Hand and Eye Gaze Analysis for the Objective Assessment of Open Surgical Dexterity

by

Simon Charles Byrns

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

Master of Science

in

Experimental Surgery

Department of Surgery
University of Alberta

Abstract

Objective assessment of technical skill remains a challenging task. Paper based evaluations completed by expert assessors have been criticized for not accurately or consistently describing a surgeons' technical proficiency due to inter-observer variability and subjective bias. In the laparoscopic or minimally invasive surgical domain, technology assisted evaluation has been shown to provide a reliable and objective measure of performance based on motion analysis, focusing on instrument movement and gestures. Aided by the miniaturization of motion tracking technology, this thesis focuses on the development of novel techniques for acquiring synchronized hand motion and eye tracking data in *open* surgical procedures.

An overview of motor learning theory is provided as a basis for segmenting or decomposing surgical movements into constituent gestures. An empirical study investigating the learning effects of a visuospatial intensive video game as a substitute for traditional practice was performed, and showed that video gaming, can in some conditions, enhance or reinforce traditional simulator based practice.

Existing motion capture techniques are reviewed along with an analysis of computational models used in high level motion analysis. A second empirical study was completed to investigate the application of one of these computer models to hand motion captured via an optical marker-less tracking device. Hidden Markov Models applied to the motion data was able to discriminate between participants emulating different levels of dexterity.

Finally, the development of a technology-assisted assessment system for evaluating a surgeons' performance based on synchronized hand motion, eye gaze and force application

in open surgical techniques is presented. Several empirical studies designed to validate this system are described. The novel aspects of this system include the ability to capture eye gaze in a 3-dimensional environment as well as highly detailed hand motion based on a surgical glove system where 6D electromagnetic sensors are embedded. The design and assembly of this apparatus is described including an overview of the software required for achieving spatial and temporal coherence.

The thesis concludes with a summary of findings and a brief discussion of planned experiments necessary to validate the clinical utility of a surgical motion and eye tracking system for both objective assessment and training purposes.

Preface

This thesis is an original work by Simon Byrns. The research project, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board, Project Name "Surgical Motion Analysis as a Performance Metric and Training Tool", Study ID. Pro00056225, September 21, 2015; and Project Name "Feasibility of Transferring Videogame Skills to Laparoscopic Surgery", Study ID. Pro00048207, May 28, 2014.

A version of Chapter 3 has been submitted for publication as X. Sun, S. Byrns, I. Cheng, B. Zheng, A. Basu. "Smart Sensor-Based Motion Detection System for Hand Movement Training in Open Surgery," *Journal of Medical Systems*. I was responsible for the design of the experiment, assembly of the simulation apparatus, literature review, assisted with data collection, and composition and significant revisions of the manuscript. X. Sun was responsible for designing the software necessary for data collection and analysis as well as manuscript composition. I. Cheng, B. Zheng and A. Basu were supervisory authors who assisted with revision of the manuscript.

A version of Chapter 4 was submitted for publication as S. Byrns, B. Zheng. "Combining Traditional Lap Box Practice with Video Gaming: A Randomized Control Trial," to *Surgical Endoscopy*. I was responsible for the literature review, design of the experiment, data collection and analysis and composition of the manuscript. B. Zheng was the supervisory author responsible for revision of the manuscript.

The empirical studies presented in Chapter 7 and 8 were the result of a collaboration with M. Feist and X. Jiang and were withheld from publication pending a patent application for the technology described in these sections. I was responsible for the design of the experiments, data collection, data analysis and drafting of the manuscripts. M. Feist developed the software necessary for data collection. X. Jiang assisted with data analysis.

Acknowledgements

I am indebted to several collaborators from the Department of Computer Science for their assistance in completing the software implementations necessary to conduct my experiments and subsequent data analysis.

Xinyao (Alvin), thank you for your help designing the first experiments we completed with optical tracking and the first implementation of HMM for studying hand motion and gestures. You worked tirelessly to implement all the features I requested and were always available to troubleshoot and discuss improvements to the software.

Xianta, thank you for all of your help both with all the MATLAB scripts and coding and feedback regarding the design of the experiments used to test the accuracy of the multimodal system.

Michael, thank you for allowing me to use your software for collecting and storing all the synchronized data, for instantaneous responses to emails, and for recompiling the recording software so many times.

A special thanks to Christine for supporting me throughout my research. Thank you for all the home cooked meals and for your patience listening to all my frustrations.

Finally, a big thanks to my supervisory committee for providing support and guidance, attending 6 month reviews and providing the resources necessary for completing this work. I must also acknowledge the generous support I received in the form of Scholarships from the Faculty of Medicine at the University of Alberta and the Canadian Institutes for Health Research (CIHR).

