tablet repositioning, the annotations drift in between annotation anchoring updates, as they are overlaid on more recent frames than the frames where they are anchored. The drift increases the perceived annotation anchoring error. The maximum annotation anchoring error occurs just before annotation anchoring completes.

For an annotation anchoring frame rate of 10Hz and video rate of 30Hz, the annotation's position is 3 to 5

Table 3 Annotation anchoring error in display image pixels and annotation anchoring success rate for various video frame downsampling factors.

	1:1 1920 x 1080	**1:2 960 x 540**	**1:4 480 x 270**	**1:8 240 x 135**
Error [2,560 x 1,600 pix]	8.14	8.35	3.65	12.08
Success rate	60%	78%	89%	38%

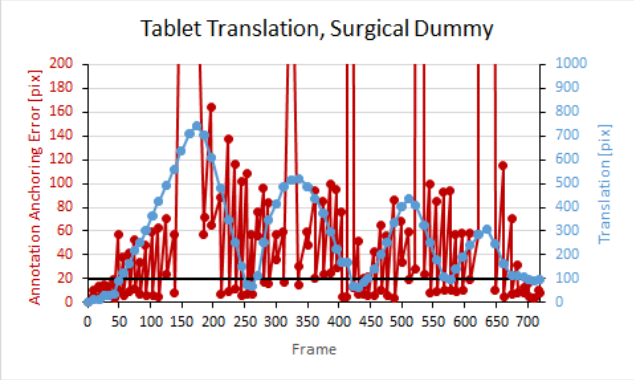

Fig. 16 Total perceived annotation anchoring error in decoupled mode.

video frames behind: 3 for the frame when the annotation anchoring data has just been updated, and 5 when the it is about to be updated. The frame latency translates to annotation anchoring errors according to tablet repositioning speed. If the tablet moves quickly, a 5 frame latency can lead to an annotation anchoring error of hundreds of pixels. Once the tablet stabilizes, the additional annotation anchoring error due to latency decreases, vanishing after 6 frames.

Fig. 16 shows the total, perceived annotation anchoring error in the case of the surgical dummy scene for tablet translation. If the annotation anchoring algorithm is run on frame F_a and if the frame that is displayed is F_d, the total error is computed as the sum of two errors: the error with which the annotation is anchored in F_a, plus the latency error due to how much the annotation has drifted from F_a to F_d. The total error increases when the tablet moves at a faster rate (e.g. frame 196) and decreases when the tablet stabilizes (e.g. frame 709).

In a second mode, display and annotation anchoring frame rates are coupled. The system only displays a new frame when annotation anchoring completes. The advantage is that there is no annotation anchoring error due to latency, but this comes at the cost of less frequent updates of the trainee's hands and instruments, and of the visualization of the surgical field.

Mentor system frame rate The mentor system's performance depends on the transfer rate of frames from the trainee to the mentor. A video frame is downsampled with an 1:4 factor, losslessly encoded as a PNG image, and transmitted via WiFi direct. In our experiments we measured a sustained mentor system frame rate of 5fps, and a maximum frame rate of 10fps. Once the mentor annotates a reference frame, the mentor system computes features and descriptors in the reference frame, which are sent along with the annotation data to the trainee system. Compared to the reference frame itself, this metadata is of negligible size.

6.3 System Usability

We tested an initial system prototype with surgeons from the Indiana University School of Medicine trauma team. First we demonstrated the system to the surgeons in a conference room using the anatomical poster. This initial demonstration conveyed the system functionality, and how the system is to be used by the mentor and the trainee surgeon. Then we asked two surgeons to use the system in the context of a cricothyrotomy and of a lower limb fasciotomy using a euthanized porcine model. (The porcine model was used during a regularly scheduled third year surgical resident training laboratory course independent of our research.) The mentor indicated the location of the incisions, and the trainee replicated those incisions following the annotation lines that were directly overlaid onto the surgical field.

The formative evaluation revealed several shortcomings of the system that should be removed in the next iterations of system refinement. The trainee surgeon did not find usable the coupled mode that displays the surgical field at annotation anchoring rate. The delay between actual hand motion and the appearance of the hand motion on the tablet was disconcerting. The trainee surgeon would prefer updates to the surgical field at the highest frame rate possible. This shortcoming has already been addressed with the creation of the decoupled mode described above.

