ARssist: augmented reality on a head-mounted display for the first assistant in robotic surgery

Long Qian ✉, *Anton Deguet, Peter Kazanzides*

Johns Hopkins University, Baltimore, Maryland, USA
✉ *E-mail: long.qian@jhu.edu*

Published in Healthcare Technology Letters; Received on 13th August 2018; Accepted on 20th August 2018

In robot-assisted laparoscopic surgery, the first assistant (FA) is responsible for tasks such as robot docking, passing necessary materials, manipulating hand-held instruments, and helping with trocar planning and placement. The performance of the FA is critical for the outcome of the surgery. The authors introduce *ARssist*, an augmented reality application based on an optical see-through head-mounted display, to help the FA perform these tasks. *ARssist* offers (i) real-time three-dimensional rendering of the robotic instruments, hand-held instruments, and endoscope based on a hybrid tracking scheme and (ii) real-time stereo endoscopy that is configurable to suit the FA's hand–eye coordination when operating based on endoscopy feedback. *ARssist* has the potential to help the FA perform his/her task more efficiently, and hence improve the outcome of robot-assisted laparoscopic surgeries.

1. Introduction: In a *da Vinci*® robot-assisted surgery, the main surgeon sits at the console teleoperating the robot, while the patient-side assistant stands or sits at the bedside assisting the operation (see Fig. 1). The patient-side assistant also called a bedside assistant, scrubbed surgeon [1], or first assistant (FA) [2], plays an important role in the robotic laparoscopic surgery. Before the main surgeon starts tele-operation, the FA is responsible for or takes an important role in trocar placement, docking of the robot, and preparing the operative field. During the surgery, the FA exchanges the instrument for the main surgeon manipulates certain laparoscopic instruments, e.g. gripper and vessel sealer, and extracts specimen [1–3].

The outcome of a robotic surgery is dependent on the performance of the FA. Through an analysis of 222 urologic cases, researchers have identified that the mean operative time for all robotic procedures showed a consistent trend of reduction with increasing experience of the FA [4]. In another study comparing the performance of well-trained and less-trained FAs among 280 different robotic surgical interventions, the authors concluded that interventions with a well-trained FA are more rapid and secure [3].

The *da Vinci*® system restores the hand–eye coordination for the main surgeon by providing an immersive endoscopic operative field and letting him/her intuitively control the robotic instruments, which appear registered with his/her hand motion [5]. However, this configuration and resulting benefits are not provided to the FA. For example, when the FA needs to install or exchange an instrument for the main surgeon, he/she has to manually and blindly adjust the robotic arm in order for the instrument to appear in the operative field or have the console surgeon reposition the endoscope to visualise the instrument until it arrives at the desired location. As another example, when the FA is manipulating instruments inside the patient body, he/she has to look at the monitor mounted on the vision cart that is not near the operative field, which leads to an awkward hand–eye coordination.

We propose to use augmented reality (AR), based on an optical see-through head-mounted display (OST-HMD), to address the aforementioned problems of current laparoscopic robots. We choose an OST-HMD because it provides the user with an unhindered and instantaneous real-world view [6], with computer graphics presented to the user on top of the real-world view through optical combiners. An OST-HMD is fail-safe, in that the user is still able to operate even if the system completely fails to deliver any augmentation. In addition, many OST-HMD products have recently entered the commercial market, which enables good performance in visualisation, computation, tracking and ergonomics [7].

In this Letter, we present our system *ARssist*, an application based on the integration of a surgical robot and an OST-HMD. *ARssist* provides various AR information to the FA, including (i) three-dimensional (3D) real-time rendering of the endoscope, robotic instruments and hand-held instruments within the patient body, and (ii) real-time stereo endoscopy that is configurable for the FA's preferred hand–eye coordination. *ARssist* has the potential to help FAs perform their tasks better, and hence improve the outcome of robotic laparoscopic surgeries.

2. Background and related work: AR has been used for a number of medical applications [8], including laparoscopic surgery [9]. AR in laparoscopic surgery offers various advantages to the surgeon: (i) it provides guidance to critical targets and structures, (ii) it reduces the surgeon's cognitive load, (iii) it can display pre-operatively planned trajectories on a virtual model, and (iv) it increases the surgeon's spatial awareness [9]. Most laparoscopic AR applications are composed of two components: registration of the augmentation to the scene and tracking to maintain the accuracy of the visualisation. In addition, the video endoscopy is sometimes further processed to match the viewpoint of the surgeon to enhance hand–eye coordination [10], using either 3D surface reconstruction [11, 12] or 2D image-based processing [13, 14].

In the setup of a da Vinci robot-assisted laparoscopic surgery, AR can be implemented on the video display of the surgeon console by overlaying virtual information on the real-time endoscopy [15]. However, to our knowledge, the benefit of AR has not previously been implemented or investigated by other members of the surgical team, especially for the FA who plays a critical role in a robotic surgery. We introduce the *ARssist* application to provide this benefit.

3. Methods

3.1. Components and transformation map in ARssist: To offer visualisation of robotic instruments and hand-held instruments at the correct location and orientation with respect to the viewer, the system must track them in real time. Fig. 2 shows some *ARssist* components. We assume that these components are rigid bodies and affix a Cartesian coordinate system to each one.

