TOWARDS UNDERSTANDING EXPERT EYE BEHAVIORS IN LAPAROSCOPIC SURGERY

by

Geoffrey Tien
B.Sc., University of British Columbia, 2005
M.Sc., Simon Fraser University, 2009

a Thesis submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

in the
School of Computing Science
Faculty of Applied Sciences

SIMON FRASER UNIVERSITY
Spring 2015

APPROVAL

Name:	Geoffrey Tien
Degree:	Doctor of Philosophy
Title:	Towards Understanding Expert Eye Behaviors in Laparoscopic Surgery
Examining Committee:	**Chair:** Dr. Ze-Nian Li Professor

Dr. M. Stella Atkins
Professor
School of Computing Science
Simon Fraser University
Senior Supervisor

Dr. Arthur Kirkpatrick
Associate Professor
School of Computing Science
Simon Fraser University
Supervisor

Dr. Bin Zheng
Endowed Research Chair
Department of Surgery
University of Alberta
Supervisor

Dr. Lyn Bartram
Associate Professor
School of Interactive Arts and Technology
Simon Fraser University
Internal Examiner

Dr. Andrew T. Duchowski
Professor
School of Computing
Clemson University
External Examiner

Date Approved: February 17, 2015

Partial Copyright Licence

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the non-exclusive, royalty-free right to include a digital copy of this thesis, project or extended essay[s] and associated supplemental files ("Work") (title[s] below) in Summit, the Institutional Research Repository at SFU. SFU may also make copies of the Work for purposes of a scholarly or research nature; for users of the SFU Library; or in response to a request from another library, or educational institution, on SFU's own behalf or for one of its users. Distribution may be in any form.

The author has further agreed that SFU may keep more than one copy of the Work for purposes of back-up and security; and that SFU may, without changing the content, translate, if technically possible, the Work to any medium or format for the purpose of preserving the Work and facilitating the exercise of SFU's rights under this licence.

It is understood that copying, publication, or public performance of the Work for commercial purposes shall not be allowed without the author's written permission.

While granting the above uses to SFU, the author retains copyright ownership and moral rights in the Work, and may deal with the copyright in the Work in any way consistent with the terms of this licence, including the right to change the Work for subsequent purposes, including editing and publishing the Work in whole or in part, and licensing the content to other parties as the author may desire.

The author represents and warrants that he/she has the right to grant the rights contained in this licence and that the Work does not, to the best of the author's knowledge, infringe upon anyone's copyright. The author has obtained written copyright permission, where required, for the use of any third-party copyrighted material contained in the Work. The author represents and warrants that the Work is his/her own original work and that he/she has not previously assigned or relinquished the rights conferred in this licence.

Simon Fraser University Library
Burnaby, British Columbia, Canada

revised Fall 2013

Ethics Statement

Abstract

Laparoscopic surgery is a visually-guided manual task requiring mastery of non-intuitive motor mapping and detailed procedural knowledge for decision-making. Earlier research studies with unfamiliar motor tasks have shown novices make distinct changes in gaze behavior as the necessary manual skills are acquired. Although the basic manual skills are used repeatedly, successful completion of a specific task instance is still dependent on execution of a well-informed motor plan based on internalized visual information.

While other researchers have been able to use certain eye metrics to quantify expertise for some tasks, we have conducted various additional studies exploring different eye movement patterns as well as a combination of eye and manual parameters for identifying differences between expert and novice eye-hand coordination patterns. Still, expertise is consistently correlated with a clear difference in task completion time.

This research covers a series of eye tracking studies conducted in laparoscopic training environments and in the real operating room. Subsequent analyses prompted efforts to improve data quality and led to development of an instantaneous measure combining eye tracking and manual movement data, to describe expert and novice eye-hand coordination behavior. This knowledge suggests the possibility of applying training protocols in the future to directly manipulate the development of eye-hand coordination in surgical trainees for rapidly improving task performance.

Acknowledgments

I wish to thank my supervisor Dr. Stella Atkins for her guidance and instruction throughout my studies, and especially for her patience during this final stretch towards the completion of this thesis.

I also thank my family for their love and encouragement; to whom I owe credit for all my achievements.

To beloved Abigail, I am eternally grateful for her tireless support and love through my struggles.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Laparoscopic surgery

Laparoscopic surgery is a minimally-invasive surgical procedure involving the use of remote instruments to operate inside a patient's abdominal cavity through a view offered by a remote optical device. The instruments are inserted through relatively small incisions called ports, which reduce morbidity and recovery times in contrast to the large incisions traditionally needed to access the abdominal cavity in open surgery. Learning these surgical skills is particularly difficult compared with open surgery, where the operating site is visible with full depth perception.

Now a common technique used for a variety of diagnostic and intervention procedures, laparoscopy began around the start of the 20th century with experiments on controlling bleeding in the small bowel by applying pressure from the peritoneal cavity inflated with air. In these experiments, endoscopes which were straight-tubed optical devices held by the surgeon were used only for observation. It was not until the mid-1960s following few slow technological developments that laparoscopy gained momentum for both diagnosis and limited treatment involving manual interactions with patient tissues.

The most significant technological development was made in the mid-1980s with the introduction of electronic videoendoscopy. A small video camera at the end of the endoscope (Figure 1.1) projected a bright, clear image of the peritoneal cavity onto video monitors, allowing the entire operating room staff to view the scene (Figure 1.2). An effect of this was that the operating surgeon was no longer restricted to using one hand to control his personal view of the scope and manipulating surgical instruments with his remaining hand, but an assistant could control the videoendoscope and allow the surgeon to operate two manual instruments at once, setting the stage for more complex interventions to take place endoscopically. Importantly, such procedures required a new set of motor skills previously not encountered in traditional surgery or other occupations.

Figure 1.1: A modern laparoscope. The end at the left is inserted into the patient, the end at the right contains attachment points for a light source and power/video transmission. Source: stryker.com, accessed July 08, 2014.

1.2 Laparoscopic cholecystectomy

The gallbladder is a small sac-like organ mainly responsible for storing bile produced from the liver. Sometimes, cholesterol and bile salts can form solid, concentrated deposits called gallstones, a condition affecting up to 10% of the adult North American population. In some cases gallstones may be asymptomatic, but often can cause severe abdominal pain requiring treatment. Treatment can be given non-invasively in the form of oral medication for dissolution of the stones, or extracorporeal shock wave lithotripsy which breaks the stones into fragments which may then pass freely through the digestive tract. The disadvantage with these non-invasive treatments is the possibility of stone reoccurrence, requiring repeated future treatments.

A surgical removal of the affected gall bladder (cholecystectomy) is thus often the prescribed intervention. Prior to the adoption of laparoscopic interventions, cholecystectomy was performed as an open surgical procedure. In open cholecystectomy, a single incision roughly 5 to 7 inches in length is made just below the patient's ribcage, through which the cystic duct and cystic artery are severed and the gall bladder is removed. The incision leaves a large scar on the patient's abdomen, and typically requires a recovery time of up

Figure 1.2: A laparoscopic operation in progress, performed by a team who share a view of the surgical site displayed on overhead monitors.

to 4 days in hospital.

With laparoscopic cholecystectomy, only three or four incisions each roughly one inch are required - one for the laparoscope, two for surgical instruments, and one for an auxiliary instrument for displacing the patient's other internal organs during the operation. The laparoscopic incisions through which the instruments enter leave smaller scars which are preferred by the patients. Hospital stays for recovery from laparoscopic cholecystectomy are dramatically reduced compared to open cholecystectomy, providing a great benefit to hospital bed allocation if cholecystectomy patients can be treated as outpatients requiring no overnight hospital stay. Especially due to these advantages, laparoscopic cholecystectomy has become the standard for gallstone treatment, with over 600,000 operations performed yearly in the United States [52].

1.2.1 Challenges in laparoscopic cholecystectomy

Due to the large volume of cholecystectomy cases which must be performed, clearly laparoscopic skills are an essential component of the general surgery repertoire. However, as illustrated in Figure 1.3 showing a typical scene viewed during a laparoscopic cholecystectomy, there are many challenges causing laparoscopic cholecystectomy to be a vastly different interactive experience compared to its open counterpart.

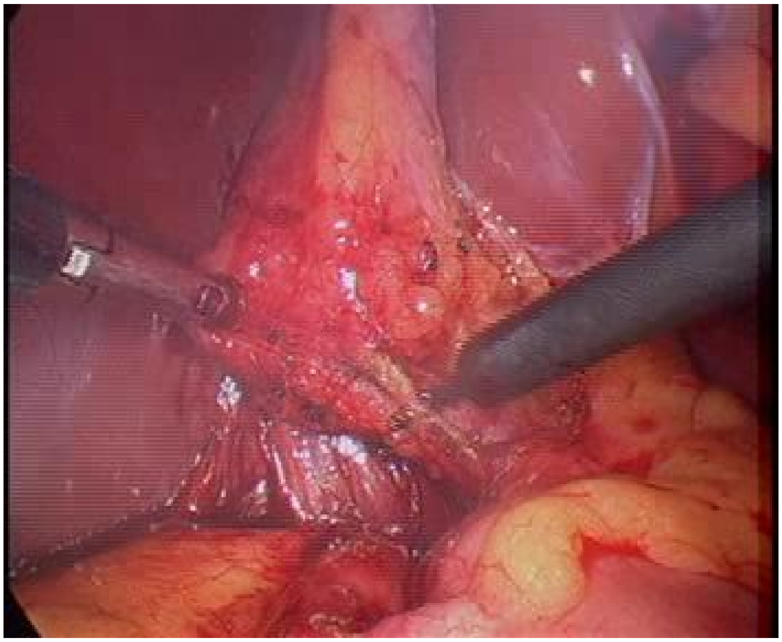


Figure 1.3: Surgical site viewed through a laparoscope, with interaction using laparoscopic instruments.

Visual challenges

The first and most obvious difference is the disconnection between the point of regard and the location of operation. The surgeon manipulating the laparoscopic instruments has no direct vision of the surgical site, relying instead on a projected image displayed on a video monitor which can be quite remote from the patient's location. This creates an unnatural mapping between action and consequence, especially amplified when the horizon of the captured image becomes tilted due to rotation of the laparoscope.

The field of view captured through the laparoscope is limited, offering little reference in terms of global patient anatomy, yet the surgeon must precisely navigate both scope and instruments based only on what is visible through the scope, which can become clouded due to fluids on the end of the scope, smoke from cautery, or obstruction from patient tissues.

Furthermore, navigation is made even more difficult due to the monocular nature of the captured scene. With the monocular view projected onto a two-dimensional display, many natural depth cues are lost, and the operator must learn to make use of other visual cues in order to construct the necessary depth information for safe completion of the operation.

Manual challenges

The long and slender laparoscopic instruments are difficult to move precisely simply due to their length, which causes gain effects in the amount of input movement required for a particular action such as opening a grasper. The instruments also have reduced haptic feedback which can cause dangers when squeezing and pushing soft tissues. Furthermore, movements are severely restricted due to the anchored nature of the trocars through which they enter the patient. The trocars produce a fulcrum effect which can move the end effector of an instrument in a direction opposite that in which the hand moves.

Thus, in training general surgeons to perform laparoscopic operations, there is a learning curve and a period of eye-hand coordination development during which mappings between motor inputs at the instruments and the desired visual consequences at the video monitor are linked.

Since there are both visual and manual challenges to overcome while learning how to perform laparoscopic operations, the development of eye-hand coordination applicable specifically to the laparoscopic environment has several unique features. Manual proficiency can be measured with tool trackers and expert evaluation of training exercise results (i.e. performance measures), though the task of recording and evaluating the "proficiency" of eye movements remains, and this can be accomplished with eye tracking.

1.3 Eye tracking

Eye tracking is an area of human-computer interaction (HCI) that encompasses the study and use of the user's gaze as part of an interaction with a machine system [8], where the gaze is a point in world space determined as the combination of the user's eye orientation and head position.

Eye tracking technologies vary in implementation and invasiveness, and different technologies can serve well in different applications - however, many recent research studies employ video-based eye trackers due to a number of advantages of the platform.

Modern eye trackers now typically use video oculography, by processing a digital video stream from a camera aimed at the wearer's eye or head. Video processing is performed for each video frame to locate the wearer's pupil which in turn is used to determine the direction of gaze. While there exist head-mounted video-based eye trackers, video oculography can also be done using a remote camera, not requiring any equipment to be in physical contact with the subject being tracked. A remote camera is typically used when viewing a stimulus where a user's head is not expected to move much, and in some cases head motion may even be limited by using a restraint. With remote video eye trackers, the user's point of regard relative to a situated real-world stimulus (typically a video monitor) can be recorded. Video-based eye trackers are convenient but often their spatial and temporal resolutions are limited by the camera resolution and frame rate. In spite of these limitations, the commercial availability and convenience of video eye trackers for both experimenter and subject as well as the analytical power of commercial eye data processing software have allowed video eye trackers to become the preferred equipment for studies not requiring the fine temporal resolution of more invasive eye tracking hardwares.

1.3.1 Types of eye movements

The movement of the human eye is a continuous phenomenon, and can be categorized into several types of movements based on distinct physiological characteristics. Firstly, gaze fixations are characterized when the focus of the eye remains fixed on a stationary object or location for an interval of time typically lasting 120 ms or longer [8, 11, 30]. During a fixation, the eye is subject to small natural jittery movement observable with high-resolution eye trackers; as a result the point of gaze will actually be constantly moving about the object of interest over the course of a fixation.

The next types of eye motion are saccades, which are very rapid ballistic eye movements from one fixation to another. These are characterized by a large rotational velocity of the eyeball, during which visual acuity is greatly reduced. These occur in foveation which angles the eye to bring a visual target into the sharp central field of vision.

Smooth pursuits are characterized by a steady rotational movement of the eye, typically occurring when the gaze is locked to a small but slowly moving or distant target, and predictive mechanisms are in place to overcome latencies in the visual system when target motion can be anticipated [30].

Other minor eye movements include those such as vergence motions used to orient both eyes to targets at varying distances, and motions generated by the vestibulo-ocular reflex which maintains a steady gaze position in space under head movement [30].

1.4 Tool motion tracking

In order to study eye-hand coordination, there needs to be some way to record the motion of the hand. In other studies, motions can be measured directly with motion trackers attached to a subject's hands. In an endoscopic surgical setting, this would be impractical. For the purposes of this study, manual motion is tracked at the end effector (laparoscopic instrument) from surgical videos recorded from the laparascope, inferring the motion of the actuator (operators hand). References to eye-hand coordination in our own experiments may be more precisely described as eye-tool coordination and are tracked only in the two-dimensional projection viewed from the laparoscopic camera.

1.5 Motivation for this thesis

Training and evaluation in laparoscopy have been mainly focused on manual skills, and the training can be done with a variety of equipment, including physical training boxes with synthetic or animal tissues, computer-based virtual reality (VR) simulators, and observation of actual cases recorded from the laparoscope video camera.

Physical training boxes have the advantage of low initial cost, accurate instrument motion and more haptic feedback, since actual instruments and tissues are used; however, ex vivo tissues do not react completely like live tissues (e.g. bleeding), and repetitive use and preparation can be costly since consumables such as tissues and sutures must be prepared and expended with each use.

VR trainers have extremely high initial costs, but can reproduce visually realistic surgical scenes with high fidelity, although haptic interactions can be inferior to those experienced with physical trainers, and the high purchase price leads to limited units available to share among trainees, restricting training throughput. VR however can provide evaluations of task performance, for example with automatic recording of task completion time, efficiency of tool motion trajectories, and occurrences of unsafe actions such as collateral cautery on unintended anatomy.

An important aspect of laparoscopic training involves watching pre-recorded laparoscopy cases. Watching these cases, either with or without instructional commentary, provides the most realistic visual depiction of an operation. While these videos are readily produced from equipment already installed for performing laparoscopic operations, watching them lacks any manual interaction so there is no opportunity to practice eye-hand coordination skills.

The common theme between these training setups is the focus on manual skill, which is only one component of eye-hand coordination. Expert and novice operators have been demonstrated to exhibit some different eye movement patterns in various task domains [4, 11, 24, 32, 42, 58], so it follows that there is not only a period of task learning during which manual skills improve, but also a shift in gaze behaviour from novice to expert during the natural course of task learning. This has been notably demonstrated in the work of Law et al. [29] who observed tool-tracking behaviour in novice operators to direct target-gazing in experts.

To expedite this shift in gaze behaviour, a technique of so-called gaze training may be applied, which focuses on the other component of eye-hand coordination not covered by current manual-focused training protocols. In practice, gaze training describes the use of implicit or explicit means to influence the gaze patterns of novice operators to more closely match the gaze patterns of expert operators with the hope that expert-like gaze patterns can correlate with expert-like manual task performance.

Towards the goal of gaze training in laparoscopy, we first identified a sub-goal of making educational surgical videos with expert gaze overlaid, with the reasoning that trainees would improve their own performance by looking where the experts would look. Since expert gaze is difficult to obtain during live surgery, we postulated that we can instead overlay expert gaze collected post-hoc while reviewing pre-recorded surgical videos.

Another approach of gaze training based on directly shifting gaze towards expert patterns was implemented and evaluated.

With a focus on the domain of laparoscopic surgery, the following chapters in this thesis aim to identify various eye tracking parameters where novices and experts differ, leading to the application of gaze training. With such gaze training we hope to accelerate the development of eye-hand coordination in the initially non-intuitive environment of laparoscopic surgery.

1.6 Overview of thesis

Chapter 2 surveys a selection of eye tracking studies of task expertise and eye-hand coordination in natural and medical tasks.

Chapter 3 presents an experiment using a simulated surgical environment to observe attentional differences reflected in the point of gaze between expert and novice operators across multiple displays.

Chapter 4 explores expert and novice points of regard on a single display in the live operating room and on replayed surgical videos.

Chapter 5 continues studying the point of regard for task execution and observation with novice participants, using a simplified laparoscopic peg transport task which became the archetypical manual task of our future experiments.

Chapter 6 repeated the same peg transport task with a different group of subjects including expert surgeons and novices. Eye movements were synchronized with tracked motor events throughout the task for our first eye-hand coordination measure.

Chapter 7 presents an experiment in gaze training on a modified peg transport task, using visual or verbal cues in an effort to influence the visual patterns and manual behaviour of novice operators.

Chapter 8 presents discussions addressing my hypotheses and closes this thesis.

Appendix A details a data filter applied to smooth the recorded eye data prior to analysis. This filter was applied in the experiments of Chapters 6 and 7.

Appendix B contains forms and paper materials as used in the various experiments.

1.7 List of research publications

1.7.1 Chapter 3

- Geoffrey Tien, M. Stella Atkins, Bin Zheng, and Colin Swindells. Measuring situation awareness of surgeons in laparoscopic training. In *Proceedings of the 2010 Symposium on Eye Tracking Research & Applications*, ETRA '10, pages 149–152, 2010.

- Geoffrey Tien, Bin Zheng, and M. Stella Atkins. Quantifying surgeons' vigilance during laparoscopic operations using eyegaze tracking. In *Studies in Health Technology and Informatics*, MMVR 18, 658–662, 2011.

- Bin Zheng, Geoffrey Tien, M. Stella Atkins, Colin Swindells, Homa Tanin, Adam Meneghetti, Karim A. Qayumi, and O. Neely M. Panton. Surgeons' vigilance in the operating room. *The American Journal of Surgery*, 201(5):673–677, 2011.

- Bin Zheng, Xianta Jiang, Geoffrey Tien, Adam Meneghetti, O. Neely M. Panton, and M. Stella Atkins. Workload assessment of surgeons: Correlation between NASA TLX and blinks. *Surgical Endoscopy*, 26(10):2746–2750, 2012.

1.7.2 Chapter 4

- Rana S. A. Khan, Geoffrey Tien, M. Stella Atkins, Bin Zheng, O. Neely M. Panton, and Adam T. Meneghetti. Analysis of eye gaze: Do novice surgeons look at the same location as expert surgeons during a laparoscopic operation? *Surgical Endoscopy*, 26(12):3536–3540, 2012.

- M. Stella Atkins, Geoffrey Tien, Rana S. A. Khan, Adam Meneghetti, and Bin Zheng. What do surgeons see: Capturing and synchronizing eye gaze for surgery applications. *Surgical Innovation*, 20(3):241–248, 2012.
- Geoffrey Tien, M. Stella Atkins, Xianta Jiang, Rana R. S. Khan, and Bin Zheng. Identifying eye gaze mismatch during laparoscopic surgery. In *Studies in Health Technology and Informatics*, 184:453–457, 2013.

1.7.3 Chapter 5

- Geoffrey Tien, M. Stella Atkins, and Bin Zheng. Measuring gaze overlap on videos between multiple observers. In *Proceedings of the 2012 Symposium on Eye Tracking Research & Applications*, 309–312, 2012.
- M. Stella Atkins, Xianta Jiang, Geoffrey Tien, and Bin Zheng. Saccadic delays on targets while watching videos. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, 405–408, 2012.

1.7.4 Chapter 7

- Geoffrey Tien, M. Stella Atkins, Xianta Jiang, Bin Zheng, and Roman Bednarik. Verbal gaze instruction matches visual gaze guidance in laparoscopic skills training. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, 331–334, 2014.

Chapter 2

Related work

Eye tracking studies comparing the eye movements of novices and experts in various domains have been conducted. This chapter summarizes a selection of eye tracking research performed in natural and direct-vision, direct-manipulation tasks, to then focus on studies of expertise and eye movements in image-guided tasks, in particular to health care occupations and laparoscopic surgery.

2.1 Eye movements and expertise in direct-vision domains

2.1.1 Sports

Recreational and professional sports is an area in which one can find clear differences in skill between individuals. Vickers [57] sought to investigate gaze differences between golfers of differing skill levels, motivated by earlier research showing that stroke mechanics are highly invidual whereas the relative position of the ball and player's head is quite constant across different players. Vickers recruited five low-handicap (highly skilled) and seven higher handicap golfers to perform numerous consecutive putts on level astroturf to a cup 3m away. According to standard golf handicaps though, even the higher handicap players were considered to be very skilled, so there were no truly novice players in this study. A head-mounted eye tracker with first-person scene recording was used, and for task analysis the entire putt was delineated into four phases: preparation, backswing/foreswing, contact, and follow-through. Items in the visual field extracted from the first-person video were marked as the player's own feet, the ball, the club head, the cup, and the putting surface.

Vickers found that the low-handicap players, as confirmed in other studies of expertise, fixated more on the ball and cup during the preparation phase and longer on the ball during the swing, whereas the high-handicap players looked more at the club. Interestingly, low-handicap players tracked the moving ball for a longer duration in the follow-through phase despite no longer having any control over the ball's movement.

Vickers also applied her eye tracking study to the high-pressure, high-precision area of basketball free throw shooting, to see if the expected target-focused gaze strategy observed in golf putting held for highly skilled foul shooters. Once again, highly skilled and moderately skilled players were recruited for the study; there were no novices. Three possible strategies were proposed: position-only where gaze is continuously focused on the target hoop or backboard, movement-only where gaze is focused on and synchronized to the player's hands, and a combined strategy using one at the beginning of the throw and the other at the end.

Similar to the golf study, the free throw action was divided into four phases: preparation, pre-shot, shot, and flight. Expert shooters were found to take longer in the preparation phase and showed steadier gaze behavior during the first half of the shot, but contrary to Vickers' expectation, the experts' gaze became more mobile than the near-experts' as the shot occurred. Notably, experts spent much longer gazing very steadily at the target during preparation, a period which Vickers calls "quiet eye". Additionally, this extended quiet eye period seems to be responsible for experts to maintain their consistent aiming even during the shot phases when the movement of the arms and ball could clutter the visual field and affect aiming. This knowledge sets a foundation for quiet eye training to improve skills in sports. A key technique which made such analyses and results possible was the finer division of an overall task into several logical phases.

Land and McLeod [28] also studied eye movements in sports, focusing on how cricket batsmen read an approaching ball to plan their swing of the bat. Unlike golf putting where the putter has no time constraint to prepare his swing, a fast cricket bowl can reach the batter 600 ms after leaving the bowler's hand. Because a reaction based on new visual information can take about 200 ms and the inertia of the bat takes even more time to overcome, cricket batsmen must essentially, from the first part of the ball's flight, make a predictive judgment to guide the swing of his bat.

A head-mounted video eye tracker with first-person scene camera recorded at 25 Hz. An automatic bowling machine delivered balls to the batter at 25 m/s, bouncing the ball at a number of fixed distances. The gaze recording showed that initially the gaze was focused on the aperture of the bowling machine, and then as the ball appeared the gaze quickly shifted below the ball in the field of view to where the ball was predicted to land, and the gaze continued to stay ahead of the ball along its predicted trajectory in a smooth motion aided by a slight head movement. This behavior of initial fixation followed by a smooth motion demonstrated that cricket batsmen do not actually watch the ball continuously at all times but use the initial fixation to gather the necessary trajectory information. Furthermore, it was found that the best batter in the study's cohort was able to initiate a saccade soonest after the appearance of the ball, indicating a quick reaction time and an ability to make the necessary prediction rapidly. Unfortunately the eye tracking equipment used in this study has a very low temporal resolution which is suboptimal for tracking the high velocity of the bowl. Also, only high speed bowls were observed whereas it was earlier stated that slower, spinning balls would require a different batting strategy.

2.2 Driving, flight, and mechanical inspection

2.2.1 Driving

Driving is an activity during which a person finds himself in highly dynamic surroundings which can sometimes be hazardous. Being able to recognize potential hazards before they occur can be key to road safety.

Chapman and Underwood [4], in an effort to avoid logistical difficulties of studying gaze habits in actual dangerous driving situations, devised an experiment where young novice drivers and older experienced drivers watched a number of short film clips recorded from a driver's point point of view. The clips were recorded on rural roads, suburban roads, and urban roads and each contained one to four hazardous events. The timing of each hazardous event within a video clip was delineated by experimenters as a "danger window". The subjects were tracked with an infrared corneal reflection eye tracker with their heads stabilized on a chin rest as they watched the film clips. When they saw a hazardous event occur, they were required to push a button. Latency between the beginning of each danger window and a button press was recorded in addition to fixation duration, saccade length, horizontal and vertical gaze angles and variance. Fixations used a special combined dispersion and velocity definition since virtually all objects on screen were in motion and traditional fixation calculations would yield very low numbers, based on the assumption fixations tend to last around 100 ms on a stationary point.

Fixation durations were found to be shorter in both groups as road conditions became busier. Although there were no differences in the button-pressing latency between groups, the authors' visualization of horizontal and vertical variance as two-dimensional fixation plots of the first-person scene surprisingly showed that the novice drivers looked further ahead of the vehicle, contrary to earlier research. Nonetheless, as dangerous scenarios cannot be accurately reproduced by passively watching a video with a limited field of view, it is still difficult to confirm the results.

Having intuitions about the gaze habits of novice and experienced drivers, Dishart and Land [7] began to study how eye movements develop as driving experience is gained. This study involved head-mounted gaze recording of subjects using both a driving simulator and actual public road driving. No head movement restriction was applied but visible markers were inserted throughout the field of vision so that head orientation could be estimated from the recorded first-person video. The driving task was focused on curve driving and lane position maintenance. As driving experience increased, subjects on the simulator showed a wider horizontal spread and fewer, longer fixations. One very novice subject drove on the actual road, and demonstrated noticeable changes in gaze behaviour at 4 hours experience and after 12 hours experience, producing longer fixation durations alternating between the far road edge tangent point and the center line tangent point. This strategy is presumed to alternately scan the far road edge as this is where potential hazards are most likely to appear, and the center line was used to continue guiding the steering of the car.

2.2.2 Flight

Similar to driving is air flight, where a pilot is in control of a vehicle and must navigate it in a safe manner according to the surroundings. Visual flight rules (VFR) flight is such a situation where a pilot must take visual cues from several locations in the environment, to determine the aircraft's position and orientation. Using a computer flight simulator and a remote video eye tracker with head position tracking, Kasarskis et al. [24] conducted a study on aircraft landing during VFR flight. Subjects each performed 15 landings onto a straight airstrip using VFR rules. The quality of each landing was given a calculated score based on deviation from the optimal landing point at touchdown. The pilot's field of view was divided into 4 areas of interest (AOIs) which included the view outside the cockpit through the window and 3 different flight instruments; gaze fixations were classified as being focused on these AOIs.

In terms of performance, not surprisingly experts landed consistently nearer the optimal landing point. They did this with many more fixations with lower average dwell time on the direct runway view and airspeed instrument, alternating between these with a clearly structured sequence while novices used the altimeter more along with the runway view and airspeed instrument without any clear pattern. As well, it was found that both experts and novices made fewer fixations when they performed the lower quality landings.

To go slightly further than searching for expert and novice gaze pattern differences, Schriver et al. [47] sought to investigate how experts and novices perceive and respond to different diagnostic cues in simulated situations of aircraft failure. The flight simulator was a high-fidelity immersive simulator with a replica cockpit and wide field of view from three projected displays. The view from the cockpit was divided into 26 AOIs for the various instruments and surroundings. In each failure scenario, problem cues occurred singly or in groups. Where multiple cues were present, these were either correlated or uncorrelated. Aside from the expected performance differences in diagnostic accuracy and latency, experts were found to attend to high diagnostic cues in the multiple uncorrelated scenarios whereas novice pilots showed the opposite behavior. This can be attributed to the expert pilots having a greater knowledge of instruments, but with additional study, novice pilots could learn to reverse their behaviors to respond to flight problems more quickly and more appropriately.

2.2.3 Mechanical inspection

Pilots may be able to diagnose and correct for some problems in flight, but repair and maintenance must be done on the ground by highly trained mechanical crews. There are several models of training people to perform inspections for defects in aircraft, and gaze feedback training has been used, which can be described concisely with the phrase “you looked here”. However, a novel approach taken by Sadasivan et al. [44] is the concept of using expert gaze movements as a feedforward training strategy which can be expressed as “you should look here”.

In a simulated task of inspecting an aircraft cargo hold for structural defects, experts were demonstrated to have a very systematic strategy of scanning the hold. The authors predicted that providing examples of this gaze strategy to trainee inspectors would allow the trainees to more effectively identify defects in a large search area. The simulator equipment was a binocular head-mounted display (HMD) with built-in eye tracker. The display was fully immersive and movements of the subjects’ head and body were reflected in the HMD. A trained inspector used the simulator and eye tracker to provide the feedforward training data. The eye tracking data were processed and visualized as a scanpath sequence of AOIs with a bar next to each AOI indicating the amount of time spent at the AOI, and present defects were highlighted. A controlled experiment was set up to test the effect of the feedforward training, so each subject in the treatment and control groups performed a pre-test inspection, treatment, and then post-test inspection.