Table of Contents

List of Tables

List of Figures

List of Symbols and Abbreviations

2D	2-Dimensional
3D	3-Dimensional
API	Application Program Interface
DIP	Distal Interphalangeal (Joint)
DOF	Depth of Field
EM	Electromagnetic
FLS	Fundamentals of Laparoscopic Surgery
FOV	Field of View
FSR	Force Sensitive Resistor
HMM	Hidden Markov Model
ICSAD	Imperial College Surgical Assessment Device
IR	Infrared
LED	Light Emitting Diode
MCP	Metacarpophalangeal (Joint)
MIS	Minimally Invasive Surgery
MM	Markov Model
OSATS	Objective Structured Assessment of Technical Skill
PIP	Proximal Interphalangeal (Joint)
PVC	Polyvinyl Chloride
XML	Extensible Markup Language

Chapter 1 – Introduction

1.1 Modern Surgical Training

Surgical training has been steeped in tradition for more than a century. The original Halstedian model introduced by Sir William Halsted at Johns Hopkins Hospital and School of Medicine in 1889 replaced the inconsistent and unstructured apprenticeship model of the time. Halsted's model (1) involved an integration of basic sciences into the training curriculum, along with intense and repetitive exposure to the care of surgical patients. Trainees continued to work under the tutelage of a senior staff surgeon, but were granted increasing responsibility depending on their skill level. More than a century later, our current resident training programs continue to adopt this approach. The often repeated mantra 'see one, do one, teach one' aptly describes our contemporary model, where residents participate in a large volume of cases and through repeated exposure develop competency in performing surgical procedures. However, this approach (2) is extremely time consuming and inconsistent between residents, especially for less frequently performed procedures. In addition, this particular teaching method has long been criticized for not providing a standardized means of assessing skill between trainees, making any assessment of competency difficult. These challenges have prompted surgical educators (3) to question whether this training style is sustainable given the increasing technical nature of surgery and the growing time and legal pressures on surgical trainees.

Adding to the complexity of modern surgical training, the use of laparoscopy has dramatically increased in the last two decades. Minimally Invasive Surgery (MIS) procedures (4) demand an exceptional level of coordinated psychomotor activity. Dissecting the components of even the simplest maneuvers reveals a complex path involving hand-eye coordination, hand-instrument coordination and extrapolation of a 2-dimensional video image into a 3-dimensional working environment. This has motivated surgical educators (5) to look for additional teaching strategies that address these specific challenges that are unique to laparoscopy. Consequently, the curriculum for surgical trainees is expanding as graduates are expected to demonstrate competency in both open and certain laparoscopic techniques.

Less time in the operating room, the desire to improve operating room throughput, and increasing emphasis on mitigating medical error have all contributed to less operating experience (6) for current surgical trainees. In the United States, new labour laws have been introduced in many medical centres that limit the resident work week to 80 hours. These pressures (7,8) have led to more emphasis on simulated surgical training in current curricula in order to supplement and prepare surgical trainees for experience in the operating room. Luckily, surgical simulation has been around for some time. In the 1990s when MIS became more widespread, the challenge of retraining surgeons already in practice was solved in part by surgical simulation. By shifting the learning curve outside of the OR (9), simulation allowed for a safe and standardized environment in which to teach and assess new technical skills. Evidence began to show that practicing surgical skills in a virtual or simulated environment resulted in improved performance in the operating room.(10,11)

Let us now review some of the literature describing the application of a specific training technique, video gaming, for improving dexterity and performance in laparoscopic techniques. There is a paucity of studies evaluating the use of video gaming as a training technique for laparoscopic novices. We hypothesized that video gaming could substitute traditional simulated laparoscopic practice, resulting in similar performance improvements in fundamental laparoscopic tasks. We designed a randomized control trial to explore the effect of a standardized training program on two simulated laparoscopic tasks.

While MIS simulators improved the efficiency of learning laparoscopic techniques, few simulators have been developed to improve surgical skill in traditional *open* techniques. Even with an increasing prevalence of MIS procedures, many surgical educators acknowledge the fact that current trainees should be competent in the equivalent traditional open procedure prior to learning MIS.

Simulation allows for more deliberate practice, where a task can be repeated in a controlled environment with the aid of an instructor to provide feedback (12). However, simply making new simulation technology available is unlikely to improve surgical skill alone. Objective assessment comprises the second and perhaps the more crucial part of the equation necessary for success. Improved methods to evaluate and measure performance will give educators an ability to target specific deficiencies in each trainee and perhaps provide guidance as when selecting remedial tasks. Fortunately, existing technology available in some laparoscopic simulators permit the recording of instrument motion and forces that can be analysed to compute a quantitative measure of performance. For example, computer generated virtual reality simulators such as the LAP Mentor system generates performance data based on accuracy of dissection, time to complete each step of a procedure, and critical errors that might result in injury to a patient (2).

While progress has been made in generating more reliable measures for performance in MIS procedures, few *open* surgical simulation models have been validated. A review of the current use of open surgery simulation reveals that there is a need for simulation based studies on less frequently performed open procedures (13). In addition, further research is required to evaluate the transferability of skills honed on open surgical simulators to the operating room in order to determine how open simulators should be designed.

