Training, Education, and Learning Sciences, Vol. 155, 2024, 174–184
https://doi.org/10.54941/ahfe1005402

Multimodal Extended Reality for Laparoscopic Surgery Training

Yang Cai

University of California San Diego, La Jolla, CA, USA

ABSTRACT

In this study, we explore a multimodal extended reality system for laparoscopic surgery training. The system contains multimodal feedback and holographic overlays. The haptic organs are integrated into the simulator to fill the gap of the mixed reality interfaces for realistic training needs. The holographic overlay synchronizes with the simulated or actual tissue. The 3D objects from the CT DICOM data are overlaid to the live simulated tissues in the cavity. The 3D object registration can be controlled by hand gestures. The machine vision algorithms are designed to enable the dynamic overlay process on the live laparoscopic surgery video. For example, we overlay the symbolic Calot's Triangle based on the Visual or Near-Infrared video data. Machine vision also includes virtual reality to capture, model, and render 3D objects from 2D or 3D image and video sources, including the live 3D scope camera, CT, MRI, NIR images, and tissue scanning. The photorealistic virtualized reality emphasizes that the image data looks natural rather than the synthetic imagery used in virtual reality. The 3D reconstruction results show that the 2D laparoscopic camera can achieve reasonable accuracy in measured distances. Our experiments indicate that multimodal extended reality can increase the fidelity of the laparoscopic surgery simulation and potentially improve training efficiency.

Keywords: Stereo, 3D, Field of view, Augmented reality, Virtual reality, Extended reality, Usability, Interaction, HCI, Human-computer interaction

INTRODUCTION

Learning is our instinctual behavior for surviving and understanding. A learning process is multimodal. We use sensation, vision, hearing, movement, balance, and cognition in the learning process. Learning also involves tools that assist learning, such as books, videos, demonstrations, and simulators.

Laparoscopic surgeries are minimally invasive surgeries (MIS) that became increasingly common in recent decades. They reduce infections, scars, recovery time, and costs. However, it requires sophisticated surgical skills to operate through a few trocar ports on the skin. It is extremely challenging for novice surgeons because of impairments in spatial and haptic perceptions, in the ability to perceive depth, to sense the difference in tissues, to develop mental models of the anatomical structures in hand-eye coordination, and to make decisions in response to adverse situations (Lin and Chen, 2013; Matern et al., 2005). Figure 1 shows the typical laparoscopic surgery

environment and the close-up view of the operation area. From the illustrations, we can see the surgeons have to coordinate the hand and eyes as well as multiple instruments.

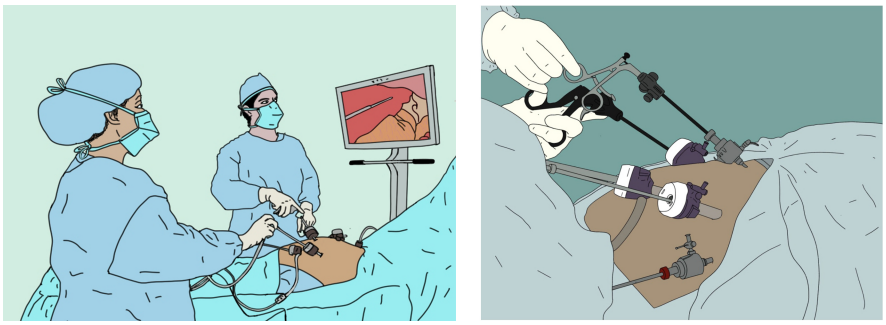

Figure 1: Typical laparoscopic surgery (left) and a close-up of the instruments (right).

Virtual Reality (VR) simulation is a prevailing approach for training, for example, the Fundamental Laparoscopic Surgery (FLS) training boxes with the pegs and beads for hand-eye coordination training and simulated tissues for cutting and stitching. Augmented Reality (AR) enables the virtual organs or events to be overlaid on top of the real-world scene, for example, overlaying the recorded 3D CT object to the mannequin for training or pre-surgery planning. Furthermore, Extended Reality (XR) superimposes real-time sensory data to the physical scene, for example, the near-infrared image of the veins projected onto the patient's skin. Extended Reality is where reality and simulation are so seamlessly blended together that the borderline between the two worlds is blurred. Figure 2 shows the spectrum of the reality technologies from VR, AR, to XR. From the left to right, we can see the evolution of the technologies. The more toward the right, the more involved real-time environment and object sensing, fusing, and overlay processes.