As hoped, the feedforward training group found more defects in the inspection task than did the control group. This came with a speed-accuracy tradeoff, where the gaze trained individuals made more and longer fixations, having adopted a slower-paced systematic search strategy. The method of feedforward gaze training can potentially be applied to training in many areas of skilled human performance through visualization of pre-recorded expert scan paths.

2.3 Eye movements and expertise in image-guided tasks

Unlike the examples presented in the previous section where the visual stimuli are objects situated in the real world, decisions made while performing image-guided tasks are informed by cues presented on a virtual display.

2.3.1 Video gaming

An early work by Shapiro and Raymond used gaze training to stream study participants into groups of varying video game-playing expertise [50]. The game used in the study is called "Space Fortress", fully detailed in [32]. Space Fortress takes place in a frictionless two-dimensional Euclidean space, where the player controls a ship to fire missiles at and destroy a central fortress while avoiding hostile fire. While the full Space Fortress game has many complex mechanisms dictating the appearance and movement of objects on the screen, Shapiro and Raymond hypothesized that drilling subjects on simpler exercises using 'efficient' or 'inefficient' oculomotor strategies would have an impact on the player's final score in the real game, if oculomotor habits are generalizable and eye movement behavior has a role in performance. 33 male subjects who passed a skills screening for the game participated in the study and were assigned to control or treatment groups based on their screening score so that there was no significant difference in the mean score across groups. For the gaze-monitored group, gaze was recorded using an ASL Eye Trac 210 head-mounted infrared monitor with a sampling rate of 1000Hz and stabilized with a chin rest. The game screen was video recorded and the gaze location of each subject was later superimposed on the screen video for analysis.

Based on the knowledge that visual detection is greatly reduced during saccades, drill exercises done by the efficient gaze strategy groups were designed to minimize the number of saccades made while playing, encouraging use of peripheral vision to detect pertinent information in the playing area. Conversely, the inefficient gaze strategy group performed drills which encouraged making many saccades around the playing field. After many playing sessions, it was indeed found that the group receiving instruction on efficient gazing produced higher scores than the group with inefficient gaze strategies and the control group which received no instruction. As hoped, the efficient gaze group consistently performed fewer numbers of foveations compared to the controls and the inefficient group.

The strength of the study is in its relatively large and persistent subject pool, and two different control groups following the same protocol at separate physical locations. Each subject participated in nearly 400 game minutes. However, due to such a strenuous demand, only subjects who passed an initial screening test were admitted into the study, to reduce the chance of a participant withdrawing from the study prematurely. Since some level of expertise with playing the game was required to admit entry, the study was left without an opportunity to study players with lower gaming abilities. Moreover, the drills for the inefficient group were inherently risky, as subjects were instructed to wait for an additional stimulus to appear before taking action whereas subjects in the efficient group were allowed to eliminate threats as soon as they appeared. Such an artificial constraint was not present in the full game, so the trained inefficient gaze strategy would not be so transferable. While the experimenters were able to secure committed participation from their subjects, this highlights the need for training conditions to be transferable to the real-world task.

Although the ASL 210 provided eye tracking data with very high temporal resolution, at the time of the study there were not readily available automatic methods of processing the data, so foveations on various screen elements were actually manually categorized by humans viewing the screen video tapes with the gaze location superimposed. Five reviewers were available though not enough to cover all videos by multiple observers. Inter-rater reliability was found to be high for videos on which multiple scoring was available.

2.3.2 Cursor pointing

Continuing with the screen manipulation theme, Sailer et al. conducted a study on eye-hand coordination for a cursor-pointing task using a novel input device [45]. The task was to use a custom-built device to quickly point an on-screen cursor to a small target. The input device was a small rigid box with cylindrical ends held in both hands, and was sensitive to torsional and compressional forces applied by twisting, pushing, or pulling the hands in opposite directions. Torsional force moved the cursor along one primary axis on screen, and compressional forces moved the cursor along an orthogonal axis. The amount of force applied translated to the distance from which the cursor would be displaced from the center of the screen. As such, the input device acted like an auto-centering joystick but with an unique input mapping which experimental participants were required to discover on their own. 10 subjects each participated in two sessions with a 5-minute break in between. A single session lasted as long as necessary for a participant to hit 500 targets, and the control-to-screen mapping was reversed for the second session, requiring subjects to re-learn the mapping to hit targets quickly. As this was done, their eyes were tracked using a 120 Hz ISCAN video-based tracker combined with a bite bar for head stability. A pseudorandom sequence of 44 points scattered across the screen quadrants was repeated until 500 targets appeared. The sequence of target positions was chosen to be unpredictable and that consecutive targets would be separated by at least 18° of visual angle. A monetary compensation which scaled with increased target hit rate encouraged subjects to learn and perform the task as quickly as possible.

For the data analysis, the authors of this work were able to distinguish three stages of eye-hand coordination, which they called exploratory, skill acquisition, and skill refinement. In the exploratory stage, there were many cursor movements along the cardinal axes and gaze tended to pursue the cursor in whatever direction it traveled, which was often not towards the target. In the skill acquisition stage, the scanpaths gradually straightened towards the target as the cursor movement also did, and gaze fixations landed more on the actual cursor position or later ahead of the cursor along its expected trajectory. In the skill refinement stage, cursor movement became much more direct to the target, and fixations were primarily on the target as soon as a new one appeared. The authors clearly took precautions to ensure that consecutive targets would be located far enough away to produce adequate saccades and fixations. However, in doing so, the unpredictability of the next target location would actually be reduced, since nearby locations can be quickly eliminated from consideration.

2.3.3 Virtual block stacking

As a precursor to the theme of gaze training via observation of another individual's gaze behavior, Gesierich et al. [13] conducted a computer-based study on gaze differences between action execution and observation of a recorded action. This idea was investigated earlier by Flanagan and Johansson [10] in physical form, where the task involved using one's hand to reach and grasp a wooden block to stack on another block.

In summary, in Flanagan and Johansson's study, action participants were eye tracked while performing the physical block stacking task, and observer participants were eye tracked while watching an “invisible” human agent do the same physical task. They found that eye movements of observers were proactive instead of reactive, producing scan patterns similar to those performing the actual actions, and concluded that during action observation, humans instinctively implement motor programs as if they were performing the task, and the eye motor system gazes accordingly. A technical shortcoming of this physically-situated study is that it was not possible for different observers to be eye tracked on the exact same stimulus, an issue which is resolved in Gesierich et al.’s study. The Gesierich task was a two-dimensional representation of Flanagan’s block stacking task, where three blocks of different width appear on one platform, and must be dragged using a computer mouse to form a pyramidal stack on another platform, sometimes with a vertical obstacle between the platforms to force a curved path of block movement.

For this task, the 23 experimental participants were presented with four conditions:

1. Task action
2. Observation of a pre-recorded task which subject believes is executed by a visible human agent
3. Observation of a pre-recorded task which subject believes is executed by a hidden human agent
4. Observation of a pre-recorded task which subject is told is a computer-automated action

The visual stimulus was the same pre-recorded task for all the observation conditions, but the two deceptive conditions were presented to identify any psychological effect on watching behavior that the knowledge would cause. Eye tracking was done using a video-based Tobii x50 tracker recording at 50 Hz.

The results of this study are in partial contradiction to the findings of Flanagan and Johansson. While Gesierich et al. found anticipatory eye movements to be significantly higher in active conditions, there were far fewer proactive eye movements in the passive conditions, albeit with high individual variability. Particular subjects did consistently anticipate the block movements while others did not, and no differences in behavior were found between the three passive conditions.

The measures used in the study were percentage of time tracking the block, and percentage of time anticipating the block. These were defined on a per-sample basis with Euclidean distance of the gaze point being ahead of or behind the block position along its trajectory. There was no mention of using aggregate gaze measures such as fixation. No measure of expertise was included in this study; the task was very simple with an identical setup each time, so there was no opportunity to apply any complex decision-making in regards to block movement. Still, it is an important finding that was able to compare gaze patterns on identical visual stimuli, even if there was no direct comparison between the action and observation recordings.

2.3.4 Program debugging

Spanning the areas of task expertise, training, and joint attention is Stein and Brennan's work on computer program debugging [51]. Debugging computer code can be thought of as a visuo-spatial task, in that pertinent information must be expertly gathered from certain areas of the screen, often informed by cues appearing elsewhere. It is clear that some individuals are more skilled at code debugging than others, yet such expert knowledge can be difficult to articulate and communicate from one programmer to another, hence the goal of this study was to use an expert's gaze as a visual cue to guide for solving programming problems.

The experiment was conducted in two phases. To begin, three Java programs were written, each with a number of intentional logical bugs. The programs varied in complexity and modularity. In the first phase, a group of four professional programmers were asked to visually scan the buggy code with a think-aloud protocol, and had their eyes tracked with a head-mounted corneal reflection eye tracker. Their gaze provided the stimuli for the second phase. In the second phase, a different group of six professional programmers were asked to debug the same code from Phase 1. In half the cases they were shown a gaze trace before debugging, and in the other half they began debugging immediately.

The gaze of the Phase 2 participants was not recorded; the focus of the study was entirely on the amount of time required to identify and correct the program errors. The authors found that on average, having an eye trace available greatly reduced the amount of time programmers needed to find bugs, but a limitation of the technique is that it can be difficult to remember a lengthy gaze trace and apply it to solve the problem at hand. Furthermore, the gaze cue is impractical as an actual debugging tool – it would be more effective to simply allow the first round of programmers to remove the bugs directly. The authors stress the work's value as a teaching tool to reduce the amount of time to needed for novice programmers to find program errors. It is difficult to conceive the generalizability of this method, taking the large stimulus problem as an example. After watching the gaze trace of first phase programmers, the second phase programmers were able to jump directly to the code block containing errors, eliminating the need to scan through irrelevant code blocks. However, the training value of this can be exposed only upon a second review, for the novice programmer to look at the irrelevant code to understand what makes it irrelevant. The development of the ability for the novice programmer to identify program errors independently was not addressed in this paper, but understanding expert gaze indeed is an important early step in training.

2.3.5 Radiology

Reading of medical images in radiology and telepathology put forth a variation on image-guided task, as the images viewed are static and not affected by any user actions. Nonetheless, decisions made by viewers are still driven by information gathered from the available image, and how these images are viewed can affect the diagnosis returned by the viewer.

Nodine and Kundel spent many years studying the eye movements of radiologists viewing X-ray images. During their research using a custom head-mounted infrared reflection eye tracker, they developed a three-stage visual search and detection model of radiologists' decision process and three classes of diagnostic error [37]. The first stage of visual search described is the overall pattern recognition stage where the radiologist gains a global impression of the image at a brief glance and mentally flags potential target sites for detailed investigation. In the second scanning stage, central vision is used to examine possible target sites, producing clusters of closely spaced fixations. In the third decision stage, the radiologist stops to report a tumor after closely studying a target site, or continues the visual search.

The diagnostic error classes arise from individual characteristics of the second and third visual search stages, and are also tightly based on the analysis of aggregated gaze samples. The authors defined fixations at the lowest level as an aggregation of individual gaze samples with duration totalling up to one second and having a radius of 0.5 degrees. At a higher level, fixations are aggregated into clusters with a radius of 2.5 degrees using a running-mean rule. With these definitions, the authors identified the error classes as sampling error where tumors existed in parts of the image not covered by a fixation cluster, recognition error where a tumor is in a part of the image which is gazed at for no longer than normal anatomy, and decision error where a true tumor was studied intently but classified as normal.

The contribution of the work is in using the available gaze data to improve radiological diagnosis using what the authors call “feedback-assisted visual search”. After a complete round of inspection and diagnosis, based on the image coverage by fixation clusters the system can invite the radiologist to view areas not viewed previously (reduce sampling error) and reinspect clusters which received 3-4 fixations but diagnosed negatively (reduce decision error as false negatives), and finally offering a chance to revise the diagnosis if necessary. With this feedback system in place, the authors found that for a set of 60 normal and 60 tumor chest X-rays, 19% of false negative decisions were revised, which is an important result in this application affecting the well-being of human patients.

Vitak et al. [62] continued a variation of this feed-forward gaze instruction in what is called “gaze-augmented think-aloud” (GATA). In their study, a Tobii 1750 remote video eye tracker was used to record the eye movements of an expert scanning magnified images of cell histology for abnormalities. The expert was asked to speak aloud his train of thought while searching within the images. The think-aloud voice recording was saved and used to generate two sets of training stimuli – a set of histology slides shown together with the voice recording, and a set of the same histology slides shown with voice and the expert’s point of gaze overlaid (GATA).

The experimenters found that novice participants who were asked to identify abnormal cells on a new set of slides after viewing the GATA training slides did so with significantly fewer false positive identifications than those novices who viewed the training slides with only the voice recording. This suggests a powerful utility of expert point of gaze for education, since when combined with a think-aloud protocol, may prevent trainees from becoming “lost in space” when listening to the voice recording without any visual direction.

2.3.6 Telepathology

Krupinski et al. [27] performed a comprehensive gaze analysis in telepathology, to investigate gaze characteristics of professional pathologists, residents, and medical students while reading digital slides of a breast core biopsy. Each of these skill groups had 3 participants in the study, which tracked the readers’ eyes using a video-based head-mounted ASL SU4000 tracker with a sampling frequency of 60 Hz and accuracy within 1 degree. Fixations were defined using a running-mean distance calculation with a 0.5 degree radius. Regions of interest (ROIs) with a radius of 2.5 degrees were manually selected by each subject. The chosen regions of interest along with how the members of each skill group spent their time making fixations and saccades were the main interests of this study.

The authors found that most ROIs chosen for magnified inspection contained relevant information, and the irrelevant ROIs were contributed mostly by medical students and residents. Of the measures pertaining to gaze, expert pathologists fixated less when selecting their ROIs, indicating they selected using peripheral vision. This is supported by the observation that pathologists had few saccades with long saccade durations, and short scanpath distances. Some measures reported in the study seem to be slightly redundant, such as the average saccade duration and mean saccade velocity. Intuitively for a given saccade between two fixed points, if one measure increases then the other measure is expected to decrease. The saccade duration is also dependent on the saccade distance which thankfully is reported. In this work where the overall visual behavior of different skill groups is important to understand, the authors presented detailed results on fixation data and especially saccade data where many other works focus only on fixations.

2.3.7 Anesthesiology

Anesthesia in the operating room (OR) presents a unique situation in that it is highly visually loaded and it is necessary for an anesthesiologist to gather information from both digital displays as well as other physical instruments, making this both a direct-vision and image-guided task.

Loeb [31] performed an important experiment on monitor surveillance and vigilance on anesthesia residents in the OR. Although no special eye tracking equipment was used, the paper was included in this survey as it did in a way measure the amount of time during a procedure that an anesthesia resident's gaze was directed at various points around the OR during actual operations. In this study, two trained observers watched the anesthesia residents, very coarsely categorizing the point of gaze as being directed towards four activities: observing digital displays, writing on the anesthesia record, busy with another activity, and idle, once per second in each of the three 15-minute stages of the operation. The stages were defined as induction – the initial period of bringing a conscious patient to an unconscious state, maintenance – keeping the patient unconscious, and emergence – waking an unconscious patient. In addition to their regular OR duties, the residents were asked to perform a vigilance task which involved pressing a button when they noticed a number shown on another digital display changed.

While the authors found no correlation between experience and response time on the vigilance task, they at least found that monitor observation was low during the induction phase compared to maintenance and emergence. A clear limitation of the study is the method of gaze estimation used. Requiring a human observer to constantly make a reliable categorization every second for three 15-minute periods is a heavy demand on the observer and results could be affected by fatigue of the observer. On the other hand, a remote observer requires no calibration and is perhaps the most non-invasive tracking method encountered so far. Next, all three stages of every operation were fixed to last 15 minutes. The mean values of monitor observation could be greatly perturbed in cases where actual induction lasted for far fewer than 15 minutes. Nonetheless, this study opened the way for further anesthesiology studies using more sophisticated eye tracking equipment.

This was accomplished by Seagull et al. [48], who used a head-mounted eye tracker with first-person video to test their model of monitoring frequency. They predicted that monitoring frequency increases before an anticipated event, drops during a high workload event, then rapidly increases again after the event has passed and finally returning to normal levels after recovering necessary information from the monitors. The task of interest was routine anesthesia induction and completion of intubation of the patient's airway. Fixations were extracted from videotapes of the eye tracker recordings using MacShapa software; fixation parameters were not reported and are not specific to the general-purpose MacShapa. Like Loeb's study, fixations were coarsely categorized based on the activity being performed, and when plotting the frequency of monitoring against the progression of the anesthesia task, a plot similar to the one predicted by the authors' monitoring frequency model was obtained.

Although it was not thoroughly discussed, the authors had the foresight to deploy a lateral-view video camera to record an overall third-person view of OR activities. This would provide additional insight to the conditions influencing attentional shifts as well as a backup data source in case of eye tracker failure.

A more technical application of eye tracking in anesthesia is presented by Segall et al. [49]. This is less of a study in expertise or gaze behaviors of anesthesiologists, but rather a study in designing visualizations for the vast multidimensional eye tracking data collected from a head-mounted video eye tracker in a real or simulated OR, and thus it is worthwhile to mention the benefits it provides to focused expertise and training studies.

Four visualizations were proposed for the collected data. In the first, Studiocode software is used to manually annotate fixations by categorizing a given fixation into one of several pre-defined ROIs. After all fixations have been labeled, a timeline can show at what time and for how long each time the subject fixated on each ROI. Another visualization is a scatterplot which plots a ROI as a point in a 2D space with the number of fixations on one axis and the mean dwell time on the other axis. For example, one could quickly see that a subject gazed infrequently at the ECG waveform but spent a relatively long time looking at it whenever he did. The third visualization uses a timeline and fixation ROIs like the first, but includes a mark to show the transition made from one ROI to another, and two such timelines can be drawn on the same graph to compare differences in attention and transitions before and after some event. Lastly, a single-step gaze path is proposed, where digital representations of each ROI are laid out on a 2D display, and an arrow of varying thickness is drawn to show the frequency of a transition from one ROI to another.

The visualizations ought to be powerful to show trends in attention and changes in attention at a glance, and can even be applied to areas outside of anesthesia. There is however a high labour cost in manually annotating fixations, but in the future this may be greatly aided by computer vision.

2.4 Minimally-invasive surgery

Like the screen manipulation tasks described earlier in this chapter, minimally-invasive surgery presents a manual task using an interface which is foreign to most people, thus requiring a period of dexterity training before an expert skill level can be attained.

Law et al. [29] investigated gaze habits during a simulated laparoscopy task. Five novice subjects and five experienced surgeons had their eyes tracked using an ASL 504 remote video-based eye tracker while performing a virtual aiming task. The interface used was an Immersion Laparoscopic Impulse Engine which simulates the movement of a laparoscopic instrument in three dimensions. The task was to use the right-handed tool to point the tip of the instrument on a virtual cube while avoiding tool contact with the boundaries of the allowable movement space. Due to differences in tracking resolution between the eye tracker and the manual interface, an involved synchronization procedure was performed for the offline data analysis, as well as a mapping of the tool tip and target box from virtual space to screen space. After processing, gaze measures such as gaze time on tool and gaze time on target were compiled, and the authors discovered that expert surgeons hardly spent any time gazing at the tool while novices exhibited frequent tool-tracking gaze behavior. An additional unique form of data presentation from the authors is the movement profile which plots the distance of both the eyes and the tool to the target against the elapsed time. This provides an immediate striking visual difference between the target gazing and tool following behaviors.

In terms of performance measures, it was stated that surgeons did not make significantly fewer errors than novices. It may have been worthwhile to observe the movement paths chosen by experts, if they performed movements to avoid clipping the target with the instrument shaft which was not categorized as an error but would be a potentially undesirable incident in a real surgery.

Kocak et al. [26] also studied the eye motion parameters of operators with varying levels of laparoscopic experience, with the goal of introducing an objective measurement expertise at the surgeon-monitor interface, to accompany the established measurements of expertise at the surgeon-instrument interface.

This study used more realistic bimanual tasks inside a standard laparoscopic training box. Three different tasks were devised, requiring coordinated movement of both hands on all axes of motion. 24 subjects self-classified into 3 skill groups had their eyes tracked using a head-mounted infrared saccadometer. The eye tracker was not electronically linked to the visual stimulus from the training box. The gaze data collected from the eye tracker are mainly saccade parameters although a fixation duration was also reported. The parameters defining saccades and fixations were not stated, but authors mentioned that the outputs of the eye motion parameters were displayed in real time, so it is assumed the saccade and fixation parameters were provided by the developer of the eye tracking hardware and software.

A positive correlation between surgical experience and fixation duration, saccadic peak velocities, and saccadic amplitude was found, and surgical experience was negatively correlated with saccadic rates. As the authors noted, because the image from the training box was not recorded, it was impossible to infer any connection between the subjects' gaze habits and the actions being executed inside the training box. To investigate this, the three tasks were well-devised to exercise skills that are readily applicable to the OR, so an additional study to record only the training box image on the same tasks could suffice for analysis, assuming the gaze-experience correlation still holds. The authors also claim that their head-mounted eye monitor is non-invasive, but it is likely that the tether could still get in the way of normal operations of the primary surgeon or OR staff or otherwise cause discomfort when used for a prolonged period. Finally, the authors agree that although a correlation was found between the observed gaze parameters and the level of surgical experience, the gaze parameters alone are not sufficient for predicting an individual's laparoscopic skill.

Richstone et al. [42] present another effort to provide an objective assessment of surgical skill using eye metrics. The authors recruited 21 surgeons to be eye tracked during simulated and live laparoscopic operations. Performance in the live surgeries was rated by expert judges according to the objective structured assessment of technical skills (OSATS) rating scale. A head-mounted SR Research Eye Link II infrared video system was used to record blink rate, fixation rate, index of cognitive activity (ICA) [33], and vergence. ICA is based on the notion that pupil diameter fluctuates as a person performs a task under different cognitive workloads.

Recorded eye metrics were given to linear discriminate function analysis (LDA) and nonlinear neural network analysis (NNA) to classify the owner of a given recording as expert or non-expert. Ground truth for this purpose was taken as the rating from OSATS. For both the simulated and live cases, the statistical methods had high classification accuracy. There was also a single longitudinal case where one subject was recorded over 18 months, and the classifiers became less able to distinguish his eye metrics from those of an expert's as his skill progressed. This raises the question that clearly skill develops over time, and people all possess differing skill levels – that is, expertise is continuous rather than easily cut into two classes. At least, further investigation could be done with quantizing skill into a higher number of discrete classes. With the high classification accuracy of the statistical methods though, this work could well provide a standardized and automatic way of rating performance as long as training institutions have the necessary eye tracking equipment locally.

Continuing with the investigation of the link between the surgeon-tool interface and surgeon-monitor interface, Wilson et al. [63] conducted an eye tracking study of surgeons at a computer-based laparoscopic surgical simulator. The interface is called Lap Mentor and includes realistic instruments for performing simulated OR procedures in addition to simpler exercises aimed at improving individual manual skills. The authors used a head-mounted ASL video-based eye tracker which was calibrated to the Lap Mentor screen, and recruited 6 novice surgeons and 8 expert surgeons to do a Lap Mentor exercise which involves using the tips of the left and right instruments to point at different colored balls in the virtual 3D space. An advantage of using the Lap Mentor over the training equipment used by Kocak et al. and the live operations of Richstone et al. is that the Lap Mentor automatically records all instrument movements made during the procedure, and at the end compiles a list of manual performance metrics such as the number of delineated tool movements, and economy of movement. While such manual metrics are quite similar to those used for expert judging of surgical skill, Wilson et al. added fixation rate and percentage of time fixation on critical locations to comparison.

In terms of the manual metrics, the authors found no difference in number of movements between the novices and experts, but experts displayed high economy of movement in both hands while novices could only use their dominant hand well. For the eye metrics, fixation rates were the same across both groups, but while novices spent equal time fixating on the tool and the target ball, experts fixated mostly on the target balls, confirming the findings by Law et al.

Setting their earlier eye metrics study as a precursor, Wilson et al. in 2011 [64] designed a new study, this time defining actual gaze-based and movement-based training protocols for the same ball-pointing task on the Lap Mentor apparatus. In this study, 30 novice participants were divided into gaze-trained, movement-trained, and discovery learning groups. Over the course of several trials of ball pointing on Lap Mentor, the gaze-trained subjects were shown their own eye movement patterns and received instructional feedback for adjusting their eye behaviors. In contrast, movement-trained subjects received feedback regarding their manual actions. Discovery learning subjects received no feedback but reviewed their movement metrics reported by Lap Mentor.

In addition to the ball pointing task, subjects at the same time were asked to multitask by counting the number of times an audible beep occurred during the procedure. The authors showed that the gaze-trained group produced faster completion times and this remained consistent in the multitasking scenario. Furthermore, the gaze trained group produced lower beep counting errors than the other groups. Not surprisingly, the movement-trained group produced the lowest target fixation times, even lower than the discovery learning group. Obviously this is because the movement trained group was explicitly instructed to focus on the instruments; the implication is that traditional instructional feedback is given regarding a trainee's actions and this produces gaze movements entirely contrary to pattern exhibited by expert surgeons, thereby possibly inhibiting the development of expert skills.

With the abundance of evidence demonstrating the differences in novice and expert gaze behaviors, Wilson et al. have successfully shown that gaze training can greatly expedite both gaze and manual skill acquisition for a laparoscopic pick-up and drop task.

Following the lead of these prior research findings and considering the advantages and disadvantages of applying certain data collection and analysis methods to specific tasks, we have been inspired to make new observations and discoveries in the challenging domain of laparoscopic surgery.

Chapter 3

Vigilance in minimally-invasive surgery

This study was the first in a series to investigate gaze pattern differences between experts and novices in a minimally-invasive surgical setting. In general, vigilance is defined as a state of alertness or watchfulness to avoid danger. In a surgical setting for the purposes of this study, vigilance was defined as the ability of the primary operating surgeon to detect abnormal changes in the simulated patient's vital signs and to monitor the patient's condition throughout the operation. The primary operator is referred to as the "subject" in the experiment, and wore a head-mounted eye tracker while performing the simulated operation.

3.1 Hypotheses

We hypothesize that:

1. Expert surgeons will be more vigilant than novices when performing a simulated laparoscopic cholecystectomy, measured as glances to a secondary vitals monitor,
2. Both novice and expert subjects will glance toward the vitals monitor more frequently when operating on a patient with an unstable condition, and
3. Expert surgeons will complete the simulated laparoscopic operation in shorter time than novices.

3.2 Independent variables

For this study we manipulated the following independent variables:

- Expertise (expert / novice)
- Patient condition (stable / unstable)

The specific experimental conditions tested are as follows:

- Expert operator × stable patient
- Expert operator × unstable patient
- Novice operator × stable patient
- Novice operator × unstable patient

3.3 Dependent variables

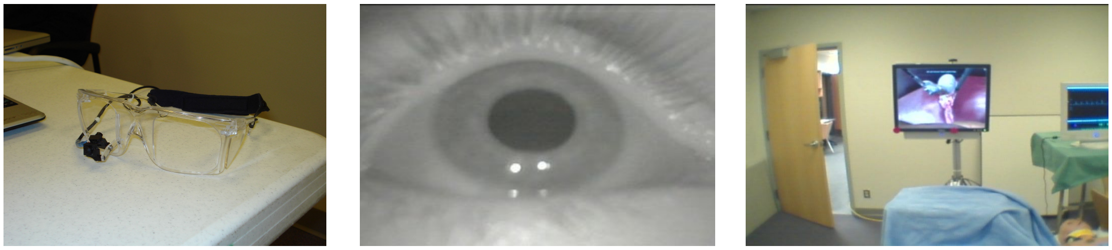
The dependent variables observed are as follows:

- Number of glances to secondary vitals display
- Task completion time

The number of glances to the secondary display is used as our measure of surgical vigilance regarding the patient condition.

3.4 Apparatus

The main interactive interface was provided by a SurgicalSim Virtual Reality simulator (Medical Education Technologies International, Inc.). The SurgicalSim VR has a foot pedal and two multi-purpose instruments with magnetic motion tracking. Using a variety of physical attachments, the instruments can be adapted to simulate various tools such as a laparoscope, graspers, needle drivers, or electrocautery to suit different training procedures. For the simulated laparoscopic cholecystectomy task selected in this study, the left instrument was designated as a grasper and the right instrument with foot pedal supplied cautery. A tower-mounted 17" LCD monitor displayed the virtual laparoscopy task at a viewing distance of approximately 180 cm.



(a) (b) (c)

Figure 3.1: (a) Locarna PT-Mini headgear, (b) eye camera screenshot, and (c) scene camera screenshot

For a more realistic scenario, a life-sized training mannequin controlled by an Emergency Care Simulator (ECS) system (METI, Inc.) was placed next to the SurgicalSim VR unit. The mannequin was programmed to simulate a breathing patient, and a simulated heart rate, blood pressure, and blood oxygen saturation were displayed on a 15" LCD monitor above the patient; the heart rate was augmented with audible beeps.

In order to record visual attention on multiple separate objects providing visual information, it was necessary for the eye tracking hardware in this study to track eye movements relative to objects situated in physical space. A head-mounted PT-Mini eye tracker produced by Locarna Systems, Inc. was chosen for this study for satisfying this requirement. The system consists of a headgear tethered to a modified netbook computer serving as the recording unit. The headgear can be fitted over prescription eyeglasses and performs eye tracking by synchronizing video recorded from two cameras one aimed at the wearer's right eye, and one facing forward to capture the scene from the wearer's point of view. The eye camera performs dark pupil recording at a resolution of 352×240 pixels, and the scene camera records at 720 × 480 with a 60° horizontal field of view. Both cameras capture at a frame rate of 29.97 Hz, with focal depth adjustable by manual rotation of the aperture barrel. The headgear and sample frame captures from the scene and eye cameras are shown in Figure 3.1.

The experimental task to be performed on the SurgicalSim VR was a partial laparoscopic cholecystectomy. In relation to the overall procedure, the cystic duct and cystic artery were already isolated, clipped, and cut; subjects only needed to use the virtual grasper and electrocautery hook to dissect the connective tissue attaching the gall bladder to the liver bed. During the procedure, ECS was programmed to present one of two patient conditions. One patient displayed all stable vital signs, and the other patient was programmed to display occasional mild cardiac arrhythmia. The complete apparatus (PT-Mini, ECS, SurgicalSim

VR) in use is illustrated in Figure 3.2.

