A Cloud-Based Multi-User Immersive Teleoperation Framework for Robotic Surgery

Mohamad Shaaban[1], Nicola Piccinelli[2], Riccardo Muradore[2], Jesus Ortiz[1], Veronica Penza[1], Nikhil Deshpande[3], Darwin Caldwell[1], Leonardo De Mattos[1], and Yonas Tefera[1] *†‡

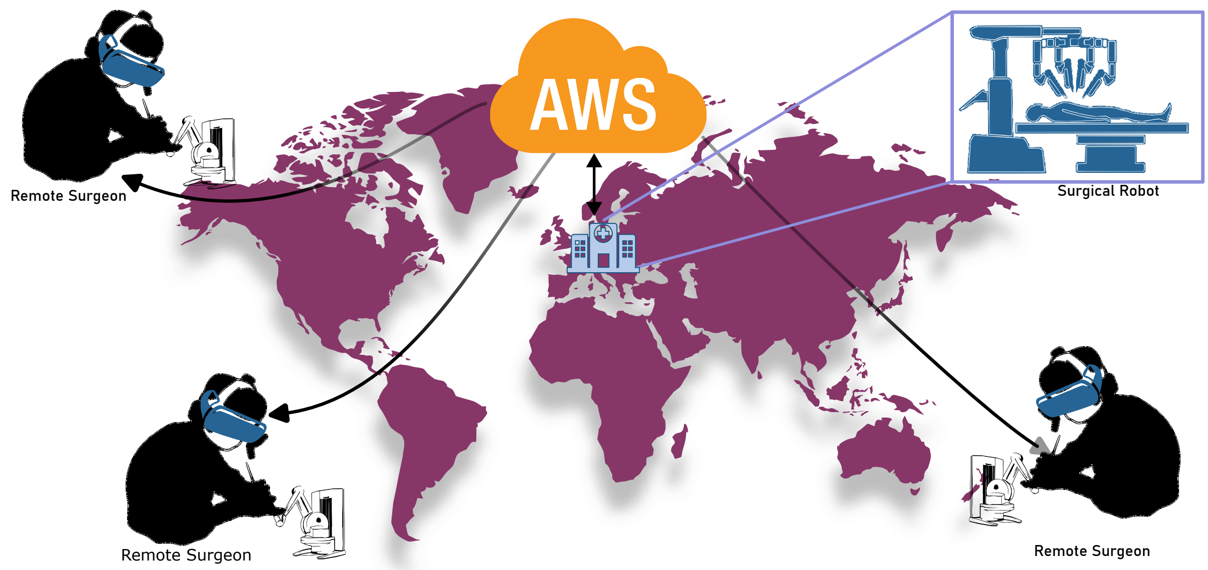

Figure 1: Distributed XR system for remote teleoperation of the da Vinci Research Kit, enabling three geographically distributed surgeons to collaborate in real time through a cloud-based coordination server.

Abstract

Extended reality (XR) technologies offer new opportunities for remote collaboration in robotic-assisted surgery; however, most existing systems focus on single-user interaction and local operation, thereby limiting their applicability to distributed surgical workflows. This paper presents a multi-user remote XR teleoperation framework for robotic surgery based on the Da Vinci Research Kit (dVRK), integrating immersive mixed reality visualization, unilateral haptic teleoperation, and cloud-based synchronization to enable real-time collaborative interaction among geographically distributed participants. The system employs a role-based teleoperation paradigm, in which a primary surgeon performs bi-manual instrument manipulation using a haptic device, a secondary surgeon controls the endoscopic camera through a dedicated haptic interface, and a support user facilitates coordination via intent communication and spatial annotation, without directly actuating the robot. Robot state, user inputs, and visual information are synchronized through a centralized cloud infrastructure hosted on Amazon Web Services (AWS), which enforces role-based access, separates control channels, and maintains a consistent shared XR representation across users under networked conditions. A preliminary experimental evaluation on the integration of this systems shows the round-trip teleoperation latency, including robot command transmission and force feedback, is approximately 38 $\pm$ 5.4 ms with a mean video transmission latency of 112 $\pm$ 13 ms.

Index Terms: Extended Reality (XR); Robotic-Assisted Surgery; Teleoperation; Multi-User Collaboration; Cloud Computing

*[1]Istituto Italiano di Tecnologia (IIT), Via Morego 30,16163 Genova, Italy

†[2]Department of Engineering for Innovation Medicine, University of Verona, Strada le Grazie 15, 37134, Verona, Italy

‡[3]School of Computer Science, University of Nottingham, Nottingham, NG8 1BB, UK

1 Introduction

Historically, Extended Reality (XR) technologies have been used to bridge the gap between digital and physical environments, enabling the creation of immersive and interactive experiences for various healthcare applications, including training, education, and pre-surgical planning. While these systems offered rich local immersion, their ability to support real-time remote experiences and meaningful interaction across distances was limited [21, 12, 11].

