Design of interactive augmented reality functions for robotic surgery and evaluation in dry-lab lymphadenectomy

Maria-Paola Forte1 | Ravali Gourishetti1 | Bernard Javot1 | Tobias Engler2 | Ernest D. Gomez3 | Katherine J. Kuchenbecker1

1Haptic Intelligence Department, Max Planck Institute for Intelligent Systems, Stuttgart, Germany  
2Tübingen University Hospital, Tübingen, Germany  
3Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA

Correspondence  
Maria-Paola Forte. Haptic Intelligence Department, Max Planck Institute for Intelligent Systems, 70569 Stuttgart, Germany. Email: forte@is.mpg.de

Funding information  
Max-Planck-Gesellschaft

Abstract

Background: Augmented reality (AR) has been widely researched for use in healthcare. Prior AR for robot-assisted minimally invasive surgery has mainly focussed on superimposing preoperative three-dimensional (3D) images onto patient anatomy. This article presents alternative interactive AR tools for robotic surgery.

Methods: We designed, built and evaluated four voice-controlled functions: viewing a live video of the operating room, viewing two-dimensional preoperative images, measuring 3D distances and warning about out-of-view instruments. This low-cost system was developed on a da Vinci Si, and it can be integrated into surgical robots equipped with a stereo camera and a stereo viewer.

Results: Eight experienced surgeons performed dry-lab lymphadenectomies and reported that the functions improved the procedure. They particularly appreciated the possibility of accessing the patient’s medical records on demand, measuring distances intraoperatively and interacting with the functions using voice commands.

Conclusions: The positive evaluations garnered by these alternative AR functions and interaction methods provide support for further exploration.

Keywords  
augmented reality, laparoscopic surgery, robotic surgery

1 | INTRODUCTION

Augmented reality (AR) seamlessly integrates the user’s sensory perception of the real world with the power of digital information and technology. Following recent advances in optics, sensing and computer systems, applications of AR are being vigorously researched and successfully deployed in several areas of the healthcare industry, including medical education and training1 and intravenous catheterisation.2

AR was being used in surgery long before this recent surge of more general medical interest. The first AR-assisted intervention was performed in neurosurgery in 1986.3 A computer-based vision system registered and overlaid computerised tomography (CT) image data in the operating microscope so that the surgeon could locate
small targets in the brain intraoperatively without a stereotaxic frame. Enriched by more than three decades of research, neurosurgery has continued to push the frontiers of AR technology in the operating room (OR), largely following this early paradigm of image-based surgical navigation.1

Experts believe AR has the potential to reduce the surgeon’s cognitive load and thereby increase their focus and efficiency during both minimally invasive surgery (MIS, also known as laparoscopic surgery) and robot-assisted minimally invasive surgery (RMIS).5 Even though the introduction of AR into laparoscopy has been widely researched, AR has not gained widespread adoption in clinical practice.5,6 Most prior work applies AR to RMIS rather than MIS: in RMIS, the primary surgeon typically sees the operative field through a stationary stereoscopic viewer, avoiding tremors and technical issues related to the tracking of the surgeon’s head movement. Furthermore, the surgical instruments are controlled via in-hand manipulators and are rigidly attached to the robotic arms. This design choice allows for precise and stable calibration and registration.7 In contrast, the heads-up or head-mounted displays used in MIS or open surgery8 have the potential to benefit the assistant surgeon standing at the bedside in RMIS.9

Besides the ‘off-screen indicator’ function integrated into the software of the da Vinci system itself,10 surgical AR innovations have predominantly been limited to research prototypes that have not yet been transformed into commercial products. The lack of clinical translation can mainly be attributed to the regulatory and medical-legal issues associated with modifying surgical devices and instruments.11

Bernhardt et al.5 Qian et al.,6 and Makhataeva and Varol12 reviewed the literature that has applied AR to MIS and RMIS. The trend has been to use AR in the same way as it has been successfully applied in neurosurgery, that is, for surgical navigation; in the brain, only small deformations occur between the preoperative and intraoperative states, and stable bony landmarks can be automatically maintained throughout a procedure. In contrast, the presence of soft tissues and deformable organs in the abdomen makes accurate superimposition and tracking extremely challenging in MIS and RMIS.7 This complexity has stymied automation and led research teams to manually or semi-manually control the images overlaid on the patient’s anatomy. For example prior AR work done on the da Vinci robot used the auxiliary tile of the da Vinci TilePro multi-input display13 to show the three-dimensional (3D) model of the patient’s prostate superimposed on the intraoperative view.13 Preoperative data were manually controlled using a dedicated application and input device. Manually controlled virtual images have also been overlaid on the primary tile in real time using a video mixer14 or a tablet. This last method was recently integrated into da Vinci systems15; however, the preoperative data are not superimposed on the patient’s anatomy and are displayed in the auxiliary tile.

A minor line of research has focussed on using AR for sensory substitution, that is, to compensate for the lack of haptic feedback whilst operating at the surgical console. Akinbiyi et al.16 developed an AR system for presenting contact force information measured through strain gauges mounted on the da Vinci instruments. The 3D position of the virtual contact was identified through external kinematic trackers. Yamamoto et al.17 overlaid the tissue stiffness, which was obtained by comparing the recorded kinematics and force data with those of the selected mathematical tissue model. AR was generated using a hue-saturation-luminance representation on a semi-transparent disc at the tissue surface. Similar work was proposed by Zevallos et al.18 and developed with a da Vinci Research Kit (dVRK): a phantom tumour, identified in an area selected by the surgeon, was visually augmented with overlaid stiffness information obtained with a miniature force sensor.

Independent of the final application, Sielhorst et al.19 identified three essential requirements needed to ensure the quality of a medical AR system:

- **Reliability**: correct functioning of the system during a specific time duration.
- **Usability**: ease of use for surgeons.
- **Inter-operability**: use of generic data and protocols to enable high compatibility with other equipment.

Previous literature has focussed mainly on reliability with some detriment to usability and inter-operability5 because of the significant medical-legal implications associated with technology that impacts surgical performance. However, the acceptance and utility of an assistance system can be harmed by the visualisation of unnecessary, obsolete or redundant information that might distract or confuse the end user,20 or by unintuitive user interfaces that require too much of the user’s attention. Thus, a poorly designed human-machine interface will prevent integration into the clinical workflow. Furthermore, the proposed systems often have complex protocols composed of tedious setups and lengthy calibration, reducing the likelihood of clinical use.5

Inspired by a desire to explore alternative uses and interaction methods of AR for RMIS, and driven to meet all the requirements defined by Sielhorst et al.,19 we present a robot-independent hardware and software system that provides four intuitive AR functions through the use of computer graphics and computer vision. The first two functions bring additional visual information into the surgeon’s view, and the other two functions leverage computer vision to provide more sophisticated computational capabilities. Our approach relies only on vision rather than robot kinematics to guarantee the inter-operability of the platform, to limit the necessary hardware and software, to provide precise visual alignment between AR markers and the tip of the instrument and to avoid lengthy calibration procedures that are challenging for non-technical personnel to perform.7 An early version of this system was described in the master’s thesis of Forte21 and reported at SAGES 2019.22
2 | METHODS

2.1 | Developed functions

We iteratively designed and implemented four AR functions that can be used in both lab and clinical settings: external camera view, medical records, distance computation and instrument position warning system.

We intended to create easy and intuitive interactions that could improve surgical procedures by speeding up processes, increasing effectiveness, helping in decision making and reinforcing safety. Concise descriptions and motivations of each function follow, and Figure 1 shows our functional prototypes.