Figure 3.2: A subject being eye tracked with a PT-Mini while operating the SurgicalSim VR with patient vitals displayed using ECS.

3.5 Description of participants

The data for this study were collected in two separate rounds, both using a within-subjects design. The initial round involved eight subjects (4 novice, 4 expert) from the Centre of Excellence for Surgical Education and Innovation (CESEI) at Vancouver General Hospital (VGH).

Twenty-five new subjects from CESEI and VGH were recruited for the second round with two omitted from analysis due to low recorded data quality. Of the remaining 23 included in analysis, 13 were categorized as experts (medical residents post-graduate year 4–6, fellows, practicing surgeons) and 10 were novices (PGY 1–3 residents).

3.6 Procedure

Subjects were first briefed on the nature of the study and then gave signed consent to participate in the study. A copy of the participant instructions appears in Appendix B.1.1. The instructions advised subjects to be aware of the patient's vital signs, but also to perform the operation as they saw fit. Every subject first completed a background questionnaire

(Appendix B.1.1) to determine their level of experience with performing laparoscopic operations and with using head-mounted eye trackers. Following completion of the questionnaire, subjects were immediately placed into different expertise groups based on the response and pseudo-randomly assigned the order in which to perform two trials (stable and unstable patient condition). After putting on the PT-Mini headgear, a calibration procedure was performed to obtain images of the subject's eye at 9 different locations in the scene cameras field of view. Subjects then were allowed to read their first patient history on a sheet of paper according to their assigned task order and began to perform the virtual gall bladder dissection. Patient histories are detailed in Appendix B.1.1.

The ECS was activated to present the appropriate patient vital signs. The stable patient displayed a healthy heart rate without much fluctuation, and the unstable patient had some mild arrhythmia which cycled through stable-unstable-stable at roughly one minute intervals. The order in which the patient conditions were presented to the subjects was counter-balanced.

After completion of the first virtual operation, the eye tracker was stopped and subjects were asked to complete a paper-based copy of the NASA Task Load Index (TLX) [35] survey to assess the workload level experienced during the operation. If the eye tracker was not moved or adjusted during this time, subjects proceeded to operate on their second patient; otherwise the eye tracker was re-calibrated before the next operation. The NASA TLX was completed again following the second operation, and subjects were dismissed.

The overall procedure was not changed between the first and second rounds of data collection with the exception of using the unweighted TLX in the first round as opposed to the electronic TLX of the second round. The electronic weighted TLX includes an additional series of questions which attributes the overall workload to six factors, offering more detailed insight into the various factors contributing to the overall workload level.

3.7 Results and analysis

Raw eye tracking data were processed with Locarna's Pictus software, generating fixations relative to the scene camera defined as consecutive points of gaze within 40 pixels (approximately 3.3° visual angle) for a minimum of 3 frames (100 ms). Reported fixations were coarsely annotated as being on the main display, vitals, display, or elsewhere. A saccade to the vitals display was marked by a fixation on the vitals display preceded by a fixation the main display.

		Completion time (s)		# glances to vitals		TLX workload score	
		Novice	Expert	Novice	Expert	Novice	Expert
Round 1 (4N, 4E)	Stable	168.7 ± 56.3	218.7 ± 134.2	0	20	65.2 ± 25.0	31.9 ± 9.4
	Unstable	219.8 ± 113.2	200.8 ± 96.5	2	21	60.6 ± 29.4	35.2 ± 14.9
Round 2 (10N, 13E)	Stable	162.8 ± 43.4	195.4 ± 71.5	11	32	50.0 ± 15.5	51.5 ± 14.1
	Unstable	173.0 ± 55.6	188.8 ± 70.8	21	38	54.7 ± 16.1	53.8 ± 16.6

Table 3.1: Completion time (mean ± SD), groupwise total saccades to vitals monitor, and NASA Task Load Index score (mean ± SD).

3.8 Discussion

Hypothesis 1 is supported by the results in Table 3.1, but hypothesis 2 is not well-supported, with experts monitoring both stable and unstable patients approximately the same amount. Hypothesis 3 was also not found to be true – the overall completion time between novices and experts was not significantly different by t-test, though experts tended to produce slightly higher task completion times. There are several possible reasons for this. The lack of negative consequences to tissue perforation errors could have lead novices to perform the task with less care than they might have with a stricter system. Additionally, the instabilities of the risky patient were pre-programmed; due to the isolation of ECS and SurgicalSim VR from one another, there was no way for adverse events in SurgicalSim VR to affect the state of ECS. In other words, if operators detected abnormal patient conditions, they could still only continue the operation with no options for intervention. Another likely contributing factor is that attending to the patient vitals display almost necessarily pauses the operation briefly, and thus operators more observant of the patient condition took longer to complete the task. The SurgicalSim VR, while being visually realistic, lacks haptic feedback which may have caused expert surgeons some difficulty in transferring their years of clinical experience to the simulator.

This suggestion can be supported by the reported workload from the second round of data collection. Although similar overall workloads were reported by experts and novices, in a 2 (group) × 2 (patient condition) mixed ANOVA, experts reported a higher level of frustration (46 ± 20 vs 33 ± 17, $p = 0.07$) and a lower level of physical demand (36 ± 18 vs 54 ± 20, $p = 0.03$) than did the novices.

Results here indicate that task completion time alone, which is often inversely correlated with expertise is insufficient in distinguishing operating skill, although this study indicates that when patient information is available, experts are more attentive of the patient condition even if this is usually monitored by an anaesthesiologist in an actual operating room setting.

3.8.1 Threats to validity

The arrangement of the task is not entirely realistic – in most real OR situations, the patient's vital signs are displayed on the anaesthesiologist's console and would not be directly monitored by the primary surgeon. Nonetheless, a surgeon must still be aware of other cues pertaining to the patient's condition and communicate with the surgical team in the event of any instabilities not automatically managed by the anaesthesiologist. The use of task completion time as a measure of expertise possibly may not be directly comparable between expert and novice subjects. Novices focused intently on the primary task while completely ignoring the secondary monitoring task are then performing only one task, in contrast to experts or other novices who perform both the primary task and the secondary task and whose primary task completion time is thus affected.

3.9 Research contributions

In this chapter we found expert surgeons were able to more comfortably divide their attention between their primary surgical task and monitoring patient condition, perhaps due to perceiving lower levels of physical strain with operating the laparoscopic trainer. Moreover, the degree of realism of a surgical simulator may affect performance of experts and novices differently. This implies that completion time alone may not be the most important factor in distinguishing expertise, as indicated by surgeons applying their cautious clinical performance to the simulator. Nonetheless, the lengthy and complex task led to confounds in this study design, precluding firm conclusions about the development of eye-hand coordination in laparoscopy. A simpler, more rigidly-defined task will be explored in Chapter 5.

Chapter 4

Eye tracking during live laparoscopic cholecystectomy

Following from the results of the vigilance study showing attentional differences between experts and novices, eye tracking data from practicing laparoscopic surgeons were collected during actual laparoscopic cholecystectomy cases scheduled at UBC Hospital. The eventual goal was to use the expert's eye gaze to train novices to look at the correct task-relevant anatomy during an operation's critical moments. Ideally, this would be achieved using the expert gaze recorded live in the OR during the surgical operation. Since a surgical scene could contain different logical anatomies separated by less than two degrees of visual angle, this application required higher resolution eye tracking on the laparoscopic scene than could be achieved using the head-mounted Locarna eye tracker, which can only reliably be used to identify gaze on fewer than nine coarse regions of the surgical display which occupies only a portion of the captured first-person view. Furthermore, the tethered nature of the Locarna tracker and the difficulty in sterilizing it for OR safety was a deterrant to its use in the live OR. Therefore a Tobii x50 remote eye tracker was chosen for collecting live data for later use in training firstly due to its higher tracking resolution and additionally for its flexibility in accommodating a variety of displays including those installed in the OR at UBC Hospital.

We understood that live OR data was difficult to collect consistently, so we also planned an alternate method of collecting expert eye gaze data on the surgical video. This involved having experts watch pre-recorded cholecystectomy cases while recording their gaze on the same video seen by the surgeon during the live operation. With recorded eye data available, this allowed us also to compare watching vs. doing gaze patterns by the same operator as well as the watching gaze pattern of external viewers.

4.1 Hypothesis

For this study we hypothesize that:

1. Expert surgeons watching a recording of their own surgical case will have a high amount of overlap with the gaze recorded when they performed the live operation, and
2. there will be more overlap with the recorded gaze between the expert watching their own recording than novices watching the same case.

The reason for investigating the difference between doing and watching was towards the thesis sub-goal of producing educational videos with expert gaze overlaid, and this expert gaze can be collected at a much higher quality in a controlled laboratory setting instead of the dynamic OR environment.

4.2 Independent variables

Recorded surgical cases were watched by the operating surgeon and medical residents, for a single expertise independent variable.

4.3 Dependent variables

The following parameters described in Section 4.5.2 are observed as dependent variables:

- Pairwise overlap summary
- Number of mismatch intervals

4.4 Method

4.4.1 Live OR eye tracking

Consent to be eye tracked was obtained from the operating surgeons before their scheduled procedures.

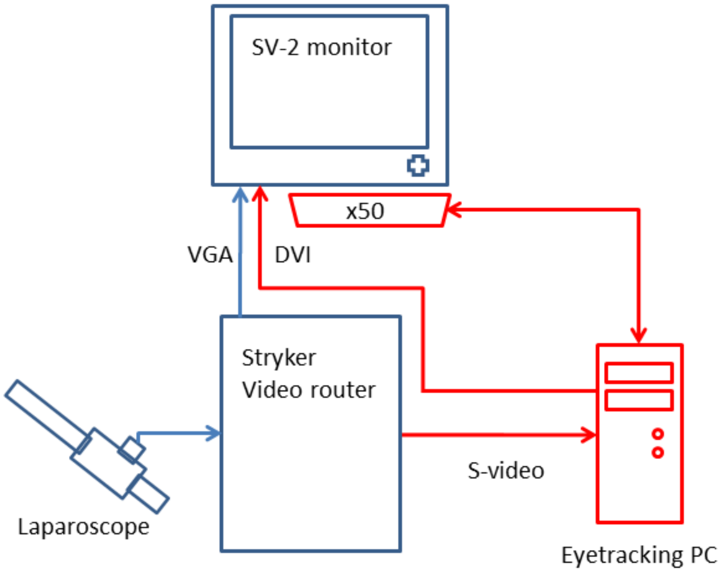


Figure 4.1: Diagram of OR eye tracking setup. Blue components show existing equipment and connections in the regular OR workflow; red components are components and connections introduced for eye tracking. None of the existing connections or components were interrupted.

The laparoscopic equipment in the OR interfacing with the eye tracking equipment included a straight laparoscope with a 90° field-of-view. The video from the scope was sent to a Stryker video router which projected the surgical scene to a 19” Stryker SV-2 video monitor ceiling-mounted to an articulating arm. The SV-2 was connected to the video router via VGA connection at 1280 × 1024 resolution. There was an unused DVI connection on the monitor, and pushbuttons on the front of the monitor could be used to select the input to drive the display.

The eye tracking hardware consisted of a Windows XP-based PC running Tobii Clearview 2.7.1 software, and a Tobii x50 remote eye tracker, performing dark-pupil video eye tracking for both eyes at 50Hz, accurate to 0.5° visual angle. Video capture was performed by a Hauppauge HD 1600 tuner card and the “external video” stimulus in Clearview.

The Stryker video router had an open S-video port which was connected to the input channel of the TV tuner card on the eye tracking PC. The PC was then connected to the DVI port of the SV-2 monitor. Figure 4.1 presents a stylized representation of the OR eye tracking setup.

All electrical and computer connections were made in the OR before the first scheduled operation of the day. Upon arrival of the OR staff and patient, a height-adjustable shelf was covered with a sterile drape and the x50 eye tracker was placed on top and moved out of the way while patient preparation went underway.

Once the patient was anesthetized and all laparoscopic ports were prepared, the shelf with the x50 was slid into place below the primary display monitor. The SV-2 input was switched to DVI to display the Clearview 9-point calibration procedure and the eye tracking began. Due to a slight lag introduced by video capture as well as to avoid safety issues which could arise in the case of PC failure, the SV-2 input was switched back to the original VGA input immediately after eye tracking was started. In this way the surgeon could continue the operation using same video output as usual from the Stryker video router, shown as blue VGA line in Figure 4.1. Eye tracking was stopped when the patient's gall bladder and the laparoscope were removed, at which point the display monitor was no longer the primary focus of operation.

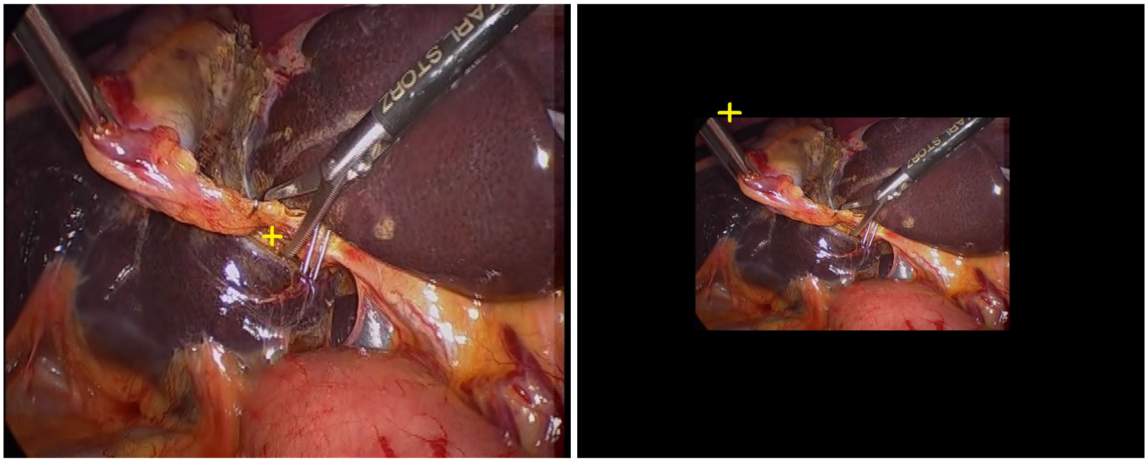
4.4.2 Post-operation watching

During the operations described in Section 4.4.1, although the recorded video from the laparoscope was captured at 720×480 - much lower than the SV-2's native resolution of 1280×1024, the image was automatically expanded to fill the display's viewable area. Recorded points of gaze in screen space still occupied the range of [0,720] horizontally and [0,480] vertically.

For recording gaze location using the AVI video stimulus in Clearview 2.7.1, the video stimulus is displayed in its native resolution in the centre of the display, with black borders filling the remainder of the display. However, eye tracking coordinates are still recorded in the range of the display's native resolution, allowing gaze location to be tracked in the empty black space. This undesirable effect is demonstrated in Figure 4.2. To have the recorded surgical videos fill the screen for the Clearview AVI video stimulus as they were in the OR required a simple non-uniform resizing of the video and using the new video as the stimulus.

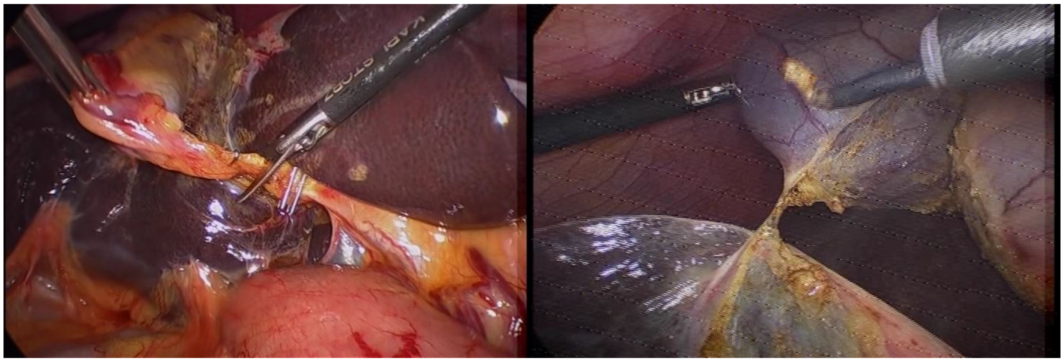
Cases to be reviewed were split into two logical parts. The first part covers the beginning of the procedure from the time when all laparoscopic ports are complete, to the point when the patient's cystic duct and cystic artery have been isolated from the surrounding tissue, clipped and severed. The second part covers the procedure from the moment of cutting the cystic duct and cystic artery to complete dissection of the gall bladder from the liver bed. Screenshots of a procedure near these events are shown in Figure 4.3.

The available cases were scaled and trimmed according to these event markers. At



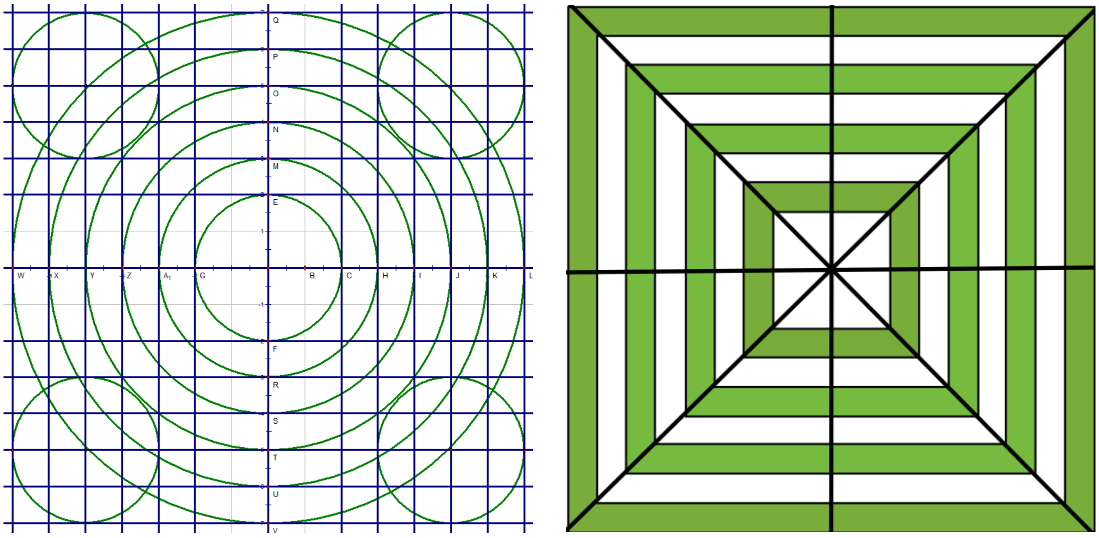
(a) (b)

Figure 4.2: (a) The surgical scene as viewed during the live operation, with eye tracking performed at 720 × 480. A yellow cross shows an example point of gaze placed at coordinates (341, 245). (b) How the raw recorded 720 × 480 would appear during the watching phase in Clearview, eye tracking at 1280 × 1024. The yellow cross is placed at what would be coordinates (341, 245).



(a) (b)

Figure 4.3: (a) Part 1 of each case ended when the cystic duct and cystic artery were cut using scissors. (b) Part 2 ended when the last piece of connective tissue between the gall bladder and liver was cauterized.



(a) (b)

Figure 4.4: Partial calibration grid samples. Actual grids used were larger. (a) Circular pattern for observing aspect ratio properties, (b) right-angle grid for determing field-of-view differences.

least 3 months following the dates of the surgical cases, the operating surgeon was asked to return to the lab to watch their own case while being eye tracked. The same PC and x50 eye tracker were used, with a different 17” 1280 × 1024 LCD monitor in a quiet, enclosed room. The display was arranged to be viewed comfortably while seated. Seven cases from three surgeons were reviewed, and one of the cases was reviewed by three different surgical residents.

In order to compare the eye tracking coordinates from the OR (range $x = [0, 720]$ and $y = [0, 480]$) with the coordinates from post-operative review in the lab (range $x = [0, 1280]$ and $y = [0, 1024]$), it was necessary to map points in one scene to the coordinate scale of the other. To complicate matters, it was found that due to peculiarities in the output from the Stryker video router in the OR, the scene displayed on the SV-2 had slightly different fields of view between the original VGA and eye tracked DVI inputs.

To more precisely document the behaviour of the Stryker video system, two regular calibration grids with 10mm grid spacings were created and brought back to the OR. One grid contained a series of circles, aimed to determine the scaling behaviour of the Stryker system, and the other grid used a series of concentric squares which could be counted to determine the visible area of the scene. Samples of these grids are given in Figure 4.4.

The eye tracking setup was re-installed into the OR as before, and the laparoscope was fixed in place in a horizontal orientation, facing a vertical surface. A calibration grid was adhered to the surface, and a photograph of the SV-2 monitor was taken on each of the VGA (original) and DVI (eye tracking) inputs. The second calibration grid was attached to the vertical face without moving anything else, and again photographs of the display monitor were taken and the eye tracking equipment was shut down and removed from the OR. Lastly, screenshots of the captured digital video were taken for each of the two calibration grids and copies of these screenshots were scaled up to 1280×1024 resolution.

In the captured screenshots at the original 720×480 resolution, the circles in the calibration grid appeared normal. In the upscaled screenshots as well as both of photographs from the VGA and DVI inputs, the circles appeared vertically elongated. From this it was concluded that roughly the same non-uniform scaling was performed by Stryker on the VGA channel and by Clearview on the DVI channel.

Next, it was found using the square calibration grid that more of the scene was visible along the horizontal axis when using the eye tracker, compared to what was seen by the surgeon operating on the VGA channel. Thus for example, if the surgeon gazed at some anatomy at the extreme left boundary of the display during the operation on the VGA channel, this would map to the anatomy being slightly closer to the display centre if he were viewing through the DVI channel nonetheless his gaze would reported at the screen boundary and would fall on the incorrect anatomy if overlaid onto the captured video. Figure 4.5 offers a visual explanation of this effect.

By counting the number of grid lines visible in the photographed and screen-captured scenes, a simple transformation of the horizontal gaze location given in Equation 4.1 was devised to map the reported gaze horizontal gaze coordinates from the range $x = [0, 720]$ to a slightly narrower range of approximately $x = [26, 694]$ to account for the parts of the captured scene at the left and right sides which were not visible while the surgeon was operating. The values 127 and 137 were determined from the number of visible grid lines in the photographed and captured frames, and 360 is one half of the captured video frame's horizontal resolution. After the transformation was applied, the x-coordinate could then be scaled to match display's native resolution. No such transformation was necessary for the y-coordinate and a simple scaling operation could be done immediately. With the gaze coordinates from the OR scaled and corrected, they could then be directly compared to the gaze coordinates obtained from the post-operation sessions in the lab.

$$X_{VGA} = [(X_{DVI} - 360) \times (127/137)] + 360 \tag{4.1}$$

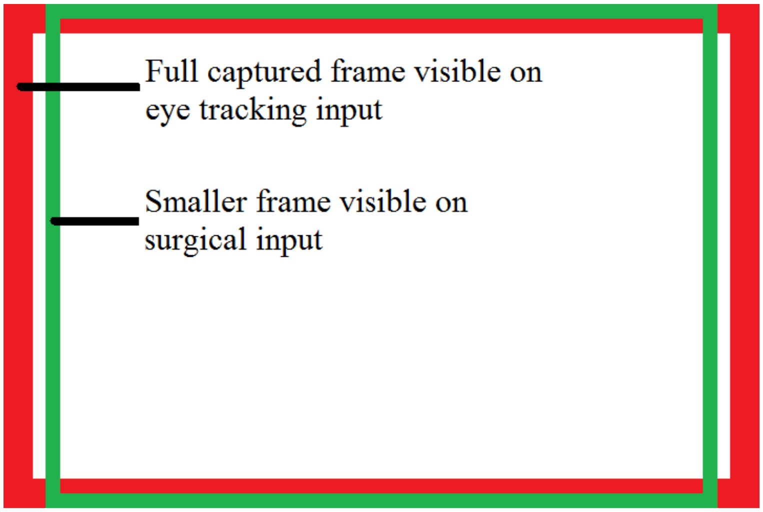


Figure 4.5: The full frame captured from the Stryker video router via S-video was scaled to fit the full display dimensions when the eye tracking (DVI) input was used. The smaller inner frame was scaled to the full display dimensions when the original display connection (VGA) was used. As a result, the screen space coordinates of surgical anatomies between the two display inputs would begin to deviate as the physical location of the anatomies approach the horizontal periphery, with the extreme periphery of the captured frame not even visible on the original VGA connection. A scaling correction was required to map recorded gaze points back onto the location which was actually viewed through the VGA connection during the live operation.

4.5 Data analysis

4.5.1 Inactivity periods

While the surgeons and trainees watching the previously recorded videos in the laboratory environment were expected to have their attention focused on the display monitor at most times, this was not the case for the surgeon performing the operation live. In particular, the surgeon was not likely looking at the display monitor while instruments were being changed - thus, periods of inactivity were defined as intervals during the recorded video when the surgical instruments were not in the scope's field of view. These intervals were manually annotated and were omitted from comparison between the operator and viewers.

Due to a tendency for the operating surgeon to move his head during the operation, there were many occasions when the surgeon's head was near the boundary of the eye tracker's operating range. This caused periods where the eye tracking data flickered rapidly between valid and missing data, even if in actuality the surgeon's gaze remained focused on the display monitor.

To reduce the detrimental effects of this missing data, a simple gap-filling operation was performed on the recorded data. Small gaps of missing data at most 10 samples in duration (200 ms at 50 Hz) were filled in using linear interpolation between the valid data points enclosing the gap. This technique is covered in further detail in Appendix A.

4.5.2 Pairwise gaze overlap

Pairwise gaze overlap is a novel technique we developed for providing a summary of the proportion of time two viewers of the same dynamic visual stimulus spent looking at the same place. With two temporally aligned gaze recordings, a gaze data sample from one recording is paired with a gaze sample from the second recording with the nearest timestamp. If both gaze samples are valid, the Euclidean distance between them in screen space is calculated. If the calculated distance is less than some specified threshold in screen space or visual angle, the pair is said to be overlapping; otherwise the pair is said to be mismatched. Examples of overlapping and mismatched pairs are shown in Figure 4.6.

The overlap was used in two ways. First, the number of overlapping gaze sample pairs was summed and divided by the total number of valid gaze sample pairs to return an overall summary of the duration two people spent looking at similar parts of the display. Second, as spurious mismatches could occur frequently, of greater interest is contiguous intervals of mismatched pairs, since this indicates prolonged gaze at different surgical anatomies and in

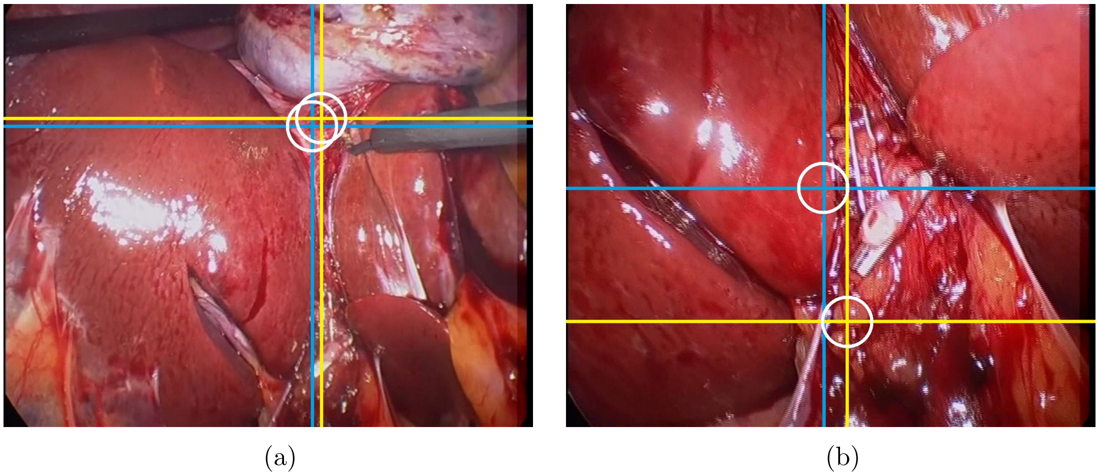

(a) (b)

Figure 4.6: (a) Two viewers on a common stimulus with gaze overlapping within 3° visual angle (diameter of white circle). (b) Gaze mismatch where viewers are looking at different parts of the scene.

turn possibly different cognitive processing between viewers. Such mismatch intervals were identified by post-processing timestamped mismatch gaze sample pairs and highlighting any intervals lasting longer than one second. Finally, the pairwise overlap values can be plotted against the task timeline as shown in Figure 4.7, allowing a visual analytic approach to identify mismatch or overlap periods.

4.6 Results

Using the summary measure of overlap duration, when expert surgeons reviewed the cases they personally performed, they showed a 59.9% overlap of gaze within approximately 3° visual angle. For the cases reviewed by junior residents, there was only a 54.4% overlap. All subjects showed a higher amount of overlap (58.7%) when watching the initial isolation of the cystic duct and cystic artery which required higher precision, than when watching the comparatively straightforward gall bladder and liver separation in the latter part of the procedure (55.2%). These results are repeated in Table 4.1.

Correspondingly, mismatched intervals seemed to occurr less frequently in the first part of the surgical procedure (31 intervals in 598 seconds = 0.05/s) than in the second part (9 intervals in 121 seconds = 0.07/s). Detailed results are given in Table 4.2, although with this small sample size, no tests of statistical significance were performed.

For the cholecystectomy case with multiple observers, mismatch intervals were aligned

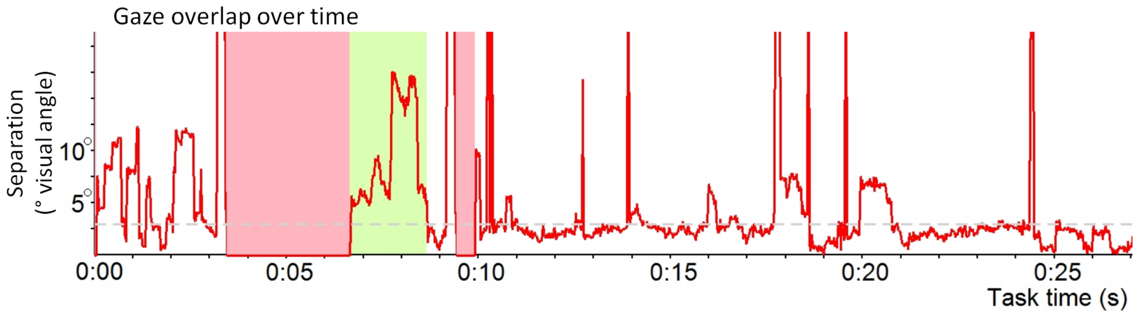


Figure 4.7: A portion of a pairwise gaze overlap plotted on a timeline. A light grey horizontal dashed line indicates roughly 3° visual angle. Pink highlighted regions are missing data, and the green highlighted region shows a detected period of gaze mismatch lasting approximately two seconds.