Over time, however, the increasing need to create a strong sense of presence in remote locations, commonly referred to as telepresence, particularly in the context of performing surgical operations, has driven a significant transformation in the use of immersive technologies for remote surgery [14, 2]. These advancements are especially driven by robotics, networking, sensing, visualization, and human–computer interaction, which have progressively shifted the focus from isolated immersive experiences to connected, collaborative, and spatially aware remote environments [22, 23]. Despite these advances, the ability to collaborate remotely using XR remains relatively underexplored, particularly in the context of robotic-assisted surgery. Although a limited number of studies have begun to study multi-user XR approaches in surgical environments, these systems often require additional training and adaptation by

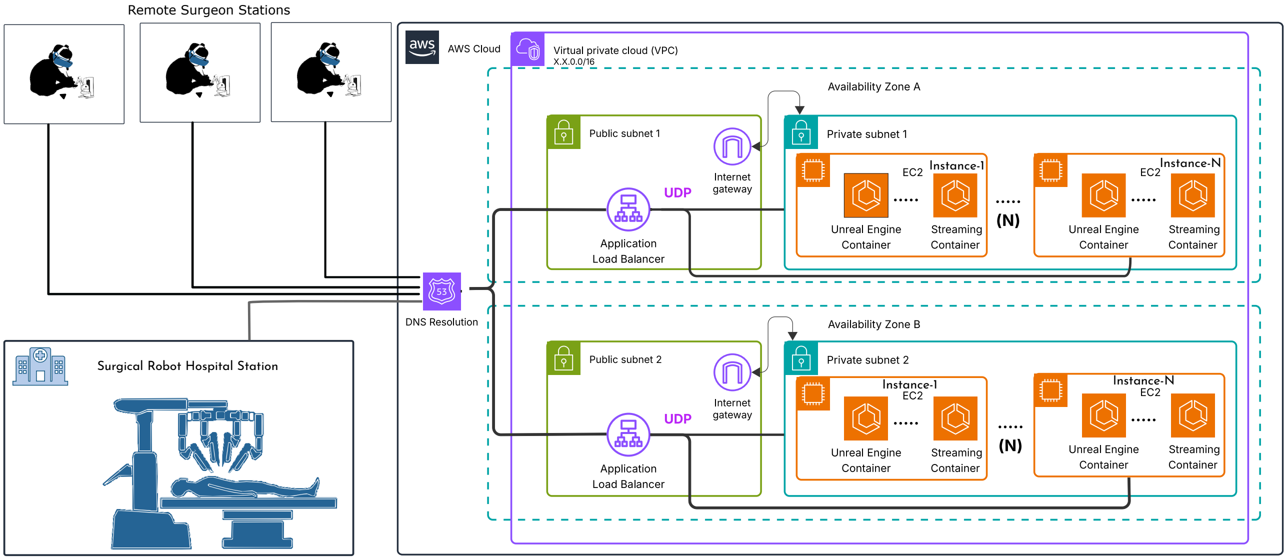

Figure 2: Overall software and hardware framework on AWS infrastructure: Left: remote environment and pilot stations, including teleoperated robots and user-side teleoperation interfaces. Right: AWS EC2–based cloud architecture providing centralized coordination, synchronization, and shared XR state management.

users to surgeons [18, 4, 24].

In this paper, we propose a multi-user remote XR teleoperation framework for robotic surgery using the Da Vinci Research Kit (dVRK). The proposed framework allows multiple surgeons to collaboratively interact with the robotic system using haptic devices, while simultaneously monitoring through real-time video streams. In addition, it provides coordination tools that support shared awareness and interaction around a common, real-time XR representation of the robotic surgical system. The main and supplementary contributions of this work are summarized below:

1. ***Main:*** A multi-user XR-based teleoperation framework for robotic surgery using the da Vinci Research Kit (dVRK), enabling collaborative surgeon interaction through haptic devices and shared real-time XR representations, supported by using Amazon Web Services (AWS) for low-latency communication and real-time state synchronization.

2. ***Supplementary:***

    (a) A preliminary assessment of cloud-based infrastructure for distributed surgical collaboration, identifying performance characteristics, limitations, and scalability considerations.

    (b) Demonstration of key collaborative surgical scenarios across different cities, including real-time team interaction, contextual visualization, remote telementoring, and XR-enabled telepresence.

2 Related Work

The proposed framework and its underlying concepts encompass multiple research domains, including extended reality (XR) technologies, cloud computing, data compression, real-time streaming, and the application of XR in surgical environments. To maintain a focused and relevant scope, this section reviews the most recent related works on the use of XR technologies in the operating room.

Immersive visualization in Robotic Surgery has emerged as a key application in robotic-assisted surgery, aiming to enhance situational awareness, overlay information, provide depth perception, and spatial understanding during complex procedures. Providing surgeons with three-dimensional views of the surgical scene to improve task performance and safety. Early research works in this area focused on stereoscopic visualization and head-mounted displays (HMDs) to improve depth perception and hand–eye coordination during minimally invasive procedures [15, 20, 21, 12, 11]. As XR hardware matured, researchers began integrating real-time endoscopic video, robot kinematics, and anatomical models into immersive environments, allowing surgeons to visualize both the operative field and the underlying robotic system in a unified spatial context [16, 24, 6]. In addition, these approaches have demonstrated the feasibility of overlaying virtual information, such as tool trajectories, safety boundaries, and anatomical landmarks, directly onto the surgeon's view of the surgical scene to guide and avoid errors. In parallel, immersive simulation frameworks that integrate physics-based models with ROS-compatible interfaces have supported user training and algorithm development for surgical robotics [13].

