2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

A Digital Twin-Driven Immersive Teleoperation Framework for Robot-Assisted Microsurgery

Peiyang Jiang, Dandan Zhang

Abstract— **This paper presents a novel digital twin (DT)-driven framework for immersive teleoperation in the domain of robot-assisted microsurgery (RAMS). The proposed method leverages the power of DT with mixed reality (MR) technology to create an interactive, immersive teleoperation environment for surgeons to conduct RAMS with higher precision, improved safety, and higher efficiency. More specifically, the MR device can provide operators with the 3D visualization of a digital microsurgical robot mimicking the motions of the physical one as well as the 2D real-time microscopic images during microsurgical operation. We evaluated the proposed framework through user studies based on a Trajectory Following task and conducted comparisons between scenarios with and without using the proposed framework for RAMS. The NASA-TLX questionnaire, along with additional evaluation metrics such as total trajectory, time cost, mean velocity, and a predefined collision metric, were used to analyze the user studies. Results indicated that the proposed DT-driven immersive teleoperation framework could enhance the precision, safety, and efficiency of teleoperation, and provide a satisfactory user experience to operators during microsurgical operation.**

I. INTRODUCTION

Robot-Assisted Microsurgery (RAMS) represents an evolution that enables unparalleled precision for microsurgery and reduces invasiveness for patients in microsurgical procedures [1]. Despite the significant advancements of RAMS, several challenges remain, which demand innovative solutions to further optimize its implementation in microsurgery [2]. Currently, the majority of robotic platforms designed for microsurgery are based on teleoperation [3], which enables patient care in remote or inaccessible locations [4], [5]. However, teleoperation often faces challenges like data transmission issues that lead to potential errors and unintuitiveness of operation. A promising solution is the use of a digital twin (DT), a real-time virtual replica of the physical system [6]. DT has shown significant potential for enhancing surgical teleoperation [7]. For example, it has been used to enable remote users to interact with a simulated Robotic Surgical Assistant before actual operations [8]. However, few studies have explored the integration of DT technology into teleoperated microsurgical robotic systems to ensure safety and precision during microsurgical operations.

In addition, further enhancements are needed to overcome certain limitations, such as the lack of ergonomic human-robot interaction interface and intuitive visual feedback during teleoperation [9]. The use of two-dimensional (2D) screen-based visual feedback has a high potential to cause fatigue in operations, which may lead to surgical errors. To mitigate this issue, Augmented Reality (AR) technology has been introduced in RAMS [10]. By overlaying computer-generated virtual objects onto real-world scenes, AR enhances the user's sensory perception. This technology not only augments the visualization of lesions but also aids in instrument guidance [11], significantly increasing surgical precision and safety. More recently, Mixed Reality (MR), with its ability to facilitate real-time interactions between physical and digital objects, has been used to enrich surgeons with enhanced three-dimensional (3D) information [12]. This technology can create an immersive experience for surgeons and improve the efficiency of teleoperation in RAMS. Moreover, a significant proportion of telesurgery relies on motion-tracking devices that use mechanical linkages to capture surgeons' hand movements and generate control commands. [13]–[15]. Such a method of control can often be unintuitive, potentially increasing surgeons' workload. Therefore, we will integrate an ergonomic human-robot interaction interface for microsurgical robot remote control based on hand gestures and speech commands, aiming to foster more intuitive teleoperation.

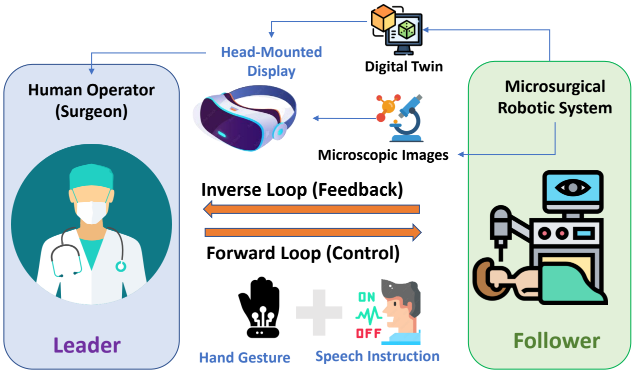

Fig. 1. Concept Overview of the Digital Twin-Driven Immersive Teleoperation Framework for RAMS.

In summary, this paper introduces the development of a DT-driven immersive teleoperation Framework to improve the teleoperation experience of RAMS. The concept of which is illustrated in Fig. 1. The main contributions are listed as follows:

- We develop a DT of the microsurgical robot and the tissue intended for manipulation and leverage MR technology to enhance visual feedback through the display of a 3D virtual robot overlaid alongside a 2D real-time microscopic view. This aims to provide human operators

---

Peiyang Jiang is with the Department of Engineering Mathematics, University of Bristol, Bristol, United Kingdom. Dandan Zhang is with the Department of Bioengineering, Imperial-X, Imperial College London, London, United Kingdom. Corresponding: Dandan Zhang, d.zhang17@imperial.ac.uk

with an immersive teleoperation experience.

- We construct a microsurgical robotic system that enables teleoperation through hand gestures and speech instructions. An experimental evaluation is conducted using a trajectory-following task to assess the effectiveness of the proposed method through comparative studies.