Component of surgical procedure	Viewer	
	Operator ($N = 3$)	Resident ($N \leq 3$)
Isolation and cutting of cystic duct & artery	61.0%	56.4%
Dissection of gall bladder from liver bed	58.4%	50.2%
Full procedure	59.9%	54.4%

Table 4.1: Proportion of procedure time when viewer's gaze overlapped with operator's gaze within 3° visual angle.

Component of surgical procedure	Viewer	# mismatched intervals	Mean interval duration (s)	% task time mismatched
Isolation and cutting of cystic duct & cystic artery, 598 seconds	Operator	20	1.47	4.9
	Viewer 1	49	1.48	12.1
	Viewer 2	25	1.42	5.9
	Viewer 3	30	1.52	7.6
	Mean	31	1.47	7.6
Dissection of gall bladder from liver bed, 121 seconds	Operator	5	1.58	6.5
	Viewer 1	14	1.58	18.3
	Viewer 2	7	2.09	12.1
	Viewer 3	9	1.75	13.0
	Mean	9	1.73	12.5

Table 4.2: Mismatched intervals when comparing operator's gaze to viewer's gaze

across all viewers. It was found that two intervals were common to all viewers, and there were four intervals common to all viewers but one.

4.7 Discussion

While the technique of measuring overlap is simple in design, it shows that experts watching their own cases produced greater overlap with their operating gaze than external reviewers. Therefore, for the purpose of creating instructional videos, it would be better to collect gaze from the operating surgeon.

The utility of identifying mismatch intervals comes from its educational value. The simple existence of periods during an operation where external viewers all deviate from the gaze location of the operator implies some unique decision being taken at the time of operation which was not apparent to viewers. Knowing when such intervals occur can allow an instructor to investigate in detail the circumstances which could have led to such disconnection. In the particular case for this study, such disconnects occurred when the grasper had released its captive tissue and was seeking another location to grasp - this interval could be used to teach learners where a good grasping site could be quickly found.

From the results presented here, although the sample sizes are small, it seems that hypothesis 2 is supported both in the higher overlap and fewer mismatch intervals when surgeons watch recordings of cases they personally performed. However, hypothesis 1 is not well-supported as the actual overlap amount is below 60%.

4.7.1 Challenges with eye tracking during live surgery for training

Practical application of this technique as a new training tool comes with challenges. As the efficacy of new training programs ultimately must be evaluated based on clinical outcomes, a long-term study with a cohort of surgical trainees should be done for proper evaluation using a set of defined clinical criteria of task proficiency and training time. If students are able to reach expert-level proficiency in fewer laparoscopic operations with the additional gaze-based mentorship, then this should lead to improved patient outcomes.

Data collection itself for such a training protocol poses a challenge. Ideally the expert surgeon's gaze will be recorded live in the OR, but the dynamic nature of the environment leads to increased likelihood of data loss. Eye tracking the expert offline in the lab can yield higher quality gaze data, but the surgeon may not be in an operating mentality and most importantly this doubles the time commitment required of the surgeon, whose schedule often cannot accommodate this.

Though still in its early stages, the research presented in this chapter established the process for live OR data collection and built upon the research of Gesierich et al. [13] to bring eye tracking action execution and observation into the surgical domain, and more clearly identified necessary challenges to overcome for future pursuits in this direction.

4.7.2 Threats to validity

The use of pairwise overlap in its current form makes it greatly affected by the quality of recorded gaze data which in turn affects experimental results. For example, if comparing two temporally-aligned fixations on the same physical location on the stimulus, one jittering with extremely large dispersion, and one with high precision, the Euclidean distance between each sample pair may be larger than the overlap threshold and not be counted as overlapping even if the true points of gaze overlapped exactly. Additionally, periods of missing data in one or both gaze recordings will produce no overlap, and when a significant portion of any recording is missing, the reported overlap summary may not be accurate to ground truth. Furthermore, as pairwise overlap is counted without consideration for fixations and saccades, a saccade in one gaze recording which passes through the location of a fixation in another recording may generate an overlap for one time sample, even though technically the saccading subject was not looking at the other subject's fixation location.

4.8 Research contributions

In this chapter we successfully collected gaze data during live laparoscopic surgical operations using a remote eye tracker. We designed and employed a new technique for comparing the similarity of two gaze recordings on a common visual stimulus and for identifying periods of gaze mismatch when watching a surgical recording, which hints at some potential for educational use.

Chapter 5

Eye movement patterns in execution and observation of simulated laparoscopic tasks

In the previous chapter, eye tracking data were collected from the expert primary operator during real laparoscopic cholecystectomy cases as well as the expert's gaze while reviewing the case post-hoc, and a method of identifying gaze pattern discrepancies was introduced. The live surgical task was complex and offered great flexibility in terms of possible action sequences which could achieve the task goal, and only a limited number of such cases were recorded with great logistical difficulty with scheduling and data collection. We were still motivated by the sub-goal of making educational surgical videos proposed in the previous chapter from expert gaze recorded either live or post-hoc. For this we needed to answer the question of how similar a laparoscopic operator's first-hand gaze patterns will be to those of his gaze while watching the recorded operation and to the gaze patterns of a different person watching the same operation. A much simpler task was devised to tease out these differences, and was carried out under more controlled conditions. Like the live surgical setting before, this study was designed in two phases with an active task execution phase followed by a passive observation phase.

A number of considerations were made in designing a suitable experimental task. Firstly, the task needed to be simple enough with clearly delineated sub-goals that subjects would be able to complete in the correct order with minimal intervention from the experimenter to avoid interrupting the focus of the participant. For simplicity of data analysis, a fixed-position internal camera in a laparoscopic training box provided a static field of view for the task scene. Trials selected for review needed to have enough reviewers for statistical power, yet the total time spent by each reviewer had to be kept low to reduce effects of fatigue. The reason for collecting multiple reviewers on a common video was to establish whether there would be any common watching patterns between multiple viewers; if this turns out to be the case, we could distribute the generation of training videos across several experts, thereby reducing the time commitment from any one expert. Furthermore, to reduce effects of ownership during reviewing, subjects needed to be blind to whose trials were being reviewed. For this study we decided to recruit student volunteers who were naïve to laparoscopy, with the assumption that the gaze behaviour from these novices would be transferable to the gaze behaviour of surgical trainees and also with experts.

The chosen laparoscopic task was a series of unimanual transfers with a fixed camera angle providing the operator with a consistent field of vision. This task is based on a component of the Fundamentals of Laparoscopic Surgery (FLS) evaluation curriculum. Use of a laparoscopic grasper is novel enough that naïve operators experience difficulty with motor mapping, yet the transfer subtasks were short and geometrically arranged such that the course of the next subtask was always clear and always in the field of view.

5.1 Hypotheses

Based on results from the previous chapter, we hypothesized that:

1. gaze overlap (as described in Section 4.5.2 would be different when actively performing a simple laparoscopic task and when watching a recorded video of the same task, and
2. that there will be a smaller difference in overlap between all watchers (including the original operator) than between an operator and a watcher.

If these can be shown, then they may imply that expert gaze must be collected live for making gaze training videos.

5.2 Independent variables

The source of gaze recordings (operator live, operator watching, third party watching) used for pairwise overlap analysis is used as the independent variable of this study.

The specific pairwise overlap comparisons done are:

- Operator live × operator watching

- Operator live × third party watching
- Operator watching × third party watching

5.3 Dependent variables

The pairwise overlap summary is reported in this study's results.

5.4 Apparatus

For the active task execution phase, the task display and eye tracking were performed by a Tobii 1750 remote eye tracker with an integrated 17" LCD monitor. This was raised on a height-adjustable platform and set above a physical laparoscopic training box with a fixed internal camera (3D Technical Services, Franklin, OH), as shown in Figure 5.1. These items were arranged so that the display was at eye-level with a viewing distance of approximately 60 cm, and a single laparoscopic grasper and trocar (Ethicon Endosurgery, Cincinnati, OH) inserted into the upper right-hand port of the laparoscopic training box could be used comfortably in a standing position. The Tobii 1750 has the same 50 Hz eye tracking capability as the x50 used in the live surgical setting of the previous study. A Logitech C510 web camera was installed near the base of the display monitor, capturing video of subjects head movements at 30 Hz / 640 × 480 resolution. The video stream from the training box was saved on the eye tracking PC as 352 × 288 resolution video at 29.97 Hz through a Hauppauge HVR 2250 PCI capture card through the NTSC composite video input channel. The eye tracking and video captures were coordinated by Tobii Clearview 2.7.1 software.

A circular wooden peg board of approximately 15 cm diameter was securely taped to the inside of the laparoscopic training box. Three coloured plastic cups were set into the peg board in a roughly equilateral arrangement, approximately 7 cm apart. A red cup was set in the northern position; the southwestern position held a green cup, and a blue cup was fixed in the southeastern location. Additionally, a small 1 cm square of medical hook & loop tape (hook side) was fixed in the centre of the triangle, serving as a home position, as shown in Figure 5.2. A small green rubber cylinder approximately 1 cm long and 3 mm thick served as the item to be transferred between the cups using the laparoscopic grasper.

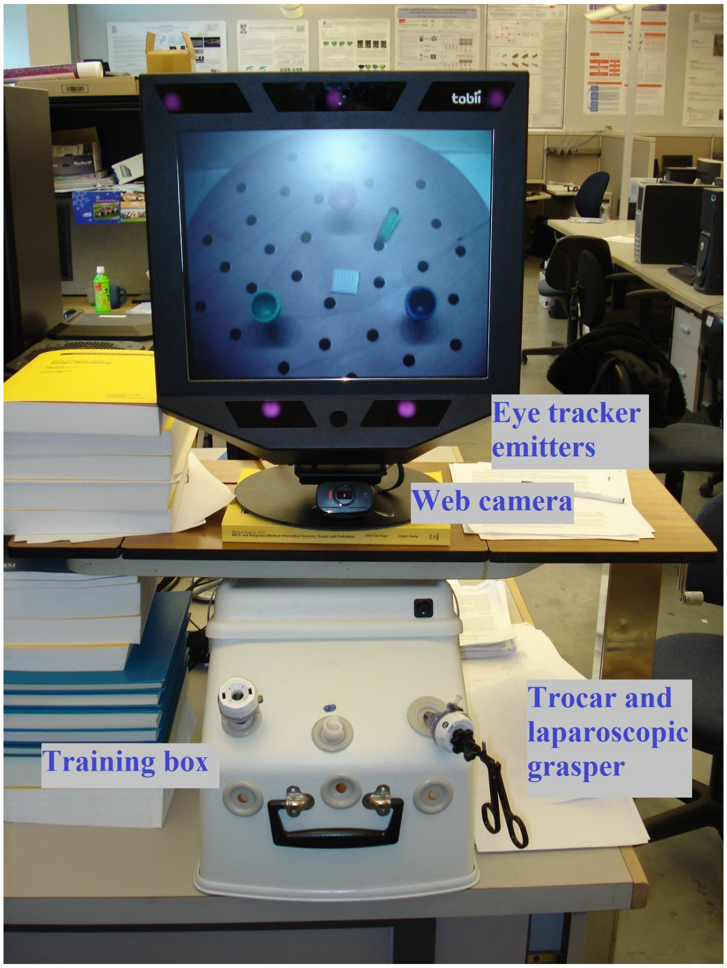


Figure 5.1: Phase 1 experimental setting for an active laparoscopic peg transfer task.

Figure 5.2: Screen capture of the peg transfer physical arrangement. While the physical arrangement is uniform, this view presents a perspective transformation due to the oblique viewing angle of the training box's fixed internal camera.

5.5 Description of participants

Fourteen subjects were recruited from the Computing Science and Engineering graduate student laboratories at Simon Fraser University. As the instrument of the apparatus entered from the right side of the training box, all chosen subjects were right-handed, and the average age was 28 years. There were 9 males and 5 females, and 8 subjects wore eyeglasses for corrected-to-normal vision. None had any experience with any form of laparoscopy.

5.6 Procedure

5.6.1 Phase 1: task execution

In this phase of the study, the required manual task was to use the laparoscopic grasper to transfer the rubber cylinder between the three coloured plastic cups on the peg board. Initially the rubber cylinder was located in the northern red cup. The cylinder was to be transferred to each cup in counter-clockwise order ending at the initial configuration with the cylinder returned to the red cup. After each successful transfer to a cup, the empty grasper was to make contact with the central home square. Thus a complete trial consisted

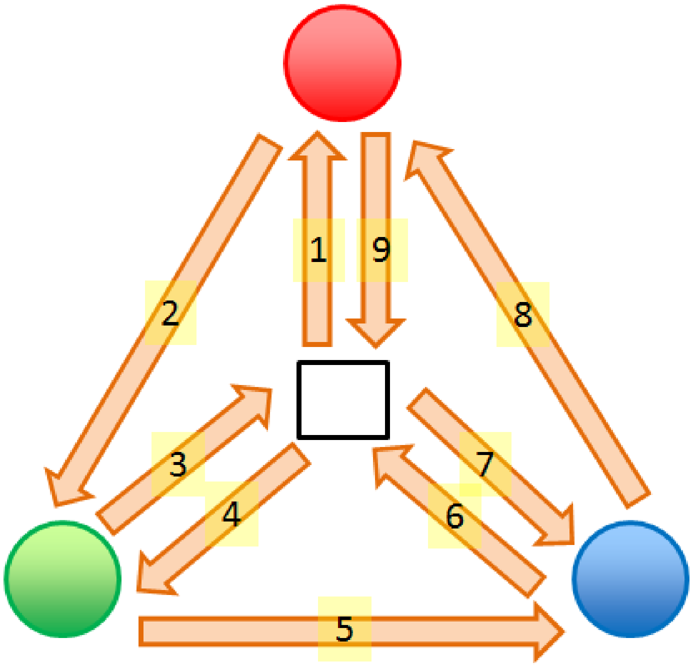


Figure 5.3: Order of peg transfer subtasks. Subtasks 1, 4, and 7 involve reaching for and grasping a peg from a cup; subtasks 2, 5, and 8 involve transferring and releasing a peg into a destination cup; homing of the empty grasper is done in subtasks 3, 6, and 9.

of three sets of subtasks each involving one reaching and grasping action where the cylinder is picked up from the originating cup, one transfer action where the cylinder is released into the target cup, and one homing action where the empty grasper is returned to the home position. The subtask sequence is outlined in Figure 5.3.

Subjects were briefed on the nature of the study and informed that their participation would involve returning for another observation session. After signing their consent to participate, subjects were given a background questionnaire for collecting demographic data. A sample of this questionnaire is attached in Appendix B.

The experimenter performed a demonstration of the required task and subjects were allowed to practice for five minutes. Once practice time expired or subjects felt they were ready to proceed, they were calibrated on the eye tracker and performed five recorded trials of the peg transfer task with short breaks in between each trial. After completion of the fifth trial, a brief questionnaire was administered to assess subjective task difficulty and stress level, and the task execution session was concluded.

5.6.2 Phase 2: task observation

In the task execution phase, 14 subjects performed 5 trials each for a total of 70 recorded trials. From these 70 trials, a bank of 10 recordings was selected to be reviewed by multiple subjects in the task observation phase. The 10 recordings were selected by the following process: to begin, trial 3 for each subject was added to the pool. If trial 3 was recorded with poor quality (i.e. large amount of missing data), then the better of trial 2 or trial 4 was selected. Finally, the 10 best quality recordings were kept. Thus, 10 different subjects each owned one recording in the bank while the remaining 4 did not own any video in the common pool for the observation phase.

Subjects were required to watch 10 recordings altogether, comprised of their own 5 trials and 5 pseudo-randomly selected non-owned videos from the common pool. Per subject, the order of videos to be watched was pseudo-randomized, with the common videos distributed in such a way that each common video was reviewed by exactly 7 subjects including the owner. All videos were assigned a pseudo-random 3-digit numeric identifier to obfuscate ownership from the participants.

As was needed in the UBC OR study of the previous chapter, all videos for review had to be resized to fit the display monitor's native resolution. However, unlike the UBC OR study, there was no input switching in this experiment. Thus there were no anomalous differences in the perceived and recorded fields of view and a simple non-uniform resolution scaling was done.

At least two weeks following their participation in the task execution phase, each subject was contacted by e-mail to return for the second component of the study. For this phase, the laparoscopic training box was removed and the eye tracker and LCD monitor were lowered so that subjects could view the recorded tasks from a seated position. Subjects were refreshed with an explanation of the previously recorded task, calibrated for the new viewing position, and watched their 10 assigned recordings in the experimenter's prescribed order. After viewing each recording, subjects completed a brief questionnaire (Appendix B.2) to assess the subjective stress endured during viewing, as well as to inquire whether subjects felt they recognized the recording as their own. Once the tenth recording and questionnaire were complete, subjects were dismissed from the study.

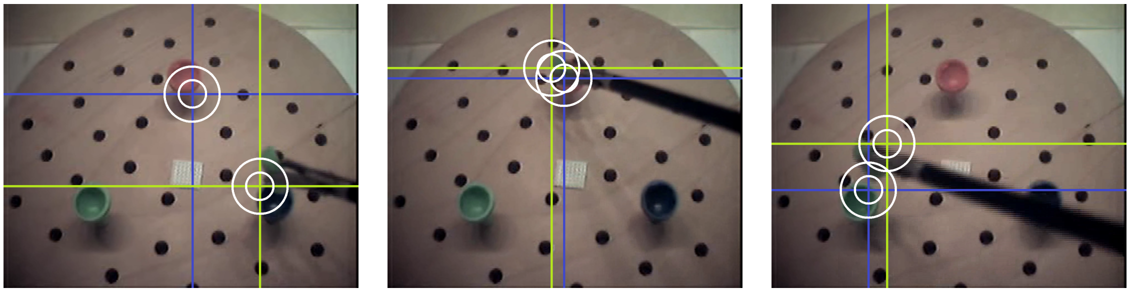
Subject #	Usable trials	2.5° overlap (% time)	5° overlap (% time)
1	5	65.4	81.8
2	4	57.3	73.2
3	2	72.6	86.9
4	4	76.6	80.9
5	5	82.0	86.8
6	5	42.4	86.0
8	5	75.6	82.5
9	4	74.9	86.6
10	5	55.6	77.9
12	5	69.1	76.9
13	4	84.2	89.2
14	3	64.6	85.8
Mean, std.deviation			
		(% time)	(% time)
Doing vs self-review		67.9 ± 15.4	82.5 ± 6.4
Doing vs 3rd-party review		70.1 ± 10.6	81.2 ± 6.7
Self-review vs 3rd-party review		74.7 ± 9.2	86.9 ± 6.4

Table 5.1: Mean doing vs self-review overlap, with mean 3rd-party overlap.

5.7 Results

For this study, gaze overlap was calculated in the same way as described in Chapter 4.5.2. Recordings compared included operator's active gaze (doing) vs operator's passive gaze (self-review), doing vs 3rd party review, and self-review vs 3rd party review. Due to a loss of eye tracking during a large proportion of the task time in 19 trials including all trials from subjects 7 and 11, the trials where fixations accounted for less than 72% of the task duration were omitted from the doing vs self-review analysis. The threshold of 72% fixation time over total task time was chosen in accordance with empirical data of fixations during scene observation [40] with allowance for some reduced fixation detection due to recorded data loss.

Table 5.1 lists the percentage of task time where the active and passive gaze for self-review overlapped, averaged over each subject's trials for overlap parameters of 2.5° and 5° visual angle. Overlap and non-overlap at these two parameters are demonstrated in Figure 5.4. Note that the central homing position was estimated to be separated from the cups by about 9.2° visual angle and the cups were separated from one another by roughly 17°. Hence both 2.5° and 5° for analysis are sufficient to distinguish gazing on separate objects in the scene while still forgiving some gaze jitter.



(a) (b) (c)

Figure 5.4: Example of (a) non-overlap and overlapping points of gaze at (b) 2.5° and (c) 5° visual angle. The blue lines indicate the point of gaze of the operator, and the green lines indicate the observer's point of gaze.

5.8 Discussion

From Table 5.1, the overlap amounts are similar for the cases doing versus both self-review and 3rd-party review, supporting hypothesis 1. Moreover, gaze patterns show higher concordance when both data streams are from passive video review, supporting hypothesis 2. This indicates that most people will view a task in a similar way, regardless of whether or not there is ownership of the task. Conversely, performing a task first-hand produces different eye movement overlap that cannot be fully reproduced simply from a passive review of the recorded task. The results for 2.5° are also less stable than for 5°, likely with vulnerability to jitter or suboptimal calibration being contributing factors.

However, rigorous statistical analysis of the results is difficult because the self-review and 3rd party review averages were obtained in different ways. More specifically, every subject produced 5 recordings, all of which were used for self-review. In an effort to increase viewership for 3rd party review without making participation prohibitively lengthy, only 10 of the original 70 videos were used; each one was viewed by 7 other participants. Due to this viewership imbalance and the lack of participation from experienced laparoscopic operators, only a descriptive analysis is provided.

In this study, when a video was watched by multiple reviewers including the owner, the gaze points overlap within 2.5° for 75% of the task time, suggesting that a common gaze pattern was employed by all reviewers. In contrast, overlap was lower when comparing an operator's gaze to self-review (68%) and 3rd-party review (70%), which may be explained by a gap in visual reaction as reported by Atkins et al. [1] or a lack of planning and control while watching passively, regardless of procedural knowledge. With the main task broken down into 9 discrete tool movements, the total watching delay can comprise roughly 5-20%

of the task duration, reflected in the reported overlaps. Such visual delays in pursuing moving objects have also been observed by other researchers [45, 6].

Atkins et al. [1] showed that for this particular peg transfer sequence involving 9 subtasks, the saccadic lag during video watching could range from 100 to over 700 milliseconds, with typical delays around 550 milliseconds per subtask. For subtasks lasting on average approximately 3 seconds each, it is conceivable that up to 16% of the non-overlapping task time can be attributed to the lag by watchers.

Having the tightly-defined sub-task sequence eliminated any instances of subjects completing the required sequence out of order; however there was little opportunity for deeper decision making, so higher performance was more likely attributed to better manual control rather than a better understanding of the necessary actions.

For the purpose of studying manual skill acquisition, the five execution trials each subject performed would not have provided sufficient repetition to reach a more proficient skill level. However, even with only novice participants, we learned that gaze behaviour differs sufficiently between doing and watching. As such, overlaying watching gaze on a recorded video may not be suitable for training.

For gaze training using expert gaze overlaid on surgical videos to be effective, there must first be a difference between expert and novice gaze patterns, which we need to identify.

Since this study still lacked participation from subjects already possessing laparoscopic skills training, an extension study was needed involving expert participants, described in the following chapter.

5.8.1 Threats to validity

In addition to the existing issues with pairwise overlap previously discussed in section 4.7.2, the subjects in this study were all students from the Computing Science and Engineering graduate laboratories and represent neither the general population nor those potential future surgeons. With only 5 repetitions of the transport sequence performed for each subject, the results could be confounded by learning effects as subjects were still familiarizing themselves with the mechanics of laparoscopic manipulation.

5.9 Research contributions

In this chapter we showed that novices watching a simple laparoscopic task showed more similarity to one another than they did to the gaze pattern of the active operator. Part of the dissimilarity resulted from a temporal lag in the watcher's gaze compared to the movements of the operator's point of gaze. A second study was still needed to determine whether this observer lag still holds with expert laparoscopic operators.

Chapter 6

Eye-hand coordination in a simulated laparoscopic task

The peg transfer study detailed in the previous chapter established some fundamentals of eye movement behaviour differences during execution vs observation, discouraging us from continuing towards the use of 3rd-party observation for generation of surgical training videos. However, the statistical analysis was difficult as the manner of video distribution to observers led to unequal counts of watching self and watching others. Furthermore, all subjects were effectively laparoscopic novices and so it was not possible to make observations about differences between novice and expert eye-hand coordination.

The protocol was revised to address these points and a follow-up experiment was conducted at Vancouver General Hospital, using both experts and novices. From this data set covering a range of expertise, we hoped to identify an eye-hand coordinatio measure to distinguish levels of laparoscopic ability.

As stated in the introduction, earlier research showed that one common approach for novices is to use "tool tracking", rather than experts who use a target locking approach [29]. To emulate these results in our setting would require knowing the distance between the tool tip and the eyegaze location at each gaze point, which was not readily measurable with our available experimental equipment.

Other expertise measures used in similar environments include a "target-locking" score [60, 59] and fixation characteristics [63], although these are more appropriate for the head-mounted eye trackers used in their respective studies.

For a remote video eye tracker and an experimental setting where there is a high possibility of data loss due to subjects making gross changes in body position, a non-fixation-based gaze parameter capable of distinguishing expert and novice behaviours was needed. Because overall fixation measures reported over the duration of a recording can be greatly skewed by missing or poor-quality data, an instantaneous measure using raw gaze points was more suitable for the analysis in this study.

Given our experimental setting carried over from Chapter 5 has a fixed scene with dynamic movement only provided by the tool and the peg, we proposed to identify a simple measure to separate novices and experts which only required the eye gaze points provided by the remote eye tracker (in order to see when the eye gaze landed on the target), and the moments when the tooltip reached the new target. For such a measure, we only need to annotate the moment when the tool tip arrives at the target.

6.1 Hypotheses

For this study we hypothesize:

1. that expert laparoscopic operators will complete the same laparoscopic transport sequence from the previous study in shorter time than novice operators, and
2. that experts will perform the transport task while exhibiting a different eye-hand coordination behaviour from the novice operators. This eye-hand coordination parameter is described in Section 6.8.2.

6.2 Independent variables

Experts and novices underwent the same procedure for a single expertise independent variable.

6.3 Dependent variables

The following dependent variables are measured in this study:

- Task completion time
- Tool-target separation (Section 6.8.2)

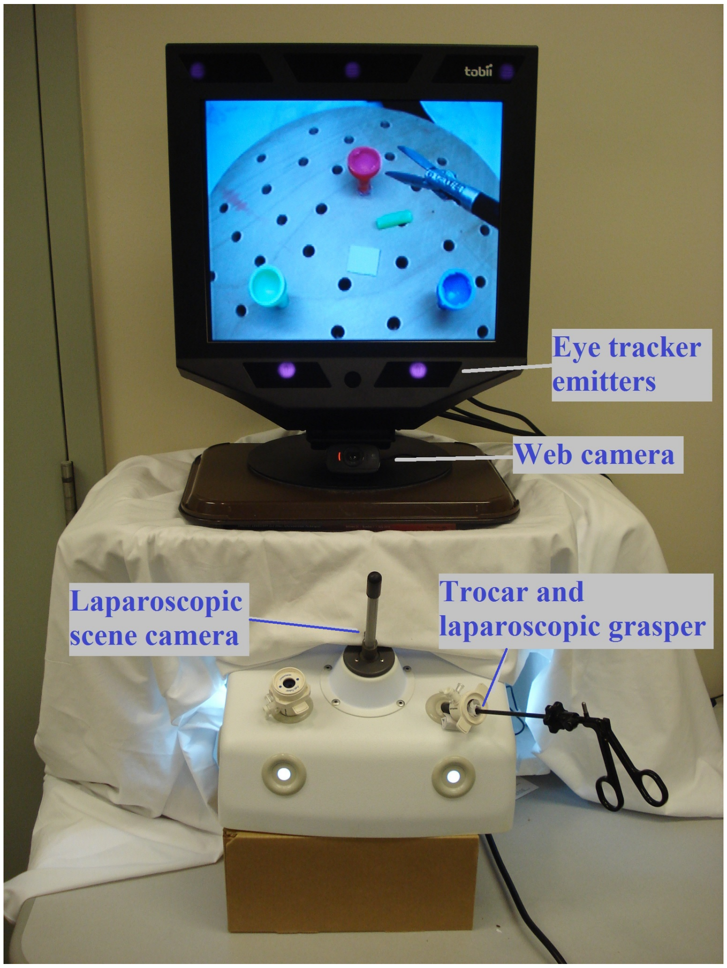


Figure 6.1: Experimental setting for second active laparoscopic peg transfer study. The laparoscopic scene camera can be repositioned, but was kept relatively still throughout the study.

6.4 Apparatus

The apparatus setup was largely unchanged from the study described in Chapter 5, except the available laparoscopic training box had a movable camera, whereas the box used earlier had a fixed camera. Throughout this study, the camera was kept in approximately the same position although there were some minor changes in the field of view between different subjects. The camera was not moved between a subject's individual trials. Figure 6.1 shows the experimental setup.

6.5 Description of participants

Surgical staff, residents, and non-medical staff were recruited from CESEI and Vancouver General Hospital. Six participants were classified as laparoscopic experts and six were novices. Of the novices, two were practicing general surgeons with no laparoscopic training. All participants were right-handed males.

6.6 Procedure

6.6.1 Phase 1 – Task execution

The required task was unchanged from the previous peg transfer study, described in Chapter 5.6.1.

Subjects were briefed on the nature of the study including the requirement of their participation in a follow-up session and gave signed consent to participate. They were given a demonstration of the task and given 5 minutes to practice. After practice was complete, they were calibrated on the eye tracker and performed 5 trials of the experimental task with short breaks in between. Subjects were dismissed after completion of the fifth trial.

6.6.2 Phase 2 – Task observation

After phase 1 data collection was complete, the subjects were distributed into two expert groups and two novice groups, with each group containing three experts or three novices. For the observation phase, after at least 2 weeks subjects watched all available recordings within their own group - specifically, their own 5 recordings plus 10 recordings from the other group members, in pseudo-random order. Video names were not completely anonymized but subjects were assigned a cryptic name which was not explicitly told.

As in the earlier study, for the observation phase, the laparoscopic training box was removed and the display monitor and eye tracker were arranged for comfortable viewing in a seated position. Subjects were calibrated on the eye tracker and then watched the videos in the prescribed order. There was no questionnaire administered between videos, and subjects' participation was concluded after the fifteenth video finished playing.

6.7 Data analysis

From the 12 subjects recruited for this study, 60 "execution" gaze recordings were obtained in phase 1. Each subject watched his own 5 trials plus the 10 trials of the other two subjects in his group, for a total of 180 "observation" gaze recordings. Trials typically lasted between 20-120 seconds depending on the expertise of the subject.