Teleoperation Interfaces in Robotic Surgery, are a foundational component of robotic-assisted surgery, allowing surgeons to control patient-side robotic instruments with precision and stability. Traditional teleoperation in surgical robotics relies on master–slave architectures, where surgeon-side master devices map hand motions to patient-side manipulators. Systems such as the da Vinci Surgical System employ kinematically matched master tool manipulators (MTMs) combined with stereoscopic visualization to provide intuitive control and high dexterity[10, 17, 19, 5, 1]. In addition, hand gesture and hand pose-based teleoperation methods have also been explored as alternatives to traditional control interfaces [3, 25, 9]. Wen et al. [25] proposed a gesture-guided surgical system using Kinect-based recognition to control a surgical robot and interact with an augmented reality interface. Fu et al. [9] introduced an inertial measurement unit (IMU)-based approach that captured wrist motion for robotic instrument control.

Despite extensive research on XR-based visualization and teleoperation, relatively few systems explicitly address collaborative and multi-user interaction in robotic-assisted surgery. Existing approaches predominantly focus on single-user control paradigms and do not support role-based interaction or geographically distributed collaboration. The framework proposed in this work addresses this gap by enabling multi-user, role-based collaboration within a shared XR environment.

3 System Overview

As illustrated in Figures 1 and 2, the proposed system provides a multi-user, remote XR-based telesurgery framework using the dVRK. It provides a high-resolution MR visual interface and a unilateral haptic teleoperation device for surgeons or users. Data from the user's side is networked on an Amazon Web Service with mechanisms for synchronization across users and the patient-side manipulator, helping geographically distributed participants interact with a shared, real-time representation of a robotic surgical system.

This section presents the overall system architecture and explains how the robotic components and their states are modeled, communicated, and synchronized within the proposed framework. It first describes the ***Patient-Side Systems***, including the dual Patient Side Manipulators (PSMs) and their associated kinematic and interaction states, as well as the modeling and control of the Endoscopic Camera Manipulator (ECM). The section then presents the ***User-Side Components***, focusing on the mechanisms used to capture user interactions, manage immersive visual data streams, and maintain a consistent shared representation of the robotic system across geographically distributed users. Finally, it will present the ***Network and Synchronization*** mechanisms, showing how robot state information, user inputs, and visual content are transmitted over the cloud, and how synchronization mechanisms guarantee temporal consistency and coherent system behavior under varying network conditions.

3.1 *Patient-Side Systems*

The physical setup of the da Vinci Research Kit (dVRK) typically consists of three main subsystems: the passive positioning structure, referred to as the Setup Joints (SUJ); the two Patient Side Manipulators (PSMs), which enable bimanual instrument manipulation; and an Endoscopic Camera Manipulator (ECM). The proposed system utilizes the full kinematic model of this structure, and each PSM is modeled as a six-degree-of-freedom serial manipulator, excluding the gripper, providing real-time access to joint states and end-effector pose information. This kinematic representation enables accurate tracking and representation in XR environments, allowing for the creation of replicas and the teleoperation of robotic instruments within the shared XR environment.

To ensure accurate replication of the robotic system and consistent state synchronization across distributed users, the proposed framework adopts a unified coordinate frame convention and notation. The coordinate frames are defined as follows. The frame $\{W\}$ defines the world frame (e.g., operating room reference), $\{S\}$ the SUJ base frame, and $\{B_k\}$ the base frame of subsystem $k \in \{\mathrm{PSM}_L, \mathrm{PSM}_R, \mathrm{ECM}\}$. For each subsystem k, let $\{g_k\}$ denote the tool (end-effector) frame and ${}^{A}\boldsymbol{T}_B \in SE(3)$ is used for the homogeneous transform from frame $\{B\}$ to $\{A\}$.

The Setup Joints (SUJ) provide passive positioning of each patient-side manipulator base, allowing flexible placement of the robotic arms within the operating environment. For each robotic subsystem k, the SUJ configuration is represented by the joint vector

$$\boldsymbol{q}_{\mathrm{suj}}^{(k)} \in \mathbb{R}^{n_{\mathrm{suj}}}, \tag{1}$$

where n_{suj} is the number of passive SUJ joints used to position the corresponding arm. The SUJ defines the transform from the world to the subsystem base frame:

$${}^{W}\boldsymbol{T}_{B_k}\left(\boldsymbol{q}_{\mathrm{suj}}^{(k)}\right) = {}^{W}\boldsymbol{T}_S \, {}^{S}\boldsymbol{T}_{B_k}\left(\boldsymbol{q}_{\mathrm{suj}}^{(k)}\right). \tag{2}$$

Here ${}^{W}\boldsymbol{T}_S$ is a fixed calibration transform, and ${}^{S}\boldsymbol{T}_{B_k}(\cdot)$ is the SUJ forward kinematics. In this paper, ${}^{W}\boldsymbol{T}_{B_k}$ is treated as part of the authoritative state for consistent multi-user visualization.

