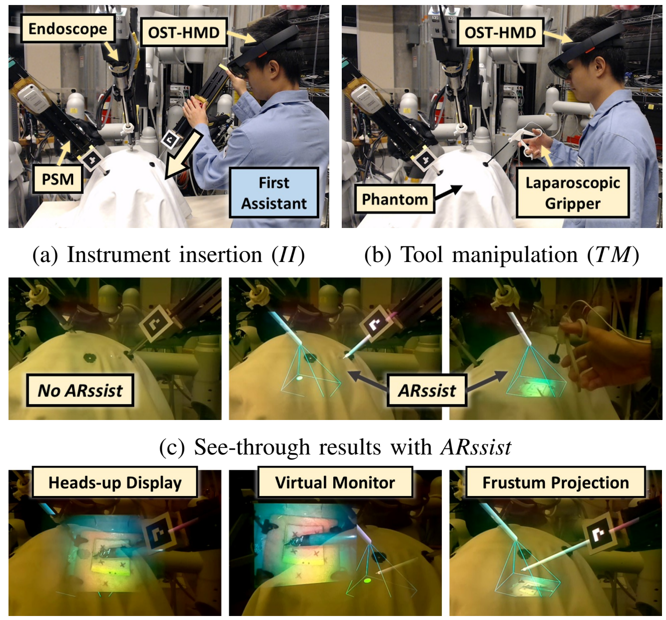
2019 International Conference on Robotics and Automation (ICRA)
Palais des congres de Montreal, Montreal, Canada, May 20-24, 2019

Augmented Reality Assisted Instrument Insertion and Tool Manipulation for the First Assistant in Robotic Surgery

Long Qian[1], Anton Deguet[1], Zerui Wang[2], Yun-Hui Liu[2] and Peter Kazanzides[1]

***Abstract*— In robotic-assisted laparoscopic surgery, the first assistant (FA) stands at the bedside assisting the intervention, while the surgeon sits at the console teleoperating the robot. Tasks for the FA include navigating new instruments into the surgeon's field-of-view and passing in or retracting materials from the body using hand-held tools. We previously developed *ARssist*, an augmented reality application based on an optical see-through head-mounted display, to aid the FA. In this paper, we refine the system and first perform a pilot study with three experienced surgeons for two specific tasks: instrument insertion and tool manipulation. The results suggest that *ARssist* would be especially useful for less experienced assistants and for difficult hand-eye configurations. We then perform a multi-user study with inexperienced subjects. The results show that *ARssist* can reduce navigation time by** 34.57%**, enhance insertion path consistency by** 41.74%**, reduce root-mean-square path deviation by** 40.04%**, and reduce tool manipulation time by** 72.25%**. Thus, *ARssist* has the potential to improve efficiency, safety and hand-eye coordination, especially for novice assistants.**



(a) Instrument insertion (*II*) (b) Tool manipulation (*TM*)

(c) See-through results with *ARssist*

(d) Different visualization methods of endoscopy in *ARssist*

Fig. 1: Instrument insertion and tool manipulation w/ *ARssist*

I. Introduction

Robotic-assisted laparoscopic surgery has various advantages compared to conventional laparoscopic surgery, including the restoration of hand-eye coordination, improved dexterity of instruments, and 3D immersive visualization [1]. As the most successful commercial surgical robot product, the *da Vinci*® [2] system has been used for, and demonstrated good results in, prostatectomies [3], gynecologic procedures [4], vascular surgeries [5] and others. In a *da Vinci* surgery, the surgeon sits at the teleoperation console and a first assistant (FA) is required at the patient side to help perform some tasks such as to change the robotic instrument or pass additional materials into the surgical site.

Researchers have identified that the performance of the FA is critical to the overall outcome of the surgery [6], [7], [8]. However, the environment, workflow and facilities for the FA to perform his/her tasks have not significantly changed since the introduction of surgical robots. In our prior work, we described *ARssist*, an augmented reality application integrated with a surgical robot, which aims to improve the perception and performance of the FA [9]. *ARssist* provides visualization of the robotic instruments and endoscope "inside" the patient body, and various ways to render the stereo endoscopy on the head-mounted display.

In this work, we refine the implementation of *ARssist* and evaluate the FA's performance for two frequent tasks during interventions: instrument insertion (*II*) and tool manipulation (*TM*)[†]. We first conduct a pilot study and interview with 3 surgeons who have worked as a FA. Although their performance does not show statistical significance, they are positive about the experience and future integration with *ARssist*. Based on their feedback, we designed and conducted a multi-user study to compare user performance in the two tasks with and without *ARssist*. Based on a quantitative analysis of the performance of 20 inexperienced users, we find that *ARssist* is able to significantly reduce the danger of potential collision during instrument insertion and significantly improve hand-eye coordination during tool manipulation.

