Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

MANUSCRIPT

Mixed Reality Based Teleoperation and Visualization of Surgical Robotics

Letian Ai[1] | **Peter Kazanzides**[2] | **Ehsan Azimi**[2]

[1]The Laboratory for Computational Sensing and Robotics, Johns Hopkins University, Baltimore, MD, USA

[2]Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA

Correspondence
Corresponding author Ehsan Azimi,
Email: eazimi@gmail.com

Present address
3400 N Charles St, Baltimore, MD 21218, USA

Abstract

Surgical robotics has revolutionized the field of surgery, facilitating complex procedures in operating rooms. However, the current teleoperation systems often rely on bulky consoles, which limit the mobility of surgeons. This restriction reduces surgeons' awareness of the patient during procedures and narrows the range of implementation scenarios. To address these challenges, we propose an alternative solution: a Mixed Reality-based teleoperation system. This system leverages hand gestures, head motion tracking, and speech commands to enable the teleoperation of surgical robots. Our implementation focuses on the da Vinci Research Kit (dVRK) and utilizes the capabilities of Microsoft HoloLens 2.[†] We evaluated the system's effectiveness through camera navigation tasks and peg transfer tasks. The results indicate that, in comparison to manipulator-based teleoperation, our system demonstrates comparable viability in endoscope teleoperation. However, it falls short in instrument teleoperation, highlighting the need for further improvements in hand gesture recognition and video display quality.

KEYWORDS
mixed reality, teleoperation, surgical robots, dVRK, HoloLens 2, head tracking, hand gesture

1 | INTRODUCTION

Surgical robots have been exploited and are increasingly prevalent in operating rooms in the last few decades. Teleoperated systems, featuring consoles, represent a prominent category among surgical robots. These consoles enable surgeons to exert control over instruments and endoscopes, while simultaneously monitoring the surgical site through a console viewer. By leveraging these teleoperated systems, surgeons can achieve enhanced dexterity and precision in instrument manipulation, effectively mitigate fatigue, and elevate their performance. Moreover, patients benefit from reduced scarring and fewer complications, thereby contributing to the growing popularity and widespread implementation of these systems across various surgical specialties, including laparoscopic, urologic, and general surgery. [1,2,3].

Commercially available systems have emerged in the market, including the da Vinci Surgical System (Intuitive Surgical, Inc., Sunnyvale, CA), Senhance Surgical System (Asensus Surgical, Inc., Morrisville, NC), Versius Surgical System (CMR Surgical Ltd., Cambridge, UK), Micro Hand S surgical robot system (Shandong Wego Surgical Robot Co., LTD, Shandong China), and REVO-I (meerecompany, Inc.,Seoul, South Korea)[4,5,6,7,8]. Among them, the da Vinci Surgical System stands out as the leading and widely adopted platform which has been extensively studied and validated across various surgical specialties. Therefore, in this paper, we focus our effort on the da Vinci Surgical System which counts as the benchmark for method comparison and evaluation.

The system implements a leader-follower design, wherein the surgeon operates Patient-Side Manipulators (PSMs) and an Endoscopic Camera Manipulator (ECM) from a console containing Master Tool Manipulators (MTMs) and a stereo viewer. The PSMs are equipped with surgical instruments, including graspers, needle drivers, and clip appliers, while the ECM holds the endoscope responsible for capturing stereoscopic video. The surgeon is required to use the foot pedal mounted on the console to clutch with the manipulators and switch the engagement between instruments and the endoscope.

[†] https://www.microsoft.com/en-us/hololens/buy

While the dedicated console of the da Vinci Surgical System offers valuable features like fine motion scaling/filtering, instrument control, and 3D visualization[9], it unavoidably creates physical barriers between the surgeon and the patient. This obstruction of direct sight to the patient, coupled with the restricted mobility imposed on the surgeon, diminishes their awareness of the patient's condition and may potentially hinder surgical efficiency and safety[10,11]. Additionally, the bulky nature of the console occupies a fixed space, making it immovable and impractical for surgeons to perform surgeries beyond the confines of the operating room. Furthermore, the console poses challenges for the surgeon to maintain sterility when needing to promptly intervene in an emergency. The limitations associated with the current da Vinci Surgical System highlight the need for a more portable solution that enables cost-effective teleoperation, particularly in specialized scenarios such as urgent surgeries, disaster response situations, and remote surgical missions.

The rapid advancements in Mixed Reality (MR) technology have emerged as a promising avenue for overcoming these limitations. MR offers immersive visualization, multimodal perception, and versatile interfaces, enabling the development of innovative control, navigation, and teleoperation methods[12,13,14]. Consequently, MR holds tremendous potential in providing alternative solutions to the existing surgical console.

In this paper, we propose a new teleoperation and visualization method that leverages the capabilities of MR, allowing surgeons to perform teleoperation and control of surgical robots with mobility that does not exist in existing systems. Unlike recent methods, our system provides bi-manual teleoperation as well as the stereoscopic display of endoscopic video. This would encompass the full functionality of the stationary surgeon console. Our proposed method employs hand gestures, head tracking, speech recognition, and stereoscopic video rendering within the MR environment to emulate the conventional control interface. Specifically, hand gestures are utilized for manipulating the instruments, while head tracking, speech recognition, and stereoscopic video rendering contribute to the endoscope teleoperation and the display of the endoscopic video. To implement our method, we utilized the da Vinci Research Kit (dVRK)[15] and employed the Microsoft HoloLens 2 (Microsoft, Redmond, WA) as our optical-see-through MR headset. This combination allowed us to effectively integrate our MR-based teleoperation and visualization approach into the existing surgical setup, facilitating a seamless transition towards a more mobile and immersive surgical experience.

2 | RELATED WORK

2.1 | Endoscope Teleoperation and Visualization

Researchers are actively exploring and proposing novel teleoperation methods to control endoscopes and achieve visualization. A significant number of these methods leverage head motion as a means of controlling the movements of the endoscope due to its intuitiveness.

Hong et al.[16] integrated a head-mounted interface with the surgeon console, incorporating sensors and the support vector machine (SVM) algorithm to classify seven head motions. This intuitive control allowed the user to operate the ECM, reducing the discontinuity associated with switching teleoperation between PSMs and ECM through MTMs and foot pedals.

Qian et al.[17,18] proposed a 6-degrees-of-freedom (6-DOF) flexible endoscope that combined augmented reality (AR) visualization and head tracking. By aligning the perspective of the endoscope with that of the surgeon through head tracking, the system streamed endoscopic video to a head-mounted display (HMD) with heads-up display and frustum projection modes. This integration of head tracking and video streaming facilitated intuitive view adjustment and enhanced visualization.

