2023 IEEE International Conference on Robotics and Automation (ICRA 2023)
May 29 - June 2, 2023. London, UK

Augmented Reality-Assisted Robot Learning Framework for Minimally Invasive Surgery Task

Junling Fu[1], Maria Chiara Palumbo[1], Elisa Iovene[1], Qingsheng Liu[1,2], Ilaria Burzo[1], Alberto Redaelli[1], Giancarlo Ferrigno[1] and Elena De Momi[1]

Abstract— **This paper presents an Augmented Reality (AR)-assisted robot learning framework for Minimally Invasive Surgery (MIS) tasks. The proposed framework exploits an external optical tracking system to collect human demonstration. Gaussian Mixture Model (GMM) and Gaussian Mixture Regression (GMR) are utilized to encode and generate a robust desired trajectory for transferring to the real robot for the MIS task. The HoloLens 2 Head-Mounted-Display (HMD) is integrated for intuitive visualization of the robot configuration under the constraint of a small incision on the patient's abdominal cavity during the demonstration phase. Experiments are conducted to verify the feasibility and performance of the proposed framework and compared it with the kinesthetic teaching-based modality in a tumor resection MIS task. The results illustrate that the proposed AR-assisted robot learning framework requires lower workload demand, achieves higher performance and efficiency, and ensures the feasibility of the learned results for reproduction on a real robot for MIS tasks.**

I. INTRODUCTION

Robot-Assisted Surgery (RAS) technique, which has great potential in improving positioning accuracy, consistency, and reduce fatigue for the surgeon, has been integrated into many medical scenarios over the past decades [1], [2]. Currently, conventional RAS allows the surgeon to manipulate the surgical robot to complete surgical tasks, such as knotting, puncture, tumor resection, etc, using either teleoperation control or hands-on control models [3].

Although hands-on or teleoperation control modalities allow surgeons to perform surgical tasks with better outcomes, several critical issues still remain. For instance, safety is the most critical issue when the surgeon manually operates the real robot because of the possible unexpected collision with surrounding tissue or organs [4]. One potential solution to deal with the above-mentioned issues is to enhance the Level of Autonomy (LoA) of surgical robotics [5]–[7].

Robot Learning from Demonstrations (LfD) is a promising paradigm in which robots acquire new skills by learning to imitate an expert, mainly categorized as kinesthetic teaching, teleoperation, and passive observation modalities [8]. LfD has the capability of facilitating non-expert robot teaching which makes it popular for robot learning in many applications. The LoA of robotic systems can be further enhanced by transferring the learned results to the physical robot [9].

[1] Junling Fu, Maria Chiara Palumbo, Elisa Iovene, Qingsheng Liu, Ilaria Burzo, Alberto Redaelli, Giancarlo Ferrigno, and Elena De Momi are with Department of Electronics, Information and Bioengineering, Politecnico di Milano, 20133, Milan, Italy. (junling.fu@polimi.it)

[2] Qingsheng Liu is also with the College of Information Science and Engineering, Ocean University of China, 266100, Qingdao, China.

Kinesthetic teaching-based robot learning is a physical human-robot interaction modality, which allows the operator to manipulate the robot to perform demonstrations for specific tasks [10]. Calinon *et al.* implemented a robot learning interface on the flexible robotic-assisted surgical application to enhance surgical robotics autonomy in an MIS cutting task [11]. Rafii-Tari *et al.* proposed an LfD framework for a robot-assisted cardiac catheterization task. The demonstration was performed by expert surgeons and the GMM and GMR were utilized to model and generate a smooth trajectory for robot execution [12]. However, precision demonstration for delicate tasks is difficult to be achieved due to the non-negligible mass, inertia, and configuration of the real robot, especially for redundant manipulators [13].

Teleoperation control-based robot learning modality does not require humans to be co-present with the robot during the teaching phase [14], but external input devices, like the joystick, haptic device [15], keyboard, etc are required. Zhang *et al.* proposed a novel simulation-to-real robot learning framework for surgical subtasks automation in MIS and Gaussian Process Regression (GPR) was employed to model the learned trajectory [16]. In addition, the passive observation-based modality is also a practical solution, including optical, electromagnetic tracking, etc [17]. Passive markers were attached to the surgical gripper and tracked by the Nokov motion capture device to collect the suturing skills from the human surgeon and transfer them to the physical robot [18]. Huang *et al.* designed a vision-based LfD framework for multi-robot cooperation and task reproduction in a robot-assisted bimanual sewing task, infrared markers were adopted to record human demonstrated trajectories [19]. Typically, the motion scaling and mapping strategy between the input device and the real robot should be properly established to ensure the effectiveness of human demonstration and robot reproduction [14].

Recently, AR has been integrated into many robotic applications by superimposing a computer-generated hologram model over the physical one [20]–[22]. AR-based robot learning modality provides the user with a more intuitive interface by showing the holographic robot configuration and verifying robot behavior before transferring it to the real robot. Diehl *et al.* developed an AR-assisted visualization framework to verify the feasibility of the robot learning results, also the robot behavior was verified in the AR environment before reproducing on the real robot [23]. Luebbers *et al.* integrated the novel AR visualization framework for constrained robot learning from demonstration, which