Another system shortcoming was a perceived complexity of the mentor system user interface. The mentor

favored simplifying the interface to only line-based annotations. Although it was not the case in this particular test, we foresee scenarios where the nature of the surgery and the trainee's level of expertise could require a rich interface with many annotation types.

The test revealed deficiencies in the first implementation of the mechanical arm holding the tablet above the surgical field. One deficiency was the inability to hold certain desired tablet positions and orientations. Also, the arm lacked the required range of motion: the porcine model's position during the fasciotomy surgery required lifting the tablet high above the table to leave enough room for the trainee to operate, a high position poorly suited for the arm. This shortcoming has already been addressed by redesigning the mechanical arm for increased stability and range of motion.

The test revealed that tablet repositioning is a relatively rare event, and therefore future work on annotation anchoring robustness should probably focus on occlusion and deformation conditions. The infrequent substantial repositioning of the tablet can be handled by asking the mentor to recreate or manually anchor the annotations for a new reference frame.

Finally, a practical telementoring system requires establishing and observing an interaction protocol between mentor and trainee. For example, the trainee should not occlude the surgical field such that the mentor can annotate a suitable reference frame. Capturing a reference frame with transient occlusions, for example with the trainee hands moving in the field of view, will unnecessarily weaken annotation anchoring.

6.4 Pilot Test

A pilot user study was conducted to compare the hand-eye coordination, task accuracy and task completion time of participants when using our augmented reality system (AR), compared with using a conventional system for telementoring based on displaying mentor feedback on a nearby monitor (Conventional). Fig. 17 shows the AR and Conventional setups.

Participants Twenty-two participants were recruited from graduate students of computer science and industrial engineering programs at Purdue University. The participants were randomly divided into two equally-sized groups and assigned to the AR and the Conventional conditions. Each participant wore a Google Glass head-mounted camera, which acquired a video of the task from the participant's point of view.

Task A medically relevant aim of this study was to assess a trainee's ability to identify regions in the neck area of a patient, which usually is a necessary condition to conduct a cricothyrotomy. The participants' task was to place seven circular paper stickers (6.35 mm in diameter) near the neck region of a patient simulator at precise locations indicated one at a time by the mentor. The task was repeated three times with different paper sticker location patterns. Each participant was given verbal instructions on how to complete the task before the actual experiment. As part of the task description, the participants were asked to place the stickers as quickly and accurately as possible. The instructions took approximately two minutes.

AR condition For the participants that used our telementoring system, the mentor indicated the location of the next sticker with a virtual annotation on the transparent display. The participant would see their hand and the sticker through the transparent display and would guide the sticker to coincide with the virtual annotation. The tablet was placed at the same relative position and orientation with respect to the patient simulator for each participant using a robotic arm. This allowed interleaving experiments for participants in the AR and Conventional groups.

Conventional condition For the control condition, a 46-inch LCD monitor was used to display the position of the markers prescribed by the mentor. The participant would look at the LCD and then back at the patient simulator for guidance as to where to place each sticker.

Methods For each condition, each participant, and each seven sticker trial, the following data was recorded: (1) the time needed to place all seven stickers; (2) the number and duration of focus shifts, which was obtained by analyzing each video recorded by the Google Glass head-mounted camera worn by the participant during the experiment; (3) the sticker placement error in pixels, which was computed by taking a photograph of the seven stickers placed on the patient simulator and by measuring the distance between the actual and the mentor prescribed position of the stickers.

Results and discussion The average (max, min) placement error was 59.6 (467.8, 4.3) pixels for the Conventional condition, and 32.0 (168.5, 1.0) pixels for the AR condition (for an image resolution of 2,560 x 1,600 pixels). To provide real-world context for these results, given the pose of the tablet camera in relation to the patient simulator, this translates to an average error of approximately 0.97 cm for the Conventional condition, and and average of 0.52 cm for the AR condition. Fig. 18 shows sticker placement accuracy for the Conventional

Fig. 17 Experimental setup for the AR (left) and Conventional (right) conditions.

