Acquisition and Remote Transfer of Operative Field View During Open Surgery

Malek Anabtawi
Department of Surgery
Hamad Medical Corporation
Doha, Qatar
maleka@alumni.cmu.edu

Jhasketan Padhan
Department of Surgery
Hamad Medical Corporation
Doha, Qatar
JPadhan@hamad.qa

Abdulla Al-Ansari
Department of Surgery
Hamad Medical Corporation
Doha, Qatar
AALANSARI1@hamad.qa

Zhigang Deng
Department of Computer Science
University of Houston
Houston, Texas, USA
zdeng4@central.uh.edu

Elias Yaacoub
Department of Computer Science
Qatar University
Doha, Qatar
Elias@qu.edu.qa

Amr Mohammed
Department of Computer Science
Qatar University
Doha, Qatar
amrm@qu.edu.qa

Nikhil V. Navkar
Department of Surgery
Hamad Medical Corporation
Doha, Qatar
nnavkar@hamad.qa

***Abstract*—Telementoring in open surgery has emerged as an alternative to on-site teaching. To ensure effective communication between the mentee and the mentor, the view of surgical operations in the operating room needs to be shared remotely in real-time. This study is carried out to propose a multi-threaded system that takes input from multi-color/depth cameras, aligns and creates a single point cloud representing the operative field in three-dimension, transfers it over the network, and reconstructs it for visualization for mentors. The system's performance is measured in terms of execution time of each threaded unit, and latency for transfer of multi-color/depth data over the network. The achieved registration accuracy is bench-marked for multiple cameras. The results show the potential usage of the proposed system for near real-time telementoring during open surgeries.**

Keywords—surgical telementoring, open surgery, operative field, 3D visualization, point-cloud registration

I. Introduction

Telementoring involves utilizing telecommunication technology to facilitate guidance from an expert (known as a mentor) located remotely to a less experienced learner (known as a mentee) [1]. In situations where open surgery is being performed, a surgical trainee must acquire proficiency in the intricate movements of surgical instruments while operating on tissue. Ideally, an expert surgeon is physically present in the operating room to provide guidance. However, due to logistical constraints such as scheduling conflicts, travel limitations, or legal barriers, it may not always be feasible to have the expert on-site. In such cases, a telementoring session can be conducted, where the remote expert imparts surgical skills to the trainee [2]. To ensure effectiveness, the telecommunication technology employed in telementoring needs to be customized to suit the specific domain.

In order for a practical telementoring application in open surgery to be successful, the mentor must have access to accurate and comprehensive information about the operative field. This information must include details such as the location of incisions, the specific tissue being operated on, and a clear view of the mentee's hands holding the surgical instruments. While guiding the mentee using only voice instructions is a possibility, it can be unnecessarily complex without any visual cues. An alternative approach is to utilize annotations overlaid onto a screen that displays a live video feed of the operative field. This involves transferring the video captured by a camera in the operating room to a remote site where the mentor is located. However, relying solely on a single camera in the operating room may not be sufficient, as the mentee's hands can potentially obstruct the mentor's view.

Several studies have showcased various approaches for sharing the mentee's working space with the mentor, highlighting the benefits it brings to the telementoring process. In one instance, Mitsuno et al. [3] used Skype on the mentee's HoloLens to share the surgical field with the mentor. This method, facilitated by the wearable HoloLens device, offered a valuable "first-person" view of the field, albeit in a two-dimensional format, allowing the mentor to closely experience what the mentee sees. Similarly, in another study conducted by Munoz et al. [4], a video stream of the operating room was shared with the mentor. In this case, a top-down camera positioned above the operating site captured the visuals and transmitted them to the mentor. Dewitz et al. [5] explored the use of the HoloLens, but with a different approach. In their study, the HoloLens was utilized to transmit both color and depth information to the mentor, enabling the construction of a live-point cloud representation of the mentee's space in real-time on the mentor's computer. Gasques et al. [6] made an assumption in their study that both the mentor and the mentee shared a similar operational setup. This assumption had a significant advantage as it eliminated the need for transmitting information about the room itself. Instead, the focus was directed towards conveying an accurate point-cloud representation of the actual patient.

The contribution of this study can be summarized as follows. Firstly, our application introduces the utilization of multiple point clouds to offer a more comprehensive representation of the mentee's work space. To the best of our knowledge, this approach has not been extensively explored in the context of telemedicine. Secondly, we provide a detailed analysis of the architecture deployed.

