Creating a Smart Eye-Tracking Enabled AR Instructional Platform: Fundamental Steps

by

Yerly Paola Sanchez Perdomo

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in
Surgical Education

Department of Surgery
University of Alberta

Abstract

Surgical procedures are demanding to learners because they need to perform and excel on the task to succeed, and eventually, provide a high standard of care to patients. During training, safety and proficiency are emphasized. However, the limited numbers of hours and instructor feedback for learning and practicing procedures in certain medical schools make it challenging to master surgical techniques. As a result, simulation has been implemented, and new technologies such as augmented reality (AR) are employed in medical and surgical education. In the design of a useful AR platform for learning and practicing surgical skills, it is necessary to provide only essential information, thereby reducing cognitive load to the user. In this project, we provide the foundation of a smart eye-tracking enabled AR platform for teaching the multistep procedure of chest tube insertion. In this thesis, we answered two core questions. First, can eye tracking and AR devices be synchronized and integrated into one platform to be potentially used for the practice of a chest tube insertion? Second, can eye tracking identify the moment of performance difficulty during a multistep surgical procedure, specifically a chest tube insertion? This project lays the foundation for a customized eye-tracking AR teaching platform for chest tube insertion.

Preface

This thesis is an original work by Paola Sanchez. The research projects in this thesis received ethics approval from the University of Alberta Health Research Ethics Board, “Augmented Reality Assistance During a Chest Tube Placement Task”, No Pro00080600, May 24, 2018. Funding from NSERC Discovery 2016-2021 and U of A TLEF Award 2016-2019 was used to support the research studies in this thesis.

Some parts of the thesis are under review or will be published. A version of the technical component of Chapter 4 is under consideration in the *Journal of Medicine System* as Lu S, Sanchez P, Jiang X and Zheng B. “Integrating Eye-tracking to Augmented Reality System for Surgical Training.” I was responsible for incorporating the AR and eye-tracking devices hardware, testing the accuracy of the devices, research design, concept formation, manuscript writing and editing, and supervision of a computer science student, Lu S. Data was not collected in this chapter. Lu S contributed with the technical integration of the systems and manuscript composition. Jian X assisted with the concept formation and manuscript editing, and Zheng B was the supervisory author and contributed with the concept formation and manuscript composition. Versions of Chapters 3 and 5 will be published in journals to be determined. Literature review, research design, construction of the chest tube model, data collection and analysis of the research data were my original work. Jian X wrote the bases of the MATLAB code utilized in Chapter 5. The literature review in Chapter 2, as well as Chapter 6 and all the writing in this thesis are my original work.

Acknowledgments

I am immensely grateful to the people who supported and encouraged me to carry out this thesis. First, my supervisor, Dr. Zheng, has been a great support during my master's program. He always listened to my ideas and concerns and provided constructive feedback and guidance to choose the most appropriate steps in my research. Also, he offered me pieces of advice during moments of struggle and encouraged me to choose the best for my family and me. Thank you for not only being a supervisor but a mentor.

I was always nervous and full of questions before my committee meetings. However, I was fortunate to have a variety of committee members who were knowledgeable and experts in their fields, Dr. Jason Harley from Educational Psychology, Dr. Jonathan White from Surgery, and Dr. Pierre Boulanger from Computer Science. After each committee meeting, I was relieved and enthusiastic about continuing my master's. My committee members provided insightful comments to help me overcome difficulties and come out with ideas.

I want to thank the members of the Surgical Simulation Research Lab, especially, Shuqi for assisting me with the skin pads and Xianta, who helped me with the coding needed for one of my projects and his responses on questions about computer science. Similarly, I want to express my appreciation to Shang, a computer science student who worked in the lab with me in the technical component of the integration of eye tracking and augmented reality devices.

I cannot thank enough my soulmate, my husband. He listened to my worries and frustrations and cheered on my achievements. I would not have gone through a master's without his unconditional love and support to fulfill my professional and personal goals.

Table of Contents

List of Tables

List of Figures

Figure **Page**

List of Abbreviations

MIS: Minimally Invasive Surgery

SBT: Simulation-Based Training

VR: Virtual Reality

AR: Augmented Reality

HMD: Head-Mounted Display

CT: Computed Tomography

MR: Magnetic Resonance

US: Ultrasound

TEL: Technology-Enhanced Learning

VIPAR: Virtual Interactive Presence with Augmented Reality

AOIs: Areas of Interest

TMD: Table-Mounted Display

QE: Quiet Eye

HMM: Hidden Markov Models

CAI: Computer-Assisted Instruction

ATLS: Advanced Trauma Life Support

NASA TLX: NASA Task Load Index (TLX)

MPD: Moment of Performance Difficulty

NMP: Normal Moment of Performance

HPU: Holographic Processing Unit

FOV: Field of View

UDP: User Datagram Protocol

IPC: Inter-Process Communication IPC

Chapter 1 General Introduction

In this first chapter, I introduce my experiences as a medical student, emergency physician, and research assistant that led me to the research topic of this thesis. Then, I define the problem and research questions. Finally, I conclude this chapter with the outline of the thesis.

Why did I Choose this Research Topic?

During medical school, learning new surgical procedures was difficult for me. I was intimidated by the complexity of the tasks and the precision with which they needed to be delivered. In the past, one way to become proficient in these procedures was to practice with live patients, animal models or with cadavers. Practicing on live patients and animal models are not optimal options given the moral issues attached to the procedure and the jeopardization of patient safety. I learned to perform surgical procedures on cadavers donated to the university. Using cadavers has ethical, economic, and availability issues. First, even though I treated cadavers with maximum respect, I could not stop thinking that they were once living humans. Second, the preparation of a cadaver for teaching purposes requires many hours and has a high economic cost. Another issue is that not every medical student is comfortable working with cadavers either for moral or emotional causes. Lastly, there are not enough cadavers available for the number of medical students who need to learn procedures; therefore, the opportunity to prepare students to perform essential procedures when they are working independently becomes questionable. For example, I noticed during my work as a physician in the emergency room that I wished I could have more opportunities to practice surgical procedures, especially complex and multistep procedures such as intubation or chest tube insertion. I was afraid that I could not perform the procedure correctly, and the patient would not obtain the benefits of this critical intervention.

All the problems and limitations associated with the use of animal models or cadavers for learning made me wonder about other ways to learn surgical procedures. When I was introduced to the Surgical Simulation Program at the University of Alberta in 2016, I decided to learn more about simulation and its role in surgical skill training.

How Can Simulation Help to Solve the Problem?

During my time working as a research assistant at the Surgical Simulation Research Lab (SSRL), I realized that there are various methods and resources to teach surgical procedures. I was fascinated with the idea of introducing technology, such as simulation, into learning. Simulation offers the opportunity to practice surgical procedures in inanimate models in a broad spectrum of complexity, from practicing simple maneuvers such as a suture to more intricate ones like putting a tube into a chest cavity and fixing it onto the chest.

Most learners struggle during the practice of complex surgical procedures, and they do not have the opportunity to repeat the procedure many times in order to master it. Using simulation and the latest technology could allow students to have more opportunities to practice and learn multistep procedures such as chest tube insertion. Simulation is a field that is growing, and simulators are becoming more realistic as the technology evolves. At the SSRL, I learned about the different types of simulators, from physical task trainers to computer-based virtual models for learning and practicing surgical procedures. Among these models and training systems, I was interested in the role of augmented reality and eye tracking technology in surgical simulation. In brief, augmented reality technology offers new types of simulators allowing the users to interact with the real-world while receiving computer-generated information to guide them through a task. Eye tracking is a technology that offers an objective assessment of performance and identifies moments where participants are having difficulty by tracking eye movements.

In this thesis, I want to contribute to the ever-growing field of simulation and introduce the steps towards the construction of a smart simulation platform for the practice of a multistep surgical procedure, chest tube insertion, without the ethical limitations of using cadavers or hybrid models. This will ensure the future of patient safety as the trainee obtain proficiency in surgical procedures.

Problem Definition and Research Questions

In healthcare, all procedures include multiple steps, such as a chest tube insertion. These entire procedures should be performed correctly and competently on all the steps and in the shortest period of time possible. Traditional methods to learn these procedures consist of either watching a video, observing a live expert after reading about the task or practicing in a simulated model. However, trainees face several challenges during the learning and practicing of surgical tasks, thereby, affecting their performance. For example, students can easily forget the steps or amy not know how to perform a specific action.

How can new technology change the way we teach surgical procedures?

Utilizing eye tracking data as input signal may detect the moment of learning difficulty, such as fixations and saccades towards non-related task areas. Eye tracker technology offers the possibility to understand people's behaviours and thinking by analyzing eye patterns. After the eye-tracker identifies the moment of the learner's struggling, guidance could be given to the trainee to complete a task. Taking it one step further, Augmented Reality (AR) has the potential to facilitate the interaction of users with the real world and virtual elements. Therefore, combining AR and eye tracking technologies could allow us to display assisting information at the right moment of the procedure without affecting the learning process. Achieving a functional eye tracking enabled AR platform requires intricate and multiple steps. This thesis is focused on the first steps of this process; we will answer the following research questions:

1) Can eye tracking and AR devices be synchronized and integrated into one platform to be potentially used for the practice of a chest tube insertion?

2) Can eye tracking identify the moment of performance difficulty during a multistep surgical procedure, specifically a chest tube insertion?

Outline of the Thesis

This thesis is composed of six chapters. As you already known, in Chapter 1, I provide a general introduction, and post two major research questions of this thesis. In Chapter 2, I review current literature on AR and eye-tracking technology. In Chapter 3, I describe the way of constructing a simulator model for chest tube insertion, which is used in the next chapters of the thesis. In chapter 4, I describe the challenges of integrating eye tracking and AR into one platform. Technical details were reported on how I found the solution with the assistance from Computing Science to incorporate the eye tracker to the AR goggles. In Chapter 5 I report my effort of extracting data from eye tracking metrics to define the moments of performance difficulty (MPDs) during the performance of a chest tube insertion. Defining MPDs is a critical step before we can use eye tracking data as a signal to prompt appropriate instructional message on the AR goggles while a trainee is performing a task. I am delighted to report my achievement on this portion of work based on human eye movement data. My supervisor and I believe that the future step of linking MPDs to the delivery of instructional message to the AR goggles is heavily relied on the expertise of computing programming, which is beyond my scope of knowledge. Therefore, in Chapter 6, I offer a general discussion of my current work, and provide a vison on the future work.

Chapter 2 Literature Review

Chapter 2 provides background information on the main topics in this thesis. I provide a brief description of Surgical Education focusing on simulation, followed by what is known about augmented reality and eye tracking technologies and their role in surgical education. I will conclude this chapter with the literature review of the combination of eye tracking and augmented reality, including potential challenges.

2.1 The Current State of Surgical Education

In North America, the formal education program is given to those who determine to become a physician from the beginning of their medical school and continues during residency training. Medical education involves all medical specialties, including surgery. Surgical education has emerged to address the needs of any physician who wants to become a proficient surgeon (e.g., dexterity skills).

In 1889, influenced by his experiences in Europe, Dr. William Halsted developed the first surgical training program in America while he was the chief of the department of surgery at Johns Hopkins Hospital. Dr. Halsted's surgical program was based on long hours of work under the supervision of an experienced surgeon, allowing surgical residents to have extensive clinical exposure and receiving increased responsibility through the years of residency (Carter, 1952; Gallagher, O'Sullivan, & O'Sullivan, 2012). The Halsted model has been implemented for long time since its creation. However, surgical education started to change after the establishment of the Accreditation Council for Graduate Medical Education (ACGME) in 1982. The ACGME introduced in 1999 six core competencies: medical knowledge, patient care, interpersonal and communication skills, professionalism, practice-based learning and improvement, and system-based practice (Sachdeva, 2007). Another significant change in surgical training was the reduction of working hours for residents. This regulation was implemented in 2003 and modified in 2011 to improve patient safety and quality of life for the trainees (Moonesinghe, Lowery, Shahi, Millen, & Beard, 2011). Now, residents need to achieve the six competencies introduced by the ACGME, and surgical programs are focused on evaluation of outcomes.

In Canada, surgical education has evolved to meet the needs of modern society, switching from a time-to an outcome-based training program. First, the Royal College of Physicians and Surgeons of Canada (RCPSC) employed the Canadian Medical Education Directions for Specialists (CanMEDS) project. CanMEDS encompasses seven physician roles: Medical Expert, Communicator, Collaborator, Manager, Health Advocate, Scholar, and Professional. Initially, CanMEDS was adopted in 1996 by the RCPSC; then, it was reviewed and updated to be re-implemented in 2005 (Frank & Danoff, 2007). Implementing CanMEDS into standards ensures that competence is more than medical expert responding to the demands of society (Frank & Danoff, 2007).

In 2015, the CanMEDS had its latest renewal and the Competence By Design (CBD) framework was introduced, redefining the curriculums for the Canadian medical and surgical specialties (Skinner, Ho, & Touma, 2017). In the CBD framework, trainees must achieve outcomes (Entrustable Professional Activities (EPAs) and CanMEDS milestones) to move to their next stage of training. EPAs are the tasks that a trainee can perform without supervision once the supervisors consider the trainee competent (Ten Cate, 2013). Therefore, CBD provides a certain degree of confidence that trainees are well prepared as they need to demonstrate competency during their training before graduation (Frank et al., 2010).

Surgical education continues evolving as practicing in animal models, increasing public awareness on patient safety and the introduction of more complex surgical procedures, such as those in Minimally Invasive Surgery (MIS). Therefore, surgical simulation has been integrated into the traditional education system (Gallagher & O'Sullivan, 2011). Simulation increases hours and settings for practicing, facilitating achievement and maintenance of skills.

Introducing Simulation-Based Training (SBT) into surgical education facilitated achievement of ability, maintenance of skill, and development of proficiency-based training models, and plenty of research in MIS has validated the translation of technical skill obtained via simulation to the operating room (Pinzon, Byrns, & Zheng, 2016; Stefanidis et al., 2015).

Among different SBTs employed for teaching surgical skills, physical model-based, computer-generated and hybrid are the most used. Physical model-based simulators use an inorganic representation of human tissues. Computer-generated models create simulations using digital animations, such as virtual reality (VR). The hybrid model is a combination of physical objects and computer-generated graphics (Kneebone, 2003); this is often found in Augmented Reality (AR) environments. Satava, one of the pioneers in using Virtual Reality (VR) in surgical education, introduced a classification of simulators base on the skill they are intended to train to surgeons (Satava & Satava, 2001). This taxonomy can be seen in Table 2.1.

Models that use VR and AR, bring innovative advantages to the learning process of surgeons, such as in-vivo and objective feedback of performance. However, these educational tools cannot be valid until the knowledge and skill acquired using VR and AR models are transferable to the real world; as Satava mentioned "the student runs the risk of practicing the skill to perfection but perfectly incorrect"(Satava & Satava, 2001, p. 1484).

For SBT, several virtual-based models have been validated in surgical training, such as LapVR (CAE), LapMentor (Simbionix), and LapSim (Surgical Science) (Loukas, 2016). However, SBT using augmented reality technology has not been studied extensively. Nevertheless, AR offers the option to become competent in complex surgical psychomotor skills by breaking down complex procedures into parts to enable automation (S. M. B. I. Botden & Jakimowicz, 2009). The unique feature of AR is that learners receive instructional messages while practicing a task in physical models facilitating the performance of the task. Furthermore, including AR into MIS may alleviate the hindrance of vision and tactile feedback by enforcing depth perception via image-guided surgery (Bernhardt, Nicolau, Soler, & Doignon, 2017). I will present an overview of AR technology in the next section and highlight what we already know about the application of AR in education.

Table 2.1.

Classification of simulators (Adapted from Satava)

Skill	Manual requirement	Examples
Precision placement	Direct needle/instrument to a point	Intravenous needle placement, spinal anesthesia, needle biopsy
Simple manipulation	Guide catheter or endoscope	Coronary angioplasty, endoscopy, ultrasound
Complex manipulation	Perform a single complex task	Anastomosis, MIST
Integrated procedure	Perform multiple tasks of the entire procedure	Anesthesia, laparoscopy, arthroscopy

MIST: Minimally Invasive Surgery Trainer-Virtual Reality

2.2 Augmented Reality

What is augmented reality?

Augmented reality belongs to the spectrum of Reality-Virtual continuum (Milgram et al., 1995). In this model, *Virtual Environment* stands on one side of the continuum. In virtual reality, computer-generated graphics are designed to replace the real world. These artificial objects can be displayed on a monitor or designed into an immersive environment where the user is disconnected to the real world, receiving a stimulus to the visual, auditive, and sometimes tactile channels to make the illusion of reality. As a result, virtual environments interfere with the interaction between users and the real world. On the other end of the continuum, the *Real Environment* represents the physical world everyone has access to and interacts with. Between the virtual and real environments, the *Mixed Reality* of both virtual and real-world elements is part of the Reality-Virtuality continuum, Figure 2.1 (Milgram et al., 1995). Mixed reality is composed of two parts: one that is predominantly virtual with fewer real components called Augmented Virtuality (AV), and the other which has more authentic elements than virtual, called Augmented Reality (AR). AR is the compilation of technologies that improve a person's ability to interact with the real world via the superimposition of images produced by a computer. It offers the advantage of interacting with the real world while users access digital information at any moment they want. Conventionally, AR displays as a visual enhancement of an environment, but it can also take the form of any of the human senses (Carmigniani et al., 2011).

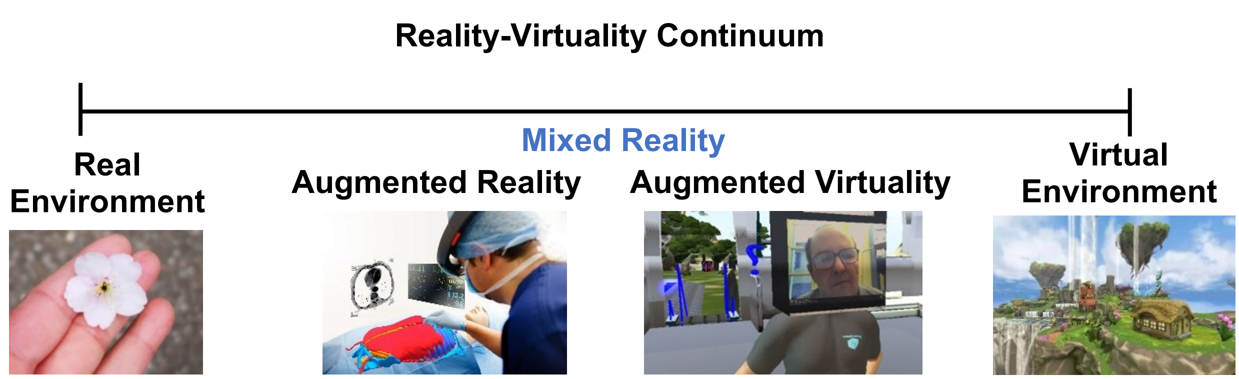

Figure 2.1. Augmented reality: a class of displays on the Reality-Virtuality continuum. It is adapted from (Milgram et al., 1995).

Typically, three types of technologies allow the display of virtual information into the real world to create an AR environment: Head Mounted Display (HMD), handheld display, and projective dispaly (Azuma et al., 2001). As the name implies, the HMD can project a computer-generated image onto a screen directly in front of the user's eyes. HMD can either display on a semi-transparent mirror (optical see-through) or onto a darkened mirror casting video captured from a head-worn camera as the environment for the AR. Handheld devices can display virtual objects directly on the top of real objects, quickly creating AR scenarios (Azuma et al., 2001). For instance, using a tablet, computer-generated images of the internal structures of a brain can be superimposed on a patient's head during brain surgery. In the projection display, the computer-generated objects can be directly overlying on the surface of the real object (Wang, Watts, & Zheng, 2017). In any of these three models, the virtual information can be display either in text or graphics fashion.

In the next section, the displays and usages of AR in surgery and surgical/medical education are expanded.

Mapping virtual images to the real world.

Overall, HMD and handheld devices are the most frequently employed to create AR environments in medical training (Barsom, Graafland, & Schijven, 2016; Kamphuis, Barsom, Schijven, & Christoph, 2014). There are two possible ways the computer-generated message can be visualized and mixed with the real world. The first one is to superimpose the desired virtual image onto real objects, guided by markers that can be identified with a camera (Cheng & Tsai, 2012). Once the computer detects the markers, it can adjust the desired image on the screen based on the location of the markers in the real world. The second one is markerless AR that launches models based on position data concerning the display, wireless network or global positioning systems (GPS) (Cheng & Tsai, 2012). In most AR platforms, markers placed on the real world are used for guiding the mapping process.

Instructional messages for guidance in a surgical procedure can be displayed employing four different methods. A monitor or *fixed video* is one of the most utilized techniques to display the virtual information in a video feed, such as an endoscopic video (Nicolau, Soler, Mutter, & Marescaux, 2011). A more practical technique is the AR *projection of anatomical features* onto the patient (Wu, Wang, Liu, Hu, & Lee, 2014). This last method is excellent for instrument placing but has some limitations. The projection angle often is obstructed by the surgeon, can become inaccurate with the pneumoperitoneum, and can be different from the operator's point of view (Bernhardt et al., 2017). A more sophisticated approach that is gaining popularity is the *online presentation* of AR data in an HMD of live video (Cutolo et al., 2016). This see-through video technology allows for surgical planning on a mobile display with motion tracking. Lastly, a different technique for AR display is the *image overlay* on the field or "AR windows"(Bernhardt et al., 2017). This setup can be either on an HMD or on semi-transparent mirrors. I have described the general concepts of AR display in the real world. In the next two sections, I will expand on AR in the surgical field.

Creating digital 3D models for AR displaying in surgery.

The first step in creating an AR environment is to use the computer to generate a 3D image close to the real anatomy of a patient. Typically, such a virtual model is derived from a series of cross-sectional images taken by the high-resolution Computed Tomography (CT) or Magnetic Resonance (MR) imaging devices. These imaging sources are the highest quality available. A CT scan is a consolidation of several x-ray measurements from different perspectives diverging into a cone to create cross-sectional virtual slices of a body part (Goldman, 2007). By digitally processing these slices, a 3D rendering of the inside of the organ is obtained. This 3D rendering can be further projected as a 2D image on a computer screen.

The fundamental 2D picture element, or pixel of each image, is represented on a scale of radiodensity. Each of these pixel densities represents a type of tissue corresponding to an attenuation scale known as Hounsfield (Goldman, 2007). The aggregation of this digitized data from these imaging devices reproduces a copy of the human anatomy. The anatomy in this type of image can further be enhanced via a contrast either venous or arterial, during angiography for vessels. A similar procedure is performed to exalt tumours, organs, and bones with rendering by methods of segmentation (Bernhardt et al., 2017; Kersten-Oertel, Jannin, & Collins, 2012). Acquiring a high resolution of the cross-sectional image and developing an effective segmentation algorithm is important for creating the 3D virtual model.

Introducing digital models during surgery.

Augmented Reality technology has been added to Surgery to enhance preoperative planning and intraoperative exploration. The data to generate 3D images in AR are obtained from the previously mentioned visuals sources (CT scan, MR, and the US), but the difference from the traditional methods is that in AR, the 3D images could be displayed on top of the patient. Intraoperative AR systems facilitate surgical navigation of complex anatomy as it has been shown in tumour resection in breast surgery (Sato et al., 1998). Also, Lovo et al. utilized co-registration of AR to display the cerebral cortex, and venographic images in patients undergoing surgery are in its early stages. In this study, eight patients had the concordant 3D reconstruction of venous vessels, lesions, and cerebral cortex by employing fiduciaries markers displayed in the surgical field. Confirmation of the AR overlay precision of this changing anatomy is "adequate" via Intraoperative ultrasound and stereotaxis (Lovo et al., 2007). The preoperative and intraoperative applications of AR have been applied to Neurosurgery, Orthopedic Surgery, General Surgery, Maxillofacial Surgery, Otolaryngology, and Cardiovascular and Thoracic Surgery (Shuhaiber, 2004).