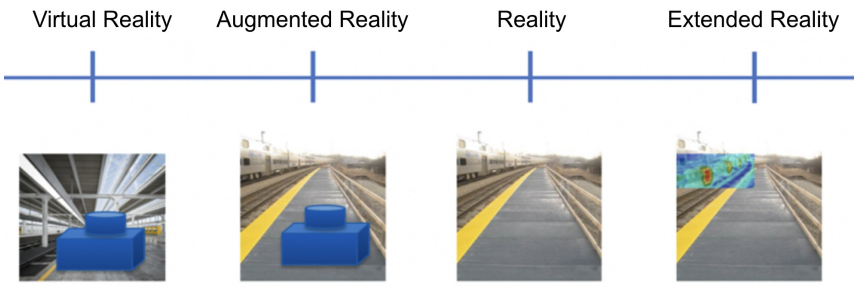

Figure 2: Spectrum of reality technologies in representing a railway station: overlaying a virtual object to the virtual background (VR), overlaying a virtual object to a physical background (AR) and overlaying a live thermal image onto the physical background (XR).

Most current VR and AR training systems for laparoscopic surgery are focused on individual skills such as precision cutting, and individual organs

such as gallbladder. They normally do not provide real-time feedback. Overall, the simulation is distant to the real-work laparoscopic surgeries, especially, the training for laparoscope entry, anomalous anatomic structures, and disastrous scenarios.

In this study, we explore Multimodal Extended Reality (XR) for laparoscopic surgery training. The new system includes haptic perception simulation with physical simulated digestive organs in the cavity, spatial audio that reflects 3D soundscape in the OR environment with head tracking, the XR simulator that combines the physical model with virtual surroundings instruments and surgical assistants, and the real-time machine vision overlay on the live laparoscopic surgery monitor.

HAPTIC SIMULATION

Haptic perception is a critical compotent in laparoscopic surgery because the surgeon's direct vision is obscured. Haptic feedback can be extremely important for the surgeon to evaluate the nature of tissue – for example, to find a tumor and its boundaries prior to excision. In laparoscopy, it is hard using the haptic feedback of current laparoscopy instruments (Eskel et al., 2011). Besides, laparoscopic surgery creates discordance between the visual and haptic systems. This causes incorrect sequencing of psychomotor output that requires a significant period of compensatory change – for example, the perceived inversion of movement of the tip of the laparoscopic actuator (Crothers et al., 1999). We want to simulate these haptic and visual effects with Extended Reality interfaces.

We have developed realistic artificial tissues and organs from available human CT data. We started with objectively measure the hardness of tissues or organs with a durometer for the Shore Unit (SU). We then produce the soft tissues and organs based on the reference SU values. We arrange the organs and tissues according to CT data and general anatomic structures in a 3D printed digestive cavity, which is also modeled from the CT data in a realistic shape and size (Cai and Perez, 2024). For example, the bottom of regular Fundamental Laparoscopic Surgery (FLS) training boxes are flat, but ours is curvy in form of a spinal cord shape. The top skin is designed to be soft that can be cut for inserting the laparoscopic instruments. This expands the FLS training from basic elements to advanced ones such as port location selection based on surgery type.

We use the simulated laparoscopic surgery instruments with the haptic organ cavity. For example, we connect the laparoscopic camera to the laptop and intert graspers to simulate the live surgical procedure. Figure 3 shows the prototype of the hapic simulation with the monitor. This simulation system alone can be used as a training platform (Cai and Perez, 2024).