Each Patient Side Manipulator (PSM), indexed by $j \in L, R$, is modeled as a serial kinematic chain with six actuated degrees of freedom, excluding the gripper. This representation is consistent with standard modeling and identification approaches for the dVRK PSMs [7]. The joint configuration of each PSM is defined as

$$\boldsymbol{q}_{\mathrm{psm}}^{(j)} = \begin{bmatrix}\theta_1 & \theta_2 & d_3 & \theta_4 & \theta_5 & \theta_6\end{bmatrix}^{\top} \in \mathbb{R}^6, \tag{3}$$

where the third joint is prismatic, and the remaining joints are revolute. The PSM tool pose in world coordinates is:

$${}^{W}\boldsymbol{T}_{g_{\mathrm{psm}}^{(j)}} = {}^{W}\boldsymbol{T}_{B_{\mathrm{PSM}_j}}\left(\boldsymbol{q}_{\mathrm{suj}}^{(\mathrm{PSM}_j)}\right) {}^{B_{\mathrm{PSM}_j}}\boldsymbol{T}_{g_{\mathrm{psm}}^{(j)}}\left(\boldsymbol{q}_{\mathrm{psm}}^{(j)}\right). \tag{4}$$

The manipulator Jacobian is defined as

$$\dot{\boldsymbol{x}}_{\mathrm{psm}}^{(j)} = \boldsymbol{J}_{\mathrm{psm}}^{(j)}\left(\boldsymbol{q}_{\mathrm{psm}}^{(j)}\right)\dot{\boldsymbol{q}}_{\mathrm{psm}}^{(j)}, \tag{5}$$

where $\dot{\boldsymbol{x}}_{\mathrm{psm}}^{(j)}$ is the 6D spatial twist of the tool.

The Endoscopic Camera Manipulator controls the viewpoint of the endoscopic camera, and it is modeled as an actuated kinematic chain whose joint configuration is represented by

$$\boldsymbol{q}_{\mathrm{ecm}} \in \mathbb{R}^{n_{\mathrm{ecm}}}. \tag{6}$$

The ECM is modeled with yaw/pitch, insertion, and roll and the world pose of the camera frame $\{g_{\mathrm{ecm}}\}$ is:

$${}^{W}\boldsymbol{T}_{g_{\mathrm{ecm}}} = {}^{W}\boldsymbol{T}_{B_{\mathrm{ECM}}}\left(\boldsymbol{q}_{\mathrm{suj}}^{(\mathrm{ECM})}\right) {}^{B_{\mathrm{ECM}}}\boldsymbol{T}_{g_{\mathrm{ecm}}}(\boldsymbol{q}_{\mathrm{ecm}}). \tag{7}$$

This transform provides the camera extrinsic used to render consistent viewpoints across all XR clients.

3.2 *User-Side Components*

To demonstrate collaborative interaction and emulate a realistic operating room (OR) workflow on the user side, we defined a role-based unilateral teleoperation paradigm involving three remotely connected, distributed participants. The first role is the primary user (surgeon), who is responsible for operating the dual Patient Side Manipulators (PSMs) for bimanual instrument control using two haptic devices. The second role is that of a secondary user (another surgeon), who controls the Endoscopic Camera Manipulator (ECM) using a single haptic device. The third role is a support user (e.g., a nurse), who serves as a coordinator and facilitates interaction.

The primary surgeon operates two PHANTOM Omni devices, one assigned to each Patient Side Manipulator. The master-side Cartesian poses are denoted as

$$\boldsymbol{x}_m^{(L)}(t), \quad \boldsymbol{x}_m^{(R)}(t) \in \mathbb{R}^6, \tag{8}$$

corresponding to the left and right haptic devices, respectively. Each master input is independently mapped to the desired Cartesian pose of the corresponding PSM through a device-to-robot mapping function:

$$\boldsymbol{x}_{\mathrm{cmd}}^{(\mathrm{PSM}_j)}(t) = \Phi^{(\mathrm{PSM}_j)}\left(\boldsymbol{x}_m^{(j)}(t)\right), \quad j \in \{L, R\}. \tag{9}$$

The resulting Cartesian commands are converted into joint-space references via inverse kinematics [8]:

$$\boldsymbol{q}_{\mathrm{psm,cmd}}^{(j)}(t) = \mathrm{IK}^{(\mathrm{PSM}_j)}\left(\boldsymbol{x}_{\mathrm{cmd}}^{(\mathrm{PSM}_j)}(t)\right), \tag{10}$$

and transmitted unilaterally to the robot-side controller. No force or haptic feedback is rendered to the master devices.