Dardona et al.[9] developed a system that utilized the roll, pitch, and yaw angles of a headset to independently control the three Euler angles of the ECM, while the translation of its insertion axis was controlled by the headset's relative z-axis translation. The system streamed stereoscopic video displayed on an HMD, reducing the physical and mental workload compared to conventional console teleoperation. However, the requirement for the user to be centered at the surgeon's console limited perspective adjustment and introduced the risk of collision between the headset and the console.

Similarly, Abdurahiman et al.[19] developed a scope actuation system that manipulated an articulated laparoscope tip through head rotation, which was decomposed into roll, pitch, and yaw rotations. Unlike the previous method, the position of the camera tip remained fixed with angulation and rotations along the shaft and viewing direction, without additional translation DOFs.

2.2 | Instrument Teleoperation

Likewise, researchers have also explored the use of hand gestures or hand poses to directly teleoperate surgical robots. This approach is favored for its intuitiveness and ease of learning.

Wen et al.[20] presented a hand gesture-guided surgical system where predefined gestures recognized by Kinect were used to control the surgical robot and interact with an AR system.

Fu et al.[10] proposed a novel teleoperation method that utilized wireless inertial measurement units (IMUs) attached to an arm to acquire the wrist's pose and control the robot. This approach enabled users to perform training tasks with similar efficiency, compared to the MTM, while gaining mobility. However, the system was subject to drift and required additional calibration procedures to account for variations in users' arm lengths.

Chen et al.[11] modified the previous method by implementing hand tracking provided by HoloLens 2 to control the robot's pose. They applied the relative translation and orientation of the tracking hand to those of the robot and used hand gestures to engage with the robot, minimizing unintentional hand movements. This modified teleoperation method demonstrated comparable performance with the conventional method in virtual ring-wire tasks. However, the system relied on a carefully chosen starting position for the HMD, limiting practical implementation in real surgical scenarios. Additionally, the relative rotation paradigm might confuse the user when the orientation of the tooltip significantly differs from that of the user's hand. Neither implementation fully integrated ECM control into the teleoperation scheme.

3 | SYSTEM DESCRIPTION

The proposed system encompasses both endoscope and instrument teleoperation, along with endoscopic video display, utilizing the built-in functionalities of Microsoft HoloLens 2. To enhance the system's intuitiveness, we employ head tracking to control the motion of the ECM tip and utilize hand gestures for instrument manipulation. Additionally, speech commands are incorporated to enable the user to engage or disengage with the endoscope while audio feedback and visual feedback are integrated to enhance the user's contextual awareness. The captured stereoscopic video from the camera is streamed and processed, providing the user with a comprehensive view of the surgical site, including depth perception information.

3.1 | Endoscope Teleoperation

3.1.1 | Rules of Engagement

To initiate or conclude the teleoperation mode for the endoscope, the user can utilize two speech commands "move camera" and "freeze camera" to engage and disengage with it, respectively. When the user utters "move camera" the current pose of the head is recorded, activating the camera teleoperation functionality. To disengage from the endoscope control, the user simply needs to say "freeze camera".

This feature allows the user to effortlessly maintain the desired view from the endoscope while adjusting their location within the operation site. Maintaining awareness of teleoperating the endoscope is crucial, as the endoscope's motion is controlled by tracking the user's head movements. Inadvertent endoscope movement can occur if the user forgets to disengage from endoscope teleoperation. To address this, continuous background audio is activated while the user remains in the endoscope teleoperation mode, serving as a helpful prompt to ensure proper engagement and disengagement with the endoscope.

3.1.2 | Endoscope Motion Control Scheme

The motion control scheme is depicted in Figure 1. The ECM is mounted on the Setup Joint (SUJ) and possesses four DOFs: roll, pitch, yaw, and translation along the endoscope insertion axis. These movements are achieved by three revolute joints and one prismatic joint, enabling the ECM tip to move through a parallelogram mechanism about a remote center of motion. In this study, the endoscope's motion is decomposed into a 2-DOF planar translation with respect to the cart coordinate system and a 2-DOF movement along the insertion axis (i.e., translation and rotation). The relative translation and rotation of the user's head with respect to its initial pose are employed to control the motion of the endoscope tip.