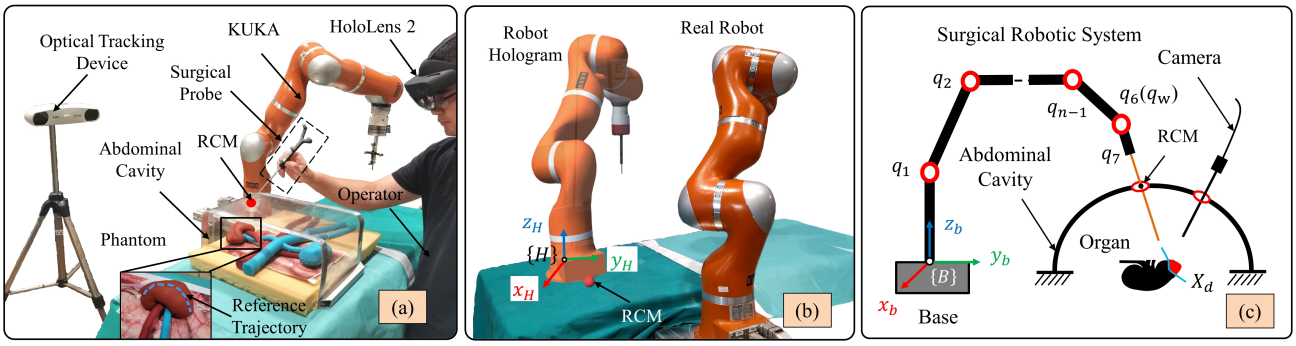

Fig. 1. Overview of proposed AR-assisted robot learning framework for MIS task. (a) Phase 1: Human performs demonstrations for a tumor resection task on a phantom with the hand-held surgical probe, an external optical tracking system records the demonstrated trajectories; (b) User's view with HoloLens2 HMD (Before alignment); (c) Phase 2: Surgical robot semi-autonomously reproduces the learned trajectory $\boldsymbol{X}_d$ for the MIS task.

also allowed the users to maintain, update, and adapt the learned skills via the AR interface [24]. Compared with the kinesthetic teaching-based modality, the AR-assisted robot learning framework had been demonstrated to require lower effort, provide more intuitive and transparent experiences, as well as efficiency improvement [25].

Considering the aforementioned issues, this paper proposes a novel AR-assisted robot learning framework for MIS tasks. The contributions are summarized as follows: (i) The robot can learn from human demonstration using an optical tracking system instead of manipulating the physical robot; (ii) AR-assisted visualization is integrated into the framework to intuitively display robot configuration and the Remote Center of Motion (RCM) constraint during demonstration phase; (iii) The real robot semi-autonomously reproduce the learned results for the tumor resection MIS task.

The remainder of this paper is organized as follows. Section II describes the methodology, including coordinates transformation and system calibration, proposed AR-assisted framework, robot learning, and reproduction. Section III presents the system design and experimental protocol. Following that, Section IV gives the results and discussion. Finally, Section V concludes this work.

II. METHODOLOGY

A. Overview of Proposed Framework

The proposed AR-assisted robot learning framework for MIS tasks mainly includes two phases, namely the demonstration phase and the reproduction phase, which are illustrated in Fig. 1 (a) and Fig. 1 (c), respectively. In detail, $\{\boldsymbol{B}\}$ is the robot base coordinate, $\boldsymbol{q}_i (i = 1,2,\ldots,7)$ is the robot joints, $\boldsymbol{X}_d \in \mathbb{R}^3$ is the learned desired trajectory from human demonstrations. The RCM is a kinematics constraint that is introduced by the small incision hole on the patient's abdominal cavity for inserting the surgical instrument and should be always respected during the MIS task.

Firstly, during the demonstration phase, as shown in Fig. 1 (a), the expert performs ideal demonstrated trajectories $\boldsymbol{X} = \{X^i \mid i = 1,2,\ldots,N\}$, where N is the number of demonstrations, on the soft silicon phantom using a localized handheld tool equipped with remotely tracked passive markers. The demonstrated trajectories are recorded using the optical tracking system with respect to the robot base $\{\boldsymbol{B}\}$. Moreover, as shown in Fig. 1(a) and (b), the HoloLens 2 AR-HMD shows the configuration of the surgical robot and the RCM position during this phase. The RCM position is collected before starting the demonstration from the patient's abdominal cavity. With the AR-assisted visualization interface in the proposed framework, the robot configuration verification of human-demonstrated trajectories on the physical robotic system can be verified in the meantime during the demonstration phase.

Afterwards, during the reproduction phase, the real robot semi-autonomously reproduces the learned results (Manual insertion first), as shown in Fig. 1(c). Firstly, the robot is manually inserted into the patient's body via the small incision hole (RCM) on the abdominal cavity and the robot tip is located to the start point of the learned trajectory $\boldsymbol{X}_d$. Then, the robot performs two levels of tasks, the primary task is to track the desired trajectory $\boldsymbol{X}_d$, while maintaining the RCM constraint.

B. Coordinates Transformation and System Calibration

When the surgeon performs demonstrations, the handheld surgical probe tip's positions are recorded by the optical tracking device. The coordinate transformations between the robot, the optical tracking device, and the surgical probe are illustrated in Fig. 2(a). The transformation matrix from robot base $\{\boldsymbol{B}\}$ to surgical probe tip $\{\boldsymbol{P}\}$ is represented as ${}^B_P\boldsymbol{T}$. From Fig. 2(a), ${}^B_P\boldsymbol{T}$ can be calculated as follows:

$$ {}^B_P\boldsymbol{T} = {}^B_N\boldsymbol{T} * {}^N_S\boldsymbol{T} * {}^S_P\boldsymbol{T} \tag{1} $$

where ${}^B_N\boldsymbol{T}$, ${}^N_S\boldsymbol{T}$, and ${}^S_P\boldsymbol{T}$ are the transformation matrix from $\{\boldsymbol{B}\}$ to the optical tracking system $\{\boldsymbol{N}\}$, from $\{\boldsymbol{N}\}$ to the stylus of the surgical probe $\{\boldsymbol{S}\}$, and from $\{\boldsymbol{S}\}$ to the tip of the surgical probe $\{\boldsymbol{P}\}$, respectively. From Fig. 2, ${}^B_N\boldsymbol{T}$ is expressed as follows:

$$ {}^B_N\boldsymbol{T} = {}^B_T\boldsymbol{T} * {}^T_R\boldsymbol{T} * {}^R_N\boldsymbol{T} \tag{2} $$

where ${}^T_R\boldsymbol{T}$ represents the matrix from the Tool Central Point (TCP) $\{\boldsymbol{T}\}$ of the robot to the reference tool $\{\boldsymbol{R}\}$. ${}^R_N\boldsymbol{T}$ is the