Unfortunately, once an operation starts, CT and MRI are not always available for creating a 3D model. Instead, intraoperative data is taken in by devices with lower image resolution. This hindered condition arises from the low invasiveness of these devices, such as open MR scanners, ultrasound (US), and flat-panel cone-beam CT (CBCT) (Bernhardt et al., 2017). Contrast agents and online digital processing can be applied to increase the quality of the imaging in these former cases. The image fusion technology can be applied to merge intraoperative images taken by devices, such as ultrasound to the 3D virtual model built using preoperative CT or MR data (El-Hariri, Pandey, Hodgson, & Garbi, 2018; Prada et al., 2014).

Impact of AR on human operators.

Keeping visual presentations as natural as possible for human operators has technological challenges during the mapping of a virtual model into real targets. Several perception tactics can be utilized to make AR depth more credible to the human performer (Welchman, Deubelius, Conrad, Bülthoff, & Kourtzi, 2005). The most powerful tactic is occlusion. Occlusion is often presented as a virtual window and partially obstructs the image in the background to give the illusion of layers (Bichlmeier, Heining, Feuerstein, & Navab, 2009). In some cases, inverse realism superimposes the actual image of the surgical field onto the AR image to offer a sense of order (Pratt et al., 2012). Occlusion, although being the strongest of perception cues for depth only provides an understanding of the order of the objects but not the distances between them.

Two other perceptual cues that provide limited depth information are motion parallax and stereo disparity. Motion parallax contributes a sense of depth in surgery by perceiving monocular objects closer to the observer to move faster than those far away (Bichlmeier et al., 2009; Reichelt, Häussler, Fütterer, & Leister, 2010). In stereo disparity, the brain takes advantage of both retinal images from each eye (stereopsis). The strength of the latter from motion parallax is that stereopsis attains information from static data.

However, the accommodation and vergence desynchronization creates discomfort for some users in the AR environment (Reichelt et al., 2010). Shading and perspective may also improve depth cues in AR. By casting a shadow on other objects, we learn the object shadowing the other is closer to the light source (Cutting & Vishton, 1995). Perspective can be achieved by rendering the correlation of tumour contour and vasculature to their depth (Hansen, Wieferich, Ritter, Rieder, & Peitgen, 2010).

Although the above technology can enhance the visualization of the AR scenario, artificial images superimposed on the real world can still be a distraction. Dixon et al. evaluated the effect of inattentional blindness caused by augmented reality in a surgical scenario. In the task, two groups of participants (AR and control group) were asked to perform a laparoscopic task where unexpected objects were introduced. They found that the AR group had more accuracy in task performance than the control group, but they identified fewer unexpected findings ($p < 0.001$) (Dixon et al., 2013). Therefore, AR could sometimes disrupt human attention and interfere with skill learning.

2.3 AR in Surgical and Medical Education

With the advancement in computing, technology-enhanced learning (TEL) research has been increasingly focused on AR immersion for motivation purposes during the learning process (Dunleavy & Dede, 2014). Especially in surgery, simulation-based TEL has come to the forefront of skills training. The challenge becomes to find the right balance between the type of TEL required for the desired educational outcomes as surgeons advance during their training (*Surgical*

Education, 2011). For example, benchtop suturing simulation models would be ideal for junior surgery residents but not appropriate for senior residents.

Educational experts have made efforts to develop educational theories that optimize the training environments in surgery, given its dynamic nature. One well known theory is that problem-based learning (PBL) applies a wide range of educational ideas within a context, for instance, a clinical decision-making scenario. Overall, three educational theory models are supporting the efficacy of AR in learning: *Situated Learning, Constructivist Learning*, and *Experiential Learning*.

Situated learning theory refers to the learning that occurs when a trainee is immersed in a specific environment, and the knowledge acquired is the result of interactions within this given context (Lave & Wenger, 1991). In this theory, learning in the context allows the students to learn through work (Fry, 2011), which is related to the learning by doing in surgery. Situated learning theory is applied in immersive learning environments where the student participated in a realistic experience. Augmented reality interfaces offer the opportunity for vivid experiences through interaction with the real world. Therefore, students can experience near-transfer of knowledge (Dunleavy & Dede, 2014). In other words, AR immersive environments offer an ideal practice environment where the skills learned in this AR environment can be applied to a similar situation. For instance, a task like making an incision learned under AR environments can be repeated in a real-world scenario like an actual incision.

In *Constructivist Learning,* students create new knowledge by linking new information to previous knowledge and experiences (Dunleavy & Dede, 2014). It offers the advantages of a self-driven and active learning process and the application of knowledge-based to new environments. The experience of sensory input at the same time of interaction with the real world in AR environments can facilitate the construction of knowledge as the students are actively participating and interacting with new information during task performance or learning new concepts (Dunleavy & Dede, 2014; Sommerauer & Müller, 2018). In some cases, depending on the AR learning environment design, students can collaborate with peers, building new experiences and facilitating knowledge construction as described in the social constructivist learning theory (Cheng & Tsai, 2013).

Augmented reality scenarios allow the user to interact with the real world while having real-time feedback and virtual instructional information; trainees can apply experiential learning methods. Knowledge gained from practice or *Experiential Learning* progresses in four phases of a learning cycle.

A. Completing a task (i.e., completing a surgical knot)
B. Reflecting on the task by self-criticism and external feedback
C. Recognizing areas that need improvement
D. Adding these new improvements to the task and thus restating a new learning cycle with the new and improved task (David A. Kolb; David A Kolb, 2014).

AR as an anatomy learning tool.

Before AR was used in surgical education, it was employed for teaching anatomy. Differing from textbooks and other classical methods, AR offers interactive and convenient ways of learning anatomy. Ma et al. (2016) created an interactive AR system using *magic mirrors*. It allowed in-situ visualization of a user's anatomy based on a CT dataset in real-time. In the technique of magic mirrors, the user's image is reflected on the screen (resembling a mirror), and computer-generated images are superimposed to the user's image. Ma et al. also utilized the Microsoft Kinect that contains a depth camera to track the motion and position of the user to improve the overlay accuracy of the AR images. To assess the AR magic mirrors system, seven clinicians and 72 students answered a survey. The participants approved the system with 86.1 % for educational purposes and precision of 0.96 cm of overlay precision (Ma et al., 2016).

AR simulators in surgical skill learning.

Even though augmented reality has been used in several surgical fields, AR simulators for surgical training have been introduced recently and focused mostly on MIS. At the moment, only ProMIS (Haptica, Dublin, Ireland) has been largely validated (face and construct validity) as an AR simulator for laparoscopic surgery (S. Botden, Berlage, Schijven, & Jakimowicz, 2008; S. M. Botden, S. N. Buzink, M. P. Schijven, & J. J. Jakimowicz, 2007; Broe, Ridgway, Johnson, Tierney, & Conlon, 2006; Van Sickle, McClusky III, Gallagher, & Smith, 2005). ProMIS meets the requirement of an AR system by combining the features of the real-world through an upper-body mannequin that contains visual-tracking to track the real instruments and hand movements

(Vasilelios Lahanas, Georgiou, & Loukas, 2016). The input obtained from the real world is connected to computer-generated graphics displayed on a computer screen. The ProMIS displays a vision tracking system measuring instrument and hand movement thereby identifying position, direction, and velocity of real instruments. The displayed tasks are a VR task, a real task on a tray, or combination of VR and real. The ProMIS has shown benefits for learners in orientation, dissection, and basic suturing tasks (S. M. B. I. Botden & Jakimowicz, 2009).

In neurosurgery, AR has also influenced the way residents learn. A virtual interactive presence with augmented reality (VIPAR) device was designed to allow surgeons who are remotely placed to offer real-time assistance. In this feasibility study, a local (learner) and remote (expert) station shared the same screen; each site had a camera that allowed stereoscopic streaming, and the images were displayed in a common virtual screen with selected elements of remote station in the local monitor. Additional imaging from volumetric MRI was superimposed to the learners' monitor to enhance the viewing field. This assistance was offered online via the Internet to local surgeons performing two cadaveric procedures: a pterional craniotomy and carotid endarterectomy (Shenai et al., 2011). Both dissection procedures were successfully identified. However, time lag and resolutions presented some limitations.

AR for motor skill learning.

There is great potential for AR for motor skill training and the essential mechanism by which AR benefits motor learning is multifactorial. Learners can profit the most from AR training in complex motor tasks, whereby extra information aids task completion (Sigrist, Rauter, Riener, & Wolf, 2013). Overall, one benefit the AR systems offer over other training systems is the option of realistic haptic feedback, essential for translation of skills (S. M. B. I. Botden & Jakimowicz, 2009). Botden et al., (2007) found that the users perceive the ProMIS AR laparoscopic simulator as more realistic than the LapSim VR laparoscopic simulator (Surgical Sciences) (Figure 1.2, right) in necessary skills task and suturing task, p (< 0.001). The reason for this was established as better haptic feedback provided by the AR simulator (S. M. B. I. Botden, S. N. Buzink, M. P. Schijven, & J. J. Jakimowicz, 2007). Also, several studies have demonstrated that displaying AR information into a task can aid people in learning motor skills outside healthcare (Sigrist et al., 2013). The question remains whether AR technology can assist in learning new surgical skills.

Movement coordination, built on the connection from visual input to motor output, is the foundation for skill learning (Neggers & Bekkering, 2000). AR creates an optimal visual channel, which is connected to motion output, facilitating motion learning. This visual-motor association may be the reason why AR could surpass VR for teaching motor skills. Nevertheless, AR may have a potential in the motor training realm. The full effect of AR in visual-motor coordination is still not fully understood. Another aspect of AR in learning motor skills such as those in surgical procedures is the frequency of feedback provided to the learner. Even though AR guides in real-time, providing instructions too often can hinder the performance and learning experience of the students (Wulf & Shea, 2002). In this thesis, we estimate that with the addition of eye tracking in AR, we can facilitate the display of the instructions at the proper moment.

AR as a tool for skills assessment.

An essential step for validation of a tool in skills learning is the differentiation between novices and experts. Lahanas et al. (2015) combined AR and a training box to assess two groups of experienced surgeons and trainees; each group was composed of 10 participants. The devices used were a training box, electromagnetic sensors to track the position of the tool, a camera that captured the real scene (set inside the training box) and a computer screen to display the combination of virtual object images and the 2D capture of the setting. Researchers found that expertise can be identified using this AR simulator. The experts performed the three tasks: instrument navigation, peg transfer, and clipping more efficiently and successfully compared to the novice group ($p < 0.01$) (Vasileios Lahanas, Loukas, Smailis, & Georgiou, 2015).

Ritter et al. (2007) found that the level of expertise can be discriminated using the ProMIS AR simulator. They assessed the performance of the peg transfer task using the FLS score, instrument path length, and smoothness. Differences that corresponded to the level of expertise were found in three groups: novices, intermediates, and experts ($p <0.001$) (Ritter, Kindelan, Michael, Pimentel, & Bowyer, 2007). It shows that Augmented Reality simulators can successfully differentiate between novices and experts in the MIS task.

Summary of the AR section.

Augmented reality-based learning settings offer essential advantages to the users during the learning skill process, from learning anatomy to becoming competent in complex surgical psychomotor skills by breaking down complex procedures into parts to facilitate automation (S.

M. B. I. Botden & Jakimowicz, 2009; Khor et al., 2016). Also, AR promotes engagement during learning through vivid experiences as it combines virtual elements into the real world. Even though augmented reality scenarios seem to have valuable advantages in learning that could surpass VR, only ProMIS has been considered a valid AR skill learning system. The need for research in AR as a learning tool is essential, especially with the introduction of more versatile wearable devices in the market such as HoloLens (Microsoft) and Meta 2 (Meta Company, California, USA). These new devices offer hands-free navigation into mixed reality. However, the small field of view of these devices limits their use in education. As a result, the implementation of other technologies such as eye-tracking to enhance the power of AR as a learning tool needs to be explored further.

2.4 Eye Tracking

The study of eye movements started in the 1800s. In 1879, Louis Emile Javal noticed by the naked-eye observation that readers did not move their eyes smoothly through the text. Instead, they performed rapid movements and short pauses (Wade, 2010). The direct observation of eye movements was used to assess the reading process and continued until 1901 when Dodge and Cline built a non-intrusive eye tracking that used the technique of corneal reflection. In 1905, Judd, McAllister, and Steel employed motion picture photography to record the eye movements in its temporal aspects (R. J. K. Jacob & Karn, 2003). Advances in these two techniques (corneal reflection and motion picture capture) and their combination allowed them to develop the method that is used in modern eye trackers (Wade, 2010).

In early research, reading was the only area of applied eye tracking, but this started to change as the understanding of the relation between eye movements and cognitive processes increased. In 1967, the Russian psychologist Alfred Yarbus was the first researcher to elicit the underlying factors of eye movements. He performed an experiment where a single participant was asked to look at a painting several times, but each time Yarbus asked a different question to the observer (Figure 2.2). Yarbus found that the eye movements depended on the interest in the stimulus and the task given to the observer (Yarbus, 1967). Later, in 1980, the eye-mind hypothesis was introduced, strengthening the concept of the connection between eye movements and cognitive process.

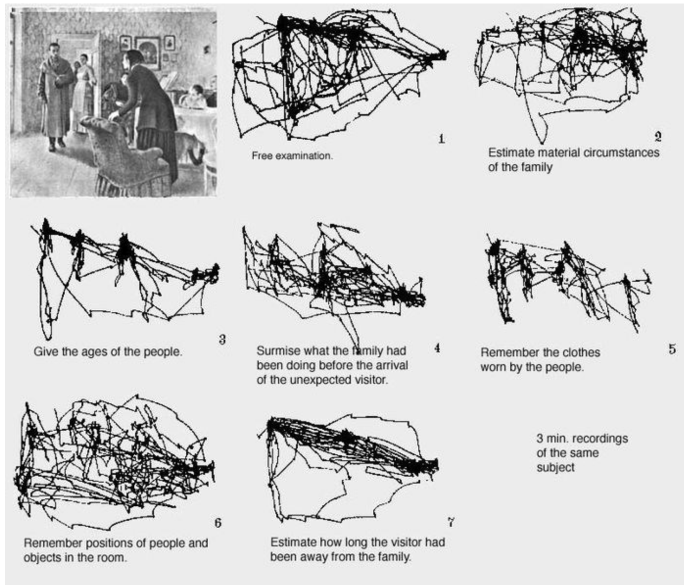

Figure 2.2. Visualization of the gaze path in Yarbus' experiment (1967). Retrieved from Wikipedia under the public domain.
https://commons.wikimedia.org/wiki/File:Yarbus_The_Visitor.jpg

The eye-mind hypothesis establishes a relationship between where the eye is looking and a person's mental processing, attention, and cognitive state (Duchowski, 2007; Just & Carpenter, 1980). In other words, people look at objects that are interesting for them, and the cognitive process (e.g., thinking) is produced regarding that object while their gaze is placed on it. An essential aspect of this hypothesis is that the object needs to be relevant to the observer while they look at it. The correlation between eye and mind processes has led to the study of the eye and its activity, precisely eye movements. The research of eye movements can be done via direct or mechanical observation (Olk & Kappas, 2011). Through mechanical observation, researchers obtain accurate objective and unbiased data about eye movements to use in quantitative studies. Eye tracking has been applied to many fields, including neuroscience, psychology, human factors, marketing, and human-computer interaction (Duchowski, 2007).

Eye tracking metrics.

The primary metrics employed in eye tracking are fixations, saccades (see Figure 2.3, left), and smooth pursuit. Fixations are the eye movements fixed to a stationary object; saccades are the rapid movements from one fixation to another; and smooth pursuit is the eye movements that follow a moving object (Duchowski, 2017). Fixations are commonly associated with visual attention and active cognitive processing while saccades is associated with fatigue (Di Stasi et al., 2014). Eye tracking metrics can be analyzed by their characteristics, such as the amplitude and direction of saccades or the number of fixations. The eye tracking metrics are analyzed based on areas of interest (AOIs) that the researcher establishes before the data collection (see Figure 2.3, right) (Blascheck et al., 2014; Duchowski, 2007). Also, there are two other components of eye tracking data that are often analyzed: gaze paths and dwell time. Gaze paths are defined as a group of fixations and saccades, and dwell time is the amount of time a person allocates several fixations on an object (Holmqvist et al., 2011). Lately, pupil diameter changes have been widely employed to analyze cognitive load (Krejtz, Duchowski, Niedzielska, Biele, & Krejtz, 2018; Mathôt, 2018).The importance of the use of eye tracking is that its metrics are related to the cognitive and psychological process and attention.

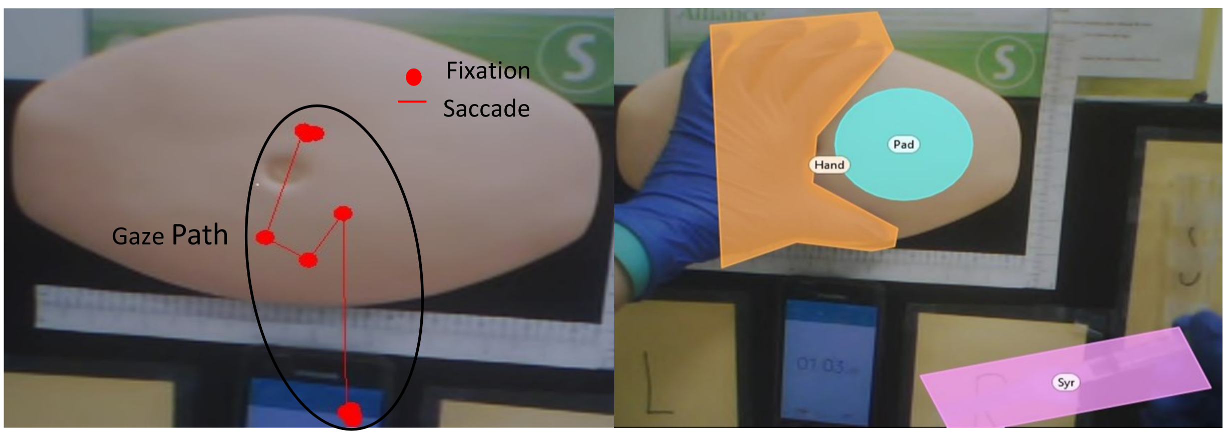

Figure 2.3. Left, representation of fixations (red spots), saccades (the line between two fixations), and gaze path (a group of fixations and saccades). Right: an example of AOIs selected by the researcher to facilitate the data analysis. Pictures were retrieved from Sanchez, YP archived (2017).

Eye tracking technology.

To consider eye tracking as a research method in quantitative studies, we need to be familiar with eye-tracking technology in order to foresee possible limitations and challenges. There are two components of eye tracking technology, hardware (eye trackers) and software. Eye trackers are divided into two categories Head-Mounted (HMD) and Table-Mounted (TMD) displays (Duchowski, 2017). HMD could be tethered or wireless. I have worked with Tobii eyeglasses and the Pupil Lab eye trackers. These devices can either be connected to a smartphone or a hand-held recording box facilitating their use in field research projects. However, the rate of data recording is slow, and the calibration may be affected by external factors such as environment light and the use of prescribed eyeglasses. On the other hand, TMD is preferred in lab-based research because of the high rate of data collection and accuracy.

Most of the eye trackers use video-based pupil and corneal reflection tracking to track the eye movements. The eye tracker shines an infra-red light to the eye to produce the reflection on the user's eye. The pupil reflection alone could be enough to track eye movements, but the corneal reflection is used as a reference to compensate for small movements of the head (Holmqvist et al. 2011). After the eye tracker detects the reflections, geometrical calculations are done using software that provides the x and y —in some cases z— positions of the eye with respect to the stimuli (Holmqvist et al. 2011). Depending on the algorithm including in the software, several types of data visualization are accessible to the researcher. Blaschechk et al. (2014) describe two methods to visualize eye tracking data: point-based and AOI-based. The former is the eye movement in its temporal and spatial location, and the last is the representation of gaze transition among AOIs (Blascheck et al., 2014). Also, the software provides eye tracking data on file type excel for statistical analysis and statistical graphics.

2.5 Eye Tracking in Surgical Education

Eye behaviour is paramount to study decision making in human beings and to understand the process of human perception. In the surgical field, eye tracking has been used to assess expertise, cognitive workload, vigilance, and fatigue in the operating room.

Eye-hand coordination.

Eye tracking can evaluate eye-hand coordination, and the mastering of this skill is a distinction of expertise in surgery. When humans perform movements, a synchronization of both hands and the movement's destination is needed. Since eye movements are faster than the hand, the target's visual information acquired reaches the brain faster and direct the overall direction of the motor system. This eye-hand coordination is not necessarily precise but oversees the target's zone, creating a proactive gaze behaviour that is task-specific. The promptness and precision of the eye fixation on the target will determine the success of the movement (Neggers & Bekkering, 2000).

With practice, eye-hand coordination can be influenced by proprioception. Experts utilize fewer visual fixations than novices (B. Law, Atkins, Kirkpatrick, & Lomax, 2004; Richstone et al., 2010), possibly by guiding movement with eye saccades complemented by proprioceptive feedback. Eye movements are closely related to the motor task. During task performance, eye movements lead the sequence of the actions. Land and Hayhoe (2001) classified the fixations involved in a task in four groups: *Localization-fixations* identify the objects used for the activity, *Direction-fixations* are hand movements preceded by a fixation, *Guidance-fixations* occur during the interaction of two or more objects, and *Checking-fixations* confirm an action is complete. Another concept in the eye-hand coordination is the *Quiet Eye* (QE) defined as the last fixation on the target before a hand movement involved in the intended task (Vickers, 1996). More extended QE is associated with expertise, and as a result QE is utilized as a tool for gaze training in surgical procedures (S. J. Vine, R. S. Masters, J. S. McGrath, E. Bright, & M. R. Wilson, 2012; Wilson et al., 2011). For instance, QE training showed an improvement in the effectiveness and efficiency of performance during a knot-tying task (Causer, Vickers, Snelgrove, Arsenault, & Harvey, 2014). Similarly, Wilson et al. (2011) found that verbally instructing trainees about keeping a target-locking strategy (longer QE) improved their performance in a laparoscopic task compared to the students receiving video feedback about their movement performance.

Workload assessment via eye metrics.

As eye tracking input is linked to attention and cognitive process, eye metrics have been utilized as a predictor of cognitive workload. Kodappully et al. (2016) demonstrated that gaze patterns are related to the process of identification and response to anomalous events. Also, eye tracking can work as a performance monitor and as a training tool (Kodappully, Srinivasan, & Srinivasan, 2016). In similar studies, the data obtained from eye tracking helped to identify gaze patterns whereby individuals prone to errors have the tendency to be distracted easily by the environmental stimulus (Larson & Perry, 1999), whereas individuals that identify errors have usually organized scan patterns and fewer distractions (Marquard et al., 2011).

Many elements can increase a surgeon's mental workload, for instance, the high volume of working hours, the complexity of the surgery, and a critical step during the procedure (Zheng et al., 2012). Previous research demonstrated that a higher level of cognitive workload was present in novices whereby an increased gaze entropy is noticeable in the AOI's compared to experts. Experts would encounter less cognitive workload associated with them performing the task several times before finding it less demanding (Tien et al., 2015). Another helpful eye metric that measures mental workload is blinking. Increased frequency and duration of eye blinking during a surgical task is associated with less cognitive workload (Berguer, Smith, & Chung, 2001; Zheng et al., 2012). Pupil size changes are a more recent eye tracking metric employed to assess cognitive load.