- The external camera view function allows the surgeon to see the surrounding OR environment without removing their head from the robotic console. As shown in Figure 1A, the external camera might point at the patient cart (e.g., to look for collisions between the robotic arms), or it could look elsewhere in the OR.
- The medical records function aims to expedite the procedure by providing the surgeon with preoperative data that can affect the surgery, for example CT scan, histology report and consent form. Staying in the console rather than removing one's head reduces sight diversion, a common issue in image-guided surgery. To avoid a shift of focus and any size reduction of the intraoperative view, we overlay the preoperative data directly in the primary tile rather than on an auxiliary window. The surgeon can define the position and size of the displayed medical records.
- The distance computation function allows the surgeon to measure 3D Euclidean distances and sizes intraoperatively, such as during tumour enucleation or tissue replacement. Similar to a previous patent, the surgeon uses the tip of the surgical instruments (with closed jaws) to select the points between which the distance should be measured. Once a point is defined, a virtual marker appears and the system computes the distance between two markers. The surgeon can choose between two methods to place the markers:

- Method 1: placing both markers simultaneously. The distance is measured between the two instrument tips, as shown in Figure 1C.
- Method 2: placing the markers sequentially. The surgeon applies one marker at a time with either the left or the right instrument. The distance is calculated after the placement of the second marker.
- The instrument position warning system function has the goal of improving safety. Moving a surgical instrument out of the camera’s view can be dangerous since unseen tissue could be damaged. A visual indicator appears right before either robotic instrument moves out of the surgeon's field of vision. The warning persists until the missing instrument returns. In contrast with the built-in 'off-screen indicator' of the da Vinci Xi, our implementation relies on vision rather than robot kinematics for improved accuracy and easier integration into different surgical robots.

Sample videos of all four functions recorded during the user study can be seen in the Supporting Information S2–S5 of this article. The audio was distorted to protect the identities of the surgeon participants.

2.2 | Visual augmented reality

To meet the inter-operability requirement defined by Sielhorst et al., we developed a platform that does not require access to any custom data interfaces of the surgical robot or information on the kinematic operation of its instruments. Instead, our platform uses visual augmented reality and image processing. Visual AR consists in the superimposition of computer-generated visual information on a user’s view of the real world, resulting in a composite view. To achieve this superimposition, two-way video communication between the surgical robot and a workstation computer is required: one direction streams images of the surgical field as they are captured by the stereo camera (the user’s view of the real world), and the other

![Figure 1](image-url)
direction displays the augmented images in the stereo viewer (the composite view). Our requirements for video communication are:

- to have real-time processing for the vision pipeline,
- to be able to read the synchronised stereo images acquired by the surgical robot and
- to output the processed content in HD into the stereo viewer of the surgical robot.

A description of our setup, its integration with the da Vinci system and the computer, and the superimposition of the computer-generated images follows.

2.2.1 Video capture and playback card

To acquire the video signal, most research groups working with the dVRK use off-the-shelf USB frame grabbers. This method suffers from high latency in the grabbing process. Furthermore, most low-cost frame grabbers cannot read digital video signals (HDMI, DVI and SDI) and are limited to noise-sensitive analog video formats. These issues make this approach unsuitable for our requirements. In addition to capturing the images seen by the stereo endoscope, the video interface must also display the virtual content in the stereo viewer.

To fulfill the video communication requirements, we selected a Blackmagic Design DeckLink Quad 2 card that can perform keying, that is, composite two full-frame images together. We use this technique to overlay other video signals, images and computer-generated stereoscopic content over the source video with minimal latency. As shown in Figure 2, we create the augmented world by compositing each real-world frame with a transparent frame of the same size wherein the virtual content is placed at the desired position; the left and right stereoscopic frames are processed individually.

It is essential to consider that most surgical stereo endoscopes have high-definition (HD) resolution; only advanced video playback cards can perform keying in HD.

The DeckLink Quad 2 can be used simultaneously for up to eight independent I/O ports that can be configured to be either input (reading), output (writing) or keying. This means that it is not possible to read and perform keying simultaneously from the same channel. Nonetheless, we need to read the real-world images to process the captured images, and at the same time, we need to perform keying to superimpose the transparent frame we have created on top of the images acquired by the stereo endoscope. To overcome this challenge, we split the left and right input video signals taken from the da Vinci system with two SDI splitters, one for each channel of the stereo camera.

2.2.2 Integration of the video card with the robotic system

The left and right camera control units (CCUs) provide HD-SDI output for the respective cameras. Normally, this output is connected to the HD-SDI input of the Core (the processing center for the robotic system) through BNC cables. Thus, it is possible to intercept the stream of the standard vision pipeline by placing the video capture and playback card in this location, between the CCUs and the Core (the exact location varies depending on the generation of the da Vinci system). The modified pipeline is shown in Figure 3. This setup requires the computer to be booted and processing in order for the da Vinci to work in its standard mode. In an OR, additional hardware would be required to provide a video connection.

![Real world + Virtual content = Augmented world](Image)

**Figure 2** The keying technique is used to superimpose the virtual content on the real world. The real-world image and the transparent frame with the virtual content are composited to generate the augmented-world image.

![Diagram of video setup](Image)

**Figure 3** The standard video connection between the left and right camera control units and the Core is replaced with two SDI-splitters and the DeckLink Quad 2 video capture and playback card. I, input; L, left; O, output; R, right.
that does not go through the video capture and playback card and, in turn, through the computer, to avoid losing the intraoperative video input in case of computer failure. Specifically, one would need three-output splitters instead of the two-output splitters, plus a switcher before each channel of the Core.

In contrast to the da Vinci video outputs, the selected card has mini-BNC connectors, so mini-BNC-to-BNC cables are required. Further, RG-59 coaxial cables with BNC connectors of impedance 75 Ω are used in order to maintain a consistent impedance throughout the system and avoid reflections.

2.2.3 | Integration of the video card with the workstation computer

The selected video capture and playback card requires a second-generation PCI Express slot (8 or 16 lanes). To interface with the vision system in real time and in HD, we developed appropriate drivers that can be downloaded from our GitHub site.26 These drivers expose DeckLink cards to a ROS network by leveraging libdecklink, a higher-level interface to the BlackMagic Design SDK used to control the cards.

2.2.4 | Overlaying of the computer-generated images

As shown in Figure 4, the images acquired by the stereo endoscope and, in turn, by the video capture and playback card, have a resolution of 1920 × 1080 (16:9 ratio). Importantly, the stereo images shown in the stereo viewer are cropped by the da Vinci system itself to a resolution of 1340 × 1072 (5:4 ratio). The cropping of the internal and external edges (respectively, lighter and darker areas in Figure 4) reduces problems related to the keystone distortion, typical of the toe-in stereo-rendering method, as distortions are more severe towards the edges and in particular at the corners.27 This strategy is also confirmed by the patent of the endoscope.28 Furthermore, the cropping of the internal edges compensates for the small endoscope baseline (~6 mm), which is the distance between the two cameras. Because of this cropping, the virtual content has to be positioned in the 1920 × 1080 transparent frame considering the resolution of the images seen by the surgeon, that is, 1340 × 1072. Thus, researchers adapting our technology to a new system would need to provide as input the cropping of the images shown in their stereo viewer, if any. A further consideration is needed when using a da Vinci robot; the x and y origin of the crop varies slightly for both channels every time the da Vinci stereo endoscope 3D calibration is performed to correct for the misalignment of the optical axes.28 Consequently, the stereo window position, that is where the disparity of the raw stereo images is zero, also varies. These subtle shifts must be taken into account to properly overlay the virtual content on the raw stereo images, and in particular when overlaying directly in the raw images.

2.2.5 | Inter-operability

DeckLink cards support the most popular video formats and can detect input resolution and pixel format. Our drivers have been validated on several variants of DeckLink cards with different generations of da Vinci systems; every tested combination worked. Furthermore, the described video communication is not limited to work with da Vinci systems; it can be extended to any SDI-compatible devices. Our platform can thus be integrated into any surgical robotic system equipped with a stereo camera and a stereo viewer.