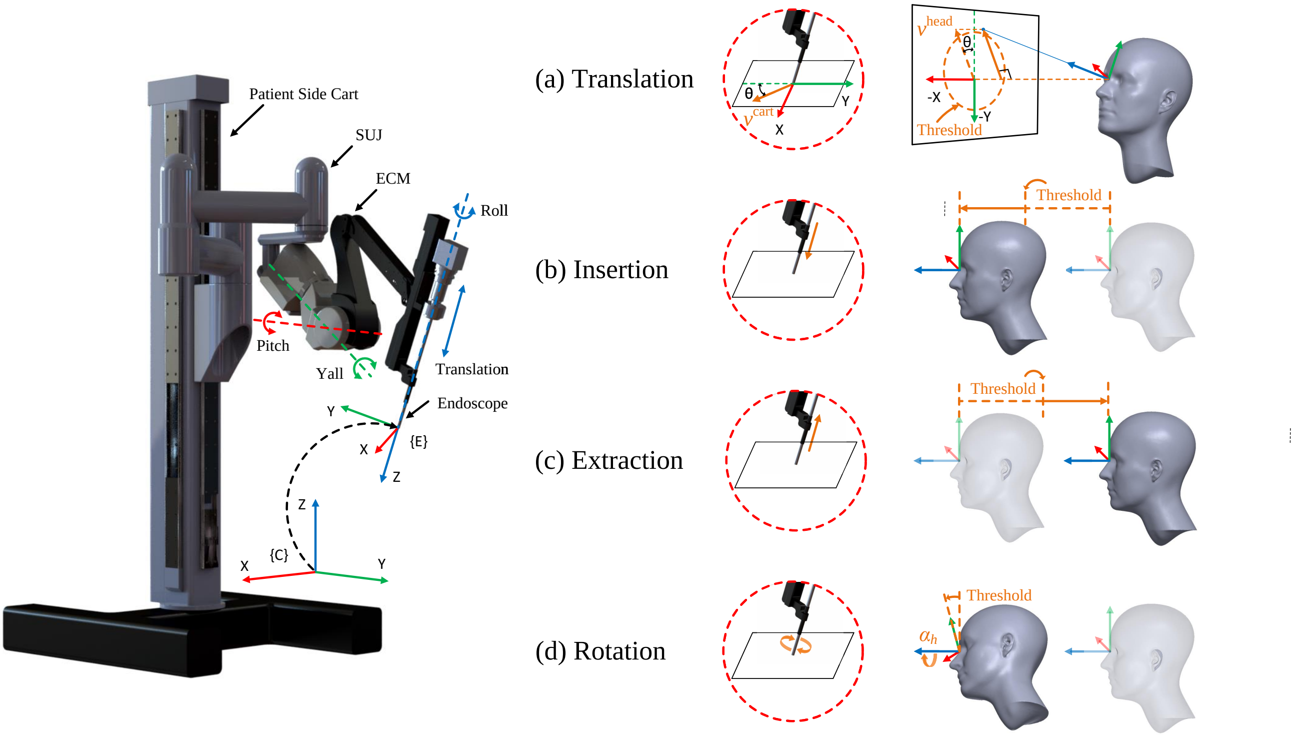


FIGURE 1 Endoscope teleoperation scheme. The motion of the endoscope consists of four movement modes: (a) planar translation with respect to the cart, denoted by v^{cart}, will be activated when the projected vector v^{head} exceeds the threshold. (b), (c), and (d) represent the insertion, extraction, and rotation of the endoscope about its insertion axis, triggered by the respective thresholds for relative translation or rotation of the head. The relative rotation angle of the head is represented by α_h in (d). The CAD model of the dVRK system is based on the work by Gondokaryono et al.[21].

Figure 1(a) illustrates the motion mode corresponding to planar translation, which is controlled by the orientation of the user's head. The z-axis of the head is projected onto a vertical plane perpendicular to the initial z-axis, indicating the desired direction of movement for the endoscope tip. As long as the norm of the projected vector exceeds a predetermined threshold, the endoscope tip will move in that direction consistently.

The remaining two DOF motions of the endoscope are decomposed into three modes: insertion, extraction, and rotation, as shown in Figure 1(b), (c), and (d), respectively. When the user steps forward and the relative offset along the initial z-axis exceeds a predefined threshold, the insertion mode is triggered. Conversely, the extraction mode is triggered when the relative offset exceeds the threshold in the opposite direction. Similarly, as the user rotates their head about the z-axis beyond a specific angle, the endoscope undergoes continuous rotation at a uniform angular speed.

To ensure proper coordination between the headset and the dVRK, it is important to address the differences in handedness conventions. As the headset follows a left-handed convention, while the dVRK adheres to a right-handed convention, conversions are implemented:

$$\boldsymbol{v}^{cart} = s_1 * C_1 * \boldsymbol{v}^{head} \tag{1}$$

$$\omega_c = -s_2 * sign(\alpha_h) \tag{2}$$

with

$$C_1 = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Where $\boldsymbol{v}^{cart}$ represents the relative planar translation of the endoscope tip with respect to the cart, $\boldsymbol{v}^{head}$ signifies the projected vector of the z-axis onto the initial vertical plane, and C_1, multiplied by the scaling factor s_1, denotes the conversion matrix for the projected vector. The $\boldsymbol{v}^{cart}$ will be sent to dVRK at a constant frequency, resulting in a consistent movement. Additionally,

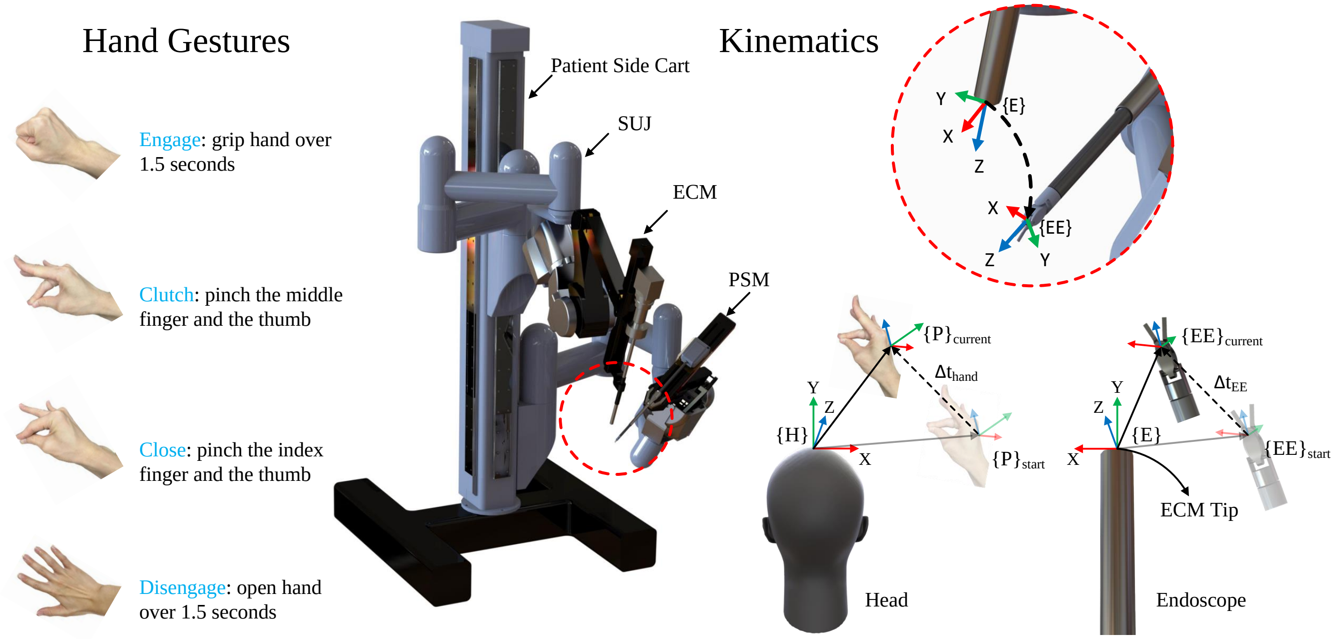

Hand Gestures
Engage: grip hand over 1.5 seconds
Clutch: pinch the middle finger and the thumb
Close: pinch the index finger and the thumb
Disengage: open hand over 1.5 seconds
Patient Side Cart
SUJ
ECM
PSM
Kinematics
Y
X
Z
{E}
X
{EE}
Z
Y
{P}current
Δthand
Y
Z
{H}
X
{P}start
Head
{EE}current
ΔtEE
Y
Z
{E}
X
{EE}start
ECM Tip
Endoscope

FIGURE 2 Instrument teleoperation scheme. Four hand gestures are utilized to engage, clutch, close the gripper, and disengage with the instrument mounted on the PSM. The hand's relative translation with respect to the head is scaled down to control the relative translation of the end effector with respect to the endoscope. The orientation of the hand with respect to the head is employed to govern the orientation of the end effector with respect to the endoscope. The palm of the hand is selected as the reference point, with an orientation offset to ensure that the tool's pose aligns more closely with the natural pose of the index finger and thumb.

the angular velocity of the endoscope, denoted as ω_c, is obtained by multiplying the scalar factor s_2 with the negative sign of the relative rotation angle of the head, represented by α_h.