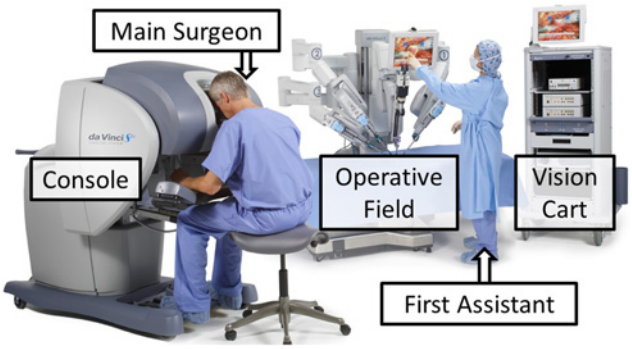

Fig. 1 *Surgery team with a da Vinci S® surgical robot; image © 2018 Intuitive Surgical, Inc.*

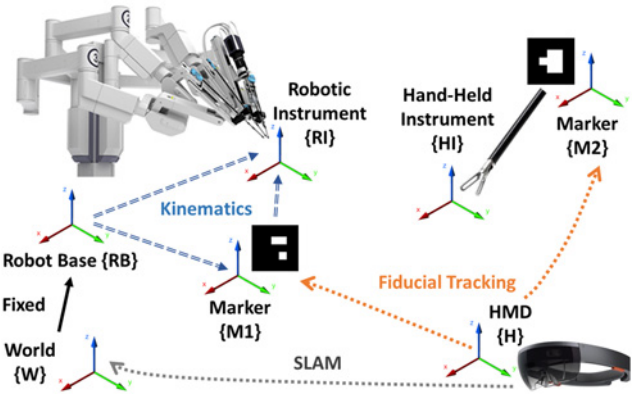

Fig. 2 *Components of ARssist and their relative transformations*

Different components in *ARssist* are geometrically linked in various ways. In a robotic surgery, once docked the robot remains stationary within the operating room. Therefore, it is safe to assume that the transformation of the world and the robot base is fixed, i.e. $\boldsymbol{T}_{\mathrm{W}}^{\mathrm{RB}}$ is a constant. The robotic instruments are controlled precisely by the robot during the surgery. The transformation between the robot base and robotic instrument, $\boldsymbol{T}_{\mathrm{RB}}^{\mathrm{RI}}$, is obtained from the robot model and real-time kinematics data. We attach fiducial markers to certain parts of the robot and to the hand-held instruments to support optical tracking on the HMD. The markers cannot be attached directly to the tool tip because they will not be visible to the HMD during the surgery. As a result, the fiducial markers are 'plugged' into the robot kinematics chain. The poses of the markers are dependent on joints that are closer to the base of the robot. For a robotic instrument, $\boldsymbol{T}_{\mathrm{RB}}^{M_1}$ and $\boldsymbol{T}_{M_1}^{\mathrm{RI}}$ are both obtained from robot kinematics, and for hand-held instruments, the transformation $\boldsymbol{T}_{M_2}^{\mathrm{HI}}$ is fixed. The transformations between the markers and the HMD, $\boldsymbol{T}_{M_1}^{\mathrm{H}}$ and $\boldsymbol{T}_{M_2}^{\mathrm{H}}$, are computed at runtime through vision-based tracking algorithms. In addition, it is notable that recent HMDs, such as Microsoft HoloLens [16] and Meta 2 [17], offer inside-out localisation. The HMD can compute $\boldsymbol{T}_{\mathrm{W}}^{\mathrm{H}}$ at runtime through inside-out tracking methods, e.g. simultaneous localisation and mapping (SLAM) [18].

Therefore, the transformation between the HMD and a robotic instrument can be computed in two ways:

$$\boldsymbol{T}_{\mathrm{H}}^{\mathrm{RI}} = \boldsymbol{T}_{M_1}^{\mathrm{RI}} \cdot \boldsymbol{T}_{\mathrm{H}}^{M_1}, \tag{1}$$

$$\boldsymbol{T}_{\mathrm{H}}^{\mathrm{RI}} = \boldsymbol{T}_{\mathrm{RB}}^{\mathrm{RI}} \cdot \boldsymbol{T}_{\mathrm{W}}^{\mathrm{RB}} \cdot \boldsymbol{T}_{\mathrm{H}}^{\mathrm{W}}. \tag{2}$$

The transformation between the HMD and a hand-held instrument is obtained by

$$\boldsymbol{T}_{\mathrm{H}}^{\mathrm{HI}} = \boldsymbol{T}_{M_2}^{\mathrm{HI}} \cdot \boldsymbol{T}_{\mathrm{H}}^{M_2}. \tag{3}$$

Equation (1) uses the fiducial tracking, the kinematics data, the model of the robot, and the pivot calibration that determines the pose of the marker relative to a certain joint of the robot. Equation (2) uses the inside-out tracking capabilities of the HMD, the robotic model, kinematics data, and the calibration. Equation (3) uses the fiducial tracking and pivot calibration. It is notable that there exists redundancy in the tracking of robotic instruments.

3.2. Hybrid tracking scheme for robotic instruments: In *ARssist*, we take advantage of the redundancy and employ a hybrid tracking scheme, derived from [19], to localise the robotic instruments. Our tracking scheme is composed of three steps. First, the prioritisation of each transformation is determined with prior knowledge, so that reliable and accurate transformations are given higher priority. Then, we prioritise different tracking methods, which are constructed by composing transformations with different priorities. These two steps are conducted in an offline stage. Finally, at the online stage, we always use the tracking method of the highest priority when it is available. When the highest priority method is not available, e.g., due to the line-of-sight loss, we model the discrepancy between the lower and higher priority tracking methods as a static error and compensate for it when switching from the high-priority tracking method to a low-priority tracking method.