Jiang, Atkins, Tien, Zheng, and Bednarik (2014) found that an increase in pupil diameter is related to high task difficulty, which has more cognitive load. Conversely, less concentration demand is represented by less pupil and gaze entropy. In this way, cognitive workload relates to confidence in performance. Experts possess wired coordination of movement that blooms from countless times practicing, thus decreasing mental workload (Zheng, Cassera, Martinec, Spaun, & Swanström, 2010).This decreased workload may be presented not only in less attentional switches and increased blinking but also in requiring less time to perform the task.

Vigilance in the operating room.

Eye tracking has been used to measure vigilance in the operating room. Zheng et al. (2011) designed an experiment where the vigilance of surgeons and novices was assessed during surgery on an unstable patient. They found that experienced surgeons paid more attention to the vital signs of the patient compared to learners (Zheng et al., 2011). Also, saccades have been associated with fatigue. Di Stasi et al. (2014) measured the eye movements of 12 surgical residents before and after a 24-hour call day and found that longer saccades were related to fatigue, but it did not affect the performance (Di Stasi et al., 2014).

Skill evaluation and gaze training.

Eye tracking has emerged as a leading technology for assessment in surgery. Gaze analysis identifies differences in surgical training, contrasting inexperienced from skilled surgeons and skill evaluation (Hermens, Flin, & Ahmed, 2013); thus, gaze training has the potential to improve motor skill (S. J. Vine, R. S. W. Masters, J. S. McGrath, E. Bright, & M. R. Wilson, 2012). Different research groups have considered what causes the difference in eye behaviour between different levels of experience. In their study, Law et al., (2004) elucidated the tool following behaviour of typical novices during the task, and target focused gaze pattern of experts while attempting to touch a cube with a laparoscopic instrument in a virtual environment (B. Law, Stella Atkins, Kirkpatrick, & Lomax, 2004)

Based on this notion, computational categorization approaches have been attempted to rank performance. Hidden Markov Models (HMM) compute the probability that a given item may repeat itself in a sequence of labels; in other words, a probabilistic sequence model, and are widely employed for language recognition (Eddy, 2004). A research group utilized HMM to assess the skill level of experienced versus novices in a surgical navigational task. Although the experiment was not successful in differentiating the two teams, possibly given to the fact that they utilized still images instead of a live video, the HHM separated high-performance individuals from those with low performance exclusively by eye behaviour (Sodergren, Orihuela-Espina, Clark, Darzi, & Yang, 2010).

Richstone et al. (2010) utilized eye-tracking information to correlate eye behaviour with the level of expertise. In this study, 21 surgeons performed in a simulated and a live surgical scenario utilizing Linear Discriminant Analysis (LDA) and nonlinear Neural Network Analysis (NNA) to analyze and differentiate skill levels. Moreover, LDA and NNA identified experts and non-experts with 91.9% and 92.9% correctly. Similarly, in the live surgery environment, it recognized experts and non-experts 81% and 90.7%, respectively (Richstone et al., 2010).

Training eye behaviour to improve performance has been favourable in laparoscopic tasks (Wilson et al., 2011). In one study, a group of novices trained in a laparoscopic task under two different environments: self-learning and gaze training. Although in both teams, all trainees decreased the number of errors and task duration, it was more significant for those who received gaze training (Samuel J. Vine et al., 2012). The primary rationale behind the effectiveness of gaze training for skill acquisition is the quiet eye. Quite-eye is the last fixation before a pivotal motion during a performance (Vine, Moore, & Wilson, 2014). This fixation is shorter for novices and it is trained via verbal feedback or gaze guidance (Samuel J. Vine et al., 2012).

2.6 AR and Eye Tracking

Currently, research on mobile eye tracking and AR is quite limited. Besides, several publications on this topic in our search related to other areas not associated with medical training. We believe applying eye tracking in AR may offer opportunities for active research more realistic to typical surgical training environments. Commercially, eye-tracking technologies can be added to wearables AR devices such as Google glasses, Epson Moverio BT-200, and HoloLens. A good real-world example of the benefits of AR is in the work of Barakonyi et al., (2007) In this experiment, researchers implemented a fixed eye tracking device into a screen-based augmented reality to facilitate the interaction in a remote video conference. This pilot study provided the concept of eye tracking as a tool in augmented reality environments where two users can interact specifically within an educative context (Barakonyi, Prendinger, Schmalstieg, & Ishizuka, 2007).

There is ongoing research applying eye-tracking data in order to predict actions and improve the human-robot interaction in robotic assistive devices (Admoni & Siddhartha, 2016). Studies employing eye tracking as an input signal to display information could increase in number. Toyama et al. (2013) proposed a novel system to assist readers using real-time eye tracking data: a document retrieval method and a wearable AR device. In this experiment, the researchers used the eye-tracking technology to identify the moment the reader was close to a previously identified keyword that could require the display of additional information to guide the reader through the document. The augmented information was displayed on a monocular HMD (Brother Airscouter) screen, and users could select the information needed through their gaze (Toyama, Dengel, Suzuki, & Kise, 2013). Another study explored the use of eye tracking in virtual environments to improve the retrieval of information using gaze position, to decrease the clustering of information and to improve the interaction in the virtual world. Even though the researchers in this later study used a virtual reality headset (Oculus) with and add-on eye tracking device (Tobii Studio), it provides insight of how eye tracking can alleviate the congestion of information in virtual and possible mixed realities (McNamara, Boyd, Oh, Sharpe, & Suther, 2018).

AR and eye tracking as a guidance tool to improve performance.

AR technology can improve performance since it provides information live-stream, mainly, during a manual task with an HMD allowing a user to be hands-free. In a 75 participant study, Tang et al., (2003) compared four types of media to display instructions during an assembly task: “printed media,” “computer-assisted instruction (CAI) on an LCD monitor,” “CAI on a see-through HMD,” and “spatially registered AR.” The authors found that AR users improved the accuracy of the task by reducing errors and the workload when assessed through the NASA TLX score. However, the performance time was not significantly different among the different media used. Also, the researchers found that one limitation of AR utilization to display information is the size of the field of view (Tang, Owen, Biocca, & Mou, 2003). Renner and Pfeiffer (2017) tried to tackle this issue by employing peripheral vision and eye tracking data input for guidance. This last research group reported eye tracking was relevant when the information was displayed in advance to find objects (Renner & Pfeiffer, 2017).

AR can be used preoperative and intraoperative to guide surgeons. Stoyanov et al. (2008) designed a gaze-contingent AR system to provide effective and real-time surgical guidance during MIS procedures. Researchers used fixations to determine the areas of interest for the surgeons during procedures, including soft tissue location, to extract 3D information and provide the corresponding feedback (Stoyanov, Mylonas, Lerotic, Chung, & Yang, 2008).

Potential challenges.

Some authors have employed AR technology to enhance the experience during task performance (Vasileios Lahanas et al., 2015; Sigrist et al., 2013). Unfortunately, AR glasses have a limited field of view that might interfere with the user-environment interaction; hence, task performance could be affected. Renner et al. (2017) implemented the eye gaze to introduce flickering as a clue for guidance through the AR. However, it showed that there was only a benefit if the information was in advance (Renner & Pfeiffer, 2017). Mobile eye tracking is the best method for studying mixed reality environments, but its analysis can be burdensome. Some of the challenges faced in this type of analysis are fixation times in AOI’s, recording individual video frames during gaze resting, and high workloads of data can sum up to the deterrents for eye tracking analysis (Meißner, Pfeiffer, Pfeiffer, & Oppewal, 2017).

Another challenge with gaze-controlled systems is to distinguish between gathering visual information and the gaze intended to activate commands. There is a definite issue when the users enable a command once moving their eyes known as “The Midas Touch Problem” (R. J. K. Jacob, 1995). Midas Touch problem is defined as triggering the information or activating command intentionally through gaze data (R. J. Jacob, 1993). For example, individuals naturally scanned the environment even if they are not interested in something specific. However, in environments where eye tracking is implemented as visual feedback to activate commands or display information, it is paramount to differentiate between relevant gaze and just a simple gaze scanning.

Several attempts have been made to correct this issue, such as duration of dwelling time and blink, but are found too distracting during task performance.

One surprising finding with applying AR in the surgical field was the presence of inattentional blindness during a task. Introducing virtual information during real surgical procedures can cause a phenomenon of “attentional tunneling,” where the participants are not able to recognize the virtual objects (Dixon et al., 2013).

Chapter 3 Development and Evaluation of a Chest Tube Insertion Model

To fulfill the goals of this thesis, first, I needed to build a simulator of a surgical multi-step procedure. A chest tube insertion was chosen because it is a common and vital procedure learned by students across several medical specialties. A completion of chest tube insertion included about twenty steps. Students need to perform correctly all the steps sequentially which gives us a good chance to capture the moment of performance difficulty.

There is not a standard model to learn and practice a chest tube insertion. Most of the time, this procedure is performed on animal models, cadavers, or on the popular simulator model used for Advance Trauma Life Support (ATLS) courses (Trauma-man). These models are expensive to use in research projects. As a result, I designed and developed a model suitable for research purposes. In this chapter, I discuss the construction, assessment, and implementation of this chest tube insertion model.

3.1 Introduction

Most surgical procedures consist of multiple steps. Proper completion of these steps is fundamental to deem the procedure successful. Besides, most surgical procedures are performed under high acuity and have risk conditions, making them high-stake procedures necessary to save a patient's life. Such is the case of a chest tube insertion or placement of a tube into the chest cavity. A chest tube removes air, and fluid, such as blood and pus, to recover or optimize a person's breathing ability under different circumstances. Although chest tube insertion is essential to preserve life, it can also be hazardous to the patient if performed incorrectly (Hernandez, Zeb, Heller, Zielinski, & Aho, 2017), increasing the procedure complications up to 10% (Ariffin, 2018). These complications derive from patients' factors or trainees' lack of knowledge and dexterity and can result in issues with tube insertion or positioning (Dev, Nascimiento Jr, Simone, & Chien, 2007; Kashani, Harati, Shirafkan, Amirbeigi, & Hatamabadi, 2017). Not surprisingly, the performer's complication rate varies if completed by a trainee or a specialist (Kashani et al., 2017). There is no current literature about the number of complications among Canadian trainees. However, Ball et al., in 2007, performed research about complications of chest tubes placed by residents from different specialties and found that from the 61 patients, there were 17 complications such as insertional (35%), positional (53%) and ineffectiveness (12%). Also, they found that the rates of complications varied according to the training discipline. Specifically, most of the complications were mainly in emergency medicine and other surgical disciplines (40 % and 25% respectively). On the other hand, general surgery residents had 7% complications and internal and family medicine residents had a 13% complication rate (Ball et al., 2007). Therefore, optimizing training of novices is paramount to preparing them to perform surgical procedures, such as chest tube insertion, and thereby, reduce the possibility of complications.

3.1.1 Chest tube insertion models.

Several simulation models have been employed such as animal models, hybrid models (synthetic and animal components), and synthetic simulators (mannequins). Most of the hybrid models are composed of a synthetic thorax or supportive structure and a piece of either chicken or pig meat to simulate the soft tissues of the rib cage (Ghazali, Breque, Léger, Scépi, & Oriot, 2015; Tatli et al., 2017). Even though these Do-It-Yourself (DIY) models are affordable, they have several drawbacks. These models lack external feedback for the direction of the tube during insertion. For example, chest tubes need to be inserted upwards and posterior in clinical cases of pneumothorax (air between the lungs and rib cage), but downwards in the cases of hemothorax (blood between lungs and rib cage). Also, these models do not effectively represent human anatomy landmarks, such as the axillar lines and the nipple.

On the other hand, mannequin simulators can provide better anatomical references and feedback. Some studies have shown that mannequin simulators have comparable results to animal models and cadavers (Ali, Sorvari, & Pandya, 2012; Chung, Kim, You, & Chung, 2016; Hishikawa et al., 2010). Also, synthetic simulators mitigate the ethical implications of employing animal or hybrid models. There are several commercially available models such as Surgical Trauma Training Manikin (SimuLab Seattle, WA, USA), Chest Drain Simulator (3B Scientific, Budapest, Hungary), Chest Tube Trainer (Trucorp, Belfast, Northern Ireland) or just pieces of skin pads such as the Chest Tube Insertion Modules (Laerdal, Stavanger, Norway). However, the model most used is the Trauma-Man (SimuLab), as this model is used to teach ATLS courses.

Despite the benefits of chest tube insertion simulators, the main limitation for using them is the cost.

Therefore, we needed to design a chest tube model capable of offering a realistic environment to the students to learn the vital steps that are associated with complications, such as identification of the area of insertion and direction of the tube, and at the same time being affordable and accessible for multiuse. I describe the design and construction of a simulator where the trainees can learn the basic steps and complete a chest tube insertion.

3.2 Methodology

This section is composed of three parts. First, I describe the design and construction of the model. Second, the feedback received from experts to validate the model is discussed. Lastly, the implementation of the chest tube is performed.

3.2.1 Construction of the model.

The main components of the initial chest tube model are a torso, pleura, and a silicone-based skin pad.

Torso and pleura.

We purchased a standard endurable plastic male torso mannequin (Eddie’s Hang-Up Display Ltd, Canada) with measurements of 55 cm length, 48 cm shoulder to shoulder and 33cm wide on the chest. Part of the right lateral wall of the torso was modified to recreate three average male ribs and their corresponding intercostal spaces. Sticks were glued to the back of each rib to give support to the ribs. Then, the torso was painted to simulate the skin color. Also, to recreate the thoracic inner layer close to the ribs called parietal pleura, a thin plastic layer was placed inside the mannequin’s rib cage (see Figure 3.1). This step was important as we wanted to create the sensation of penetrating the parietal pleural during a chest tube insertion. This plastic layer was replaced in every chest tube placement trial.

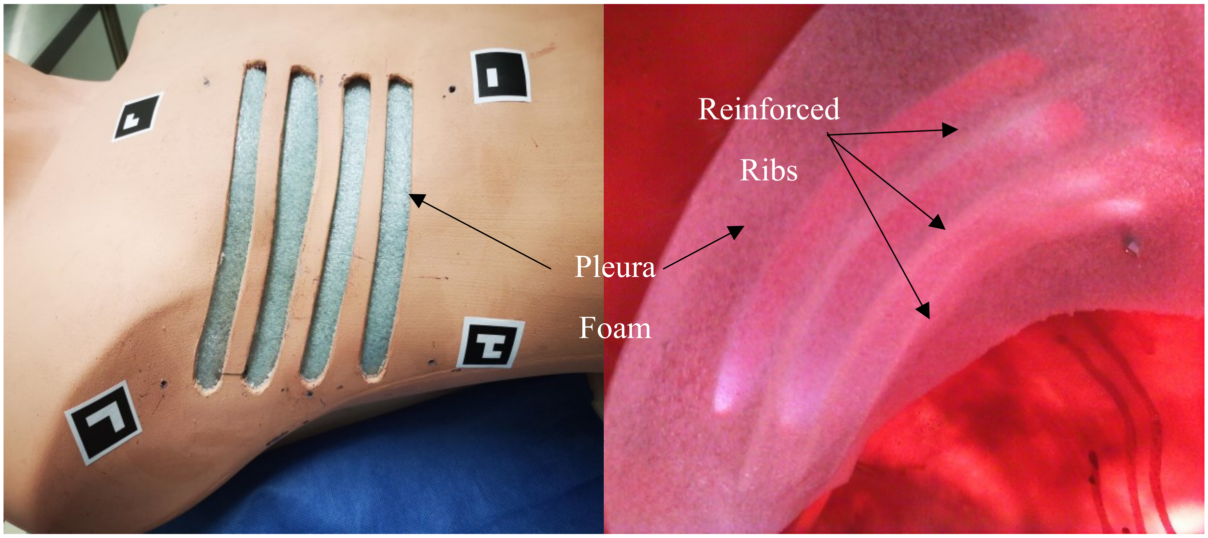

Figure 3.1. Left: visualization of the rib cage and the layer of foam. Right: internal view of the reinforced ribs and the foam lining to simulate the parietal pleura.

Skin Pad.

Several 16 cm x 18 cm skin pads were created using customized known materials in the world of simulation (silicones) that replicate human skin sensation and resistance. The skin pads contain three layers: skin, fat and two-layers of muscle. The silicones utilized were the Ecoflex 30 (Smooth-On Inc, Easton, PA) for the skin and fat layers and Dragon Skin, which is a harder silicone for the muscle layers. A piece of red mesh was located between the two layers of muscle to provide more structure between the layers. The silicones come in a clear color, so acrylic paint was used to provide the corresponding color of the layers. Each skin pad was made using a flat wooden mold covered by a texturized vinyl adhesive paper to recreate the texture of human skin. A power mesh was attached to the wooden mold to give support and structure to the pads (see Figure 3.2). Two to three skin pads were created at the same time, and it took around 24h for each pad to dry. On the finalized skin pad, a silicone nipple was attached at the level of the 5th intercostal space (Figure 3.4). A skin pad can be used up to three times, changing the insertion site of the chest tube.

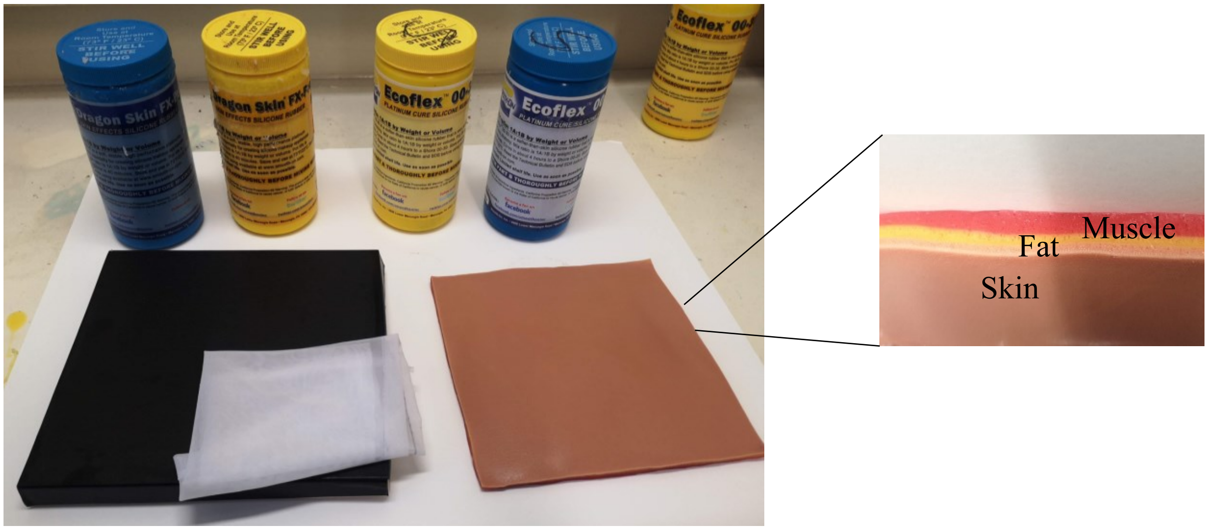

Figure 3.2. Left: materials to make a skin pad, silicones: Ecoflex 30 and Dragon Skin (Smooth-On Inc, Easton, PA), a wooden base, and a power mesh. Right: three layers of a finished skin pad (skin, fat, and two-layers of muscle).

3.2.2 Content validity of the model.

To validate the model (content validity), three experts (2 men and 1 woman, age mean: 33 years) defined as healthcare providers who have performed 30 or more chest tube insertion procedures, volunteered to perform a chest tube insertion and provide written feedback. Before the experts carried out the procedure, they read the research project information and signed a written consent form. The written feedback form had two major components. First, using a 5-point Likert scale, the experts ranked the similarity of the model to a real patient (1 = Not very similar, and 5 = Very similar), and its usefulness for novices to practice a chest tube insertion in this model (1 = Not useful at all, and 5 = Very useful). In the second component, experts provided written comments or recommendations to improve the model. The written feedback form can be found in Appendix A in this thesis. The performance of the experts was timed and recorded using a room camera and the world camera of a wearable eye-tracker device (Pupil Labs. Berlin, Germany). The data from the eye tracker was not analyzed, but the video of the world camera was used to make the explanatory video for the novices.

The completion time was 8.5 ± 1.06 minutes, and the initial model received good feedback in each one of the items assessed. All three experts agreed on the usefulness of the model to be used to teach novices how to perform a chest tube insertion. The results of the first component are described in Table 3.1.

Table 3.1.

Experts' written feedback for the initial chest tube insertion model.

ITEM	**Mean**	**SD**
Anatomical Landmarks	3.7	1.53
Skin Incision	3.7	0.58
Dissection of Tissues	3.7	1.53
Identification Pleural Space	3.7	1.53
Tube Insertion	3.7	1.53
Chest Tube Fixation	4.7	0.58
Drainage Connection	4.7	0.58
Overall	4.0	0.00
Usefulness	4.3	0.58

Regarding the second part of the written feedback, experts made the following recommendations to improve the model:

"Benefit from [a] balloon inside the cavity to mimic the presence of lungs."

"Needs a fixation point near the tube insertion site."

"[The] main difficulty is the dissection of tissue."

"…the skin needs to be improved…. the surface of the skin is a little harder" "the skin of the tissue of the model is too dry for me to insert the tube…. a real human has blood and fat which will lubricate the tube".

3.2.3 Modifications.

Based on the comments, the model was improved by adding a synthetic lung, better fixation of the pad to the torso, and constant and better quality among the skin pads to address the issues of difficult dissection and skin quality.

Synthetic Lung.

During the insertion of a chest tube, the feedback of the correct direction of the tube is, in part, given by the resistance of the lung when the tube is in the chest cavity. Therefore, having a synthetic lung inside the mannequin is paramount to providing more realistic and effective feedback to the trainee. The synthetic lung was made of fabric and filled with foam that measured 22 cm in length, 9cm at its base, 8cm on the top, and 13cm wide (Figure 3.3). The synthetic lung was fixed to the posterior wall of the mannequin using double-sided tape.

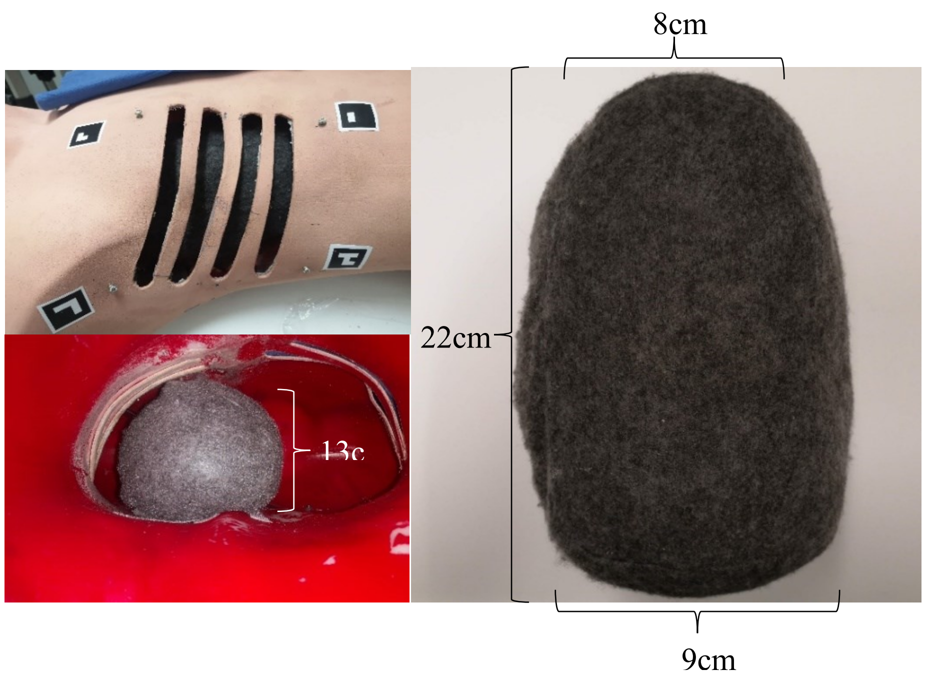

Figure 3.3. Right: the synthetic model with its measurements (length: 22cm, top: 8cm, and bottom 9cm). Left top: external visualization of the lung. Left bottom: internal visualization of the lung in the chest cavity, including its width measurement of 13cm.