The planar translation mode and the orientation mode provide the user with intuitive navigation capabilities, allowing them to effortlessly position the endoscope to the desired spot of interest. Additionally, the insertion and extraction modes facilitate zoom adjustments for the endoscope. By activating these modes, users can effectively focus on specific areas of interest and work with precise details during the surgical procedure.

3.2 | Instrument Teleoperation

In Chen et al.'s work[11], they implemented hand gestures to teleoperate the instrument mounted on the PSM and achieved comparable performance with the conventional method in the virtual ring-wire task. However, their implementation did not integrate the endoscope and only considered the motion of the instrument with respect to the world coordinate system, which was determined by the starting position of the HoloLens. Consequently, the coordination between the instrument and the endoscope was not taken into account. In this work, we build upon their methodology and modify the kinematic aspect of the system.

The engagement rule is depicted in Figure 2. To engage with the instrument, the user must grip his hand for over 1.5 seconds, and to disengage, he should open the hand for over 1.5 seconds. Clutching the instrument is achieved by pinching the middle finger and thumb while closing the gripper requires pinching the index finger and thumb.

As shown in the kinematics part of Figure 2, the reference frame of the instrument end effector is set to be the camera frame (ECM tip). We use the absolute orientation of the hand with respect to the head frame to control the orientation of the end effector with respect to the camera frame. The palm of the hand is selected as the reference point, with an orientation offset to ensure that the tool's pose aligns more closely with the natural pose of the index finger and thumb, improving the intuitiveness and ease of teleoperation. The position of the jaw is determined by adding the scaled relative translation of the hand to the previous position of the end effector. Likewise, we perform a conversion of the translation and orientation from the left-handed coordinate system to the right-handed coordinate system.

$$\boldsymbol{t}_{EE_{new}}^{camera} = \boldsymbol{t}_{EE_{start}}^{camera} + s_3 * C_2 * \Delta\boldsymbol{t}_{hand}^{head} \tag{3}$$

with

$$C_2 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The variables in equation 3 are described as follows: $\boldsymbol{t}_{EE_{start}}^{camera}$ and $\boldsymbol{t}_{EE_{new}}^{camera}$ represent the starting and new positions of the instrument end effector with respect to the camera frame, respectively, while s_3 denotes the scaling factor, C_2 denotes the handedness conversion matrix for hand translation, and $\Delta\boldsymbol{t}_{hand}^{head}$ represents the relative translation of the hand with respect to the head.

For simplicity, the conversion of rotations is presented in quaternion format as follows:

$$\boldsymbol{q}_{EE}^{camera} = q_\omega + \boldsymbol{i}q_x^{head} - \boldsymbol{j}q_y^{head} - \boldsymbol{k}q_z^{head} \tag{4}$$

with

$$\boldsymbol{q}_{hand}^{head} = \boldsymbol{q}_{palm}^{head} * \boldsymbol{q}_{offset} = q_\omega + \boldsymbol{i}q_x^{head} + \boldsymbol{j}q_y^{head} + \boldsymbol{k}q_z^{head} \tag{5}$$

In equation 5, unit quaternions $\boldsymbol{q}_{hand}^{head}$ and $\boldsymbol{q}_{palm}^{head}$ represent the orientation of the hand and palm with respect to the head, respectively, while $\boldsymbol{q}_{offset}$ denotes the orientation offset. The $\boldsymbol{q}_{hand}^{head}$ is composed of a scalar component q_ω and a vector component $\boldsymbol{i}q_x^{head} + \boldsymbol{j}q_y^{head} + \boldsymbol{k}q_z^{head}$ which represents the Euler axis described in head frame. In equation 4, unit quaternion $\boldsymbol{q}_{EE}^{camera}$ denotes the orientation of the instrument end effector with respect to the camera frame. This quaternion shares the same scalar component as $\boldsymbol{q}_{hand}^{head}$, whereas its Euler axis is mirrored across the origin and converted from the left-handed coordinate to the right-handed coordinate.

3.3 | Network and Endoscopic Video Display

The network architecture between the dVRK and HoloLens 2 is depicted in Figure 3. It comprises two threads: one for teleoperation (upper blue flowchart) and another for video streaming (lower green flowchart).

In the teleoperation thread, the kinematic states of the PSMs and ECM are transmitted from the dVRK to Unity[†], utilizing the sawSocketStreamer[‡] package, through User Datagram Protocol (UDP). Unity receives user inputs, such as head motion, hand gestures, and speech commands from HoloLens 2, and generates corresponding motion commands. These motion commands are then sent back to the sawSocketStreamer.

In the video streaming thread, the endoscopic video captured by the dVRK is published via Robot Operating System (ROS). The left and right images are horizontally concatenated into a single image, which is then transmitted from ROS to Unity using the ROS-TCP-Connector[§] via Transmission Control Protocol/Internet Protocol (TCP/IP). In Unity, a customized shader[¶] performs video rendering based on the eye index and displays the appropriate images on the left and right lenses of the HoloLens 2. To accommodate the computational demands of video rendering, the process is conducted within Unity, and the scene is subsequently streamed to the HoloLens 2 with a 30 Hz frame rate, using the Holographic Remoting[#] functionality.

4 | EXPERIMENTS AND RESULTS

4.1 | Experiment Design

To assess the effectiveness and intuitiveness of the proposed system, we conducted two tasks: the endoscope navigation task and the peg transfer task. These tasks were designed to evaluate the viability of the endoscope teleoperation method and the overall system independently.