Current evaluation methods in surgical specialist training involve the use of paper based assessments that guide expert evaluators when scrutinizing trainees. Examples of these include the Objective Structured Assessment of Technical Skills (OSATS) (14), the Global Rating Index for Technical Skills (GRITS) (15), Direct Observation of Procedural Skills (DOPS) (16), and the Ottawa Surgical Competency Operating Room Evaluation (O-SCORE) (17). Despite attempts to improve the objectivity of these assessments, these tools are subject to a number of biases including inter-observer variability, central tendency bias, and lack of blinding of the assessors to a candidate's level of training (18). In addition, many of these evaluations are procedure specific, limiting their generalizability to other procedures.

While there is still a significant disparity between current simulation models in MIS and open surgery, some of the techniques and technology developed for acquiring and analysing instrument motion data in the MIS domain are transferrable. An additional aim of this thesis is to demonstrate that motion analysis, a method for capturing and analysing human motor movement, can be applied to the development of novel assessment methods for open surgical techniques.

We hypothesized that a technology-assisted evaluation system could be as good or better than current expert or paper based evaluations of surgical trainees for assessing open surgical dexterity. We sought to show that decomposition of surgical tasks using this technology could be used to compare gestures between individuals and that this could be the basis of a new objective assessment strategy. In addition, we hypothesized that this granular level of analysis could be used to provide useful feedback to trainees. In order to accomplish this, we required the development of a novel motion capture system with a spatial resolution high enough, on the order of a few mm, to reliably detect hand gestures during complex bimanual tasks such as open surgical maneuvers used for hand tying or suturing. Similar technology has been applied to analyse instrument movement in MIS, but the complexity of hand and eye movement has limited the extension of this approach to open surgery. We sought to develop a system capable of collecting highly precise hand and finger position and combine this with eye tracking to be able to provide enough detail to a computer model capable of generating a set of performance metrics.

1.2 Contributions and Thesis Organization

This thesis is divided into three major parts. The first part provides an overview of motor skill acquisition and the theories describing the development of technical expertise (Chapter 2). The use of and current evidence for simulation to enhance surgical dexterity is also discussed. To support this theoretical framework with experimental evidence, a control laboratory study investigated whether video gaming could be used to substitute traditional simulated laparoscopic practice in novices (Chapter 3). This study represents a significant contribution to the field of surgical education, as it is the largest study of its kind to date, and consisted of a randomized control trial designed to investigate the effect of a video gaming training program on simulated laparoscopic skill. The results would be of interest to surgical program directors considering the integration of video gaming into modern training curricula.

The second major part of this thesis involves a review of current technology and analytical methods used in surgical motion analysis (Chapter 4). Language models, a method for decomposing movements into a series of characteristic gestures are introduced. It has been more than a decade since the introduction of an electromagnetic tracking system for analysing movement during rudimentary surgical tasks. An empirical study (Chapter 5) was completed to investigate whether the same data could be acquired with the use of a marker-less optical tracking system while participants completed a surgical hand tie. This study also demonstrated the application of language models to the hand motion data for differentiating between individuals with different levels of expertise. While the optical capture technique had some limitations mainly due to occlusions, the analysis was independent of the acquisition technique.

The third and final section of this thesis (Chapters 6-9) describe the development of a novel system capable of capturing and recording spatially and temporally synchronized three-dimensional eye tracking, hand kinematics, and instrument forces. This system represents a novel technology, capable of recording extremely precise and accurate human-instrument-task interactions. While our intended purpose for this technology is surgical skill assessment and training, the highly detailed recording of human interaction with the 3D environment has tremendous implications for improving our understanding of motor skill acquisition, execution and impairment. Following a discussion of the design and implementation of this system, two validation experiments were completed (Chapters 7 and 8) to demonstrate the accuracy of the system. Chapter 9 summarizes the conclusions drawn from the studies presented and provides an overview of future research applications and planned experiments utilizing this new system.

Chapter 2 – Learning Technical Skills

2.1 Motor Skill Acquisition

The process of human motor skill acquisition has been studied extensively, and these theories can be directly applied to modern surgical skills training. Adams introduced a summation of several theories in experimental psychology in the early 1970s regarding learning skilled movement that gave rise to a 'closed loop theory of motor learning' (19). This described feedback derived from proprioceptive, auditory, memory and verbal queues that allows individuals to refine their movement. Adams described having a knowledge of results (KR) based on dichotomous feedback from multiple queues such as 'right' or 'wrong' or 'too long' or 'too short' that provided feedback to refine motor movement. Previous open loop theories had postulated that motor movement was learned as a result of trial and error or evolution of multiple attempts at skilled movement until expertise was obtained. An open loop system has no feedback or compensatory mechanisms. The closed loop theory in contrast is analogous to a physiologic feedback loop with various external queues that provide either positive or negative feedback to refine motor movements. There was a strong view at that time that motor movement and verbal processing were intrinsically related and that our internal voices provide a significant amount of feedback while learning and refining motor movement. Each individual has a perceptual and memory outline or 'trace' of a motor movement. As they work to refine their motor skill, the closed loop feedback provides insight into how accurate one's movements mimic their previous and an ideal perceptual or memory trace.