A secondary surgeon is responsible for controlling the Endoscopic Camera Manipulator (ECM) using a single PHANTOM

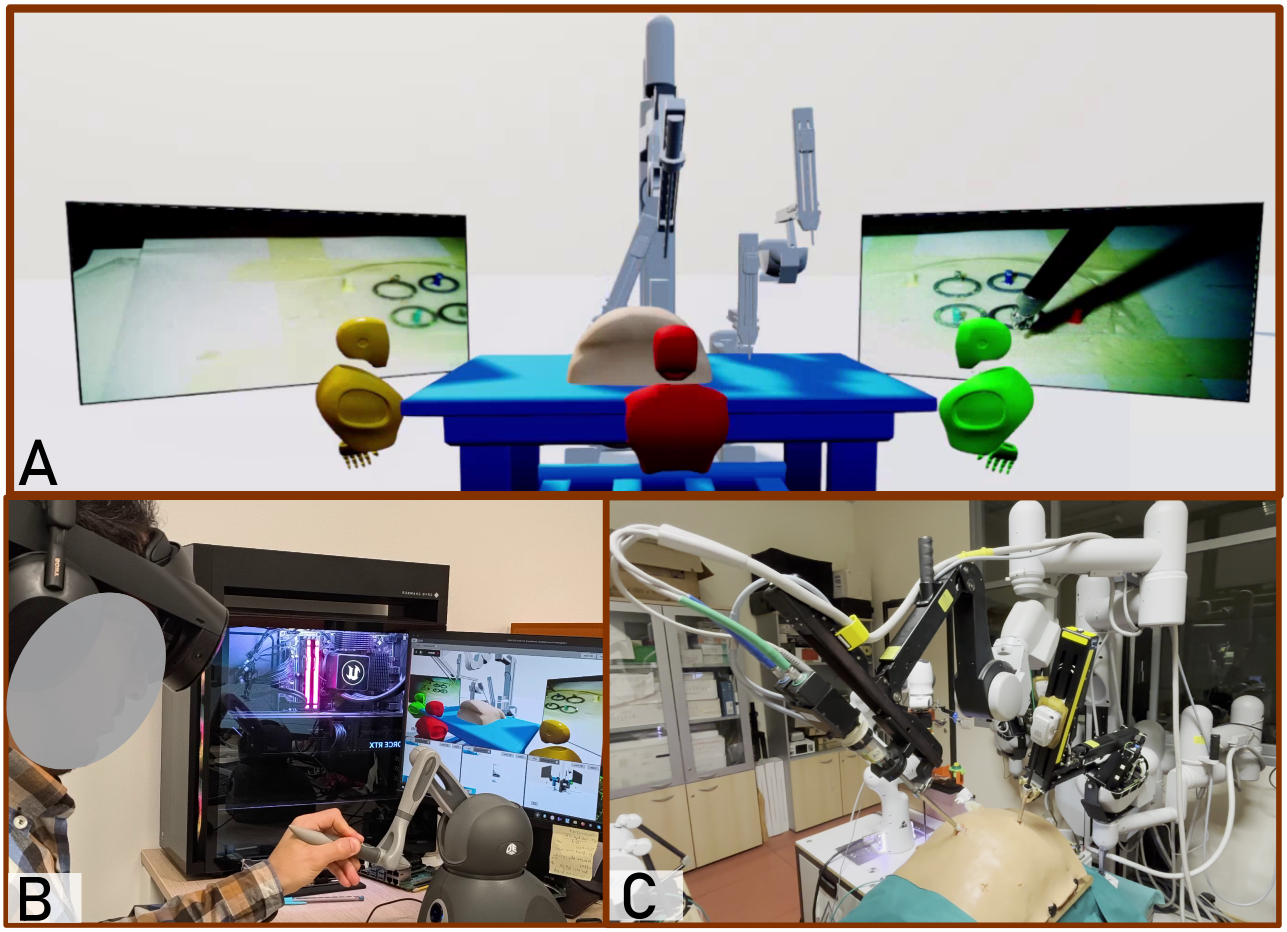

Figure 3: Surgeon-side and patient-side telesurgery setup. (A) Mixed reality (MR) view of the operating room, showing the virtual model of the robotic system and real-time rendering of the endoscopic video streams. (B) Surgeon-side teleoperation setup, where the primary surgeon uses a Phantom Omni haptic device to control the position of the da Vinci Research Kit (dVRK) instruments. (C) Patient-side telesurgery configuration, illustrating the robotic manipulators and experimental setup used during remote operation.

Omni haptic device and does not issue commands to the Patient Side Manipulators, to ensure a clear separation between instrument teleoperation and camera control. The master-side Cartesian input provided by the haptic device is denoted as

$$\boldsymbol{x}_m^{(\mathrm{ECM})}(t) \in \mathbb{R}^6. \tag{11}$$

This input is mapped to a desired endoscopic camera pose through a device-to-camera mapping function:

$$\boldsymbol{x}_{\mathrm{cmd}}^{(\mathrm{ECM})}(t) = \Phi^{(\mathrm{ECM})}\left(\boldsymbol{x}_m^{(\mathrm{ECM})}(t)\right), \tag{12}$$

The resulting Cartesian command is converted into joint-space references using the inverse kinematics provided by the dVRK driver [8]:

$$\boldsymbol{q}_{\mathrm{ecm,cmd}}(t) = \mathrm{IK}^{(\mathrm{ECM})}\left(\boldsymbol{x}_{\mathrm{cmd}}^{(\mathrm{ECM})}(t)\right). \tag{13}$$

The third participant assumes the role of a charge nurse, acting as a liaison between the two surgeons. This role does not involve direct control of any robotic subsystem. Instead, the charge nurse supports the collaborative workflow by facilitating intent communication, coordinating task sequencing, and mediating potential conflicts between operators.