II. Background and Related Work

A. Augmented Reality for Robotic-Assisted Surgery

Augmented Reality (AR) has been used in many image-guided surgeries, including robotic-assisted laparoscopic surgeries [10]. Volonte et al. integrated 3D volume-rendered

* This work is supported by a 2018 Technology Grant from Intuitive Surgical Operations, Inc, and in part by the HK RGC under T42-409/18-R.
[1] Long Qian, Anton Deguet and Peter Kazanzides are with the Department of Computer Science, Johns Hopkins University, Baltimore, Maryland, United States. Email: {long.qian, anton.deguet, pkaz}@jhu.edu
[2] Zerui Wang and Yun-Hui Liu are with the T Stone Robotics Institute and Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong, China. Email: {zrwang,yhliu}@mae.cuhk.edu.hk

[†] The words "instrument" and "tool" are interchangeable. We specifically use "tool" to refer to the hand-held laparoscopic instrument, thereby emphasizing the distinction between the two tasks.

images onto the da Vinci surgeon console, displaying the pre-operative DICOM images alongside the endoscopy [11], [12]. The system increased surgeon comfort and security in right colectomy. In [13], the 3D virtual anatomical model is manually overlaid on the endoscopic video for the console surgeon. The same setup was also used for robotic duodenopancreatectomy [14]. Su et al. studied the use of AR in robotic partial nephrectomy by registering 3D-CT with stereoscopic video [15]. Mohareri et al. integrated the da Vinci robot with a transrectal ultrasound (TRUS) robot to create intraoperative AR with a 3D ultrasound image overlaid on endoscopy [16]. The literature in AR and robotic-assisted surgery primarily focuses on providing improved visualization for the console surgeon. To the best of our knowledge, our work of *ARssist* is the only work that aims to provide an AR interface for the FA.

B. Instrument Insertion (II) and Tool Manipulation (TM)

The involvement of the FA can be categorized based on the three phases of the surgery. Before the surgery, the FA usually works with the surgeon for trocar placement and docking of the robot [17]. During the surgeon's operation, the FA is responsible for instrument exchange, manipulating laparoscopic tools depending on the specific procedure, and other tasks [17]. After the operation, the FA removes the instruments, undocks the robot, and performs port closure [18].

Instrument insertion (II) and tool manipulation (TM) are relatively frequent and repetitive tasks for the FA. II occurs after the robot is docked, and when a robotic instrument needs to be changed due to a procedure requirement. TM refers to the maneuver of normal laparoscopic tools. For example, when the surgeon performs suturing, the FA must hold the suture with a laparoscopic gripper and pass the suture to the robotic instrument controlled by the surgeon. TM is also required for retraction, suction, and specimen extraction [6]. In the current surgical workflow, the endoscopy is displayed on a stationary monitor that is located far from the operation site and provides the only feedback to the FA. In II, the FA has to blindly navigate new instruments into the endoscopy, or ask the surgeon to drive the endoscope to look at the inserted instrument to ensure safety. For TM, the FA also watches the external monitor for guidance and, depending on the location and orientation of the endoscope, the operation may create an awkward situation for the FA's hand-eye coordination [19].

III. Overview Of *ARssist*

In this section, we present an overview of the *ARssist* application [9]. *ARssist* is implemented on Microsoft HoloLens, an optical see-through head-mounted display (OST-HMD), and requires access to the da Vinci kinematic data and endoscopic video. In our current implementation, these are obtained via the open source da Vinci Research Kit (dVRK) [20]. Through *ARssist*, the FA is able to view:

- Real-time 3D robotic instruments and endoscope with its field-of-view indicator inside the patient body, visually aligned with the real ones (Fig. 1c)
- Real-time stereo endoscopic video displayed in three different modes: heads-up display, virtual monitor, or projected in the viewing frustum of endoscope (Fig. 1d).

A. Visualization of Robotic Instruments and Endoscope

In order to visualize robot-controlled instruments on the OST-HMD, it is required to obtain the real-time pose of each instrument relative to the HMD. This can be achieved via marker-based or markerless tracking [22], or via the self-localization of the HoloLens. *ARssist* takes advantage of the redundancy and uses a prioritization-based fusion algorithm to obtain the most reliable result [23]. For the endoscope, the viewing frustum is also visualized, extending the tip of the endoscope (Fig. 1c). Lastly, we apply the display calibration of the OST-HMD in order to align the virtual content with the human's perception [24]. The average error of the overlay was determined to be $4.27mm$ [9].

B. Visualization of Stereo Endoscopic Video

The stereo endoscopic video is obtained from the video pipeline of the robot, and then streamed to the OST-HMD for rendering. On the OST-HMD, we developed three methods to display the video (Fig. 1d):

- Heads-up display: The video is displayed in the "screen coordinates". The video stays in the same area of the screen regardless of the user's head rotation.
- Virtual monitor: The video is displayed on a virtual 3D plane that is stationary in the world coordinate system [25]. It appears like a "virtual stereo monitor".
- Frustum projection: The video is displayed in the viewing frustum of the endoscope.

In all three methods, the left and right eye see the left and right channel of the stereo endoscopy, respectively.

C. Refinement of Implementation

We made several refinements to the implementation described in [9]. First, prior to both the pilot run and user study, we added a voice-based user interface to control the behavior of the application. This allows the assistant to select the desired endoscopy display by saying "*heads-up*","*virtual monitor*", or "*frustum*". The word "*next*" can iterate over the three methods, in case the assistant does not remember the specific command. When the endoscopy is displayed as a virtual monitor, the additional voice commands "*move back*", "*move forward*", "*larger*", and "*smaller*" can adjust the position and scale of the virtual monitor.

Based on the pilot run experience, we made two additional refinements prior to the user study. One change was to support a 30°-angled endoscope in addition to the straight endoscope (Fig. 2).