[†] https://unity.com

[‡] https://github.com/jhu-saw/sawSocketStreamer

[§] https://github.com/Unity-Technologies/ROS-TCP-Connector

[¶] https://docs.unity3d.com/Manual/SinglePassInstancing.html

[#] https://learn.microsoft.com/en-us/windows/mixed-reality/develop/native/holographic-remoting-overview

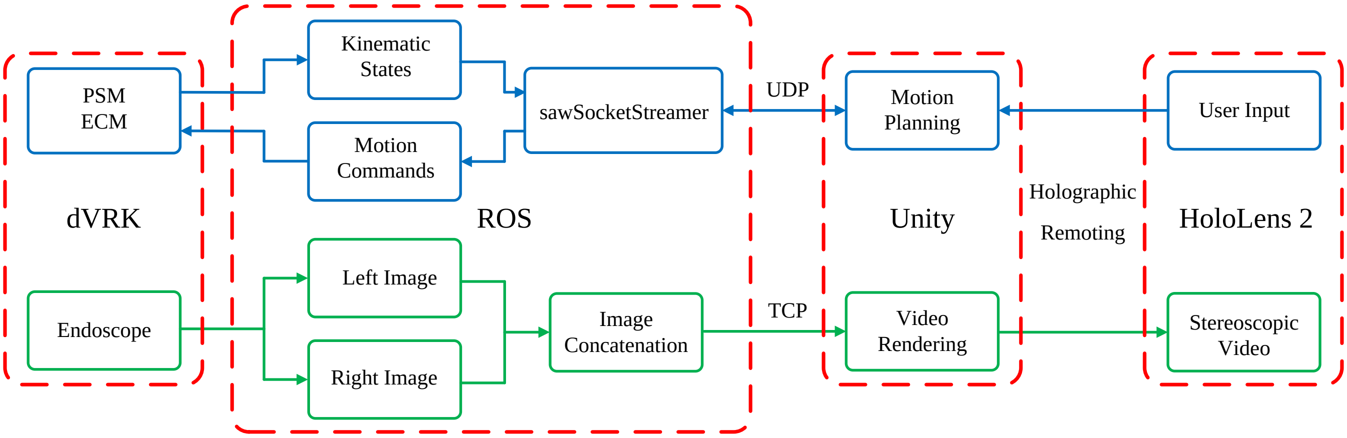


FIGURE 3 The network architecture between dVRK and HoloLens 2. The upper flowchart illustrates the teleoperation data transmission while the lower flowchart represents the video streaming. To accommodate the computational demands, Unity running on a PC processes the data from ROS and HoloLens 2 and streams the video to HoloLens 2 through Holographic remoting functionality.

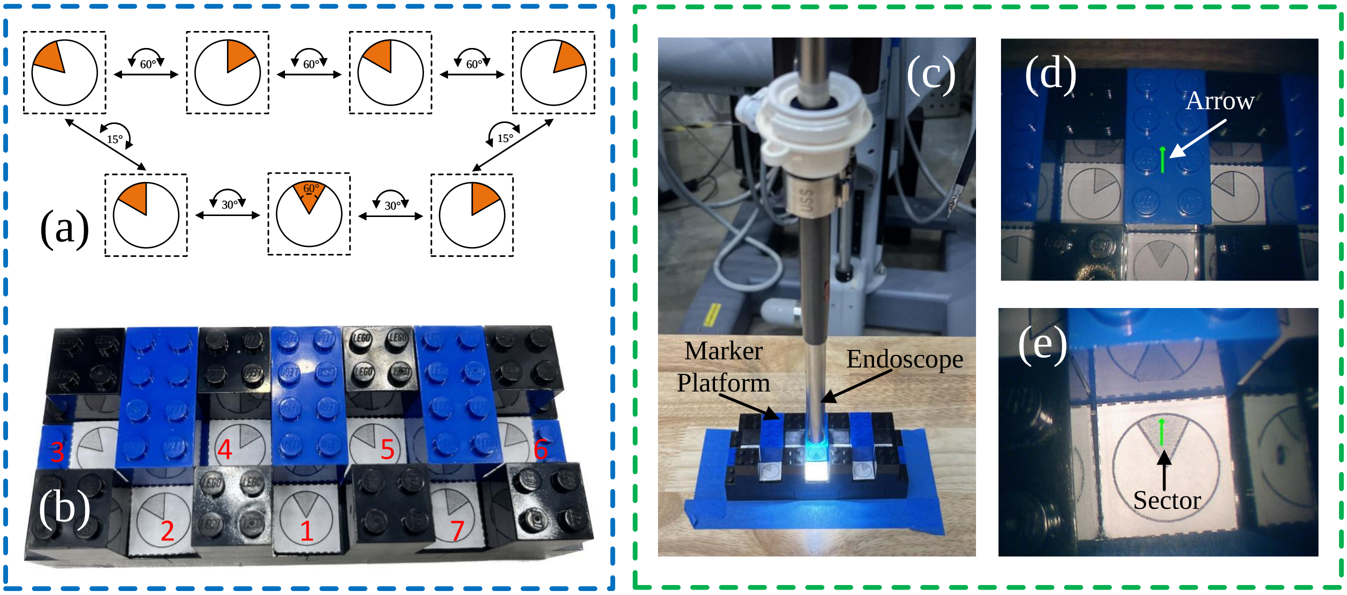


FIGURE 4 Endoscope navigation task: (a) Seven markers embedded with rotated sectors. (b) Marker platform housing markers. (c) The endoscope is positioned above the marker platform. (d) An arrow is overlayed on the endoscopy video. (e) The alignment of the arrow with one of the sectors.

We compared the proposed method with the conventional MTM teleoperation method based on completion time, usability, and workload. Usability was measured using the usability scale (SUS)[22], while workload was assessed using the NASA Task Load Index (TLX)[23] through questionnaires.

In the endoscope navigation task, we created a marker platform, as depicted in Figure 4 (a) and (b), with 7 circles labeled with rotated sectors. The protocol is as follows:

1) Participants should navigate the endoscope from the initial position (Figure 4 (d)) to align an arrow, displayed on the video, with the sectors.
2) The arrow has to be positioned within the sector while maintaining a similar orientation to the target sector (Figure 4 (e)).