Improvement of the skin pad quality and its fixation to the torso.

The silicones are versatile and easy to obtain for the construction of simulated skin pads. However, several factors can affect the silicones such as temperature, amount of product, and time for layering and drying. A very detailed instructional guide was created to obtain consistent quality among skin pads to be used in each simulation. An expert mentioned that the dissection of the tissue was difficult. This was corrected by removing the mesh between the two layers of muscle as it becomes rigid after the contact with the silicone, and thereby impedes the dissection. Besides, the fixation of the skin pad to the mannequin was improved by using five fixation points, and the nipple was fixed using non-absorbable surgical suture. The final model and set-up are pictured in Figure 3.4.

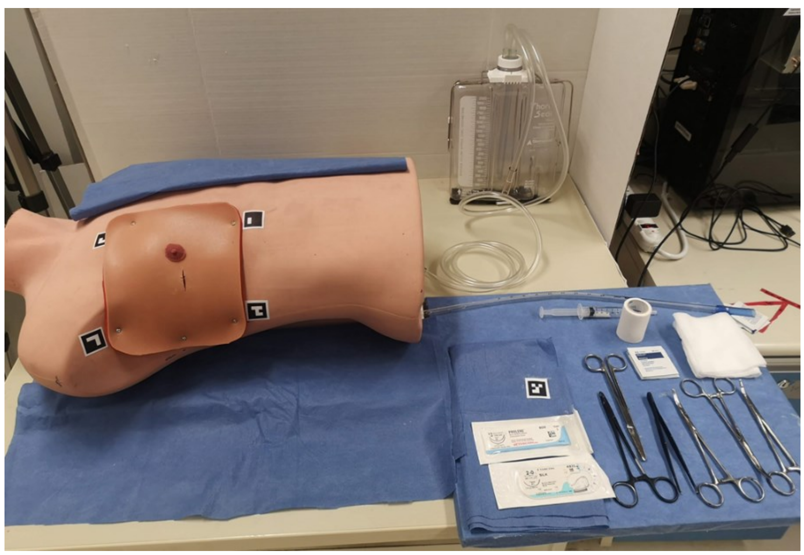

Figure 3.4. Chest tube insertion scenario: simulated human thorax with a skin pad attached, which includes a nipple, surgical instruments, and a drainage system.

3.3 Implementation of the Chest Tube Model

After changes were made on the model based on the experts' feedback, first-year medical students performed a complete chest tube insertion at the Surgical Simulation Research Lab located at the Heritage Medical Research Centre at the University of Alberta.

3.3.1 Participants.

The data collection was performed for three months, having an initial target of 40 medical students as participants in order to have a considerable group size and to foresee possible dropouts. However, only twenty medical students (50 % female, 95% right-handed, and age 24 ± 2.7 years) without surgical experience or muscular deficit were recruited. Medical students were in their first or second year of medical school, allowing for measurement of performance unaffected by prior clinical experience.

3.3.2 Performance measures.

Efficiency was measured through time to complete the task. Effectiveness was measured by the number of errors committed during the chest tube insertion. I employed the Chest Tube Insertion Performance Assessment Scale developed by Ghazali et al., (2016). This scale has a maximum score of 20, and it was validated to be applied to trainees working in a simulated model in a case of pneumothorax. The reliability coefficient, Cronbach alpha, was 0.747, and the intraclass coefficient was 0.966. The cut-off score of the success rate of chest tube insertion is ≥ 14, where 100% of the chest tubes are functional. This scale has an unrestricted use for assessment during simulation practices and research purposes.

3.3.3 Procedures and task.

The data collection was gathered individually. Each participant filled out an anonymous questionnaire about demographics and previous experience in surgical procedures. To maintain consistency of the instructions among the participants, a 9-minute standardized explanatory video of the task using our chest tube insertion simulator was displayed once to each participant. The video was given in the context of a chest tube insertion in a 32 years old patient diagnosed with a pneumothorax and the procedure described was based on the Advanced Trauma Life Support (ATLS) recommendations. The video footage and photos used to make the instructional video were obtained from the performance videos of the three experts and some pictures from the internet. There was no limit on time for watching the explanatory video. Participants could stop the video at any moment to read the captions or to look in more detail the graphics, but the video could not be played back. All participants received a suturing workshop at the beginning of their medical school and did not practice hand knots. Therefore, before starting the task, the trainees practiced hand knots until they made two consecutive hand knots without assistance. This step was considered necessary to ensure that all the participants were able to do the suturing step.

In this experiment, the participants wore an AR headset, HoloLens (Microsoft), with an eye-tracker device to record eye movement data. More details about the AR and eye-tracking systems are described in Chapter 4, and the eye-tracking data is analyzed in Chapter 5. In this chapter, only the videos of each session recorded using the world camera of the eye-tracker were used to calculate the duration of the task and to facilitate the completion of the chest tube score scale. Also, a room camera worked as a secondary source of information; a go-pro camera was placed inside the simulator to visualize the insertion of the tube inside the chest of the mannequin. None of the videos contain the face of the participants.

Task.

After watching the explanatory video and eye-tracker calibration, the participants were asked to perform a complete chest tube insertion using the chest tube model created and real surgical instruments, which included a 28 Fr chest tube. The tube size was chosen based on the size of the model. Also, there is no evidence that a smaller or bigger tube in chest trauma has a difference in the effectiveness of the chest tube (Tanizaki et al., 2017). As this was the first time for them to carry out this procedure, written instructions were placed within hand's reach for them to review. The main steps of the procedure were: (1) identification of landmarks, (2) antiseptic procedure, (3) local anesthesia, (4) incision, (5) dissection, (6) confirmation of the pleural space, (7) introduction of chest tube with forces clamp, (8) securing the chest tube, and (9) connection to the drainage system and wound dressing.

3.3.4 Data analysis and results.

Data were organized in an excel sheet, and descriptive analysis was carried out using SPSS 24.0. All participants were able to complete all the steps of a chest tube insertion. However, they took a longer time than expected for this procedure. The completion time was 32.98 ± 10.40 minutes. As it was expected in a group of novices, most of the participants did not pass the chest tube insertion assessment scale (Performance Scale score 13.80 ± 2.46). The rate of success calculated as *Chest Tube Score/Completion Time*100* had a mean of 46.50 ± 17.67%.

A Shapiro-Wilk's test ($p > 0.05$) and a visual inspection of the histogram showed that the results of completion time ($p = 0.357$), chest tube performance score ($p = 0.106$), and the rate of success ($p = 0.673$) were approximately normally distributed. Skewness and kurtosis values are shown in Table 3.2.

Table 3.2.

Shapiro-Wilk's test results of chest tube performance metrics.

	N	**Mean**	**Skewness**	**SE**	**Kurtosis**	**SE**	**p-value**
Completion Time (min)	20	32.98 ± 10.40	0.714	0.512	0.251	0.992	0.357
Performance scale scores	20	13.80 ± 2.46	0.115	0.512	-1.386	0.992	0.106
Success Rate (%)	20	46.49 ± 17.67	0.409	0.512	0.008	0.992	0.673

3.4 Discussion

Our main objective was to design a model sufficiently accurate for experts to think that was real enough for learners to get a sense of what a real-life procedure would feel like.

This was assessed by estimating the experts' opinion of how real the task was and associate those results with the learners' performance.

The experts in our experiment thought that our model was accurate in the feeling of the skin incision, chest tube fixation, and drainage connection. There was discordance regarding the accuracy of feeling the tube insertion, dissection of tissues, and identifying the pleural space. We used the feedback from the experts to improve our model, focusing on realistic landmarks, dissection and tube insertion steps. We achieved this by creating a silicone nipple fixed on the 5^{th} intercostal space, improving the skin pad, and placing a synthetic lung inside the chest cavity. Improving these steps was crucial to provide realistic and useful practice for the trainees as we know most of the complications related to chest tube insertions are associated with the insertion and fixation of the tube. Ideally, the experts who provided the initial feedback should have been consulted to re-assess the model. Unfortunately, this was not possible in this study because experts did not have the availability to assess the model again. However, overall, all the experts in their overall scores felt that the chest model was useful to offer learners a primer to the chest tube insertion task. This model offered the elements necessary for the medical students to become familiar with a chest tube insertion procedure. We expected that the medical students did spend more time performing the task than the experts and the results supported this prediction.

Completion time was approximately normally distributed amongst participants. The standard deviation was around 30 % of their mean value. Motivation and self-achievement drive could be a reason for the variability. During the task, we observed that some participants did not invest much time in performing a correct procedure. The reason for this could be that the task was part of a simulation task. During simulation tasks, it was expected that participants did not perceive the need for performing correctly as the tasks were not graded. However, there is no correlation between short completion time and lower performance scores. Another observation during the trials was that participants who showed more hesitation and self-doubt had longer completion times.

Nevertheless, we were able to attain a better idea of the trainees' chest tube insertion ability via the performance score. As it was expected, most participants achieved a performance scale score less than the passing threshold. The performance score seemed to be closer amongst different participants with less variability amongst them. The participants' success rate for chest tube insertion was directly influenced by their completion time since it made up part of the equation. This can be seen by the variability of their values.

Overall, it was very reassuring to see that for the experts our model was satisfactory, and it was implemented to assess the ability of a group of novices to carry out a chest tube insertion. Also, our model has several advantages over other synthetic models. It provides realistic landmarks of the human chest, feedback during dissection, and chest tube position through the synthetic lung. This model contains the necessary elements to teach the basics of a chest tube insertion procedure. As a result, the model will be used in our future research involving AR and eye-tracking devices.

Limitations.

First, this study could face the limitation of external validity as the results obtained from this study cannot be generalized to other healthcare trainees such as surgical residents. Also, the chest tube insertion model was created for researcher purposes; therefore, it limits the transferability of the findings to the current training setting (cadavers) or the clinical setting.

3.5 Conclusion

We were able to design a model that was sufficiently believed to be close to reality by experts and that accurately associated with the learners' score. Our primary motivation was to make a model affordable and self-sufficient that could be utilized by different groups to train learners and not be limited by accessibility. The work dedicated to this project can help other researchers in the construction of models for teaching surgical and medical procedures.

Chapter 4 Tracking Eye Movements in an AR System.

The purpose of this chapter is to describe the methodology employed to incorporate the eye tracking and augmented reality (AR) systems into one platform. The first section of the chapter is dedicated to describing challenges and solutions of adapting the two devices and keeping an appropriate eye tracking accuracy. In the second section, I describe steps of pairing the eye tracker to the HoloLens AR goggles.

4.1 Introduction

Augmented Reality (AR) offers textbooks, graphic images or 3D material to a user's instant perspective of the real world to produce a mixed reality and it has been used frequently to direct surgical procedures (Bernhardt et al., 2017; Darko Katić et al., 2013; Navab, Traub, Sielhorst, Feuerstein, & Bichlmeier, 2007) and to create skills training systems (Barsom et al., 2016; Bernhardt et al., 2017; Dickey et al., 2016; Pelargos et al., 2017). Many researchers agree that AR can provide surgeons with graphical guidance in real-time, help surgeons match surgical scene photos, display trajectories or cutting margins, and minimize anatomical ambiguities(Barsom et al., 2016; Bernhardt et al., 2017). All of these AR utilizations produce substantial ways of improving operational precision to maximize patient safety in the operating room.

AR can lead learners to look for advice in training centers when practicing sophisticated procedures without frequent breaks during skill coaching. The necessary instructions can be provided to the learners on their peripheral vision. This AR feature is handy when performing a multi-step procedure like delivering newborns or inserting a tube into the chest to save a life. Displaying the educational data at the right times to maximize the learning outcome without causing diversion is a critical matter when using AR for instructional presentation. A continuous display of teaching materials can hinder the self-learning ability of a novice. Some learning schemes ask the learners to signal their learning difficulty times via a hand or speech demand to safeguard the natural learning method (Eid, Giakoumidis, & El Saddik, 2016). Displaying unnecessary information or making hand gestures to enable information on the AR screen are not the most ideal techniques because these cause interruptions of the task. Other researchers used object surveillance to define the principal moments of a procedure and to build a context-based

AR system for displaying an ongoing instructional message at separate learning stages (Esposito et al., 2015; D Katić et al., 2010; Darko Katić et al., 2013; Navab et al., 2007).

It is vital to provide useful instructional information on AR scenarios to avoid distraction and extraneous load to the user. In this research, with our new AR platform, we aim to recognize the vital, challenging moments of a learning task depending on the user's eye behaviour. The AR system should automatically display the necessary instruction once the opportunity is detected. This detection will allow students to concentrate on their practice without distraction and disruption.

To achieve this goal, we added an eye tracker to a pair of AR goggles. Eye tracking technology provides useful data on eye behaviours, revealing cognitive information, including human mental workload (Kosch, Hassib, Buschek, & Schmidt, 2018), attention shifts (Steil, Müller, Sugano, & Bulling, 2018), and visual search behaviour. Eye tracking has been used in medical simulation studies (Kottayil, Bogdanova, Cheng, Zheng, & Basu, 2016; B. H. Y. Law et al., 2018; Patel & Zurca, 2018) to produce essential data to portray the surgeon's practices over different stages of their skill. With the growth of the computational algorithms, eye tracking can also be used as an input signal to guide medical devices such as an endoscope (Langbehn, Steinicke, Lappe, Welch, & Bruder, 2018).

Adding eye trackers to the AR scheme goes hand in hand with several difficulties, varying from hardware assembly and system calibration to data processing. In this part of my thesis, we add the Pupil Lab Eye-tracker to the HoloLens and calibrate the AR-Eye tracking model to acquire precise eye monitoring signals and attempt to find the time of learning difficulty using eye monitoring information.

4.2 Augmented Reality and Eye Tracker Hardware

As AR is blooming and its uses have been extended to the research and medical field, the possibility of improving this technology has lured researchers and companies by implementing eye tracking into AR. For example, researchers have used eye tracking to enhance the signalling and selection of objects in the real world to embed computer-generated messages into them (B et al., 2016; Kyt et al., 2018; Renner & Pfeiffer, 2017). HoloLens (Microsoft, U.S.), which is the most robust augmented reality device commercially available was used in our research. At the moment of this project, the HoloLens did not include eye tracking to gather data; however, Pupil Lab (Berlin, Germany) has launched an accessible and open source binocular add-on eye tracker for HoloLens.

HoloLens.

Microsoft HoloLens is a mixed reality instrument for displaying virtual information on the screen of the headset. This 580 g device with a field of view of 30°×17.5° degrees has a 32-bit Intel Atom processor and Microsoft's custom Holographic Processing Unit (HPU), which creates a 3D version of the environment using four spatial-mapping cameras and a depth camera (Furlan, 2016). Users interact with the apps and holograms by voice commands, head gaze, and hand gestures. Voice commands are given to a Microsoft virtual assistant, Cortana. HoloLens uses the position and orientation of the head to determine a head gaze vector to create a laser pointer seen on the screen to interact with the holograms. There are two main hand gestures, air tap and bloom; and three composite gestures, tap and hold, manipulation, and navigation (Zeller, 23 Feb 2019). We substituted these gesture interactions by gaze location, ground detection, and pupil information obtained from the eye tracker.

Pupil Lab add-ons eye tracker.

Pupil Lab Eye-tracker devices with a collection of add-on components can be readily mounted on HoloLens. The eye tracker contains three cameras, one for world perspective (world camera) and the other two for eye movement monitoring (eye cameras). The globe camera has a 100-degree FOV wide-angle lens that generates images with a resolution of 1280 ~720 pixels and 30 frames per second. The eye cameras capture data at a maximum 200hz with a resolution of 192 or 192 pixels. In this project we set up the eye camera video at 120 frames per second.

HoloLens with the add-on eye tracker.

Even though Pupil Lab built an eye tracker specifically to work with the HoloLens, this eye tracking was not designed to account for the HoloLens hardware structure. Precisely, the eye tracker world camera is placed on the top of the HoloLens, increasing the distance between the world and eye cameras. As a result, the accuracy of the eye tracker is severely affected (see Figure 4.1).

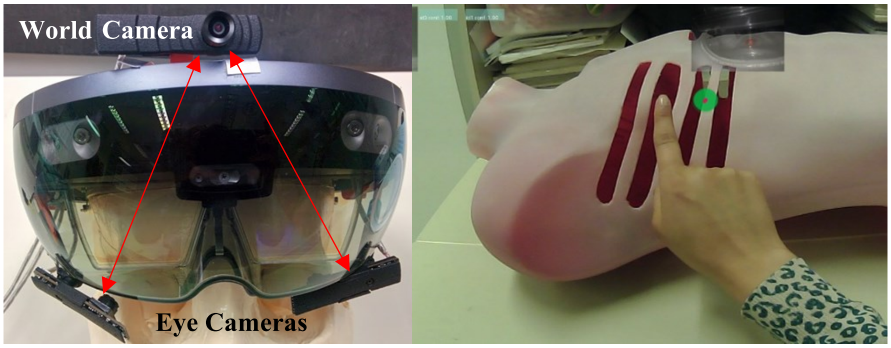

Figure 4.1. Left: position of the world camera on the top of the HoloLens. Right: Gaze is significantly off despite an appropriate calibration.

I had two options to solve this issue: (a) use the scene camera of the HoloLens or (b) change the position of the eye tracker world camera. The first option was not feasible because the Microsoft HoloLens does not have an open-source code to modify or interfere with the connection of the scene camera. Therefore, I used the conventional eye tracking design from Pupil Lab (Figure 4.2) to relocate the world camera. Head-mounted pupil lab eye trackers placed the world camera close to the eyes to decrease the parallax error, which is described as a geometrical error caused by the position of the scene camera and the individual's eyeball in a non-common axis (Su, Li, & Xiong, 2016). In human vision, parallax error does not exist because the visual input is received directly to our eyes.

Figure 4.2. Pupil Lab head-mounted eye tracker with an accuracy of 0.60° and a precision of 0.02. It is retrieved from https://pupil-labs.com/products/core/tech-specs.

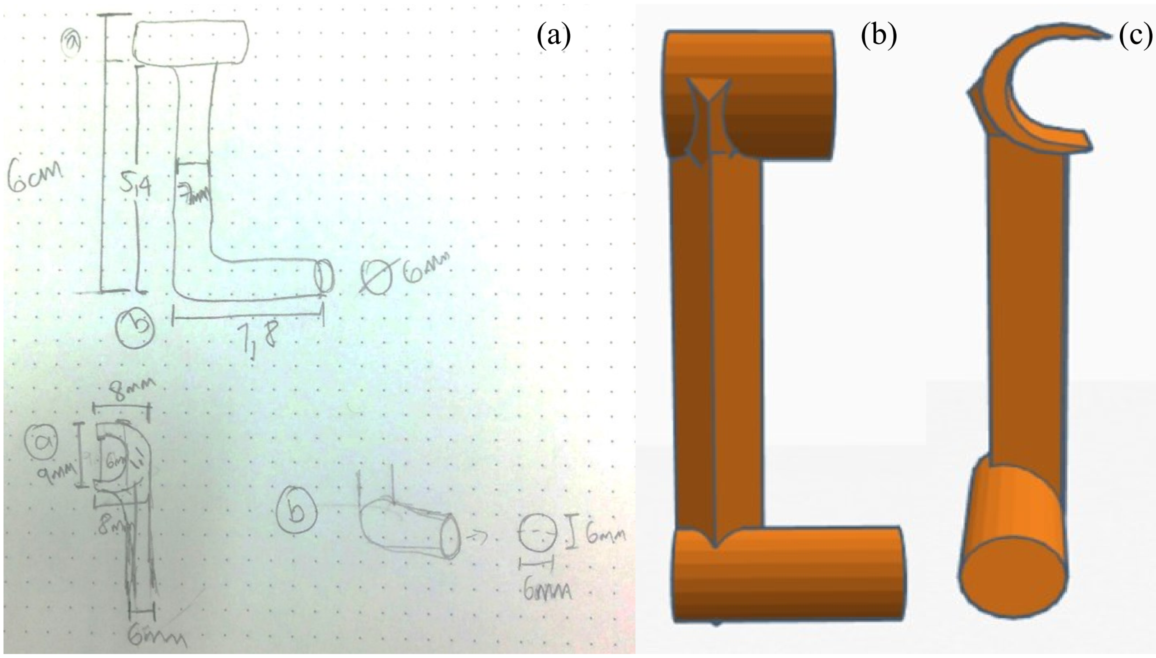

Figure 4.3. (a) sketch of the arm with its measurements, (b) anterior view of the arm, and (c) lateral view.

After measuring the distances between the scene and eye cameras, a 3D extender arm was designed using the TinKercad 3D Design Software (San Francisco, U.S.) and printed to place the world camera closer to the eyes (Figure 4.3).

Several tests were performed confirming that with the new position of the world camera, the add-on eye tracker had a similar accuracy achieved with a conventional head-mounted pupil lab device (see Figure 4.4).

Figure 4.4. Left: AR headset with the new position of the world camera of the add-on eye tracker. Right: Screenshot of the visualization of eye gaze, now more accurate.

4.3 Integration of AR and Eye Tracking Systems

After the HoloLens and the add-ons pupil lab were successfully attached, and the proper accuracy of the eye tracker was achieved, the next step was to establish a communication between the two devices. The entire system was built in the Surgical Simulation Research Lab (SSRL) at the University of Alberta using a high-performance computer and testing a simulation model for the system. We used the chest tube simulation model described in Chapter 3. In the next paragraphs, I will describe the software characteristics and implementation of the system.

We used Unity Editor software (Unity Technologies, U.S) to construct HoloLens apps and link the HoloLens and eye tracking technology. Unity is a game development application containing a variety of tools to create 2D and 3D scene designs. Scripts are written in C++ using Visual Studio (Microsoft Corporation, U.S).

Pupil lab software.

Pupil Lab eye-tracking software is composed of two elements: Pupil Capture and Pupil Player. Pupil Capture records, in real-time, data of eye movements, including gaze location and pupil diameter, as well as ground detection information (Kassner, Patera, & Bulling, 2014). Pupil Capture contains several plugins, including the "HoloLens Relay" plugin to pass Pupil Lab eye tracking information to HoloLens using the User Datagram Protocol (UDP). This plugin, however, only reads and transports data about the eye location. In this research, to obtain and transfer pupil data, a change was introduced to this plugin. On the other hand, Pupil Player is used to playback video and gaze data recorded using the pupil capture.

4.4 Implementation

Areas of interest and surface tracker.

We used the three areas of interest (AOIs) created during the chest tube insertion project (Chapter 5): Skin Pad, Instructions, and Drape. These AOIs were selected for the reason that participants should spend most of the time looking at these areas to either perform the task or review the instructions for guidance. Pupil Lab Capture software has a plugin called "Surface Tracker," which enables people to identify objects and monitor them in real-time via markers in the setting. Figure 4.5 demonstrates how the surface is ready, as determined by using sensors and the identification outcome in the implementation of the Pupil Lab. Surfaces are automatically recorded and saved as "surface definitions." Once the Pupil Lab eye tracker starts, the surfaces appear on the screen like in Figure 4.5 if the “Surface Tracker” plugin is on. As shown in Figure 4.5, even some markers are not detected due to decreased light, but the surfaces are still detected, showing square windows upon the surfaces. Surface data and the gaze position relative to it are pushed into data flow in Pupil Lab application. Our system can subscribe to such data and broadcast them to HoloLens (see the section below).