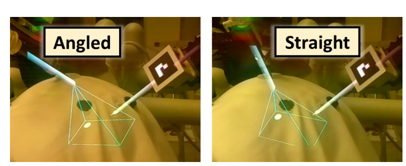
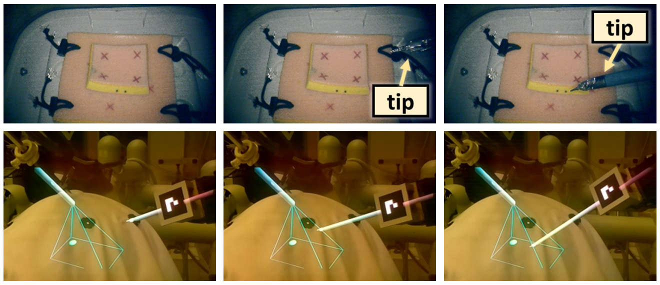


Fig. 2: 30°-angled and straight endoscope for *ARssist*



(a) Before insertion (b) During insertion (c) Finish insertion

Fig. 3: II_{AR}: instrument insertion with the help of *ARssist*

In addition, we removed the tracking and overlay of hand-held tools, which simplified the system. Although *ARssist* is capable of tracking the hand-held instrument if a fiducial marker is attached to it, during the pilot run (Sect. V) we discovered that the tracking often failed due to the limited tracking field-of-view (FOV) of the HoloLens front-facing camera. The tracking failure caused the overlay of the hand-held tool to disappear, which was distracting for the user. This issue could potentially be solved by using an external tracking system, but that would complicate the system setup. Also, placement of fiducial markers on a hand-held instrument reduces its ergonomics.

IV. Experiment Setup

We evaluate instrument insertion (II) and tool manipulation (TM), both with (AR) and without (NA) the aid of *ARssist*, forming four scenarios: II_{NA}, II_{AR}, TM_{NA}, TM_{AR}.

A. Instrument Insertion (II): Procedure and Data Analysis

For II_{NA} and II_{AR}, the users are asked to insert the robotic instrument attached on the Patient-Side Manipulator (PSM) into the endoscopy's right side, assuming that the surgeon intends to control it using the right Master Tool Manipulator (MTMR). In II_{AR}, the user has the additional view of the AR content "within" the patient body, as shown in Fig. 3.

For data analysis, we manually annotate the time that the user starts insertion t_S, and the time when the insertion completes t_E based on the PSM kinematics data. We also annotate the time when the instrument tip appears in the endoscopy video, t_M. We compute the trajectory of the instrument tip (Fig. 4) based on the kinematics data and the Denavit-Hartenberg parameters of the PSM: $Q(t), t_S \le t \le t_E$.

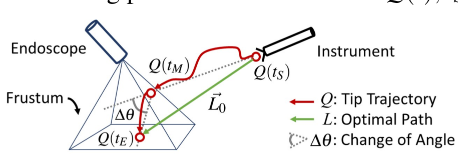


Fig. 4: Subjective metric for evaluation of II

We define three objective metrics to evaluate user performance: navigation time (t_{Nav}), change of angle ($\Delta\theta$) and root-mean-square (RMS) distance of the trajectory (d_{RMS}).

1) Navigation Time (t_{Nav}): Before the instrument appears in the endoscopy, the user is navigating the instrument with the guidance of *ARssist* (II_{AR}) or without any guidance (II_{NA}). This amount of time can be calculated as: $t_{Nav} = t_M - t_S$. If the user wrongly orients the instrument, the insertion will not be successful, and therefore, the user needs to spend more time to find out "where is the instrument", leading to longer navigation time. t_{Nav} determines the efficiency of the user.

2) Change of Angle ($\Delta\theta$): We define $\Delta\theta$ as:

$$\vec{L_1} = Q(t_M) - Q(t_S), \quad \vec{L_2} = Q(t_E) - Q(t_M) \tag{1}$$

$$\Delta\theta = arccos(\vec{L_1} \cdot \vec{L_2}) / (\|\vec{L_1}\| \cdot \|\vec{L_2}\|) \tag{2}$$

The vector $\vec{L_1}$ represents the user's intended direction to insert the instrument before it appears in the endoscopic video. In case of II_{AR}, this intended direction is guided by the virtual instruments and FOV of endoscope rendered inside the phantom (Fig. 3), but in case of II_{NA}, it is based on the user's "feeling" or experience. When the instrument appears in the endoscopy (after t_M), the motion of the instrument will change if the user realizes that the instrument's path diverts from his/her intended path. $\vec{L_2}$ represents the subsequent corrected direction of motion. Therefore, the change of angle $\Delta\theta$ is defined as the angle between $\vec{L_1}$ and $\vec{L_2}$, as an indicator of the consistency of the insertion path (illustrated in Fig. 4).

3) Root-Mean-Square (RMS) Distance (d_{RMS}): We treat the line between $Q(t_S)$ and $Q(t_E)$ as the optimal trajectory $\vec{L_0}$ (see Fig. 4). During the insertion, for each point on the user's trajectory $Q(t)$, we compute its distance to $\vec{L_0}$, and then calculate the RMS value:

$$d(t) = DistancePointToLine(Q(t), \vec{L_0}) \tag{3}$$

$$d_{RMS} = \sqrt{\frac{1}{N} \sum_{t_S < t < t_E} \|d(t)\|^2} \tag{4}$$

where N is the total number of sample points on the trajectory. d_{RMS} represents the extent that the instrument tip diverts from the optimal path. Larger d_{RMS} means that the instrument tip moves further away from the desired path, which increases the potential of collision with an organ or other tissue. Thus, d_{RMS} is an indicator of operation safety.

B. Tool Manipulation (TM): Procedure and Data Analysis

For TM_{NA} and TM_{AR}, the users manipulate a laparoscopic gripper to retract one or more rubber rings out of the body. The rubber rings are visible in the endoscopy. During the procedure, the user needs to first navigate the gripper into the FOV of endoscopy, then pick up the ring based on the feedback of endoscopy, and finally retract it from the site. In TM_{AR}, *ARssist* offers flexible configurations to display the endoscopic video (Fig. 5), while in TM_{NA}, the user can only watch the endoscopy on a monitor.