The transformations that are fixed or derived from kinematics data are given high priorities because they are most reliable in terms of accuracy and latency. They can be reliably calculated within a few millimetres [20]. Transformations obtained from fiducial tracking are assigned medium priority. The accuracy of fiducial marker tracking will suffer when the relative motion between the HMD and the marker is more significant, as the latency caused by camera exposure and computation is not negligible. Futhermore, the accuracy of camera-based fiducial tracking is affected by the distance from the object to the camera, and specific software algorithm. It could be around several centimetres [21]. At last, we assign a low priority to the self-localisation of the HMD. Note that we assign these priority levels based on the current generation of HMD hardware and software. We summarise the transformations and the priorities between each component shown in Table 1.

3.2.1 Visual overlay on OST-HMD: A video see-through head-mounted display has access to every pixel that the user sees and therefore is able to blend the overlaid graphics perfectly with the background. However, an OST-HMD does not have access to the user's retina, which makes visual alignment a non-trivial problem. This problem is usually referred to as the display calibration of OST-HMD, and many solutions have been proposed to solve it, e.g. single point active alignment method (SPAAM) [22], interaction-free calibration method (INDICA) [23] and display relative calibration (DRC) [24].

In *ARssist*, a binocular OST-HMD is used; therefore, we chose to use a variant of the SPAAM method [25], which better supports a binocular OST-HMD and is integrated with development tools such as rendering engines. Assume that a 3D point p is defined in the

Table 1 Transformations and priorities between components of *ARssist*

Transformation	Method of computation	Priority
world to the robot base $\boldsymbol{T}_{\mathrm{W}}^{\mathrm{RB}}$	fixed	high
robot base to robot instrument $\boldsymbol{T}_{\mathrm{RB}}^{\mathrm{RI}}$	kine. + model	high
robot base to marker $\boldsymbol{T}_{\mathrm{RB}}^{M_1}$	kine. + model + pivot calib.	high
marker to the robotic instrument $\boldsymbol{T}_{M_1}^{\mathrm{RI}}$	kine. + model + pivot calib.	high
marker to a hand-held instrument $\boldsymbol{T}_{M_2}^{\mathrm{HI}}$	pivot calib.	high
marker to HMD $\boldsymbol{T}_{M_{\{1,2\}}}^{\mathrm{H}}$	fiducial tracking	medium
world to HMD $\boldsymbol{T}_{\mathrm{W}}^{\mathrm{H}}$	SLAM	low

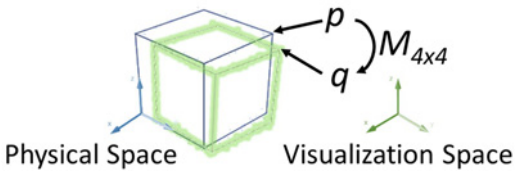

Fig. 3 *Illustration of display calibration in ARssist*

coordinate system of the HMD $\{H\}$ and is aligned with a point q in the visualisation space. The visualisation space is usually defined in the rendering engine. The display calibration method aims to find a linear mapping between the physical space and the visualisation space (Fig. 3):

$$\hat{q} = \boldsymbol{M}_{4\times4} \cdot \hat{p}, \tag{4}$$

where $\hat{q}$ represents a 3D point q in homogeneous coordinates and $\boldsymbol{M}$ is a 4×4 matrix. A calibration procedure needs to be conducted before the use of *ARssist*, where the user manually collects a few corresponding point pairs $\{(p_i, q_i)\}$. The calibration matrix $\boldsymbol{M}$ is then solved using direct linear transformation (DLT) [26]. At runtime, the calibration matrix $\boldsymbol{M}$ is able to offset any real world vertex p defined in $\{H\}$, so that its adjusted 3D location q used for rendering will appear aligned with the physical point.

To evaluate the accuracy of the overlay, we can determine the average planar re-projection error $\bar{e}$ by projecting the residue of the DLT algorithm $\boldsymbol{r}$ to the XY plane, with the assumption that the camera principle axis is $\boldsymbol{z}$, and n denotes the total number of point pairs

$$\boldsymbol{r}_i = (r_{i,x}, r_{i,y}, r_{i,z}) = \hat{q}_i - \boldsymbol{M}_{4\times4} \cdot \hat{p}_i, \tag{5}$$

$$e_i = \|(r_{i,x}, r_{i,y})\|, \tag{6}$$

$$\bar{e} = \frac{1}{n}\sum e_i. \tag{7}$$

For evaluation purposes, we rigidly attach a pair of eye-simulating cameras behind the OST-HMD to 'see' through the OST-HMD. In this way, the planar re-projection error can be objectively obtained after the display calibration.

3.3. Visualisation of stereo endoscopy: The endoscopy serves as the primary feedback both for the console surgeon and for the FA, in a robotic laparoscopic surgery. The console surgeon has an immersive stereoscopic view of the surgery field as if he/she were tele-ported into the patient body. The hands of the surgeon and the instruments are intuitively coupled. However, for the FA, the endoscopy is displayed on an external monitor mounted on the vision cart or the ceiling and is single channel even if dual channels are available. Therefore, the FA does not have depth perception of the endoscopy and hand–eye coordination is hindered because the FA has to constantly switch the view between the monitor and the patient.

A binocular OST-HMD can present the left and right endoscope channel to the left and right eye, respectively, thereby restoring the depth perception of the endoscopy to some extent. To guarantee the synchronisation of the left and right channels, *ARssist* concatenates both channels in the robot vision system and streams the combined image to the HMD, which utilises a stereo shader for rendering.

ARssist offers three options that can potentially mitigate the hand–eye coordination and ergonomical issue: (i) heads-up display, (ii) stereo *virtual monitor*, and (iii) endoscopy registered with the endoscope frustum.