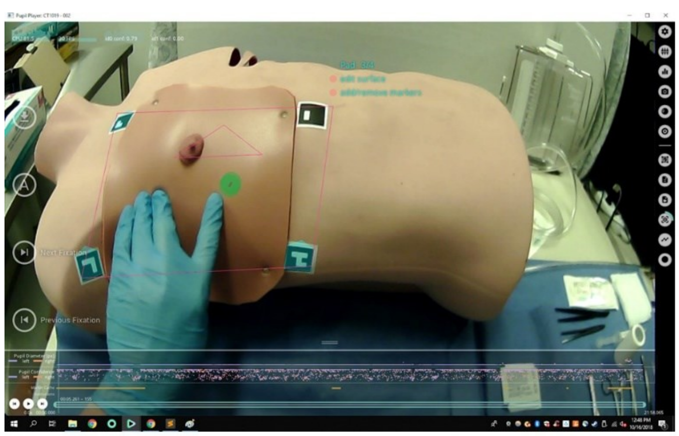

Figure 4.5. The Surface Tracker plugin identifies the Skin Pad surface in real-time.

Virtual data displayed on the HoloLens screen can be activated and deactivated by using the gaze location information. We took the eye's location relative to surfaces as the performance status marker for the trainee. If the eye is inside the skin pad, this implies that the student is likely to focus on the job at hand. Then, to eliminate distractions, the AR data is impaired. If the eyes are inside the instruction sheet area, this implies that the student is seeking assistance at that moment. Therefore, displaying AR data would be ideal at the time of seeking information.

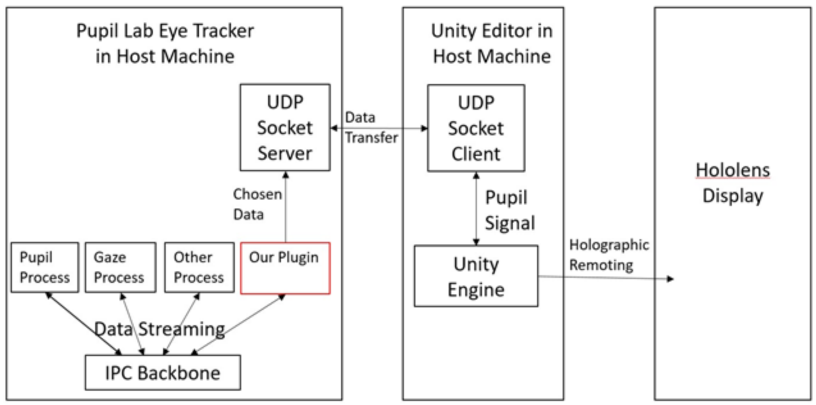

Figure 4.6. The flow of the data between eye tracker and HoloLens through the Unity Editor.

Eye-tracking signal acquiring and processing.

We used the Pupil Lab's network communication system to stream data obtained from the eye tracker to the AR. Pupil software employs the ZeroMQ network library as the basis of the communication bus, known as the Inter-Process Communication (IPC) backbone. In other words, the IPC is the mainstay of all communication within, from and to Pupil apps (Prietz, Sae-Tan, Yang, & Patera, 2018).

A single backbone binds to various procedures, thus creating a one-to-many communication proxy (Figure 4.6). Each method can then force the signal into the backbone and remove texts from the mainframe of others. For example, each frame is generated with surface data and the gaze position relative to it and pushed under the "surface" topic into IPC backbone.

An "IPC sub" connector is constructed in the plugin we altered, through which our plugin subscribes to IPC mainframe information. In the socket, several subjects are introduced.

Consequently, our plugin[1] captures this information whenever the information under these subjects appears in the core of the IPC. A distant plug with UDP protocol for transmitting information to HoloLens (Unity Editor) is constructed in our plugin. Our plugin repacks the information into byte format after subscribing to the information from the IPC backbone and provides an additional header to the information stating the sort of information. The plugin then transfers such information to the distant socket and transmits it to HoloLens (Unity Editor) (see Figure 4.6).

Because Pupil Lab Eye-tracker is open-source software, our plugin was placed straight in the folder "pupil catch settings/plugins." This plugin can be activated through the primary menu of the Pupil Capture application.

The workflow in HoloLens.

In the HoloLens implementation, Pupil Lab Eye-tracker software offers Unity plugins for eye tracking. However, since this research altered the Pupil Lab Eye-tracker plugin, a change to the fresh information format was required in the HoloLens Unity set. A UDP client socket is constructed in the HoloLens implementation (Unity Editor), then it listens to a distant IP address set as Pupil Lab Eye-tracker's IP. When running the software HoloLens (Unity Editor), it writes an "I3" byte to the IP address of the Pupil Lab Eye-tracker. Pupil Lab Eye-tracker reacts to the broadcast socket with a byte "0I" after getting this byte. When HoloLens (Unity Editor) UDP customer socket receives this byte, the link is created (Figure 4.6).

All calibration and recording procedures in our scheme are performed via Pupil Lab Eye-tracker, so we ignore or remove associated parts in the HoloLens Unity set. HoloLens implementation in Unity Editor writes byte "S" to the IP address of the eye tracker after the link is formed, showing that it is ready for information transfer. When the information under chosen subjects is driven into the core of IPC after answering with byte "0S," the Pupil Lab Eye-tracker transmits all this information to the distant socket. Application HoloLens (Unity Editor) gets all of this information via UDP client. For interpreting information from byte format to digit or

[1] Because the plugin was largely changed in the rest of this paper, the modified plugin is called "our plugin".

string, a feature is constructed. Based on the data header, the digit or string is collected under distinct variables and is modified whenever new information is obtained.

4.5 System Setup Procedure

Now that the AR and eye-tracking software are communicating by IP addresses and UDP sockets (server and client) using Unity Editor as an intermediary, the system setup process is described below.

Calibration.

Pupil Capture contains a calibration application to map the center of the pupil identified in the eye pictures to gaze coordinates in the scene picture. Calibration needs to be done before using the eye tracker for data recording. The first step for an adequate calibration is to make sure the eye monitoring cameras are correctly placed to capture the pupils in all directions the eye may move. This step is crucial because the Pupil Lab software employs the dark pupil detection method, which detects the darkest structure of the eye (pupil) in an eye camera image lighted up with infrared light (Kassner et al., 2014). We use the most robust calibration method, the “Six-point Screen Marker.” A participant sits in front of a computer screen, and a target appears on six different positions where the participant needs to gaze keeping the head as still as possible. The calibration step can be repeated as needed until the desired calibration is obtained.

Connection setup.

The software HoloLens operates properly in Unity Editor if the following settings are enabled and supported in Unity: "Universal Windows Platform," with "Windows Mixed Reality" and "internet client," "InternetClientServer," "PrivateNetworkClientServer," "Microphone," and "Spatial Perception." It is necessary to connect HoloLens and the host machine to the same Wi-Fi network with only one router. The eye tracker has as well its IP address “Pupil Remote IP” to connect with other devices. We used the “PupilSettings” created in Unity to put in the eye tracker’s IP address. Then, the HoloLens is linked to the host machine by using the "Holographic Remoting" app in the headset and "Holographic Emulation Mode" in Unity. Now, the three instruments: HoloLens, host computer (containing Unity Editor), and Pupil Lab Eye-tracker are linked.

4.6 Control of Displaying AR Information

After connecting the three devices, the scene created on Unity for the HoloLens can operate on the host machine in Unity Editor play mode at the same time as being broadcast on the HoloLens via "Holographic Remoting." Now, the student can carry the apparatus (mounted HoloLens with Pupil Lab Eye-tracker devices) and execute the simulation job on the model. The holographic information is controlled using the gaze position on the previously determined surfaces (AOIs). Specifically, if the position of the gaze is inside the "Instructions" surface, computer-generated texts, images, or videos appear on the HoloLens screen to provide the learner with instructional information. On the other hand, the holographic data vanish when the student switches back to the surgical site.

4.7 Discussion

The purpose of this document was to portray the first steps of creating an eye-tracking enable AR platform. We demonstrated the concept of AR and eye-tracking technologies integration to facilitate the display of computer-generated instructional information. We selected the Pupil Lab Eye-tracker and HoloLens in our project. The eye tracker is a user-friendly and open source. Since HoloLens is a famous and robust AR tool on the market, it has several advantages: 1) it has a decent cost; 2) it is easy to operate; 3) in HoloLens we can operate various instruction modules to practice distinct procedures; 4) assets and plugins are available to integrate the HoloLens with the Pupil Lab Eye-tracker.

In this project, we first fixed the issue of eye tracker accuracy due to the position of the work camera of the add-ons on the HoloLens. Then, the devices were connected through Unity using IP addresses and ICP backbones to transfer information from the eye tracker to the AR headset. Later, we employed moments of vision lingering on specified surfaces to turn on/off educational data in the HoloLens. All of these steps were performed towards the goal of automatically displaying instructional information on the HoloLens during moments of performance difficulty in a chest tube placement procedure. Hence, in our future work, we intend to determine parameters and thresholds using the data of *Gaze Dispersion* obtained in the research project in Chapter 5 to trigger the information on the AR headset as it is needed.

While this novel system offers excellent opportunities to create a smart platform to practice a surgical task, there are some constraints to consider. First, the system depends on a UDP link between the host computer and HoloLens. The UDP link is obstructed if the host machine and HoloLens connect to separate routers. Consequently, the scheme operates pretty well under one Wi-Fi router while losing the link under a Wi-Fi multi-router. For example, our system can be operated using university Wi-Fi. Second, it is imperative to design an effective environment learning to select the relevant information at the proper size because the HoloLens has a small screen (30°). Third, the versatility of the wireless Microsoft HoloLens comes with the caveat of its weight. After a while of using it, the HoloLens becomes cumbersome and even produces discomfort or pain.

Also, we mentioned that the gaze position and pupil diameter are available in our system. However, researchers need to consider the trouble of working with pupil data. Many factors produce involuntary pupil diameter changes no related to mental workload during moments of difficulty. For example, pupil size is strongly affected by slight light changes. In AR technology, illumination will never be constant because the virtual information on the screen will produce light at different intensities making the pupil data not reliable.

We believe that this research provides the groundwork for the next steps of creating a smart platform to learn surgical procedures such as chest tube placement.

Chapter 5 Using Eye Tracking to Predict Performance Difficulty during Chest Tube Insertion

In the previous chapter, I described the first step of the integration of AR and eye-tracking systems and how gaze position information can be used to control holographic information. The next step in the process of building a smart eye-tracking enabled AR platform is the identification of a suitable eye-tracking metric to automatize the AR system. Therefore, the main goal of this chapter is to examine eye-tracking metrics that can be used to identify the moments of performance difficulty (MPDs) during a multi-step surgical procedure.

5.1 Introduction

When trainees are learning a surgical procedure with multiple steps, such as the chest tube insertion, they may not be able to remember all the steps. In the middle of the task, students may need to read the written instructions or check with instructors. Therefore, the flow of the performance is paused. With the new technologies such as augmented reality (AR), instructional messages can be displayed to the visual field of the trainee without the need for moving their gaze away from the surgical site and disrupting the task.

Augmented reality technology has been used as a training platform during assembly tasks. These types of tasks could be compared to those in medicine as they involve eye-hand coordination and cognitive requirements to succeed in the task. Augmented reality has been used as an instructional and supportive platform during the assembly tasks. Henderson and Feiner (2010) employed an AR head-mounted display (HMD) to assist users during the maintenance of a military vehicle. The AR elements were attention-directing information, text instructions, location labels, and 3D models of tools. The researchers found that AR reduced the time for locating a task and the number of head and neck movements. However, frequently display instructions to the AR users may also distract their attention. It is noticeable that one major issue is the obstructiveness of the real environment caused by the virtual information (Henderson & Feiner, 2010).

In this project, we offer the foundation of a smart eye-tracking enabled AR platform for teaching a chest tube insertion. We do that by using eye-tracking technology to identify *moments of performance difficulty* (MPDs), where the cognitive load increases.

Learning and performing multi-step surgical procedures, especially for the first time, could be demanding for most students. Surgical procedures have components that can add cognitive load to the learner, leading to struggle during the tasks. They need to not only learn and execute the procedure flawlessly but manage resources and task demands. Cognitive load is defined as the load imposed on the learner during the performance of a task and the individual needs to use resources and capacities to respond to the demand given by the task. Several factors can increase the cognitive load, such as the characteristics of the task, environment, and subject (Paas & Van Merriënboer, 1994). For example, a trainee performing a new task under limited time may significantly raise the cognitive load.

Increased cognitive load can hinder the performance of a procedure. Therefore, it is imperative to identify and assess mental workload experienced by novices during the learning process of a surgical procedure. There are different subjective and objective methods to study cognitive load. A systematic review showed that the most employed method to measure cognitive load in surgeons was the NASA Task Load Index (TLX) (Dias, Ngo-Howard, Boskovski, Zenati, & Yule, 2018). NASA TLX is one of the most popular subjective assessment techniques. NASA TLX is a multidimensional assessment tool that scores the mental workload perceived by a user during or after performing a task (Sandra G. Hart, 2006). This tool has been used widely in different fields obtaining results correlated to other measures of cognitive load. Also, NASA TLX has the advantages of being portable, easy to implement (there is a software version), and providing a detailed analysis of the source of the workload (Cao, Chintamani, Pandya, & Ellis, 2009). However, NASA TLX is not exempted from drawbacks. Applying this scale could be time-consuming, repetitive, and it is prone to biases because of its subjectivity. For example, the responses are affected by the state of the users and how much they remember of their performance during the task. Also, this method cannot be applied in real-time, decreasing its validity (Dias et al., 2018). As a result, NASA TLX scores go along with other objective measures of mental workload, such as eye tracking.

The relation between eye tracking and cognitive load found in MPDs has continued to evolve, thanks to the advancement in technology. At the moment, several metrics have been used to measure cognitive load such as fixation (duration, ratio, horizontal and vertical dispersion), saccades (velocity, rate, amplitude), dwell time, gaze patterns and distribution, blinks (duration, frequency, and interval), and pupil diameter changes (Coral, 2016; Holmqvist et al., 2011).

The measurement of pupil diameter changes as a metric for mental workload has gained popularity among researchers in the surgical field (Bednarik et al., 2018; Zhang et al., 2018; Zheng, Jiang, & Atkins, 2015). However, pupil diameter is strongly influenced by several factors independent of the user, such as illumination of the environment (Mathôt, 2018). In mixed environments, computer-generated images are created and displayed on screens of an AR device, such as the HoloLens. Introducing virtual images provides an extra source of light that is not constant, in turn, producing changes in the user's pupils size to adapt to the amount of light received. This variation of pupil diameter due to light is difficult to filter, and there are not enough studies that explore the data cleaning in AR environments. As a result, we did not consider pupil diameter as a suitable measure in our project.

In this study, we focused on the use of the spatial location of the gaze, or *gaze dispersion*. Gaze dispersion as a measurement of cognitive load that started in the field of driving and pilot safety in fixed images, resulting in a more focused gaze during challenging moments. This construct has been supported by several researchers in the fields of aviation and driving where the dispersion of the gaze is decreased (tunneling effect) when individuals face an increase in cognitive workload especially during secondary tasks situations (Marquart, Cabrall, & de Winter, 2015; Reimer, 2009; Savage, Potter, & Tatler, 2013).

Nevertheless, opposing results have been found when eye gaze spatial distribution is analyzed using another method, or it is applied to other types of dynamic environments. Di Nocera et al. (2007) employed the Nearest Neighbor Index (NNI), which is the ratio between point distribution over an entire scene. Researchers found that higher values of NNI were linked to high workload phases of an aviation simulation including dynamic and static interfaces (Di Nocera, Camilli, & Terenzi, 2007). Gaze dispersion has been employed recently to study the effect of automatization during car driving. It has been demonstrated that during episodes of difficulty, such as when visualization was poor and cognitive load was high, gaze dispersion increased (Louw & Merat, 2017). In another study related to surgical procedures, Di Stasi et al., (2017) performed a study using gaze-based metrics to assessed the response towards the increased workload during laparoscopic procedures in a group of eight experts and eight novices, finding that during more complex tasks the participants had a more disorganized and unpredictable gaze (entropy) (Di Stasi et al., 2017).

Contradictory results in the gaze dispersion studies are due to the influence of the characteristics of the task on the eye tracking data. Based on the research done in dynamic environments and surgical tasks using gaze dispersion, we hypothesized that the gaze dispersion during MPDs in a chest tube insertion task would be larger than the NMPs representing more cognitive load.

5.2 Methods

The study was performed in a controlled environment at the Surgical Simulation Research Lab at the University of Alberta under the Ethics approval number Pro00080600. Twenty medical students (50 % female, 95% right-handed, and age 24 ± 2.7 years) in their first or second medical year at the University of Alberta were recruited with normal to corrected vision and without surgical experience. After all participants read the instructions and signed the consent form, they filled out the demographic questionnaire and practiced surgical hand knots. A pre-recorded explanatory video about a chest tube placement on the simulator, described in Chapter 3, was presented to the participants.

All participants wore an AR headset (HoloLens) with an eye-tracker (Pupil Lab) attached to it (Figure 5.1). In this part of the research, we did not display any information on the HoloLens. However, it was essential to assess the functionality of the HoloLens and eye tracking for future steps in this research. The eye tracker required calibration, which was performed using the screen market calibration method of the pupil lab software on a computer screen placed at the same level of the chest tube simulator and one arm length from the participant. The reason to do the calibration at this position was to get as close as possible as to the location the participant was going to be during the task performance to ensure the working area was covered within the calibration range for best tracking accuracy. The calibration accuracy was visualized with the accuracy visualizer plugin of the Pupil Lab. Also, participants were asked to look at specific points of the working area to confirm that the gaze position matched the intended gaze.

Figure 5.1. Left: AR headset with a set of eye tracker. Right: Screenshot of the visualization of eye gaze during the procedure

After the eye-tracking calibration, the participants performed the chest tube insertion once on the chest tube simulator using real surgical instruments, including a 28 Fr chest tube and a non-absorbable suture 2.0 (Figure 5.2).

Figure 5.2. Simulated chest tube scenario including a panel of instruments

On the chest tube simulator, we defined three surfaces (*Skin Pad, Instructions, and Drape*) that can be identified by the eye-tracking system. These surfaces were used to categorize the positions of the gaze on the physical environment, and to track the areas of interest. The surfaces were defined using printed surface markers downloaded from the Pupil Lab website (https://docs.pupil-labs.com/#pupil-capture) and the Surface plugin in Pupil Capture. We used more than three markers to define each of the surfaces to increase the robustness in the eye tracking of these surfaces (see Figure 5.3).

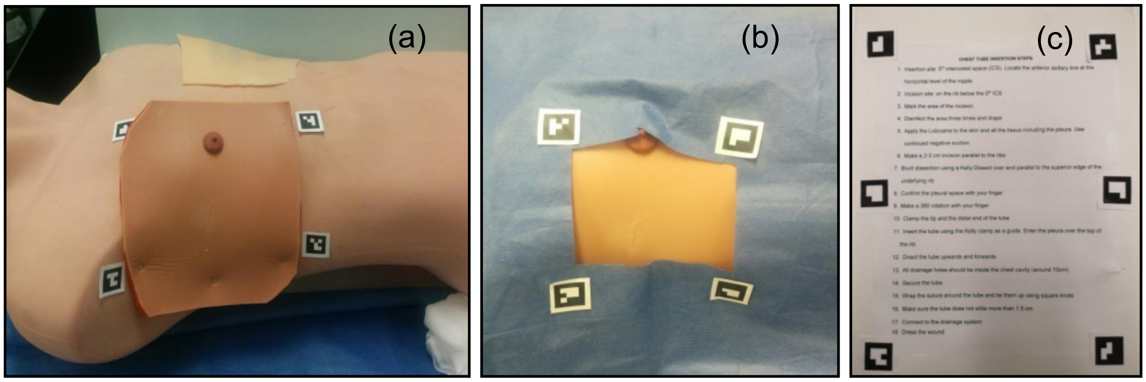

Figure 5.3. Four markers were to define the Skin Pad (a) and Drape (b) surfaces, and six markers for the Instructions (c) sheet.

During the performance of the task, written instructions and guidance from the instructor were available to the participant upon request. However, they were encouraged to perform the task with the least help possible. Also, a printed red circle was placed in front of the participants when a step was performed incorrectly (*Mistake*), or they had forgotten one (*Lapse*). The terms Mistake and Lapse were used based on the description of Reason (1995) of human failures where mistakes are due to failure of intention and lapses belong to failures of execution due to failure in memory (Reason, 1995). They had to correct the mistake or lapse before continuing the task.

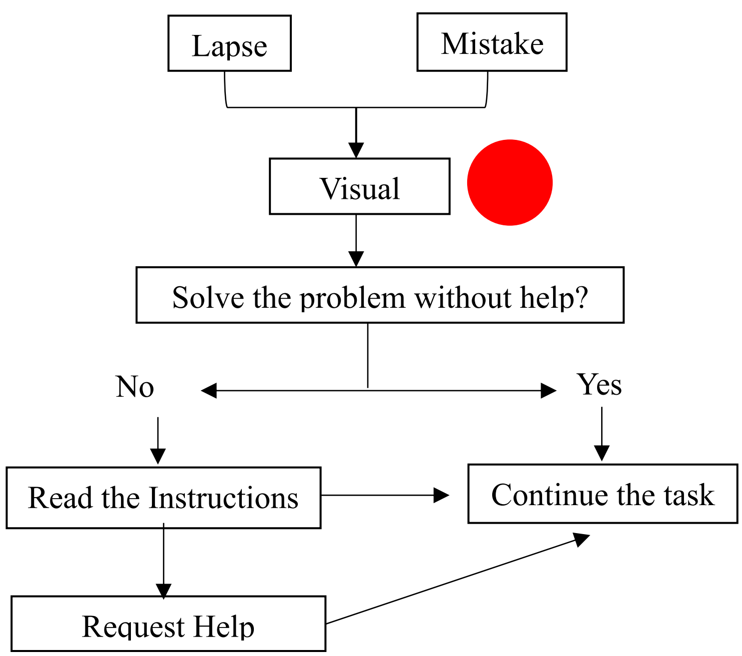

Figure 5.4. Flowchart for solving lapses and mistakes to continue the task.

Only if they were not able to solve the problem even with the assistance of the written instructions, could the participants ask for guidance. In Figure 5.4, the flow of the problem-solving process is shown.

After the participants finished the task, they were asked to fill out a computerized version of the NASA-TLX (Task Load Index) mental workload score. All participants received a gift card for participating.

Metrics.

Number of errors and help requested.

In each subtask (defined in the section of video analysis), the total number of mistakes, lapses, instructions check, and help requested were registered to identify the most challenging subtasks.

Eye tracking.

In this project, we used *Gaze Dispersion* to characterize the distribution of the gaze location during the difficult moments in the task. The unit of gaze dispersion was provided in percentage to facilitate data analysis and interpretation. For instance, a gaze dispersion of 30% is larger than one that of 5%, which means in the latter, the participant's gaze was focused on one small area of the simulated scenario (world view). Pupil Lab software computes the gaze position using the eye position information recorded with the eye cameras with respect to the world camera image. The eye positions are given in the X, Y, and Z axes in normalized numbers (min 0 to max 1). In this project, I utilized the normalized X and Y eye positions of both eyes.