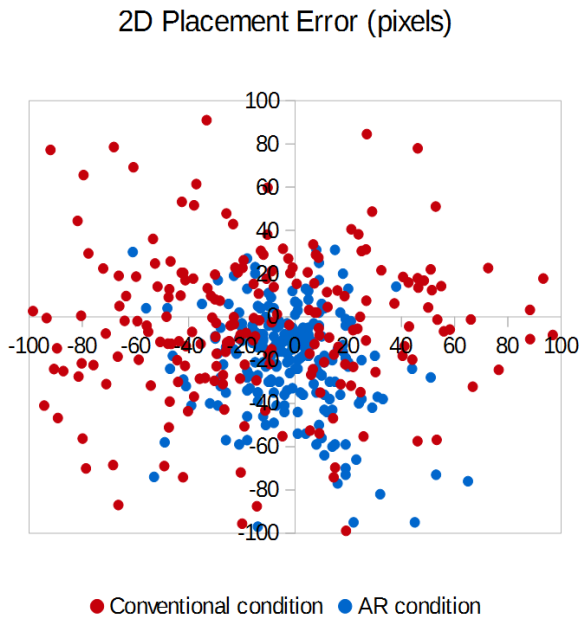

Fig. 18 2D placement error for individual stickers for the AR (blue) and Conventional (red) conditions.

(red) and the AR (blue) groups. Participants for the Conventional condition shifted focus away from the operating field an average (max, min) of 13.8 (26.0, 7.0) times per seven sticker placement trial, and focus was shifted for an average of 34% (43%, 21%) of the trial duration. Participants for the AR condition shifted focus away from the operating field an average (max, min) of 6.6 (15.0, 2.0) times, for 14% (48%, 0%) of the task duration. The average (max, min) completion time for each trial was 41.31 (97.70, 25.70) seconds for the Conventional condition, and 53.44 (80.70, 31.52) seconds for the AR condition.

On average, the placement error was considerably smaller when using the AR system than when using a separate screen. The tablet provides precise feedback as to where the sticker should be placed and the participant leverages this feedback to minimize placement error. Several non-tablet participants commented that, in cases when they were not already looking at the screen when a new virtual annotation was displayed, they had difficulty identifying which annotation was the newest one. This is an expected shortcoming of conventional systems, where the need for a focus shift implies that a trainee may not receive information as soon as it arrives. No participants for the AR condition indicated such a difficulty.

Focus shifts were greatly reduced when using the tablet system as opposed to the conventional system. This is a reasonable result, given that a participant in the Conventional condition is required to shift focus in order to access the instruction, while in the AR condition accessing the instruction does not require shifting focus. Although for some participants in the AR condition there was no focus shift, somewhat surprisingly, the focus shifts were not zero for all participants. For example, we noted during the experiment that one participant in the AR condition repeatedly shifted focus to look under the tablet at the real scene below. Some participants who performed the task for the AR condition commented that a lack of depth perception from the tablet screen, as well as a slight latency in the camera, caused difficulty with hand-eye coordination. For consistency we opted for using the same relative position between the tablet and the patient simulator, although participants varied in height and therefore the selected relative position might not have been ideal.

One interesting result is that the task completion time was slightly longer for the AR condition than it was for the Conventional condition. Possible causes could be deficiencies in hand-eye coordination due to the lack of a fully transparent effect on the display, or the positioning of the tablet being cumbersome for some users. However, when taken with the result that placement error was worse for the Conventional condition, this could indicate that participants spent more time when they

had more immediate feedback and had the potential to be more accurate, as in the AR condition. In contrast, the Conventional condition provides no live feedback of the user's correct positioning, and so participants may elect to quickly place the stickers at their best guessed location in the absence of feedback.

Conclusion The study provides a preliminary indication that the AR system allows trainees to follow some mentor instructions more accurately. According to a surgeon on our team, a reasonable upper bound for accuracy on surface-level surgical actions is approximately 1 cm. As such, the reduction in average placement error from 0.97 cm (in the Conventional condition) to 0.52 cm (in the AR condition) suggests that the AR system can provide meaningful improvements to the accuracy of surgical tasks. We hypothesize that the biggest shortcomings of this initial implementation of our AR system is the lack of perfect transparency (i.e. the tablet image is not seamlessly aligned with the parts of the surgical field directly observed), and the lack of depth perception. These issues will be addressed in future versions of the system.

6.5 Limitations

Two interdependent shortcomings of the first system prototype are low frame rate and limited annotation anchoring robustness. In addition to reducing the latency annotation anchoring error, a higher frame rate will also allow computing annotation anchoring in higher resolution frames, which will decrease the annotation anchoring error and will increase robustness. We will pursue the acceleration of annotation anchoring by parallelizing the implementation, leveraging the multiple cores and the GPUs available on the tablets.