II. RELATED WORK

A. *Digital Twin for Robotic Surgery*

DT has been investigated for surgical training and practice [16]. Customizable to specific applications, DT can provide an intuitive, low-latency interface for teleoperation in RAMS [7]. For example, Laaki et al. developed a DT system to control a UR3 arm in a surgical environment by transmitting the robot's end-effector pose via user datagram protocol (UDP) [17]. Their experiments showed that the real arm could effectively mirror the virtual arm's end-effector pose with minimal delay. In addition, DT has demonstrated a significant potential for enhancing remote surgeries, providing precise and real-time guidance even when the surgeon and patient are not co-located. Hagmann et al. [18] explored a Shared Control Parametrization Engine that retrieves procedural context information from a DT to help novice surgeons proficiently use a robotic system. A pilot study was conducted using basic surgical training tasks on the DLR MiroSurge system. Shu et al. [19] had developed Twin-S, a DT framework designed for skull base surgeries, which combines high-precision optical tracking and real-time simulation to enhance situational awareness and improve surgical operation accuracy. This work demonstrated the promising potential for applying DT in image-guided surgical interventions.

However, the aforementioned research didn't implement immersive manipulation with the support of MR technologies, while ergonomic human-robot interaction interfaces for teleoperation have not been investigated.

B. *Mixed Reality for Robotic Surgery*

Previous research has indicated that MR can assist surgeons in gaining a better understanding of the spatial relationship between the surgical robot and the lesion during robotic surgery procedures. For example, MR has been proven to help improve surgical outcomes by enhancing spatial awareness during surgery by revealing occluded and obscured organ boundaries and other key anatomical structures [20]. Other studies have demonstrated that MR can enable immersive teleoperation [21], thus enhancing the efficiency of surgical training or robotic surgery. A simulation-based feasibility study was conducted using an MR headset for endoscopic surgery [22]. This research showed that the device contributed to enhanced performance among novices. In [23], MR was used to enhance orthopedic training and surgical procedures, which led to improved surgical precision, reduced complications, and advanced training simulations. Similarly, Lu et al. [24] utilized a visualization system based on MR technology in orthopedic surgery, which demonstrated that MR notably reduced users' workload compared to conventional 2D methods. Reis et al. [25] introduced the synergy of MR with the da Vinci Surgical System to enhance precision in procedures like radical prostatectomy, which improves intraoperative navigation.

However, the aforementioned studies primarily focus on the application of MR in the context of traditional surgeries, which include pre-operative planning and intra-operative navigational imagery presentation. The previous research did not explore DT-driven immersive manipulation methods for potential advancements and applications in RAMS.

III. METHODOLOGY

A. *Architecture*

In our research, we design a real-time DT that accurately reflects the movements of a microsurgical robot within a virtual environment. The DT of the microsurgical robot acts as a bridge between the operator and the physically controlled robot. Taking robot-assisted eye surgery as an example, human operators can utilize a simulated eyeball in the DT as a point of reference. In the virtual environment, the positions of the robot and the virtual eyeball can be collectively manipulated by the operator according to individual preferences within the virtual space. During this operation, the relative relationship between the virtual eyeball and the robot remains unchanged. This allows operators to comprehend the relative position between the microsurgical tooltip and the eyeball, which can enhance safety and precision during microsurgical operations. This method also helps to prevent accidental contact of the microsurgical tool with the delicate tissues of the actual eyeball. Additionally, DT not only provides operators with visually magnified perception but also employs a pre-trained You Only Look Once (YOLOv5) model to recognize target objects and needle tips [26]. During the actual motion process, the distance between the needle tip and the target object can be displayed on the screen for the operator. We further develop a DT-driven MR system with the aim of enriching the immersive manipulation experience for users and providing comprehensive support for operators when conducting complex microsurgical procedures.

The hardware system and workflow of the DT-driven MR-based immersive teleoperation framework are illustrated in Fig. 2. The figure also demonstrates the real-time capturing of hand gestures and processing of speech information, and illustrates how these data are utilized to control the microsurgical robot. The diagram also presents the real-time rendering of the DT within the virtual environment. The DT and the real-time microscopic view are displayed on the head-mounted headset as visual feedback to the human operator.

B. *Hardware System*

A microrobot (Sensapex-UMP) equipped with four linear high-precision piezo actuators is used as the micromanipulator, which can mount different microsurgical tools for different microsurgical sub-tasks. The microsurgical robot is assembled on an anti-vibration optical table to enhance