Adams identified several limitations that the closed loop theory could not rationalise. One was a need for peripheral feedback in error processing, and he suggested that a portion of error processing might be completed centrally i.e. in the brain without the need for peripheral feedback. In addition, anecdotal evidence pointed to examples where feedback obtained from some sensory queues actually impaired or worsened performance. His example was with experienced musicians such as a pianist or violinist whose movement proficiency declines when they watch their fingers performing highly skilled movement.

Another prevalent theory was introduced by Schmidt in the mid 1970's who described the *schema theory* for discrete motor learning. A schema is a concept that groups similar events into a particular class that subsequently defines what additional events should be included in the defined class (20). Schmidt defined four sources of information necessary when performing a discrete motor task. The first is knowledge of the initial conditions of the performer in relation to the environment prior to executing a movement. The second are the response specifications needed to execute the task. The third is knowledge of the desired outcome of the movement, and the final component is the sensory consequences that result from performing a movement. Two additional 'schemata' are employed in this theory. The first one, recall schemata, describes the relationship between response specifications and the outcome that develops from experience. A novel response is generated by matching recall schemata with appropriate initial conditions and a desired outcome to generate response specifications which can be executed by a motor program.

The second scheme called the 'recognition schema' is the relationship between sensory consequences and outcomes that develops from experience. An individual will choose a desired outcome and using a recognition schema will be able to predict the expected sensory consequences of a given response, then without using knowledge of the outcome compare the actual sensory response with the expected and use this information to further refine the recall schemata. The strength of the recall and recognition schema is dependent on the volume and variability of experience or practice producing a given motor response. Schmidt also theorised that individuals with stronger schema would produce less errors in motor movement when performing novel movements and also demonstrate a more rapid learning curve. Margolis and Christina were able to validate the theory of improved initial performance of a motor task in subjects that experienced greater variability in previous practice or training sessions (20).

A third theory of motor skill acquisition has survived Adams and Schmidt's theories to become the most prevalent contemporary theory for learning a skilled motor movement. The three stage theory of motor skill acquisition first described by Fitts and Posner in the 1960s demonstrates the necessary steps required to obtain proficiency and eventually expertise for a specific task (21). The first stage, the *cognitive stage*, involves intellectualization of the task. This requires a trainee to break down a task into discrete steps in order to understand the mechanics of the skill. Reznick provides an example of tying a surgical knot where a learner in the cognitive stage must first understand the mechanics of how to hold the tie, throw the suture and move each hand to complete a knot (6). Performance is unpredictable and an observer is more likely to identify each discrete step of a task. With additional practice, the *integrative* stage describes translation of knowledge into learned motor behaviour. This involves stringing together small portions of motor movements and assembling them to create the desired overall movement. In this stage the steps of the task are no longer distinct and movements are more fluid. Returning to the suturing analogy, the optimal position of a suture on finger, the degree to which the hand is supinated or pronated and the amount of force used to tie each throw of the suture are refined in this stage. Finally, in the third and final stage, the *autonomous* stage, movements are smooth and efficient and the trainee no longer has to think about the individual components of the task. This learning process is not exclusively linear, but rather recursive, requiring trainees to revisit the cognitive and integrative stages when exposed to additional layers of complexity. For example, introducing the additional step of using an index finger to apply more precise counter tension when hand tying requires trainees to revisit the previous learning stages in order to properly integrate this maneuver into a previously learned skill.

Expertise can eventually be obtained through gradual improvement in performance with enough practice in a particular motor skill, representing the highest level of skill acquisition (6). However, not all surgeons are equal. Ericsson defines experts as surgeons with consistently better outcomes than non-experts, but argues that most surgeons probably do not attain true expertise (12). Instead, surgeons likely reach an average level of performance and remain at this level for the remainder of their careers.

2.2 Kinesthetic Learning Theory

Kinesthetic motor learning describes learning by methods where instruction focuses on how to achieve a particular spatial configuration of one's body or in the case of surgery, one's hands, by paying special attention to sensory input and control of motor output. This is similar to the concept of deliberate practice or being mindful of the subtleties of each movement necessary to attain expert performance (22). The kinesthetic approach was further explored in a recent randomized control trial that compared the traditional method of teaching surgical knot tying with a kinesthetic teaching method (23). Students who were trained to tie sutures using a kinesthetic approach performed significantly better as assessed on a global rating scale by blinded expert reviewers. Decomposition of the knot tying task into discrete kinesthetic elements is analogous to the segmentation of kinematic data afforded by computational modeling techniques. These will be described in detail in Chapter 3. Comparison between individuals at this level has the potential for generating a reliable performance metric and provides a foundation for providing pertinent feedback to trainees by identifying deficiencies in specific surgical gestures.