2023 IEEE 5th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS) | 979-8-3503-2097-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ECBIOS57802.2023.10218549

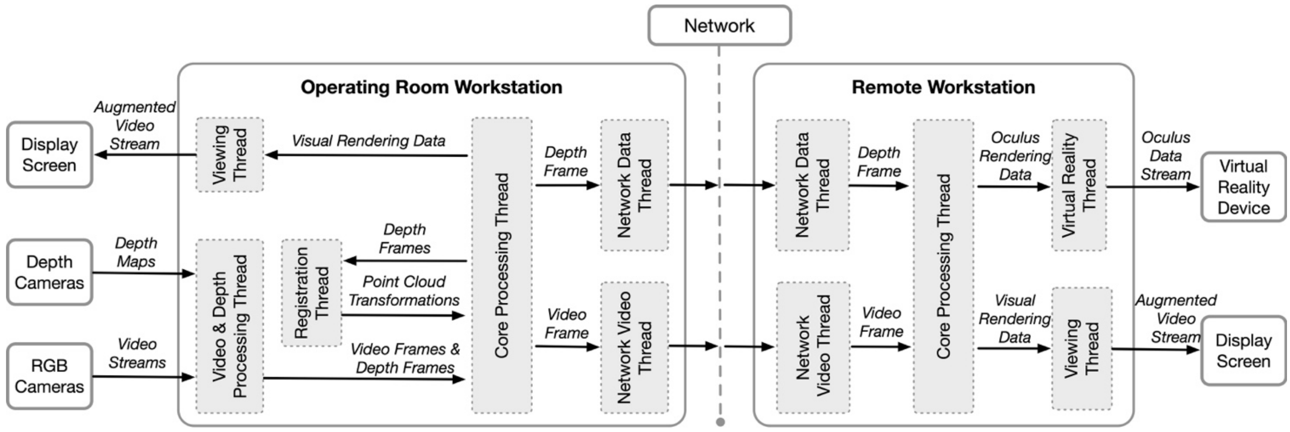

Fig. 1. Architecture of the proposed system. The operating room and the remote workstation are connected over network. The processing on the workstations is performed by parallel running threads, communication with each other, and the hardware units.

II. Materials and Methods

A. Visualizing Operating Field

The system architecture, as depicted in Fig. 1, is based on the telementoring framework proposed by Shabir et al. [7– 9]. It consists of two main components: the operating room workstation and the remote workstation, which are initiated independently by the mentor and mentee, respectively. On the mentor's side, multiple Azure Kinect cameras, up to three in this case, are connected to the system. These cameras serve the purpose of capturing color information and depth maps of the operating room. To avoid interference, each camera is positioned to capture a unique angle of the room while avoiding infrared (IR) interference [10].

After reading the color and depth information from each connected Azure Kinect camera, we construct individual point clouds, denoted as $P_1, P_2, ..., P_n$, where n represents the number of cameras, and Pi represents the point cloud generated from the color and depth data of the i^{th} camera. The resulting points are rendered in a separate window, providing the expert with the flexibility to view the space according to their preferences. However, directly rendering all the point clouds together without any adjustments does not yield a coherent representation. This is because each camera has a unique position and orientation. The point clouds can be combined by a set of transformations $T_1, T_2, ..., T_n$, where for each i, transforming P_i by T_i aligns all the different views as closely as possible. In order to find the above transformation, the mentor initiates a one-time *registration process*, which runs as follows.

- Capturing one frame of color/depth information for each camera
- Computing an initial set of transformations $T_1 , T_2 , ..., T_n$ that roughly align the cameras together. We implemented two ways of doing this: (1) Using a ChArUco board to estimate the board's pose relative to each camera, and (2) Using Optitrack's Motive to detect the pose of the rigid body attached to each camera.
- Refining the above transformations using multiple stages of the ICP algorithm [11]

Once the registration process is completed, and after the mentor establishes a connection with the mentee's workstation, the remote workstation starts receiving multiple 2D camera views. These views can be cycled through by the user, providing different perspectives of the operating room. Additionally, the remote workstation receives the corresponding depth maps, which are transformed using the provided list of transformations. The complete point cloud, incorporating the aligned and transformed data, is then transmitted to a Unity application running on a connected Oculus Quest 2 device. The Unity application is specifically designed to render the received point cloud, enabling the mentor to observe the operating room in a three-dimensional manner using the virtual reality headset.