3.3.1 Heads-up display: When the HMD is used as a heads-up display for the stereo endoscopy, the endoscopy window occupies a large portion of the screen. The visualisation of the endoscopy does not rely on the external environment, e.g., the location or orientation of the HMD. The benefit of heads-up display visualisation is that it enables better visual acuity of the endoscopy since more pixels are dedicated to the rendering. The drawback of this technique is that it greatly reduces the see-through ability of the user. The heads-up display visualisation may find the best usage when the FA does not need to pay significant attention to the external condition of the surgery site.

3.3.2 Virtual stereo monitor: With real-time localisation, the HMD is able to place an object that appears fixed to the world coordinate system. *ARssist* can display the endoscopy on a "virtual stereo monitor' [27]. The virtual monitor is more flexible than the traditional monitor in that the scale and pose of it can be adjusted. The FA can "move' the virtual monitor to an arbitrary place that he/she feels the most comfortable with. For example, the virtual monitor can be placed on top of the trocar, so that the FA can see both the endoscopy and the external condition of the patient without turning his/her head. The virtual monitor resembles the traditional setup, but brings more flexibility to the placement of the endoscopy monitor and enables stereo display. An example of a virtual stereo monitor is shown in Fig. 4*e*.

3.3.3 Endoscopy registered with the endoscope frustum: We propose a novel visualisation technique that renders the endoscopy at the end of the endoscope frustum, which is able to inform the FA not only about the endoscopy video itself, but also the geometry of the endoscope. Since the endoscope is also held by a robotic arm, *ARssist* obtains the kinematics of the endoscopic arm and calculates the pose of the endoscope at runtime. With a standard camera calibration of the endoscope, we calculate the horizontal and vertical

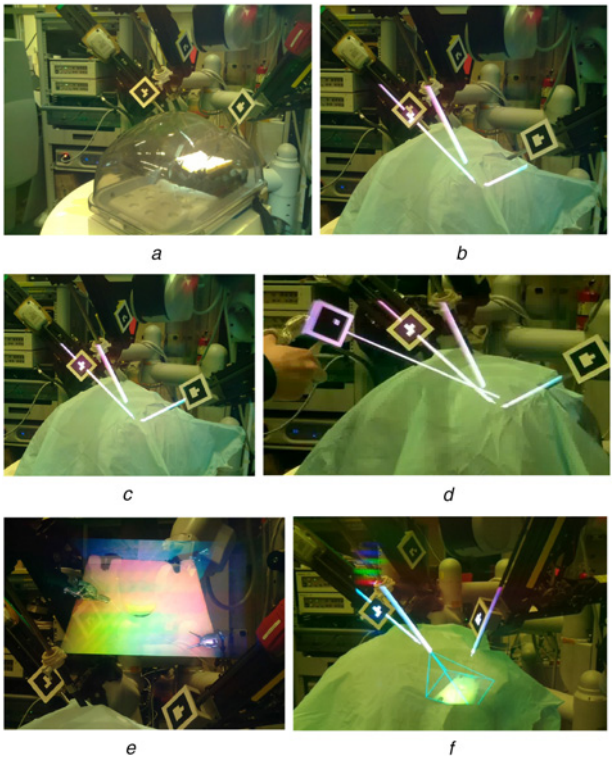

Fig. 4 *Visualisation results of ARssist*
a Transparent body phantom
b Before display calibration
c With display calibration
d Overlay with a hand-held instrument
e Virtual monitor visualisation of the endoscopy
f Endoscopy visualisation registered with viewing frustum

doi: 10.1049/htl.2018.5065

field-of-view (FOV) of the endoscope. Combiining the pose and FOV, *ARssist* renders a frustum extending the tip of the endoscope and projects the endoscopy on a clipping plane of the frustum. The visualisation result is shown in Fig. 4*f* in Section 5. In this way, the disorientation issue of traditional laparoscopic surgery [10] is solved because the endoscopy is displayed in the correct orientation with respect to the world coordinate system.

4. System

4.1. Implementation of ARssist: We chose the da Vinci Research kit (dVRK) [28] as the robotic platform and Microsoft HoloLens as the HMD. We note, however, that *ARssist* could be implemented on a clinical da Vinci system, using the available research interface [29] to obtain the kinematics data.

4.1.1 da Vinci research kit: dVRK is an open-source surgical robotics research kit based on the first-generation *da Vinci*® system and robot operating system (ROS) [30]. In our configuration, it provides runtime access to the kinematics data at 200 Hz, and both channels of the endoscopy at 30 Hz, with a resolution of 1920 × 1080 pixels. Two custom ROS nodes (*udp_relay* and *arssist_streamer*) are implemented to separately process the kinematics data stream and endoscope video stream. The first ROS node (*udp_relay*) receives the joint state message from ROS topics, filters unnecessary data, serialises in JSON and sends the packet to the HoloLens through user datagram protocol (UDP) protocol. UDP is chosen because it has low latency and intermittent packet loss is not a concern for high-rate kinematics data. The second ROS node (*arssist_streamer*) fetches the two channels of endoscopy, downscales the original images (to 640 × 480), concatenates both channels, and streams it to the HoloLens through Motion-JPEG protocol. A desktop computer, with Ubuntu 16.04 operating system, Xeon(R) E5-1620 CPU and 28.8 G RAM, hosts the programs for dVRK and our custom ROS nodes. The dVRK setup is shown in Fig. 5.

4.1.2 Microsoft HoloLens: Microsoft HoloLens is a binocular OST-HMD featuring a holographic waveguide-based optical system, stable self-localisation capability, sufficient computational power for tracking, and good support from development tools [7]. A unity application runs on the HoloLens as part of *ARssist*. Fiducial marker tracking is implemented based on ARToolKit [31, 32]. The front-facing camera of HoloLens is configured for a resolution of 1344 × 768, 67° FOV [33] and 15 fps. The application communicates with *udp_relay* and *arssist_streamer*. We use the robot model of [34] for transformations between robot joints. A special shader is implemented to handle the stereo rendering of the endoscopy. The rendering, socket communication, and tracking are handled with different threads on HoloLens.