Previous studies have established the utility of gesture-based analysis for identifying the portions of a particular task responsible for a performance deficit. For example, following segmentation of a laparoscopic instrument motion during a ring transfer task into a number of discrete gestures using affine velocity, Cifuentes et. al. compared the relative potential energy of each gesture to reliably discriminate between novices and participants with previous experience in laparoscopic surgery (24,25). Two of the gestures analysed had a significantly lower potential energy when performed by expert surgeons and compared to novices. This granular level of analysis was able to identify the particular portions of a task that require additional practice and serves as means for objectively monitoring the performance improvement over time.

2.2 Role of Simulation

Simulation addresses two critical concerns in current healthcare education. The first is the need to provide an ethical opportunity for novice healthcare providers to learn new skills. This prevents clinicians in training from ‘learning on’ patients (26). Second, it provides a controlled environment in which to offer repeated practice in a particular technique until proficiency is obtained. Kneebone describes an important distinction between simulations which recreate an entire clinical event and simulators which focus on a particular task and are typically physical models (26). Traditionally, technical skills were introduced to novice surgeons in the operating room, exposing patients to more harm. The last decade has seen an explosion of simulator based surgical training. A variety of devices have been used to simulate procedures from low fidelity bench top models used to teach suturing to computer based laparoscopy simulators that provide haptic feedback. The complexity or fidelity of a simulator, as will be described later, does not need to be enormously intricate in order to provide and effective construct on which to learn a particular technical skill. However, Kneebone does emphasize the need for a simulator to be engaging or challenging enough in order to relay relevance to a trainee and provide motivation for its continued use.

An extension to the concept of fidelity in simulation is the specificity of practice hypothesis first introduced by Henry in the late 1960s (27) and later expanded upon by Proteau and collegues in the late 1980s (28,29). This hypothesis describes how motor skills are best learnt when the practiced movement and environmental conditions closely mimic those of the target context. A movement plan or motor program is developed that is specific to the afferent sensory information available during practice. Consequently, if the environment and/or afferent sensory information changes, the learned motor program is no longer applicable to the new conditions, resulting in a reduction in performance. For example, an individual may practice a specific motor skill on a surgical simulator, but given the significant change in context, may not be able to execute this same skill in the operating room at the same level of proficiency. This emphasizes an important consideration in the design of any simulators or simulation environments. Simulators intended to teach technical skills should provide a trainee with afferent sensory feedback similar to what one would obtain when completing the procedure on a real patient. Hybrid simulation may be the answer to providing more accurate contextual information and improving the realism of a simulator (26). Kneebone describes combining inanimate models with standardized patients in order to provide an integrated simulation that allows practice in both technical and non-technical skills. Not only does this increase the realism of the practice exercise but improves engagement, and likely improves the transferability of the skills to work with real patients. Other simulators combine ex-vivo organs harvested from animals that can be embedded in a plastic torso to simulate a human abdomen. A porcine liver was utilized in this fashion in a recent study simulating an open cholecystectomy procedure (30).

Surgical simulation appears, at first glance, to be ideal in providing a setting for practicing surgical motor skills until automaticity is obtained. Unfortunately, volume alone does not account for differences in skill. Achieving expertise in performing a surgical procedure is also highly dependent on 'deliberate practice (31),' which consists of focusing on a specific task to improve performance while being scrutinized by an instructor. This suggests that decomposition of a task, and perhaps revisiting the cognitive or intellectualization phases of learning a motor skill are crucial to improving performance and obtaining expertise. While there is evidence for improved clinical outcomes for surgeons with higher operative volume (32), there is still a significant amount of variation in performance amongst surgeons with high and very high volumes (6). This may be due to surgeons that are more cognizant of deliberate practice and work to continually hone their skills (31). This also suggests that improved performance metrics that are able to measure performance of the individual components of a given surgical task may provide further insight into improving current simulation-based training strategies. With respect to providing useful feedback to surgical trainees, understanding the critical components of an expert movement could identify additional targets for improving surgical dexterity.

A review of the current state of surgical simulation reveals significant evidence for simulation improving basic skills of trainees and transferability of these skills to the operating room (33-35). There is growing evidence for the ideal components of curriculum design that improve skill transfer to the clinical environment, some of which include training to expert-derived performance criteria (proficiency based training), deliberate practice, multiple training episodes and training to automaticity. However, there are still significant challenges related to optimal implementation of currently available simulation technology into surgical training programs. In addition, there are several flaws in current research related to demonstrating the validity of simulators that make comparisons of different technology difficult (33).

With respect to open surgical simulation, some promising new advanced haptic models have been developed that can be used to teach suturing, liver biopsies and lumbar punctures (36). Davies and colleagues go on to describe how these exercises are a step towards the development of more refined haptic simulations of open surgical procedures, and that a reproducible simulation with instant device-generated feedback is a reasonable expectation for the future of this training modality. While improved 3D technology and fidelity are possible, we are reminded of the Grober *el al.* study (37), which tested the impact of fidelity of simulator effectiveness and showed that low fidelity bench top models are just as effective as expensive high fidelity systems in teaching basic skills to novice surgical trainees. Going back to the specificity of practice hypothesis, a greater understanding of the components of a simulator that are most valuable in mimicking the movements and context of a particular skill will allow us to design smarter, more relevant teaching tools in surgery.