B. Networking

The main challenge encountered during the system development revolved around efficiently transmitting a large number of point clouds over the internet for an extended period. At a depth resolution of 512×512, with two bytes per depth "pixel", gives us a lower bound of $\sim$500Kb per point cloud. Assuming a frame rate of 30 FPS is maintained [12], it requires sending $\sim$15Mb per camera per second without the video stream. For a single camera, this is expensive but manageable. For two or three cameras, a more sophisticated method was required. We attempted to solve this problem by compressing all depth frames using a delta compression algorithm [13] and designed a protocol to ensure that both workstations have the same copy of the reference used by the compression algorithm (each camera has a different reference). This approach, when combined with other simple optimizations and ensured that the reference buffer used per camera is continually updated, resulted in compression rates of up to 95%.

III. Results and Discussion

We measured the performance of our system along three fronts: execution time, network delay, and alignment fitness. Because the number of connected Azure Kinects greatly affects the system's performance, we repeated our experiments with n = 1, 2, 3 connected cameras. We briefly summarize our procedure for each experiment.

TABLE I. PROCESSING PERFORMED ON OPERATING ROOM AND REMOTE WORKSTATIONS

Thread	Processing performed by the thread
Viewing	Takes all the color and depth information and renders *(i)* the 2D video stream of the selected camera, and *(ii)* the combined point cloud after applying the computed transformations
Registration	Implements the registration process and caches the resulting point cloud transformations
Video & Depth Processing	Reads the video/depth streams from the cameras.
Core Processing	Calls the other components in the workstation and routes the data fed from one component to the next.
Network Data	Manages transfer of the point clouds between this workstation and the remote one. This entails compression, updating the available reference (for delta compression), and sending the compressed point clouds over the internet, using TCP.
Network Video	Compresses and sends all available video streams over he internet, using WebRTC.
Virtual Reality	Renders video and depth stream in a virtual reality environment.

- The execution time of several key components described in Table I was recorded over a period of five minutes each. This was repeated three times per component.
- For network latency, both workstations were located in Doha, Qatar, and connected to separate routers. Both workstations were synchronized to the same clock. The sent and received timestamps were recorded and analyzed upon completing the experiment.
- We performed the registration process ten times, each time noting the final fitness score produced. This score is simply the Root Mean Squared Error (RMSE), defined as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_1(i) - P_2(i))^2} \qquad (1)$$

where n is the number of points in both point clouds (assumed to be the same), P_1, P_2 are the transformed point clouds, and $P_1(i)$ is the *(x, y, z)* and *(r, g, b)* values of the i^{th} point.

Table II provides a comprehensive overview of the results obtained in our study. Additionally, Fig. 2 illustrates the experimental setup and the achieved alignment. It is worth noting that the number of connected point clouds significantly impacts the performance of various components within the system. As the number of point clouds increases, there is a noticeable increase in execution time and latency. Even with just one camera, the viewing thread necessitated approximately 200 ms to render a single- point cloud. This translates to an approximate frame rate of 5 FPS, which falls below the recommended frame rate [12].

TABLE II. SYSTEM PERFORMANCE

Recorded Parameters of the System		RGBD Cameras		
		n = 1	n = 2	n = 3
Execution Time	Video and Depth Processing	33.4 ms	33.4 ms	34 ms
	Viewing Thread	217.8 ms	467.2 ms	727.7 ms
	Network Video & Data Threads	60.25 ms	129.75 ms	194.5 ms
	Registration Thread	N/A	~ 2s	~ 5s
Latency	Depth Frames	77 ms	335 ms	815 ms
	Video Frames	72 ms	162 ms	284 ms
Fitness Score	Motive Registration	N/A	0.21	0.30*
	ChArUco Registration	N/A	0.25	0.31*

[(*) Refers to the average registration score of both pairs.

The existing implementation of our system follows a relatively simplistic structure, wherein the creation, transformation, and rendering of the point cloud(s) take place on a single thread. This approach is rudimentary and imposes limitations on the system's overall performance. In future study, we intend to improve the system architecture by further dividing its components into separate threads, allowing for better utilization of parallel processing capabilities [14]. Furthermore, our current implementation relies solely on the CPU for processing these tasks. In future iterations, we plan to leverage the power of the GPU (Graphics Processing Unit) for rendering the point cloud.

In the following study, we aim to integrate preoperative image data [15,16] along with guidance contours derived from the image data [17,18] into the point cloud. This integration holds particular significance for dynamic environments such as the beating heart, as it enables the creation of a comprehensive visual representation for both the mentor and mentee, providing them with valuable insights. To accomplish this, conducting clinical studies involving end users [19–22] is essential. These studies help us gain a deeper understanding of how users perceive and interact with the visual environment that is generated.