We annotate the time that the hand-held gripper first appears in the endoscopic video, t_S, and the time that it leaves the scene, t_E, and compute the following objective metric.

1) Manipulation Time (t_{Mani}): The manipulation time t_{Mani} is defined by $t_E - t_S$, as the total amount of time that the user is guided by the endoscopic video to grab and retract the

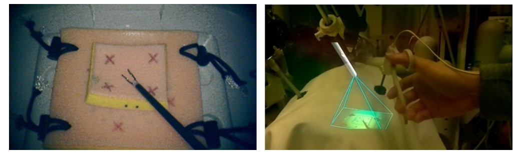
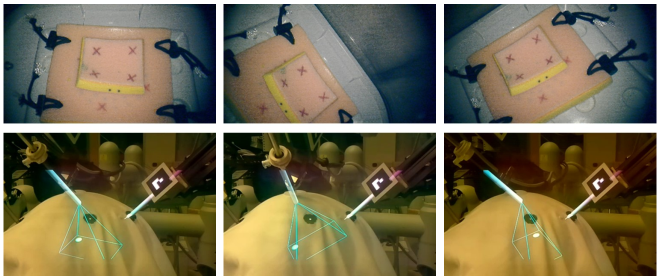


Fig. 5: TM_{AR}: tool manipulation with the help of *ARssist*



(a) EP_1 (b) EP_2 (c) EP_3

Fig. 6: Different poses of the endoscope for the experiment

rubber ring. We do not consider the time spent on inserting the gripper into the FOV of endoscope, which is mainly addressed in the task of instrument insertion.

C. Pose of Endoscope

We use a straight endoscope for the pilot run and a 30°-angled endoscope for the multi-user study, both manufactured by Intuitive Surgical, Inc. For each of the four scenarios, we position the endoscope programmatically at three different poses: EP_1, EP_2 and EP_3 (Fig. 6). Therefore, there are 12 trials in total for each user:

$$II_{NA}^{EPn},\ II_{AR}^{EPn},\ TM_{NA}^{EPn},\ TM_{AR}^{EPn},\ n = 1,2,3$$

Among the three poses, EP_1 is the most normal configuration, where the horizontal axis of the endoscopic video is parallel to the transverse axis of the phantom. EP_2 and EP_3 are two less common, but still clinically possible, endoscope setups. In EP_2, the endoscope is oriented so that it and the user have a similar perspective, whereas EP_3 is the most awkward setup for the user's hand-eye coordination.

D. Experimental Procedure

The experimental procedure for both the pilot run and user study is as follows:

1) Complete the consent form (user study only) and pre-experiment survey
2) Training of II & TM, with and without *ARssist*, repeated multiple times, with multiple endoscope poses
3) Perform II_{NA}^{EPn} & II_{AR}^{EPn}, $n = 1,2,3$ in randomized order
4) Complete the post-experiment survey for II
5) Perform TM_{NA}^{EPn} & TM_{AR}^{EPn}, $n = 1,2,3$ in randomized order (2 rings for pilot run, 1 ring for user study)
6) Complete the post-experiment survey for TM
7) Conduct an informal interview (pilot run only)

The post-experiment questionnaire includes: *i*) self-reported ratings of *outcome*, *speed*, *confidence*, *satisfaction*, *fatigue*, *interest* and *hand-eye coordination* for performance with and without *ARssist* ($0 \sim 5$), *ii*) preference of the three endoscopy visualization methods ($0 \sim 5$), and *iii*) whether *FOV*, *smoothness*, *latency*, *accuracy* of the virtual overlay or any other factor is limiting the current application (Y/N).

We record all related data for each trial, including the robot kinematics, tracking results, stereo endoscopic video, and questionnaire results. All data are timestamped with millisecond accuracy.

Tab I: Background information for the invited surgeons

ID	Age	Gender	Most practiced surgeries	Frequency
#1	39	Female	Hepatectomy	2-4/mo.
#2	36	Male	Radical Prostatectomy	3-4/mo.
#3	33	Male	Prostatectomy, crystectomy	3-4/mo.

V. Pilot Run and Interviews with Surgeons

We invited three surgeons who frequently perform surgeries with the da Vinci robot for the pilot study and interview. All of them have experience working as the first assistant. Their background information is listed in Tab. I.

The surgeons followed the experimental procedure outlined in Sec. IV-D. Their performance data and subjective feedback are shown in Fig. 7.