Within the proposed framework, the intention of each participant n is represented as

$$\boldsymbol{i}_n(t) = \left(k_n(t),\, \boldsymbol{x}_{n,\mathrm{ref}}(t),\, \mathscr{A}_n(t)\right), \tag{14}$$

where $k_n(t)$ denotes the relevant subsystem (PSM_L, PSM_R, or ECM), $\boldsymbol{x}_{n,\mathrm{ref}}(t)$ denotes a reference pose, direction, or region of interest, and $\mathscr{A}_n(t)$ represents auxiliary actions such as annotations, warnings, or coordination cues.

3.3 *Network and Synchronization*

To establish the network from the patient side and across all users, Amazon Web Services (AWS) is used as a centralized server using Unreal Engine 5 dedicated server implementation, as illustrated in Figure 2. This cloud server acts as the authoritative coordination layer, managing state distribution, synchronization, and access control among all connected users and enforces the following coordination constraints:

- Control of Patient Side Manipulators (PSMs) and Endoscopic Camera Manipulator (ECM) is explicitly decoupled, preventing interference between instrument teleoperation and camera control.

- The control of the Patient Side Manipulators (PSMs) and the Endoscopic Camera Manipulator (ECM) are decoupled.

- Only the designated operator for a given subsystem is permitted to issue control commands, to ensure clear responsibility and avoid conflicting inputs.

- All participants receive similar robot and camera states, subject to bounded synchronization error introduced by network latency and update rates.

- Intent signals generated by the charge nurse are visible and shared with all participants.

The relationship between the cloud server state and a users (client's) local replica can be represented by the following model:

$$\hat{\boldsymbol{s}}_i(t) = \boldsymbol{s}(t - \tau_i(t)) + \varepsilon_i(t), \tag{15}$$

and the equation captures the core challenges of networked synchronization within the Unreal Engine architecture, where $\boldsymbol{s}(t)$ is an authoritative server state maintained solely by the dedicated server. All critical logic, such as user location, robot locations, and interaction outcomes, is resolved. $\hat{\boldsymbol{s}}_i(t)$ is the user replica state, and it denotes the perceived state of users i at time t. The user's view is always an approximation derived from data received from the server. $\tau_i(t)$ is the network latency/Delay, and it accounts for variable network latency (ping) between the server and users i. Data cannot arrive instantaneously, meaning the client's replica $\hat{\boldsymbol{s}}_i(t)$ inherently represents the server's state as it was a moment ago, $\boldsymbol{s}(t - \tau_i(t))$. $\varepsilon_i(t)$ is the prediction and smoothing error to deliver a smooth user experience despite delays, and the Unreal Engine clients use techniques like client-side prediction and interpolation. This introduces a controlled error $\varepsilon_i(t)$, as the client must extrapolate between received server updates. The engine strives to minimize this error and reconcile the predicted state with the delayed server truth as new data arrives.

4 RESULTS

When all components are integrated, the overall system configuration is shown in Fig. 3.(A) showing the mixed-reality (MR) operating room view, including the virtual robot model and real-time endoscopic video rendering. (B) showing the surgeon-side teleoperation interface, where a Phantom Omni haptic device is used to control the da Vinci Research Kit (dVRK) instruments and (C) shows the patient-side setup, including the robotic manipulators and the experimental environment used for remote operation.

To evaluate the effectiveness of the proposed system, we analyzed improvements in Quality of Service (QoS) in terms of bandwidth utilization and replication latency, as these factors directly impact the overall Quality of Experience (QoE). For evaluation, the system was deployed on Amazon Web Services (AWS) using a t3.medium virtual machine instance hosted in the **Milan, Italy** region, equipped with 2 vCPUs and 2 GB of RAM. Two teleoperators connected remotely from **Verona city** and **Genova city**, approximately 292 km apart, while the patient-side manipulators and a patient mannequin were located in **Verona city**. All patient-side devices were controlled in real time by the operator from **Genova city**. Consequently, time-critical data, including robotic control commands, force feedback, and two video streams, was exchanged continuously between the two locations.

A video recording of the experimental setup and execution is available online [1].

Latency measurements for the robot control loop, corresponding to command signals transmitted from the surgeon-side interface to the remote robotic system and the associated feedback, are shown in Figure 4, and it shows a mean round-trip latency of 38.17 ms, with a minimum of 28 ms and a maximum of 48 ms. The measured latency

[1] https://youtu.be/xi0Yo5iz1xA

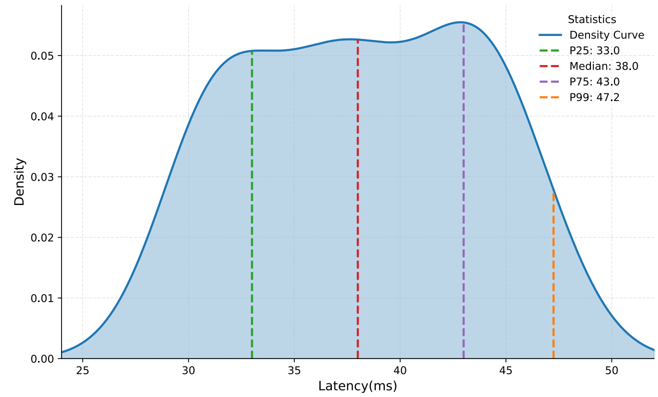

Figure 4: Normalized distribution of round-trip teleoperation latency with percentile markers.