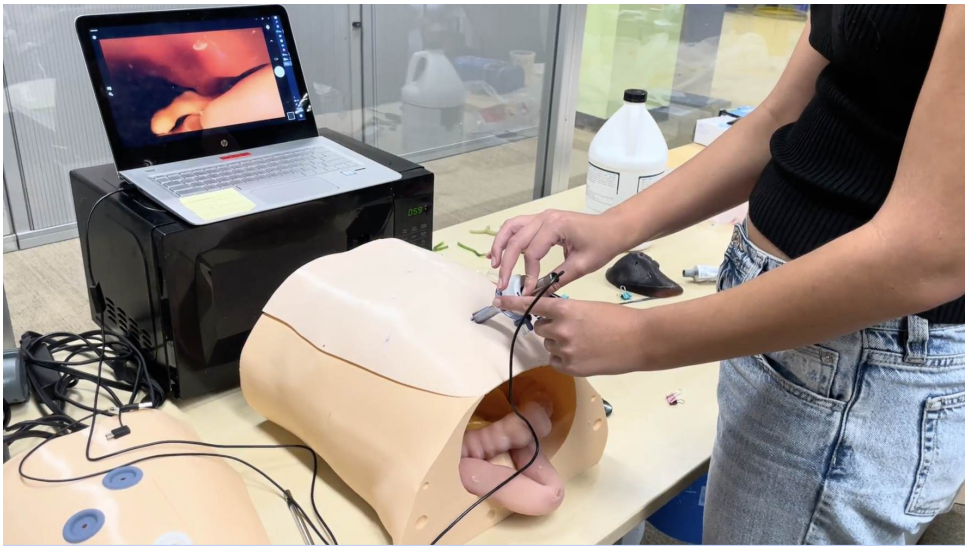

Figure 3: Simulated physical digestive cavity with laparoscopic camera.

Our survey shows that surgeons and trainees prefer physically realistic simulation rather than non-haptic virtual reality training systems. In order to mimic the physical procedures, we still have a lot of work to do, including the simulation of electrical knife cutting, interconnected vessels, bleeding, inflammation, and fatty tissues.

SPATIAL AUDIO

Indian musicology professor Inayat Kahn said: "In sound, the wise can interpret the secret of the nature of the working of the whole universe" (Kramer, 1994). In fact, audible sound and ultrasound have been widely used in medicine, especially in diagnoses. In Extended Reality, sound is certainly another dimension of reality. As we know, the soundscape in a surgical theater in the real world is complex, dynamic, and noisy. There are verbal communications between surgical team members, the sound from the vital signs such as ECG, the sound from procedures such as electrical knife, and ambient noises, etc. Each sound source normally has a particular location and we can use the spatial information as a reference for attention, event detection, and pattern recognition.

Spatial audio is a location-referenced recording and playback technology. It is an audio experience for enhancing immersive perception by simulating a surrounding sound setup. The 3D surrounding sound can be achieved by an array of speakers placed systematically around the room, or using headphones.

In our case, we used a four-channel microphone to record the sound from a laparoscopic surgical theater in a combined WAV file. We then split the channels into five spatial audio channels and play back in five speakers in the form of a circle. The spatial audio enables users and listeners to identify the locations of sound sources despite their orientation and position. Considering the

physical space conflict between the AR/VR headset and the headphones, we selected the plug-in style spatial audio earplugs, Apple's Air Pod, Generation Three, which tracks user's head movement and playback the surrounding sound either from a mobile phone or a laptop.

During the data collection process, we discovered that surgeons often use the vital signs to adjust the surgical pace or procedure. For example, when the surgeon heard the patient's pulse rate increased from the ECG sound, the surgeon would realize the patient might be in pain and would ask the anesthesiologist to give the patient more anesthesia. In order to simulate the dynamic events in the surgery, we need to articulate the ECG sound. Instead of using the recorded ECG sound data, we developed an adjustable ECG sound that can be merged into the spatial audio and can be tuned in real-time to simulate the patient's conditions.