A. Results and Discussion for Pilot Run

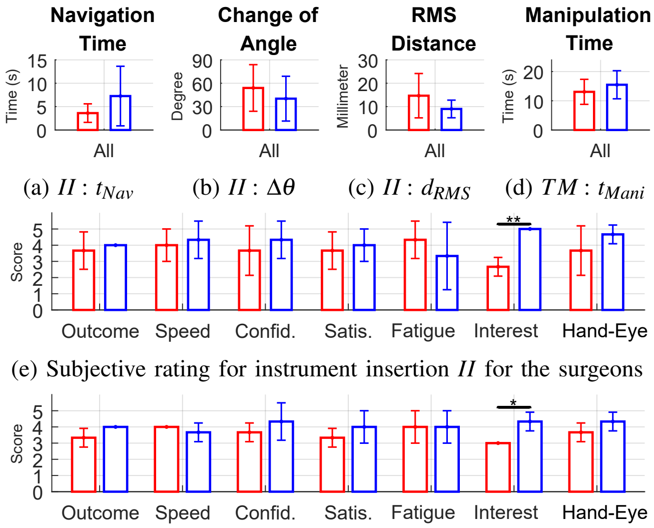
1) Instrument Insertion (II): Figs. 7a, 7b and 7c show the results of the objective metrics. The navigation time t_{Nav} for these experienced users is $7.26 \pm 6.37\,s$ in II_{AR}, which is longer than their normally practised condition ($3.61 \pm 1.92\,s$). The trajectory profile indicates improvement using *ARssist*. The average change of angle $\Delta\theta$ is reduced by 25.44% from $53.95° \pm 29.90°$ in II_{NA} to $40.22° \pm 28.74°$ in II_{AR}. The RMS distance d_{RMS} of the trajectory is reduced by 38.64% on average, from $14.68 \pm 9.48\,mm$ for II_{NA} to $9.01 \pm 3.77\,mm$ for II_{AR}. However, with two-way ANOVA tests, none of the above differences are statistically significant.

2) Tool Manipulation (TM): Fig. 7d shows the average time for the surgeons to manipulate the gripper to retract a rubber ring from the phantom in TM_{NA} ($13.06 \pm 4.28\,s$) and TM_{AR} ($15.48 \pm 4.81\,s$). It is still slightly easier for the surgeons to manipulate under the guidance of endoscopy displayed on a traditional monitor, partially due to the fact that laparoscopic surgeons are especially trained for instrument maneuver under challenging hand-eye coordination conditions [26]. The difference in t_{Mani} is insignificant with a two-way ANOVA test.

3) Subjective Feedback: Figs. 7e and 7f show the subjective ratings of the surgeons for II and TM, which indicate no substantial preference between the AR and non-AR cases. The surgeons do find the task with *ARssist* to be more interesting than without AR guidance.

4) Interview Details from Pilot Run: Surgeon #1 performs liver resection more frequently. Her FA is usually in charge of changing instruments, applying clips/staples, and passing needles. She pointed out some major limitations of the current implementation: i) the HoloLens is too heavy to wear for a long time, ii) there is perceivable lag for the virtual overlay, especially for the hand-held gripper, and iii) the FA needs some training to use the AR system. She agreed that if the limitations are addressed, *ARssist* can be integrated into hepatectomy and will be useful for the FA.

Surgeon #2 performs robotic radical prostatectomy and also occasionally works as a FA. His FA helps bring in instruments, apply suction/retraction and pass in clips and needles. He thinks that *ARssist* can be beneficial for the FA: i) during initial insertion of an instrument at the start



(a) $II : t_{Nav}$ (b) $II : \Delta\theta$ (c) $II : d_{RMS}$ (d) $TM : t_{Mani}$

(e) Subjective rating for instrument insertion II for the surgeons

(f) Subjective rating for tool manipulation TM for the surgeons

Fig. 7: Results for the pilot run with the surgeons. (a)(b)(c): objective results for instrument insertion II. (d): objective results for tool manipulation TM. (e)(f): subjective results for II and TM. Red: without *ARssist*. Blue: with *ARssist*.

of the surgery, ii) during instrument insertion in the middle of surgery, and iii) especially when an angled-endoscope is used. The major limitation is the small FOV of HoloLens. He believes that *ARssist* can mainly benefit a less experienced FA in terms of operation time and safety.

Surgeon #3 performs robotic radical prostatectomy, robotic radical crystectomy and related procedures. During his operation, the FA helps in traction, passing sutures, and clipping vessels. He thinks *ARssist* can improve the performance of the FA: i) in initial setup of the robotic instruments, and ii) facilitate the insertion of the hand-held instruments for passing suture and clipping vessels. The resolution of HoloLens is a current limitation. He believes that the system will be especially useful for an inexperienced FA.

B. Summary of Pilot Run

In the pilot run with the surgeons, *ARssist* extended the navigation time in II and the manipulation time in retracting the rubber ring, but improved the path consistency and operation safety in II. From the subjective ratings, there is not much significant difference between using and not using AR guidance. In the subsequent interview, all surgeons agreed that the FA plays an important role in robotic-assisted surgeries and that our system can be beneficial for the FA, especially for an inexperienced FA. If the current limitations are solved, *ARssist* will be useful for their current tasks, and therefore would improve the overall quality of the surgery.

VI. User Study

After completing the pilot study and further refining the experiment setup, we conducted a multi-user study (HIRB 00007467) with 20 inexperienced subjects (gender: 17 male, 3 female; age: mean 26.65, std 8.00). The rationale for using inexperienced (non-medical) subjects is that the goal of *ARssist* is to improve spatial awareness and hand-eye coordination, rather than to provide medical information, and therefore medical knowledge is not a prerequisite.

A. Results and Discussion of User Study

1) Instrument Insertion (II): Fig. 8a shows that on average, *ARssist* reduced the navigation time t_{Nav} by 34.57%, from $3.58 \pm 4.50 s$ in II_{NA} to $2.34 \pm 2.10 s$ in II_{AR}. For each endoscope pose EP_i, we use a t-test to determine whether the difference between II_{AR} and II_{NA} is significant. For EP_3, the users spent $2.24 \pm 1.63 s$ to navigate in II_{AR}, which is significantly ($p = 2.96 \times 10^{-2}$) shorter than the amount of time spent in II_{NA}: $5.85 \pm 6.96 s$. EP_1 and EP_2 do not show significance with a t-test. The user's navigation time is dependent on two categorical variables: i) II_{AR} or II_{NA}, and ii) pose of endoscope: EP_1, EP_2 or EP_3. Therefore, we use a two-way ANOVA to test whether t_{Nav} is significantly affected by the two variables. The result shows that both factors are significant for t_{Nav} ($p_{AR} = 4.53 \times 10^{-2}, p_{EP} = 2.09 \times 10^{-2}$).