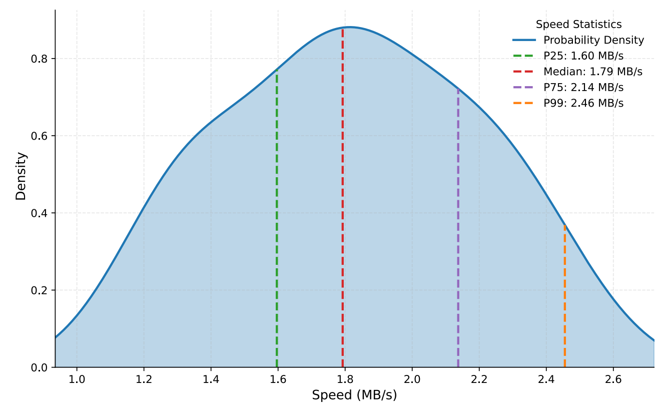

Figure 5: Normalized distribution of per-user bandwidth consumption (MB/s), including all data transmitted to each participant, such as dual video streams and bidirectional teleoperation data.

exhibits a standard deviation of 5.45 ms, indicating limited variability during operation. As seen in the figure, the 25th, 50th (median), and 75th percentiles are observed at approximately 33 ms, 38 ms, and 43 ms, respectively, while the 99th percentile remains below 48 ms. This distribution shows that the majority of control-loop latency values are tightly clustered around the mean, with only a small tail of higher-latency events. Two video streams, each with a resolution of 640×420 pixels, were used for visual data transmission between the surgeon and patient sites. Video compression was performed using H.264 codec at a target bitrate of 20 Mbps per stream to ensure high visual fidelity. Video data were transmitted using the Real-time Transport Protocol (RTP) to maximize throughput and minimize transmission overhead, which is critical for real-time surgical visualization. Under these conditions, the measured end-to-end video transmission latency was approximately 112 ± 13 ms.

Figure 5, illustrates the normalized distribution of per-user bandwidth consumption during the experiment. The reported bandwidth usage includes all data transmitted to each participant, encompassing dual video streams and bidirectional teleoperation data such as robot control commands and force feedback. The distribution shows a median bandwidth consumption of approximately 1.79 MB/s, with the 25th and 75th percentiles at 1.6 MB/s and 2.14 MB/s, respectively. The upper tail of the distribution remains bounded, with the 99th percentile observed at approximately 2.46 MB/s.

Another important performance metric is CPU utilization, which represents the percentage of processing capacity used by the cloud instance during operation. The system was deployed on a virtual machine equipped with two virtual CPUs (vCPUs). During the experimental evaluation, CPU usage remained low and stable, with a minimum utilization of 4.62%, a maximum of 7.70%, and a mean utilization of 6.08%, accompanied by a standard deviation of 0.8%.

5 Conclusion

In this paper, the telesurgery setup and interface were presented, introducing a multi-user mixed-reality (MR) framework for remote robotic-assisted surgery built on the da Vinci Research Kit (dVRK). The proposed system integrates immersive MR visualization, cloud-based synchronization, and role-based teleoperation to enable real-time collaborative interaction among geographically distributed participants. While the proposed mixed-reality–based collaborative telesurgery framework demonstrates promising preliminary performance, several directions remain for future investigation. First, a comprehensive experimental evaluation involving a larger number of participants is required to assess system scalability, multi-user interaction dynamics, and collaborative efficiency under realistic scenarios. In particular, replication latency and bandwidth utilization should be analyzed independently to better characterize their respective impact on system performance and Quality of Experience (QoE). Second, the current implementation relies on unilateral teleoperation. Future work will focus on integrating bilateral teleoperation with haptic feedback, supported by delay-aware and stability-preserving control strategies, to enhance transparency and realism in the presence of network-induced latency. Finally, the mixed-reality environment can be further improved through the integration of real-time three-dimensional scene reconstruction using external 3D sensing technologies. The generation and rendering of remote 3D information are expected to significantly improve situational awareness and depth perception.

References

[1] A. Abdurahiman, S. Zuo, and Z. Wang. Head rotation based articulated laparoscope control. *International Journal of Medical Robotics and Computer Assisted Surgery*, 2021. 2

[2] D. G. Caldwell, A. Acemoglu, and L. S. Mattos. 5g telesurgery in phonomicrosurgery. In *Handbook of Robotic and Image-Guided Surgery*, pp. 931–939. Elsevier, 2026. 1

[3] X. Chen, J. Huang, and Y. Wang. Hand tracking based surgical robot teleoperation using hololens 2. *IEEE Robotics and Automation Letters*, 2020. 2

[4] S.-Y. Chiou, T.-M. Liu, and H.-L. Liu. Multi-user surgical navigation platform based on mixed reality. *IEEE Access*, 2024. 2

[5] S. Dardona, A. Pietrabissa, and E. De Momi. Head-motion-based endoscopic camera control for robotic surgery. *IEEE Transactions on Medical Robotics*, 2020. 2

[6] S. Das and S. Vyas. The utilization of ar/vr in robotic surgery: A study. In *Proceedings of the 4th International Conference on Information Management & Machine Intelligence*, pp. 1–8, 2022. 2