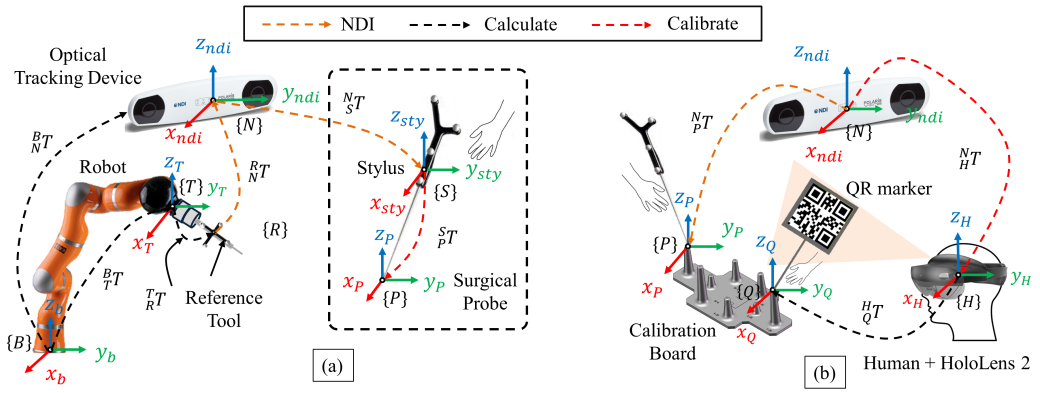

Fig. 2. System coordinate transformation and calibration of the proposed AR-assisted robot learning framework. (a) The coordinate transformation between different components. (b) Calibration between the optical tracking device and the HoloLens 2 HMD.

matrix from the reference tool to the optical tracking device.

By moving the robot to different configurations in its workspace, the following equation can be obtained:

$$ {}^{B}_{T}\boldsymbol{T}_2 * {}^{B}_{T}\boldsymbol{T}_1^{-1} * {}^{B}_{N}\boldsymbol{T} = {}^{B}_{N}\boldsymbol{T} * {}^{R}_{N}\boldsymbol{T}_2^{-1} * {}^{R}_{N}\boldsymbol{T}_1 \tag{3} $$

Further, Equation (3) can be simplified of solving the following formulation, which is fomulated as:

$$ \boldsymbol{AX} = \boldsymbol{XB} \tag{4} $$

where $\boldsymbol{A} = {}^{B}_{T}\boldsymbol{T}_2 * {}^{B}_{T}\boldsymbol{T}_1^{-1}$, $\boldsymbol{B} = {}^{R}_{N}\boldsymbol{T}_2^{-1} * {}^{R}_{N}\boldsymbol{T}_1$, and $\boldsymbol{X} = {}^{B}_{N}\boldsymbol{T}$. After moving the robot to several configurations, a set of equations in the formate of Equation (3) can be obtained. Tsai's algorithm [26] is adopted to solve the matrix ${}^{B}_{N}\boldsymbol{T}$.

C. Augment Reality-Assisted Visualization

In our framework, AR is integrated to intuitively display the robot configuration and the RCM position. The alignment can be achieved after the calibration between the HoloLens 2 HMD $\{\boldsymbol{H}\}$ and the optical tracking device $\{\boldsymbol{N}\}$. The tools used for calibration are shown in Fig. 2(b) Furthermore, the detailed calibration steps can be found in [27], [28]. The 3D point set in these two coordinate system are recorded as $\boldsymbol{P}'_H = \{\boldsymbol{P}_{Hi}, i = 0,1,2,\ldots,M\}$ and $\boldsymbol{P}'_N = \{\boldsymbol{P}_{Ni}, i = 0,1,2,\ldots,M\}$.

In Fig. 1(b), the alignment problem can be transformed into finding the optimal virtual to real alignment rotation matrix $\boldsymbol{R} \in \mathbb{R}^{3\times3}$ and translation matrix $\boldsymbol{T} \in \mathbb{R}^3$ to minimize the distance between the two point sets. The objective function is defined as follows:

$$ \min_{\boldsymbol{R},\boldsymbol{T}} J = \sum_{i=1}^{M} \|\boldsymbol{P}_{Hi} - (\boldsymbol{R}\boldsymbol{P}_{Ni} + \boldsymbol{T})\|^2 \tag{5} $$

Then, the Singular Value Decomposition (SVD) algorithm [29] is utilized to calculate $\boldsymbol{R}$ and $\boldsymbol{T}$ in Equation (5).

D. Robot Learning from Human Demonstration

Then, GMM and GMR [30] extract the features from human demonstration. The demonstrated trajectory is expressed as $\boldsymbol{X} = \{X^i \mid i = 1,2,\ldots,N\}$, where N is the number of demonstrations. Each demonstration is composed of the time series ξ_t^i and the 3D position ξ_p^i. Hence, it is recorded as $\boldsymbol{X}^i = \{\xi_t^i, \xi_p^i\}$, where $\boldsymbol{\xi}_t^i = \{\xi_t^i(t) \mid t = 1,2,\ldots,T\}$ and $\boldsymbol{\xi}_p^i = \{\xi_p^i(t) \mid t = 1,2,\ldots,T\}$. Then, the probability density of the demonstrated trajectories is expressed as:

$$ P(\xi_i) = \sum_{k=1}^{K} \omega_k N(\xi_i \mid \mu_k, \Sigma_k) \ \text{ s.t. } \sum_{k=1}^{K} \omega_k = 1, \omega_k \in [0,1] \tag{6} $$

where k is the number of GMM components, ω_k, μ_k and Σ_k denote the prior probability, the mean matrix, and the covariance matrix of the Gaussian component k. $N(\xi_i \mid \mu_k, \Sigma_k)$ is the k-th Gaussian distribution of conditional probability.