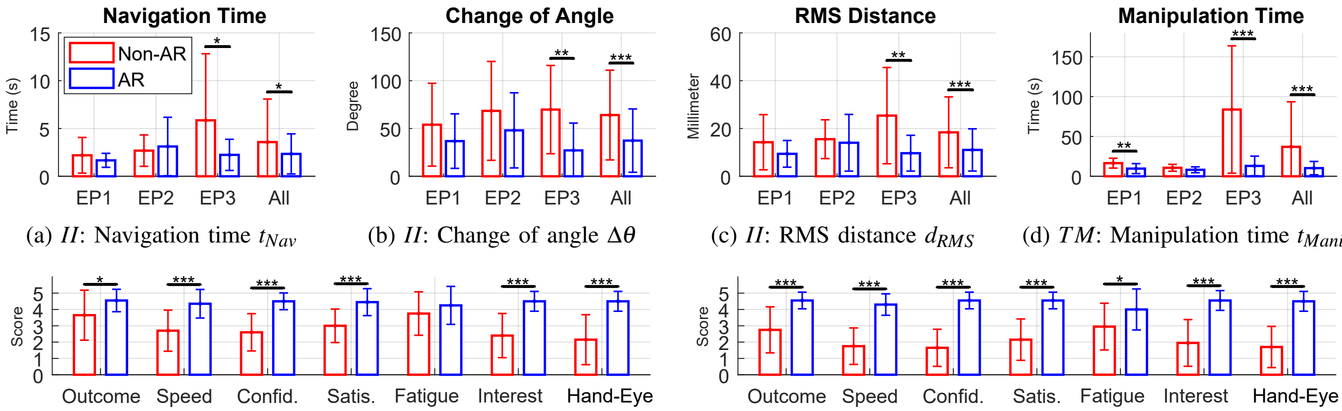
As seen in Fig. 8b, the change of angle $\Delta\theta$ is reduced by 41.74% from $64.04 \pm 46.93°$ for II_{NA} to $37.31 \pm 33.11°$ in II_{AR}. Within all groups of endoscope pose, the average $\Delta\theta$ is smaller in II_{AR}. With a t-test, only EP_3 shows significant improvement ($p = 1.15 \times 10^{-3}$). A two-way ANOVA test of $\Delta\theta$ shows that the use of *ARssist* significantly affects user performance ($p_{AR} = 4.51 \times 10^{-4}$).

The results of RMS distance d_{RMS} are shown in Fig. 8c. The mean d_{RMS} is reduced by 40.04%, from $18.39 \pm 14.83 mm$ in II_{NA} to $11.02 \pm 8.83 mm$ in II_{AR}. This reduction is achieved for all tested endoscope poses EP_1, EP_2 and EP_3, but is only significant for EP_3 ($p = 2.26 \times 10^{-3}$). The two-way ANOVA shows that, with *ARssist*, the reduction of d_{RMS} is quite significant ($p_{AR} = 8.59 \times 10^{-4}$).

We extract the participants' subjective ratings for their experience with and without *ARssist* from the questionnaire. For II_{NA}, their average ratings are 3.7 (*outcome*), 2.7 (*speed*), 2.6 (*confidence*), 3.0 (*satisfaction*), 3.8 (*fatigue*), 2.4 (*interest*) and 2.2 (*hand-eye coordination*). For II_{AR}, their average ratings are 4.6, 4.4, 4.5, 4.5, 4.3, 4.5 and 4.5. The t-test shows that users significantly prefer II_{AR} in *outcome*, *speed*, *confidence*, *satisfaction*, *interest* and *hand-eye coordination*. In terms of *fatigue*, the average rating in II_{AR} is not significantly higher than in II_{NA}.

2) Tool Manipulation (TM): The tool manipulation time t_{Mani} is shown in Fig. 8d. It is $10.26 \pm 8.36 s$ in TM_{AR}, which is significantly shorter than $36.96 \pm 56.52 s$ in TM_{NA} (72.25%). The two-way ANOVA test shows that the reduction of t_{Mani} is significant ($p_{AR} = 2.50 \times 10^{-5}$). The average time reduction is also very significant for EP_1 ($p = 1.32 \times 10^{-3}$) and EP_3 ($p = 3.63 \times 10^{-4}$) after t-test. Especially for EP_3 in TM_{NA}, many users spend a lot of time adjusting the gripper to approach the rubber ring under bad hand-eye coordination. With *ARssist*, the FOV of the endoscope is directly visualized for the user, therefore the spatial relationship between the gripper's motion and its appearance in the endoscopic video is naturally registered.