2.3 | Image processing

In the external camera view and medical records functions, the video acquired by an external camera (Figure 1A) and the patient’s pre-operative data (Figure 1B), respectively, are overlaid through keying on the stereo window plane, the same plane as the da Vinci overlaid elements, for example, instrument status, critical messages.

The instrument position warning system and distance computation functions include image processing of the surgical field captured by the stereo endoscope to identify the instruments and their tips. An

![Figure 4](image-url) The left and right images acquired by the stereo endoscope have a resolution of 1920 × 1080. In the stereo viewer, they are cropped mostly on the horizontal internal edges (lighter area) and on the horizontal external edges (darker area), leaving a visible area of 1340 × 1072 (area outlined by the red dashed rectangle). The axes are oriented as shown by the arrows.
alternative strategy would be to use robot kinematic data either alone or combined with vision; however, previous work showed that accurate detection of the tip of an articulated instrument is not achievable with only kinematic data due to friction, flexibility and other non-idealities in the robot’s kinematic chain. Achieving positioning errors on the millimetre scale requires at least two-dimensional (2D) images. Furthermore, relying on robot kinematic data could undermine the platform’s inter-operability and would also require accurate hand-eye calibration.

The instrument position warning system function uses the neural network TernausNet-16 to locate the surgical instruments. This neural network can be used to perform three different tasks: binary segmentation, part segmentation and instrument segmentation. In our function, the raw left-channel image is given as input to TernausNet-16 for binary segmentation: the left and right instruments are extracted from the background and assigned to two different labels. When the distal end of either instrument reaches a distance of about 160 pixels from the left or right edge of the $1340 \times 1072$ frame shown to the surgeon, or a distance of about 190 pixels from the top or bottom edge of the same frame, the relevant left or right attention symbol is overlaid on the zero-disparity plane (Figure 1C). To decrease the computing cost and take advantage of the wider captured view, we use only one channel of the $1920 \times 1080$ stereo-pair images to detect out-of-view instruments, and we assume that the right instrument is always visible in the right-most 200 pixels of the cropped image, and the left instrument in the left-most 200 pixels.

The distance computation function recognises the tips of the robotic instruments to place virtual markers. Each placement consists of four steps, which are summarised in Figure 5 and described here.

In the first step, the stereo images acquired with the toed-in cameras are rectified, and the lens distortions are corrected.

In the second step, the rectified left-channel image is given as input to the neural network TernausNet-16 for part segmentation: the robotic instruments are extracted from the background, and three articulated parts of each instrument are identified, that is, the rigid shaft, the articulated wrist and the claspers. We remove wrongly classified areas based on thresholds related to their sizes. The projection of the tip of an instrument ($P$) on the rectified left image plane is defined as the corner of the claspers, that is farthest from the centre of the wrist. If two corners have a similar Euclidean distance from the centre of the wrist, tip $P$ is defined as their midpoint. Importantly, these two conditions depend on how the surgeon holds the instrument tip (Figure 6). The aforementioned corners are detected with the Harris corner detector and refined in small search windows (best results were obtained with $11 \times 11$ search windows). Finally, an $81 \times 61$ region-of-interest (ROI) image centred on tip $P$ is created. The selected dimensions guarantee that the ROI entirely includes tip $P$ when the instrument is held at different distances from the stereo endoscope.

![Diagram](image)

**Figure 5** Placing a virtual marker for the distance computation function consists of four steps. Step 1 rectifies the acquired raw images and corrects the lens distortions. In step 2, the tip (marked in blue) of the selected robotic instrument is identified in the left image plane. As shown on the left, the instruments are segmented with deep neural networks in the case of real tissues and with colour-based segmentation in the case of simulated tissues. Step 3 retrieves the tip in the right image plane through template matching. In step 4, the identified position of the tip is transformed to be shown in the raw stereo images that will be displayed in the surgeon console.
In the third step, the ROI image extracted from the left-channel image is used as a template to search for tip $P$ in the rectified right-channel image using template matching. Since we rectified the images, the same pixels lie (approximately) on the same epipolar line in the left and right images, so the vertical searching area is reduced. Consideration of the maximum feasible disparity decreases the horizontal searching area.

In the fourth step, the pixel locations of tip $P$ identified in the rectified stereo-pair image are retrieved in the raw stereo-pair image through the undistortion and rectification transformation maps. A white circle of radius 5 pixels centred on the retrieved tip positions is superimposed over each raw image through keying, thereby allowing the user to visualise the virtual marker placed with tip $P$ in 3D space (Figure 1D). The $x$, $y$- and $z$-coordinates of the virtual marker are computed by combining the identified disparity $d = x_L - x_R$ with the parameters of the intrinsic right camera matrix for the rectified images (obtained with a stereo calibration of the stereo endoscope\(^{31}\)) as follows:

$$z_P = \frac{T_x}{d},$$

$$x_P = \frac{(x_L - c_x) \cdot z_P}{f},$$

$$y_P = \frac{(y_L - c_y) \cdot z_P}{f},$$

where $T_x$ is the translation along the $x$-axis of the optical centre of the right camera in the left camera’s frame, $x_L$ and $x_R$ are the $x$-coordinates of the projections of tip $P$ on the left and right image planes, $y_L$ is the $y$-coordinate of the projection of tip $P$ on the left image plane (Figure 4), $c_x$ and $c_y$ are the optical centres (principal points) and $f$ is the camera’s focal length.

The neural network TernausNet-16 produces good results when working on images acquired from real surgeries, similar to those it was trained on\(^{30}\), however, during this design phase and for the sake of reproducibility and live demonstration, we chose to evaluate the functions with phantom tissues. Thus, we replaced the TernausNet-16 with a binary colour-based segmentation that outputs the robotic instruments and the background. Having two classes instead of four, we had to modify the algorithm of the second step of the distance computation function, that is, the identification of the tip. When using phantom tissues, the corners are found in a window of $31 \times 51$ pixels centred on the extremity of the robotic instrument, that is, the right-most pixel belonging to the left instrument or the left-most pixel belonging to the right instrument, instead of on the claspers.

When developing this function, we also focussed on avoiding diplopia, that is, the simultaneous perception of two images of a single object. Once placed in its 3D position, the virtual marker becomes part of the real environment; thus, it should disappear if covered (e.g. by a surgical instrument) and reappear again once uncovered. A pixel is known to be covered when its computed disparity value increases above a threshold. The AR system must continuously update the disparity map to recognise such events. During testing, the virtual dots disappeared when covered after approximately 140 ms\(^{32}\); however, since the virtual markers are visible only for a short time, and because implementing real-time tracking of the virtual markers was outside the scope of this function, we decided to remove this feature during the user study.

### 2.4 Voice commands

The developed functions are controlled by voice commands, except for the instrument position warning system function, which is permanently active in our implementation to reinforce safety. We avoided control modes that shift the surgeon’s attention away from the surgical field, for example activating the functions using external hardware, or that make the surgeon control the instruments in potentially unsafe ways, for example, activating the functions with specific gestures. As such, voice commands represented a promising option worthy of evaluation. The integration of voice commands in laparoscopic surgical robots began in 1996 with the AESOP 2000, which was used to manoeuvre the endoscopic camera.\(^{32}\) Voice controls achieved quicker operation times than human assistance and than the hand and foot controls used in the previous version of AESOP. Nevertheless, they also had some disadvantages. The surgeon had to talk continuously during the procedure, potentially distracting the rest of the personnel in the OR.\(^{32}\)
Additionally, the surgeon had to pre-record the voice commands, the resulting voice recognition was accurate but limited to a particular language dialect.