$$ N(\xi_i \mid \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{3/2}|\Sigma_k|^{1/2}} e^{-1/2(\xi_i-\mu_k)^T \Sigma_k^{-1}(\xi_i-\mu_k)} \tag{7} $$

The Expectation-Maximization (EM) algorithm [31] is used to estimate ω_k, μ_k and Σ_k in Equation (7). μ_k and Σ_k of the k-th Gaussian component is calculated by:

$$ \mu_k = \begin{bmatrix} \mu_t^k; \mu_p^k \end{bmatrix}, \Sigma_k = \begin{bmatrix} \Sigma_{t,t}^k & \Sigma_{t,p}^k \\ \Sigma_{t,p}^{k\,T} & \Sigma_{p,p}^k \end{bmatrix} \tag{8} $$

The mixed conditional probability is estimated by:

$$ P(\xi_p \mid \xi_t) = \sum_{k=1}^{K} \beta_k N\left(\xi_t \mid \hat{\xi}_p^k, \hat{\Sigma}_{p,p}^k\right) \tag{9} $$

where β_k is represented as follows:

$$ \beta_k = \frac{\omega_k * N\left(\xi_t \mid \mu_t^k, \Sigma_{t,t}^k\right)}{\sum_{i=1}^{K} \omega_i * N\left(\xi_t \mid \mu_t^i, \Sigma_{t,t}^i\right)} \tag{10} $$

$\hat{\xi}_p^k$ and $\hat{\Sigma}_{p,p}^k$ are the estimated mean matrix and covariance matrix, which is calculated by:

$$ \begin{aligned} \hat{\xi}_p^k &= \mu_p^k + \Sigma_{t,p}^k \left(\Sigma_{t,t}^k\right)^{-1} \left(\xi_t - \mu_t^k\right) \\ \hat{\Sigma}_{p,p}^k &= \Sigma_{p,p}^k - \Sigma_{p,t}^k \left(\Sigma_{t,t}^k\right)^{-1} \Sigma_{t,p}^k \end{aligned} \tag{11} $$

Then, GMR generates a robust desired trajectory $\boldsymbol{X}_d$, which takes a time step ξ_t as input and calculates the estimated conditional expectation of the 3D position of the

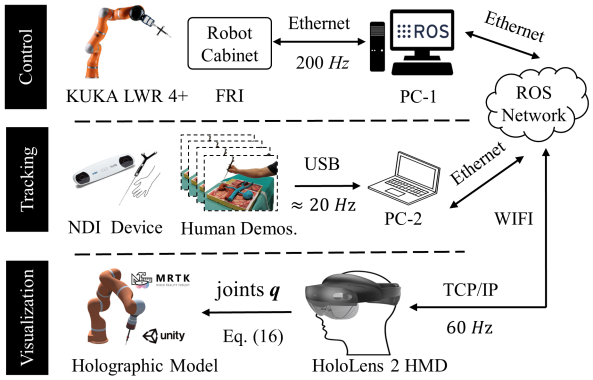

Fig. 3. System configuration and communication protocol.

surgical probe tip $\hat{\xi}_p$, which is calculated by $\hat{\xi}_p = \sum_{k=1}^{K} \beta_k \hat{\xi}_p^k$. The generated desired trajectory $\boldsymbol{X}_d$ from the human demonstrations is expressed as $\boldsymbol{X}_d = \left[\xi_t, \hat{\xi}_p\right]$ at the time step ξ_t.

E. Robot Reproduction with RCM Constraint

In Fig. 1 (c), the forward kinematics of a 7 Degree-of-Freedoms (DoFs) serial redundant robot is expressed as:

$$\boldsymbol{X}_d = \varphi(\boldsymbol{q}) \tag{12}$$

where $\boldsymbol{X}_d \in \mathbb{R}^3$ is the position of the robot end-effector in the Cartesian space, and $\boldsymbol{q} \in \mathbb{R}^7$ is a vector of the joint angles.

Furthermore, a null-space projector is implemented to deal with the task priority issue by leveraging robot's redundancy. The velocity of the robot joints is calculated by:

$$\dot{\boldsymbol{q}} = \boldsymbol{J}^{\dagger} * \dot{\boldsymbol{X}}_d + \boldsymbol{N} * \dot{\boldsymbol{q}}_w \tag{13}$$

where $\boldsymbol{J}^{\dagger} \in \mathbb{R}^{7\times3}$ donates the general inverse of Jacobian matrix $\boldsymbol{J}$ and one possible solution is to utilize the pseudo-inverse of $\boldsymbol{J}$ to calculate it as $\boldsymbol{J}^{\dagger} = \boldsymbol{J}^T \left(\boldsymbol{J} * \boldsymbol{J}^T\right)^{-1}$. where$\boldsymbol{N} \in \mathbb{R}^{7\times7}$ is a null-space projection matrix and calculated by:

$$\boldsymbol{N} = \boldsymbol{I} - \boldsymbol{J}^{\dagger}\boldsymbol{J} \tag{14}$$

where $\boldsymbol{I} \in \mathbb{R}^{7\times7}$ is the identity matrix. $\dot{\boldsymbol{q}}_w \in \mathbb{R}^7$ represents the wrist joint velocity and is calculated by

$$\dot{\boldsymbol{q}}_w = \dot{\boldsymbol{X}}_w / \boldsymbol{J}_w \tag{15}$$

where $\dot{\boldsymbol{X}}_w \in \mathbb{R}^3$ donates the robot's wrist velocity in Cartesian space and $\boldsymbol{J}_w \in \mathbb{R}^{3\times7}$ is the Jacobian matrix from the robot base system $\{B\}$ to the wrist $\boldsymbol{q}_w(q_6)$. The wrist velocity $\dot{\boldsymbol{X}}_w$ of the robot is calculated using the method in [32].