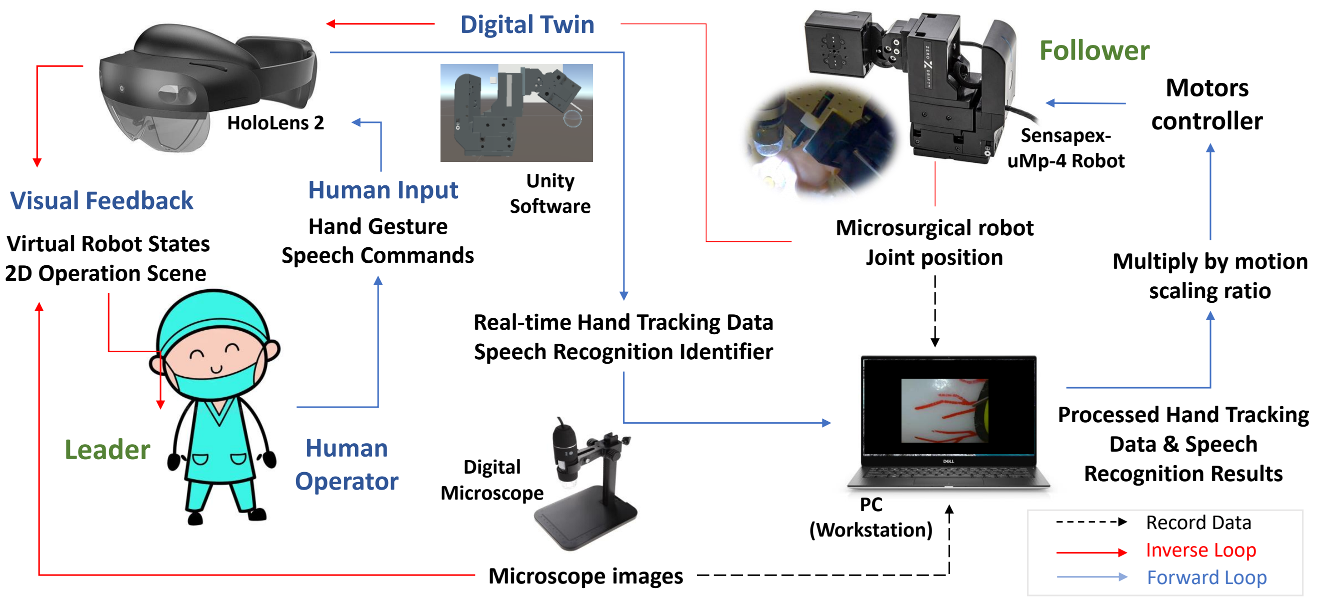

Fig. 2. System Overview. This figure illustrates the hardware, the forward/inverse loop for teleoperation, and the data transmission workflow of the proposed system. The key components of the framework include i) the microsurgical robotic platform with mounted microsurgical tools, ii) digital microscope for capturing 2D operation scene, iii) MR device (HoloLens 2) for immersive visualization of the 3D view of the DT and 2D microscopic images, iv) Unity software utilized for DT construction and updating the virtual robot states for real-time display to the human operator, v) the Python interface on PC used for processing hand-tracking data, vi) speech recognition and hand-tracking data used for generating robot control commands.

stability during micromanipulation. Two microscopes (Opti-Tekscope Digital USB Microscope) are used to capture the real-time operation image and provide visual feedback to human operators. Although the microscope may limit depth perception, our MR application aims to supplement this by providing additional spatial cues and enhancing 3D visualization.

Microsoft's HoloLens 2 is used as the MR device in this framework, which can provide operators with immersive experiences augmented by data generated from DT. Compared to using a VR headset, an MR headset is more convenient for surgeons to observe the surrounding conditions during microsurgery. In MR, the virtual and real content coexist and interact in real-time. With our system, the operators are not merely navigating a purely virtual space; they're navigating a virtual space that is anchored and relevant to their real-world position and perspective.

The MR device (HoloLens 2, Microsoft) features an Optical See-Through Head-Mounted Display (OST-HMD) and a suite of sensors, including cameras and depth sensors. Advanced hand-tracking, eye-tracking, and speech recognition technologies are incorporated to enhance user interactions with virtual content visualized in OST-HMD. The sensors can interpret operator gestures, offering more intuitive control of the robotic system. We therefore utilize hand-tracking and speech recognition to allow human operators to remotely manipulate the microsurgical robot. In this system, speech recognition serves as a contactless instruction for the surgeon to perform repositioning operations with ease, while the hand-tracking method is used for fine motion control of the robot.

C. *DT-Driven Immersive Teleoperation Framework*

1) Overview: A leader-follower control paradigm is used to control the microsurgical robot, where surgeons act as the leader and the microsurgical robot acts as the follower. The teleoperation system consists of two parts: the forward loop for robot control and the reverse loop for operator perception based on visual feedback. The forward loop for teleoperation is fundamental, where the operators can control the microsurgical robots remotely with the assistance of ergonomic human-robot interaction interfaces. To understand the current state of the robot and prepare for the next step of robotic control, users need to access real-time sensory feedback. To fulfill this requirement, the development of the reverse loop helps to reconstruct users' perceptions.