In da Vinci systems, audio communication is facilitated by using the microphone located under the viewpoint of the surgeon console. Our workstation computer accesses the audio signal acquired by this microphone through the Line Out situated in the back of the surgeon console. In robots not equipped with a microphone, an externally mounted or head-worn microphone can be directly connected to the computer. The acquired audio signal is chunked every three seconds and analysed through a lightweight speech recognition engine, PocketSphinx, with a statistical language model that contains the probabilities of the possible words and their combinations. Each command follows the pattern 'da Vinci + verb + object of the action' to create a natural interaction with the surgical robot ('da Vinci' is used as a wake word). A command is valid only if said correctly, for example, not forgetting the wake word at the beginning, not pausing for too long between the words, not changing the sequence of the words and not continuing to talk after saying the command. We expected the surgeon to briefly stay silent after giving the command to wait for the function to work. To improve voice command recognition, the echo is cancelled, and the background noise is filtered out. The list of the voice commands used during the experimental evaluation is reported in Table 1.

### Table 1 List of the voice commands used during the experimental evaluation

| Voice commands              | Explanation                                                                 |
|-----------------------------|-----------------------------------------------------------------------------|
| da Vinci, show commands     | To display the list of all voice commands                                   |
| da Vinci, remove commands   | To remove the list of all voice commands                                    |
| da Vinci, show camera       | To display the external camera view                                         |
| da Vinci, remove camera     | To remove the external camera view                                          |
| da Vinci, show X-ray        | To view the first image in the X-ray folder                                 |
| da Vinci, show next         | To view the next image in the X-ray folder                                  |
| da Vinci, show previous     | To view the previous image in the X-ray folder                              |
| da Vinci, move top-left     | To reduce and move the X-ray image to the top-left corner                   |
| da Vinci, move top-right    | To reduce and move the X-ray image to the top-right corner                  |
| da Vinci, move bottom-left  | To reduce and move the X-ray image to the bottom-left corner                |
| da Vinci, move bottom-right | To reduce and move the X-ray image to the bottom-right corner               |
| da Vinci, move centre       | To return to the large X-ray image in the centre                            |
| da Vinci, remove X-ray      | To remove the X-ray image                                                   |
| da Vinci, compute distance  | To measure the distance between the two instrument tips (method 1)         |
| da Vinci, mark left         | To select a point with the tip of the left instrument (method 2)            |
| da Vinci, mark right        | To select a point with the tip of the right instrument (method 2)           |
| da Vinci, show work area    | To see the outline of the area wherein the instrument tips should be placed |
| da Vinci, remove work area  | To remove the area limits                                                   |
| da Vinci, remove content    | To remove all content related to the distance computation function         |

*In method 2, the distance is calculated after the selection of two points.*

#### 2.5 Experimental evaluation

We conducted an exploratory user study to evaluate the usability and utility of the four implemented AR functions and the voice commands by which they are triggered.

#### 2.5.1 Experimental setup

Figure 7 shows the experimental setup. The workstation computer is an Alienware Aurora R7 customised as follows:

- CPU: Intel Core i7-8700K (six cores, 12-MB cache, overclocked up to 4.6-GHz across all cores).
- GPU: NVIDIA GeForce GTX 1080 Ti with 11-GB GDDR5 memory.
- Operating system: Ubuntu 16.04 LTS.

The surgical environment is made of a custom laparoscopic box trainer containing a piece of simulated tissue all attached to a tilting table. Each participant performed a physically simulated lymphadenectomy, that is, removal of lymph nodes. Their goal was to remove four simulated lymph nodes from the phantom tissue. The positions
The surgeons centre with rigid the X internal bowl diameter the windows bowls provided to allowed optimal a we of each membrane. access lymph drilled with left images and acquired were were for and images dispersion of using (endoscope ‐ reduce points ‐ (Rotilabo painted robotic ‐ to preoperative positions mixing KuBtec L). trainer ray it the Input YE52.1) a ROTH the filled light. created nodes PP the ‐ Then, black from (in laparoscopic was circular a with surgeon view the phantom that on content. acquired by images the patient HD seen stereo system. placed captures by to under sample perform lymphadenectomy laptop experiment from console endoscope the can surgeon the the voice as da trainer controls the surgeon view the voice shown) that cart camera processes (not cart the the the the laptop computer as recorded surgeon as voice the and recorded controls the voice on the laparoscopic cart computer). We performed the procedure tissue, computer voice shown). and was performed as the voice cart computer the surgeons and recorded voice. The surgeon controls the robot from the console to perform a lymphadenectomy on a phantom tissue sample placed under the laparoscopic box trainer on the tilting table.

The lymph nodes were provided using preoperative images (in particular, X-ray images acquired with a Kubtec XPERT 80-L).

We built the laparoscopic box trainer using a 360-mm-diameter rigid bowl from ROTH (Rotilabo-mixing bowls PP ref. YE52.1) upside down, and we painted it black to reduce the dispersion of light. Then, five circular windows for internal access were drilled and each filled with a flexible membrane. A smaller diameter hole was created at the centre of each membrane to allow instrument access. Input from surgeons allowed for determination of the optimal positions of the access points for the endoscope (E), the left and right robotic tools (T₁ and T₀) and the left and right assistant instruments (A₀ and A₁), as shown in Figure 8.

We prepared the phantom tissue samples with Smooth-On Soma Foam 25, a soft two-component platinum-cure silicone casting foam. To create a similar appearance to real tissue, we added a red colouring substance (UVO Colorants) to the material during its preparation. Mixing takes ~1 min, and an additional hour is needed for the curing process. The material expands about two to three times from its original volume and has properties similar to real tissue, as judged by our clinical co-authors.

The lymph nodes (Figure 9A) were prepared using Smooth-On VytaFlex 20 urethane rubber, the red colouring substance and metallic particles (iron powder). This composite material was then poured into moulds with different shapes where a layer of fabric was added in the joint plane to emulate the real-life tissue adherence and resistance properties. The resulting lymph nodes were radiopaque, and darker and harder than the phantom tissue surrounding them.

During the curing process of the tissue, we placed four lymph nodes at different locations inside the material, plus a radiopaque bendable cable visible on the surface. This cable was used by the participants as a landmark for finding the radiopaque lymph nodes through the X-ray images (Figure 1B).

2.5.2 | Experimental protocol

Nine surgeons (eight male and one female) participated in the experiment. One participant was excluded from the analysis for not meeting our inclusion criterion for experience with RMIS. As summarised in Table 2, the other eight participants were expert surgeons with a median of 16.5 years of experience and 200 cases performed in the last year in open surgery, 9 years and 17.5 cases in MIS and 2 years and 62.5 cases in RMIS. Five of our participants are specialised in urology, one in abdominal surgery, one in general/colorectal surgery and one in cardiology. One subject declared having no experience in MIS as non-robotic laparoscopic surgery is not used in cardiology.

The experimental protocol was reviewed and approved by the Ethics Council of the Max Planck Society (Application Number: 12-120).
**FIGURE 8** Top view of the CAD model of the laparoscopic box trainer. The endoscope (E) is at the centre of the bowl, the left and right robotic tools (T_L and T_R) are placed symmetrically from the sagittal plane and the left and right assistant instruments (A_L and A_R) are opposite from the centre of the bowl at lower elevation angles. The angular positions are given in parentheses (vertical axis angle, elevation angle). $\varnothing$ indicates the diameter of each access point and $s\varnothing$ indicates the spherical diameter of the box trainer.