Chapter 3 – Evaluation Methods for Assessing Surgical Dexterity

Evaluation methods comprise an essential component of any education system. In order to develop and design appropriate curricula for teaching surgical skills, a reliable evaluation method is necessary to discriminate between trainees and identify areas for improvement. Prior to implementing any new training strategies or refine current simulation based surgical training tools, an assessment tool must be selected in order to identify who simulation would benefit most and if a current simulation curriculum is delivering its objectives effectively.

3.1 Paper based evaluation

Paper based grading is the principle evaluation technique currently employed in surgical training. The prototypical evaluation involves direct observation where a senior surgeon will scrutinize a trainee and provide feedback (38). As would be expected, these evaluation methods can be highly subjective. Some attending surgeons rely on a 'gut feeling' for in-training assessments that are usually made at the end of a trainee's rotation and are based on recollection of their observations (39). While these surgeons feel they are capable of judging surgical dexterity, these evaluations are made without any specific criteria. Shah and Darzi (39) also suggest that a significant bias could be introduced into these assessments by personal differences or a 'clash of personalities.' This method also has poor inter-observer agreement, as evidenced by studies demonstrating markedly different evaluations of the same trainee by two surgeons (40).

Standardization of the direct observation method is accomplished via the used of checklists and global rating measures. Examples of these methods include The Objective Structured Assessment of Technical Skills (OSATS), The Global Rating System (GRS) and error score card analysis. OSATS, developed by Martin *et al.* was introduced in 1995 (41). This assessment is comprised of an itemized checklist for each procedure with binary outcomes for correct execution of each step in the task, combined with a global rating score (42). The itemized checklist lists specific maneuvers deemed crucial to successful completion of the procedure, for example, Figure 3.1 lists the steps required for a small bowel anastomosis. The second portion of the OSATS assessment is the global rating form as seen in Figure 3.2. This scale is more generalizable to a variety of procedures and asks evaluators to rate trainees on a 5 point Likert scale in several categories including efficient movement, appropriate respect for tissue, and effective use of assistants, etc. Explicit descriptions of performance are provided for scores of 1, 3 and 5 in order to provide a reference for evaluation. Both components have been evaluated for construct validity and show increasing levels of competence as years of surgical experience or training increase (43). However, the simultaneous use of a checklist and global rating scale demonstrated that checklists were less reliable in gauging differences in performance. Consequently, the more objective component of this assessment method failed to provide a reliable performance metric.

STATION3

SMALL BOWEL ANASTOMOSIS

INSTRUCTIONS TO CANDIDATES

You have just resected a segment of small bowel. Perform a single layer, interrupted, end to end anastomosis to restore continuity

ITEM		Not Done or Incorrect	Done Correctly
1.	Bowel oriented mesenteric border to mesenteric border, no twisting	0	1
2.	Stay sutures held with hemostats	0	1
3.	Selects appropriate needle driver (Gen surg, medtip/med or short length)	0	1
4.	Selects appropriate suture (atraumatic, 3.0/4.0, PDS/Dexon/Vicryl/silk)	0	1
5	Needle loaded 1/2 to 2/3 from tip	0	1
6.	Index finger used to stabilize needle driver	0	1
7.	Needle enters bowel at right angles 80% of bites	0	1
8.	Single attempt at needle passage through bowel 90% of bites.	0	1
9.	Follow through on curve of needle on entrance on 80% of bites	0	1
10.	Follow through on curve of needle on exit on 80% of bites	0	1
11.	Forceps used on seromuscular layer of bowel only majority of time	0	1
12.	Minimal damage with forceps	0	1
13.	Uses forceps to handle needle	0	1
14.	Inverting sutures	0	1
15.	Suture spacing 3 to 5 mm	0	1
16.	Equal bites on each side 80% of bites	0	1
17.	Individual bites each side 90% of bites	0	1
18.	Square knots	0	1
19.	Minimum three throws on knots	0	1
20.	Suture cut to appropriate length (does not interfere with next stitch)	0	1
21.	No mucosal pouting	0	1
22	Apposition of bowel without excessive tension on sutures.	0	1
MAXIMUM TOTAL SCORE			(22)
TOTAL SCORE			

EXAMINER ____________

Figure 3.1 Itemized checklist from OSATS for a small bowel anastomosis from (42). Used with permission.

GLOBAL RATING SCALE OF OPERATIVE PERFORMANCE

Please circle the number corresponding to the candidate's performance in each category, **irrespective of training level**.