EXTENDED REALITY SIMULATOR

An Extended Reality Simulator is to fuse the physical model with the digital model in real-time through interactive interfaces. The goal is to blur the borderline between these two types of models and to make the simulator look like real. The physical model and the digital model can compliment each other, for example, the 3D model of a liver or a stomach can be manipulated by the hand tracking sensor for translation, rotation, and scaling, and seamlessly to be placed on the proper location in the monitor of the live laparoscopic surgery video. Figure 4 shows the 3D liver model was placed onto the live laparoscopic view in the physical cavity. The monitor is a light-field based stereo display without stereoglasses.

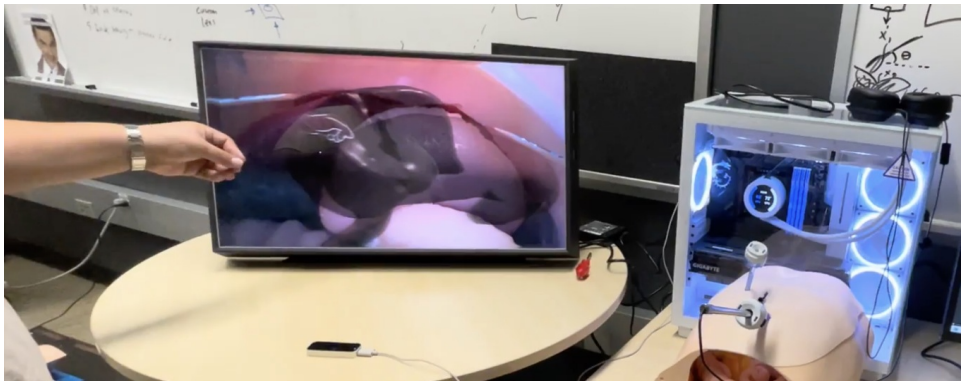

Figure 4: The live laparoscopic video is overlaid with CT 3D liver model by hand tracking.

The ultimate prototype of the Extended Reality (XR) is to simulate the surgical theater with physical and virtual details, including the haptic model, spatial audio, surgical team avatars and surgical equipment. Naturally, we overlaid virtual reality scenes onto the physical simulation model. In this study, we overlaid the laparoscopic surgery bed with the patient and related vital signs equipment on the 3D printed digestic cavity. Behind the surgical

bed, the physical laptop screen shows live video from the simulated laparoscopic camera inside the cavity. Figure 5 shows the screenshots from the Augmented Reality (AR) goggle HoloLens 2 in two different angles. With the hand tracking interface in the HoloLens 2, the user is able to move the virtual reality objects such as the surgical bed around with the hand gestures, such as grasp, move, rotate, rescale, and release. This allows the user to adjust the AR layout to fit the training tasks and the surrounding see-through environment.

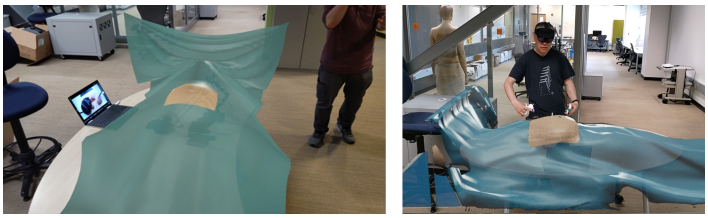

Figure 5: The view from the AR goggle of the virtual bed with the physical simulated cavity.

There are two significant human factors and ergonomic design issues in the interaction between hand control and surgical hand motion, and the interaction between the physical and virtual objects on the holographic screen on the headset. The surgical hand gesture for holding a laparoscopic instrument is easy to be confused with the hand gesture control gesture without an instrument. This problem can be resolved by turning off the hand gesture control during the surgical simulation, or improving the hand with instrument tracking. Furthermore, the animated 3D virtual hand model can be corrected with articulation, for example, bending the index figure while holding a laparoscope without straight out.