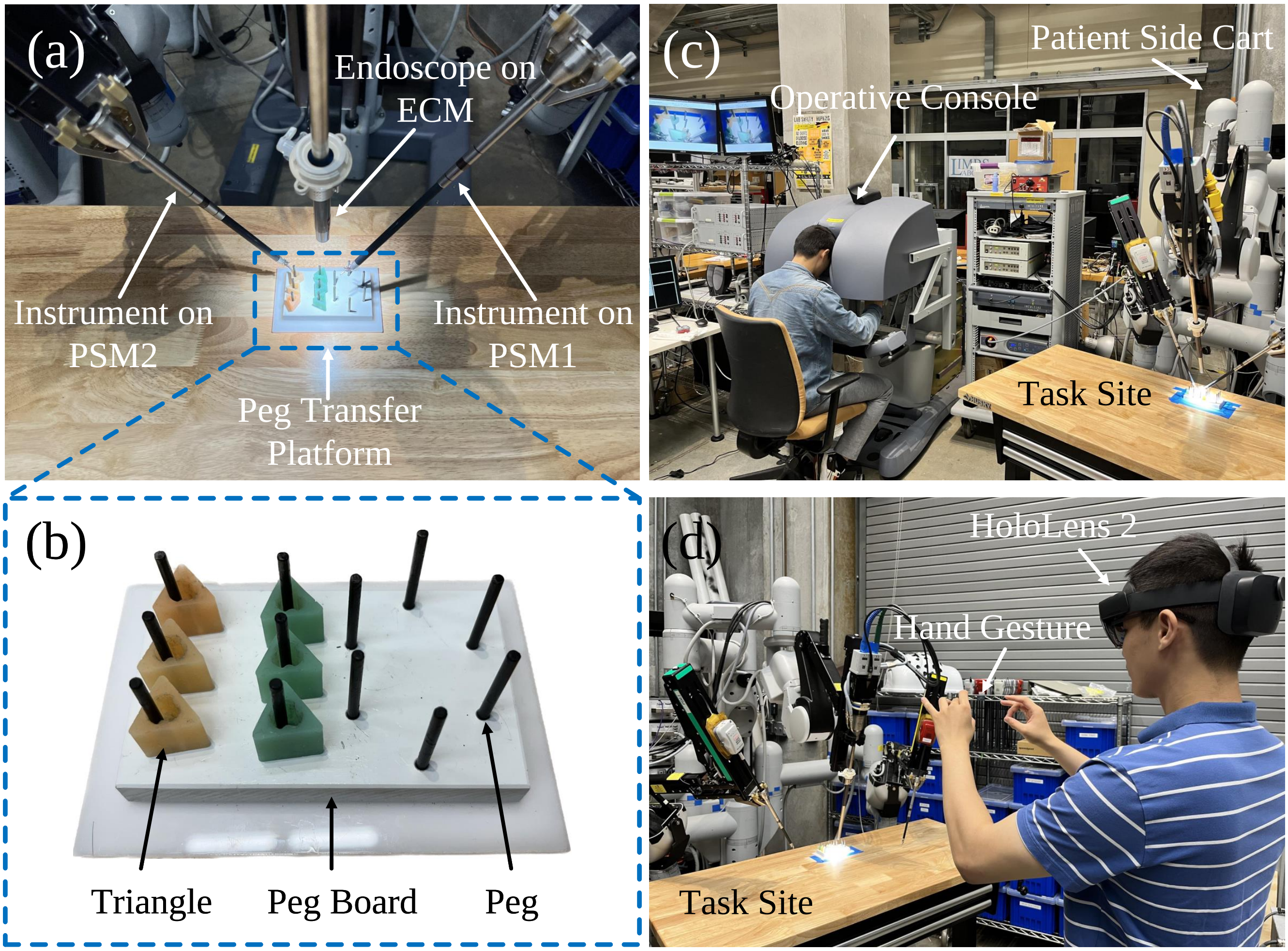


FIGURE 5 Peg transfer task: (a) Task site with peg-transfer platform, endoscope, and instruments. (b) Components of the peg transfer platform. (c) User performing tasks using the conventional console. (d) User performing the task using HoloLens 2.

3) The arrow is required to traverse all sectors and go back to the first vector in sequential order (clockwise or counter-clockwise).

The task required participants to effectively utilize the motion modes of the endoscope and complete the alignment as quickly as possible within a 6-minute time limit. Each task consisted of four trials, comprising two clockwise alignments and two counter-clockwise alignments.

We implemented a customized peg transfer task derived from the Fundamentals of Laparoscopic Surgery (FLS) exam[24] to evaluate the performance of the overall system while comparing it with the conventional MTM-based method. The customized peg transfer task has three steps which are:

1) Lift the six triangles using a gripper initially teleoperated by the non-dominant hand
2) Next, move each triangle to the gripper teleoperated by the dominant hand.
3) Position every triangle onto a peg located on the opposite side of the board.

The initial position of the endoscope was deliberately set to provide a limited view of the platform, compelling participants to fully engage in teleoperating both the endoscope and the instruments. Each trial in the task has a time limit of 6 minutes, and a total of 3 trials are performed for each task.

4.2 | Experiment Setup

A within-subjects user study was conducted, in which a group of 15 participants (13 males, 2 females; mean age: 23.27, standard deviation: 1.10) were recruited from the community. Before the experiment, participants underwent pre-experiment surveys that indicated their limited experience with both conventional and MR-based operational methods and confirmed the absence of any physical or mental impairments. Prior to the main study, a pilot study was carried out, which involved 3 endoscope navigation tasks and 3 peg transfer tasks. The pilot study allowed us to make necessary adjustments to parameters such as the endoscope moving speed, orientation offset, and video properties (contrast and window size). Including the pilot run, a total of 10 endoscope navigation tasks and 10 peg transfer tasks were conducted (5 participants took both tasks). The study was conducted with approval from our Institutional Review Board (IRB).

The proposed system was implemented on a host PC with the following specifications: Intel Core i7-12700H Processor, 16 GB DDR5 RAM, and an NVIDIA GeForce RTX 3060 Laptop GPU. The system also utilized a Microsoft HoloLens 2 device with the Holographic Remoting Player app 2.9.1 installed. The MR method we designed was developed using Unity 2021.3.8f1 in conjunction with the Mixed Reality Toolkit (MRTK) 2.8.2. The development environment also included Visual Studio 2019, and the application was operated on the Windows 10 operating system.

4.3 | Results

In addition to assessing workload and usability, we analyzed the average time taken to align a single marker in the endoscope navigation task and the average time taken to transfer a single triangle in the peg transfer task. The evaluation results for both tasks are presented in Table 1 and Figure 6.