III. Systems and Experimental Setup

A. System Description

1) Hardware System: The hardware system of the proposed framework is shown in Fig. 1 (a) and mainly consists: (1) An external optical tracking system (NDI Polaris Vicra, Northern Digital Inc., Canada), with 0.5*mm* tracking and repetition accuracy at 95% confidence interval, is utilized to track 3D positions of the handheld surgical probe; (2) An Optical See-Through HoloLens 2 HMD AR glass (Microsoft, WA, USA) is adopted to visualize hologram robot and RCM position; (3) A serial lightweight redundant robot (LWR 4+, KUKA, Germany) with 7-DoFs is utilized.

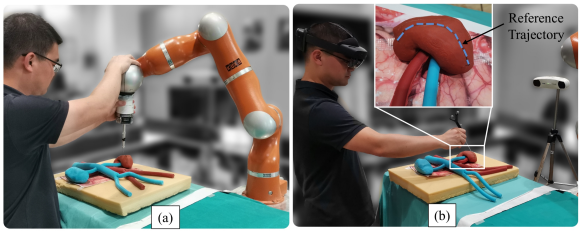

Fig. 4. Comparison experiment setup with two robot learning methods. (a) *M1*, Kinesthetic teaching-based modality; (b) *M2*, AR-assisted modality.

2) Software System: As shown in Fig. 3, the Robot Operating System (ROS) is employed with Ubuntu desktop 16.04 to facilitate the data exchange. PC-1 is connected to the KUKA LWR 4+ robot through a User Datagram Protocol (UDP) and controls the robot through the FastResearchInterface (FRI) Library. The PC-2, with Windows 10 installed and i7-9750H CPU 2.60GHz and 16GB RAM, is connected to the optical tracking system via USB and the ROS network using the Ethernet cable. The probe tip position is sent to HoloLens2 via TCP/IP protocol and wireless network. The AR application is developed based on the Universal Windows Platform (UWP), with Unity3D v2020.3.27 LTS version and Mixed Reality Toolkit (MRTK) v2.8.2 employed.

B. Experimental Setup

To verify the performance of the proposed robot learning framework, a tumor resection MIS task scenario is employed and comparative experiment is conducted with traditional kinesthetic teaching-based modality.

1) Experimental Setup: As shown in Fig. 4(b), a reference trajectory on the patient's silicon phantom is illustrated as the profile of the tumor resection task. With Modality 1 (***M1***), the operator manipulates the real robot and follow the reference trajectory to perform demonstrations. The robot is running in the Gravity Compensation (GC) mode, with which the gravity force of the robot is compensated and the robot can be manually guided with little external force. The robot tip position is recorded during the demonstration phase. Modality 2 (***M2***), the operator performs the demonstrations with the proposed AR-assisted framework by manipulating the handheld surgical probe, which is tracked by the optical tracking device. During the demonstration process, the robot configuration is calculated and visualized in real-time under the RCM constraint, and the operator can observe the robot configuration from the holographic model.

2) Participants: $P = 10$ healthy participants (8 males and 2 females), aged between 24-31 (Avg = 26.3, Std = 2.16) are invited to participate the experiment. All of them gave their informed consent about the experiments. Before the experiments, they are instructed to be familiar with the tracking system, surgical probe, robot manipulation, and the AR HMD. Moreover, the participants did practice with the task trajectories to avoid unfair comparison. In

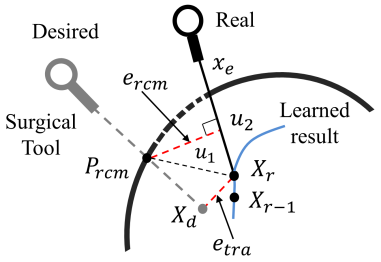

Fig. 5. Trajectory tracking error e_{tra} and RCM constraint error e_{rcm}.

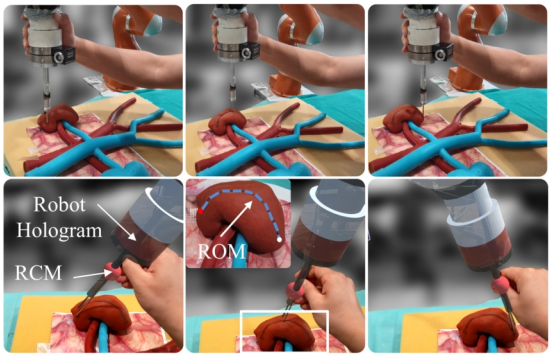

Fig. 6. Snapshots during human demonstration phase with ***M1*** and ***M2*** frameworks, respectively. Row 1 with ***M1*** and row 2 with ***M2***.

all the experiments, the robot is initialized with the same configuration. After completing the demonstration task, the participants were asked to fill in a NASA-TLX [33] form to evaluate the perceived cognitive workloads.

3) Performance Metrics: According to the 3 objectives of this paper, corresponding performance metrics are adopted to analyze the results. (i) The metrics to evaluate the performance during the demonstration phase, including the time cost $\boldsymbol{T}_{tot}$, trajectory length $\boldsymbol{L}_{tra}$:

$$L_{tra} = \sum_{t=1}^{T} \|\boldsymbol{X}_r - \boldsymbol{X}_{r-1}\| \tag{16}$$

where $\boldsymbol{X}_r$, $\boldsymbol{X}_{r-1}$ are the position of two samplings. And the average velocity V_{avg}, which is calculated by $V_{avg} = L_{tra}/T_{tot}$. (ii) The metrics to verify the real robot reproduction performance, including the tracking error $\boldsymbol{e}_{tra}$, and RCM constraint error $\boldsymbol{e}_{rcm}$ in Fig. 5 is expressed as:

$$e_{tra} = \|\boldsymbol{X}_d - \boldsymbol{X}_r\|, \quad e_{rcm} = \frac{\|\boldsymbol{u}_1 \times \boldsymbol{u}_2\|}{\|\boldsymbol{u}_2\|} \tag{17}$$

where $\boldsymbol{X}_d$, $\boldsymbol{X}_r$ are the desired and actual real robot tool tip position, $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ are the vector from RCM position P_{rcm} to $\boldsymbol{X}_r$ and vector of robot surgical tool. (iii) The total perceived workload from the NASA TLX questionnaire score is scaled in the range from 0 to 100. For statistical analysis, the non-parametric statistical significance test, namely the Wilcoxon Rank-Sum test is exploited. A significant difference can be assessed with p-value < 0.05.