IV. CONCLUSION

The results obtained in our study showcase the practicality of transferring multiple point clouds over the internet, serving the purpose of telemedicine effectively. While our current application may not be optimized for real-time usage when utilizing three cameras, the concept's feasibility has been established. These findings demonstrate the potential and viability of leveraging this approach for telemedicine applications.

ACKNOWLEDGMENT

This study was supported by National Priority Research Program (NPRP) award (NPRP12S-0119-190006) from the Qatar National Research Fund (a member of The Qatar Foundation). All opinions, findings, conclusions, or recommendations expressed in this work are those of the authors and do not necessarily reflect the views of our sponsors.

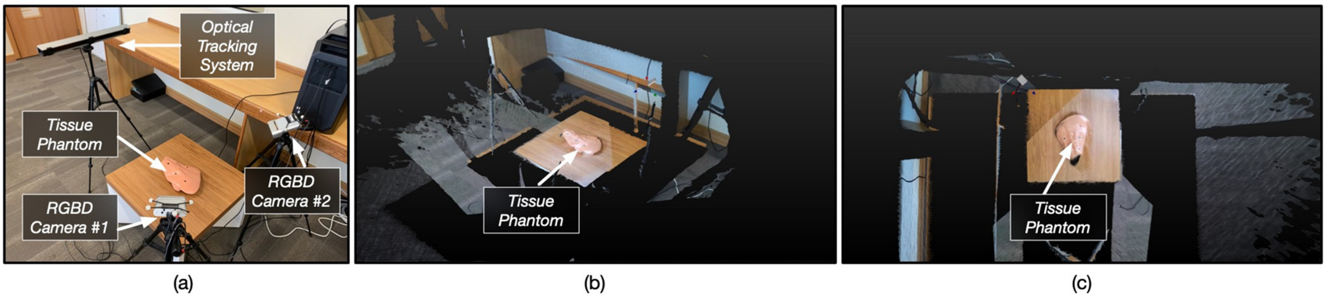

Fig. 2. (a) Experimental setup mimicking a tissue phantom at the operating room. Two RGBD Kinect cameras are connected to operating room workstation and pointing towards the phantom are placed. An OptiTrack optical tracking system captures the poses' of the rigid bodies on the RGBD Kinect cameras. (b) Side view of the tissue phantom after alignment. (c) Top-down view of the tissue phantom after alignment.

REFERENCES

[1] C. M. Schlachta, N. T. Nguyen, T. Ponsky, and B. Dunkin, "Project 6 summit: Sages telementoring initiative," Surgical endoscopy, vol. 30, pp. 3665–3672, 2016.

[2] S. Erridge, D. K. Yeung, H. R. Patel, and S. Purkayastha, "Telementoring of surgeons: a systematic review," Surgical innovation, vol. 26, no. 1, pp. 95–111, 2019.

[3] D. Mitsuno, Y. Hirota, J. Akamatsu, H. Kino, T. Okamoto, and K. Ueda, "Telementoring demonstration in craniofacial surgery with hololens, skype, and three-layer facial models," Journal of Craniofacial Surgery, vol. 30, no. 1, pp. 28–32, 2019.

[4] E. Rojas-Mun˜oz, M. E. Cabrera, C. Lin, D. Andersen, V. Popescu, K. Anderson, B. L. Zarzaur, B. Mullis, and J. P. Wachs, "The system for telementoring with augmented reality (star): A head-mounted display to improve surgical coaching and confidence in remote areas," Surgery, vol. 167, no. 4, pp. 724–731, 2020.

[5] B. Dewitz, R. Bibo, S. Kalkhoff, S. Moazemi, A. Liebrecht, C. Geiger, F. Steinicke, H. Aubin, and F. Schmid, "Towards 5g telementoring in vr-assisted heart transplantation using hololens 2," in GI VR/AR Workshop 2022. Gesellschaft fu¨r Informatik eV, 2022.

[6] D. Gasques, J. G. Johnson, T. Sharkey, Y. Feng, R. Wang, Z. R. Xu, E. Zavala, Y. Zhang, W. Xie, X. Zhang et al., "Artemis: A collaborative mixed-reality system for immersive surgical telementoring," in Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 2021, pp. 1–14.

[7] D. Shabir, N. Abdurahiman, J. Padhan, M. Trinh, S. Balakrishnan, M. Kurer, O. Ali, A. Al-Ansari, E. Yaacoub, Z. Deng et al., "Towards development of a tele-mentoring framework for minimally invasive surgeries," The International Journal of Medical Robotics and Computer Assisted Surgery, vol. 17, no. 5, p. e2305, 2021.