In this paper, a DT-driven MR framework is constructed for leader-follower control mode in RAMS. The DT serves as a dynamic representation of the physical robot within a virtual environment, which offers multiple benefits:

- Enhanced Spatial Awareness: The DT provides a 'third person' perspective, allowing the operator to view both the robot and its surroundings and gain an enhanced understanding of the spatial configuration of the operating scene.
- Better Predictive Analysis: The DT facilitates real-time simulation based on the robot's current movements. This predictive capability can alert the operator to potential collisions or misalignments before they occur in a real-world scenario, which enhances safety.
- Improved Feedback Loop: The DT functions as a continuous feedback loop, ensuring that adjustments made in the virtual environment are mirrored in real-

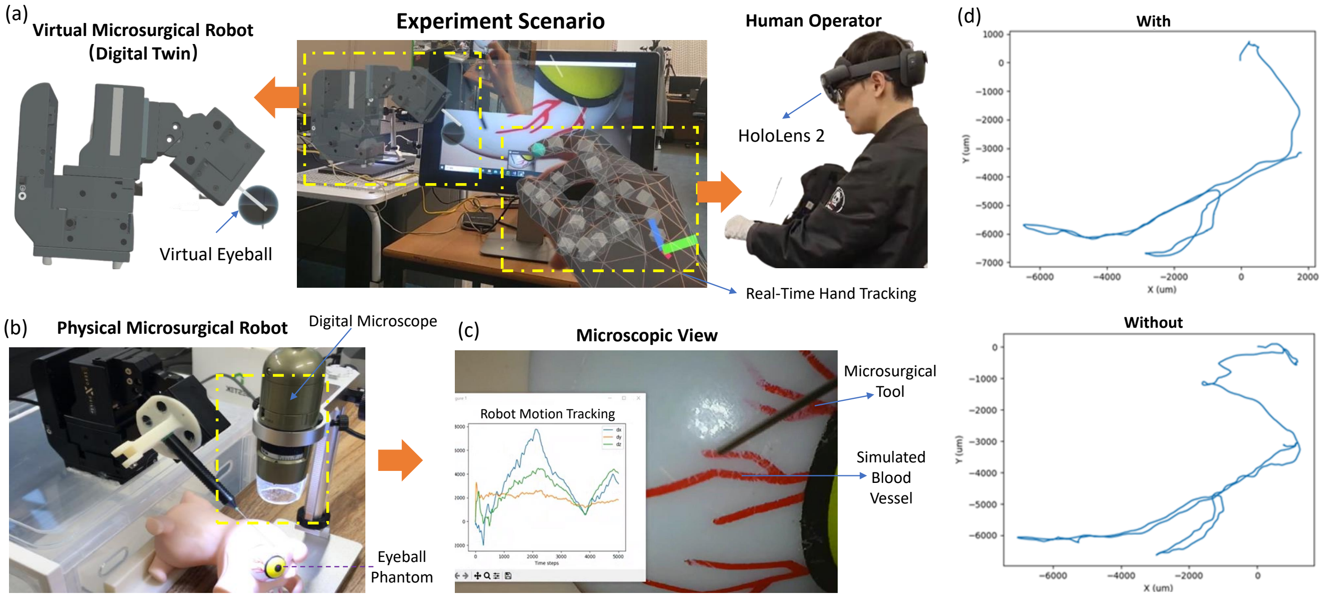

Fig. 3. Experiment scenario. (a) illustrates the experimental setup for user studies, the DT visualized during teleoperation, and the interaction between the human operator and the robot via hand tracking. (b) demonstrates the physical microsurgical robot, digital microscope, and eyeball phantom used in the experiments. (c) shows the microscopic view of the microsurgical tool on the eyeball, and the visualization of robot motion-tracking sequential data. (d) Trajectories of the end-effectors of the microsurgical robot with and without using the proposed DT-driven immersive teleoperation framework.

time robot control, and vice versa. This bi-directional interaction guarantees more precise control.

2) Forward Loop: During teleoperation, the human control input can be acquired by the MR device via hand-tracking and speech recognition techniques. The mathematical formula for mapping the leader robot's motions to the follower robot's end-effector in the forward loop of the teleoperation framework is as follows:

$$P_{\mathrm{f}}^{i} = \alpha \left(P_{\mathrm{l}}^{i} - \ P_{\mathrm{l}}^{i-1}\right) + P_{\mathrm{f}}^{i-1} \tag{1}$$

where i indicates the timestep, P_{f}^{i} is the new position of the end-effector of the microsurgical robot (the follower), P_{f}^{i-1} is the previous position of the microsurgical robot. P_{l}^{i} and P_{l}^{i-1} are the new and previous positions of the human's hand coordinate captured by the hand-tracking technique, respectively. The α is a motion scaling factor that adjusts the ratio of the end-effector position mapping between the leader and follower [27]. Through this motion mapping method, the movement trend of the robot's end effector can be synchronized with the movement trend of the hand.

The speech recognition implemented through the MR device can provide additional instructions such as 'start' to engage the robot for operation or 'clutch' to disconnect the robot from hand gesture control [27]. Hand gesture data and speech instruction data are collected using Unity software and sent to a computer via a User Datagram Protocol (UDP) socket connection. The 3D hand movements in the world coordinate are processed by the computer. To eliminate noise, the Kalman filter is used for post-processing while the filtered data is subsequently scaled to enable precise control of the microsurgical robot based on the adjustable scaling factor α.