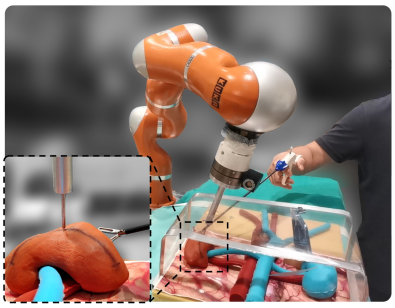

Fig. 7. Robot reproduction of human demonstrated trajectory for Minimally Invasive Surgery task in an semi-autonomous manner.

IV. Experimental Results

A. Robot Learning and Reproduction Results for MIS task

The surgical probe matrix ${}^{S}_{P}\boldsymbol{T}$ is obtained using the Pivot Calibration from 3D Slicer, with a maximum error of 0.384 mm. Then, Tasi's algorithm calculates the transformation matrix between the NDI optical tracking system and the KUKA robot by adjusting robot to 50 different configurations. The eye-to-hand calibration result is verified with a median value of 1.93 ± 0.34 mm in 8 different Cartesian positions within the primary workspace of the NDI device, in the dimension of approximately $60 \times 50 \times 50$ cm. The RCM position is collected as [-334.3, 523.5, 224.8] mm.

During the demonstration phase, the robot configuration and the RCM constraint can be observed from the HoloLens2 HMD in the meantime, as shown in Fig. 6. Also, the Range of Motion (ROM) in Fig. 6 illustrates the start point (red dot) and the end point (white dot) of the reference trajectory for the MIS task. Benefits from the real-time display of the AR robot hologram, the demonstrated trajectories are feasible for transferring to the real robot since the RCM is fulfilled during the demonstration phase.

Fig. 7 gives an illustration of the scenario during the real robot reproduction phase with the learned trajectory. Fig. 8(a) illustrates the demonstration results for the tumor resection task with 6 repetitions. Then, Fig. 8(b) shows the GMR result and the real robot reproduction trajectory. Furthermore, the trajectory tracking errors $\boldsymbol{e}_{tra}$ and the RCM error $\boldsymbol{e}_{rcm}$ for the reproduction of the learned desired trajectory $\boldsymbol{X}_d$ are calculated and shown in Fig. 8(c) with 0.4 mm and 4.5 mm maximum, respectively.

B. Comparison Results between Different Modalities

The comparison results of the robot learning performance metrics with these two modalities are illustrated in Fig. 9 including the trajectory length L_{tra}, the time cost T_{tot}, and the average velocity V_{avg} with both $M1$ and $M2$ modalities. As shown in Fig. 9(a), the AR-based robot learning framework achieves a lower Median and Standard Deviation (Std) value in the trajectory length L_{tra} performance index. $p < 0.01$ is observed between these two robot learning modalities, which indicates that the participants can perform demonstrations on the phantom using the AR-based framework with higher consistency. Besides, a significant difference with $p < 0.001$

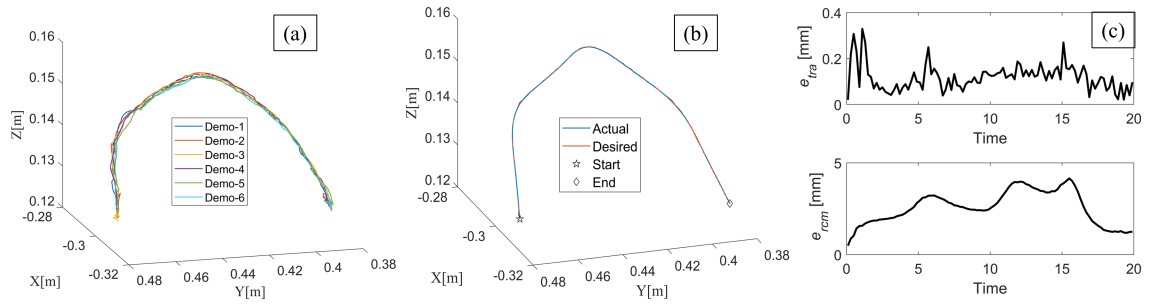

Fig. 8. Human demonstration and robot reproduction results. (a) Human demonstrated trajectories; (b) Regression results from GMR and real robot trajectory; (c) Trajectory tracking error and RCM error during reproduction phase;

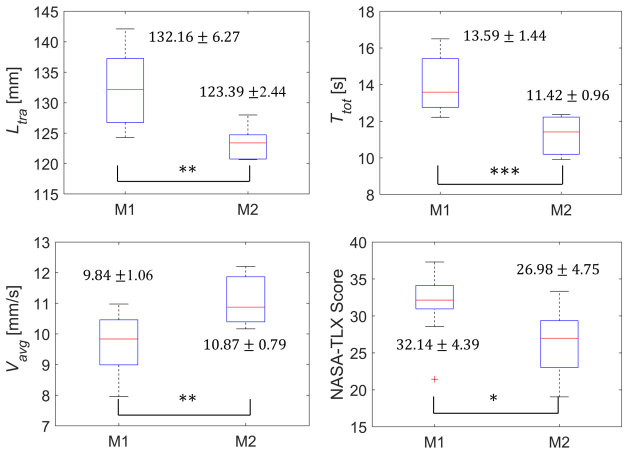

Fig. 9. Performance matrix comparison results. (a) Trajectory length L_{tra}; (b) Time cost T_{tot}; (c) Average velocity V_{avg}; (d) Total NASA-TLX user evaluation results. ($*$, $p < 0.05$; $**$, $p < 0.01$; $***$, $p < 0.001$)