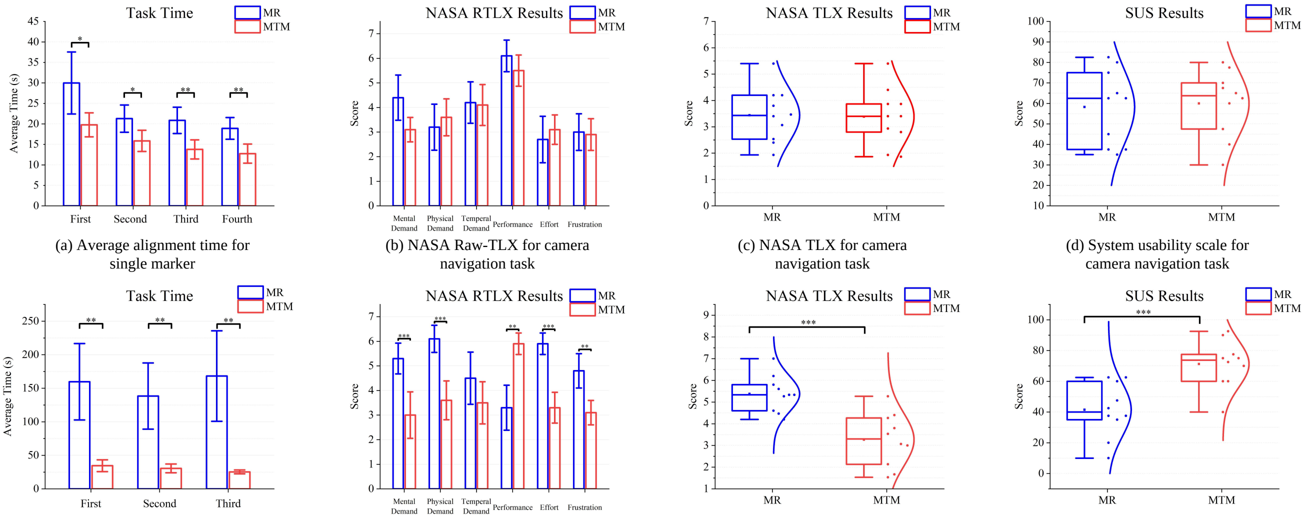
In the endoscope navigation task, despite the MR method having a higher average alignment time compared to the MTM method, the results shown in Figure 6 (a) indicate that participants rapidly improve their performance after the initial trial when using the MR method. It suggests that participants quickly grasp the skills required for the MR method once they become accustomed to it. Additionally, it is worth noting in 1 that the first three participants in the pilot study were assigned a slower endoscope-moving speed in comparison to the remaining seven participants. As a result, there is a noticeable decrease in average time from 29.49 sec to 19.88 sec.

Furthermore, Figures 6 (b), (c), and (d) demonstrate that there is no statistically significant difference ($p > 0.05$) in the evaluation of both methods by the participants. This implies that the MR-based method is generally accepted and has a comparable level of viability to the conventional method.

In the peg transfer task, the conventional MTM method has better performance than the MR method in average time, workload, and usability. Figure 6 (f) shows that the MR-based method resulted in significantly high mental demand, physical demand, and effort while leading to obvious frustration and poor performance. Because of the accumulative fatigue and frustration, in Figure 6 (e), the third trial shows an increased time compared to the second trial with the MR method. However, after revising the orientation offset of the hand and adjusting the property of the endoscopic video, the time for participants to complete the task significantly decreased from 280 sec to 102.07 sec on average.

TABLE 1 Evaluation results for the endoscope navigation task and peg transfer task. A total of 10 endoscope navigation tasks and 10 peg transfer tasks were conducted, including 3 pilot tasks for each task type. Lower scores on the NASA TLX indicate a lower workload, while higher scores on the SUS reflect better usability.

Tasks	Statistics	Average Time (s)				NASA TLX		SUS	
		MR			MTM	MR	MTM	MR	MTM
		Pilot	Revised	Overall					
Endoscope Navigation	Mean	29.49	19.88	22.76	15.54	3.44	3.39	58.25	60.00
	Std	11.46	7.67	9.88	5.62	1.02	1.08	18.26	16.16
Peg Transfer	Mean	280.00	102.07	155.45	30.11	5.38	3.27	41.5	71.25
	Std	99.50	68.61	113.38	13.19	0.84	1.23	17.61	15.29



(a) Average alignment time for single marker

(b) NASA Raw-TLX for camera navigation task

(c) NASA TLX for camera navigation task

(d) System usability scale for camera navigation task

(e) Average transferring time for single peg

(f) NASA Raw-TLX for peg transfer task

(g) NASA TLX for peg transfer task

(h) System usability scale for peg transfer task

FIGURE 6 Evaluation results for the endoscope navigation task and peg transfer task. Bar plots (a) and (e) illustrate the average time taken to align one marker and transfer a triangle in each trial, respectively. Bar plots (b) and (f) represent the NASA Raw-TLX results that eliminate the weighting process whereas box plots (c) and (g) are NASA TLX results derived from weighted rating[23]. Box plots (d) and (h) denote the usability scale in the two tasks, respectively. While the proposed MR teleoperation system falls short of the performance achieved by the conventional MTM-based method in the peg transfer task, it exhibits comparative NASA TLX and SUS scores in the endoscope navigation task. NOTE: The asterisks represent the significance levels of p-values, where "*" indicates $p \leq 0.05$, "**" indicates $p \leq 0.01$, and "***" indicates $p \leq 0.001$. The error bars in bar plots represent the standard deviation with a coefficient of variation of 0.5.

5 | DISCUSSION AND ANALYSIS

As revealed by the experiment results, our proposed camera teleoperation method has comparable functionality to the conventional method while the instrument teleoperation method still has great potential of being improved. There are several reasons that caused the MR-based method to exhibit inferior performance.

Unlike using a mechanical device controlled by dexterous hands, participants with the MR method have to take more effort to precisely perceive the relative pose of their head and navigate the endoscope by rate control. However, it was reported by participants that the endoscope teleoperation method is easy to learn and intuitive to implement. We can expect that with a more delicate parameter setting, including the speed and movement trigger threshold, the MR-based method can have a better performance.

Several factors affect the performance of MR-based instrument teleoperation. First, the huge discontinuity of teleoperation, caused by wrong hand gesture recognition and the limited recognition range of HoloLens 2, confused and frustrated the participants. Most cases happened when the index finger and middle finger are occluded, also resulting in a limited rotation range of the tool. While the user unintentionally moves their hands out of the line of sight, the sensor on HoloLens 2 would easily lose track of them and force the user to reposition their hands. Second, although the endoscopic video was stereoscopically displayed on HoloLens 2, the depth information is weaker than that of the stereo viewer of the dVRK console and led to uncertainty and hesitance when the user was transferring the triangle. This weakness was mainly caused by the low resolution and limited coloration of the HoloLens 2. Third, the latency of the system, introduced by the complexity of the network between the dVRK and HoloLens 2, is about 260 ms to 300 ms, resulting in a perceptible lag as well as higher mental demand and physical demand. However, deploying the application on HoloLens would make it worse due to its limited computational power. Lastly, participants were instructed to keep their hands raised in the air while manipulating the instruments through the MR interface, as opposed to resting their elbows on the dVRK console's support. The participants reported perceptible fatigue after 6 minutes of engagement, and this fatigue appeared to accumulate even with breaks lasting 5 to 10 minutes between each trial. It indicates that the current MR teleoperation method may not be suitable for extended surgical procedures without physical support.