The second problem is so-clalled "occlusion," which is more challenging because many AR devices assume there is nothing in between the overlaid virtual object and the headset. Normally, AR systems detect the empty flat surface in front of the headset and then place the virtual object on the surface. Most of AR devices do not detect or track small physical objects such as the laparoscopic surgery instrument. Therefore, the occlusion problem has been an open-ended problem across the industry. In our case, the physical laparoscopic instruments would "disappear" in the Extended Reality scene. In order to display the physical objects in front of the virtual object, we have at least two approaches: better instrument tracking interface, or better virtual hand overlay algorithms. The tracking interface can be improved by multiple sensors such as multiple cameras, or multimodal sensors such as inertial motion unit (IMU) sensors, or near-infrared light reflective trackers, e.g. OptiTrack (Natural Point, 2024), or 3D cameras, e.g. Intel's RealSense depth cameras (Intel, 2024). Figure 6 shows the hand-tool tracking problem and the occlusion problem.

Figure 6: The hand-tool tracking problem (left) and the occlusion problem (right).

MACHINE VISION OVERLAY

The real-time visual overlay of machine vision results is perhaps the most challenging and expanding task for the multimodal laparoscopic surgery simulator. The real-time overlaid contents are designed to assist the user to identify the critical anatomic structures, to get familiar with the new surgical procedures, instrument control dynamics, and the panoramic stereo view of the targeted area. The results of the machine vision algorithms can be overlaid to the regular live laparoscopic screen that can be converted to the in-vivo operations in the real-world.

The near-infrared (NIR) imaging is a rapidly growing method for laparoscopic surgery, especially for identifying critical vessels and organs. In this study, we simulated the live NIR video that merges physical simulated organs and digitally articulated highlights. The user can experience the haptic perception and the visualized near-infrared image at the same time. This is a one step toward the extended reality (XR) at the physical spectrum. Figure 7 shows the simulated near-infrared imaging over the live laparoscopic video on the monitor.

At the visible spectrum, we have developed vision algorithms for detecting and tracking the critical anatomic structure such as Calot's Triangle (Abdalla, 2013) and laparoscopic camera motion. Calot's Triangle is a critical anatomic structure near the gallbladder and liver where critical artery vessels pass through the area. This is important to laparoscopic surgeries such as cholecystectomy. In real-time, the machine vision overlays the bounding box around the Calot's Traiangle area to alert the user. Sometimes, the feature does not exist, the algorithm would attend a few times and produce the warning text.

The surgical instrument tracking, on the other hand, is useful for surgical coordination training. For example, normally the surgeon's assistant holds the laparoscopic camera and tries to aim at the tip of the laparoscopic instrument such as a needle grasper. Ideally, the surgeon wants the camera follows the tip of of the instrument at the center of the screen. We adopted the score system from the clinical practice, for example, the tip-in-center position is 10/10 and the tip-outside-range would be 0. The user can see the real-time feedback on the screen. Figure 8 shows the Calot's Triamge detection and tracking with the extended reality cavity and the laparoscopic camera tracking.

Figure 7: The simulated near-infrared imaging over live laparoscopic video.

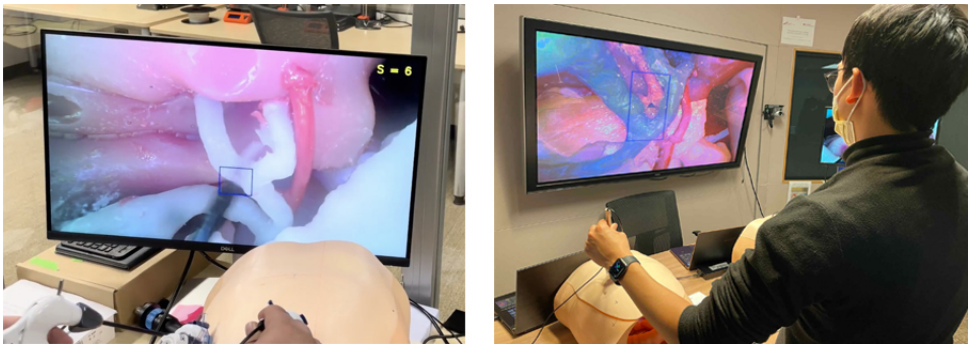

Figure 8: Calot's Triangle detection and tracking with the extended reality cavity (left) and the laparoscopic camera tracking (right).