The subjective results for tool manipulation are shown in Fig. 8f. For TM_{NA}, the average ratings for *outcome*,



(a) II: Navigation time t_{Nav} (b) II: Change of angle $\Delta\theta$ (c) II: RMS distance d_{RMS} (d) TM: Manipulation time t_{Mani}

(e) Subjective rating for instrument insertion (II) (f) Subjective rating for tool manipulation (TM)

Fig. 8: Results for the instrument insertion (II) and tool manipulation (TM) in the user study. With *ARssist* (II_{AR}), the navigation time t_{Nav}, change of angle $\Delta\theta$, and RMS distance d_{RMS} are significantly reduced, leading to better efficiency, path consistency, and improved operation safety. In TM_{AR}, the manipulation time t_{Mani} is significantly reduced due to improved hand-eye coordination. Subjective ratings prefer *ARssist*. NOTE: Bar plot: mean value. Error line: standard deviation. EP_1, EP_2 and EP_3 are different poses of the endoscope. The stars indicate the level of significance of the difference between the two columns, calculated with a t-test for each group of EP_i, with two-way ANOVA for combined data ('All' column).

speed, *confidence*, *satisfaction*, *fatigue*, *interest* and *hand-eye coordination* are 2.8, 1.8, 1.7, 2.2, 3.0, 2.0 and 1.7, respectively. For TM_{AR}, the average ratings are 4.6, 4.3, 4.6, 4.6, 4.0, 4.6 and 4.5. The improvement with *ARssist* is significant in the users' ratings in all listed aspects.

3) Preference for Endoscopy Visualization: As described in Sect. III-B, *ARssist* provides three visualization methods to render the stereo endoscopic video. We extract the user's selection of visualization methods and their subjective ratings for them, as listed in Tab. II.

Note that some users switched the visualization method during the operation, therefore, the total number of selections exceeds the number of trials (60). Most users selected the frustum projection method for both II and TM. This method orients the video at the correct geometric pose and restores the hand-eye coordination of the user, which is especially helpful for inexperienced users who have not been trained for hand-eye coordination.

VII. Limitations and Future Work

In the questionnaire, the users (both inexperienced users and surgeons) are asked about the limitations of the system. Overall, 47.8% of the users believe the *FOV* of HoloLens is a limiting factor, 17.4% selected *smoothness*, 19.6% selected *latency*, and 41.3% selected *accuracy* of virtual overlay. Apart from these, 4 users reported that the resolution of endoscopy is not high enough, and also 4 users reported difficulty seeing the depth information in the endoscopy. We have recently significantly improved endoscopy streaming to achieve stereo 1080P resolution. The limited training time may as well negatively impact the consistency of the users' performance. While some of the issues (FOV) are limited by the current generation of OST-HMD hardware [27], we believe with more advanced or dedicated hardware entering the market, these problem can be gradually alleviated.

Tab II: Total amount of time, number of selections, and user's rating for each visualization method of the stereo endoscopy

Method	Instrument Insertion			Tool Manipulation		
	Time	No.	Rating	Time	No.	Rating
Heads-up display	12.39s	2	2.0	31.15s	3	2.0
Virtual monitor	42.47s	8	2.2	21.08s	2	1.9
Frustum projection	223.77s	54	4.5	552.87s	58	4.9

We would also like to bring more augmented reality content into *ARssist*, such as real-time endoscopy-based reconstruction and visualization. Furthermore, we plan to evaluate our system in more complete surgical procedures, and with more specialized users.

VIII. Conclusions

In this paper, we refine and evaluate *ARssist* [9], an augmented reality application to aid the first assistant in a robotic-assisted surgery. We chose two frequently occurring tasks, instrument insertion and tool manipulation, and conducted a pilot run with three experienced surgeons and a user study with 20 relatively inexperienced users. In the pilot run, the experienced users did not show a significant difference with *ARssist*, but they agreed that it could be beneficial for the FA's performance, especially for inexperienced users. The user study showed that *ARssist* can benefit inexperienced users by improving efficiency (34.57% shorter navigation time), navigation consistency (41.74% less change of angle), and safety (40.04% lower RMS path deviation) for instrument insertion, and by enhancing hand-eye coordination (72.25% less time) in tool manipulation.

Acknowledgment

We acknowledge Dr. Yuen-Chun Teoh, Dr. Ching-Ning Chong and Dr. Ka-Fung Chiu for participating in the pilot study and providing valuable feedback, and Prof. Zheng Li for initiating the connection. We thank Dr. Simon DiMaio, Cortney Jansen, and Govinda Payyavula from Intuitive Surgical for their valuable feedback and support.

REFERENCES

[1] A. R. Lanfranco, A. E. Castellanos, J. P. Desai, and W. C. Meyers, "Robotic surgery: a current perspective," *Annals of Surgery*, vol. 239, no. 1, p. 14, 2004.

[2] S. DiMaio, M. Hanuschik, and U. Kreaden, "The da Vinci surgical system," in *Surgical Robotics*. Springer, 2011, pp. 199–217.

[3] C.-C. Abbou, A. Hoznek, L. Salomon, L. E. Olsson, A. Lobontiu, F. Saint, A. Cicco, P. Antiphon, and D. Chopin, "Laparoscopic radical prostatectomy with a remote controlled robot," *The Journal of Urology*, vol. 197, no. 2, pp. S210–S212, 2017.

[4] R. Sinha, M. Sanjay, B. Rupa, and S. Kumari, "Robotic surgery in gynecology," *Journal of Minimal Access Surgery*, vol. 11, no. 1, p. 50, 2015.

[5] P. Stadler, L. Dvoracek, P. Vitasek, and P. Matous, "The da Vinci robot in the field of vascular surgery," *Journal of Vascular Surgery*, vol. 61, no. 2, p. 584, 2015.

[6] O. Sgarbura and C. Vasilescu, "The decisive role of the patient-side surgeon in robotic surgery," *Surgical Endoscopy*, vol. 24, no. 12, pp. 3149–3155, 2010.

[7] A. M. Potretzke, B. A. Knight, J. A. Brockman, J. Vetter, R. S. Figenshau, S. B. Bhayani, and B. M. Benway, "The role of the assistant during robot-assisted partial nephrectomy: does experience matter?" *Journal of Robotic Surgery*, vol. 10, no. 2, pp. 129–134, 2016.