Subjective mental workload.

Each participant filled out a computerized version of the NASA-TLX (Task Load Index) subjective mental workload score. NASA-TLX is composed of two sections, ratings and weights. First, the participants rated each of the six subscales (Mental Demand, Physical Demand, Temporal Demand, Own Performance, Effort and Frustration levels) from 0 to 100 (least to most demanding). Then, the participant chooses the subscales that they thought contributed more to the workload through 15 pair-wise comparisons of the six subscales. Each selected subscale was tallied, and tallies range from 0 to 5 (not significant to most significant). Lastly, the ratings and weights were combined to obtain the overall mental workload score ranging from 0-100 (least to most mental workload) (Sandra G Hart, 1986). Even though there was not a limit of time to complete the task, the temporal component of the NASA-TLX evaluates the workload participants feel regarding the time they spent on the task regardless of whether it was timed or not.

Other metrics.

Completion time and performance scores were also gathered but analyzed separately in Chapter 3.

Data collection.

The data was gathered using the Pupil Lab eye tracker and a room camera in case a reference source was needed. The recordings were performed using the Pupil Lab software composed of two parts: Pupil Capture and Pupil Player. Pupil Capture allows for data collection in real-time, while Pupil Capture is used to visualize media and data. At the moment of data collection, the versions of the eye-tracking software did not allow long recordings of data. The maximum time allowed without crashing the system was around 15 minutes. However, a chest tube insertion procedure can last more than that if the operator is a novice. As a result, the data was collected in two or more files. The task was not interrupted to start a new recording as it can be done by clicking the recording bottom on the screen and a new calibration is not needed. However, in only two cases, the task was interrupted voluntarily by the participant because they wanted to take a break due to the heaviness of the HoloLens. In those cases, a new calibration of the eye tracker was required. Before eye-tracking data analysis, the data folders of each participant were merged using a customized MATLAB (MathWorks, Inc. Massachusetts) algorithm.

5.3 Data Analysis

The data were analyzed in two sections: video analysis and eye tracking data analysis.

Video analysis.

First, the videos of the eye tracking that were recorded using the world camera were analyzed using Pupil Player, which allows us to create annotations and trim videos to select the segments to analyze. Each trial was divided into subtasks and annotations were created accordingly: Landmarks; Disinfection; Anesthesia; Incision; Dissection; Insertion; Securing; and Connection (to the drainage system); and Dressing of the wound.

In each subtask, events of interest were identified and labelled on the video using the pupil lab annotation application. To have a clear identification of the events, they were defined as follows:

- *Moments of Performance Difficulty (MPDs)* are the instants when participants voluntarily stop the task to seek instructions by either reading the instructions sheet or asking for help from the instructor (*Instructions Check and Help Request*).
- *Normal Moments of Performance (NMPs)* are the periods where the participants perform the task without mistakes and help, which excludes the moments when participants are correcting a mistake or lapse.

Selection of NMPs.

NMPs were selected once for each subtask with the following criteria:

(a) No presence of MPDs, mistakes, or lapses

(b) The gaze is on the working site

(c) Data quality > 70%

(d) Events related to instrument handling were not included. For example, reaching for the instruments or grabbing the suture with the needle driver

By considering the criteria above, we avoided the overlapping of gaze patterns between events and subtasks. Therefore, we were confident that we compared true MPDs and NMPs. For comparison of the gaze patterns during MPDs vs NMPs, the time chosen to compare was five seconds before the MPDs and five seconds of NMPs in each subtask. We found that segments lasting more than five seconds had a higher chance of overlapping events. Also, selecting five seconds before the MPDs (Instruction check and Help Requested) gave us a better insight into the eye behaviours during the moments of struggling that led the participant to seek guidance to complete the task. We did not include the data where there were mistakes or lapses because, at that moment, there was an intervention which influenced and changed the eye behavior.

Frequency of the number of MPDs (Help Requested, and Instructions check) was calculated based on the video analysis to determine the more challenging subtasks; the eye-tracking data was further explored to find out if gaze dispersion reflected the level of difficulty of the tasks. Dissection, Insertion, and Securing were the most challenging subtasks accounting for 15.4%, 20.5%, and 51.3% of the total Help Requested events, respectively. Also, these three subtasks represented more of the Instruction Check events, Dissection 24.0%, Insertion 20.2, and Securing 22.1% (Table 5.1).

Table 5.1.

Frequency of MPDs among the subtasks.

Subtasks	**Help Requested (N)**	**%**	**Instructions Check (N)**	**%**
Landmarks	7	9.0	12	11.5
D&A	2	2.6	15	14.4
Incision	1	1.3	5	4.8
Dissection	12	15.4	25	24.0
Insertion	16	20.5	21	20.2
Securing	40	51.3	23	22.1
C&D	0	0.0	3	2.9
Total	**78**	**100**	**104**	**100**

D&A: Disinfection and Anesthesia; C&D: Connection and Dressing the wound

Eye-tracking data processing.

A MATLAB (MathWorks, Massachusetts, USA) script was used to clean the eye-tracking data. A data confidence filter of more than 0.7 was used to discard not reliable data. In the same script, a linear interpolation function was run to deal with the missing data. Linear interpolation function estimates data points between two known points. Also, a third-order media filter was used to clean the data.

The data was also visualized using MATLAB by plotting the gaze positions on the surfaces coded by colours (Blue: gaze on the Skin Pad surface, Red: gaze on the Drape surface, and Green: gaze on the Instructions sheet) and the annotations. This visualization helped us to identify the accurate interposition of gaze positions on the surfaces. Data were considered accurate when the gaze on the surfaces corresponded to the timestamps (seconds) of the annotations (see Figure 5.5).

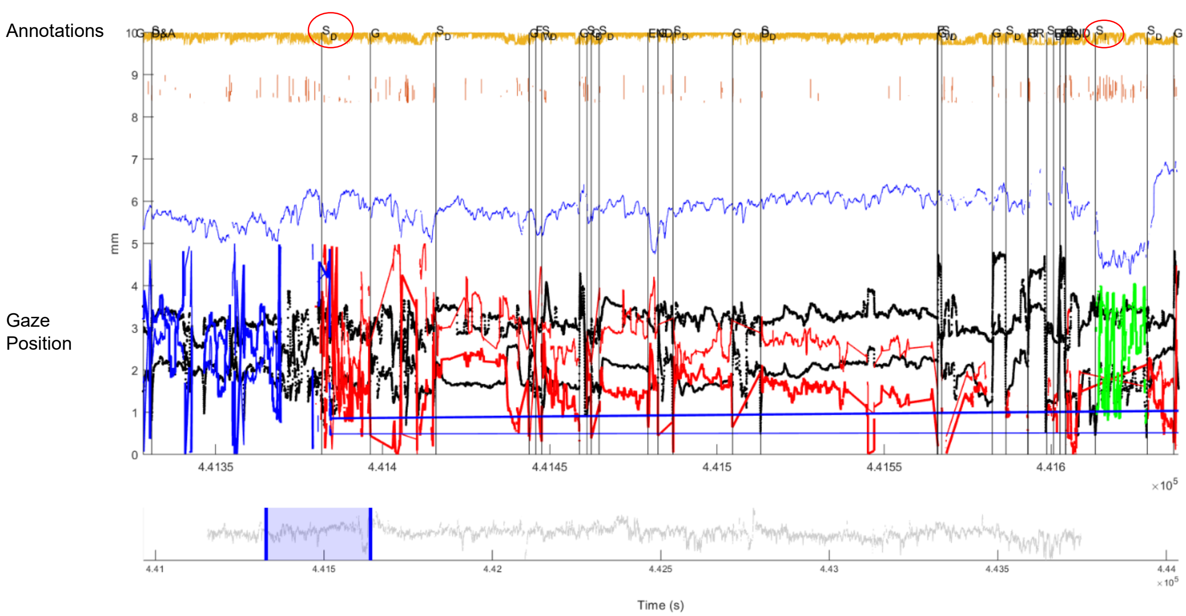

Figure 5.5. Gaze position and annotation plot. Gaze position on surfaces coded as blue for the Skin Pad, red for the Drape, and green for the Instructions sheet. The X-axis represents the timestamp in seconds, and the Y-axis are the units of the normalized gaze position exponentiated to a maximum of 5. The thin blue line in the middle of the graph represents the pupil data (not included in this thesis) in mm on the Y-axis.

Finally, gaze dispersion was calculated for each subtask and event. The gaze dispersion for the five seconds during the NMPs (subtasks) and the five seconds before the MPDs events were calculated utilizing the minimum and maximum value of X and Y to create a rectangle to show which area corresponds to how much the gaze was dispersed during the time selected (Figure 5.6). The formula was:

dx =Xmax – Xmin (width of the rectangle)

dy = Ymax – Ymin (height of the rectangle)

s = dx*dy (area of the rectangle. Area is proportional to the gaze dispersion)

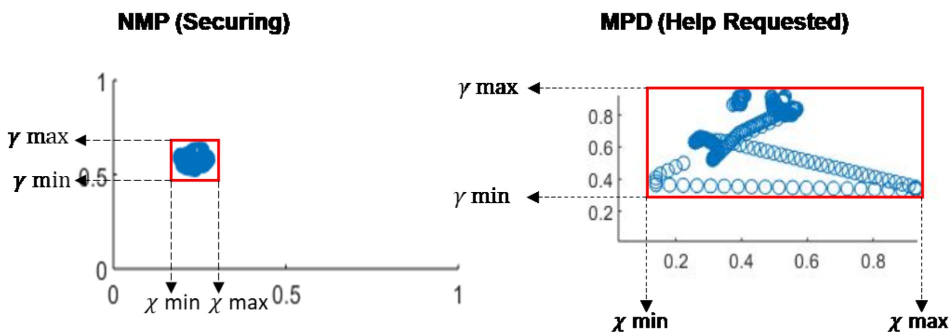

Figure 5.6. Examples of gaze dispersion. Gaze dispersion of a NMP during securing (left). It is a small gaze dispersion compared to the gaze dispersion of a MPD before requesting help (right).

Statistical analysis.

SPSS 24 was used for the statistical analysis, and a $p < 0.05$ was considered for statistical significance. Results are expressed as mean and standard deviation (SD). First, the gaze dispersion means of three NMP (Dissection, Insertion, and Securing) were analyzed using one-way ANOVA to examine whether there was a significant difference among them. NMPs and MPDs gaze dispersions were analyzed and compared using the non-parametric test, Kruskal-Wallis. Lastly, Spearman's correlation was used for the MPDs and NASA TLX. We used Spearman's correlation test because the data violated two assumptions to use Pearson's test: normality and linear correlation.

NMP.

Hypothesis.

The means of the gaze dispersion are significantly different among the three subtasks: Dissection, Insertion, and Securing.

Assumptions.

The assumptions of normality and homogeneity of variances were assessed. The results of the Shapiro-Wilk normality test were not statistically significant for Dissection ($p = 0.085$), Insertion ($p = 0.241$), and Securing ($p = 0.134$) meaning that the data can be considered normally distributed. The results of Levene's test were not significant $F(2, 51) = 0.150$, $p = 0.473$, indicating that equal variances can be assumed, and the ANOVA test can be used.

MPD.

Hypothesis.

The means of the gaze dispersion are significantly different among the two MPDs: Help Requested and Instructions Check.

Assumptions.

The assumptions of normality and homogeneity of variances were assessed. The normality of the Help Requested and Instructions Check gaze dispersion data were assessed using the Shapiro-Wilk normality test. Help Requested was normally distributed ($p = 0.23$). However, Instructions Check gaze dispersion data were not normally distributed ($p = 0.014$). Also, the assumption of equal variances was not met. As a result, a non-parametric version of the ANOVA, Kruskal-Wallis test, was used to compare the NMPs with the MPDs.

5.4 Results

First, there were no statistically significant differences in Gaze Dispersion (%) between Dissection, Insertion, and Securing subtasks as determined by one-way ANOVA ($F(2,51) = 0.128$, $p = 0.880$), Table 5.2. Therefore, the mean of these three subtasks was used as the NMPs to compare with the MPDs (Help Requested, and Instructions Check).

Table 5.2.

ANOVA results of the gaze dispersion mean of the three most challenging subtasks.

Subtasks	**N**	**Mean ± SD of Gaze Dispersion (%)**	***F***	***p* value**
Dissection	18	3.48 ± 2.78	0.128	0.880
Insertion	18	3.08 ± 2.07		
Securing	18	3.37 ± 2.32		

In Table 5.3, a Kruskal-Wallis test showed that the participants' gaze dispersion (%) of the NMPs and the MPDs groups are significantly different, H (2) = 32.835, $p < 0.001$. The gaze dispersion during Help Requested was larger (*Mdn* = 19.57) than Instructions Check (*Mdn* = 15.01) or NMPs (*Mdn* = 3.57). Dunn's Post-Hoc pairwise tests (Bonferroni correction) were employed for the three pairs of groups. The difference in gaze dispersion between the NMPs and the Instructions Check group was statistically significant ($p = 0.001$). Similarly, in the pairwise comparison, the NMPs and the Help Requested group were statistically significantly different ($p < 0.001$). No statistically significant differences were found between the Instructions Check and Help Requested groups ($p = 0.269$).

Table 5.3.

Kruskal-Wallis results of the gaze dispersion mean of the MPDs (Instructions Check and Help Requested) versus NMPs.

Event		**Median**	**Mean Rank**	***df***	***H***	***p-value***
MPDs	Instructions Check	15.01	28.43	2	32.835	< 0.001
	Help Requested	19.57	37.12	2		
NMPs		3.57	10.22	2		

NASA score was obtained from the overall task instead of each subtask. Therefore, we correlated the NASA scores with the gaze dispersion of the average of all MPD events during the tasks. The Spearman's test indicated that there was a not statistically significant weak positive association between the NASA_TLX scores and the gaze dispersion of the Help Requested (r (2) = 0.134, p = 0.595), Figure 5.7. A similar result was found between NASA NASA_TLX scores and Instructions Check (r (2) = 0.005, p = 0.985).

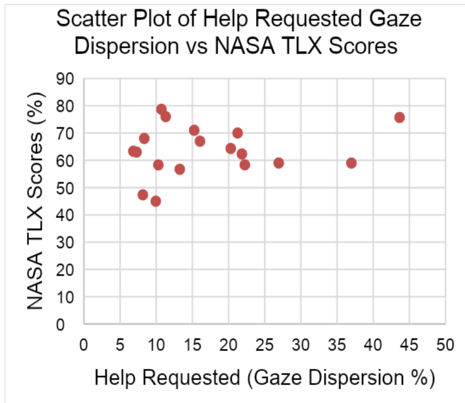

Figure 5.7. Scatter Plot of correlation between NASA TLX Scores (%) and gaze dispersion of the MPDs, Help Requested.

5.5 Discussion

In this study, we first identified the most challenging subtasks (Dissection, Insertion, and Security) by the frequency of MPDs. The number of Help Requested and Instructions Check during the three more challenging subtasks represented 87.2% and 66.3%, respectively, of MPDs in the whole task. The findings were not a surprise for us as most of the complications of a chest tube insertion are associated with the insertion and position of the chest tube (Ball et al., 2007; Kashani et al., 2017).

The results of eye gaze analysis supported our hypothesis. At an elementary level, we detected and characterized areas of performer's difficulty by using eye tracking added to an AR headset. In our study, participants had more dispersion of their gaze during the MPDs compared to the low demanding subtasks. The scores were higher for MPDs, around ten times than NMPs, suggesting higher inattention/lower focus and looking for more information to perform the task successfully. Therefore, gaze dispersion showed to be a measure of cognitive workload. Our results are aligned with other studies using gaze position randomness (entropy) in demanding surgical tasks (Di Stasi et al., 2016; Di Stasi et al., 2017; Tien et al., 2015). As well as other studies where unexperienced performers had a more attention variability, positioned their gaze on different areas during difficult moments (KATAOKA, SASAKI, & KANDA, 2011; Sanchez, Wilson-Keates, Conway, & Zheng, 2019)

More importantly, given the difference in values between the gaze dispersion means of the NMPs and MPDs, it makes it easy to differentiate them. This has the potential to be computerized in AR to synchronize it with eye tracking. Once eye tracking can determine large gaze dispersion means from the baseline, it can indicate the AR device to send automated task instructions to the performer to assist during the task. Another observation is that during the MPDs the participants did not look at the instruments.

Increase in the gaze dispersion during complicated steps is explained by one of the main differences between experts and novices. Experts have the ability to decrease cognitive load while performing tasks. In contrast, novices lack or have less capacity to deal with stressful or new situations. During complex or difficult procedures, the individual needs to deal promptly with the information required to continue with the task, increasing the variability of the gaze position (Di Nocera et al., 2007).

Contrary to other studies (Di Stasi et al., 2017), we did not find a relation between the gaze dispersion during the MPDs and the subjective NASA TLX scores. Even though most of the participants classified the task as mentally demanding, participants did not consider the task either physical or time demanding. It could be that having unlimited time to complete the task reduces the subjectively demand of the task. Also, the NASA TLX score was provided for the whole task, not for each subtask. Perhaps, having time constriction or providing subjective feedback for each subtask would have provided a correlation between the NASA and the gaze dispersion.

In the future, comparing the gaze dispersion metric with another objective tool will be ideal. A NASA TLX for each subtask can be employed to quantify the level of validity of the gaze dispersion. Also, MPDs could be compared to qualitative data about how difficult the task was for the trainees to determine whether MPDs match the learners’ true need for instructions/help during the task.

5.5.1 Limitations.

Annotations during the video analysis were crucial for further analysis of the eye tracking data. Although the annotations were made by one operator leading to possible errors, mislabeling, and bias, the annotations were verified using a MATLAB algorithm, matching the annotations to the true location of the gaze on the surfaces which were colour-coded to visualization to decrease error.

The disadvantage of this is project is that a large amount of data obtained was required to be strictly filtered and the MPDs needed to be verified visually from the raw data videos and then using a MATLAB algorithm, which increased the time of data analysis. We hope that with the improvement of wearable eye trackers, the data filtering process will be reduced. Also, for future projects, a surface marker on the instruments panel can be added to facilitate the processes of filtering false-positive events.

5.6 Conclusion

We completed one of the fundamental steps towards an adequate instructional design for an AR eye-tracking teaching platform for chest tube insertion. In this study, the gaze dispersion was significantly larger during MPDs compared to the NMPs, demonstrating that this eye tracking metric can be used to identify an increase in cognitive workload during difficult subtasks.

Chapter 6 Discussion and Future Work

AR is a promising platform for skill training and is becoming more popular in the field of surgical education. AR offers the advantage of allowing the user to interact with an enhanced real world linked to virtual images preserving the visual-motor feedback needed during procedures. In surgery, this advantage is crucial because the contact with the real world cannot be restricted. Its utility in education has been supported by different learning theories: Situated Learning, Constructivist Learning, and Experiential Learning (see Chapter 2, pages 12--13). Also, AR head-mounted devices, such as HoloLens, offer the hands-free mixed reality interaction. However, we need to carefully consider what kind of instructional messages are displayed on the screen and most importantly at what time. Giving too much information or displaying at the wrong moment may be detrimental to the learner.

For this reason, it is important for us to identify the MPDs. Up to date, we believe metrics of eye movement can help us to search for MPDs. The main purpose of this thesis was to establish the foundation of an eye tracking enabled AR platform to facilitate the practice and learning of a widely practiced and relevant multistep surgical procedure, chest tube insertion. I am glad to report in this thesis, although challenging, that answers to the two research questions have been found and satisfied.

6.1 First Research Question.

Can eye tracking and AR devices be synchronized and integrated into one platform to be potentially used for the practice of a chest tube insertion?

I answered this question in Chapter 4. First, the hardware issues of the eye tracker camera position on the HoloLens was resolved, thereby improving the tracking accuracy and the identification of the surfaces within the field of view of the camera during the performance of the chest tube insertion task. After the devices were successfully integrated, the display of the information into the AR screen was enabled using the surface tracker in Pupil Lab.

Figure 6.1 shows two different surfaces, activating and deactivating. The virtual information is activated and displayed on the HoloLens screen only when the user looks at the activating surface (placed on the right), and it disappears as soon as the user places the gaze on the deactivating surface located at the working area to avoid interference with the task.

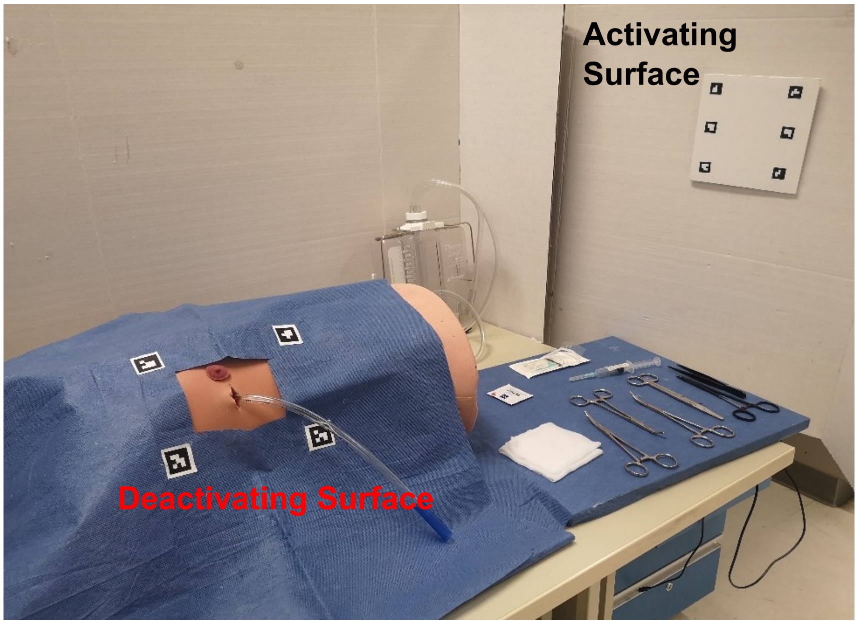

Figure 6.1. Workstation, including the chest tube model, instruments, and the surfaces placed to activate and deactivate the display of information on the AR goggles using gaze positions within the surfaces

In this chapter, we achieved the proof of concept of using gaze to activate/deactivate the virtual information display on an AR device combining them into one platform. The results of this work lay the ground for future researchers to implement eye tracking metrics such as *gaze dispersion* to automate the process of displaying instructional information in AR environments.

6.2 Second Research Question

Can eye tracking identify the moment of performance difficulty during a multistep surgical procedure, specifically a chest tube insertion?

In Chapters 3 and 5, I answered the second research question. Before developing an empirical research project to identify the MPDs during a chest tube insertion, I needed to build a simulation model to carry out our project. Finding the best simulation scenario to accurately portrayed the chest tube insertion experience is close to impossible. The next best scenario was to design a simulation model that was easy and affordable to replicate, as close as possible, as the "real" feel by experts and tailored to the learners so that human performance measurement devices could be used.

I explained, in detail, in Chapter 3, the challenges of current models in this type of simulation and why we think our model contributes to enhancing the learning experience of this life-saving procedure. The material used had a close resemblance to actual human soft tissue as evaluated by experts. Also, the silicones derived product is easy and affordable to acquire in any material store. The only challenging part was to find the correct mixture of proportion among polymers, which we hope by publishing this thesis, other groups will be able to utilize. Of course, I am aware of the model's imperfections, and I believe that as technology advances, the chest tube model can be improved to create a higher affinity model.