Virtualized Reality is to capture, model, and render 3D objects from 2D or 3D image and video sources, e.g. live 3D scope camera, CT, MRI, NIR images, and tissue scanning. The photorealistic *virtualized reality* emphasizes that the image data was natural rather than the synthetic imagery used in virtual reality. It refers to the production of views of a rendering model where the geometry and appearance content is derived from measurements of a real scene. For visualization, such a model makes it possible to render synthetic views of the scene from arbitrary perspectives that may never have been the site of any real sensor. Such techniques represent an extreme on a spectrum of real data content with augmented or mixed reality somewhere in between and virtual reality at the other extreme. Virtualized reality enables a new capacity to address many of the simulation problems by providing a photorealistic, synthetic, line-of-sight view to the operator based on the content of location-augmented real-time video feeds.

We experimented with multiple 3D capture methods and assess their visual and interactive effects in virtual reality. Our methods include photogrammetry and Neural Radiance Field. Photogrammetry (or Structure from Motion)

is to reconstruct structure from motion. It takes input from images or lidar to generate a 3D point cloud and then connect them into many triangles (mesh), and overlay color and texture on them. The advantages of photogrammetry include compactness and compatibility with most XR systems. However, it often oversimplifies the shape during the mesh generation process so that the surfaces appear to be rigid. Besides, it doesn't work well with the reflective surfaces. Neural Radiance Fields (NeRF) is a method for synthesizing novel views of complex scenes by optimizing an underlying continuous volumetric scene function using a sparse set of input views (Datagen, 2024). NeRF renders 3D volumetric blocks that are common in many medical imaging systems such as MRI, CT, and 3D ultrasound in DICOM data formation. It works with highly reflective surfaces and complex scenes such as trees. However, it is computationally expensive and has compatibility issues with existing XR platforms.

In this study, we explore the reality capture methods and develop an efficient approach for capturing, modeling, and rendering organ tissues.

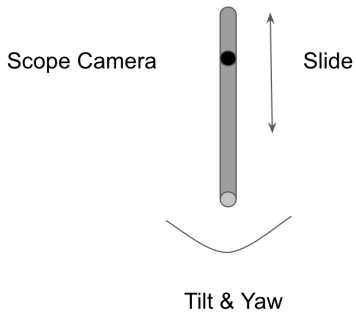

Figure 9: Panoramic views, photogrammetry, and measurements of the organs and scope status from a single scope camera port.

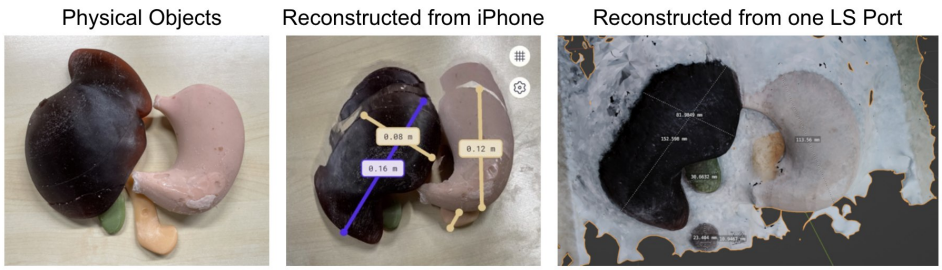

Figure 10: The physical objects and reconstruction with iPhone and laparoscopic camera.

Table 1. Measurements from reconstructed objects.

Mode	Measured Distances (MM)			
Physical	158	90	25	120
iPhone	160	80	25	120
Laparoscope	152.6	82	24	113

CONCLUSION

We developed a multimodal extended reality system for laparoscopic surgery training. The system contains multimodal feedback and holographic overlays. In particular, we have integrated haptic organs into the simulator that fills the gap of the mixed reality interfaces for surgical training needs.

The holographic overlay matches and synchronizes with the simulated or actual tissue, and it tracks in an overlapping way regardless of how the camera is moved and how organs are displayed on the screen. The 3D objects from the CT DICOM data can be overlaid to the live simulated tissues in the cavity. The object registration can be controlled by hand gestures.