Prior to analysis, all collected gaze recordings were pre-processed using a moving average filter and linear interpolation gap filling procedure in order to restore the quality of some poorly-recorded trials. Full details of the data recovery procedure can be found in Appendix A. With a 4-sample moving average filter and 10-sample gap filling interpolation applied, 59 of the 76 originally unusable cases were recovered.

A new hand-eye coordination measure was sought to distinguish levels of laparoscopic ability in this experimental setting. For this measure, we required knowing the moments when the subject's gaze landed on the target, and when the tool tip reached the target.

First, raw gaze points and fixations (detected using an implementation of the I-DT algorithm [46]) were overlaid onto each task execution video as well as observation videos. Using the same subtask notation in Chapter 5.6.1, by manual observation, each of the 9 subtasks for every recorded trial was manually annotated with (A) the timestamp when the operator's gaze first reached the appropriate target location and (B) the timestamp when the tool first reached the appropriate site of action. For each (A) timestamp, the Euclidean distance in screen space from the centre of the target site to the midpoint between the two jaws of the surgical grasper was manually obtained. A simple transformation was applied to return the target-tool distance in degrees of visual angle assuming approximately 60 cm viewing distance to the display monitor. Figure 6.2 shows an example of a video frame where a subject's eyes first reached the green target for subtask T1, and a video frame from the same recording at the first moment when the instrument reached the appropriate target.

For many tool-tracking applications, an automatic determination of the tool tip could be performed from the task video such as that reported by Jiang et al. [22]. However, potential inaccuracies could arise from Z-axis ambiguity in such an automatic 2-D determination of tool-tip intersection with the target area. An example of where this could occur is demonstrated in Figure 6.3, where the payload was not ready to be dropped into the cup until almost two seconds after the tool first appeared above the target (59 frames at 30 Hz). Therefore target positions and tool arrival timings were manually annotated, to avoid potential errors in this time-sensitive data.

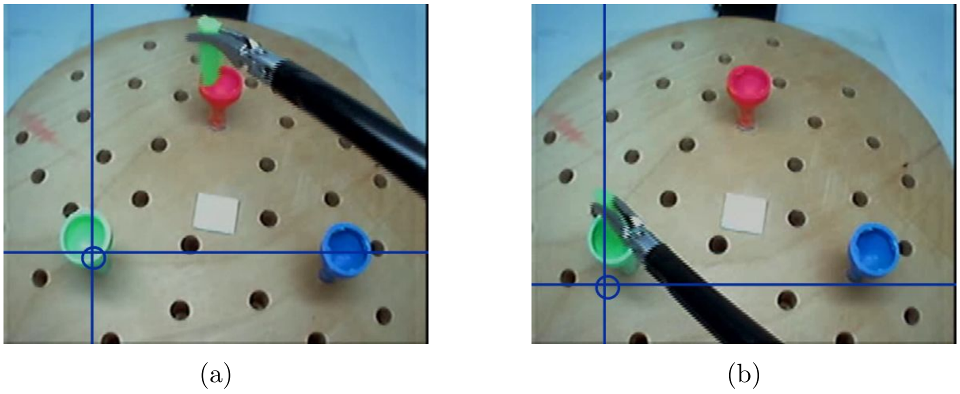

(a) (b)

Figure 6.2: (a) Moment of eye arrival to destination of transfer, (b) moment of instrument arrival to target proximity for the same transfer subtask. The circle around the point of gaze indicates a detected fixation with a radius of approximately 2.5° visual angle.

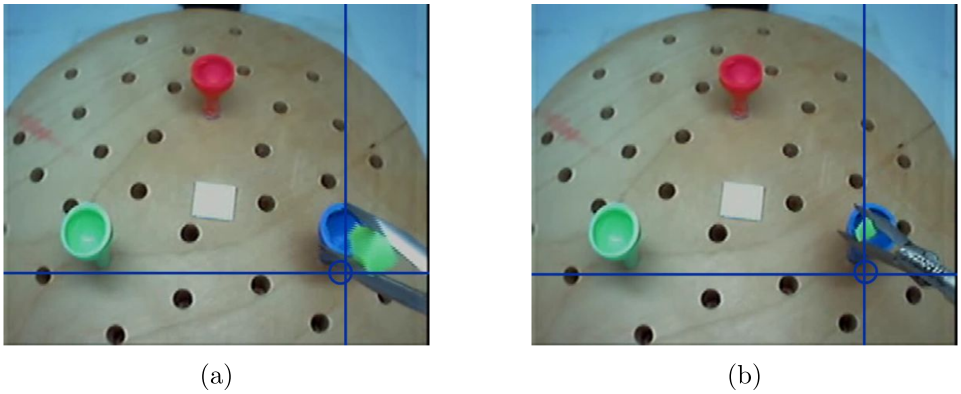

(a) (b)

Figure 6.3: (a) The 2-D location of the end effector is near the target, but is still far away along the depth axis, (b) the instrument has reached the destination in all spatial dimensions.

Subtask	Group	Mean time, std.dev (s)	p-value
Reach & Grasp (RG)	Expert	2.9 ± 0.7	< 0.001
	Novice	5.7 ± 2.4	
Transfer & Release (TR)	Expert	3.2 ± 1.1	< 0.001
	Novice	4.8 ± 1.7	
Homing (H)	Expert	1.5 ± 0.4	< 0.001
	Novice	2.4 ± 0.8	
Complete task	Expert	23.4 ± 6.4	< 0.001
	Novice	39.0 ± 13.1	

Table 6.1: The mean subtask execution times (s) separated into two groups: expert ($n = 6$) and novice ($n = 6$) in each subtask. 15 task trials were recorded for each group, for a total of 45 of each RG/TR/H subtask in each group.

6.8 Results

6.8.1 Timing

Table 6.1 and Figure 6.4 show the experts were significantly faster than the novices overall (23.4 s vs. 39.0 s) and surgeons performed each subtask faster than novices performing the corresponding subtasks. For task time, differences were revealed on the main effect of expertise ($F_{(1,174)} = 73.719, p < 0.001$) and subtasks ($F_{(2,174)} = 54.761, p < 0.001$) as well as interaction between these two ($F_{(2,174)} = 7.423, p = 0.001$). Specifically, the expert surgeons performed the task faster than novices in both the overall task and each subtask, detailed in Table 6.1 and illustrated in Figure 6.4.

A t-test indicates that over all subjects, performing the Homing subtask required significantly less time (1.9 ± 0.8 s; $p < 0.001$) than the RG (4.3 ± 2.2 s) and the TR subtasks (4.0 ± 1.6 s), but there was no significant difference between RG and TR subtasks ($p = 0.502$).

6.8.2 Eye parameters

Table 6.2 shows the mean duration and rate of fixations during the task, by expertise; significant differences between the groups are highlighted in bold text. A “quiet eye” measure similar to that used by Wilson et al. [64], here defined as the duration of the first fixation as soon as the eye acquires the target for the first time, is shown in the last row. However, as seen in Table 6.2, this measure failed to yield significant differences.

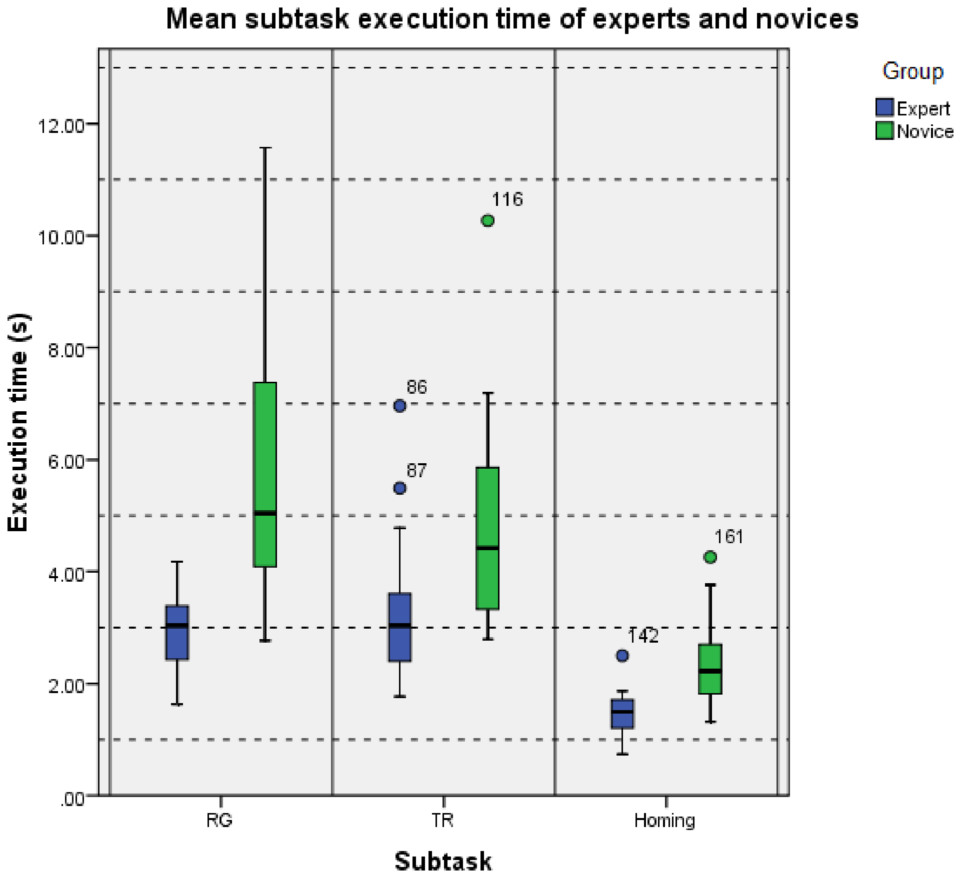


Figure 6.4: The mean subtask execution times (s) separated into two groups: expert ($n = 6$) and novice ($n = 6$) in each subtask.

Parameter (units)	Group	Value, std.dev	p-value
Mean fixation duration (ms)	**Expert** **Novice**	**824 ± 216** **707 ± 167**	**0.023**
Mean fixation rate (#/s)	**Expert** **Novice**	**1.17 ± 0.27** **1.35 ± 0.29**	**0.015**
Quiet eye fixation (ms)	Expert Novice	555 ± 485 586 ± 589	0.506

Table 6.2: Mean fixation results by expertise. t-test significant results (df=536) at the 5% level appear in bold.

Subtask type	Mean inter-target distance (°)	Group	Mean, std.dev (°) tool-target separation	p-value
Reach & Grasp (RG)	8.33	Expert Novice	3.73 ± 0.47 3.20 ± 0.82	< 0.001
Transfer & Release (TR)	11.08	Expert Novice	3.60 ± 0.65 3.02 ± 0.79	< 0.001
Homing (H)	8.33	Expert Novice	3.24 ± 0.74 2.57 ± 0.96	< 0.001
Overall	9.28 9.21	Expert Novice	3.52 ± 0.66 2.93 ± 0.90	< 0.001

Table 6.3: Expert and novice target-tool separation in degrees of foveal angle (°) at the first moment of visually acquiring the target, by subtask type. Expert-novice t-test differences for each subtask type (df=179) appear in the rightmost column.

Tool-target separation on visual target acquisition

The separation angle between the tool tip position and the target when the eye alights on the target for the first time, is defined as the tool-target separation. This can be recorded as the visual angle between the tool position and the target location at the moment captured in Figure 6.2a. Table 6.3 and Figure 6.5 show the mean tool-target separation angle for each subtask, and each group of participants. Averaged over all subtasks on an independent samples *t*-test, experts were found to visually acquire targets while the tool was farther away (3.52° of foveal angle) than novices who brought the tool closer (2.93° of foveal angle) before seeing the target.

6.9 Discussion

Prior research in surgical settings has consistently linked shorter task completion time in experts with gaze behaviors uniquely different from novices. From the expert participants in this study, we again observe rapid task completion, and also a large separation between the surgical instrument and its movement destination when the eye reaches it, supporting both of our hypotheses for this study. These results together imply that experts also operate with a greater speed of tool movement to cover the inter-target distance in a shorter time. The cumulative shorter subtask completion times allowed the experts to complete the full subtask sequence in roughly 60% of the time required for novice task completion.

In agreement with Kocak et al.'s findings [26], experts also displayed significantly longer mean fixation durations as well as a lower fixation rate defined as the mean number of

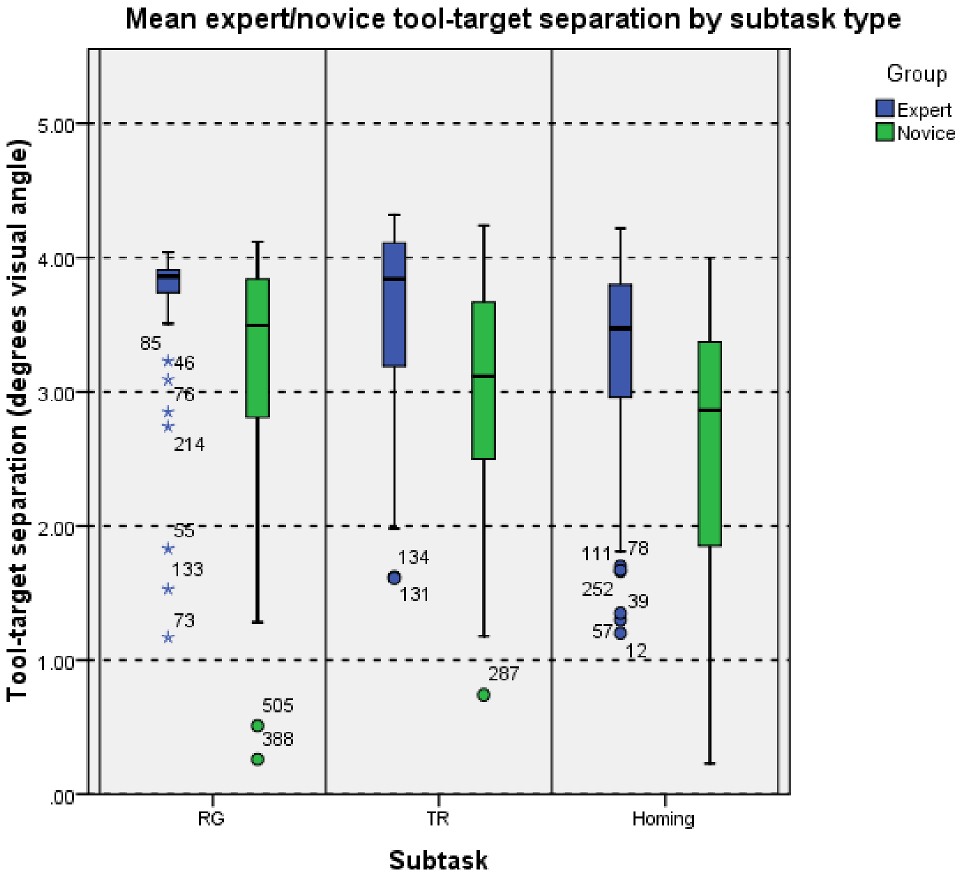


Figure 6.5: Expert and novice target-tool separation when visually acquiring the target, by subtask type. 45 subtasks per RG/TR/H were available for each group.

fixations per second.

Although experts sometimes stayed focused on the tool during movement, they tended to look at the targets soon after initiating tool movement. Experts on average saw the target with the tool farther away and did so with less variation than novices. Earlier focus on the target observed in the experts in this study allowed them to collect information on the target for guiding hand movement sooner than the novices. Although the rapid tool movements by the experts reduced the duration of "quiet eye" period operationally defined by Wilson for his task [63] thus producing a contrary result, this result still echoes the implication that a "quiet eye" phase can occur after the start of the hand movement. In a sequential movement like the surgical task we used in this study, instead of always gazing at a target before initial movement of the surgical instrument, surgeons can engage their gaze to the target while the tool is moving towards the target. Since the eye can move dramatically faster than the hand movement, experienced surgeons allocated a significant earlier gaze towards the target before the tool reached a point of interacting with the target. In contrast, novices in this study performed gaze-on-target later than expert surgeons in the hand movement stage. Without adequate visual guidance to the target, their total task times were prolonged as they struggled to place the object into the cup even though the transit of the object between its origin and destination were comparable between experts and novices. This issue is revisited in the following chapter.

In this experimental setting, target positions were known at the beginning of the task and remained static throughout each trial. Consequently subjects already familiar with the mechanics of laparoscopic manipulation may not have required as much initial visual guidance towards the targets, explaining the result seen in Table 6.3 where both experts and novices brought the tool approximately halfway to two-thirds of the way to the target before their eyes reached the target. On the other hand, as all subjects understood the task sequence well, using a more complicated, dynamic task could introduce difficulties in separating task performance from differences contributed by eye-hand coordination and task knowledge.

While it is sometimes uncertain how efforts made to distinguish surgical experts and novices purely based on eye movement characteristics can be applied to surgical education, the eye-hand coordination result presented above suggests the possibility of future skill acquisition experiments deliberately limiting vision of irrelevant areas of the scene in some form training an operator to gaze only at task-relevant areas.

The focus of our discussion has been on the distance between the tool and target upon visual target acquisition, as this is a measure not much influenced by noise in the data. Although Table 6.2 shows that the fixation data, i.e. the mean fixation duration and fixation rate were significantly different between experts and novices, these values are highly influenced by noisy data. For example, an expert individual for whom the eye tracker produces unusually jittery point-of-gaze estimates will be found to have many short fixations even if in actuality he gazed steadily. Also, missing data due to the subject moving out of the eye tracker's operating range can result in reduced numbers of fixations, in turn leading to lower fixation rates. The duration of the first fixation upon visual target acquisition is subject to the same problems as listed above; poor quality data during the moment of visual target acquisition can have a severe effect on the detected fixation, if one can be at all detected. Thus using fixation measures for distinguishing expertise requires careful consideration.

6.9.1 Threats to validity

The same novice subject selection and learning effects discussed in section 5.8.1 apply here. For this study, due to one subject with very large jitter dispersion in the recorded gaze data, it was necessary to filter the data using the process described in Appendix A, which potentially could have introduced artifacts in our analysis, particularly in determining the precise moments of saccades and target acquisition.

6.10 Research contributions

- Continuing the investigation of finding a suitable eye-based indicator of surgical expertise for the purpose of gaze training, a new instantaneous eye-hand coordination measure was defined and obtained by eye tracking expert surgeons and non-surgeons in a simulated laparoscopic task. By measuring the distance between an intended target and the current tool position, it was found that experienced surgeons looked at the target before the tool approached while non-surgeons focused their attention elsewhere until the tool had moved closer. This result promises exciting new approaches to accelerate the development of eye-hand coordination in laparoscopic surgical trainees, explored in the following chapter.

Chapter 7

Gaze training in laparoscopy

The studies described in previous chapters were designed to identify gaze characteristics for which expert and novice laparoscopic operators would exhibit different behaviors. The underlying goal throughout each study was the eventual application of using gaze behaviour for training in order to reduce the gap in gaze behaviors between novices and experts. Such "gaze training" is based on the results from gaze and performance research in domains such as sports that more rapid development of eye movement behavior towards expert-like behavior can be accompanied by more rapid manual skill acquisition.

Several different gaze characteristics were identified in each experiment described in this thesis:

- Chapter 3 – glances to a secondary display during a complex simulated surgical task
- Chapter 4, 5 – pairwise gaze point overlap in the operating room and a simple training task while actively performing a task and watching a recorded video of the same task
- Chapter 6 – tool-target separation upon visual target acquisition for experts and novices

On the basis of these results, and on others' research results, we decided that the tool-target gaze separation was a promising measure to influence performance, by implicit manipulation of the trainee's point of gaze.

On the challenge of gaze training, Vine et al. have demonstrated that explicit training of novices' eye movement behavior correlates to an accelerated rate of technical skill acquisition [60] as well as more robust retention and transfer of the acquired skills [59]. Thus we conducted our own studies to expand upon Vine's findings by training novices using a similar target-focused strategy with direct comparison to expert performance on the following measures:

- Task completion time
- Ratio of time spent looking at tool and target
- Angle of separation between tool and target when the eye first reaches the target

7.1 Hypotheses

Specifically, we hypothesize that:

1. after training, subjects receiving special gaze training will exhibit a higher tool-target separation than those receiving verbal instruction, and
2. gaze-trained subjects will have a shorter task completion time than verbally-trained subjects.

7.2 Dependent variables

All novice participants were trained by one of two gaze manipulation methods (gaze training / verbal training).

7.3 Independent variables

Task completion time and tool-target separation are reported.

7.4 Apparatus

Unlike the setting Vine et al. [60, 59] used which applied a software-based mask over the laparoscopic display, darkening non-target areas of the screen, the setting of this study allowed full vision of the scene with illuminated targets to indicate the tool movement destinations. In this way trainees were still able to make tool-tracking eye movements if they wished, but could have their focus drawn to a more salient target.

The main task resembled the peg transfer task described in Chapters 5 and 6, using a one-handed grasper to move a peg between dishes in a laparscopic training box. However,

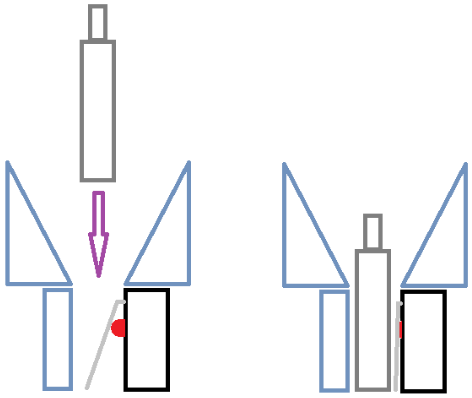


Figure 7.1: Cross-sectional diagram of target illumination switch mechanism. Insertion of a long peg activates the switch inside a target cup.

differences in the setup were needed to allow target illumination. Target illumination was accomplished using LEDs linked to a series of electrical switches activated by placement or removal of the transferred peg at the target locations. The mechanics of the physical components required some changes from the earlier study in order to ensure reliable activation of the switches. Figure 7.1 illustrates a stylized cross-section of the target switch mechanism.

The switch mechanism was constructed as follows. A levered microswitch was affixed vertically to a cup with steep sloped sides. A long metal peg inserted to the bottom of the cup applied a transverse force to depress the microswitch.

Three such cups and a simple horizontal microswitch were constructed and arranged in a similar layout as the earlier peg studies. Each target had one 5 mm red LED affixed adjacent to it. The physical peg board is shown in Figure 7.2. The cups have an outer diameter of 20 mm and are arranged in an equilateral triangle with side lengths of 103 mm. The central home position was constructed of a horizontally-mounted levered microswitch with a square of textured hooks attached. The home button measures 15 mm on each side. The transferred peg is a steel standoff with a 20 mm hexagonal shaft with a 6 mm threaded end. During the experimental trials, the threaded end was wrapped with a thin layer of cloth tape to provide more compliance when held with the laparoscopic grasper.

The sequence of target illumination was controlled by a custom program on the open-source Arduino platform [arduino.cc]. Figure 7.3 shows photographs of the pegboard showing possible illumination states.

The advantage of illumination using the additional hardware is two-fold: illumination

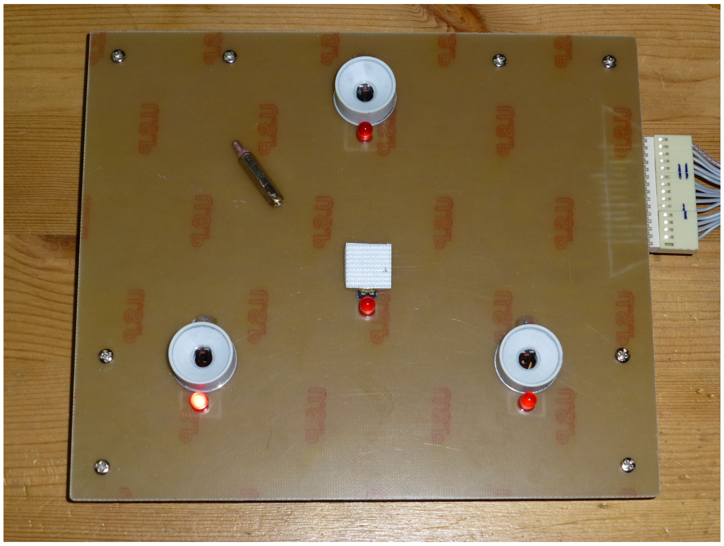


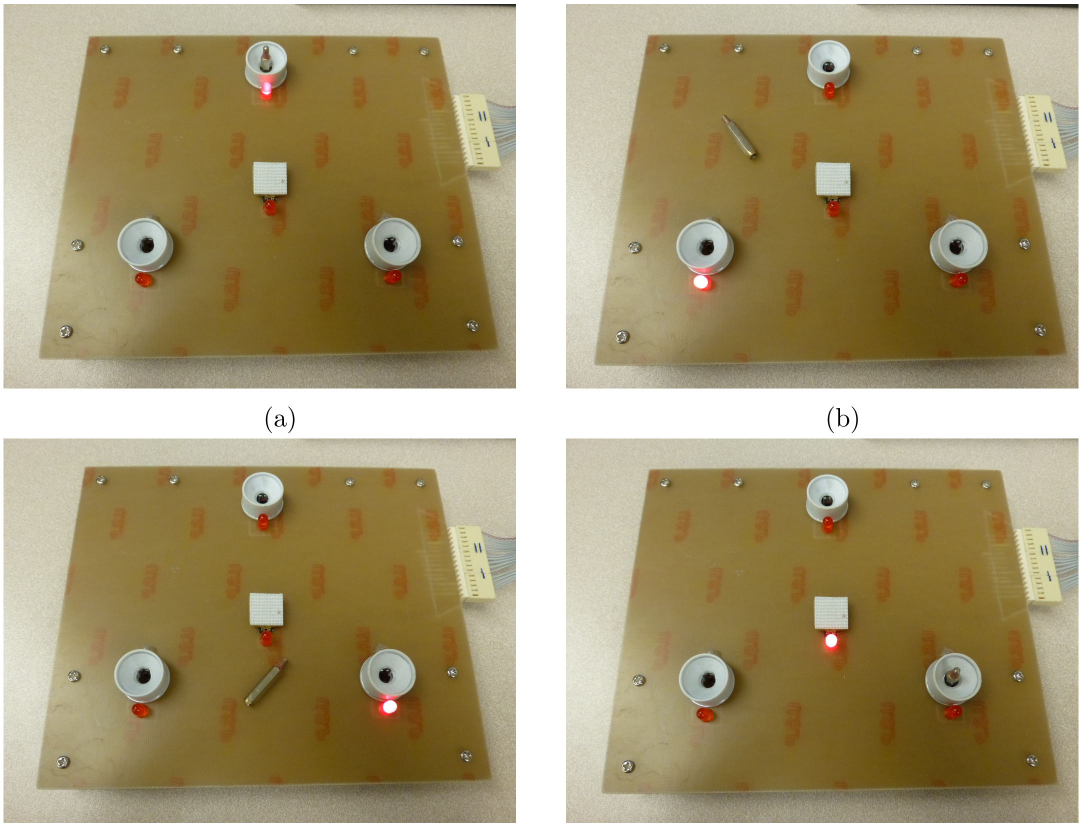
Figure 7.2: Photograph of the peg board and transferred standoff. The target at the lower left is currently illuminated.

state changes are automatic and instantaneous, transitioning as soon as a subtask is completed without any reaction time lag from an experimenter-controlled manual interface, and the subtask completion timestamps are electronically logged, eliminating the tedious step of manual subtask annotation from video done in the studies of Chapters 5 and 6.

Eye tracking duties were served by a Tobii X2-60-Wide remote eye tracker with a sampling rate of 60 Hz. The same laparoscopic training box in Chapters 5 and 6 was used, with its joystick camera removed and replaced with a high framerate web camera, providing a laparoscopic view in 640×480 resolution at 60 Hz. The data collection was managed using Tobii Studio 3.2.1 in 64-bit Windows 7. The training box contents were displayed on a Stryker Vision Elect HD 21” LCD at University of Alberta or a Dell Ultrasharp 2000FP 19” LCD at Simon Fraser University. Both displays are 4:3 aspect ratio with native resolution of 1600×1200. The experimental setup deployed at University of Alberta is shown in Figure 7.4. The setup at Simon Fraser University used the same training box, grasper, and camera arranged similarly to the Alberta setting.

7.5 Description of participants

For a between-subjects design, 8 untrained novices were recruited from the research laboratories at the University of Alberta Heritage Medical Research Centre, and one qualified surgical resident with 6 years of laparoscopic experience participated as an expert subject.



(a) (b)

(c) (d)

Figure 7.3: Illumination states of the pegboard. (a) Target dish at top illuminated, (b) target dish at lower left, (c) target dish at lower right, and (d) central target “Home” position.

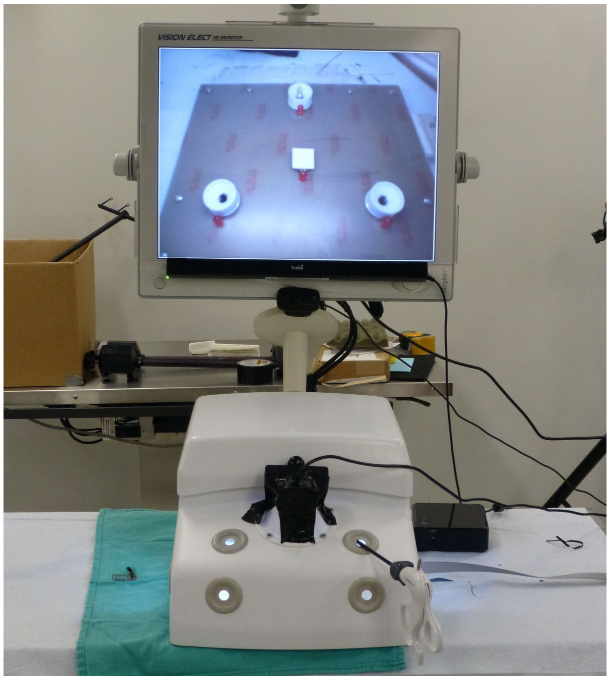


Figure 7.4: Gaze training apparatus with eye tracker attached to a position-adjustable display monitor, and peg board inside a laparoscopic training box.

An additional 9 novice participants were recruited from the graduate research laboratories at Simon Fraser University, with one subject's data omitted from analysis due to lack of focus during participation. The expert was a 29-year old male, and the novices had an average age of 29.5 years. All participants were right-handed.