3) Reverse Loop: The physical microsurgical robot's real-time end-effector positions are relayed back to the MR device, facilitating enhanced 3D visualization. This bidirectional data transfer empowers the Unity platform to construct an accurate virtual replica of the robot, reflecting the movements of the physical one in an authentic manner, thus achieving a reliable DT. The DT-driven virtual elements are superimposed onto the surgical scene [28]. We reconstruct the virtual eyeball based on the dimension of the physical eyeball, while more complex anatomies can be reconstructed using a 3D scanner. We ensure that both the virtual DT scene and the 2D view share the same coordinate frame after spatial registration [28], alleviating the need for the operators to manually align the virtual robot model to the eyeball.

In this context, the relative position between the microsurgical tooltip and the eyeball can be monitored in real-time, providing essential situational awareness to the operator. Leveraging this information, the operator can make informed adjustments to subsequent control commands, ensuring that RAMS is conducted with reliable safety and precision. Simultaneously, the head-mounted display integrates real-time operational views obtained from digital microscopes. This integration offers a comprehensive and immersive visualization of the microsurgical process. The operator observes the surgical field through a digital microscope, with the added benefit of DT-driven virtual elements overlaying the view. This enhancement improves their perception and control during microsurgical procedures. Additionally, it enables the operator to quickly locate critical operational areas during the procedure.

IV. EXPERIMENTAL VALIDATION AND RESULTS ANALYSIS

A. Experiment Design

We conducted a user study to evaluate the advantages of our proposed teleoperation framework over the traditional control method, which relies simply on the 2D microscopic images as visual perception. The experimental setup is shown in Fig. 3, where a 'Trajectory Following' task was used to simulate a microsurgical procedure [29]. The participants were required to follow a predefined trajectory that simulated a blood vessel in an eyeball model. Operators were instructed to navigate with high accuracy while applying minimal force to the eyeball. The microsurgical tool was moved back to the original position after the participants finished each trial.

The study involved seven participants (5 males and 2 females). Among these participants, two had previous experience with video games, and two were familiar with the microsurgical robotic platform. To ensure a fair comparison between different system setups, participants were instructed to alternately conduct experiments with and without using DT-driven immersive manipulation technology. Prior to the formal experiment, each participant practiced the procedures to get familiar with the proposed immersive teleoperation system. To ensure consistency in the experiment, participants were randomly assigned to begin with different types of experiments. Each participant completed six trials for statistical analyses: three using DT and three without using the DT. In the 'without DT' case, users were shown only the 2D images in the reverse loop. These images were displayed on the head-mounted display of the MR device. After receiving instructions on the experimental procedures, the participants were asked to provide their insights through questionnaires. A total of 42 trials were conducted while the experimental data was collected.

After finishing all the experiments, participants were asked to complete the National Aeronautical Space Agency-Task Load Index (NASA-TLX) [30]. NASA-TLX provides an overall workload score based on six weighted subclasses. Therefore, it was employed to evaluate the mental (E1), physical (E2), and temporal (E3) demands, as well as the frustration (E4), effort (E5), and performance (E6) of the participants using the proposed teleoperation framework. In addition, we utilized four evaluation metrics to assess the participant's performance, including total trajectory, time cost, mean velocity, and collision metric, which are detailed as follows.

Total Trajectory (TT): This metric measures the cumulative distance covered by the end-effectors of the microsurgical robot during a particular trial. It provides a summation of the path length that the robot's end-effectors take from the start of the operation until its completion. A shorter path often indicates a more direct and efficient movement of the surgical instrument. (unit: μm)

Time Cost (TC): This metric indicates the time taken for a participant to complete a single trial of the trajectory following task. A smaller time cost represents the reduction of time for microsurgery, which indicates better performance. (unit: s)

Mean Velocity (MV): The mean velocity represents the time-normalized path length of the microsurgical tool during teleoperation. A higher mean velocity indicates improved teleoperation efficiency. (unit: $\mu m/s$)

Collision Metric (CM): This metric is used to evaluate the performance of microsurgical operation based on six different needle-eyeball interaction scenarios during user studies. Each scenario corresponds to a specific score (ranging from 0 to 5), which reflects varying levels of safety and precision observed during the trajectory following task. The transformation between the tooltip and the eyeball frame was known and was used to calibrate the system before experiments. A higher score indicates better performance. More details about the definitions can be found in Table I, which quantifies the safety and accuracy of the operation.

B. Results

The results of the NASA-TLX questionnaire are summarized in Table II, with scores ranging from 0 to 10. According to the results, participants found that with our proposed approach, the tasks were less mentally demanding, involved less time pressure, required less effort, and induced less frustration. However, an exception was noted in the category of physical demand, where a slightly higher score indicated an increased physical load. A potential explanation for this outcome is that participants had to alternate their focus between the DT and the real-time 2D microscopic images, which required additional physical effort. However, we believe that the integration of the DT with 2D microscopic images remains valuable, as it leads to higher accuracy and safer operations due to a more comprehensive representation of the operating scene.

As shown in Fig. 4 (c), the reduced variability in the 'with DT' case is attributed to the enhanced depth perception provided by the DT. In the 'without DT' scenario, subjects lack depth information between the eyeball and the microsurgical tooltip. This absence of crucial depth cues can result in greater variability as operators may need to rely more on past experiences of depth estimation rather than real-time visual feedback.