The machine vision algorithms are designed to enable the dynamic overlay process on the regular live laparoscopic surgery video. For example, we overlay the symbolic Calot's Triangle based on the Visual or Near-Infrared video data. The machine vision also includes Virtualized Reality to capture, model, and render 3D objects from 2D or 3D image and video sources, including the live 3D scope camera, CT, MRI, NIR images, and tissue scanning. The photorealistic virtualized reality emphasizes that the image data was natural rather than the synthetic imagery used in virtual reality. The 3D reconstruction results show that the 2D laparoscopic camera can achieve a reasonable accuracy in measured distances.

Our experiments indicate that multimodal extended reality can increase the fidelity of the laparoscopic surgery simulation and potentially improve the training efficiency.

ACKNOWLEDGMENT

The author would like to thank the support of the Wellcome Leap SAVE Program and the NIST PSCR Program. The author would like to thank the Research Assistants for their implementations Talia Perez, Andy Cao, William Zhang, and Anthony Wang. The author also wants to thank senior software engineer Alexandra Poltorak for her editing and the artist Maria Agustina for her medical illustrations.

REFERENCES

Abdalla S, Pierre S, Ellis H. (2013). Calot's triangle. Clin Anat. 2013, May;26(4): 493–501. doi: 10.1002/ca.22170. Epub 2013 Mar 21. PMID: 23519829.

Cai, Y. (2022). Learn on the Fly. AHFE 2022.

Cai, Y. (2024). Episode memory with 3D interactive sequential graph, AHFE 2024, Nice, France, July 2024.

Cai, Y. Park, J. Hope, L. (2024). Articulated spatial audio for minimally invasive surgery training, AHFE, Nice, France, July 2024.
Cai, Y. Perez, T. (2024). Haptic perception in Artificial Tissues, AHFE, July 24–27, Nice, France, 2024.
Crothers, I. R. Gallagher, A. G. McClure, N. James, D. T. D. and McGuigan, J. (1999). Experienced laparoscopic surgeons are automated to the "fulcrum effect": An ergonomic demonstration, *Endoscopy*, vol. 31, no. 5, pp. 365–369, 1999.
Datagen (2024). Neural Radiance Field (NeRF): A Gentle Introduction. https://datagen.tech/guides/synthetic-data/neural-radiance-field-nerf/
Eskef K, Oehmke F, Tchartchian G, Muenstedt K, Tinneberg HR, Hackethal A. (2011). A new variable-view rigid endoscope evaluated in advanced gynecologic laparoscopy: A pilot study. Surg Endosc 2011;25(10).
Gallagher, A. G. McClure, N. McGuigan, J. Ritchie, J. and Sheehy, N. P. (1998). An ergonomic analysis of the fulcrum effect in the acquisition of endoscopic skills. *Endoscopy*, vol. 30, no. 7, pp. 617–620, 1998.
Intel RealSense (2024). https://www.intelrealsense.com/
Kramer, G. (1994). Auditory display. Addison-Wesley Publishing Company, 1994.
Landsberg, C. R., Mercado, A. D., Van Buskirk, W. L., Lineberry, M. & Steinhauser, N. (2021). Evaluation of an adaptive training system for submarine periscope operations. Proceedings of the Human Factors and Ergonomics Society. 56(1), 2422–2426.
Lin, C. J. Chen, H. J. (2013). The Investigation of Laparoscopic Instrument Movement Control and Learning Effect, BioMed Research International, vol. 2013, Article ID 349825, 16 pages, 2013. https://doi.org/10.1155/2013/349825
Mayer, R. E., Makransky, G., & Parong, J. (2023). The promise and pitfalls of learning in immersive virtual reality. International Journal of Human-computer Interaction, Volume 39, 2023 - Issue 11: Trends in Adaptive Interactive Training Systems. http://www.tandfonline.com/doi/abs/10.1080/10447318.2022.2108563?journalCode=hihc20
Metzler-Baddeley, C., & Baddeley, R. J. (2009). Does adaptive training work? Applied Cognitive Psychology, 23, (2), 254–266.
Natural Point (2024)., www. NaturalPoint.com.