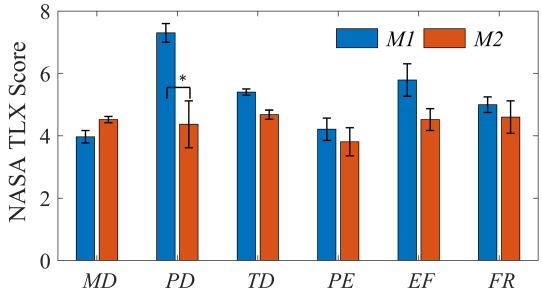

Fig. 10. NASA-TLX user evaluation comparison results of M1 and M2.

in the performance index time cost T_{tot} is observed, which means a higher demonstration efficiency for the MIS task. Furthermore, the average velocity V_{avg} in Fig. 9(c) for performing demonstrations is also higher than the kinesthetic teaching-based one, hence, a higher demonstration efficiency is achieved with the proposed framework for MIS task. This is also benefits from the lightweight and portable priorities of the NDI surgical probe adopted in our framework, and avoid suffering from the mass, inertia, etc, issues compared with the kinesthetic-based modality.

C. NASA-TLX User Evaluation Results

The total perceived workload during the task is shown in Fig. 9 (d), the kinesthetic teaching-based and the proposed robot learning framework received a score of 32.14 ± 4.39 and 26.98 ± 4.75, respectively. Also, the Wilcoxon Rank-Sum test calculates the p-value for the total score, which is 0.0231. The experimental results demonstrated that significant differences can be observed between these two different robot learning modalities. The proposed framework requires a lower workload demand compared with the traditional kinesthetic teaching-based modality.

Furthermore, the detailed scores for the 6 indexes with $M1$ and $M2$ modalities are shown in Fig. 10, including Mental (MD), Physical (PD), and Temporal Demands (TD), Performance (PE), Effort (EF), and Frustration (FR). A significant difference can be observed in the PD item, with average scores of 7.30 ± 1.53 for $M1$ and 4.37 ± 1.56 for $M2$, and with p-value result of 0.0021, which means less physical demanded from the user with the proposed framework. Besides, the MD item for $M2$ is a little higher than $M1$ and the user claimed that the HoloLens2 HMD increases the MD. No significant difference is observed for the rest workload indexes, which illustrates the proposed robot learning framework can achieve comparable performance in these indexes.

V. CONCLUSION

In this paper, an AR-assisted robot learning framework is developed for MIS tasks. Experimental results illustrate that the robot can learn from human demonstrations with the proposed framework, and requires lower workload demand from the operator compared with the kinesthetic teaching-based modality. Moreover, benefiting from the AR visualization with the HoloLens 2 HMD, the robot hologram configuration with the RCM constraint fulfilled is verified and intuitively displayed during the demonstration phase. However, additional NDI and HMD devices are involved in the proposed robot learning framework, also the occlusion issue of the optical tracking devices is a critical limitation in the following works. Future work will focus on improving the comfort of the proposed framework and develop a human-robot shared control strategy to allow the human operator to adjust the learned trajectory during the real robot reproduction phase in a simultaneous manner.

REFERENCES

[1] E. De Momi and A. Segato, "Autonomous robotic surgery makes light work of anastomosis," *Science Robotics*, vol. 7, no. 62, p. eabn6522, 2022.

[2] A. V. Iordache, A. Casella, E. Iovene, J. Fu, F. Pessina, M. Riva, G. Ferrigno, L. S. Mattos, E. De Momi, *et al.*, "Envisioning robotic exoscope: Concept and preliminary results," in *Hamlyn Symposium on Medical Robotics*, 2022.

[3] T. Kastritsi and Z. Doulgeri, "A controller to impose a rcm for hands-on robotic-assisted minimally invasive surgery," *IEEE Transactions on Medical Robotics and Bionics*, vol. 3, no. 2, pp. 392–401, 2021.

[4] M. M. Marinho, B. V. Adorno, K. Harada, and M. Mitsuishi, "Dynamic active constraints for surgical robots using vector-field inequalities," *IEEE Transactions on Robotics*, vol. 35, no. 5, pp. 1166–1185, 2019.

[5] G.-Z. Yang, J. Cambias, K. Cleary, E. Daimler, J. Drake, P. E. Dupont, N. Hata, P. Kazanzides, S. Martel, R. V. Patel, *et al.*, "Medical robotics—regulatory, ethical, and legal considerations for increasing levels of autonomy," p. eaam8638, 2017.

[6] A. Attanasio, B. Scaglioni, E. De Momi, P. Fiorini, and P. Valdastri, "Autonomy in surgical robotics," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 4, pp. 651–679, 2021.

[7] X. Zhang, M. C. Palumbo, F. Perico, M. Magro, A. Fortuna, T. Magni, E. Votta, A. Segato, and E. De Momi, "Robotic actuation and control of a catheter for structural intervention cardiology," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 5907–5913.

[8] H. Ravichandar, A. S. Polydoros, S. Chernova, and A. Billard, "Recent advances in robot learning from demonstration," *Annual review of control, robotics, and autonomous systems*, vol. 3, pp. 297–330, 2020.

[9] P. Fiorini, K. Y. Goldberg, Y. Liu, and R. H. Taylor, "Concepts and trends in autonomy for robot-assisted surgery," *Proceedings of the IEEE*, vol. 110, no. 7, pp. 993–1011, 2022.