After having the chest tube model built, the next step to answering the second research question is explained in Chapter 5. In this chapter, the eye tracking metric *gaze dispersion* was utilized to characterize the eye behaviours during the MPDs and differentiate them from NMPs. First, the MPDs were identified using video analysis and defined as the voluntary discontinuation of the task to seek guidance to continue the task by either looking at the instructions sheet (*Instructions Check*) or asking verbally for help (*Help Requested*). Then, results from the eye tracking analysis showed that during MPDs, participants had a more dispersed gaze, which could be translated into an increase of the cognitive load during a specific part of the task. Conversely, during the NMPs, defined as the performance of the task without interruptions, trainees were more focused on the working area. As a result, the gaze dispersion was lower. These findings are consistent with prior studies. Diaz-Piedra, Sanchez-Carrion, Rieiro, and Di Stasi (2017) found that during the difficult moments of a simulated surgical procedure (ureteral calculi extraction), participants had higher gaze dispersion in the visual field. Similarly, in the Tien et al. (2015) study, trainees focused less on the surgical site as compared to more experienced participants, representing that experts can perform difficult steps with a lower cognitive load having less gaze dispersion and lower shifting to other areas.

6.3 Future Work

In this thesis, I explained the importance and challenges of learning a multistep procedure such as chest tube insertion and how the moments of performance difficulty where the cognitive load is increased can effectively be identified by eye tracking through gaze dispersion. This is the phase of identification and eye tracking measures extraction. Now, the next step is the implementation of the eye-tracking data to trigger the instructional information once the MPDs were detected. To achieve this next goal, I will postulate and discuss the necessary steps. The discussion will be focused on what type of information will be displayed, how to manage cognitive load of the learner, and the interface of prompting instructional messages.

AR instructional design.

I did some work in the first component of the AR-Eye tracking implementation by creating a platform with relevant information on the four main steps for the performance of a chest tube insertion: Landmarks, Dissection, Insertion, and Securing. Designing an instructional learning platform in AR is vital to evaluate the risk of increasingly unnecessary cognitive load.

Designing a learning platform in AR possesses several challenges. One of them is to avoid overloading the user with irrelevant information distracting them from the main task. Multimedia Learning Theory addresses the problem of increasing cognitive load on users by focusing on three main aspects that influence the learning process in instructional designs: *extraneous processing, essential processing and generative processing*. First, extraneous processing is the location of cognitive processing to items not related to the learning objective, and it is caused by a deficient design of the instructional platform (Mayer, 2014). In the instructional design, I avoid unnecessary information by selecting the crucial steps to the performance of a chest tube — for instance, identification of landmarks, dissection, insertion, and securing (Figure 6.2). The steps were chosen based on the results of the study described in Chapter 5, as they were the most difficult steps for the participants. Therefore, I will avoid overloading the participants' working memory with focusing on steps that were not essential in the performance of the task, such as cutting the gauzes to dress the wound. Also, I reduced the extraneous processing by employed images only relevant to the step the learner selects to see on the AR screen.

Second, essential processing is the cognitive resources assigned to process information relevant to achieving the learning objective, and this processing can be restricted by the complexity of the material with which the learner interacts (Mayer, 2014). We designed a platform where the information of a complex surgical procedure is presented concisely. The main steps are divided into a maximum of four substeps to avoid overloading the user's working memory. Also, each substep is represented by a picture and its corresponding brief written description (Figure 6.2).

Last, in the generative process, learners have a significant role as they integrate the information received with previous knowledge to make sense of what is being presented to construct new knowledge (Mayer, 2014). From the aspect of instructional design, I try to achieve this by creating an enticing and interactive platform. In the experimental stage of the project, learners have control of the material displayed on the AR device screen. However, in future research, the implementation of the eye-tracking data to trigger the instructional information will make the platform more attractive and motivating for learners.

Designing the interface of information delivering on the AR screen.

In the next paragraphs, I will unfold some ideas for future work, considering the known limitations of current technology. First, using AR during manual tasks has the challenge of hand gestures being either impossible to perform or obstructive and time consuming, ultimately increasing the time of completing the task. A solution to this problem is the use of gaze in the AR headsets. Lee et al. used gaze dwell and half blink as the interaction tool in an AR environment built using the AR toolkit library. Similarly, to our project, the researchers employed squares markers to identify the area of interest to display the information required (S. Lee, Jae-Young, & Choi, 2011).

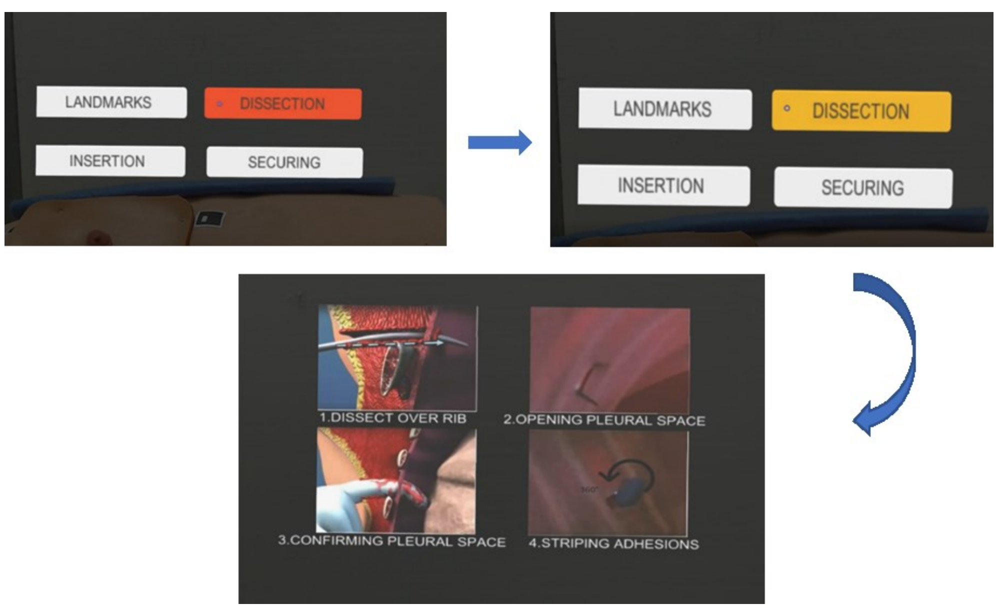

Figure 6.2. Process of information activation. After the gaze is displayed on the activating surface, a menu with four options is shown. Participants can choose with their gaze the step to review. The figure shows the activation of the dissection step.

In another study, the displaying of information on an AR headset was improved using gaze information to control the flow of information. First, small icons were displayed into the peripheral vision of the user. Then, more complex information was introduced after the user placed their gaze on the icon (Ishiguro & Rekimoto, 2011). The researchers found this method useful. However, they did not account for the Midas touch phenomenon (Chapter 2, page 25), resulting in information being display even when it was not needed. Some researchers try to avoid the problem of Midas' touch by using elaborate gaze gestures that can be implemented in the AR-user interaction such as the duration and dispersion of the gaze (Toyama et al., 2014).

Another method to use eye tracking as an input signal to trigger information into the AR headset is the neural networks training or deep learning. This method has been used already for gaze estimation in AR/VR environments (Lemley, Kar, & Corcoran, 2018). Lee et al., (2017) employed a wearable eye tracker (pupil lab) attached to an AR goggle (EPSON BT-200) and created a method to automatically turn on/off the AR device screen to avoid obstructions to the user's view. They used a machine learning algorithm for learning and prediction processes. First, the neural network was trained using the gaze data and then the machine can predict the gaze depth in real-time based on the model constructed during the learning phase with an error of 8% during gaze changing shortly after the training phase (Y. Lee, Shin, Piumsomboon, Lee, & Billinghurst, 2017). Also, the machine learning method has been employed to use eye-tracking data to modulate or interact in the augmented environment. For instance, Liu et al. employed machine learning algorithms to train a neural network to regulate the display brightness on a HoloLens using the pupil size (Liu, Plopski, Kiyokawa, Ratsamee, & Orlosky, 2018).

In summary, my supervisor and I believe that we need stronger support from people who know computer programming. As a trained physician working on this master thesis, I have limited time and capability to complete the last phase of study. My supervisor will assign the task to a student with better knowledge of computer programming. Eventually, we will develop an eye-tracking enable AR platform that will allow us to detect the MPDs in real-time and display instructional messages to learners on their AR screen at the moments they need.

References

Admoni, H., & Siddhartha, S. (2016). Predicting User Intent Through Eye Gaze for Shared Autonomy. *AAAI Fall Symposium Series*.

Ali, J., Sorvari, A., & Pandya, A. (2012). Teaching emergency surgical skills for trauma resuscitation-mechanical simulator versus animal model. *ISRN Emergency Medicine, 2012*.

Ariffin, A. C. (2018). Tube Thoracostomy Complications: More to Learn. *World Journal of Surgery, 42*(1), 310-310. doi:10.1007/s00268-017-4154-9

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Comput. Graph. Appl., 21*(6), 34-47. doi:10.1109/38.963459

B, M., #226, ce, Lepp, T., #228, nen, . . . Gomez, A. R. (2016). *ubiGaze: ubiquitous augmented reality messaging using gaze gestures*. Paper presented at the SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications, Macau.

Ball, C. G., Lord, J., Laupland, K. B., Gmora, S., Mulloy, R. H., Ng, A. K., . . . Kirkpatrick, A. W. (2007). Chest tube complications: how well are we training our residents? *Canadian journal of surgery. Journal canadien de chirurgie, 50*(6), 450-458. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/18053373

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2386217/

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2386217/pdf/20071200s00007p450.pdf

Barakonyi, I., Prendinger, H., Schmalstieg, D., & Ishizuka, M. (2007, 10-11 March 2007). *Cascading Hand and Eye Movement for Augmented Reality Videoconferencing.* Paper presented at the 2007 IEEE Symposium on 3D User Interfaces.

Barsom, E. Z., Graafland, M., & Schijven, M. P. (2016). Systematic review on the effectiveness of augmented reality applications in medical training. *Surgical Endoscopy, 30*(10), 4174-4183. doi:10.1007/s00464-016-4800-6

Bednarik, R., Bartczak, P., Vrzakova, H., Koskinen, J., Elomaa, A.-P., Huotarinen, A., . . . Fraunberg, M. v. u. z. (2018). *Pupil size as an indicator of visual-motor workload and expertise in microsurgical training tasks*. Paper presented at the Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, Warsaw, Poland.

Berguer, R., Smith, W., & Chung, Y. (2001). Performing laparoscopic surgery is significantly more stressful for the surgeon than open surgery. *Surgical Endoscopy, 15*(10), 1204-1207.

Bernhardt, S., Nicolau, S. A., Soler, L., & Doignon, C. (2017). The status of augmented reality in laparoscopic surgery as of 2016. *Medical Image Analysis, 37*, 66-90. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S1361841517300178?via%3Dihub

Bichlmeier, C., Heining, S. M., Feuerstein, M., & Navab, N. (2009). The virtual mirror: a new interaction paradigm for augmented reality environments. *IEEE Transactions on Medical Imaging, 28*(9), 1498-1510. doi:10.1109/TMI.2009.2018622

Blascheck, T., Kurzhals, K., Raschke, M., Burch, M., Weiskopf, D., & Ertl, T. (2014). *State-of-the-Art of Visualization for Eye Tracking Data.* Paper presented at the EuroVis (STARs).

Botden, S., Berlage, J., Schijven, M. P., & Jakimowicz, J. J. (2008). Face validity study of the ProMIS augmented reality laparoscopic suturing simulator. *Surgical Technology International, 17*, 26-32.

Botden, S. M., Buzink, S. N., Schijven, M. P., & Jakimowicz, J. J. (2007). Augmented versus virtual reality laparoscopic simulation: What is the difference? *World Journal of Surgery, 31*(4), 764-772.

Botden, S. M. B. I., Buzink, S. N., Schijven, M. P., & Jakimowicz, J. J. (2007). Augmented versus virtual reality laparoscopic simulation: what is the difference? A comparison of the ProMIS augmented reality laparoscopic simulator versus LapSim virtual reality laparoscopic simulator. *World Journal of Surgery, 31*(4), 764-772. doi:10.1007/s00268-006-0724-y

Botden, S. M. B. I., & Jakimowicz, J. J. (2009). What is going on in augmented reality simulation in laparoscopic surgery? *Surgical Endoscopy, 23*(8), 1693-1700. doi:10.1007/s00464-008-0144-1

Broe, D., Ridgway, P., Johnson, S., Tierney, S., & Conlon, K. (2006). Construct validation of a novel hybrid surgical simulator. *Surgical Endoscopy And Other Interventional Techniques, 20*(6), 900-904.

Cao, A., Chintamani, K. K., Pandya, A. K., & Ellis, R. D. (2009). NASA TLX: Software for assessing subjective mental workload. *Behavior Research Methods, 41*(1), 113-117. doi:10.3758/brm.41.1.113

Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., & Ivkovic, M. (2011). Augmented reality technologies, systems and applications. *Multimed. Tools Appl., 51*(1), 341-377. doi:10.1007/s11042-010-0660-6

Carter, B. N. (1952). The fruition of Halsted's concept of surgical training. *Surgery, 32*(3), 518-527. doi:10.5555/uri:pii:0039606052902195

Causer, J., Vickers, J. N., Snelgrove, R., Arsenault, G., & Harvey, A. (2014). Performing under pressure: Quiet eye training improves surgical knot-tying performance. *Surgery, 156*(5), 1089-1096. doi:https://doi.org/10.1016/j.surg.2014.05.004

Cheng, K.-H., & Tsai, C.-C. (2012). Affordances of Augmented Reality in Science Learning: Suggestions for Future Research. *J. Sci. Educ. Technol., 22*(4), 449-462. doi:10.1007/s10956-012-9405-9

Cheng, K.-H., & Tsai, C.-C. (2013). Affordances of augmented reality in science learning: Suggestions for future research. *Journal of science education and technology, 22*(4), 449-462.

Chung, T. N., Kim, S. W., You, J. S., & Chung, H. S. (2016). Tube thoracostomy training with a medical simulator is associated with faster, more successful performance of the procedure. *Clinical and experimental emergency medicine, 3*(1), 16. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5051624/pdf/ceem-15-097.pdf

Coral, M. P. (2016). Analyzing Cognitive Workload Through Eye-related Measurements: A Meta-Analysis.

Cutolo, F., Carbone, M., Parchi, P. D., Ferrari, V., Lisanti, M., & Ferrari, M. (2016). *Application of a New Wearable Augmented Reality Video See-Through Display to Aid Percutaneous Procedures in Spine Surgery*. Paper presented at the Augmented Reality, Virtual Reality, and Computer Graphics. https://link.springer.com/chapter/10.1007/978-3-319-40651-0_4

https://link.springer.com/content/pdf/10.1007%2F978-3-319-40651-0_4.pdf

http://dx.doi.org/10.1007/978-3-319-40651-0_4
Cutting, J. E., & Vishton, P. M. (1995). Perceiving Layout and Knowing Distances. In *Perception of Space and Motion* (pp. 69-117): Elsevier.
Dev, S. P., Nascimiento Jr, B., Simone, C., & Chien, V. (2007). Chest-tube insertion. *New England Journal of Medicine, 357*(15), e15. Retrieved from https://www.nejm.org/doi/pdf/10.1056/NEJMvcm071974?articleTools=true
Di Nocera, F., Camilli, M., & Terenzi, M. (2007). A random glance at the flight deck: Pilots' scanning strategies and the real-time assessment of mental workload. *Journal of Cognitive Engineering and Decision Making, 1*(3), 271-285.
Di Stasi, L. L., Diaz-Piedra, C., Rieiro, H., Sánchez Carrión, J. M., Martin Berrido, M., Olivares, G., & Catena, A. (2016). Gaze entropy reflects surgical task load. *Surgical Endoscopy, 30*(11), 5034-5043. doi:10.1007/s00464-016-4851-8
Di Stasi, L. L., Díaz-Piedra, C., Ruiz-Rabelo, J. F., Rieiro, H., Sanchez Carrion, J. M., & Catena, A. (2017). Quantifying the cognitive cost of laparo-endoscopic single-site surgeries: Gaze-based indices. *Applied Ergonomics, 65*, 168-174. doi:https://doi.org/10.1016/j.apergo.2017.06.008
Di Stasi, L. L., McCamy, M. B., Macknik, S. L., Mankin, J. A., Hooft, N., Catena, A., & Martinez-Conde, S. (2014). Saccadic eye movement metrics reflect surgical residents' fatigue. *Annals of Surgery, 259*(4), 824-829. doi:10.1097/SLA.0000000000000260
Dias, R. D., Ngo-Howard, M. C., Boskovski, M. T., Zenati, M. A., & Yule, S. J. (2018). Systematic review of measurement tools to assess surgeons' intraoperative cognitive workload. *British Journal of Surgery, 105*(5), 491-501. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5878696/pdf/nihms921855.pdf
Diaz-Piedra, C., Sanchez-Carrion, J. M., Rieiro, H., & Di Stasi, L. L. (2017). Gaze-based technology as a tool for surgical skills assessment and training in urology. *Urology, 107*, 26-30.
Dickey, R. M., Srikishen, N., Lipshultz, L. I., Spiess, P. E., Carrion, R. E., & Hakky, T. S. (2016). Augmented reality assisted surgery: a urologic training tool. *Asian journal of andrology, 18*(5), 732.
Dixon, B. J., Daly, M. J., Chan, H., Vescan, A. D., Witterick, I. J., & Irish, J. C. (2013). Surgeons blinded by enhanced navigation: the effect of augmented reality on attention. *Surgical Endoscopy, 27*(2), 454-461. doi:10.1007/s00464-012-2457-3
Duchowski, A. (2007). *Eye Tracking Methodology: Theory and Practice*: Springer.
Dunleavy, M., & Dede, C. (2014). Augmented Reality Teaching and Learning. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (pp. 735-745). New York, NY: Springer New York.
Eddy, S. R. (2004). What is a hidden Markov model? *Nature Biotechnology, 22*(10), 1315.
Eid, M. A., Giakoumidis, N., & El Saddik, A. (2016). A novel eye-gaze-controlled wheelchair system for navigating unknown environments: case study with a person with ALS. *IEEE Access, 4*, 558-573.
El-Hariri, H., Pandey, P., Hodgson, A. J., & Garbi, R. (2018). Augmented reality visualisation for orthopaedic surgical guidance with pre- and intra-operative multimodal image data fusion. *Healthcare Technology Letters, 5*(5), 189-193. doi:10.1049/htl.2018.5061
Esposito, M., Busam, B., Hennersperger, C., Rackerseder, J., Lu, A., Navab, N., & Frisch, B. (2015). *Cooperative robotic gamma imaging: Enhancing us-guided needle biopsy.* Paper

presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention.

Frank, J. R., & Danoff, D. (2007). The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. *Medical Teacher, 29*(7), 642-647. doi:10.1080/01421590701746983

Frank, J. R., Snell, L. S., Cate, O. T., Holmboe, E. S., Carraccio, C., Swing, S. R., . . . Harris, K. A. (2010). Competency-based medical education: theory to practice. *Medical Teacher, 32*(8), 638-645. doi:10.3109/0142159X.2010.501190

Fry, H. (2011). Educational Ideas and Surgical Education. In *Advances in Medical Education* (pp. 19-36).

Furlan, R. (2016). The future of augmented reality: Hololens - Microsoft's AR headset shines despite rough edges [Resources_Tools and Toys]. *IEEE Spectrum, 53*(6), 21-21. doi:10.1109/MSPEC.2016.7473143

Gallagher, A. G., & O'Sullivan, G. C. (2011). *Fundamentals of Surgical Simulation: Principles and Practice*: Springer Science & Business Media.

Gallagher, A. G., O'Sullivan, G. C., & O'Sullivan, G. C. (2012). Agents of Change. In *Fundamentals of Surgical Simulation: Principles and Practice* (pp. 1-38). London: Springer London.

Ghazali, A., Breque, C., Léger, A., Scépi, M., & Oriot, D. (2015). Testing of a complete training model for chest tube insertion in traumatic pneumothorax. *Simulation in Healthcare, 10*(4), 239-244.

Goldman, L. W. (2007). Principles of CT and CT technology. *Journal of Nuclear Medicine Technology, 35*(3), 115-128; quiz 129-130. doi:10.2967/jnmt.107.042978

Hansen, C., Wieferich, J., Ritter, F., Rieder, C., & Peitgen, H.-O. (2010). Illustrative visualization of 3D planning models for augmented reality in liver surgery. *International Journal of Computer Assisted Radiology and Surgery, 5*(2), 133-141. doi:10.1007/s11548-009-0365-3

Hart, S. G. (1986). NASA Task load Index (TLX). Volume 1.0; Paper and pencil package.

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(9), 904-908. doi:10.1177/154193120605000909

Henderson, S., & Feiner, S. (2010). Exploring the benefits of augmented reality documentation for maintenance and repair. *IEEE transactions on visualization and computer graphics, 17*(10), 1355-1368.

Hermens, F., Flin, R., & Ahmed, I. (2013). Eye movements in surgery: A literature review.

Hernandez, M. C., Zeb, M. H., Heller, S. F., Zielinski, M. D., & Aho, J. M. (2017). Tube Thoracostomy Complications Increase Cost. *World Journal of Surgery, 41*(6), 1482-1487. doi:10.1007/s00268-017-3897-7

Hishikawa, S., Kawano, M., Tanaka, H., Konno, K., Yasuda, Y., Kawano, R., . . . Lefor, A. T. (2010). Mannequin simulation improves the confidence of medical students performing tube thoracostomy: a prospective, controlled trial. *The American surgeon, 76*(1), 73-78.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*: OUP Oxford.

Ishiguro, Y., & Rekimoto, J. (2011). *Peripheral vision annotation: noninterference information presentation method for mobile augmented reality*. Paper presented at the Proceedings of the 2nd Augmented Human International Conference, Tokyo, Japan.

Jacob, R. J. (1993). Eye movement-based human-computer interaction techniques: Toward non-command interfaces. *Advances in human-computer interaction, 4*, 151-190.

Jacob, R. J. K. (1995). *Eye Tracking in Advanced Interface Design*. Paper presented at the In W. Barfield & T. A. Furness (Eds.), Virtual Environments and Advanced Interface Design. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.50.6977

Jacob, R. J. K., & Karn, K. S. (2003). Eye Tracking in Human-Computer Interaction and Usability Research. In *The Mind's Eye* (pp. 573-605).

Jiang, X., Atkins, M. S., Tien, G., Zheng, B., & Bednarik, R. (2014). *Pupil dilations during target-pointing respect Fitts' law.* Paper presented at the Proceedings of the Symposium on Eye Tracking Research and Applications.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329.

Kamphuis, C., Barsom, E., Schijven, M., & Christoph, N. (2014). Augmented reality in medical education? *Perspect. Med. Educ., 3*(4), 300-311. doi:10.1007/s40037-013-0107-7

Kashani, P., Harati, S., Shirafkan, A., Amirbeigi, A., & Hatamabadi, H. R. (2017). Comparing the quality and complications of tube thoracostomy by emergency medicine and surgery residents; a cohort study. *Emergency, 5*(1).

Kassner, M., Patera, W., & Bulling, A. (2014). *Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction.* Paper presented at the Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication.

KATAOKA, J., SASAKI, M., & KANDA, K. (2011). Effects of mental workload on nurses' visual behaviors during infusion pump operation. *Japan Journal of Nursing Science, 8*(1), 47-56. doi:10.1111/j.1742-7924.2010.00158.x

Katić, D., Sudra, G., Speidel, S., Castrillon-Oberndorfer, G., Eggers, G., & Dillmann, R. (2010). *Knowledge-based situation interpretation for context-aware augmented reality in dental implant surgery.* Paper presented at the International Workshop on Medical Imaging and Virtual Reality.