[8] D. Shabir, N. Abdurahiman, J. Padhan, M. Anbatawi, M. Trinh, S. Balakrishnan, A. Al-Ansari, E. Yaacoub, Z. Deng, A. Erbad et al., "Preliminary design and evaluation of a remote tele-mentoring system for minimally invasive surgery," Surgical endoscopy, vol. 36, no. 5, pp. 3663–3674, 2022.

[9] D. Shabir, M. Anbatawi, J. Padhan, S. Balakrishnan, A. Al-Ansari, J. Abinahed, P. Tsiamyrtzis, E. Yaacoub, A. Mohammed, Z. Deng et al., "Evaluation of user-interfaces for controlling movements of virtual minimally invasive surgical instruments," The International Journal of Medical Robotics and Computer Assisted Surgery, vol. 18, no. 5, p. e2414, 2022.

[10] "Synchronize multiple azure kinect dk devices," https://learn.microsoft.com/en-us/azure/kinect-dk/multi-camera-sync.

[11] Q.-Y. Zhou, J. Park, and V. Koltun, "Open3D: A modern library for 3D data processing," arXiv:1801.09847, 2018.

[12] E. M. Bogen, C. M. Schlachta, and T. Ponsky, "White paper: technology for surgical telementoring—sages project 6 technology working group," Surgical Endoscopy, vol. 33, pp. 684–690, 2019.

[13] D. Trendafilov, N. Memon, and T. Suel, "zdelta: An efficient delta compression tool," Technical Report TR-CIS-2002-02, Polytechnic University, Tech. Rep., 2002.

[14] E. Yeniaras, N. Navkar, M. A. Syed, and N. V. Tsekos, "A computational system for performing robot-assisted cardiac surgeries with mri guidance," in 2010 Society for Design and Process Science (SDPS), 2010.

[15] X. Gao, N. V. Navkar, D. J. Shah, N. V. Tsekos, and Z. Deng, "Intraoperative registration of preoperative 4d cardiac anatomy with real-time mr images," in 2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE), 2012, pp. 583–588.

[16] E. Yeniaras, N. V. Navkar, A. E. Sonmez, D. J. Shah, Z. Deng, and N. V. Tsekos, "MR-based real time path planning for cardiac operations with transapical access," Med Image Comput Comput Assist Interv, vol. 14, no. Pt 1, pp. 25–32, 2011.

[17] N. Navkar, E. Yeniaras, D. Shah, N. Tsekos, and Z. Deng, "Generation of 4d access corridors from real-time multislice mri for guiding transapical aortic valvuloplasties," in Medical Image Computing and Computer Assisted Intervention, 2011, pp. 249 – 257.

[18] E. Yeniaras, J. Lamaury, N. V. Navkar, D. J. Shah, K. Chin, Z. Deng, and N. V. Tsekos, "Magnetic resonance based control of a robotic manipulator for interventions in the beating heart," in 2011 IEEE International Conference on Robotics and Automation, 2011, pp. 6270–6275.

[19] J. D. Velazco-Garcia, N. V. Navkar, S. Balakrishnan, J. Abi-Nahed, K. Al-Rumaihi, A. Darweesh et al., "End-user evaluation of software-generated intervention planning environment for transrectal magnetic resonance-guided prostate biopsies," Int J Med Robot, vol. 17, no. 1, pp. 1–12, Feb 2021.

[20] C. A. Velasquez, N. V. Navkar, A. Alsaied, S. Balakrishnan, J. Abinahed, A. A. Al-Ansari, and W. Jong Yoon, "Preliminary design of an actuated imaging probe for generation of additional visual cues in a robotic surgery," Surg Endosc, vol. 30, no. 6, pp. 2641–2648, Jun 2016.

[21] J. D. Velazco-Garcia, N. V. Navkar, S. Balakrishnan, J. Abinahed, A. Al- Ansari, G. Younes, A. Darweesh et al., "Preliminary evaluation of robotic transrectal biopsy system on an interventional planning software," in 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), 2019, pp. 357–362.

[22] J. D. Velazco-Garcia, N. V. Navkar, S. Balakrishnan, J. Abinahed, A. Al- Ansari, A. Darweesh et al., "Evaluation of interventional planning soft- ware features for mr-guided transrectal prostate biopsies," in 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE), 2020, pp. 951–954