The eye-simulating cameras described in Section 3.3 are SaintSmart cameras with 1080*P* video resolution, driven by Raspberry Pi Model 3B. They are rigidly attached to the HoloLens and are separated by 64 mm, which is a typical human interpupillary distance, as shown in Fig. 6. The linkage between the cameras is 3D printed.

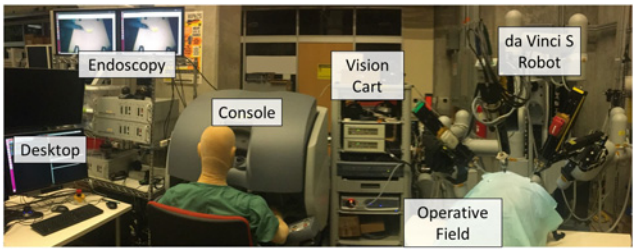

Fig. 5 *dVRK setup*

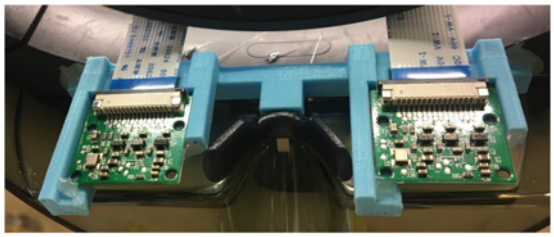

Fig. 6 *Setup of eye-simulating cameras for obtaining visualisation results (Fig. 4) of ARssist*

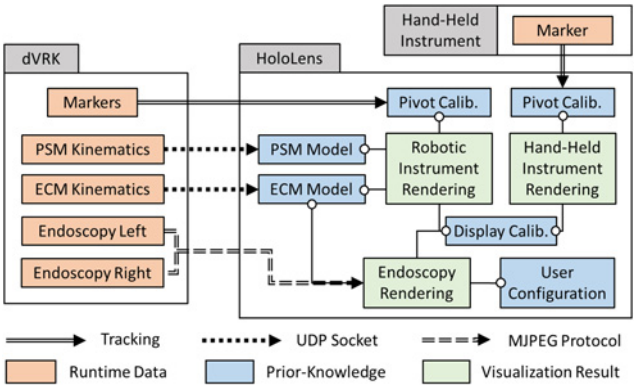

Fig. 7 *Data flow in ARssist*

4.1.3 Data flow: As a summary, the data flow in *ARssist* is illustrated in Fig. 7. Orange boxes show the data that are obtained at runtime and are updated frequently. Blue boxes identify the data that are known prior to an instance of the application. Calibration data and the robot model (e.g., Denavit–Hartenberg (DH) parameters, meshes) are considered prior knowledge. Green boxes are the visualisation results and are the destinations of the data flow.

4.2. Calibration of the system

4.2.1 Pivot calibration: We rigidly attach the fiducial markers to the robotic arms and hand-held instrument as shown in Fig. 8. The linkages are 3D printed. For the markers on the robotic arms, their positions in the robot joint hierarchy are determined and their relative transformations to the nearest joints are obtained by pivot calibration. For the marker on the hand-held instrument, the relative transformation of the marker with respect to the instrument coordinate system is also calculated using a pivot calibration.

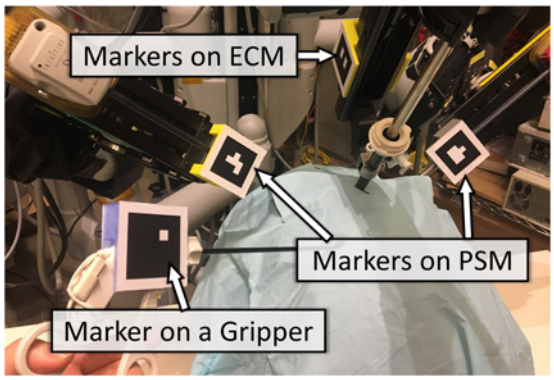

Fig. 8 *Fiducial markers on robotic arms and hand-held instrument*

4.2.2 Display calibration: As described in Secttion 3.3, we conduct a display calibration of the HoloLens, with 20 alignment points distributed within the HoloLens viewing frustum. The user holds a custom fiducial marker to align the centre crosshair with the alignment points on the HoloLens. When the user confirms alignment, the system records the corresponding p_i. After 20 alignments, we use the DLT method to solve for the linear mapping $M(\cdot):p \rightarrow q$.

4.2.3 Camera calibration: We use the standard OpenCV camera calibration algorithm to calibrate the front-facing camera of HoloLens and the endoscope camera with checkerboards. To visualise the frustum described in Section 3.4.3, we compute the vertical and horizontal FOV of the endoscope camera by

$$\begin{aligned} \mathrm{FOV}_v &= 2 \cdot \arctan\left(\frac{h}{2f_y}\right), \\ \mathrm{FOV}_h &= 2 \cdot \arctan\left(\frac{w}{2f_x}\right), \end{aligned} \tag{8}$$

where w and h are the pixel width and height of the endoscope image and f_x and f_y are the focal lengths in pixel units obtained from the camera intrinsic matrix. Ideally, the left and right channels of the endoscope camera should be identical. However, manufacturing imperfection and measurement error always exist. We take the average FOV of both channels.