The average 'total trajectory' of the microsurgical robot's end-effector across all trials was 44.1mm and 49.7mm when operated with and without our proposed framework, respectively. Our proposed framework demonstrated improved performance: the average 'time cost' was reduced (112.1s compared to 132.9s) and the 'mean velocity' increased (425.1$\mu m/s$ vs. 409.9$\mu m/s$). As per the 'collision metric', the user's performance significantly improved, evidenced by a score of 3.8 when using our immersive teleoperation framework compared to 2.9 without using it.

Before subsequent statistical analysis, normality tests (Shapiro-Wilk test) were conducted at a significance level of 0.05. The experimental data derived from the mean velocity metric revealed a non-parametric nature, while data from other metrics satisfied the assumption of normal distribution.

TABLE I. Definitions of the Collision Metric. Evaluation of the interaction between a microsurgical tooltip and an eyeball in microsurgical operations. The evaluation is achieved through a range of scenarios that measure the frequency of contact and the distance maintained between the tooltip and the eyeball. Suppose that n is the number of interactions between the tooltip and the eyeball, k denotes the distance between the tooltip and the predefined trajectory, and t represents the operation time. $k > 2mm$ represents the tooltip is far from the predefined trajectory, $k \leqslant 0.5mm$ represents the tooltip is close to the predefined trajectory, $0.5mm < k \leqslant 2mm$ indicates the tooltip slightly deviates from the predefined trajectory.

Score	Description	Safety	Precision
0	Tooltip contacts the eyeball frequently ($n > 3$) and is far from the trajectory throughout operation	Low	Low
1	Tooltip contacts the eyeball occasionally ($0 < n \leqslant 3$) and is far from the trajectory throughout operation	Medium	Low
2	Tooltip avoids contact with the eyeball ($0 = n$) and is far from the trajectory throughout operation	High	Low
3	Tooltip contacts the eyeball occasionally ($0 < n \leqslant 3$) and slightly deviates from the trajectory for a short period ($t < 10s$)	Medium	Medium
4	Tooltip avoids contact with the eyeball ($0 = n$) and slightly deviates from the trajectory for a short period ($t < 10s$)	High	Medium
5	Tooltip avoids contact with the eyeball ($0 = n$) and is close to the trajectory throughout operation	High	High

TABLE II. Questionnaire result (Average Score)

	E1	E2	E3	E4	E5	E6
With	4.14	5.71	3.71	3.14	5.43	4.00
Without	5.00	5.29	6.43	5.57	7.57	5.00

TABLE III. Results based on quantitative evaluation (Average across trials)

	TT(um)	TC(s)	MV(um/s)	CM
with	44093.0	112.1	425.1	3.8
Without	49674.9	132.9	409.9	2.9

Our user study employed a within-subject design, where all participants performed two types of experiments repeatedly. For a non-parametric statistical comparison of the 'mean velocity' metric, we used the Wilcoxon signed-rank tests. T-tests were conducted for other metrics to validate the statistical differences between the two types of experiments. A p-value of less than 0.05 was considered statistically significant.

Fig. 4 presents the results of the comparative study, as represented by four evaluation metrics. The p-values for the total trajectory, time cost, mean velocity, and collision metric are 0.58, 0.30, 0.26, and 0.03, respectively. The difference in the collision metric was statistically significant, as highlighted by its p-value of 0.03, which falls below the threshold of 0.05. Thus, the primary strength of our proposed system lies in its significant improvement in precision and safety.

V. DISCUSSIONS AND FUTURE WORK

The objective of this paper is to construct an ergonomic, intuitive, and reliable teleoperation framework for RAMS. To achieve this, we propose a novel DT-driven immersive teleoperation framework. This framework i) leverages the integration of DT methodology with MR, speech commands, and hand-tracking technology to foster ergonomic human-robot interaction, and ii) augments the real-time 3D understanding of the robot's states by superimposing the DT onto the microscope's operative view in head-mounted display. While focusing on the microscopic view, the operator can better perceive their relative positions in 3D space by observing the relative locations of the robot and the eyeball model within the DT. Paired with augmented information for microsurgery, this system potentially enhances the safety of RAMS. Its viability was validated through user studies based on a trajectory following task using an eyeball model. The proposed framework not only augments the operator's situational awareness during microsurgical procedures but also eliminates the usage of mechanical linkage-based motion-tracking devices.

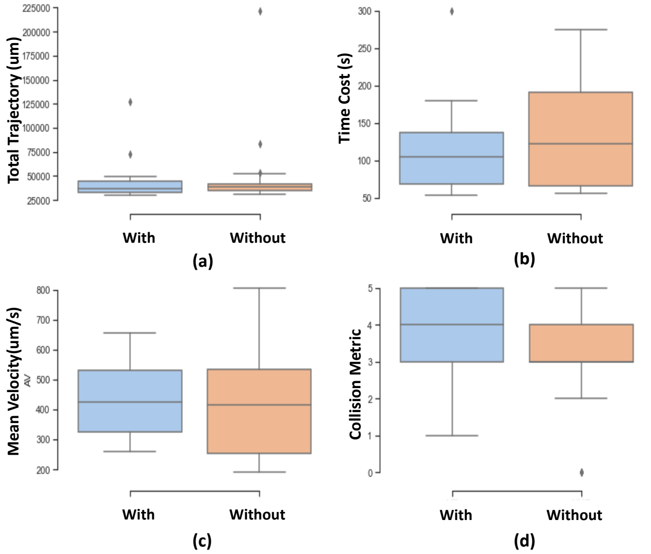

Fig. 4. Results analysis and visualization. Visualization of the box plot results of (a) Total Trajectory, (b) Time Cost, (c) Mean Velocity, (d) Collision Metric, under different experimental settings.