7.6 Procedure

Subjects were first provided a written description of the experimental goals and then asked to give signed consent to participate if they had no complaints arising from the briefing. Each completed a short questionnaire to collect demographic data; expert-level subjects were given an additional survey to determine their surgical experience score [67]. Next, subjects were given a written/pictorial description of the experimental task and shown a demonstration by the experimenter. In a seated position, they were allowed to use the grasper to practice insertion and removal of the peg from each of the cups approximately five times; they were not allowed to repeatedly practice the motion of transferring the peg from one cup to another. Subjects were then registered into Tobii Studio and a 9-point eye tracker calibration was performed. Briefing materials are given in Appendix B.3.

After completing three untrained trials of the task as a baseline performance measure, novice subjects were randomly assigned to one of two training conditions: unlit targets with verbal gaze direction (verbal training, VT, similar to discovery learning DL as in Vine et al.'s study [60]), but with the important difference that these subjects were verbally instructed to gaze at the target rather than the tool), and illuminated targets without any explicit verbal direction (gaze training, GT). The verbal training subjects in the unlit condition were asked to focus their gaze on the destination of tool movement. Subjects in the illuminated condition were only informed that the targets would be lit and were not explicitly asked to adjust their gaze in any way.

Each subject performed six training blocks of 5 trials, with rest allowed between blocks. After completion of the sixth training block, subjects completed an electronic version of the NASA TLX, and subjects in the GT groups completed an additional short questionnaire (Appendix B.3) to study their perception of the illuminated training stimulus.

When subjects were ready to continue, they performed a block of 3 trials with no illumination and no additional verbal instructions, to record their retention of the learned eye-hand coordination.

7.7 Data analysis

Each trial was divided into 9 subtasks as detailed in [1]: reaching and grasping tasks (RG) involved moving the grasper from the central home position to grasp the peg resting in one of the target cups; transfer and release tasks (TR) were done by moving the peg from its original position to the next target cup; homing tasks (H) involve returning the empty grasper to the home position once the peg is inserted into its target cup - a single trial consists of three of each of these subtask types.

7.8 Results

The data collected from the U of A expert and eight SFU novices had a mean proportion of invalid/missing data at 0.09 (min < 0.01, max 0.26, i.e. the eye tracker was able to capture on average 91% of each trial). These data were pre-processed using a moving average filter and gap-filling interpolation as detailed in Appendix A. The mean duration of subtasks through progression of training is shown in Figure 7.5. The expert performed quickest at baseline without any practice. A table of independent-samples t-test differences in subtask performance between the two training groups is provided in Table 7.1, showing that the

	Subtask type		
VT vs GT	RG	TR	H
baseline	0.437	0.096	0.294
retention	0.324	0.461	0.060

Table 7.1: Two-tailed *t*-test *p*-values for mean completion time for VT and GT groups. $df = 35$ each for VT and GT (4 subjects × 3 of each subtask type × 3 trials).

	Subtask type		
VT vs GT	RG	TR	H
baseline	0.020	0.026	0.043
retention	< 0.001*	< 0.001*	< 0.001*

Table 7.2: Two-tailed *t*-test *p*-values for tool-target separation for VT and GT groups. Values of n are the same as in Table 7.1.

training groups did not show significantly different subtask completion times. A Bonferroni correction applied for a family-wise significance level of 0.05 with 12 tests yields individual test significance at p $\approx$ 0.004.

The tool-target separation measure is illustrated in Figure 7.6. Training condition t-test differences for this measure are given in Table 7.2. Significantly different table entries at the Bonferroni-corrected $p < 0.004$ level are marked with an asterisk. Note that GT and VT appear significantly different in the figure at baseline but did not achieve significance after applying the Bonferroni correction.

7.9 Discussion

7.9.1 Performance measures

Baseline performance

The VT group at baseline exhibited more target-oriented behaviour on the tool-target separation measure than the GT group, although this may be largely attributed to population sampling - it can be seen in Figure 7.6 that while not significantly different, the VT group already demonstrated a relatively higher measure during the baseline phase. This contrasts with pre-training eye parameter data from both Vine's study [60] and a recent study by Causer et al. [3] in "quiet eye" training in open surgery knot-tying. Both of these research groups achieved improvements in gaze-trained groups, with similar performance and eye

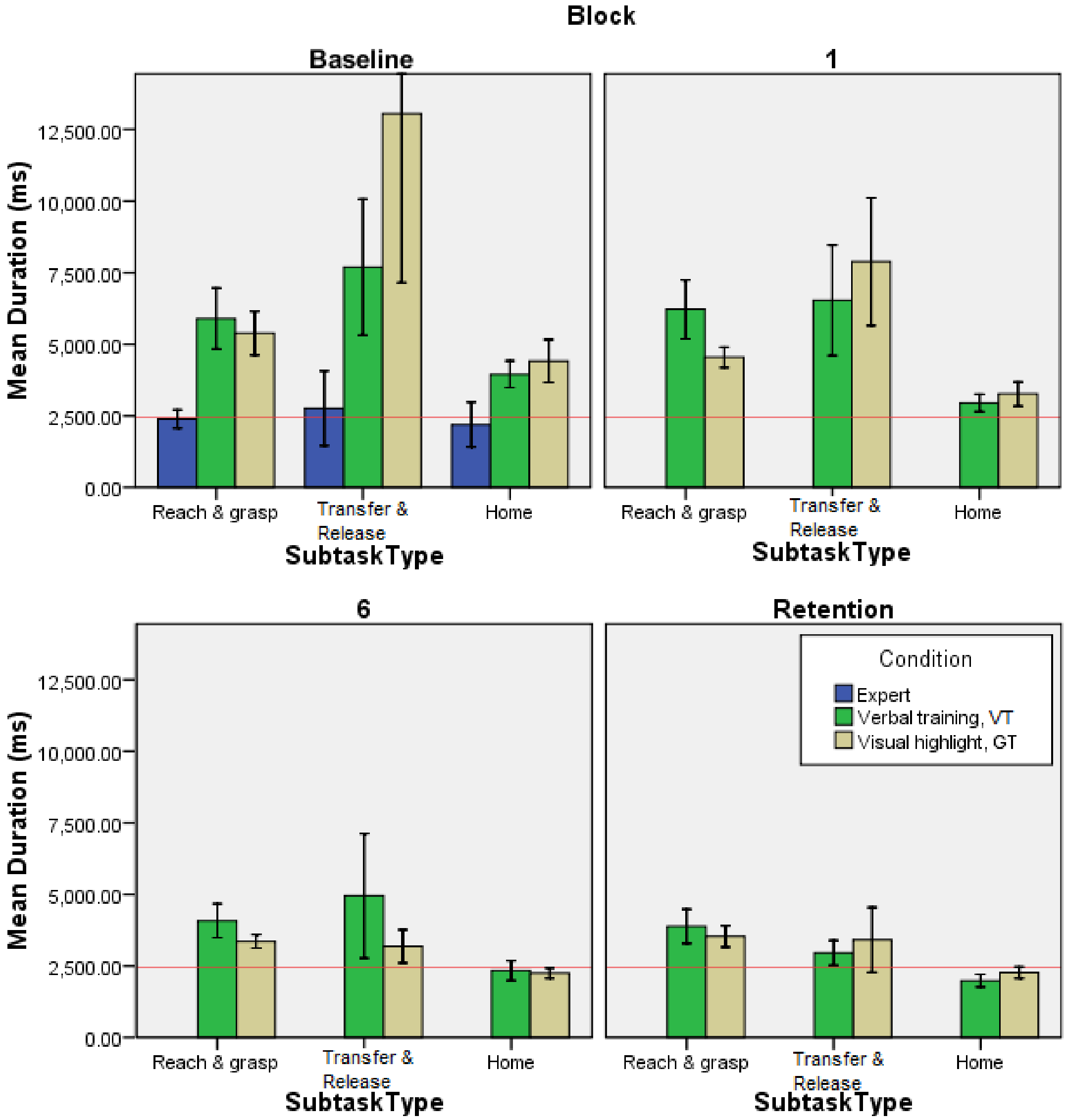


Figure 7.5: Mean subtask (± 1 S.E.) completion time by experimental block (blocks 25 omitted). The horizontal red line indicates the performance achieved by the expert surgeon. Shorter bars indicate better performance.

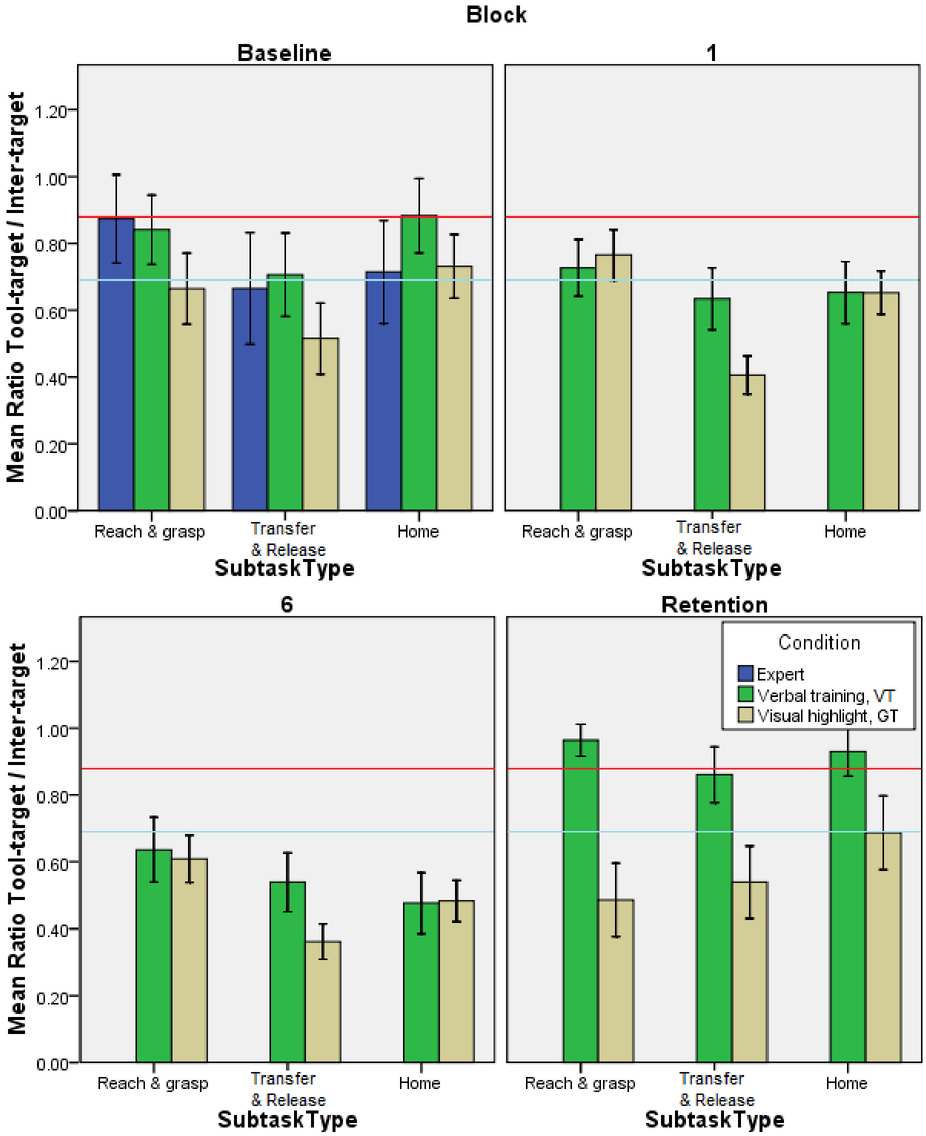


Figure 7.6: Mean tool-target separation (± 1 S.E.), normalized by the distance between the movement origin and destination. The red reference line shows expert behaviour at baseline for the RG subtask, and the turquoise line approximately marks expert behaviour at baseline for TR and H subtasks. A longer bar indicates more target-oriented gaze behavior.

parameters prior to training. In the case of open surgery, it may be that while a task-specific manual skill must be learned, it can be done without the additional layer of motor translation necessitated by laparoscopic operation.

Training and retention performance

Subjects in both training conditions approached expert-level task performance by the end of the training period, with no statistically significant performance difference between the two training groups (Figure 7.5 & Table 7.1). Thus hypothesis 1 is not supported. Surprisingly and encouragingly, the GT group did not exhibit any change in performance when target illumination was removed during the retention phase. It may be concluded that the task was repetitive enough that the task's motor requirements were predictable and already learned well enough, so that the added visual cues were no longer needed.

7.9.2 Gaze measures

Tool-target measure vs fixation measures

The use of this tool-target separation measure contrasts with the "target-locking" eye measure score used by Vine et al. [59] which counted the proportion of time spent in fixations over the required target to time spent fixating on the instrument. In Vine's study, use of this target-locking measure was made possible with a head-mounted eye tracker that could reliably maintain a steady view of the wearer's eye over the entire duration of the experimental trials. In our study, we were interested in more precisely observing the point of gaze within the surgical training scene, and this requirement for high spatial resolution (0.5° visual angle) was met with the remote Tobii X2-60-Wide eye tracker. However, with a remote eye tracker, subjects may move out of tracking range, resulting in intermittent data loss. Thus it was more appropriate to use instantaneous gaze measures such as the tool-target separation of described in Chapter 6 for each subtask rather than relying on the availability of valid fixations (which require at least 5 consecutive valid points of gaze for a 100 ms fixation with a 50 Hz eye tracker) over an entire trial to provide a measure such as a target-locking score.

Training and retention gaze measures

During the training blocks, all the subjects in both groups had a progressively lower tool target separation measure, while their performance steadily improved. This may arise due to developing a strategy of directing the tool to the target by using their peripheral vision, made easier because of the predictable nature of the target locations.

Neither group was given any specific instruction about the task before the retention trials, which were performed after a time gap during which the post-training questionnaire was completed. During the retention trials, the gaze behaviour of the two groups was significantly different (Table 7.2 & Figure 7.6). The VT group had a much higher tool-target separation than the GT group, although recall that even at baseline the VT group began with a consistently though not significantly higher tool-target separation for all subtask types. In spite of this gaze behaviour difference, both groups achieved comparable task completion times during the retention trials.

Although the VT group exhibited progressively lower tool-target separation by the end of training, they were able to change their behaviour to gaze at the target to direct the tool for the retention trials; recall that they had received the instruction to gaze at the target before their training blocks. The GT group did not change their eye behaviour on the target between the training trials and the retention block, even though the target illumination was disabled. Although we did not find any significant difference in task performance between the two training groups, the high degree of tool-target separation at retention for the VT group suggests that a target-oriented gaze behaviour is still learned eventually, and the verbal instruction may allow trainees to bypass the step of discovering this target-oriented behaviour which, while not correlated with completion time for this simple, repetitive task, may be a useful ability when performing more complex and dynamic task sequences. Nonetheless, hypothesis 2 is not supported by our findings.

The use of target illumination for gaze manipulation potentially allowed GT subjects to use peripheral vision to track the highly salient targets while maintaining foveal vision on the instrument. In contrast, the darkened mask used by Vine with more pronounced highlighting as well as partially obscured non-target areas instead encouraged foveation on target while using peripheral vision to track the instrument. Thus the two different highlighting methods could each give rise to very different gaze strategies by trainees where the roles of foveal and peripheral vision are reversed.

On the technical side, Vine et al. employed a head-mounted eye tracker, whereas a remote system was used here. One is stronger where the other is weak and vice-versa as discussed above, but as a result it is infeasible to collect the same gaze parameter measurements for direct comparison. Although it was necessary to use a different eye measure with our setup, the progress of the subjects' task performance could still be monitored, in spite of the task itself being quite different, requiring considerably high precision to insert the peg into the target cups at the correct angle.

7.9.3 Mechanism of gaze training

In spite of the overall similarity in this peg transfer task and the ball pick-up-and-drop task used by Vine et al. [60, 59], we were not able to demonstrate a correlation between our tool-target separation measure and task performance to link gaze training with skills training. In light of this result, we reflected on the design of our experimental task and the non-electronic peg transfer of Chapter 6.

As the tool-target separation by its definition occurs before the instrument reaches its destination, it seemed appropriate to investigate how long the approach towards the target could take, as a proportion of each overall movement time. For this we chose to divide each movement subtask into two phases following the examples in the classic studies of goal-directed aiming by Woodworth [66, 9].

We first revisited the results of Chapter 6. Each RG, TR, and H subtask was divided into two phases: Transport, describing the initial phase of movement from when the tool leaves its starting location to when it reaches the vicinity above its target destination, and Landing, beginning at the end of the Transport phase and ending when the action required at the end of the subtask has been completed. The beginning and end of the transport phase was determined automatically using fixed velocity thresholds based on the tool position obtained from the recorded task videos [22], and the timing of the landing phase was determined by manual annotation of tool contact with the target in the same task videos. Figure 7.7 shows the proportion of subtask time the transport and landing phases comprise for each subtask type in the peg transfer data recorded in Chapter 6.

Although the manual annotation of movement phase division is not available for the data from this chapter's gaze training experiment, Figure 7.8 shows the overall subtask completion time as determined from the electronic timing of the apparatus. Table 7.3 presents a comparison of the subtask completion times of the original peg transfer from Chapter 6 and this gaze training data set. Generally the subtask durations were slightly lower in this gaze training study, thus the potential for further reductions as a result of optimizing gaze patterns may also have diminished.

The transport step involves the transit of the grasper from its originating location to a reasonable proximity of the target location, while the landing step involves the precise placement of the transported object into the target receptable. The physical distance to

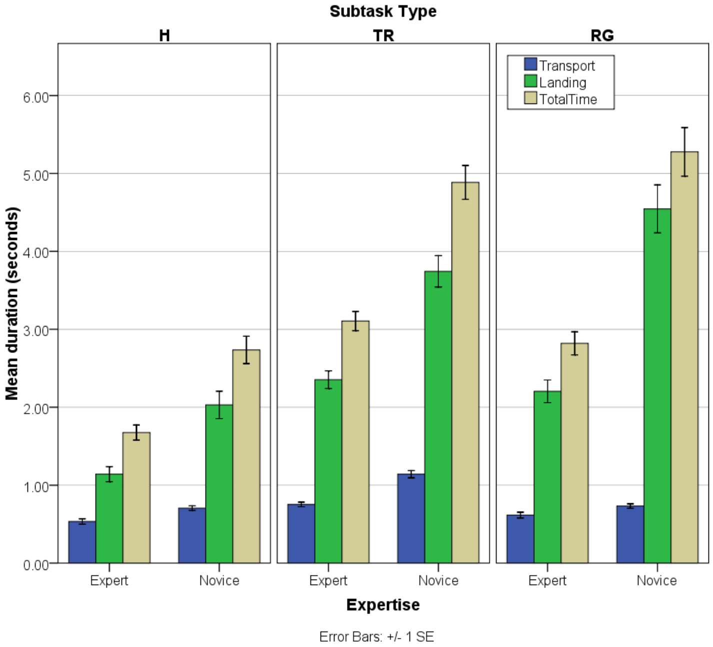


Figure 7.7: Transport and landing phase duration by subtask, from the data set of Chapter 6.

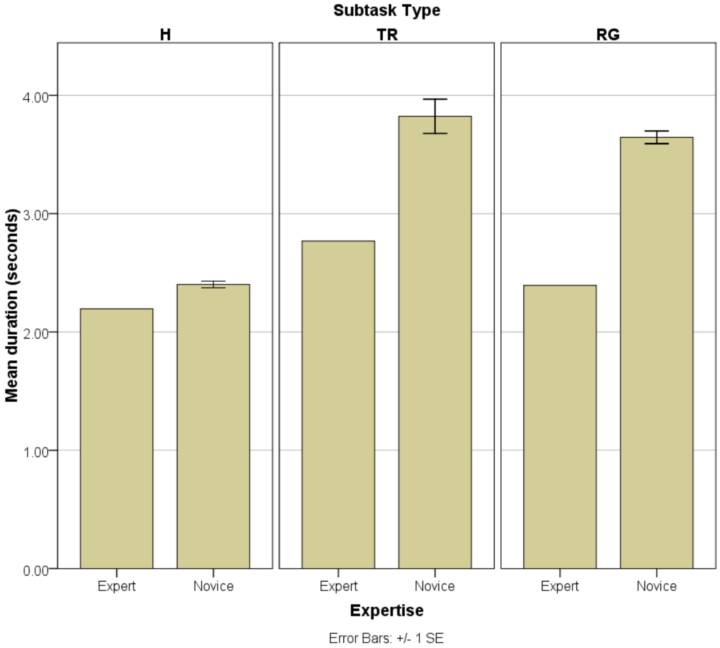


Figure 7.8: Overall subtask completion time by type, Chapter 7 data set.

		Subtask duration (s)		
	Group	H	TR	RG
Chapter 6	Expert	1.68	3.11	2.82
	Novice	2.74	4.89	5.28
Chapter 7	Expert	2.20	2.77	2.39
	Novice	2.40	3.82	3.65

Table 7.3: Mean subtask completion time comparison, Chapter 6 and Chapter 7

traverse in the transport phase of this chapter's task is roughly the same as that in the previous chapter, and the modified peg transfer apparatus of this chapter required high precision in landing to activate the electronic switches. Thus we may apply the movement phase model indicated by Figure 7.7 to show that the transport phase is only a minor portion of the overall subtask.

Since the gaze training proposed in this chapter aimed to make performance improvements during the relatively brief transport phase, any potential gains made here were likely overshadowed by the relatively lengthy process of inserting the peg into the centre of the receptacle. A better gaze training strategy could thus target other phases of manual actions where more significant improvements could be observed.

7.9.4 Threats to validity

The arrangement of the laparoscopic task and the nature of the visual highlighting leads to potential issues with transferring learned strategies to a more realistic laparoscopic environment. In contrast to the procedures in Chapters 5 and 6, subjects performed many repetitions of the laparoscopic peg transport sequence. However, due to the predictable nature of the target locations, subjects possibly developed some motor learning specifically for transports to and from these target locations, minimizing the role of initial visual guidance. Exacerbating this was the discovery that the saliency of the visual target highlighting allowed subjects to track the targets in their peripheral vision, which is not a situation which could realistically occur in a live surgical setting Additionally, subjects in the GT group were given an additional survey between the training and retention blocks which could have extended their resting period, whereas VT subjects did not take the survey and could begin their retention block quickly.

7.10 Research contributions

In this chapter we applied two methods of gaze training in order to manipulate the gaze patterns of novice operators during a simple laparoscopic peg transfer task. In the particular task scenario of this chapter's experiment, we suggest a simple verbal instruction can be a feasible and effective method to influence gaze behaviors of surgical trainees. However, larger performance gains may potentially be achieved by gaze training for manual actions where initial ballistic motions comprise a larger proportion of the overall task time.

Chapter 8

Conclusions

The central theme of this thesis is identification of consistent gaze characteristics separating levels of expertise in laparoscopic tasks. Once such characteristics are identified, this knowledge can potentially be utilized to inform methods of gaze training in an effort to expedite the acquisition of manual skills.

In the vigilance study of Chapter 3, gaze was recorded with a wide field of view across several objects in the laparoscopic operator's surroundings. Because the simulator used was not entirely realistic, for example the gall bladder did not bleed or perforate when handled roughly, the novices were faster than the more careful experts who applied their clinical experience to the simulated operation. This result highlights the need for more realistic simulators.

We learned that experts reported a lower workload than novices during the simulated surgery. Therefore experts have spare mental resources during surgery, allowing them to glance at the patient vital signs more frequently than the novices. As the behaviour of visually tending to patient vital signs on a remote digital display seemed to be more a function of experienced operators requiring fewer mental resources to the primary operating task and thus having more resources available to observe the patient vitals, this gaze characteristic is unsuitable for gaze training. For novices struggling to complete the primary operating task with intense focus, it may be counter-productive to distract them with observing a vitals display which in a live operation would be managed by an anaesthesiologist.

In Chapter 4 we successfully collected eye tracking data from the primary surgeon in live laparoscopic cholecystectomy cases, and from these data presented a method for evaluating overlap and mismatches in gaze between the operator and observers. Using these methods we found that observers matched their gaze patterns with the operator for only between

50% to 60% of the procedure's duration. Additionally, this method allows us to identify prolonged periods of gaze mismatch when the surgeon's gaze is clearly different from an observer's gaze. Recognizing the semantic events underlying such periods of gaze mismatch may have educational value in training novices to gaze at the proper surgical locations during critical moments of an operation.

Towards generating expert gaze overlaid on surgical videos as a training tool, ideally gaze should be recorded during the actual operation, a result supported by the findings of Chapter 5 which showed that in a task with predictable targets, gaze overlap between doing and watching could reach up to 80%, although a lag from the observers is still in evidence.

Chapter 6 showed that a novel gaze parameter could be used to separate experts and novices. A new instantaneous eye-hand coordination measure was defined and obtained by eye tracking expert surgeons and non-surgeons in a simulated laparoscopic task. By measuring the distance between an intended target and the current tool position, it was found that experienced surgeons looked at the target before the tool approached while non-surgeons focused their attention elsewhere until the tool had moved closer. This result promises exciting new approaches to accelerate the development of eyehand coordination in laparoscopic surgical trainees.

In Chapter 7 we applied the findings of previous chapters and designed an experiment to test the validity of the tool-target separation parameter as a target for gaze training. By administering two methods of gaze manipulation - verbal directions ("Look at the target"), and a dynamic visual illumination of the next target, we aimed to alter the gaze behaviour of novice operators in an effort to improve task performance. These methods of gaze manipulation indeed produced differing gaze behaviours after training; however there was no correlated benefit to task performance as a result of gaze training.

On observation of the task videos from the gaze training experiment, we noticed that the task time was mostly spent in landing the peg into the target at the end of the transport. We performed an analysis of the relative times spent on transporting and landing in the task of Chapter 6, and learned that only up to one-third of the task time was spent in moving the tool to the next target, where gaze training could be effective. The chosen training parameter presented little opportunity for improvement simply due to its minor proportion of the overall task time, i.e. motion of laparoscopic instruments over relatively large distances. Gaze training may perhaps be more effective if more general parameters are targeted which occur during other more significant portions of laparoscopic tasks.

The verbal training method employed had no negative effect on performance, but is still contingent on a trainee actively remembering the verbal instruction. During initial familiarization with laparoscopic manipulation, the trainee still may not have sufficient resources to remember such instruction and must rely on visual tool-tracking to learn proper motor input mapping. Allowing novice learners to tool-track as they naturally would, may still be a positive feature in gaze training techniques.

The goal of identifying a gaze behaviour which could be trained to achieve significant improvement in task performance in laparoscopic surgery was not fully realized because of several challenges, one of which is the difficulty of recording high-quality data in general eye tracking experiments. Collecting eye tracking data consistently from experimental subjects is necessary to better understand eye-hand coordination as a step towards improving performance. This is difficult when subjects are performing physical tasks using remote eye trackers, because body movements may likely lead to missing or imprecise eye tracking data.

8.1 Future work

While the performance benefits from the gaze training experiment were not significant, we believe the study may be pursued further, targeting manual actions where an initial movement phase and its associated gaze patterns account for a higher proportion of the overall task time, either by increasing transport distances or by requiring some precision and caution during the transportation.

In order to increase the likelihood of obtaining fruitful results, the quality of collected data should be ensured. Using higher-resolution or wider-angle remote systems, or high-resolution head-mounted eye tracking systems can reduce the amount of data loss from each eye tracking recording.

Bibliography

[1] M. S. Atkins, X. Jiang, G. Tien, and B. Zheng. Saccadic delays on targets while watching videos. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, pages 405–408, 2012.

[2] M. Stella Atkins, Geoffrey Tien, Rana S. A. Khan, Adam Meneghetti, and Bin Zheng. What do surgeons see: Capturing and synchronizing eye gaze for surgery applications. *Surgical Innovation*, 20(3):241–248, June 2013.

[3] Joe Causer, Adrian Harvey, Ryan Snelgrove, Gina Arsenault, and Joan N Vickers. Quiet eye training improves surgical knot tying more than traditional technical training: a randomized controlled study. *The American Journal of Surgery*, 208(2):171–177, August 2014.

[4] Peter R. Chapman and Geoffrey Underwood. Visual search of driving situations: Danger and experience. *Perception*, 27:951–964, 1998.

[5] A. S. A. Chetwood, K. W. Kwok, L. W. Sun, G. P. Mylonas, J. Clark, A. Darzi, and G. Z. Yang. Collaborative eye tracking: a potential training tool in laparoscopic surgery. *Surgical Endoscopy*, 26(7):2003–2009, July 2012.

[6] F. Deconinck, V. van Polanen, G. Savelsbergh, and S. Bennett. The relative timing between eye and hand in rapid sequential pointing is affected by time pressure, but not by advance knowledge. *Experimental Brain Research*, 213:99–109, 2011.

[7] Damion C. Dishart and Michael F. Land. The development of eye movement strategies of learner drivers. In Geoffrey Underwood, editor, *Eye Guidance in Reading and Scene Perception*, chapter 19, pages 419–429. Elsevier Science, 1998.

[8] Andrew T. Duchowski. *Eye Tracking Methodology: Theory and Practice.* Springer-Verlag London Limited, 2nd edition, 2007.

[9] Digby Elliott, Steve Hansen, Lawrence E. M. Grierson, James Lyons, Simon J. Bennett, and Spencer J. Hayes. Goal-directed aiming: Two components but multiple processes. *Psychological Bulletin*, 136(6):1023–1044, November 2010.

[10] J. Randall Flanagan and Roland S. Johansson. Action plans used in action observation. *Nature*, 424(6950):769–771, August 2003.

[11] S. Furneaux and M. F. Land. The effects of skill on the eye-hand span during musical sight-reading. *Proceedings: Biological Sciences*, 266(1436):2435–2440, December 1999.

[12] Bin Zheng Geoffrey Tien, M. Stella Atkins. Measuring gaze overlap on videos between multiple observers. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 309–312, 2012.

[13] Benno Gesierich, Angela Bruzzo, Giovanni Ottoboni, and Livio Finos. Human gaze behavior during action execution and observation. *Acta Psychologica*, 128(2):324–330, June 2008.

[14] J. Guo and G. Feng. How eye gaze feedback changes parent-child joint attention in shared storybook reading: an eye-tracking intervention study. In *2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction*, 2011.

[15] Riad I. Hammoud and Jeffrey B. Mulligan. *Introduction to Eye Monitoring*. Springer-Verlag: Berlin, 2008.

[16] Mary M. Hayhoe, Anurag Shrivastava, Ryan Mruczek, and Jeff B. Pelz. Visual memory and motor planning in a natural task. *Journal of Vision*, 3(1):49–63, February 2003.