	1	2	3	4	5
Respect for Tissue:	Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments		Careful handling of tissue but occasionally caused inadvertent damage		Consistently handled tissues appropriately with minimal damage
Time and Motion:	Many unnecessary moves		Efficient time/motion but some unnecessary moves		Clear economy of movement and maximum efficiency
Instrument Handling:	Repeatedly makes tentative or awkward moves with instruments by inappropriate use of instruments		Competent use of instruments but occasionally appeared stiff or awkward		Fluid moves with instruments and no awkwardness
Knowledge of Instruments:	Frequently asked for wrong instrument or used inappropriate instrument		Knew names of most instruments and used appropriate instrument		Obviously familiar with the instruments and their names
Flow of Operation:	Frequently stopped operating and seemed unsure of next move		Demonstrated some forward planning with reasonable progression of procedure		Obviously planned course of operation with effortless flow from one move to the next
Use of Assistants:	Consistently placed assistants poorly or failed to use assistants		Appropriate use of assistants most of the time		Strategically used assistants to the best advantage at all times
Knowledge of Specific Procedure:	Deficient knowledge. Needed specific instruction at most steps		Knew all important steps of operation		Demonstrated familiarity with all aspects of operation

OVERALL ON THIS TASK, SHOULD THE CANDIDATE: FAIL PASS

Figure 3.2 Global rating form (OSATS) from (42). Used with permission.

More recently, Doyle *et al.* (44) developed a GRS designed for evaluation of trainees in the operating room. Their Global Rating Index for Technical Skills (GRITS) was based on OSATS and the Global Operative Assessment of Laparoscopic Skills (GOALS). It was specifically designed to be generalizable to a variety of procedures without modification as it did not include a checklist with binary outcomes for procedure specific steps. Again, this tool showed excellent reliability and construct validity with scores that increased with year of training. However, in addition to providing little insight into procedure specific deficiencies, the authors commented on difficulties controlling for bias introduced by a lack of blinding within the training program. In addition, junior surgical trainees, despite their lack of experience, were assigned scores of 3 or greater in most categories. The reluctance of evaluators to assign low scores on the GRS was thought to be due to an error of central tendency. A central tendency bias is the failure of an evaluator to differentiate performance between individuals by scoring everyone around the midpoint of the scale and avoiding the extremes, i.e. 1 or 5 (45).

Other global evaluation methods, those that do not evaluate surgical dexterity directly, such as outcome measures can be misleading. Post-operative complications, mortality and morbidity following surgery have been shown to correlate with a surgeon's skills, but this data can be significantly skewed by higher risk patients as demonstrated by Bridgewater *et al.* (46). For example, a less skillful surgeon could limit their practice to low risk cases resulting in lower mortality or morbidity compared to a more experienced colleague. Bridgewater and colleagues performed their analysis on patients in northwest England undergoing bypass graft surgery for the first time and observed an overall mortality rate of 1.7%. Given the low rate of mortality, it was likely that a small proportion of high risk patients were responsible for most of the differences in mortality. They recommended against using a crude comparison of mortality and posited that this would likely result in surgeons avoiding high risk patients or encourage risk averse behaviour. Additionally, audits of crude morbidity and mortality data that attribute the outcome of the patient to the operation or technical skill of a surgeon fail to account for other factors such as local facilities and availability of other specialized services or differences in the disease process between patients (39). Operative outcomes are also attributed to the attending surgeon and not the trainee, making this a difficult tool to apply to surgeons in training.

All of the aforementioned human-based assessment systems rely on the availability and willingness of an expert surgeon to score or grade trainees. In addition to bias introduced by varying requirements for subjectivity in each evaluation, this process is expensive and time consuming. These assessments are usually completed at the end or between procedures, limiting real time feedback and lacks specificity in identifying particular components of a technical skill that need improvement. For example, a low score on "respect for tissue" or comments that suggest that a trainee may be performing dissection too aggressively fail to provide timely suggestions for improvement or alternate strategies that might be employed. While some level of oversight will likely be required in more objective assessments, limiting the amount of human-based assessment will likely improve the efficiency as well as the reliability of these tools. The following section discusses the use of motion analysis which is substantially more objective than any human based observation method in providing a metric for performance.

3.2 Motion analysis based assessment

Motion analysis applied to the assessment of surgical dexterity is based on the dynamic system theory of motor skill development (47,48). This theory describes how movements made by novices become progressively and measurably more efficient as these individuals gain experience. With higher levels of experience, or more autonomous movement, results in more timely completion of a task and decreased complexity of movement. Motion analysis has been used in other fields for some time, particularly in physiotherapy and rehabilitation for gait analysis (49). Kinetic and kinematic data from these experiments can be used to optimize gait characteristics, improve prosthetic comfort and ambulation efficiency.

Capturing motion data can be challenging, especially in the operating room where equipment should not impede the performance of the operating team or surgeon. Motion data can be captured with a variety of technologies including optical, electromagnetic and force or mechanical systems. Chmarra and Dankelman (50) provide a good overview of the essential components that make up a tracking system. The first piece of hardware required is a *source* that generates a signal, which is in turn detected by a *sensor*. Active tracking systems describe those where the sensor is attached to the object that is being tracked. An example of this would be an electromagnetic sensor placed on the dorsum of each hand in the Imperial College Surgical Assessment Device (ICSAD) (51). Alternatively, passive tracking systems localise sources that have been placed on an object and are tracked in a given field. This can be advantageous as it does not require cables or wiring attached directly to an instrument. However, most MIS tracking systems are active in order to accommodate the need for line-of-sight, as hands or the operators body can obscure the sources for tracking. The next chapter includes a discussion of some of these techniques applied to both open and minimally invasive procedures, along with their advantages and pitfalls.