[7] G. A. Fontanelli, F. Ficuciello, L. Villani, and B. Siciliano. Modelling and identification of the da vinci research kit robotic arms. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1464–1469. IEEE, 2017. 3

[8] G. A. Fontanelli, M. Selvaggio, M. Ferro, F. Ficuciello, M. Vendittelli, and B. Siciliano. A v-rep simulator for the da vinci research kit robotic platform. In *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*, pp. 1056–1061. IEEE, 2018. 3, 4

[9] Y. Fu, Z. Li, and H. Xie. Wearable imu-based teleoperation for surgical robotics. *IEEE Transactions on Human-Machine Systems*, 2018. 2

[10] N. Hong, M. Kim, C. Lee, and S. Kim. Head-mounted interface for intuitive vision control and continuous surgical operation in a surgical robot system. *Medical & Biological Engineering & Computing*, 57(3):601–614, 2019. 2

[11] A. Iacoangeli, S. Capelli, A. Held, M. Barba, A. Marasi, A. De Benedictis, M. Iacoangeli, D. Locatelli, M. M. Fontanella, C. G. Giussani, et al. Pre-surgical planning with extended reality in neurosurgery: a survey-based study describing a preliminary experience in italy. *Journal of neurosurgical sciences*, 69(3):236–244, 2025. 1, 2

[12] M. Lorenz, M. Kaluschke, A. Melzer, N. Pillen, M. Sanrow, A. Hoffmann, D. Schmidt, A. Dettmann, A. C. Bullinger, J. Perret, et al. Hips–a surgical virtual reality training system for total hip arthroplasty (tha) with realistic force feedback. *IEEE Transactions on Visualization and Computer Graphics*, 2025. 1, 2

[13] D. Nestel and S. Stevens. Surgical simulation educational research. *Surgery*, 180:109035, 2025. doi: 10.1016/j.surg.2024.109035 2

[14] C. C. Nguyen, S. Wong, M. T. Thai, T. T. Hoang, P. T. Phan, J. Davies, L. Wu, D. Tsai, H.-P. Phan, N. H. Lovell, et al. Advanced user interfaces for teleoperated surgical robotic systems. *Advanced Sensor Research*, 2(4):2200036, 2023. 1

[15] V. Penza, A. Neri, M. Koskinopoulou, E. Turco, M. Azam, and L. S. Mattos. Augmented reality navigation in robot-assisted surgery. In *Handbook of Robotic and Image-Guided Surgery*, pp. 863–877. Elsevier, 2026. 2

[16] V. Penza, A. Neri, M. Koskinopoulou, E. Turco, D. Soriero, S. Scabini, D. Prattichizzo, and L. S. Mattos. Augmented reality navigation in robot-assisted surgery with a teleoperated robotic endoscope. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4621–4626. IEEE, 2023. 2

[17] L. Qian, A. Barthel, and A. Johnson. Augmented reality based flexible endoscope control using head tracking. *IEEE Robotics and Automation Letters*, 2018. 2

[18] L. Qian, A. Deguet, and P. Kazanzides. dVRK-XR: Mixed Reality Extension for da Vinci Research Kit. In *Hamlyn Symposium on Medical Robotics*, 2019. 2

[19] L. Qian, A. Johnson, and A. Barthel. Head-tracked augmented reality visualization for robotic endoscopy. *IEEE Transactions on Visualization and Computer Graphics*, 2019. 2

[20] W. Rhee, Y. G. Kim, J. H. Lee, J. W. Shim, B. S. Kim, D. Yoon, M. Cho, and S. Kim. Unconstrained lightweight control interface for robot-assisted minimally invasive surgery using mediapipe framework and head-mounted display. *Virtual Reality*, 28(2):114, 2024. 2

[21] C. Srakoopun, S. Suebnukarn, P. Haddawy, M. Kaluschke, R. Weller, M. S. Yin, P. Aguilar, P. Phumpatrakom, K. Pinchamnankool, K. Budsaba, et al. Virtual reality simulation for learning minimally invasive endodontics: a randomized controlled trial. *BMC medical education*, 25(1):1310, 2025. 1, 2

[22] Y. Tefera, I. Sarakoglou, S. Deore, Y. Kim, V. Barasuol, M. Villa, S. Anastasi, D. Caldwell, N. Tsagarakis, C. Semini, et al. Robot teleoperativo: Collaborative cybernetic systems for immersive remote teleoperation. In *2024 IEEE Conference on Telepresence*, pp. 1–4. IEEE, 2024. 1

[23] Y. T. Tefera, Y. Kim, S. Anastasi, P. Fiorini, D. G. Caldwell, and N. Deshpande. Immersive remote telerobotics: foveated unicasting and remote visualization for intuitive interaction. *Robotica*, 42(12):4223–4248, 2024. 1

[24] E. Toni, E. Toni, M. Fereidooni, and H. Ayatollahi. Acceptance and use of extended reality in surgical training: an umbrella review. *Systematic Reviews*, 13(1):299, 2024. 2

[25] R. Wen, Y. Wang, and J. Liu. Hand gesture guided surgical robot system. *IEEE Access*, 2017. 2