Katić, D., Wekerle, A.-L., Görtler, J., Spengler, P., Bodenstedt, S., Röhl, S., . . . Speidel, S. (2013). Context-aware Augmented Reality in laparoscopic surgery. *Computerized Medical Imaging and Graphics, 37*(2), 174-182. doi:https://doi.org/10.1016/j.compmedimag.2013.03.003

Kersten-Oertel, M., Jannin, P., & Collins, D. L. (2012). DVV: a taxonomy for mixed reality visualization in image guided surgery. *IEEE Trans. Vis. Comput. Graph., 18*(2), 332-352. doi:10.1109/TVCG.2011.50

Khor, W. S., Baker, B., Amin, K., Chan, A., Patel, K., & Wong, J. (2016). Augmented and virtual reality in surgery-the digital surgical environment: applications, limitations and legal pitfalls. *Ann Transl Med, 4*(23), 454. doi:10.21037/atm.2016.12.23

Kneebone, R. (2003). Simulation in surgical training: educational issues and practical implications. *Medical Education, 37*(3), 267-277. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12603766

http://onlinelibrary.wiley.com/resolve/openurl?genre=article&sid=nlm:pubmed&issn=0308-0110&date=2003&volume=37&issue=3&spage=267

Kodappully, M., Srinivasan, B., & Srinivasan, R. (2016). Towards predicting human error: Eye gaze analysis for identification of cognitive steps performed by control room operators.

Journal of Loss Prevention in the Process Industries, 42, 35-46. doi:10.1016/j.jlp.2015.07.001
Kolb, D. A. Experiential learning : experience as the source of learning and development. In (Second edition. ed.).
Kolb, D. A. (2014). *Experiential learning: Experience as the source of learning and development*: FT press.
Kosch, T., Hassib, M., Buschek, D., & Schmidt, A. (2018). *Look into my Eyes: Using Pupil Dilation to Estimate Mental Workload for Task Complexity Adaptation*. Paper presented at the Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal QC, Canada.
Kottayil, N. K., Bogdanova, R., Cheng, I., Zheng, B., & Basu, A. (2016). *Investigation of gaze patterns in multi view laparoscopic surgery.* Paper presented at the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).
Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C., & Krejtz, I. (2018). Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PloS One, 13*(9), e0203629. doi:10.1371/journal.pone.0203629
Kyt, M., #246, Ens, B., Piumsomboon, T., Lee, G. A., & Billinghurst, M. (2018). *Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality*. Paper presented at the Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal QC, Canada.
Lahanas, V., Georgiou, E., & Loukas, C. (2016). Surgical simulation training systems: box trainers, virtual reality and augmented reality simulators. *International Journal of Advanced Robotics and Automation, 1*(2), 1-9.
Lahanas, V., Loukas, C., Smailis, N., & Georgiou, E. (2015). A novel augmented reality simulator for skills assessment in minimal invasive surgery. *Surgical Endoscopy, 29*(8), 2224-2234. doi:10.1007/s00464-014-3930-y
Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41*(25-26), 3559-3565. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/11718795

https://linkinghub.elsevier.com/retrieve/pii/S0042-6989(01)00102-X
Langbehn, E., Steinicke, F., Lappe, M., Welch, G. F., & Bruder, G. (2018). In the blink of an eye: leveraging blink-induced suppression for imperceptible position and orientation redirection in virtual reality. *ACM Transactions on Graphics (TOG), 37*(4), 66.
Larson, G. E., & Perry, Z. A. (1999). Visual capture and human error. *Applied Cognitive Psychology, 13*(3), 227-236. doi:10.1002/(sici)1099-0720(199906)13:3<227::aid-acp563>3.0.co;2-j
Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*: Cambridge university press.
Law, B., Atkins, M. S., Kirkpatrick, A. E., & Lomax, A. J. (2004). *Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment*. Paper presented at the Proceedings of the Eye tracking research & applications symposium on Eye tracking research & applications - ETRA'2004. http://portal.acm.org/citation.cfm?doid=968363.968370

http://dl.acm.org/ft_gateway.cfm?id=968370&ftid=253185&dwn=1

http://dx.doi.org/10.1145/968363.968370

Law, B., Stella Atkins, M., Kirkpatrick, A. E., & Lomax, A. J. (2004). *Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment*. Paper presented at the Proceedings of the Eye tracking research & applications symposium on Eye tracking research & applications - ETRA'2004. http://dx.doi.org/10.1145/968363.968370

Law, B. H. Y., Cheung, P.-Y., Wagner, M., van Os, S., Zheng, B., & Schmölzer, G. (2018). Analysis of neonatal resuscitation using eye tracking: a pilot study. *Archives of Disease in Childhood-Fetal and Neonatal Edition, 103*(1), F82-F84.

Lee, S., Jae-Young, L., & Choi, J. (2011, 9-11 Feb. 2011). *Design and implementation of an interactive HMD for wearable AR system.* Paper presented at the 2011 17th Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV).

Lee, Y., Shin, C., Piumsomboon, T., Lee, G., & Billinghurst, M. (2017). *Automated enabling of head mounted display using gaze-depth estimation*. Paper presented at the SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications, Bangkok, Thailand.

Lemley, J., Kar, A., & Corcoran, P. (2018, 15-17 Aug. 2018). *Eye Tracking in Augmented Spaces: A Deep Learning Approach.* Paper presented at the 2018 IEEE Games, Entertainment, Media Conference (GEM).

Liu, C., Plopski, A., Kiyokawa, K., Ratsamee, P., & Orlosky, J. (2018, 16-20 Oct. 2018). *IntelliPupil: Pupillometric Light Modulation for Optical See-Through Head-Mounted Displays.* Paper presented at the 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR).

Loukas, C. (2016). Surgical Simulation Training Systems: Box Trainers, Virtual Reality and Augmented Reality Simulators. *International Journal of Advanced Robotics and Automation, 1*(2), 1-9. doi:10.15226/2473-3032/1/2/00109

Louw, T., & Merat, N. (2017). Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation. *Transportation Research Part C: Emerging Technologies, 76*, 35-50. doi:https://doi.org/10.1016/j.trc.2017.01.001

Lovo, E. E., Quintana, J. C., Puebla, M. C., Torrealba, G., Santos, J. L., Lira, I. H., & Tagle, P. (2007). A novel, inexpensive method of image coregistration for applications in image-guided surgery using augmented reality. *Neurosurgery, 60*(4 Suppl 2), 366-371; discussion 371-362. doi:10.1227/01.NEU.0000255360.32689.FA

Ma, M., Fallavollita, P., Seelbach, I., Von Der Heide, A. M., Euler, E., Waschke, J., & Navab, N. (2016). Personalized augmented reality for anatomy education. *Clinical Anatomy, 29*(4), 446-453. doi:10.1002/ca.22675

Marquard, J. L., Henneman, P. L., He, Z., Jo, J., Fisher, D. L., & Henneman, E. A. (2011). Nurses' behaviors and visual scanning patterns may reduce patient identification errors. *Journal of Experimental Psychology: Applied, 17*(3), 247-256. doi:10.1037/a0025261

Marquart, G., Cabrall, C., & de Winter, J. (2015). Review of Eye-related Measures of Drivers' Mental Workload. *Procedia Manufacturing, 3*, 2854-2861. doi:https://doi.org/10.1016/j.promfg.2015.07.783

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition, 1*(1).

Mayer, R. E. (2014). Cognitive Theory of Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2 ed., pp. 43-71). Cambridge: Cambridge University Press.

McNamara, A., Boyd, K., Oh, D., Sharpe, R., & Suther, A. (2018, 16-20 Oct. 2018). *Using Eye Tracking to Improve Information Retrieval in Virtual Reality.* Paper presented at the 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct).

Meißner, M., Pfeiffer, J., Pfeiffer, T., & Oppewal, H. (2017). Combining virtual reality and mobile eye tracking to provide a naturalistic experimental environment for shopper research. *J. Bus. Res.* doi:10.1016/j.jbusres.2017.09.028

Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1995). *Augmented reality: a class of displays on the reality-virtuality continuum*. Paper presented at the Telemanipulator and Telepresence Technologies. http://dx.doi.org/10.1117/12.197321

Moonesinghe, S. R., Lowery, J., Shahi, N., Millen, A., & Beard, J. D. (2011). Impact of reduction in working hours for doctors in training on postgraduate medical education and patients' outcomes: systematic review. *BMJ, 342*, d1580. doi:10.1136/bmj.d1580

Navab, N., Traub, J., Sielhorst, T., Feuerstein, M., & Bichlmeier, C. (2007). Action- and Workflow-Driven Augmented Reality for Computer-Aided Medical Procedures. *IEEE Computer Graphics and Applications, 27*(5), 10-14. doi:10.1109/MCG.2007.117

Neggers, S. F., & Bekkering, H. (2000). Ocular gaze is anchored to the target of an ongoing pointing movement. *Journal of Neurophysiology, 83*(2), 639-651. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/10669480

http://jn.physiology.org/cgi/pmidlookup?view=long&pmid=10669480

Nicolau, S., Soler, L., Mutter, D., & Marescaux, J. (2011). Augmented reality in laparoscopic surgical oncology. *Surgical Oncology, 20*(3), 189-201. doi:10.1016/j.suronc.2011.07.002

Olk, B., & Kappas, A. (2011). Eye tracking as a tool for visual research. *Handbook of visual research methods*, 433-451.

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*(4), 351-371. doi:10.1007/bf02213420

Patel, R., & Zurca, A. J. C. C. M. (2018). 396: I SEE WHAT YOU SEE ADDING EYE-TRACKING TO MEDICAL SIMULATION. *46*(1), 181.

Pelargos, P. E., Nagasawa, D. T., Lagman, C., Tenn, S., Demos, J. V., Lee, S. J., . . . Ung, N. (2017). Utilizing virtual and augmented reality for educational and clinical enhancements in neurosurgery. *Journal of Clinical Neuroscience, 35*, 1-4.

Pinzon, D., Byrns, S., & Zheng, B. (2016). Prevailing Trends in Haptic Feedback Simulation for Minimally Invasive Surgery. *Surgical Innovation, 23*(4), 415-421. doi:10.1177/1553350616628680

Prada, F., Del Bene, M., Mattei, L., Casali, C., Filippini, A., Legnani, F., . . . DiMeco, F. (2014). Fusion imaging for intra-operative ultrasound-based navigation in neurosurgery. *J. Ultrasound, 17*(3), 243-251. doi:10.1007/s40477-014-0111-8

Pratt, P., Mayer, E., Vale, J., Cohen, D., Edwards, E., Darzi, A., & Yang, G.-Z. (2012). An effective visualisation and registration system for image-guided robotic partial nephrectomy. *Journal of Robotic Surgery, 6*(1), 23-31. doi:10.1007/s11701-011-0334-z

Prietz, P., Sae-Tan, N., Yang, W., & Patera, W. (2018). Interprocess and Network Communication Retrieved from https://github.com/pupil-labs/pupil-docs/blob/master/developer-docs/ipc-backbone.md

Reason, J. (1995). Understanding adverse events: human factors. *Quality & Safety in Health Care, 4*(2), 80-89. doi:10.1136/qshc.4.2.80

Reichelt, S., Häussler, R., Fütterer, G., & Leister, N. (2010). *Depth cues in human visual perception and their realization in 3D displays*. Paper presented at the Three-Dimensional Imaging, Visualization, and Display 2010 and Display Technologies and Applications for Defense, Security, and Avionics IV. http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.850094

http://dx.doi.org/10.1117/12.850094

Reimer, B. (2009). Impact of Cognitive Task Complexity on Drivers' Visual Tunneling. *Transportation Research Record, 2138*(1), 13-19. doi:10.3141/2138-03

Renner, P., & Pfeiffer, T. (2017, 18-19 March 2017). *Attention guiding techniques using peripheral vision and eye tracking for feedback in augmented-reality-based assistance systems.* Paper presented at the 2017 IEEE Symposium on 3D User Interfaces (3DUI).

Richstone, L., Schwartz, M. J., Seideman, C., Cadeddu, J., Marshall, S., & Kavoussi, L. R. (2010). Eye metrics as an objective assessment of surgical skill. *Annals of Surgery, 252*(1), 177-182. doi:10.1097/SLA.0b013e3181e464fb

Ritter, E. M., Kindelan, T. W., Michael, C., Pimentel, E. A., & Bowyer, M. W. (2007). Concurrent validity of augmented reality metrics applied to the fundamentals of laparoscopic surgery (FLS). *Surgical Endoscopy, 21*(8), 1441-1445. doi:10.1007/s00464-007-9261-5

Sachdeva, A. K. (2007). The Changing Paradigm of Residency Education in Surgery: A Perspective from the American College of Surgeons. *The American Surgeon, 73*(2), 120-129. Retrieved from https://www.ingentaconnect.com/content/sesc/tas/2007/00000073/00000002/art00005

Sanchez, Y. P., Wilson-Keates, B., Conway, A., & Zheng, B. (2019). Gaze Performance Adjustment During Needlestick Application: Can We Reduce Harm? *Nurse Educator, 44*(2), E1-E5.

Satava, M. D. R. M., & Satava, R. M. (2001). Surgical Education and Surgical Simulation. *World Journal of Surgery, 25*(11), 1484-1489. doi:10.1007/s00268-001-0134-0

Sato, Y., Nakamoto, M., Tamaki, Y., Sasama, T., Sakita, I., Nakajima, Y., . . . Tamura, S. (1998). Image guidance of breast cancer surgery using 3-D ultrasound images and augmented reality visualization. *IEEE Transactions on Medical Imaging, 17*(5), 681-693. doi:10.1109/42.736019

Savage, S. W., Potter, D. D., & Tatler, B. W. (2013). Does preoccupation impair hazard perception? A simultaneous EEG and eye tracking study. *Transportation research part F: traffic psychology and behaviour, 17*, 52-62.

Shenai, M. B., Dillavou, M., Shum, C., Ross, D., Tubbs, R. S., Shih, A., & Guthrie, B. L. (2011). Virtual interactive presence and augmented reality (VIPAR) for remote surgical assistance. *Neurosurgery, 68*(1 Suppl Operative), 200-207; discussion 207. doi:10.1227/NEU.0b013e3182077efd

Shuhaiber, J. H. (2004). Augmented Reality in Surgery. *Archives of Surgery, 139*(2), 170. doi:10.1001/archsurg.139.2.170

Sigrist, R., Rauter, G., Riener, R., & Wolf, P. (2013). Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychon. Bull. Rev., 20*(1), 21-53. doi:10.3758/s13423-012-0333-8

Skinner, T. A. A., Ho, L., & Touma, N. J. (2017). Study habits of Canadian urology residents: Implications for development of a competence by design curriculum.(ORIGINAL

RESEARCH)(Report). *Canadian Urological Association Journal (CUAJ)*(3), 83. doi:10.5489/cuaj.4132

Sodergren, M. H., Orihuela-Espina, F., Clark, J., Darzi, A., & Yang, G.-Z. (2010). A hidden markov model-based analysis framework using eye-tracking data to characterise re-orientation strategies in minimally invasive surgery. *Cognitive Processing, 11*(3), 275-283.

Sommerauer, P., & Müller, O. (2018). *Augmented Reality for Teaching and Learning-a literature Review on Theoretical and Empirical Foundations.* Paper presented at the ECIS.

Stefanidis, D., Sevdalis, N., Paige, J., Zevin, B., Aggarwal, R., Grantcharov, T., . . . Association for Surgical Education Simulation, C. (2015). Simulation in surgery: what's needed next? *Annals of Surgery, 261*(5), 846-853. doi:10.1097/SLA.0000000000000826

Steil, J., Müller, P., Sugano, Y., & Bulling, A. (2018). *Forecasting user attention during everyday mobile interactions using device-integrated and wearable sensors.* Paper presented at the Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services.

Stoyanov, D., Mylonas, G. P., Lerotic, M., Chung, A. J., & Yang, G. (2008). Intra-Operative Visualizations: Perceptual Fidelity and Human Factors. *Journal of Display Technology, 4*(4), 491-501. doi:10.1109/JDT.2008.926497

Su, D., Li, Y., & Xiong, C. (2016, 6-10 June 2016). *Parallax error compensation for head-mounted gaze trackers based on binocular data.* Paper presented at the 2016 IEEE International Conference on Real-time Computing and Robotics (RCAR).

Surgical Education. (2011). (H. Fry & R. Kneebone Eds. Vol. 2). Dordrecht: Springer Netherlands.

Tang, A., Owen, C., Biocca, F., & Mou, W. (2003). *Comparative effectiveness of augmented reality in object assembly*. Paper presented at the Proceedings of the conference on Human factors in computing systems - CHI '03. http://dx.doi.org/10.1145/642611.642626

Tanizaki, S., Maeda, S., Sera, M., Nagai, H., Hayashi, M., Azuma, H., . . . Ishida, H. (2017). Small tube thoracostomy (20–22 Fr) in emergent management of chest trauma. *Injury, 48*(9), 1884-1887.

Tatli, O., Turkmen, S., Imamoglu, M., Karaca, Y., Cicek, M., Yadigaroglu, M., . . . Turedi, S. (2017). A novel method for improving chest tube insertion skills among medical interns: Using biomaterial-covered mannequin. *Saudi Medical Journal, 38*(10), 1007. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5694633/pdf/SaudiMedJ-38-1007.pdf

Ten Cate, O. (2013). Competency-based education, entrustable professional activities, and the power of language. *Journal of Graduate Medical Education, 5*(1), 6-7. doi:10.4300/JGME-D-12-00381.1

Tien, T., Pucher, P. H., Sodergren, M. H., Sriskandarajah, K., Yang, G.-Z., & Darzi, A. (2015). Differences in gaze behaviour of expert and junior surgeons performing open inguinal hernia repair. *Surgical Endoscopy, 29*(2), 405-413. doi:10.1007/s00464-014-3683-7

Toyama, T., Dengel, A., Suzuki, W., & Kise, K. (2013). *Wearable Reading Assist System: Augmented Reality Document Combining Document Retrieval and Eye Tracking*. Paper presented at the 2013 12th International Conference on Document Analysis and Recognition. http://ieeexplore.ieee.org/document/6628580/

http://xplorestaging.ieee.org/ielx7/6627713/6628563/06628580.pdf?arnumber=6628580 http://dx.doi.org/10.1109/ICDAR.2013.15

https://ieeexplore.ieee.org/ielx7/6627713/6628563/06628580.pdf?tp=&arnumber=6628580&isnumber=6628563&ref=

Toyama, T., Sonntag, D., Dengel, A., Matsuda, T., Iwamura, M., & Kise, K. (2014). *A mixed reality head-mounted text translation system using eye gaze input*. Paper presented at the Proceedings of the 19th international conference on Intelligent User Interfaces, Haifa, Israel.

Van Sickle, K. R., McClusky III, D., Gallagher, A., & Smith, C. (2005). Construct validation of the ProMIS simulator using a novel laparoscopic suturing task. *Surgical Endoscopy and Other Interventional Techniques, 19*(9), 1227-1231.

Vickers, J. N. (1996). Visual control when aiming at a far target. *Journal of Experimental Psychology: Human Perception and Performance, 22*(2), 342.

Vine, S. J., Masters, R. S., McGrath, J. S., Bright, E., & Wilson, M. R. (2012). Cheating experience: Guiding novices to adopt the gaze strategies of experts expedites the learning of technical laparoscopic skills. *Surgery, 152*(1), 32-40.

Vine, S. J., Masters, R. S. W., McGrath, J. S., Bright, E., & Wilson, M. R. (2012). Cheating experience: Guiding novices to adopt the gaze strategies of experts expedites the learning of technical laparoscopic skills. *Surgery, 152*(1), 32-40. doi:10.1016/j.surg.2012.02.002

Vine, S. J., Moore, L. J., & Wilson, M. R. (2014). Quiet eye training: The acquisition, refinement and resilient performance of targeting skills. *European Journal of Sport Science, 14*(sup1), S235-S242.

Wade, N. J. (2010). Pioneers of eye movement research. *i-Perception, 1*(2), 33-68.

Wang, X., Watts, I., & Zheng, B. (2017). Mapping three-dimensional digital model to surgical site in facial surgery. *Digital Medicine, 3*(1), 39.

Welchman, A. E., Deubelius, A., Conrad, V., Bülthoff, H. H., & Kourtzi, Z. (2005). 3D shape perception from combined depth cues in human visual cortex. *Nature Neuroscience, 8*(6), 820-827. doi:10.1038/nn1461

Wilson, M. R., Vine, S. J., Bright, E., Masters, R. S., Defriend, D., & McGrath, J. S. (2011). Gaze training enhances laparoscopic technical skill acquisition and multi-tasking performance: a randomized, controlled study. *Surgical Endoscopy, 25*(12), 3731-3739.

Wu, J.-R., Wang, M.-L., Liu, K.-C., Hu, M.-H., & Lee, P.-Y. (2014). Real-time advanced spinal surgery via visible patient model and augmented reality system. *Computer Methods and Programs in Biomedicine, 113*(3), 869-881. doi:10.1016/j.cmpb.2013.12.021

Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychon. Bull. Rev., 9*(2), 185-211. doi:10.3758/BF03196276

Yarbus, A. (1967). Eye Movements and Vision (B. Haigh, Trans.) Plenum Press. *New York*.

Zeller, M. (23 Feb 2019). "Gestures - Mixed Reality." Mixed Reality Retrieved from docs.microsoft.com/en-us/windows/mixed-reality/gestures

Zhang, J., Liu, S., Feng, Q., Gao, J., Cheng, J., Jiang, M., . . . Zhang, Q. (2018). Ergonomic Assessment of the Mental Workload Confronted by Surgeons during Laparoscopic Surgery. *The American Surgeon, 84*(9), 1538-1543. Retrieved from https://www.ingentaconnect.com/content/sesc/tas/2018/00000084/00000009/art00064

Zheng, B., Cassera, M. A., Martinec, D. V., Spaun, G. O., & Swanström, L. L. (2010). Measuring mental workload during the performance of advanced laparoscopic tasks. *Surgical Endoscopy, 24*(1), 45-50. doi:10.1007/s00464-009-0522-3

Zheng, B., Jiang, X., & Atkins, M. S. (2015). Detection of Changes in Surgical Difficulty: Evidence From Pupil Responses. *Surgical Innovation, 22*(6), 629-635. doi:10.1177/1553350615573582

Zheng, B., Jiang, X., Tien, G., Meneghetti, A., Panton, O. N. M., & Atkins, M. S. (2012). Workload assessment of surgeons: correlation between NASA TLX and blinks. *Surgical Endoscopy, 26*(10), 2746-2750.

Zheng, B., Tien, G., Atkins, S. M., Swindells, C., Tanin, H., Meneghetti, A., . . . Panton, O. N. M. (2011). Surgeon's vigilance in the operating room. *The American Journal of Surgery, 201*(5), 673-677.

Appendix A. Chest Tube Insertion Model: Experts Feedback

1. How many chest tube insertions have you performed: ___________________________________

2. Please rate the following features of the chest tube model based on the similarity to real patients.

	1 = Not Very similar	**2 = Not similar**	**3 = Neutral**	**4 = Somewhat similar**	**5= Very similar**
Anatomical landmarks					
Skin incision					
Dissection of tissues					
Identification of the pleural space					
Tube insertion to the pleural space					
Chest drain fixation					
Connection to the drainage system					
Overall procedure					

3. Do you think this model is useful for trainees to practise a chest tube insertion?

1 = Not useful at all	**2 = Not useful**	**3 = Neutral**	**4 = Somewhat useful**	**5 = Very useful**

4. Please provide below any recommendations to improve the model or any additional comments.

____

____

____

____