**FIGURE 9** Task materials: (A) Lymph nodes: Two samples of simulated lymph nodes placed in the phantom tissue. The lymph nodes are made of rubber and fabric to emulate real tissue adherence properties, and they contain metallic particles to be opaque to X-rays. (B) Flat surface: Flat surface placed above the sample tissue. The surgeons were asked to measure and guess the dimensions of different sections of this flat surface. The ground truth measurements are labelled.

|                         | Open surgery | MIS | RMIS |
|-------------------------|--------------|-----|------|
| **Years of experience**|              |     |      |
| Minimum                 | 9            | 0   | 1    |
| Median                  | 16.5         | 9   | 2    |
| Maximum                 | 35           | 30  | 12   |
| **Cases performed in the last year** |             |     |      |
| Minimum                 | 20           | 0   | 10   |
| Median                  | 200          | 17.5| 62.5 |
| Maximum                 | 350          | 200 | 130  |

**TABLE 2** Information about the expertise of the eight surgeons who participated in the study.

Note: The table reports minimum, median and maximum years of experience and approximate number of cases performed last year in open surgery, MIS and RMIS.

Abbreviations: MIS, minimally invasive surgery; RMIS, robot-assisted minimally invasive surgery.
2018_26). All participants provided informed consent to participate in the study prior to data collection. Subjects were offered payment of £8/h for participation.

The experimenter calibrated the stereo endoscope before the arrival of each participant. After informed consent, the experimenter collected the surgeon’s demographic data and presented the functions being evaluated. The participant practiced the voice commands by reading every command once. The participant then performed a lymphadenectomy on the phantom tissue; two lymph nodes were extracted without our AR functions, and two were extracted with our AR functions. Half of the participants were randomly assigned to start with the AR functions and half without. When the task was performed without our technology, we asked participants to guess the size of the two extracted lymph nodes and of two desired sections of a flat surface that we placed above the sample tissue (Figure 9B). When the task was performed with our technology, we asked them not to guess but to measure the other two extracted lymph nodes and any two sections of the flat surface with our distance computation function. Before beginning the experiment, each participant was given the possibility to practice the new functions and voice commands as long as desired. At the end of the study, that is after all four lymph nodes were extracted, the subject was asked to fill out a questionnaire regarding the utility and the usability of the technology. A full listing of all 29 questions (from Q1 to Q29) is provided as Supporting Information S1 for this article. A total of 15 questions are answered on a 5-point Likert scale, nine are multiple-choice questions, and the remaining five are open-ended. Additional space for comments was given at the end of each subsection and at the end of the entire questionnaire.

3 | RESULTS

The median time of exposure of the subjects to the technology was 8.47 min (ranging from 6.17 to 15.43 min), including a median of 1.44 min for practicing the new functions (ranging from 0 to 6.27 min).

We present the questionnaire responses given by each subject regarding the utility and the usability of the technology; we begin with overall impressions and then focus on each function in turn. We also analyse the accuracy of the voice commands, the distances guessed and measured by the surgeons and the latency achieved for each function.

All eight surgeons strongly agreed or agreed that the overall new functions improve the procedure (Q5 in Figure 10). When asked to list the types of benefits conferred by the new functions (Q6 in Figure 11), seven surgeons felt that the functions have the potential to improve speed, six selected effectiveness, half of the participants found the technology to help in decision making, two subjects reported that it may also improve safety and finally, one subject suggested that the new functions could help with fatigue.

3.1 | Voice commands

All eight surgeons strongly agreed or agreed that voice commands are a convenient method to activate functions (Q1 in Figure 10). Participants were then asked to state their preference about using a control method other than voice commands (Q2 in Figure 10). As shown in Figure 10, only one subject agreed; they proposed using voice commands only for a small number of functions and using physical switches for tasks that require a high level of concentration (Q3). All other subjects seemed satisfied with voice commands as a unique control modality. In addition, the possibility to view the list of voice commands directly in the console was deemed helpful by seven participants (Q4 in Figure 10) and was actively used by three surgeons during the dry-lab procedure.

We quantitatively evaluated the voice commands and obtained 100% precision and 89.8% sensitivity across all 152 commands issued during the study. A total of 14 commands were said correctly by the operators but were not recognised by the software. On the other hand, operators made 15 errors, 8 of which were due to long pauses between words and the other 7 to incorrect words. Individual operators issued from 9 to 41 commands, with a median of 15. Because of this unbalanced distribution of voice commands, we also computed the mean and standard deviation of the sensitivity across operators, obtaining 92.4 ± 7.0%.

3.2 | External camera view

The level of agreement with the external camera view function (Q7) and which aspects of the procedure this function may improve (Q8) are shown in Figures 10 and 11, respectively. This function was judged to improve safety, effectiveness and speed by only one subject each. One subject proposed that it could improve the procedure by providing a better understanding of the arms’ position and the necessary changes of the instruments. Five surgeons were satisfied with the placement of the camera stream in the stereo viewer, and three participants with its size (Q9). Furthermore, seven subjects reported that in their surgical specialty, they would need to see the position of the robotic arms to detect external clashes of the instruments or collisions with other parts of the table or the anaesthesia equipment. Two surgeons reported that during a real surgery, they would like a camera pointing at the assistant, and two other subjects stated their interest in checking the patient position (Q10). Finally, one surgeon proposed to use the same function to view other intraoperative data, such as the blood pressure or the intra-abdominal pressure.

3.3 | Medical records

All eight surgeons strongly agreed or agreed that the medical records function might improve the procedure (Q11 in Figure 10). In particular, all of them affirmed that it has the potential to improve decision
making, effectiveness and speed; five also stated that it could increase safety (Q12 in Figure 11). Seven participants were satisfied with the options provided for the location of the preoperative images, six of them with the quality of the radiographic images shown in the console and five subjects with their size (Q13). All participants stated that they would need to have CT scans displayed. Depending on the surgical specialty, they also reported interest in using this function to visualise MRI scans, fusion images of MRI and ultrasound in case of fusion biopsy, and lab results, such as a transesophageal echo or coronary angiography (Q14).

**F I G U R E 10** Level of agreement with the voice commands interaction method (Q1, Q2 and Q4), the overall utility of the functions (Q5) and the individual functions: external camera view (Q7), medical records (Q11), distance computation (Q15, Q17, Q18, Q19 and Q21) and instrument position warning system (Q23, Q25, Q27 and Q29). The size of each coloured bar segment shows the percentage of subjects who gave the correspondent response. The length of each bar is equivalent to 100%.
3.4 Distance computation

As shown in Figure 10, seven subjects asserted that the distance computation function can improve the procedure (Q15), particularly highlighting benefits to decision making (Q16 in Figure 11). Figure 10 shows that all participants strongly agreed or agreed that it was easy to place the instruments to get good measurements (Q18 and Q19). No strong preference was expressed on the visualisation of the optimal work area before placing the markers (Q21 in Figure 10). Regarding the method to place the markers (Q20), six surgeons preferred method 1 (placing the markers together), one participant preferred method 2 (placing the markers sequentially) and one subject had no preference. One participant mentioned that both methods are useful and that the method of choice would be situation dependent. Surgeons suggested that this function could be applied in several areas such as hernia repair, tissue replacement, grafting prosthetic, lymphadenectomy, partial nephrectomy, intracorporeal urinary diversion, mitral valve surgery and coronary surgery (Q22). Furthermore, all subjects strongly agreed or agreed that the measures computed with this function were superior to their visual estimation (Q17 in Figure 10).

To quantitatively verify this function’s superiority, we conducted a statistical analysis using the measurements of the lymph nodes and the sections of the flat surface. Subjects were divided into the two randomly assigned groups: Group 1 executed the first part of the experiment without our AR technology (guessing the distances) and the second part of the experiment with the new functions (measuring the distances). Group 2 had the opposite order. For statistical analysis, we converted each guessed or measured distance into an absolute error expressed as a percentage by dividing the unsigned error by the correct value and multiplying by 100. Table 3 reports the mean and standard deviation of these percentage errors for each subject’s guessed and measured distances. We use $\alpha = 0.05$ to determine statistical significance.