[17] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. van de Weijer. *Eye tracking: a comprehensive guide to methods and measures*. Oxford University Press, 2011.

[18] K. Holmqvist, M. Nyström, and F. Mulvey. Eye tracker data quality: what it is and how to measure it. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, pages 45–52, 2012.

[19] H. Jarodzka, K. Scheiter, P. Gerjets, and T. van Gog. In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, 20(2):146–154, April 2010.

[20] Xianta Jiang, M. Stella Atkins, Geoffrey Tien, Roman Bednarik, and Bin Zheng. Pupil responses during discrete goal-directed movements. In *Proceedings of CHI*, pages 2075–2084, 2014.

[21] Xianta Jiang, M. Stella Atkins, Geoffrey Tien, Bin Zheng, and Roman Bednarik. Pupil dilations during target pointing respect Fitts' Law. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, pages 175–182, 2014.

[22] Xianta Jiang, Bin Zheng, and M. Stella Atkins. Video processing to locate the tooltip position in surgical eye-hand coordination tasks. *Surgical Innovation (online)*, July 2014.

[23] Roland S. Johansson, Göran Westling, Anders Bäckström, and J. Randall Flanagan. Eye-hand coordination in object manipulation. *The Journal of Neuroscience*, 21(17):6917–6932, September 2001.

[24] P. Kasarskis, J. Stehwien, J. Hickox, and C. D. Wickens. Comparison of expert and novice scan behaviors during vfr flight. In *Proceedings of the 11th International Symposium on Aviation Psychology*, 2001.

[25] Rana S. A. Khan, Geoffrey Tien, M. Stella Atkins, Bin Zheng, Ormond N. M. Panton, and Adam T. Meneghetti. Analysis of eye gaze: Do novice surgeons look at the same location as expert surgeons during a laparoscopic operation? *Surgical Endoscopy*, 26(12):3536–3540, December 2012.

[26] Ergun Kocak, Jan Ober, Necip Berme, and Scott Melvin. Eye motion parameters correlate with level of experience in video-assisted surgery: Objective testing of three tasks. *Journal of Laparoendoscopic & Advanced Surgical Techniques, Part A*, 15(6):575–580, December 2005.

[27] Elizabeth A. Krupinski, Allison A. Tillack, Lynne Richter, Jeffrey T. Henderson, Achyut K. Bhattacharyya, Katherine M. Scott, Anna R. Graham, Michael R. Descour, John R. Davis, and Ronal S. Weinstein. Eye-movement study and human performance using telepathology virtual slides: Implications for medical education and differences with experience. *Human Pathology*, 37(12):1543–1556, December 2006.

[28] Michael F. Land and Peter McLeod. From eye movements to actions: how batsmen hit the ball. *Nature Neuroscience*, 3(12):1340–1345, December 2000.

[29] Benjamin Law, M. Stella Atkins, A. E. Kirkpatrick, and Alan J. Lomax. Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment. In *Proceedings of the 2004 symposium on Eye tracking research & applications*, ETRA '04, pages 41–48, 2004.

[30] R. John Leigh and David S. Zee. *The Neurology of Eye Movements*. Oxford University Press, 4th edition, 2006.

[31] Robert G. Loeb. Monitor surveillance and vigilance of anesthesia residents. *Anesthesiology*, 80(3):527–533, March 1994.

[32] Amir Mané and Emanuel Donchin. The space fortress game. *Acta Psychologica*, 71(1–3):17–22, August 1989.

[33] Sandra P. Marshall. The index of cognitive activity: measuring cognitive workload. In *Proceedings of the IEEE 7th Conference on Human Factors and Power Plants*, volume 17, pages 75–79, December 2002.

[34] R. S. W. Masters, C. Y. Lo, J. P. Maxwell, and N. G. Patil. Implicit motor learning in surgery: implications for multi-tasking. *Surgery*, 143(1):140–145, January 2008.

[35] NASA. NASA TLX: Task Load Index. http://humansystems.arc.nasa.gov/groups/tlx/.

[36] S. F. W. Neggers and H. Bekkering. Ocular gaze is anchored to the target of an ongoing pointing movement. *Journal of Neurophysiology*, 83(2):639–651, February 2000.

[37] Calvin F. Nodine and Harold L. Kundel. Using eye movements to study visual search and to improve tumor detection. *Radiographics*, 7(6):1241–1250, November 1987.

[38] Margret Oddsdottir. Avoidance of complications in laparoscopic cholecystectomy. In Ronald C. Merrell, editor, *Laparoscopic Surgery: A Colloquium.* Springer-Verlag, New York, 1999.

[39] Anneli Olsen. The Tobii I-VT fixation filter: algorithm description, March 2012. http://www.tobii.com/en/eye-tracking-research/global/library/white-papers/the-tobii-i-vt-fixation-filter/.

[40] Jorge Otero-Millan, Xoana G. Troncoso, Stephen L. Macknik, Ignacio Serrano-Pedraza, and Susana Martinez-Conde. Saccades and microsaccades during visual fixation, exploration, and search: Foundations for a common saccadic generator. *Journal of Vision*, 8(14):1–18, December 2008.

[41] W. Leigh Ottati, Joseph C. Hickox, and Jeff Richter. Eye scan patterns of experienced and novice pilots during visual flight rules (VFR) navigation. In *Proceedings of the Human Factors and Ergonimics Society Annual Meeting*, volume 43, pages 66–70, September 1999.

[42] Lee Richstone, Michael J. Schwartz, Casey Seideman, Jeffrey Cadeddu, Sandra Marshall, and Louis R. Kavoussi. Eye metrics as an objective assessment of surgical skill. *Annals of Surgery*, 252(1):177–182, July 2010.

[43] James C. Rosser, Jr. Operative strategy for acute cholecystitis and other difficult situations. In Ronald C. Merrell, editor, *Laparoscopic Surgery: A Colloquium.* Springer-Verlag, New York, 1999.

[44] Sajay Sadasivan, Joel S. Greenstein, Anand K. Gramopadhye, and Andrew T. Duchowski. Use of eye movements as feedforward training for a synthetic aircraft inspection task. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '05, pages 141–149, 2005.

[45] Uta Sailer, J. Randall Flanagan, and Roland S. Johansson. Eye-hand coordination during learning of a novel visuomotor task. *Journal of Neuroscience*, 25(39):8833–8842, September 2005.

[46] Dario D. Salvucci and Joseph H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, pages 71–78, 2000.

[47] Angela T. Schriver, Daniel G. Morrow, Christopher D. Wickens, and Donald A. Talleur. Expertise differences in attentional strategies related to pilot decision making. *Human Factors*, 50(6):864–878, December 2008.

[48] F. Jacob Seagull, Yan Xiao, Colin F. MacKenzie, Mahmood Jaberi, and Richard P. Dutton. Monitoring behavior: a pilot study using and ambulatory eye-tracker in surgical operating rooms. In *Proceedings of Human Factors and Ergonomics 43rd Annual Meeting*, pages 850–854, 1999.

[49] Noa Segall, Jeffrey M. Taekman, Jonathan B. Mark, Gene Hobbs, and Melanie C. Wright. Coding and visualizing eye tracking data in simulated anesthesia care. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 51, pages 765–770, October 2007.

[50] Kimron L. Shapiro and Jane E. Raymond. Training of efficient oculomotor strategies enhances skill acquisition. *Acta Psychologica*, 71(1–3):217–242, August 1989.

[51] Randy Stein and Susan E. Brennan. Another person's eye gaze as a cue in solving programming problems. In *Proceedings of the 6th International Conference on Multimodal Interfaces*, ICMI '04, pages 9–15, 2004.

[52] Thomson-Reuters. In-patient and out-patient database. 2009.

[53] Geoffrey Tien, M. Stella Atkins, Xianta Jiang, Rana S. A. Khan, and Bin Zheng. Identifying eye gaze mismatch during laparoscopic surgery. In *Studies in Health Technologies and Informatics*, pages 453–457, 2013.

[54] Geoffrey Tien, M. Stella Atkins, Xianta Jiang, Bin Zheng, and Roman Bednarik. Verbal gaze instruction matches visual gaze guidance in laparoscopic skills training. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, pages 331–334, 2014.

[55] Geoffrey Tien, M. Stella Atkins, Bin Zheng, and Colin Swindells. Measuring situation awareness of surgeons in laparoscopic training. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ETRA '10, pages 149–152, New York, NY, USA, 2010. ACM.

[56] Geoffrey Tien, Bin Zheng, and M. Stella Atkins. Quantifying surgeons' vigilance during laparoscopic operations using eyegaze tracking. In *Studies in Health Technologies and Informatics 2011*, MMVR 18, pages 658–662, 2011.

[57] Joan N. Vickers. Gaze control in putting. *Perception*, 21(1):117–132, 1992.

[58] Joan N. Vickers. Control of visual attention during the basketball free throw. *American Journal of Sports Medicine*, 24(6):93–97, 1996.

[59] Samuel J. Vine, Richard J. Chaytor, John S. McGrath, Rich S. W. Masters, and Mark R. Wilson. Gaze training improves the retention and transfer of laparoscopic technical skills in novices. *Surgical Endoscopy*, 27(9):3205–3213, September 2013.

[60] Samuel J. Vine, Rich S. W. Masters, John S. McGrath, Elizabeth Bright, and Mark R. Wilson. Cheating experience: Guiding novices to adopt the gaze strategies of experts expedites the learning of technical laparoscopic skills. *Surgery*, 152(1):32–40, July 2012.

[61] Samuel J. Vine and Mark R. Wilson. The influece of quiet eye training and pressure on attention and visuo-motor control. *Acta Psychologica*, 136(3):340–346, March 2011.

[62] Sarah A. Vitak, John E. Ingram, Andrew T. Duchowski, Steven Ellis, and Anand K. Gramopadhye. Gaze-augmented think-aloud as an aid to learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 2991–3000, 2012.

[63] Mark Wilson, John McGrath, Samuel Vine, James Brewer, David Defriend, and Richard Masters. Psychomotor control in a virtual laparoscopic surgery training environment: gaze control parameters differentiate novices from experts. *Surgical Endoscopy*, 24(10):2458–2464, 2010.

[64] Mark Wilson, Samuel Vine, Elizabeth Bright, Rich Masters, David Defriend, and John McGrath. Gaze training enhances laparoscopic technical skill acquisition and multitasking performance: a randomized, controlled study. *Surgical Endoscopy*, pages 1–9, 2011.

[65] Mark R. Wilson, M. Coleman, and J. S. McGrath. Developing basic hand-eye coordination skills for laparoscopic surgery using gaze training. *British Journal of Urology*, 105(10):1356–1358, March 2010.

[66] R. S. Woodworth. The accuracy of voluntary movement. *The Psychological Review: Monograph Supplements*, 3(3):1–114, July 1899.

[67] Bin Zheng, Maria A. Cassera, Danny V. Martinec, Georg O. Spaun, and Lee L. Swanström. Measuring mental workload during the performance of advanced laparoscopic tasks. *Surgical Endoscopy*, 24(1):45–50, January 2010.

[68] Bin Zheng, Xianta Jiang, Geoffrey Tien, Adam Meneghetti, O. Neely M. Panton, and M. Stella Atkins. Workload assessment of surgeons: correlation between NASA TLX and blinks. *Surgical Endoscopy*, 26(10):2746–2750, 2012.

[69] Bin Zheng, Geoffrey Tien, M. Stella Atkins, Colin Swindells, Homa Tanin, Adam Meneghetti, Karim A. Qayumi, and O. Neely M. Panton. Surgeon's vigilance in the operating room. *The American Journal of Surgery*, 201(5):673–677, 2011.

Appendix A

Data pre-processing for eye tracking experiments

A.1 Introduction

Initial analysis of the data presented in Chapter 6 revealed unusually poor-quality data recorded from certain experimental participants. This led to investigation into the nature of the poor data and subsequently to the techniques presented in this appendix to alleviate the symptoms of poor-quality data. Two sources of inadequate quality data which may be correctable were identified in our recorded data: jitter and missing data. It is possible to identify "high-amplitude jitter" illustrated in Figure A.1a by observing scanpath replays where the magnitude of the deviation of the recorded gaze coordinate from the true point of fixation is much larger than the natural physiological jitter of the eye (Figure A.1b), and persists for the duration of the recording. Missing data occurs when the eye tracker is unable to detect both eyes. This can arise when the subject blinks, or moves his head outside of the eye tracker's range. Hovering near the tracker's range boundary leads to the tracker output flickering rapidly between eyes detected or not detected.

Furthermore, under certain unknown circumstances, eye tracking hardware can briefly produce an incorrect point of gaze. This manifests as a characteristic large, isolated "jump" in the reported point of gaze quickly followed by another "jump" back to the true gaze location, shown in Figure A.1c. This is different from jitter and the square-wave jerk of the eye gaze, described by Leigh & Zee [30] which is an actual physiological phenomenon. In the case of erroneous recording by the eye tracker, the magnitude of the jump often exceeds the maximum physiologically possible angular acceleration of the eyeball. The jump point

(a) (b) (c)

Figure A.1: Simulated examples of (a) high amplitude jitter from eye tracker, (b) natural jitter, and (c) a jump caused by the eye tracker with the same sampled gaze point sequence for all examples. The triangle indicates the true point of fixation.

of gaze also tends not to remain at the incorrect location long enough to register a fixation. As this is difficult to predict, it will be treated in the same way as jitter.

When such tracking errors occur frequently enough to invalidate a significant portion of an experimental dataset which would be costly to replace, it is advantageous to apply some data processing in an attempt to recover recordings which would otherwise be unusable. For jitter correction, a smoothing filter is applied to the raw data in order to bring the recorded gaze points nearer to the actual point of gaze so that standard fixation detection algorithms can operate correctly. Short intervals of missing data are filled by interpolating between the good data points before and after the missing data. Using these techniques targeted to the two specific patterns of data loss, well over half of previously unusable trials have been recovered in our real-world dataset. This filtering process can be done as a first step in any eye movement experiment, and has no ill effect on already high-quality data.

The methods below were developed for dealing with these problems, using data from Tobii 50-series eye trackers coupled with Tobii Clearview 2.7.X, which are still in wide use by other research groups [5, 13, 14, 19]; furthermore these methods are easily extended to other eye trackers sampling at different spatial and temporal resolutions. Similar methods have now been implemented in later Tobii software [39]. The methods described below were developed independently but are documented in a whitepaper report from Tobii [39] and first implemented in Tobii Studio 2.3 [39].

A.2 Materials and Method

A.2.1 Materials

The apparatus and data used for this study were collected as part of the study described in Chapter 6. Each of the 12 subjects performed the required unimanual laparoscopic peg transfer task for 5 trials, yielding 60 "execution" gaze recordings. Additionally, each subject was later eye tracked while watching his own 5 trials plus the trials of two other participants, yielding 180 "observation" recordings, for a total of 240 trials. Trials typically lasted between 20-120 seconds depending on the expertise of the subject. The data were found to be of insufficient quality for further analysis but were observed to exhibit the jitter and data loss characteristics described above.

A.2.2 TF/TT for data quality

A useful measure of eye-gaze data quality is the ratio (Total fixation time / Total task time), TF/TT. Otero-Millan et al. [40] empirically observed fixations to compose over 80% of one's time viewing various static stimuli. To account for motion in dynamic stimuli as well as allow for some eye tracking data loss or noise, the minimum TF/TT ratio used for analysis can be relaxed to at least 70% for visual tasks requiring looking at a display monitor; i.e. saccades and data loss should only occupy at most 30% of the elapsed time of a task. To justify this choice of threshold, consider a visual stimulus repeatedly causing fixations to last only 100 ms, followed by a long saccade lasting 80 ms. With perfect data recording, the TF/TT ratio for this stimulus would be roughly 55%. This is a greatly simplified and implausible scenario; more realistic stimuli should cause longer fixations and thus higher TF/TT. With a generous mean saccade duration of 80 ms, a mean fixation duration of only 187 ms is required to achieve a TF/TT of 70%, which should not be unreasonable to observe in typical visual stimuli.

Based on the criteria TF/TT and on observation of gaze replays in Tobii Clearview of a pool of recorded trials collected from the experiment of Chapter 6, each recorded trial was categorized as "good", "jittery/gapped", and "bad". In general, recordings with TF/TT above 70% were considered good with no need to further process the data. The bad trials resulted from excessive subject movement or circumstances otherwise leading to extended periods of tracking loss during significant portions of the task trial, with little hope of recovery. Therefore the focus here is on processing of the jittery or gapped trials, where the data have the greatest potential of being recovered.

Total	Good	Jitter	Gapped	Bad
240	164 (68.3%)	32 (13.3%)	32 (13.3%)	12 (5.0%)

Table A.1: Eye tracking quality assessment for 60 "execution" and 180 "observation" recordings (240 recordings in total).

Table A.1 shows the proportion of trials in each category of data quality. Categorization of trials with TF/TT below 70% as either "jittery" or "gapped" was decided post-hoc based on the response to application of the data recovery filters to be described. Typically trials with TF/TT increasing by greater than 20%, but with little reduction in the amount of missing data, were categorized as jitter cases. Gap cases had their missing data reduced by at least one half, but began with a more significant portion of the task time missing (5% up to 40%). Bad trials had a similarly large portion of the task missing both before and after treatment.

A.2.3 Implementation of I-DT (dispersion threshold) fixation detection

The fixation detection algorithm provided in Tobii Clearview 2.7.X is based on the I-DT dispersion threshold algorithm described by Salvucci et al. [46]. Eye tracking data recorded using Clearview are saved in an encoded format but can be exported to a plain-text format for external processing. Since it is not possible to perform any post-processing on the original encoded Clearview data, and therefore it is not possible to use Clearview to detect fixations in the modified data, the I-DT implementation is first duplicated as a substitute for fixation detection on modified raw data points.

The I-DT algorithm accepts two parameters: minimum fixation duration and maximum dispersion radius. Sequential raw gaze data points are added to a potential fixation one at a time. After adding one gaze point, the centroid of the gaze point cluster is recalculated. This is repeated until either:

- An invalid gaze sample is encountered, or
- The maximum Euclidean distance between any gaze points in the cluster has exceeded the specified dispersion threshold.

If the gaze point cluster satisfies the minimum specified duration, it is reported as a fixation, otherwise a new attempt to find a fixation begins at the next valid gaze data sample.

Although the I-DT algorithm is well-documented and understood, it is worth emphasizing the fixation candidate terminating conditions above, because the presence of jitter and missing data can severely impact the performance of I-DT - hence the need to smooth out jitter and fill in gaps prior to using I-DT.

Furthermore, different implementations of the I-DT algorithm may terminate a fixation using data from one eye, or from the average gaze position of both eyes.

A.2.4 Moving average filter

The filter uses a single parameter which is the number of samples to include in the moving average window. As an example of applying a window size of 4 samples (80 ms at 50 Hz) which ignores invalid data, for a gaze sequence $G_1, G_2, G_3, G_4, G_5, X_1, X_2, G_6, G_7, G_8, G_9$ where G_i denotes a valid gaze sample and X_j is invalid, the filtered sequence becomes $G_1, G_{12}, G_{123}, G_{1234}, G_{2345}, X_1, X_2, G_6, G_{67}, G_{678}, G_{6789}$ where for example G_{123} is the centroid of gaze points G_1, G_2, and G_3.

It should be noted that this implementation of n-samples moving average using the previous n points differs slightly from the more common implementation which uses $n/2$ previous and $n/2$ future samples. The reasoning behind this decision is so that the filtered point of gaze will not be affected by future visual stimuli. As an example, consider a stimulus which is a small visual target against a plain background, where the target disappears and reappears in a random location. With the standard moving average implementation, the filtered point of gaze will appear to be pulled towards the location a new target before it is even visible on the display. A slightly delayed saccade response in filtered data has been chosen instead to avoid the pre-visual movement effect.

A.2.5 Linear interpolation

When a subject's head approaches the boundary of the eye tracker's maximum range in any spatial dimension, the recorded samples rapidly flicker between valid and missing data, at a rate of 10 Hz or higher. During this time, the subject is still likely to be focusing on the display monitor, so the valid samples still produce points of gaze near one another. However, the presence of such gaps terminates and invalidates any candidate in the standard I-DT fixation detection, leading to many true fixations going undetected.

Filling in such gaps is simple - Given a single parameter of maximum gap duration, contiguous blocks of missing data shorter than the specified duration are detected, and a linear interpolation is performed on the points of gaze between the valid samples immediately enclosing the gap. The maximum gap duration parameter should be chosen so that short gaps should be filled in but long gaps during which attentional shifts are likely to occur will be unchanged; creating new data which may likely be incorrect can invalidate later analyses.

Interpolation is performed after the moving average filter, for the reason that data points created through interpolation should not affect the filtered point of gaze.

For both the moving average filter and linear interpolation, data points are considered invalid if neither eye can be tracked. For the purpose of averaging and interpolation, if a valid point has only been detected in one eye, its point of gaze is copied to the other eye and then used to produce averaged or interpolated data points with points of gaze for both eyes.

A.2.6 Data trimming

When recording tasks with eye gaze, there is often a period of preparation before the actual task of interest begins, and a period after the task has ended but the recording apparatus continues to operate. In order to standardize task times for analysis, the eye tracking output files can be trimmed according to the true start and end of the performed task. Furthermore, this trimming step should be performed after all filtering steps are complete, since different starting points for I-DT can lead to small differences in detected fixations which will accumulate towards the end of a recording.

A.3 Application of filters to experimental data

A.3.1 I-DT fixation detection

The implementation of custom I-DT closely matches that found in Tobii Clearview 2.7.X, but occasionally splits very long fixations (500 to 3000+ ms) into two separately detected fixations with slightly different centroids whose average position would be the location of the original Clearview fixation. In effect, the aggregate statistics for the custom fixation detection produces a slightly higher total number of fixations with a lower mean fixation duration, although total dwell time for an AOI-based analysis remains unchanged. Table A.2 illustrates this phenomenon, using an example of matched and split fixations in a short segment of an actual gaze recording. On average the custom I-DT implementation produces 10.4 more fixations per minute over Clearview when applied to raw data.

Clearview I-DT				Custom I-DT			
Timestamp	Duration	Point of gaze x	Point of gaze y	Timestamp	Duration	Point of gaze x	Point of gaze y
5825	518	652	217	5825	519	652	217
6384	518	292	636	6384	319	298	625
				6723	179	284	649
6922	1754	241	688	6922	1754	241	688

Table A.2: Comparison between fixations found by Clearview I-DT and the custom I-DT from a short sample of a single trial from the 240 data recordings examined. Most fixations are identical but some are split in the custom I-DT, such as the one highlighted.

Raw data + Clearview I-DT				Filtered data + custom I-DT			
Timestamp	Duration	Point of gaze x	Point of gaze y	Timestamp	Duration	Point of gaze x	Point of gaze y
25430	279	635	310	25530	199	632	306
25729	1854	595	219	25749	558	591	234
				26327	1535	598	215
27623	139	600	213				
27783	1276	605	226	27882	1176	605	226
29098	777	633	599	29138	798	631	599
29896	1035	603	607	29956	975	603	608

Table A.3: Filtering good data produces no harmful effect to the custom I-DT fixation detection.

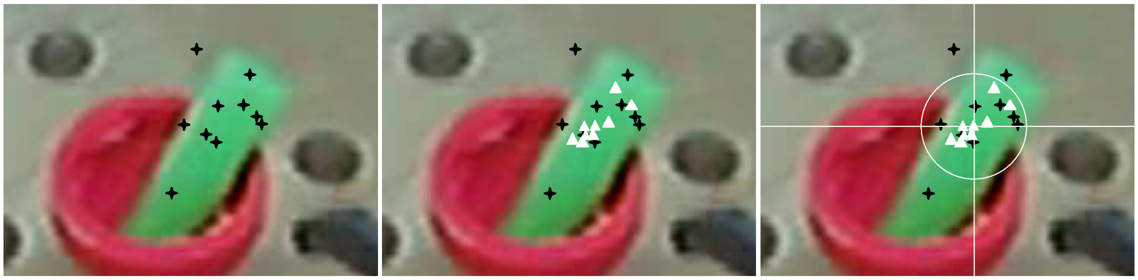
A comparison of raw and filtered fixations in a short segment of a good recording is shown in Table A.3.

TF/TT using the custom I-DT implementation increases by 3% on average across the complete 240-recording dataset, with more new fixations found in jittery recordings.

A.3.2 Moving average and linear interpolation

Figure A.2 illustrates the effect of applying the moving average filter to raw gaze data points, in the absence of linear interpolation. Part (a) shows the raw gaze points (dark crosses) which were not included in any fixation originally; part (b) shows the new filtered gaze points (light triangles) compared to the original points; part (c) illustrates the filtered points detected as a fixation of diameter indicated by the white circle using the Salvucci implementation with the same parameters as before.

Figure A.3 shows the effect of linearly interpolating small gaps of invalid data from



(a) (b) (c)

Figure A.2: Effect of moving average filter on high-amplitude jitter. (a) Original consecutive raw gaze data points (crosses) not included in any fixation, (b) New filtered data points (triangles) with original data points overlaid (crosses), (c) Filtered points now satisfy the original I-DT fixation criteria (shown within in the white circle); the three unfiltered points outside the circle would have terminated the fixation candidate.

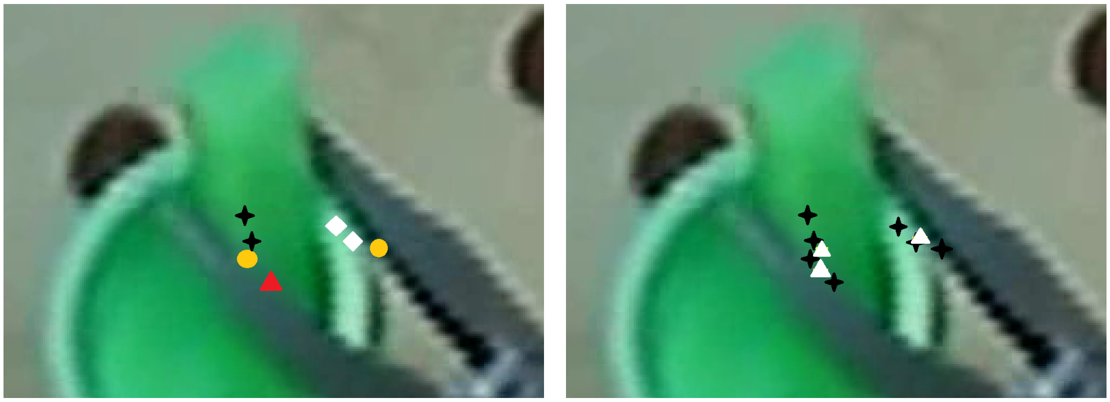
a short segment of our real-world data. Part (a) shows raw eyegaze data points, where each colour group represents consecutive raw gaze points over time of valid data. In this particular case, each group has been split by only a single invalid sample which would terminate any fixation calculation. Part (b) shows that after interpolation, the original raw points (dark crosses) have been supplemented with interpolated data points (light triangles), allowing I-DT a better chance of finding a fixation candidate.

When both the moving average and interpolation are to be applied, the final treatment of the recorded data always first applies the moving average filter, followed by interpolation. For data collected from the Tobii 50-series eye trackers, we use for the moving average a window size of 4 samples (80 ms) and interpolate gaps with a maximum width of 10 samples (200 ms). These settings can be scaled accordingly for data recorded from different eye tracking hardware at other sampling frequencies.

Table A.4 details the effect of filtering on each of the case categories. Of the 76 cases not categorized as “good”, 59 were recovered, with a final TF/TT over 70%. Note that there are 64 cases classified as jitter or gap trials; some cases with a very low TF/TT responded well to the data recovery process but the final TF/TT was not sufficient to become a “good” case for analysis.

A.3.3 Effect of pre-processing on sample-to-sample dispersion of gaze points

The Euclidean distance between adjacent gaze samples was divided by 20 ms (for 50 Hz sampling frequency) and converted to angular velocity. The distribution of these angular



(a) (b)

Figure A.3: Effect of linearly interpolating small intervals of invalid data. (a) Raw eyegaze data points, where each similarly shaped group represents consecutive raw gaze points over time of valid data. In this case, each group has been split by only a single invalid sample. The timeline of gaze samples is ✚✚_▲_●●_◆◆ where each underscore is one missing gaze sample in the timeline. (b) After interpolation, the original raw points (crosses) have been supplemented with interpolated data points (triangles), allowing I-DT a better chance of finding a fixation candidate.

	Good trials (164)	Jitter trials (32)	Gap trials (32)	Bad trials (12)
Mean TF/TT before processing	87%	42%	54%	42%
Mean TF/TT after processing	89%	80%	77%	64%
Mean TF/TT improvement	2%	38%	23%	22%
Mean missing data before (% of task time)	2.3%	1.8%	11.7%	29%
Mean missing data after (% of task time)	0.6%	0.6%	4.9%	18.5%
Percentage of missing data recovered	75%	65%	58.4%	36.3%
Number (%) of invalid recordings recovered for inclusion in analysis	n/a	30 (93.75%)	29 (90.6%)	0

Table A.4: Effect of data recovery filters on good- and poor-quality data. TF/TT denotes total fixation time / total task time.

velocities before and after smoothing is plotted in Figures A.4a, A.4b, and A.4c. Figure A.4a shows the cumulative distribution for recorded trials from all 12 subjects; Figure A.4b illustrates a distribution from a subject who produced good recordings, and Figure A.4c provides an example of the distribution from a subject with jittery recordings. A vertical line at approximately 450 degrees/second angular velocity is drawn as a physiological upper limit for saccadic velocity on this particular setup as observed by Holmqvist et al. [17].