The outcomes of the user studies highlight the efficiency, intuitiveness, and superiority of our proposed DT-driven immersive teleoperation framework over traditional approaches without using DT. Through the provision of an immersive 3D visual experience via a head-mounted display, human operators are furnished with enhanced situational awareness. The results demonstrate how operators, even those without prior training, can effectively utilize this proposed system for RAMS. The proposed technology could facilitate more precise microsurgical maneuvers and potentially enhance surgical outcomes by enhancing the safety and efficiency of teleoperation.

Future efforts could also focus on integrating haptic feedback into the system to facilitate multi-modal feedback to the operator during bidirectional teleoperation [31]. The sense of haptics allows human operators to feel as if they are touching the target object during teleoperation, which can further enhance the safety and intuitiveness of the operation during RAMS [32]. Large Language Models (LLMs) [33] could be utilized to further enhance the ergonomics and interaction efficiency between the operator and the remote microsurgical robots.

ACKNOWLEDGEMENTS

The University of Bristol Ethics Committee granted ethical approval (ref No.10389) before initiating the study. The authors would like to acknowledge all participants who contributed to the user studies.

REFERENCES

[1] W. Zhang, H. Li, L. Cui, H. Li, X. Zhang, S. Fang, and Q. Zhang, "Research progress and development trend of surgical robot and surgical instrument arm," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 17, no. 5, p. e2309, 2021.

[2] D. Zhang, W. Si, W. Fan, Y. Guan, and C. Yang, "From teleoperation to autonomous robot-assisted microsurgery: A survey," *Machine Intelligence Research*, vol. 19, no. 4, pp. 288–306, 2022.

[3] D. Zhang, Y. Guo, J. Chen, J. Liu, and G.-Z. Yang, "A handheld master controller for robot-assisted microsurgery," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 394–400.

[4] A. Gao, R. R. Murphy, W. Chen, G. Dagnino, P. Fischer, M. G. Gutierrez, D. Kundrat, B. J. Nelson, N. Shamsudhin, H. Su *et al.*, "Progress in robotics for combating infectious diseases," *Science Robotics*, vol. 6, no. 52, p. eabf1462, 2021.

[5] D. Zhang, J. Chen, W. Li, D. B. Salinas, and G.-Z. Yang, "A microsurgical robot research platform for robot-assisted microsurgery research and training," *International journal of computer assisted radiology and surgery*, vol. 15, no. 1, pp. 15–25, 2020.

[6] W. Fan, X. Guo, E. Feng, J. Lin, Y. Wang, J. Liang, M. Garrad, J. Rossiter, Z. Zhang, N. Lepora *et al.*, "Digital twin-driven mixed reality framework for immersive teleoperation with haptic rendering," *IEEE Robotics and Automation Letters*, 2023.

[7] I. A. Tsokalo, D. Kuss, I. Kharabet, F. H. Fitzek, and M. Reisslein, "Remote robot control with human-in-the-loop over long distances using digital twins," in *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019, pp. 1–6.

[8] S. Bonne, W. Panitch, K. Dharmarajan, K. Srinivas, J.-L. Kincade, T. Low, B. Knoth, C. Cowan, D. Fer, B. Thananjeyan *et al.*, "A digital twin framework for telesurgery in the presence of varying network quality of service," in *2022 IEEE 18th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2022, pp. 1325–1332.

[9] D. Zhang, J. Liu, L. Zhang, and G.-Z. Yang, "Design and verification of a portable master manipulator based on an effective workspace analysis framework," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 417–424.

[10] L. Qian, J. Y. Wu, S. P. DiMaio, N. Navab, and P. Kazanzides, "A review of augmented reality in robotic-assisted surgery," *IEEE Transactions on Medical Robotics and Bionics*, vol. 2, no. 1, pp. 1–16, 2019.

[11] J. T. Gibby, S. A. Swenson, S. Cvetko, R. Rao, and R. Javan, "Head-mounted display augmented reality to guide pedicle screw placement utilizing computed tomography," *International journal of computer assisted radiology and surgery*, vol. 14, pp. 525–535, 2019.

[12] R. P. Kumar, E. Pelanis, R. Bugge, H. Brun, R. Palomar, D. L. Aghayan, A. A. Fretland, B. Edwin, and O. J. Elle, "Use of mixed reality for surgery planning: Assessment and development workflow," *Journal of Biomedical Informatics*, vol. 112, p. 100077, 2020.

[13] D. Zhang, J. Liu, L. Zhang, and G. Z. Yang, "Hamlyn crm: a compact master manipulator for surgical robot remote control," *International Journal of Computer Assisted Radiology & Surgery*, vol. 15, no. 3, pp. 503–514, 2020.