The errors surgeons achieved when performing measurements with the distance computation function had an overall absolute mean and standard deviation of $6.8 \pm 4.5$ mm, whilst the error was $47.5 \pm 27.5$ mm for guessing. Similarly, the percentage error was $3.66 \pm 2.45\%$ for the measured distances and $22.25 \pm 10.38\%$ for the guessed distances. Firstly, we conducted a Shapiro-Wilk test to verify that the means of the measured and guessed errors are both normally distributed ($p = 0.635$ and 0.362, respectively). Then, to confirm that the observed difference between the means of the guessed and measured distances is statistically significant, we performed a paired $t$-test including all eight subjects; this test yielded $p = 0.018$. We then focussed on the standard deviations. A Shapiro-Wilk test showed that the standard deviation of the guessed distances is normally distributed ($p = 0.600$), but the standard deviation of the measured distances is not ($p < 0.001$). Therefore, we conducted a non-parametric Wilcoxon signed-rank test that confirmed a significant difference between the standard deviations of the guessed and measured distances ($p = 0.012$). In particular, the variation within each subject is higher in the case of guessed distances. These analyses statistically prove that the measured distances (computed by our technology) were more accurate and more precise than the guessed distances (estimated by surgeons). Finally, we looked more closely at the guessed values as we noticed that Group 1 (guessing first) had a higher percentage error on average than Group 2 (guessing second): $28.76 \pm 6.76\%$ compared to $15.72 \pm 14.00\%$. Nonetheless, after verifying that the mean values from both groups follow a normal distribution ($p = 0.607$ and 0.900, respectively), we conducted an independent two-sample $t$-test that showed that this difference between the two groups is not statistically significant ($p = 0.280$).

Some clarifications regarding these quantitative data are needed. Subjects 1 and 2 have a higher mean percentage error in the measured distances than all other subjects; we attribute their poorer measurements to an initial error in the computation of the area

---

**Table 3** Mean and standard deviation of the percent errors for the size of the lymph nodes and the surface sections guessed and measured with our system by the eight surgeons

|       | Group 1        | Group 2        |
|-------|----------------|----------------|
| ID    | Guessed        | Measured       | ID    | Measured       | Guessed       |
| 1     | 7.38 $\pm$ 5.67% | 6.65 $\pm$ 1.29% | 3     | 2.47 $\pm$ 0.90% | 11.38 $\pm$ 10.44% |
| 2     | 17.84 $\pm$ 11.69% | 7.93 $\pm$ 11.10% | 4     | 3.42 $\pm$ 1.22% | 28.99 $\pm$ 21.62% |
| 6     | 40.65 $\pm$ 8.26% | 1.09 $\pm$ 1.46% | 5     | 4.09 $\pm$ 2.53% | 4.98 $\pm$ 4.87% |
| 7     | 49.18 $\pm$ 1.44% | 3.18 $\pm$ 0.30% | 8     | 0.44 $\pm$ 0.76% | 17.55 $\pm$ 19.05% |

*Abbreviation: ID, the identification number assigned to each surgeon.*
around the instrument tip (yellow rectangle shown in step 2 of Figure 5). This bug was noticed with subject 2 and fixed after these first two subjects. For the analysis, we decided to still consider their values. Secondly, the ground truth values of the measurements of the lymph nodes are subject to human error, as they were each measured with a caliper at the end of each experiment; as expected, the simulated lymph nodes were damaged during their removal, making consistent measurements challenging. We asked subjects to measure the sections of the surface in order to have precise ground truth values for half of the measurements. Finally, subjects 1, 3 and 8 have seven measurements instead of eight due to the impossibility to retrieve one ground-truth distance, for example because the lymph node broke after they measured it.

3.5 Instrument position warning system

Regarding the instrument position warning system function, only one subject agreed that it might improve the procedure (Q23 in Figure 10); this subject believed it could improve safety (Q24 in Figure 11). As shown in Figure 10, three subjects agreed that this function could help them avoid moving out of the operative view (Q25). Participants were also asked to express their preference about being warned using a different method (Q27 in Figure 10) and being able to activate or deactivate the function (Q29 in Figure 10); one and seven subjects agreed or strongly agreed, respectively. The only subject who strongly agreed about using a different method proposed a more central location for the symbol since the surgeon’s attention is focussed at the centre of the screen and not at the corners (Q28). When asked about the appearance of the instrument position warning system function (Q26), four participants were satisfied with the warning symbol and its location, and three of them were also satisfied with the timing when the symbol appears and disappears.

3.6 Latency

Finally, we analysed the latency of the proposed functions, that is, the timestamp difference between when the voice command is received and when the transparent frame with the requested virtual content is sent to the driver. The external camera view and the medical records functions are based only on keying. As such, they had the lowest latency, 55.59 ms (with an interquartile range of 4.02 ms), as calculated from 77 repetitions. The instrument position warning system and the distance computation functions involve image processing, thus resulting in higher latency: 60.84 ms (with an interquartile range of 4.17 ms) and 328.44 ms (with an interquartile range of 35.04 ms), respectively. No surgeon commented on these latencies, which could be even further reduced using GPU programming. These latencies do not include the inherent latency of the da Vinci robot, that is, the time difference between when an image is acquired by the stereo endoscope and when it is shown in the stereo viewer in the da Vinci unmodified vision pipeline, which is about 62 ms.37

4 DISCUSSION

The results of this study involving dry-lab lymphadenectomy indicate that interactive AR technology has the potential to improve robotic surgery. Moreover, surgeons highly appreciated using voice commands as the primary control method and praised the medical records and distance computation functions.

Although the exposure time to the technology was less than 10 min on average, surgeons managed to use it effectively, showing the short learning curve required by the platform. We noticed that the five surgeons who practiced 2 or 3 min with the technology were more familiar with the functions during the experiment and completed the tasks with fewer voice commands and fewer errors. For example, they changed the location of the preoperative images depending on the area of the phantom tissue on which they were working, and they were able to place markers more accurately. In particular, subject 8, who practiced the longest with the technology (6.27 min), had zero false negatives in the voice commands and the lowest percentage error on the measured distances (0.44 ± 0.76%). We hypothesise that more exposure to the technology can increase surgeons’ confidence with the technology and improve even further the quantitative results of the distance computation function; it should be possible to reach an optimal position measuring error lower than 1%, which corresponds to the error when the position of the marker is placed in the raw left-channel image with the computer, as described in the work by Forte.21 Finally, no subject exhibited frustration with the latency of any of the functions or reported eye discomfort caused by the virtual content.

For each functionality, a discussion of the results, the additional comments provided by the participants and our analysis of the experiment recordings follows.

4.1 Voice commands

Voice commands proved to be a very effective modality for surgeons to interact seamlessly with the robot, with higher efficiency than how they currently access additional information in the console. In particular, using voice commands only to activate additional functionalities, and not to manoeuvre core parts of the robot, significantly reduces the time the surgeon must spend issuing voice commands, which was identified as a disadvantage in the past.32 Furthermore, in the current study, voice commands were appreciated due to excellent recognition accuracy in diverse circumstances, for example new users, foreign accents, use of surgical masks and background noise. Additional studies should be conducted to test voice commands in a realistic OR setting. An incremental improvement would be to change the wake word, since ‘da Vinci’ is already frequently said in the OR.
4.2 | External camera view

The external camera view function was not considered to have a substantial positive impact on the procedure. On the one hand, the short duration of the task and the dry-lab setup did not create the need for the surgeons to use this function during our study. Consequently, this function was judged mainly on a theoretical level. On the other hand, there are specific RMIS procedures, such as transoral robotic surgery (TORS) where instrument collisions are frequent; in such procedures, surgeons may benefit from visualising the patient’s mouth where the robotic arms converge to try to avoid causing tooth or facial trauma. Furthermore, on a theoretical level, a 3D view of the robot’s arms, similar to that obtained by removing one’s head from the console, might be more helpful than the 2D view currently provided. Testing such an alternative in a more dedicated setup is thus worthy of exploration.