[8] R. Nayyar, S. Yadav, P. Singh, and P. N. Dogra, "Impact of assistant surgeon on outcomes in robotic surgery," *Indian Journal of Urology (IJU): journal of the Urological Society of India*, vol. 32, no. 3, p. 204, 2016.

[9] L. Qian, A. Deguet, and P. Kazanzides, "ARssist: Augmented reality on a head-mounted display for the first assistant in robotic surgery," *Healthcare Technology Letters*, 2018.

[10] P. Vávra, J. Roman, P. Zonča, P. Ihnát, M. Němec, J. Kumar, N. Habib, and A. El-Gendi, "Recent development of augmented reality in surgery: a review," *Journal of Healthcare Engineering*, vol. 2017, 2017.

[11] F. Volonté, F. Pugin, N. C. Buchs, J. Spaltenstein, M. Hagen, O. Ratib, and P. Morel, "Console-integrated stereoscopic OsiriX 3D volume-rendered images for da Vinci colorectal robotic surgery," *Surgical Innovation*, vol. 20, no. 2, pp. 158–163, 2013.

[12] F. Volonté, N. C. Buchs, F. Pugin, J. Spaltenstein, B. Schiltz, M. Jung, M. Hagen, O. Ratib, and P. Morel, "Augmented reality to the rescue of the minimally invasive surgeon. the usefulness of the interposition of stereoscopic images in the da Vinci^{TM} robotic console," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 9, no. 3, pp. e34–e38, 2013.

[13] P. Pessaux, M. Diana, L. Soler, T. Piardi, D. Mutter, and J. Marescaux, "Towards cybernetic surgery: robotic and augmented reality-assisted liver segmentectomy," *Langenbeck's Archives of Surgery*, vol. 400, no. 3, pp. 381–385, 2015.

[14] ——, "Robotic duodenopancreatectomy assisted with augmented reality and real-time fluorescence guidance," *Surgical Endoscopy*, vol. 28, no. 8, pp. 2493–2498, 2014.

[15] L.-M. Su, B. P. Vagvolgyi, R. Agarwal, C. E. Reiley, R. H. Taylor, and G. D. Hager, "Augmented reality during robot-assisted laparoscopic partial nephrectomy: toward real-time 3D-CT to stereoscopic video registration," *Urology*, vol. 73, no. 4, pp. 896–900, 2009.

[16] O. Mohareri, C. Schneider, T. K. Adebar, M. C. Yip, P. Black, C. Y. Nguan, D. Bergman, J. Seroger, S. DiMaio, and S. E. Salcudean, "Ultrasound-based image guidance for robot-assisted laparoscopic radical prostatectomy: initial in-vivo results," in *International Conference on Information Processing in Computer-Assisted Interventions*. Springer, 2013, pp. 40–50.

[17] R. Kumar and A. K. Hemal, "The scrubbed surgeon in robotic surgery," *World Journal of Urology*, vol. 24, no. 2, pp. 144–147, 2006.

[18] B. J. Ardila, M. A. Orvieto, and V. R. Patel, "Role of the robotic surgical assistant," in *Robotic Urologic Surgery*. Springer, 2011, pp. 495–505.

[19] B. Wentink, "Eye-hand coordination in laparoscopy-an overview of experiments and supporting aids," *Minimally Invasive Therapy & Allied Technologies*, vol. 10, no. 3, pp. 155–162, 2001.

[20] P. Kazanzides, Z. Chen, A. Deguet, G. S. Fischer, R. H. Taylor, and S. P. DiMaio, "An open-source research kit for the da Vinci® surgical system," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 6434–6439.

[21] E. Azimi, L. Qian, N. Navab, and P. Kazanzides, "Alignment of the virtual scene to the 3d display space of a mixed reality head-mounted display," *arXiv preprint arXiv:1703.05834*, 2018.

[22] Z. Wang, S. C. Lee, F. Zhong, D. Navarro-Alarcon, Y.-h. Liu, A. Deguet, P. Kazanzides, and R. H. Taylor, "Image-based trajectory tracking control of 4-dof laparoscopic instruments using a rotation distinguishing marker," *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1586–1592, 2017.

[23] J. Wang, L. Qian, E. Azimi, and P. Kazanzides, "Prioritization and static error compensation for multi-camera collaborative tracking in augmented reality," in *Virtual Reality (VR), 2017 IEEE*. IEEE, 2017, pp. 335–336.

[24] L. Qian, E. Azimi, P. Kazanzides, and N. Navab, "Comprehensive tracker based display calibration for holographic optical see-through head-mounted display," *arXiv preprint arXiv:1703.05834*, 2017.

[25] L. Qian, M. Unberath, K. Yu, B. Fuerst, A. Johnson, N. Navab, and G. Osgood, "Towards virtual monitors for image guided interventions-real-time streaming to optical see-through head-mounted displays," *arXiv preprint arXiv:1710.00808*, 2017.

[26] A. M. Derossis, G. M. Fried, H. H. Sigman, J. S. Barkun, and J. L. Meakins, "Development of a model for training and evaluation of laparoscopic skills 1," *The American Journal of Surgery*, vol. 175, no. 6, pp. 482–487, 1998.

[27] L. Qian, A. Barthel, A. Johnson, G. Osgood, P. Kazanzides, N. Navab, and B. Fuerst, "Comparison of optical see-through head-mounted displays for surgical interventions with object-anchored 2D-display," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. 6, pp. 901–910, 2017.