Although there are many issues with the MR-based teleoperation method when performing conventional tasks, it still has advantages including portability, increased mobility, and improved situational awareness. It can potentially remove the need for assistance to work with the robot, as the surgeon can be at the patient's side. To enhance the performance of the system, we expect a more powerful MR headset that enables direct video rendering and supports higher resolution as well as better coloration. To solve the discontinuity caused by finger occlusion and the loss of hand tracking, we propose to utilize hand-attached sensors, such as IMUs and encoders, to eliminate the hand gesture misrecognition and enlarge the limited recognition range.

6 | CONCLUSION

We proposed a novel teleoperation and visualization method based on MR and implemented it on the dVRK with the use of HoloLens 2. The system leverages head tracking, hand gesture recognition, and speech commands to facilitate the teleoperation of both the endoscope and instruments of the robot while providing the stereoscopic display of the endoscopic video.

To evaluate the viability of the system, we conducted a camera navigation task and peg transfer task. The results demonstrated that the teleoperation scheme for the endoscope was comparable to the conventional method, indicating its potential for effective use. However, the MR system showed limitations in the peg transfer task, primarily due to challenges with hand gesture recognition.

The findings of the experiments highlight areas that require improvement, particularly in hand gesture recognition and video display quality, to further enhance the system's performance. Addressing these issues can pave the way for more efficient and versatile surgical procedures.

It is worth noting that the proposed system holds potential beyond surgical applications and could be implemented in other teleoperated robots such as exploration robots and rescue robots, expanding its usability across various domains.

REFERENCES

1. Peters BS, Armijo PR, Krause C, Choudhury SA, Oleynikov D. Review of emerging surgical robotic technology. *Surgical Endoscopy.* 2018;32:1636–1655.
2. Kumar A, Yadav N, Singh S, Chauhan N. Minimally invasive (endoscopic-computer assisted) surgery: Technique and review. *Annals of Maxillofacial Surgery.* 2016;6(2):159.
3. Zidane IF, Khattab Y, Rezeka S, El-Habrouk M. Robotics in laparoscopic surgery-A review. *Robotica.* 2022:1–48.
4. Guthart GS, Salisbury JK. The Intuitive^{TM} telesurgery system: overview and application. In: *IEEE International Conference on Robotics and Automation*, . 1. IEEE. 2000:618–621.
5. Adams dT, Eubanks WS, Fuente d. lSG. Early experience with the Senhance®-laparoscopic/robotic platform in the US. *Journal of Robotic Surgery.* 2019;13(2):357–359.
6. Alkatout I, Salehiniya H, Allahqoli L. Assessment of the Versius robotic surgical system in minimal access surgery: a systematic review. *Journal of Clinical Medicine.* 2022;11(13):3754.
7. Wang Y, Li Z, Yi B, Zhu S. Initial experience of Chinese surgical robot "Micro Hand S" -assisted versus open and laparoscopic total mesorectal excision for rectal cancer: Short-term outcomes in a single center. *Asian Journal of Surgery.* 2022;45(1):299–306.
8. Lee HK, Lee KE, Ku J, Lee KH. Revo-i: The competitive Korean surgical robot. *Gynecologic Robotic Surgery.* 2021;2(2):45–52.
9. Dardona T, Eslamian S, Reisner LA, Pandya A. Remote presence: Development and usability evaluation of a head-mounted display for camera control on the da vinci surgical system. *Robotics.* 2019;8(2):31.
10. Fu G, Azimi E, Kazanzides P. Mobile Teleoperation: Feasibility of Wireless Wearable Sensing of the Operator's Arm Motion. In: *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE. 2021:4238–4243.
11. Chen AC, Hadi M, Kazanzides P, Azimi E. Mixed Reality Based Teleoperation of Surgical Robotics. In: *International Symposium on Medical Robotics (ISMR)*, IEEE. 2023:1–7.
12. Moniruzzaman M, Rassau A, Chai D, Islam SMS. Teleoperation methods and enhancement techniques for mobile robots: A comprehensive survey. *Robotics and Autonomous Systems.* 2022;150:103973.
13. Suzuki R, Karim A, Xia T, Hedayati H, Marquardt N. Augmented Reality and Robotics: A Survey and Taxonomy for AR-Enhanced Human-Robot Interaction and Robotic Interfaces. 2022. doi: 10.1145/3491102.3517719
14. Yeung AWK, Tosevska A, Klager E, et al. Virtual and augmented reality applications in medicine: analysis of the scientific literature. *Journal of Medical Internet Research.* 2021;23(2):e25499.
15. Kazanzides P, Chen Z, Deguet A, Fischer GS, Taylor RH, DiMaio SP. An open-source research kit for the da Vinci® Surgical System. In: *IEEE International Conference on Robotics and Automation (ICRA)*, IEEE. 2014:6434–6439.
16. Hong N, Kim M, Lee C, Kim S. Head-mounted interface for intuitive vision control and continuous surgical operation in a surgical robot system. *Medical & Biological Engineering & Computing.* 2019;57:601–614.
17. Qian L, Song C, Jiang Y, et al. FlexiVision: Teleporting the surgeon's eyes via robotic flexible endoscope and head-mounted display. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE. 2020:3281–3287.
18. Ma X, Song C, Qian L, Liu W, Chiu PW, Li Z. Augmented reality-assisted autonomous view adjustment of a 6-DOF robotic stereo flexible endoscope. *IEEE Transactions on Medical Robotics and Bionics.* 2022;4(2):356–367.
19. Abdurahiman N, Khorasani M, Padhan J, et al. Scope actuation system for articulated laparoscopes. *Surgical Endoscopy.* 2023;37(3):2404–2413.

20. Wen R, Tay WL, Nguyen BP, Chng CB, Chui CK. Hand gesture guided robot-assisted surgery based on a direct augmented reality interface. *Computer Methods and Programs in Biomedicine.* 2014;116(2):68–80.
21. Gondokaryono RA, Agrawal A. An approach to modeling closed-loop kinematic chain mechanisms, applied to simulations of the da vinci surgical system. *Acta Polytechnica Hungarica.* 2019;16(8).
22. Brooke J, others . SUS-A quick and dirty usability scale. *Usability evaluation in industry.* 1996;189(194):4–7.
23. Hart SG. NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.* 2006;50(9):904–908.
24. Vassiliou MC, Dunkin BJ, Marks JM, Fried GM. FLS and FES: comprehensive models of training and assessment. *Surgical Clinics.* 2010;90(3):535–558.