We will also design novel anchoring algorithms that define custom descriptors at the annotation points, which are then tracked individually. This has the potential to reduce the number of features and descriptors substantially. Moreover, individually tracked descriptors eliminate the oversimplified modeling through a homography of the transformation from the current to the reference frame. The homography model essentially assumes that the surgical field is planar and rigid. The assumption does not hold in the cases of 3-D surgical fields and surgical field deformations. For example, during the large deformation shown in Fig. 13, row 4, right, an annotation anchored above the incision line should remain anchored even when the skin deforms due to the retractor's placement. The annotation should move with the skin as it deforms.

The current implementation sends individual frames from the trainee to the mentor; this is adequate for the reference frame where the mentor creates annotations, but is inadequate in terms of providing the mentor with a high-frame rate video of the surgical field. Another low level limitation is that the current system does not provide an audio connection between trainee and mentor. In our tests audio communication was provided via a speakerphone. Both of these limitations can be easily addressed by streaming both video and audio between the two sites, in addition to occasional transfers of reference frames and annotations.

7 Conclusions and Future Work

We have described an approach for improving surgical telementoring based on an AR transparent display, as well as a first implementation of this approach that reveals that the approach is promising.

In addition to addressing the low-level limitations as described above, we will work towards improving the transparent display approximation provided by the system, building upon prior work into simulated transparent displays [4, 25, 27]. The current system does not achieve perfect visual continuity between the parts of the surgical field seen through the display and the parts seen directly (Fig. 2, right). The video frame is displayed as-is, from the viewpoint of the trainee tablet's video camera. For a better simulation of transparency, the video frame must be reprojected to the trainee's point of view. The reprojection operation requires solving the following sub-problems: (1) tracking the trainee's head, (2) knowing the geometry of the surgical field, and (3) filling in color information missing from the current frame.

The possible solutions to the first problem are using the front-video camera on the trainee tablet, using an external tracking system, or using a next generation tablet that has built user head tracking capability. Such a capability is already available in Amazon's "Fire Phone" smartphone, which has four front-facing cameras, two of which are used to triangulate the user's head position [2]. Possible solutions to the second problem include external depth acquisition using a separate depth camera, or on-board depth acquisition by attaching a depth camera to the trainee tablet, such as the Structure sensor [17]. Another option that we plan to investigate is the use of the Google Project Tango tablet, which uses an integrated infrared depth sensor combined with motion sensors to provide accurate pose estimation and depth acquisition [10]. The third problem can be solved by filling in the color samples needed but not present in the current frame from older frames.

The color samples could be missing due to field of view limitations, and due to occlusion changes as the viewpoint changes from that of the video camera to that of the trainee.

Acknowledgements We thank Sthitapragyan Parida for his help with the implementation and demonstration of our telementoring system. We thank Chun-hao Hsu and Aviran Malik for their help with the tablet mount system used in our experiments. We thank Meng-Lin Wu, Xiaoxian Dong, Chengyuan Lin, and the entire computer graphics group at the computer science department of Purdue University for their feedback on this work.

This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award No. W81XWH-14-1-0042. Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the Department of Defense.