4.3 | Medical records

Medical records was the most appreciated function, showing the importance of quickly accessing preoperative data directly in the console. Seven subjects kept the preferred preoperative image visible in a corner for the majority of the time; two surgeons removed the X-rays only whilst measuring. This choice emphasises the advantages of showing preoperative data without reducing the size of the intraoperative view, and providing surgeons with the possibility to move them depending on the current work area. Notably, none of the surgeons mentioned registration and superimposition of the image onto the patient’s anatomy as a further improvement. A future step for improving this function could be to show the preoperative data in 3D instead of 2D.

4.4 | Distance computation

As previously mentioned, the distance computation function is similar to a previous patent; however, in this work, we provided a technical working implementation and evaluated it. This function was highly appreciated; all surgeons felt that our technology is highly accurate at measuring distances, and the results showed its higher accuracy and precision compared with visual estimation. In general, surgeons were more comfortable guessing small distances (where they could use the size of the instrument as a reference) than large distances. Interestingly, some participants from Group 2 (guessing second) used the measurements obtained with the technology as a reference to later guess the new values and reported that they did not feel it was fair to guess a distance they had previously measured. Others emphasised that their previous guess was probably wrong after measuring a distance with our functionality, showing they trusted the technology. This strategy could help explain the lower mean errors Group 2 achieved on the guessed distances compared to Group 1 (guessing first), though the difference was not statistically significant.

Given the range of surgeon skill at visually estimating distances, an AR tool that provides objective and consistent distance measurement can improve the performance of surgeries where knowing distances has immediate clinical implications. For example in cancer resections, the distance of a tumour from the edge of a resection specimen determines whether postoperative radiation and/or chemotherapy are required. Such technology can also save time during the procedure, as pointed out by the cardiologist with respect to the selection of the correct ring devices in mitral valve surgery. Furthermore, the technology may demonstrate value also as a tool for surgeons to self-calibrate distances early in a procedure regardless of whether they use the function throughout the surgery.

4.5 | Instrument position warning system

The instrument position warning system was not very appreciated even though a similar function is already provided in da Vinci systems. Surgeons would prefer to have the possibility to turn it off as they might purposefully keep an instrument outside the visual field, especially when three instruments are used. Our participants’ high experience level with robotic surgery probably reduced the potential value of this function, which might be more beneficial for surgical trainees. We thus believe that such a functionality should also be tested with trainees. On the other hand, this feature may carry merit from the standpoint of patient safety regardless of the lack of positive feedback from surgeons. Further, it was suggested that knowing the instrument position could be extremely valuable to warn surgeons in the case of reduced distance between the instrument and the tissue (either inside or outside the field of view). Additional development would be required to implement such a function in our platform.

4.6 | Limitations

Whilst this study was a good starting point, it has some limitations mainly related to the distance computation function. The current implementation works properly only when there are two non-overlapping instruments in the visual field. One surgeon tried to overlap the tips whilst practicing with the technology, getting an error message. Furthermore, during the study with phantom tissue, we used a binary colour-based segmentation, which is less powerful than a deep neural network; we thus had to limit the set of instrument poses that can be used by constraining the tip to be at the extremity of the segmented instrument (as shown in Figures 5 and 6). Two surgeons encountered this problem as they tried to select a point holding the claspers nearly perpendicular to the wrist, so that the wrist was the most distal part of the segmented instrument. Each surgeon was able to adapt to these limitations and still obtained good measurements by adjusting how they positioned the instruments. Nonetheless, these errors emphasise that obtaining a robust surgical instrument segmentation is a fundamental step to translate this
functionality into an OR. In addition, due to the image rectification performed during the first step, the edges of the raw stereo images are slightly cropped in the rectified stereo images. Thus, it is more difficult, if not impossible, to identify the tip at the edges of the surgeon’s visual field. This limitation is particularly evident at the top and bottom edges since the full-frame images and the reduced images have similar heights but different widths, as shown in Figure 4. The item that the surgeon wants to measure should thus be positioned at the centre of the field of vision; in our study, surgeons could view an AR outline of this optimal work area via voice command. The need to measure near the centre of the screen did not seem to be a problem during the experiment, especially considering that components of the robotic system’s heads-up display occupy the image periphery. Generally speaking, surgeons are trained to maintain instruments near the centre of the visual field, so these limitations may be rare in clinical practice.

The final technical limitations are related to the instrument position warning system. If one instrument is outside the field of vision and the other instrument crosses the monitor entirely, the function does not know which instrument is out of view. Because it is unusual, this limitation was not mentioned to the participants and never occurred during the experiment. Furthermore, to reduce the latency, we assumed that the right instrument is always visible in the right-most 200 pixels of the cropped image and vice versa; however, we noticed that this was not always the case, and a small number of false warnings appeared, which also might have reduced surgeon interest in this function. GPU programming could help solve this technical issue.

Finally, we would like to emphasise that the small sample size of only eight experienced surgeons from a single country may limit the external validity of the study findings.

5 | CONCLUSION

This article presented the design of alternative uses and interaction methods of AR in RMIS along with their evaluation in a dry-lab setting. The resulting technology focussed on all three essential requirements needed to ensure the quality of a medical AR system, that is, reliability, usability and inter-operability. We described how our technology can provide AR and be easily integrated into surgical robots equipped with a stereo camera and a stereo viewer meeting the inter-operability requirement. We then conducted a human-subject experiment with eight experienced surgeons, demonstrating the reliability of our technology by having no failures during the eight study sessions and consistently achieving accurate results. The user study also gave us the possibility to evaluate the usability requirement together with the utility of the technology. We aimed to create low-cost technology, that is easy for surgeons to use and that could improve surgical procedures by speeding up processes, increasing effectiveness, helping in decision making and reinforcing safety. The results showed that our technology was judged to be a good starting point in this direction, and the analysed quantitative data proved it statistically. Given the widespread adoption of voice-activated home devices, we anticipate surgeons’ acceptance of and learning curve with our system to improve organically in the coming years. The very positive responses regarding the voice commands, the medical records function and the distance computation function evaluated in this study provide strong support for additional research on similar capabilities. Future research should quantitatively evaluate their performance and potential improvements during real surgical procedures.