5. Results

5.1. Tracking and overlay accuracy: We finalise our hybrid tracking scheme by determining the priority ranking of multiple transformation equations, as described in Section 3.2. Equation (1) is composed of one high-priority transformation and one medium-priority transformation, while (2) has a profile of two high- and one low-priority transformation. Therefore, we prefer (1) to (2) at runtime. When (1) is not available, we use (2) as a substitute, but with an additional offset that represents the error between the two tracking methods. In other words, we prefer the fiducial tracking over the self-localisation of the HMD based on the current generation of HMD technologies.

To objectively obtain the overlay accuracy, we conduct the display calibration with respect to the eye-simulating cameras, and by applying (5)–(7), the average 2D overlay error is calculated to be 4.27 mm with a standard deviation of 3.09 mm.

5.2. Visualisation results:

(i) Transparent body phantom
(ii) Before display calibration
(iii) With display calibration
(iv) Overlay with a hand-held instrument
(v) Virtual monitor visualisation of the endoscopy
(vi) Endoscopy visualisation registered with the viewing frustum

Fig. 4 shows the visualisation results captured by the eye-simulating cameras. Fig. 4*a* shows the setup for this set of images: a transparent body model is used; two patient side manipulators and one endoscopic camera manipulator (ECM) are inserted into the model. The fiducial markers are attached to the robotic arms. Figs. 4*b* and *c* present the overlay without and with the display calibration. The calibration is able to significantly improve the overlay accuracy of the optical see-through system. Note that the overlay of the instruments is cut off due to the limited field-of-view for augmentation on HoloLens. Fig. 4*d* shows the overlay for a hand-held instrument.

Figs. 4*e* and *f* demonstrate the different configurations for visualisation of the stereo endoscopy. In Fig. 4*e*, the endoscopy is displayed on a ‘‘virtual monitor’ (Section 3.4.2) that is floating on top of the surgical field. In this way, the FA is able to see both the surgical field and the endoscopy with much less effort in terms of head rotation. Fig. 4*f* depicts the result of the visualisation technique described in Section 3.4.3. The viewing frustum of the endoscope is rendered at the tip of the ECM. The vertical and horizontal field-of-view of the endoscope are 49.18° and 61.27°, respectively. For simplicity, the distortion parameters of the endoscope are not taken into consideration. With the rendering of the instruments and endoscope inside the body, the FA is able to navigate instruments into the FOV of the endoscope intuitively, even in the case where the robot and arms are not docked in a convenient configuration.

5.3. Real-time performance: Real-time performance is critical for augmented and virtual reality applications. We measure the frame rate for rendering, tracking, and endoscopy, which are 32.91 ± 1.96, 13.64 ± 0.78, and 26.57 ± 3.10 Hz, respectively. The end-to-end latency of stereo endoscopy streaming is 220.81 ± 25.54 ms, with a downscaled image of 640×480 pixel resolution.

6. Discussion

6.1. Clinical benefits: *ARssist* has the potential to help the FA by (i) providing a direct view of the pose of the instruments inside the patient body and (ii) providing a configurable visualisation of the stereo endoscopy.

The direct view of the instruments and endoscope frustum inside the body helps the FA to understand the geometric relationship between these instruments. When the FA needs to pass a new instrument to the surgeon, he/she is able to look at these virtual tools during the manipulation, rather than trying to manoeuver blindly or having the surgeon redirect the endoscope toward the trocar. It is both safe and time-efficient. The manipulation of hand-held instruments, e.g. vessel sealer, gripper, containers for the specimen, is also facilitated by *ARssist* in this way.

The rendering of stereo endoscopy has the potential to alleviate the current awkward hand–eye coordination for the FA. For a FA who prefers the traditional setup where a 2D monitor is used for displaying endoscopy, *ARssist* enables flexible positioning of this ‘virtual monitor’ so that the FA is able to reposition it for the most comfortable viewing experience. The visualisation as a heads-up display offers the highest visual acuity and should be useful in cases that require the perception of details. Visualisation at the end of the frustum is a novel way to handle endoscopy. In addition, the stereoscopic view is available in all visualisation methods to enhance depth perception.

6.2. Hybrid tracking scheme in ARssist: Multiple sources of error exist in the pipeline of *ARssist*. The cable-driven da Vinci robot can have kinematic inaccuracy due to the changes in cable tension. The error in fiducial tracking results from the inaccuracy of camera calibration, changes under lighting conditions, and the intrinsic limitation of the number of pixels available. The self-localisation of HoloLens has an observable drift over time, which will cause an accumulation of error. In addition, rendering of the overlay is delayed due to the latency in the system pipeline, which causes inaccuracy when there is relative motion.

It is desirable to include as little error as possible for the visualisation. Therefore, we utilise the hybrid tracking scheme described in Section 3.2. We consider the kinematics data the most reliable, and fiducial tracking is still more reliable than the self-localisation of the HMD, based on current technology availability.

6.3. Limitations and future work: To evaluate the potential benefits of *ARssist* for the FA, we will first conduct a user study with experienced and novice FAs in realistic experimental settings. It is also valuable to study how *ARssist* would affect the collaboration between the FA and the main surgeon. In our

doi: 10.1049/htl.2018.5065

current implementation, the overlay accuracy (4.27 ± 3.09 mm) is still far from satisfactory. The accuracy is limited by fiducial tracking and imperfect display calibrations. One possible solution is to add an external, highly-accurate tracking module, e.g. an infrared-based tracker, into the hybrid tracking scheme and prioritise the use of the external tracker. Other display calibration methods that model the eye–HMD pair in more sophisticated ways can be used [35].