[10] J. Zhao, A. Giammarino, E. Lamon, J. M. Gandarias, E. De Momi, and A. Ajoudani, "A hybrid learning and optimization framework to achieve physically interactive tasks with mobile manipulators," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 8036–8043, 2022.

[11] S. Calinon, D. Bruno, M. S. Malekzadeh, T. Nanayakkara, and D. G. Caldwell, "Human–robot skills transfer interfaces for a flexible surgical robot," *Computer methods and programs in biomedicine*, vol. 116, no. 2, pp. 81–96, 2014.

[12] H. Rafii-Tari, J. Liu, S.-L. Lee, C. Bicknell, and G.-Z. Yang, "Learning-based modeling of endovascular navigation for collaborative robotic catheterization," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2013, pp. 369–377.

[13] J. G. Petersen, S. A. Bowyer, and F. R. y Baena, "Mass and friction optimization for natural motion in hands-on robotic surgery," *IEEE Transactions on Robotics*, vol. 32, no. 1, pp. 201–213, 2016.

[14] W. Si, N. Wang, and C. Yang, "A review on manipulation skill acquisition through teleoperation-based learning from demonstration," *Cognitive Computation and Systems*, vol. 3, no. 1, pp. 1–16, 2021.

[15] J. Fu, J. Zhang, Z. She, S. E. Ovur, W. Li, W. Qi, H. Su, G. Ferrigno, and E. De Momi, "Whole-body spatial teleoperation control of a hexapod robot in unstructured environment," in *2021 6th IEEE International Conference on Advanced Robotics and Mechatronics (ICARM)*. IEEE, 2021, pp. 93–98.

[16] D. Zhang, Z. Wu, J. Chen, R. Zhu, A. Munawar, B. Xiao, Y. Guan, H. Su, W. Hong, Y. Guo, G. S. Fischer, B. Lo, and G.-Z. Yang, "Human-robot shared control for surgical robot based on context-aware sim-to-real adaptation," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 7694–7700.

[17] J. Fu, M. Poletti, Q. Liu, E. Iovene, H. Su, G. Ferrigno, and E. De Momi, "Teleoperation control of an underactuated bionic hand: Comparison between wearable and vision-tracking-based methods," *Robotica*, vol. 11, no. 3, p. 61, 2022.

[18] D. Yang, Q. Lv, G. Liao, K. Zheng, J. Luo, and B. Wei, "Learning from demonstration: dynamical movement primitives based reusable suturing skill modelling method," in *2018 Chinese Automation Congress (CAC)*. IEEE, 2018, pp. 4252–4257.

[19] B. Huang, M. Ye, S.-L. Lee, and G.-Z. Yang, "A vision-guided multi-robot cooperation framework for learning-by-demonstration and task reproduction," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 4797–4804.

[20] L. Qian, C. Song, Y. Jiang, Q. Luo, X. Ma, P. W. Chiu, Z. Li, and P. Kazanzides, "Flexivision: Teleporting the surgeon's eyes via robotic flexible endoscope and head-mounted display," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 3281–3287.

[21] M. Boles, J. Fu, E. Iovene, C. Francesco, G. Ferrigno, E. DE MOMI, *et al.*, "Augmented reality and robotic navigation system for spinal surgery," in *Proceeding of the 11th Joint Workshop on New Technologies for Computer/Robot Assisted Surgery*, 2022, pp. 96–97.

[22] 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*, 2021.

[23] M. Diehl, A. Plopski, H. Kato, and K. Ramirez-Amaro, "Augmented reality interface to verify robot learning," in *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2020, pp. 378–383.

[24] M. B. Luebbers, C. Brooks, C. L. Mueller, D. Szafir, and B. Hayes, "Arc-lfd: Using augmented reality for interactive long-term robot skill maintenance via constrained learning from demonstration," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 3794–3800.

[25] C. P. Quintero, S. Li, M. K. Pan, W. P. Chan, H. M. Van der Loos, and E. Croft, "Robot programming through augmented trajectories in augmented reality," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 1838–1844.

[26] R. Y. Tsai, R. K. Lenz, *et al.*, "A new technique for fully autonomous and efficient 3 d robotics hand/eye calibration," *IEEE Transactions on robotics and automation*, vol. 5, no. 3, pp. 345–358, 1989.

[27] M. C. Palumbo, L. Morchi, V. Corbetta, A. Menciassi, E. De Momi, E. Votta, and A. Redaelli, "An easy and user independent augmented reality based navigation system for radiation-free interventional procedure," in *2022 International Symposium on Medical Robotics (ISMR)*. IEEE, 2022, pp. 1–7.

[28] M. C. Palumbo, S. Saitta, M. Schiariti, M. C. Sbarra, E. Turconi, G. Raccuia, J. Fu, V. Dallolio, P. Ferroli, E. Votta, *et al.*, "Mixed reality and deep learning for external ventricular drainage placement: A fast and automatic workflow for emergency treatments," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VII*. Springer, 2022, pp. 147–156.

[29] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-d point sets," *IEEE Transactions on pattern analysis and machine intelligence*, no. 5, pp. 698–700, 1987.

[30] S. Calinon, F. Guenter, and A. Billard, "On learning, representing, and generalizing a task in a humanoid robot," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 37, no. 2, pp. 286–298, 2007.

[31] L. Rozo, S. Calinon, D. G. Caldwell, P. Jimenez, and C. Torras, "Learning physical collaborative robot behaviors from human demonstrations," *IEEE Transactions on Robotics*, vol. 32, no. 3, pp. 513–527, 2016.

[32] H. Su, J. Sandoval, M. Makhdoomi, G. Ferrigno, and E. De Momi, "Safety-enhanced human-robot interaction control of redundant robot for teleoperated minimally invasive surgery," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 6611–6616.

[33] S. G. Hart, "Nasa-task load index (nasa-tlx); 20 years later," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, no. 9. Sage publications Sage CA: Los Angeles, CA, 2006, pp. 904–908.