[14] J. Lin, X. Guo, W. Fan, W. Li, Y. Wang, J. Liang, J. Liu, W. Liu, L. Wei, and D. Zhang, "Tims: A tactile internet-based micromanipulation system with haptic guidance for surgical training," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 5114–5121.

[15] D. Zhang, J. Liu, A. Gao, and G.-Z. Yang, "An ergonomic shared workspace analysis framework for the optimal placement of a compact master control console," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2995–3002, 2020.

[16] H. Ahmed and L. Devoto, "The potential of a digital twin in surgery," *Surgical Innovation*, vol. 28, no. 4, pp. 509–510, 2021.

[17] H. Laaki, Y. Miche, and K. Tammi, "Prototyping a digital twin for real time remote control over mobile networks: Application of remote surgery," *Ieee Access*, vol. 7, pp. 20 325–20 336, 2019.

[18] K. Hagmann, A. Hellings-Kuß, J. Klodmann, R. Richter, F. Stulp, and D. Leidner, "A digital twin approach for contextual assistance for surgeons during surgical robotics training," *Frontiers in Robotics and AI*, p. 305, 2021.

[19] H. Shu, R. Liang, Z. Li, A. Goodridge, X. Zhang, H. Ding, N. Nagururu, M. Sahu, F. X. Creighton, R. H. Taylor *et al.*, "Twin-s: a digital twin for skull base surgery," *International Journal of Computer Assisted Radiology and Surgery*, pp. 1–8, 2023.

[20] A. Kolagunda, S. Sorensen, S. Mehralivand, P. Saponaro, W. Treible, B. Turkbey, P. Pinto, P. Choyke, and C. Kambhamettu, "A mixed reality guidance system for robot assisted laparoscopic radical prostatectomy," in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis: First International Workshop, OR 2.0 2018, 5th International Workshop, CARE 2018, 7th International Workshop, CLIP 2018, Third International Workshop, ISIC 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16 and 20, 2018, Proceedings 5*. Springer, 2018, pp. 164–174.

[21] M. Dianatfar, J. Latokartano, and M. Lanz, "Review on existing vr/ar solutions in human–robot collaboration," *Procedia CIRP*, vol. 97, pp. 407–411, 2021.

[22] H. F. Al Janabi, A. Aydin, S. Palaneer, N. Macchione, A. Al-Jabir, M. S. Khan, P. Dasgupta, and K. Ahmed, "Effectiveness of the hololens mixed-reality headset in minimally invasive surgery: a simulation-based feasibility study," *Surgical Endoscopy*, vol. 34, pp. 1143–1149, 2020.

[23] J. Gerup, C. B. Soerensen, and P. Dieckmann, "Augmented reality and mixed reality for healthcare education beyond surgery: an integrative review," *International journal of medical education*, vol. 11, p. 1, 2020.

[24] L. Lu, H. Wang, P. Liu, R. Liu, J. Zhang, Y. Xie, S. Liu, T. Huo, M. Xie, X. Wu *et al.*, "Applications of mixed reality technology in orthopedics surgery: a pilot study," *Frontiers in Bioengineering and Biotechnology*, vol. 10, 2022.

[25] G. Reis, M. Yilmaz, J. Rambach, A. Pagani, R. Suarez-Ibarrola, A. Miernik, P. Lesur, and N. Minaskan, "Mixed reality applications in urology: Requirements and future potential," *Annals of Medicine and Surgery*, vol. 66, p. 102394, 2021.

[26] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[27] D. Zhang, B. Xiao, B. Huang, L. Zhang, J. Liu, and G.-Z. Yang, "A self-adaptive motion scaling framework for surgical robot remote control," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 359–366, 2018.

[28] Z. Lin, A. Gao, X. Ai, H. Gao, Y. Fu, W. Chen, and G.-Z. Yang, "Arei: Augmented-reality-assisted touchless teleoperated robot for endoluminal intervention," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 5, pp. 3144–3154, 2021.

[29] D. Zhang, Z. Wu, J. Chen, A. Gao, X. Chen, P. Li, Z. Wang, G. Yang, B. Lo, and G.-Z. Yang, "Automatic microsurgical skill assessment based on cross-domain transfer learning," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 4148–4155, 2020.

[30] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task load Index): Results of empirical and theoretical research," in *Advances in psychology*. Elsevier, 1988, vol. 52, pp. 139–183.

[31] X. Guo, F. McFall, P. Jiang, J. Liu, N. Lepora, and D. Zhang, "A lightweight and affordable wearable haptic controller for robot-assisted microsurgery," *Sensors*, vol. 24, no. 9, p. 2676, 2024.

[32] J. A. Griffin, W. Zhu, and C. S. Nam, "The role of haptic feedback in robotic-assisted retinal microsurgery systems: a systematic review," *IEEE Transactions on Haptics*, vol. 10, no. 1, pp. 94–105, 2016.

[33] U. B. Karli, J.-T. Chen, V. N. Antony, and C.-M. Huang, "Alchemist: Llm-aided end-user development of robot applications," in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 2024, pp. 361–370.