References

1. Agarwal, R., Levinson, A.W., Allaf, M., Makarov, D.V., Nason, A., Su, L.M.: The roboconsultant: telementoring and remote presence in the operating room during minimally invasive urologic surgeries using a novel mobile robotic interface. Urology **70**(5), 970–974 (2007)
2. Amazon.com, I.: Amazon Fire Phone (2014). URL `http://www.amazon.com/firephone`
3. Ballantyne, G.H.: Robotic surgery, telerobotic surgery, telepresence, and telementoring. Surgical Endoscopy and Other Interventional Techniques **16**(10), 1389–1402 (2002)
4. Baričević, D., Höllerer, T., Sen, P., Turk, M.: User-perspective augmented reality magic lens from gradients. In: Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology, pp. 87–96. ACM (2014)
5. Bashshur, R.L.: On the definition and evaluation of telemedicine. Telemedicine Journal **1**(1), 19–30 (1995). DOI 10.1089/tmj.1.1995.1.19. URL `http://dx.doi.org/10.1089/tmj.1.1995.1.19`
6. Bogen, E.M., Augestad, K.M., Patel, H.R., Lindsetmo, R.O.: Telementoring in education of laparoscopic surgeons: An emerging technology. World Journal of Gastrointestinal Endoscopy **6**(5), 148–155 (2014). URL `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4024487/`
7. Chou, W., Wang, T., Zhang, Y.: Augmented reality based preoperative planning for robot assisted teleneurosurgery. In: Systems, Man and Cybernetics, 2004 IEEE International Conference on, vol. 3, pp. 2901–2906 vol.3 (2004)
8. Ereso, A.Q., Garcia, P., Tseng, E., Gauger, G., Kim, H., Dua, M.M., Victorino, G.P., Guy, T.S.: Live transference of surgical subspecialty skills using telerobotic proctoring to remote general surgeons. Journal of the American College of Surgeons **211**(3), 400–411 (2010)
9. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM **24**(6), 381–395 (1981)
10. Google: ATAP Project Tango (2014). URL `https://www.google.com/atap/projecttango/`
11. Guo, Y., Henao, O., Jackson, T., Quereshy, F., Okrainec, A.: Commercial videoconferencing for use in telementoring laparoscopic surgery. Medicine Meets Virtual Reality 21: NextMed/MMVR21 **196**, 147 (2014)
12. Itseez: OpenCV (2014). URL `http://opencv.org/`
13. Khronos: OpenGL ES - the standard for embedded accelerated 3D graphics (2014). URL `https://www.khronos.org/opengles/`
14. Loescher, T., Lee, S.Y., Wachs, J.P.: An augmented reality approach to surgical telementoring. In: Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on, pp. 2341–2346. IEEE (2014)
15. Marescaux, J., Diana, M.: Robotics and remote surgery: Next step. In: K.C. Kim (ed.) Robotics in General Surgery, pp. 479–484–. Springer New York (2014)
16. Marescaux, J., Rubino, F.: Telesurgery, telementoring, virtual surgery, and telerobotics. Current urology reports **4**(2), 109–113 (2003)
17. Occipital, I.: The Structure Sensor is the first 3D sensor for mobile devices (2014). URL `http://structure.io/`
18. Ponce, B.A., Jennings, J.K., Clay, T.B., May, M.B., Huisingh, C., Sheppard, E.D.: Telementoring: Use of augmented reality in orthopaedic education. The Journal of Bone & Joint Surgery **96**(10), e84– (2014). URL `http://jbjs.org/content/96/10/e84.abstract`
19. Rosten, E., Drummond, T.: Machine learning for high-speed corner detection. In: Computer VisionECCV 2006, pp. 430–443. Springer (2006)
20. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF. In: Computer Vision (ICCV), 2011 IEEE International Conference on, pp. 2564–2571. IEEE (2011)
21. Satava, R.: Virtual endoscopy. Surgical endoscopy **10**(2), 173–174 (1996)
22. Schulam, P., Docimo, S., Saleh, W., Breitenbach, C., Moore, R., Kavoussi, L.: Telesurgical mentoring. Surgical endoscopy **11**(10), 1001–1005 (1997)
23. Shenai, M.B., Dillavou, M., Shum, C., Ross, D., Tubbs, R.S., Shih, A., Guthrie, B.L.: Virtual interactive presence and augmented reality (VIPAR) for remote surgical assistance. Neurosurgery **68**, – (2011)
24. Smurro, J.P., Reina, G.A., L'esperance, J.O.: System and method for surgical telementoring and training with virtualized telestration and haptic holograms, including metadata tagging, encapsulation and saving multi-modal streaming medical imagery together with multi-dimensional [4-d] virtual mesh and multi-sensory annotation in standard file formats used for digital imaging and communications in medicine (dicom) (2013)
25. Tomioka, M., Ikeda, S., Sato, K.: Approximated user-perspective rendering in tablet-based augmented reality. In: Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on, pp. 21–28. IEEE (2013)
26. Treter, S., Perrier, N., Sosa, J.A., Roman, S.: Telementoring: a multi-institutional experience with the introduction of a novel surgical approach for adrenalectomy. Annals of surgical oncology **20**(8), 2754–2758 (2013)
27. Unuma, Y., Niikura, T., Komuro, T.: See-through mobile ar system for natural 3d interaction. In: Proceedings of the companion publication of the 19th international conference on Intelligent User Interfaces, pp. 17–20. ACM (2014)
28. Vera, A.M., 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. Surgical endoscopy **28**(12), 3467–3472 (2014)