A.3.4 Results of pre-processing on spatial distribution of gaze points

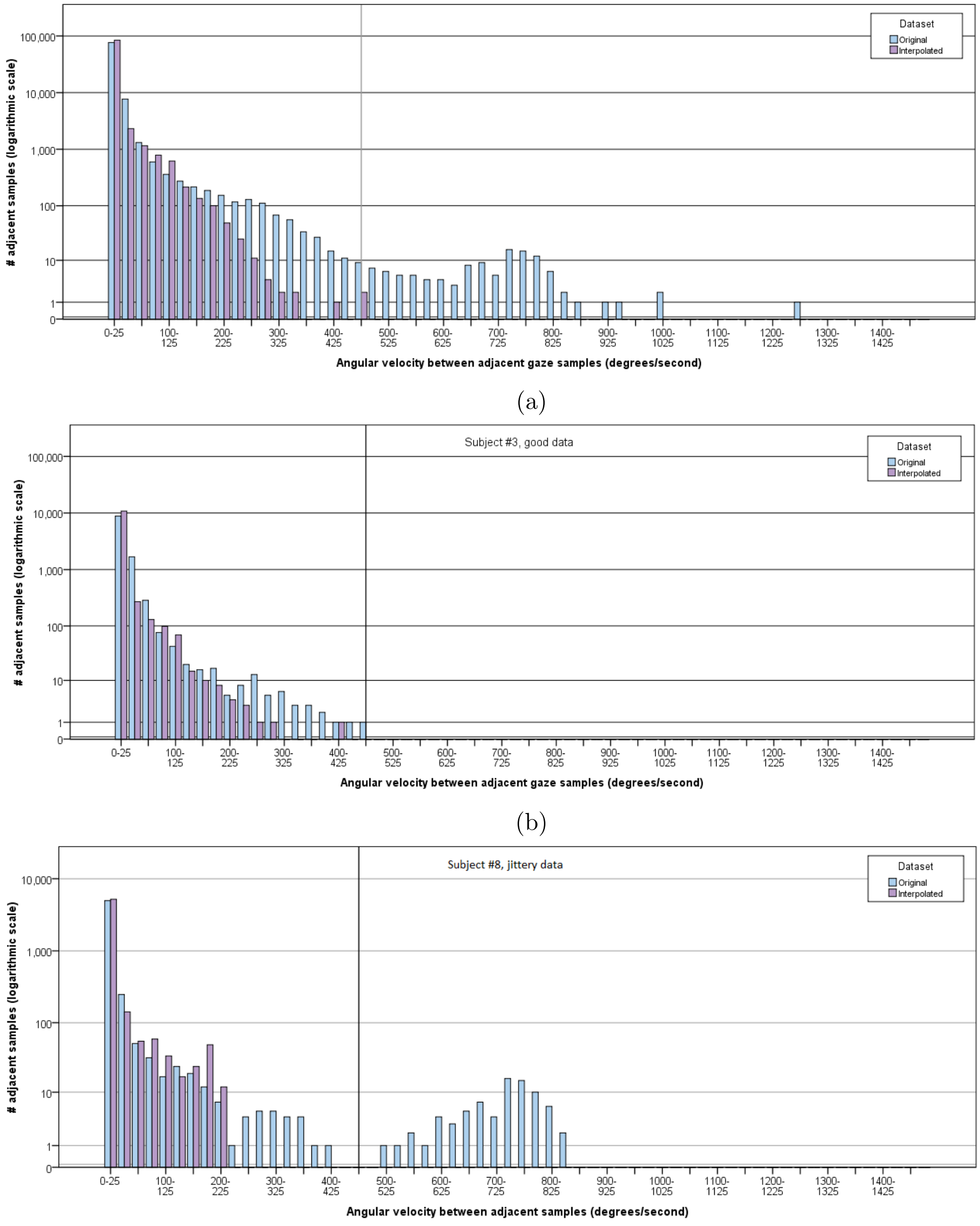
Figure A.5a shows the spatial distribution of the gaze points before and after smoothing accumulated for all 12 subjects. Filtered gaze points are those modified or new data points created from application of the moving average and linear interpolation, and the overlap points indicate where the filtered data points are located within 0.5° visual angle of the corresponding original data point. As such, filtered points created by linear interpolation have no corresponding raw data point and will always be plotted as “Filtered” rather than “Overlap”. A few new downward gaze points were introduced at about six o’clock position on resulting from interpolation on subject 11’s trials shown in Figure A.5b.

A.4 Discussion

As Table A.4 shows, 59 out of 76 initially unusable trials in our real-world data set were recovered and included in analysis, and initially good trials remained good. This indicates a possible recovery of up to 77% of unusable trials in data sets with similar noise characteristics, so it may be advantageous to perform this filtering process prior to any analysis.

However, a possible concern with the application of both the moving average filter and the linear interpolation, is dealing with saccades. At the eye trackers sampling frequency of 50 Hz, a saccade can cross a large part of the display with only one or two samples between the originating point and the target point. When the moving average is applied at the onset of a saccade, the saccade is effectively slightly delayed and is given a smoother acceleration in the filtered data. Using a window size of 4 samples produces a satisfactory balance between increased TF/TT and preservation of saccades, whose onsets will be delayed by 20 ms and appear to accelerate slowly; arrivals to targets show similar behaviour.

Interpolation could also potentially cause problems if a saccade happens to occur during the gap of missing data. If this were to happen and there are indeed two different fixations points before and after the saccade, the effect may be to cause I-DT to extend the earlier

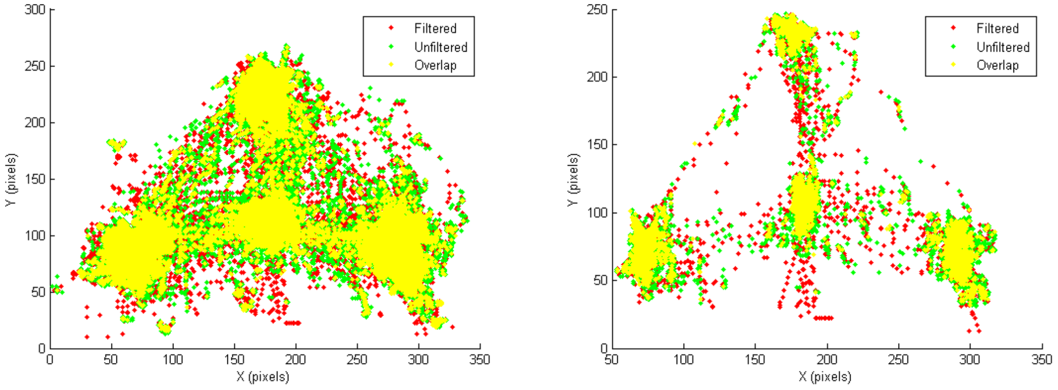


(a)

(b)

(c)

Figure A.4: Profile of sample-to-sample angular velocity before and after interpolation, accumulated for (a) all subjects, (b) a subject who produced good data, and (c) a subject who produced jittery data. No samples should exist beyond a physiological limit of about 450 degrees/second after processing.



(a)

(b)

Figure A.5: (a) Spatial distribution of gaze points pre- and post-smoothing. Only a few gaze points at the lower region of the display were created as a result of interpolating physiological blinks, (b) Spatial distribution of gaze points for subject 11, who had some downward gaze points generated from interpolation.

fixation, and either delay the later fixation or produce two shorter fixations following the saccade.

Another issue with interpolation is that it almost certainly fills in legitimate gaps in the data that occur when the subject blinks, since the duration of these blinks are often longer than rapidly flickering data loss but also shorter than longer intervals of missing data. Since blinking frequency and duration can be correlated to a subject's level of stress or fatigue [68], it is inadvisable to apply the interpolation to data where subject's blinking characteristics are of interest.

There is an additional danger in interpolation of blinks as noted by Holmqvist et al. [17] where the detected pupil centroid experiences a downward shift as the eyelid drops, followed by a rise in the centroid as the pupil reappears below the eyelid at the end of the blink. As can be seen in Figure A.5a, there were only a few gaze point samples created at the lower part of the viewing area Those few that were created all belonged to subject 11 (Figure A.5b) and out of thousands of fixations, registered only one false fixation possibly due to interpolation of blinks.

Finally, the small change to TF/TT observed after applying the filtering and interpolation to "good" gaze recordings indicates that the processing does not affect recordings that are not in need of recovery; thus it is safe to filter all recordings in a study for consistency in data treatment across trials.

For the potentially recoverable "jittery/gapped" cases, TF/TT improvement arises from the cooperation of the two data recovery processes. The interpolation usually only increases the amount of valid data by a small amount, but the added data continuity combined with tighter clustering of raw gaze points produced by smoothing can lead to much more fruitful operation of dispersion-based fixation detection algorithms.

Ideally, verification of the smoothing and gap-filling could be done by adding artificial noise to known good data, followed by application of the filters and comparison to the original unmodified data. However, the nature and origin of noise in the data is unknown and unpredictable and thus we do not have an accurate model of noise generation. Figure A.4 shows that no samples with an unreasonably high angular velocity were produced.

The design and application of these particular smoothing parameters were chosen for working well with the Tobii 50-series eye trackers. However, application of the described techniques to trackers of higher spatial and temporal resolution should merely involve linearly scaling the parameters to cover the desired time intervals. It should be mentioned that the I-DT fixation detection filters used in Tobii Clearview are available for use in Tobii Studio, but were implemented in their original form without gap-filling and smoothing. An important advantage of this data recovery is that it allows expansion and combination of experimental datasets collected using the 50-series with Clearview and the X-series with Studio + I-DT detection. A longitudinal study during which equipment was upgraded can have its data from the old and new equipment shared, as can datasets collected from different locations using different hardware. Furthermore, this smoothing step can be introduced transparently into data analysis, as has already been done successfully by Jiang et al. [20, 21] and Tien et al. [54]. In the event of encountering noisy recordings, Holmqvist et al. [17] recommend adjusting fixation parameters on a per-recording basis. With the smoothing filters, the same process can be applied to all recordings in a dataset and subsequently analyzed using the same fixation parameters for a more consistent procedure with results from different recordings directly comparable to one another.

Appendix B

Experimental documents

B.1 Vigilance in minimally invasive surgery

B.1.1 Participant instructions

B.1.2 Participant instructions

TASK INSTRUCTIONS FOR PARTICIPANT

Nov 3, 2009

We are reading these instructions to you so that all individuals receive exactly the same instructions. Please pay close attention. If you do not understand something, please stop me and I will explain.

You will be asked to perform a partial cholecystectomy on a simulated patients presented. We will provide you a brief history of this patient and you will be required to perform surgical task on this virtual reality model.

You will use a laparoscopic grasper in your left hand and laparoscopic electorcautery (hook) in your right hand. The energy of the cautery will be controlled by a food paddle placed to your right foot. With these tools, you will be able to dissect the gall bladder from the liver without causing side damages to the liver.

The image of the surgical field will be displayed in front of you on a surgical monitor. During the procedure, the patient's vital signs, including heart rate, blood pressure, and saturation of oxygen, will be display on a standard anesthetic monitor placed on the right side of surgical monitor.

You are free to choose any way to perform the task in a comfortable and accurate manner, but your performance will be evaluated by the speed of operation and safety of patient.

The equipment you are using is very safe. The special goggles you are wearing is used to track your eyes movement. It should not interfere with your task performance. Now, you will be given 1 practice trial with the system before we start the real trial. You will be asked to repeatedly perform the same task for two trials. The entire test time will be about 20 minutes. Do you have any questions?

B.1.3 Pre/post-test questionnaires

VR eye tracking study (March 2010) Pre-test Questionnaire (page 1/2)

Subject No.: __________ **Date:** _____________

Gender: Male Female **Age:** ______ **Handedness:** R L

Surgical Training Level: Student R1 R2 R3 R4 R5 Fellow Staff

Number of Years doing Laparoscopic Surgery: 0 1-3 4-6 7-9 10+

Prior experience using the SurgicalSim (hours): 0 1-2 3-5 5-10 10+

Prior experience using the eyetracker or similar device (hours): 0 1-2 3-5 5-10 10+

Do you require vision-correcting eyeglasses or contact lenses? **Yes** **No**

Please circle the number of times you have performed each of the following procedures as a Surgeon *(in column 1)* **and as an assistant** *(in column 2).*

Procedure	**Performed as Surgeon Frequency**					**Performed as Assistant Frequency**				
1) Laparoscopic Cholecystectomy:	0	1-5	6-10	11-15	>15	0	1-5	6-10	11-15	>15
2) Diagnostic Laparoscopy:	0	1-5	6-10	11-15	>15	0	1-5	6-10	11-15	>15
3) Laparoscopic Appedectomy:	0	1-5	6-10	11-15	>15	0	1-5	6-10	11-15	>15
4) Lap Nissen Fundoplication:	0	1-5	6-10	11-15	>15	0	1-5	6-10	11-15	>15
5) Laparoscopic Splenectomy:	0	1-5	6-10	11-15	>15	0	1-5	6-10	11-15	>15
6) Laparoscopic Bowl Resection:	0	1-5	6-10	11-15	>15	0	1-5	6-10	11-15	>15
7) Laparoscopic Adrenalectomy:	0	1-5	6-10	11-15	>15	0	1-5	6-10	11-15	>15
8) Laparoscopic Nephrectomy:	0	1-5	6-10	11-15	>15	0	1-5	6-10	11-15	>15
9) Laparoscopic Bariatric surgery:	0	1-5	6-10	11-15	>15	0	1-5	6-10	11-15	>15
10) Lap Inguinal Hernia Repair:	0	1-5	6-10	11-15	>15	0	1-5	6-10	11-15	>15
11) Lap Incisional Hernia Repair:	0	1-5	6-10	11-15	>15	0	1-5	6-10	11-15	>15
12) Other Advanced Procedure(s):	0	1-5	6-10	11-15	>15	0	1-5	6-10	11-15	>15

Please Name Procedure(s): ____

B.1.4 Pre/post-test questionnaires

VR eye tracking study (March 2010) Post-test Questionnaire

Subject No.: __________ **Date:** ____________
Task: __________ **Trial #:** ____________

1) Do you feel that wearing the Locarna eye tracker interfered with your performance?

Not at All		Somewhat		Very
1	2	3	4	5

2) Do you feel your performance during the simulated procedure was similar to your performance in an actual OR procedure (if applicable)

Not at All		Somewhat		Very
1	2	3	4	5

3) Please rate your performance on the task compared to your best performance.

Worst match		Average		Best match
1	2	3	4	5

4) Based on the information available to you during the procedure, what is your assessment of the patient's condition at various points during the procedure?

If you have any additional comments please write them on the space below.

This ends your participation in this study. Thank you very much for your time!

B.1.5 Patient scenario descriptions

Patient One:

35-year-old female with a history of gallstones was booked for cholecystectomy.
She complains of abdominal pain in the right upper quadrant as well as pain in the back, between the shoulder blades, abdominal bloating, intolerance of fatty food, belching, gas, and indigestion.
On physical examination the right upper quadrant was tender and Murphy's sign was positive.
Post medical history is insignificant, no previous surgeries and she is not taking any medication

B.1.6 Patient scenario descriptions

Patient two:

50 – year- old female with a sudden onset of right upper quadrant pain, low-grade fever, vomiting and nausea presented to the ER. The pain was constant and severe. On physical examination the right upper quadrant was tender and Murphy's sign was positive.
Subsequent laboratory and imaging tests confirmed the diagnosis of acute cholecystitis and the patient was booked for cholecystectomy.

Past-Medical history:
Smoker, history of hypertension, high cholesterol and gout.

Medication:
Patient is taking medication for hypertension and high cholesterol.

B.2 Eye movement patterns in execution and observation of simulated laparoscopic tasks

B.2.1 Participant instructions

Eyegaze differences between watching and doing image-guided manual tasks

Description of experiment and task instructions to participants

In this experiment we will observe eye movement patterns while performing a simulated laparoscopic task and while watching a pre-recorded video of the same task. Please read this procedure carefully and then read and sign the consent form if you agree to give your participation according to the procedure outlined below.

This procedure has two parts – one to be completed today and a second part to be completed at least two weeks after today, at your earliest convenience. With your permission, we will contact you by e-mail to schedule a time for your second visit.

Be aware that even after signing your consent to participate, you may still withdraw your participation in the study at any time and without negative consequences.

Part One

The task is to use the laparoscopic grasper to transport a small object to several locations. Your actions inside the laparoscopic training box will be visible on the computer display in front of you.

The steps of the physical task are listed below:
- begin with the grasper touching the base of the training box inside the central square.
- pick up the item which is resting on the upper peg and place it onto the left peg.
- touch the base of the training box inside the central square.
- pick up the item from the left peg and place it onto the right peg.
- touch the base of the training box inside the central square.
- pick up the item from the right peg and place it onto the upper peg.
- touch the base of the training box inside the central square.

Please try to follow this routine, but do not worry if you miss a step or mix up the step order – simply carry on with the task, ensuring that the item reaches all pegs eventually. If you drop the item, please first make an effort to retrieve it using the grasper. If retrieval is impossible, we will replace the item for you to try again.

You can practice this routine as many times as you wish before we begin the recorded trials. Before beginning the recorded trials, we will also perform a short calibration routine to ensure your eyes are tracked accurately – simply watch the dot as it moves around the screen. Your eye and facial movements and your performance will be recorded onto the PC during the trials. Once you have practiced and completed the calibration procedure, you will be asked to perform the manual task sequence 5 times. You may rest in between each trial if necessary.

B.2.2 Participant instructions

Part Two

We will first perform a calibration routine similar to the one done in Part One. Next we will watch a number of performances recorded from the first part of the study. These will include your own performances as well as those of other participants. If possible, while you are watching, please try to think of how you would carry out the task yourself.

Thank you for your participation in this user study. Knowledge from this study will help us to produce richer training videos for surgical education.

B.2.3 Pre/post-test questionnaires

Eyegaze differences between watching and doing image-guided manual tasks

Subject No.: ______________ **Date:** __________________

Gender: Male Female **Age:** ________ **Handedness:** R L

The following questions are to be completed prior to beginning Session 1.

Do you require vision-correcting eyeglasses or contact lenses? Yes No

Your experience with playing video games or performing other image-guided tasks:

Little to none 1 2 3 4 5 6 7 Abundant, regular habit

The following questions are to be completed immediately following the completion of Session 1.

How difficult did you feel it was to manipulate the surgical instrument in the manual task?

Very easy 1 2 3 4 5 6 7 Very difficult

How difficult did you feel it was to watch the objects on the monitor during the manual task?

Very easy 1 2 3 4 5 6 7 Very difficult

How difficult did you feel it was to complete the entire manual task without making mistakes (objects slipping or dropping) ?

Very easy 1 2 3 4 5 6 7 Very difficult

Please estimate your stress level during the manual task.

Not stressful 1 2 3 4 5 6 7 Very stressful

Session2 questions on reverse

B.2.4 Pre/post-test questionnaires

Eyegaze differences between watching and doing image-guided manual tasks

Subject No.: ______________ **Date:** __________________

Session 2: Video 1. (ID: _________)

How stressful or difficult did you feel it was to watch and follow the recorded task in these videos?

Low stress 1 2 3 4 5 6 7 High stress

Do you feel that you completed the task last week with better performance than one or more of the videos you watched? Yes No Same

To what degree do you feel that your performance can become better than the videos you watched?

No change 1 2 3 4 5 6 7 Significantly better

To what degree do you feel that you can improve your performance *by watching these videos*?

No improvement 1 2 3 4 5 6 7 Significantimprovement

Session 2: Video 2. (ID: _________)

How stressful or difficult did you feel it was to watch and follow the recorded task in these videos?

Low stress 1 2 3 4 5 6 7 High stress

Do you feel that you completed the task last week with better performance than one or more of the videos you watched? Yes No Same

To what degree do you feel that your performance can become better than the videos you watched?

No change 1 2 3 4 5 6 7 Significantly better

To what degree do you feel that you can improve your performance *by watching these videos*?

No improvement 1 2 3 4 5 6 7 Significantimprovement

Session 2: Video 3. (ID: _________)

How stressful or difficult did you feel it was to watch and follow the recorded task in these videos?

Low stress 1 2 3 4 5 6 7 High stress

Do you feel that you completed the task last week with better performance than one or more of the videos you watched? Yes No Same

To what degree do you feel that your performance can become better than the videos you watched?

No change 1 2 3 4 5 6 7 Significantly better

To what degree do you feel that you can improve your performance *by watching these videos*?

No improvement 1 2 3 4 5 6 7 Significantimprovement

B.2.5 Pre/post-test questionnaires

Session 2: Video 4. (ID: _________)

How stressful or difficult did you feel it was to watch and follow the recorded task in these videos?

Low stress 1 2 3 4 5 6 7 High stress

Do you feel that you completed the task last week with better performance than one or more of the videos you watched? Yes No Same

To what degree do you feel that your performance can become better than the videos you watched?

No change 1 2 3 4 5 6 7 Significantly better

To what degree do you feel that you can improve your performance *by watching these videos*?

No improvement 1 2 3 4 5 6 7 Significantimprovement

Session 2: Video 5. (ID: _________)

How stressful or difficult did you feel it was to watch and follow the recorded task in these videos?

Low stress 1 2 3 4 5 6 7 High stress

Do you feel that you completed the task last week with better performance than one or more of the videos you watched? Yes No Same

To what degree do you feel that your performance can become better than the videos you watched?

No change 1 2 3 4 5 6 7 Significantly better

To what degree do you feel that you can improve your performance *by watching these videos*?

No improvement 1 2 3 4 5 6 7 Significantimprovement

Session 2: Video 6. (ID: _________)

How stressful or difficult did you feel it was to watch and follow the recorded task in these videos?

Low stress 1 2 3 4 5 6 7 High stress

Do you feel that you completed the task last week with better performance than one or more of the videos you watched? Yes No Same

To what degree do you feel that your performance can become better than the videos you watched?

No change 1 2 3 4 5 6 7 Significantly better

To what degree do you feel that you can improve your performance *by watching these videos*?

No improvement 1 2 3 4 5 6 7 Significantimprovement

B.2.6 Pre/post-test questionnaires

Session 2: Video 7. (ID: _________)

How stressful or difficult did you feel it was to watch and follow the recorded task in these videos?

Low stress 1 2 3 4 5 6 7 High stress

Do you feel that you completed the task last week with better performance than one or more of the videos you watched? Yes No Same

To what degree do you feel that your performance can become better than the videos you watched?

No change 1 2 3 4 5 6 7 Significantly better

To what degree do you feel that you can improve your performance *by watching these videos*?

No improvement 1 2 3 4 5 6 7 Significantimprovement

Session 2: Video 8. (ID: _________)

How stressful or difficult did you feel it was to watch and follow the recorded task in these videos?

Low stress 1 2 3 4 5 6 7 High stress

Do you feel that you completed the task last week with better performance than one or more of the videos you watched? Yes No Same

To what degree do you feel that your performance can become better than the videos you watched?

No change 1 2 3 4 5 6 7 Significantly better

To what degree do you feel that you can improve your performance *by watching these videos*?

No improvement 1 2 3 4 5 6 7 Significantimprovement

Session 2: Video 9. (ID: _________)

How stressful or difficult did you feel it was to watch and follow the recorded task in these videos?

Low stress 1 2 3 4 5 6 7 High stress

Do you feel that you completed the task last week with better performance than one or more of the videos you watched? Yes No Same

To what degree do you feel that your performance can become better than the videos you watched?

No change 1 2 3 4 5 6 7 Significantly better

To what degree do you feel that you can improve your performance *by watching these videos*?

No improvement 1 2 3 4 5 6 7 Significantimprovement

B.2.7 Pre/post-test questionnaires

Session 2: Video 10. (ID: _________)

How stressful or difficult did you feel it was to watch and follow the recorded task in these videos?

Low stress 1 2 3 4 5 6 7 High stress

Do you feel that you completed the task last week with better performance than one or more of the videos you watched? Yes No Same

To what degree do you feel that your performance can become better than the videos you watched?

No change 1 2 3 4 5 6 7 Significantly better

To what degree do you feel that you can improve your performance *by watching these videos*?

No improvement 1 2 3 4 5 6 7 Significantimprovement

How drowsy did you feel while watching these videos?

Not drowsy, full wakefulness Very drowsy, falling asleep

1 2 3 4 5 6 7

Your experience with tasks involving manual dexterity (e.g. sewing, carpentry, art crafts, assembly)

Little to none Abundant, regular habit

1 2 3 4 5 6 7

At the time of this study, were you suffering any allergies or other conditions that can cause irritation or discomfort of the eyes? Yes No

Here concludes your participation in this study. Thank you for your contribution.

B.3 Gaze training in laparoscopy

B.3.1 Experiment briefing

Eye-hand coordination in laparoscopic tasks

Briefing / Experimental overview

This study aims to investigate the development of eye-hand coordination during a laparoscopic task.

Your eye movements will be recorded by a remote eye tracker and your head movements will be tracked with a web camera. The eye tracker uses near-infrared illumination and poses no known health risks. Images captured by the web camera will be used purely for motion analysis and will not be published.

You will be asked to perform a peg transport task inside a laparoscopic training box, which the experimenter will demonstrate. A single task trial is expected to take roughly 20-60 seconds. You will be allowed to practice the task for 3 trials. Then after a break you will be asked to repeat the task for a total of 30 trials. You will be allowed to take breaks between every 5 trials.

You have the freedom to fully withdraw from the study at any time.

If you agree to participate in this study, you will provide your informed consent and complete a questionnaire to collect basic demographic data. Your name will only be used for internal identification during analysis and will not be included in any published results. Your complete participation in this study will involve a short revisit to this lab in at least one month’s time following the end of today’s session. You will be asked to provide a method of contact for the experimenters to reach you at that time.

B.3.2 Pre-test questionnaire

Eye-hand coordination in laparoscopic tasks

Pre-test questionnaire

Date: ____________________ Subject group: E / R1 / N / DL / GT-B2 / GT-CF

Name: ___________________ Subject number: _____________________

Age: _____________________ Gender: M / F

Dominant hand: Right / Left / Either

Do you require vision-correcting lenses? No / Eyeglasses / Contact lenses

Amount of experience in laparoscopy: ___________ years

(Surgical residents only) Post-graduate year level: ______ Specialty: ___________

B.3.3 Post-test questionnaire

Eye-hand coordination in laparoscopic tasks

Post-test questionnaire

Date: ___________________________

Subject name: __________________ Condition: VT / GT-B2 / GT-CF

Please first complete an electronic copy of the NASA Task Load Index.

(GT-B2 and GT-CF subjects only below)

The target illumination helped me to focus on the target (circle the applicable response):

Strongly disagree 1 2 3 4 5 Strongly agree

The target illumination distracted me from completing the task:

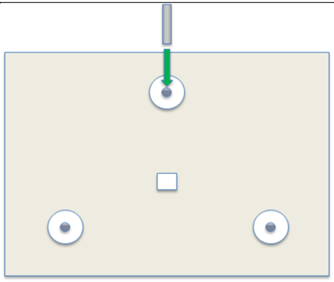
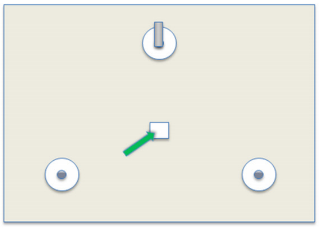
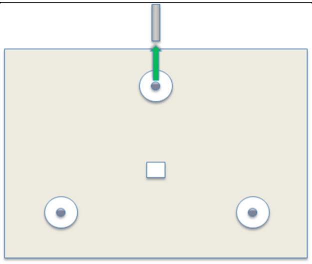
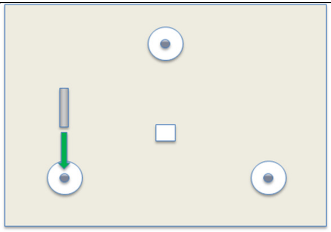
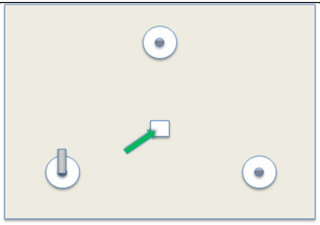
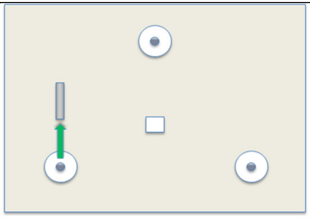
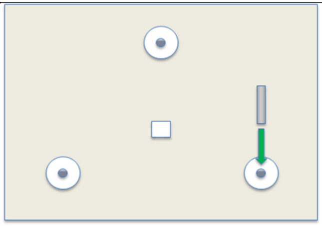
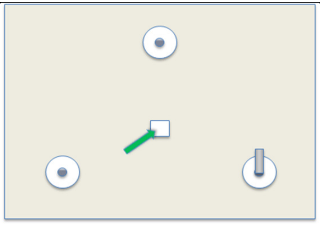
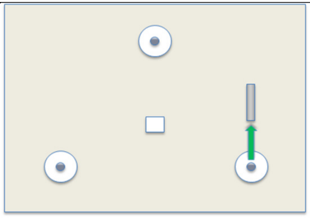
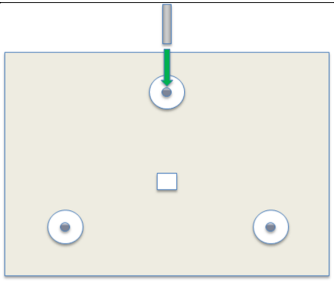
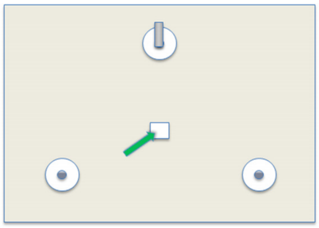
Strongly disagree 1 2 3 4 5 Strongly agree

B.3.4 General task instructions

Eye-hand coordination in laparoscopic tasks – General task instructions

The task is to use the laparoscopic grasper to move a peg between three cups in an anticlockwise direction, and pressing the central button between each transport. The experimenter will demonstrate this for you. When placing the peg into a cup, please ensure that the peg is fully inserted into the slot and not simply resting in place. If you drop the peg, please try to pick it up and continue the trial. If the peg cannot be reached, inform the experimenter and you may restart the trial.

Please follow the transport sequence illustrated below:

 1. Insert peg into upper cup	 2. Press the central button	 3. Pick up peg from upper cup
 4. Insert peg into lower left cup	 5. Press the central button	 6. Pick up peg from lower left cup
 7. Insert peg into lower right cup	 8. Press the central button	 9. Pick up peg from lower right cup
 10. Insert peg into upper cup	 11. Press the central button	

B.3.5 Verbal training condition task instructions

Eye-hand coordination in laparoscopic tasks

Task instructions (Group VT)

The task is the same as before. You will be asked to repeat this sequence for a total of 30 times, and you will be allowed to rest between each block of 5 trials.

When performing this task, please do your best to focus your gaze on the target location, and minimize the amount of time spent looking at the grasper.

B.3.6 Gaze training condition task instructions

Eye-hand coordination in laparoscopic tasks

Task instructions (Group GT-B2, Group GT-CF)

The task is the same as before. You will be asked to repeat this sequence for a total of 30 times, and you will be allowed to rest between each block of 5 trials.

When performing this task, a flashing red light indicates the location of the next target – either one of the cups, or the central button .