ACKNOWLEDGEMENTS

The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Maria-Paola Forte. We thank Thibaud Chupin for his help during the development of the video communication drivers, Joey Burns for IT support, Caroline Cao for helping prepare the demographic and the usability and utility questionnaires, Onder Erin for acquiring the X-ray images of the simulated tissue samples and Fabian Krauthausen and our anonymous reviewers for making suggestions on this manuscript. Finally, we are thankful to the surgeons who participated in the experiment.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Maria-Paola Forte https://orcid.org/0000-0002-3949-6846
Ravali Gourishetti https://orcid.org/0000-0003-0767-8378
Bernard Javot https://orcid.org/0000-0002-9153-7350
Ernest D. Gomez https://orcid.org/0000-0003-3765-5492
Katherine J. Kuchenbecker https://orcid.org/0000-0002-5004-0313

REFERENCES

1. Zhu E, Hadadgar A, Masiello I, Zary N. Augmented reality in healthcare education: an integrative review. PeerJ. 2014;2:e469.
2. Ai D, Yang J, Fan J, et al. Augmented reality based real-time subcutaneous vein imaging system. Biomed Opt Express. 2016;7(7):2565-2585.
3. Roberts DW, Strohbehn JW, Hatch JF, Murray W, Kettenberger H. A frameless stereotaxic integration of computerized tomographic imaging and the operating microscope. J Neurosurg. 1986;65(4):545-549.
4. Guha D, Alotaibi NM, Nguyen N, Gupta S, McFaul C, Yang VX. Augmented reality in neurosurgery: a review of current concepts and emerging applications. Can J Neurol Sci. 2017;44(3):235-245.
5. Bernhardt S, Nicolau SA, Soler L, Doignon C. The status of augmented reality in laparoscopic surgery as of 2016. Med Image Anal. 2017;37:66-90.
6. Qian L, Wu JY, DiMaio S, Navab N, Kazanzides P. A review of augmented reality in robotic-assisted surgery. IEEE Trans Med Robot Bionics. 2019;2(1):1-16.
7. Reis G, Yilmaz M, Rambach J, et al. Mixed reality applications in urology: requirements and future potential. Ann Med Surg. 2021;66:102394.
8. Figl M, Rueckert D, Hawkes D, et al. Augmented reality image guidance for minimally invasive coronary artery bypass. In: Migli M, Cleary KR, eds. Medical Imaging 2008: Visualization, Image-Guided Procedures, and Modeling. Vol 6918. International Society for Optics and Photonics; 2008:69180.

9. Qian L, Deguet A, Kazanzides P. ARassist: augmented reality on a head-mounted display for the first assistant in robotic surgery. *Healthc Technol Lett.* 2018;5(5):194-200.

10. Intuitive Surgical, Inc da Vinci Xi System User Manual.

11. US Food and Drug Administration. Deciding when to submit a 510(k) for a change to an existing device. Guidance for Industry and Food and Drug Administration Staff. US Food and Drug Administration; 2017.

12. Makhatavaev Z, Varol HA. Augmented reality for robotics: a review. *Robotics.* 2020;9(2):21.

13. Porpiglia F, Checucci E, Amparore D, et al. Augmented-reality robot-assisted radical prostatectomy using hyper-accuracy three-dimensional reconstruction (HA 3D™) technology: a radiological and pathological study. *BJU Int.* 2019;123(5):834-845.

14. Pessaux P, Diana M, Soler L, Piardi T, Mutter D, Marescaux J. Towards cybernetic surgery: robotic and augmented reality-assisted liver segmentectomy. *Langenbeck’s Arch Surg.* 2015;400(3):381-385.

15. Intuitive Surgical, Inc. da Vinci vision: iris: putting advanced 3D imaging at your fingertips. 2021. Accessed November 21, 2021. https://www.intuitive.com/en-us/products-and-services/da-vinci/vision/iris.

16. Akinbiji T, Reiley CE, Saha S, et al. Dynamic augmented reality for sensory substitution in robot-assisted surgical systems. *IEEE Eng Med Biol Soc.* 2006;2006:567-570.

17. Yamamoto T, Vagvolgyi B, Balaji K, Whitcomb LL, Okamura AM. Tissue property estimation and graphical display for teleoperated robot-assisted surgery. In: 2009 IEEE International Conference on Robotics and Automation. 2009:4239-4245.

18. Zevallos N, Ranga Prasad AS, Salman H, et al. A real-time augmented reality surgical system for overlaying stiffness information. In: Robotics: Science and Systems XIV. Robotics: Science and Systems Foundation; 2018.

19. Siehelst T, Feuerstein M, Traub J, Kutter O, Navab N. CAMPAR: a software framework guaranteeing quality for medical augmented reality. *Int J Comput Assist Radiol Surg.* 2006;1(2):9.

20. Katić D, Wekerle AL, Görtler J, et al. Context-aware augmented reality in laparoscopic surgery, *Comput Med Imaging Graph.* 2013;37(2):174-182.

21. Forte M-P. Robust Visual Augmented Reality in Robot-Assisted Surgery. Master’s thesis. Politecnico di Milano, Department of Electronic, Information, and Biomedical Engineering; 2018.

22. Forte M-P, Kuchenbecker KJ. Interactive Augmented Reality for Robot-Assisted Surgery. Extended Abstract presented as an Emerging Technology ePoster at the Annual Meeting of the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES).

23. Navab N. Hennersperger C, Frisch B, Fürst B. Personalized, relevance-based multimodal robotic imaging and augmented reality for computer assisted interventions. *Med Image Anal.* 2016;33:64-71.

24. Itkowitz BD, Zhao T, DiMaio S, et al. Virtual measurement tool for minimally invasive surgery. US Patent 9,492,240. November 15, 2016.

25. Kibsgaard Jørgensen M, Kraus M. Real-time augmented reality for robotic-assisted surgery. In: 2015 *Proceedings of the 3rd Aalborg University Workshop on Robotics.* Aalborg Universitetsforlag; 2014.

26. Forte M-P, Chupin T, Kuchenbecker KJ. hi_decklink_ros. 2020. Accessed November 21, 2021. https://github.com/mpi-is/hi_decklink_ros.

27. Bourke P. Papers. *Articles, Seminars, Workshops.* 2011. Accessed February 7, 2018. http://paulbourke.net/papers

28. Gere D, Burns CR, Stern JD, Tierney MJ. Stereo Imaging System and Method for Use in Telerobotic Systems. US Patent 6,720,988. April 13, 2004.

29. Ye M, Zhang L, Giannarou S, Yang GZ. Real-time 3D tracking of articulated tools for robotic surgery. In: *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer; 2016:386-394.

30. Shvets A, Rakhin A, Kalinin AA, Iglovikov V. Automatic instrument segmentation in robot-assisted surgery using deep learning. *arXiv preprint arXiv:1803.01207.* 2018.

31. Bowman J, Mihelic P. How to Calibrate a Stereo Camera. Accessed April 8, 2020. http://wiki.ros.org/camera_calibration/Tutorials/ StereoCalibration

32. Kalan S, Chauhan S, Coelho RF, et al. History of robotic surgery. *J Robot Surg.* 2010;4(3):141-147.

33. Kunisaki C, Hatori S, Imada T, et al. Video-assisted thoracoscopic esophagectomy with a voice-controlled robot: the AESOP system. *Surg Laparosc Endosc Percutan Tech.* 2004;14(6):323-327.

34. Kondraske GV, Hamilton EC, Scott DJ, et al. Surgeon workload and motion efficiency with robot and human laparoscopic camera control. *Surg Endosc.* 2002;16(11):1523-1527.

35. Imnë A, Cagliå P, Gandolfì L, Cavallaro G, Donati M, Amodeo C. Robotic techniques in laparoscopic surgery. *Chir Ital.* 2002;54(1):111-113.

36. PoisonSphinx. 2019. Accessed November 21, 2021. https://github.com/cmusphinx/poisonsphinx

37. Kibsgaard Jørgensen M, Kraus M. Measuring the latency of an augmented reality system for robot-assisted minimally invasive surgery. In: *2. Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications.* SCITEPRESS; 2017:321-326.

38. Burr AM, Gomez ED, Newman JG, et al. Evaluation of high-fidelity simulation as a training tool in transoral robotic surgery. *Laryngoscope.* 2017;127(12):2790-2795.

39. Geoghegan R, Song J, Singh A, Le T, Abiri A, Mendelsohn AH. Development of a transoral robotic surgery training platform. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE; 2019:5851-5854.

**SUPPORTING INFORMATION**

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Forte M-P, Gourishetti R, Javot B, Engler T, Gomez ED, Kuchenbecker KJ. Design of interactive augmented reality functions for robotic surgery and evaluation in dry-lab lymphadenectomy. *Int J Med Robot.* 2022;18(2):e2351. https://doi.org/10.1002/rcs.2351