The depth information of the endoscopy is presented to the FA by rendering the left and right channels separately on the HMD. However, the pose of the endoscope is not exactly aligned with the human's eye, which can lead to a disassociation issue in the FA's perception [14]. The issue can be resolved by reconstructing a 3D model of the surgical scene, projecting the endoscopy onto this model, and then rendering it from the current viewpoint of the FA, as demonstrated by Edgcumbe *et al.* [36] for an ex-vivo kidney and in our recent work for satellite servicing [37]. Both involved a mostly static scene, leaving real-time 3D reconstruction of the changing anatomy as a future challenge. Methods based on 2D image processing can also alleviate the issue [13, 14].

In addition, without the ability to tune the transparency of the OST-HMD, the endoscopy appears floating on top of the background. This issue is also known as an occlusion leak, where the transparency of the rendered graphics does not appear natural for the viewer. Recently, researchers have proposed a solution to tackle this issue with an additional spatial light modulator [38].

The current on-board computational power of the OST-HMD is also a limiting factor for the performance of our system. We downscale the video frame for endoscopy as a trade-off between the latency caused by the network and decoding, and the viewing experience. We believe the downscaling will be unnecessary in the future as the HMD hardware keeps advancing.

Our future work includes a rigorous evaluation of the error distribution of each individual transformation, which can enable fusion of multiple sensor inputs, especially in situations with comparable priorities (e.g., multiple markers on the robotic arms).

7. Conclusion: Robot-assisted laparoscopic surgery is a team effort. The FA who serves at the bedside is also critical for the outcome of the surgery. In this Letter, we present an application, *ARssist*, which can benefit the FA by offering AR visualisation based on an OST-HMD.

First, *ARssist* presents the 3D rendering of robotic instruments, hand-held instruments, and the endoscope. Our hybrid tracking scheme takes advantage of the redundancy of various sensors, including robotic kinematics, fiducial marker tracking, and self-localisation of the HMD, to establish the geometric relationship between the objects to render and the HMD. A display calibration procedure is conducted to determine a linear mapping between the physical space and the visualisation space. The visualisation of instruments within the patient body can provide the FA with an intuitive interface to understand the geometric relationship between the components and to navigate an instrument to the desired location.

The second component of *ARssist* is the configurable rendering of stereo endoscopy. Currently, in a da Vinci robot-assisted surgery, the FA only has access to a single channel of the endoscopy, displayed on a cart-mounted or ceiling-mounted monitor, which hinders hand–eye coordination when the FA is operating. In *ARssist*, we offer several choices for the rendering of stereo endoscopy: (i) as a heads-up display, (ii) as a virtual stereo monitor or (iii) registered with the endoscope frustum. The FA can choose the rendering that enables the task to be performed more efficiently and comfortably.

ARssist has the potential to help FAs perform their tasks more efficiently and hence improve the outcome of robotic-assisted laparoscopic surgeries.

8. Funding and declaration of interest: This work was supported by a Technology Research Grant from Intuitive Surgical Operations, Inc. Mr. Qian and Dr. Kazanzides have a patent Augmented Reality Display for Minimally Invasive Surgery pending.

9. Conflict of interest: None declared.

10 References

[1] Kumar R., Hemal A.K.: 'The 'scrubbed surgeon' in robotic surgery', *World J. Urol.*, 2006, **24**, (2), pp. 144–147
[2] Martin S.: 'The role of the first assistant in robotic assisted surgery', *Br. J. Perioper. Nurs.*, 2004, **14**, (4), pp. 159–163
[3] Sgarbura O., Vasilescu C.: 'The decisive role of the patient-side surgeon in robotic surgery', *Surg. Endosc.*, 2010, **24**, (12), pp. 3149–3155
[4] Nayyar R., Yadav S., Singh P., *ET AL.*: 'Impact of assistant surgeon on outcomes in robotic surgery', *Indian J. Urol.*, 2016, **32**, (3), p. 204
[5] Sung G.T., Gill I.S.: 'Robotic laparoscopic surgery: a comparison of the da Vinci and Zeus systems', *Urology*, 2001, **58**, (6), pp. 893–898
[6] Rolland J.P., Fuchs H.: 'Optical versus video see-through head-mounted displays in medical visualization', *Presence, Teleoperators Virtual Environ.*, 2000, **9**, (3), pp. 287–309
[7] Qian L., Barthel A., Johnson A., *ET AL.*: 'Comparison of optical see-through head-mounted displays for surgical interventions with object-anchored 2D-display', *Int. J. Comput. Assisted Radiol. Surg.*, 2017, **12**, (6), pp. 901–910
[8] Chen L., Day T.W., Tang W., *ET AL.*: 'Recent developments and future challenges in medical mixed reality'. IEEE Intl. Symp. on Mixed and Augmented Reality (ISMAR), Nantes, France, October 2017, pp. 123–135
[9] Bernhardt S., Nicolau S.A., Soler L., *ET AL.*: 'The status of augmented reality in laparoscopic surgery as of 2016', *Med. Image Anal.*, 2017, **37**, pp. 66–90
[10] Wentink B.: 'Eye-hand coordination in laparoscopy – an overview of experiments and supporting aids', *Minim Invasive Ther. Allied Technol.*, 2001, **10**, (3), pp. 155–162
[11] Lo B., Chung A.J., Stoyanov D., *ET AL.*: 'Real-time intraoperative 3D tissue deformation recovery'. IEEE Intl. Symp. on Biomedical Imaging: From Nano to Macro (ISBI), Paris, France, May 2008, pp. 1387–1390
[12] Maier-Hein L., Mountney P., Bartoli A., *ET AL.*: 'Optical techniques for 3D surface reconstruction in computer-assisted laparoscopic surgery', *Med. Image Anal.*, 2013, **17**, (8), pp. 974–996
[13] Koreeda Y., Obata S., Nishio Y., *ET AL.*: 'Development and testing of an endoscopic pseudo-viewpoint alternating system', *Int. J. Comput. Assisted Radiol. Surg.*, 2015, **10**, (5), pp. 619–628
[14] Koppel D., Wang Y.-F., Lee H.: 'Image-based rendering and modeling in videoendoscopy'. IEEE Intl. Symp. on Biomedical Imaging: Nano to Macro, Arlington, VA, USA, April 2004, pp. 269–272
[15] Buchs N.C., Volonte F., Pugin F., *ET AL.*: 'Augmented environments for the targeting of hepatic lesions during image-guided robotic liver surgery', *J. Surg. Res.*, 2013, **184**, (2), pp. 825–831
[16] 'Microsoft hololens', Available at https://www.microsoft.com/en-us/hololens, accessed: 6 June 2018
[17] 'Meta', Available at http://www.metavision.com/, accessed: 6 June 2018
[18] Thrun S., Leonard J.J.: 'Simultaneous localization and mapping', in Siciliano B., Khatib O. (Eds.): 'Springer handbook of robotics' (Springer, Berlin & Heidelberg, 2008), pp. 871–889
[19] Wang J., Qian L., Azimi E., *ET AL.*: 'Prioritization and static error compensation for multi-camera collaborative tracking in augmented reality'. IEEE Virtual Reality (VR), Los Angeles, CA, USA, March 2017, pp. 335–336
[20] Kwartowitz D.M., Herrell S.D., Galloway R.L.: 'Toward image-guided robotic surgery: determining intrinsic accuracy of the da Vinci robot', *Int. J. Comput. Assist. Radiol. Surg.*, 2006, **1**, (3), pp. 157–165
[21] Abawi D.F., Bienwald J., Dorner R.: 'Accuracy in optical tracking with fiducial markers: an accuracy function for ARToolKit'. Proc. of the 3rd IEEE/ACM Int. Symp. on Mixed and Augmented Reality, Arlington, VA, USA, November 2004, pp. 260–261
[22] Tuceryan M., Genc Y., Navab N.: 'Single-point active alignment method (SPAAM) for optical see-through HMD calibration for augmented reality', *Presence, Teleoperators Virtual Environ.*, 2002, **11**, (3), pp. 259–276

[23] Itoh Y., Klinker G.: 'Interaction-free calibration for optical see-through head-mounted displays based on 3d eye localization'. IEEE Symp. on 3D User Interfaces (3DUI), Minneapolis, MN, USA, March 2014, pp. 75–82
[24] Owen C.B., Zhou J., Tang A., *ET AL.*: 'Display-relative calibration for optical see-through head-mounted displays'. IEEE/ACM Intl. Symp. on Mixed and Augmented Reality (ISMAR), Arlington, VA, USA, November 2004, pp. 70–78
[25] Qian L., Azimi E., Kazanzides P., *ET AL.*: 'Comprehensive tracker based display calibration for holographic optical see-through head-mounted display', 2017, arXiv:1703.05834
[26] Hartley R., Zisserman A.: 'Multiple view geometry in computer vision' (Cambridge University Press, New York, NY, USA, 2003)
[27] Qian L., Unberath M., Yu K., *ET AL.*: 'Towards virtual monitors for image guided interventions-real-time streaming to optical see-through head-mounted displays', 2017, arXiv:1710.00808
[28] Kazanzides P., Chen Z., Deguet A., *ET AL.*: 'An open-source research kit for the da Vinci R surgical system'. IEEE Intl. Conf. on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014, pp. 6434–6439
[29] DiMaio S., Hasser C.: 'The da Vinci research interface'. MICCAI Workshop on Systems and Arch. for Computer Assisted Interventions, Midas Journal, 2008, Available at http://hdl.handle.net/10380/1464
[30] Quigley M., Conley K., Gerkey B., *ET AL.*: 'ROS: an open-source robot operating system'. ICRA Workshop on Open Source Software, Kobe, Japan, 2009
[31] Kato H., Billinghurst M.: 'Marker tracking and HMD calibration for a video-based augmented reality conferencing system'. IEEE/ACM Intl. Workshop on Augmented Reality (IWAR), San Francisco, CA, USA, October 1999, pp. 85–94
[32] 'Hololensartoolkit'. Available at https://github.com/qian256/HoloLensARToolKit, accessed: 6 June 2018
[33] 'Locatable camera'. Available at https://docs.microsoft.com/en-us/windows/mixed-reality/locatable-camera, accessed: 6 June 2018
[34] Fontanelli G., Ficuciello F., Villani L., *ET AL.*: 'Da Vinci research kit: PSM and MTM dynamic modelling'. IROS Workshop on Shared Platforms for Medical Robotics Research, Vancouver, Canada, 2017
[35] Azimi E., Qian L., Kazanzides P., *ET AL.*: 'Robust optical see-through head-mounted display calibration: taking anisotropic nature of user interaction errors into account'. IEEE Virtual Reality (VR 2017), Los Angeles, CA, USA, March 2017, pp. 219–220
[36] Edgcumbe P., Pratt P., Yang G.-Z., *ET AL.*: 'Pico Lantern: surface reconstruction and augmented reality in laparoscopic surgery using a pick-up laser projector', *Med. Image Anal.*, 2015, **25**, (1), pp. 95–102
[37] Vagvolgyi B., Niu W., Chen Z., *ET AL.*: 'Augmented virtuality for model-based teleoperation'. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), Vancouver, Canada, 2017, pp. 3826–3833
[38] Itoh Y., Hamasaki T., Sugimoto M.: 'Occlusion leak compensation for optical see-through displays using a single-layer transmissive spatial light modulator', *IEEE Trans. Visual. Comput. Graphics*, 2017, **23**, (11), pp. 2463–2473