3.3 Methods for motion capture in open surgery

As mentioned previously, in addition to learning new MIS approaches to a given procedure, current surgical trainees are expected to develop proficiency in the traditional open approach. For many procedures in surgery, open procedures remain the standard of care (52). Open surgery presents arguably the most complex environment in which to analyze motion, as surgeons are able to freely manipulate a variety of tools in both hands, resulting in many degrees of freedom (DOF) for each hand or hand-instrument interaction. Gloves with embedded sensors can be worn by a surgeon in order to generate hand and joint position data as well as velocity data. Several commercially available gloves are available for this purpose including Cyberglove (Cyberglove Systems, San Jose, California), ShapeWrap II (Measurand, Fredericton, New Brunswick) and the 5DT Data Glove (Fifth Dimension Technologies, Irvine, California) (38). While some of these gloves can also be fitted with wireless data systems in order to minimize cables that might hinder natural movements, the gloves themselves are bulky and impair the user's sense of touch or haptic feedback.

Optical tracking systems have also been used to capture surgical movements. Besides the camera used to determine object or hand position, this system is completely unobtrusive and does not impede nature hand movement. In open surgery, Glarner *et al.* (53) conducted a feasibility study in the Department of Surgery at the University of Wisconsin, Madison, where they applied a digital video analysis system to video recorded in the operating room. The major advantage of this system was the utilization of conventional digital video to capture raw data. This technique does not require any additional equipment in the operative field such as electromagnetic trackers or wires connected to sensors. Following recording of the procedure, an analyst selects a region of interest (ROI) in the video. The template matching tracking algorithm developed by Radwin (54) then follows the specified ROI and can generate kinematic data including displacement, velocity, and acceleration. In the laboratory, this system produced similar results compared to manual frame-by-frame analysis of video data as well as infrared optical tracking systems (53) . The Wisconsin group applied this technology to reduction mammoplasties where a junior and more experienced surgeon operated simultaneously on both breasts. The video data was manually segmented using Multimedia Video Task

Analysis (MVTA) software developed by Dr. Radwin during representative surgical tasks such as cutting tissue with electro-cautery, cutting with a scalpel, suturing, and instrument tying. The hand(s) of a surgeon were then marked as a ROI. The raw hand position data obtained from analysis of the video were filtered using a low-pass filter to reduced noise. The kinematic data was then analyzed for each task and descriptive statistics were generated for both the experienced surgeon and trainee. Their analysis suggested that the experienced surgeon used their non-dominant hand more while cutting with a scalpel and when suturing when compared to residents. In addition, the attending surgeons had greater economy of movement as shown by decreased hand displacement during instrument tying. The authors posited that their data suggests that experienced surgeons were more adept at assisting themselves with their non-dominant hand. The limitations mentioned in this feasibility study include parallax error introduced into the two-dimensional video data due to the varying angles between the camera and each surgeons hand or each surgeon. In addition, no formal statistical analysis was performed to test if the differences observed between the residents and attending surgeons were significantly different.

One of the most widely recognised systems for obtaining motion data in surgery is the ICSAD (51). This system is comprised of two electromagnetic 10 mm sensors (Isotrack II, Polhemus Inc, Colchester, Vermont) placed on the dorsum of each hand and a Bespoke computer software program. These sensors, placed at the midshaft point of the third metacarpal, can record hand position in six degrees of freedom at a high rate (20Hz) which can be used to generate metrics such as number of movements, speed and velocity of movements and total distance traveled. The ICSAD was appraised using two standard simulated tasks including a small bowel anastomosis and a vein patch insertion at depth with restricted access. Both of these procedures are taught to surgical novices during the Royal College of Surgeons Basic Surgical Skills course, ensuring that all participants had previous experience. Each subject was given a single attempt at completing both tasks. Computer software translates the raw movement data obtained from the trackers into three scores of dexterity including number of movements of each hand, distance traveled by each hand and the time taken to complete the task. The number of movements were determined by counting each change in velocity. Final analysis combined the number of movements and path length of each hand. To account for a large skew in the data that could not be corrected with a logarithmic transformation, nonparametric statistics were used in all subsequent analysis. As anticipated, increasing level of experience correlated well with a decreasing number of movements and time during the small bowel anastomosis task (Spearman's coefficient of correlation 0.839, unadjusted $p < 0.001$). While this was shown to be statistically significant, when compared individually the time taken to complete the task only became significantly different between basic surgical trainees (BST) and senior trainees (SSpR) (unadjusted $p<0.001$). However, when the analysis was adjusted for number of movements, the correlation between time taken and surgical experience was no longer significant. Interestingly, when controlling for time, there was still a significant correlation between number of movements and experience level (Spearman's coefficient of correlation -0.449, unadjusted $p=0.002$). In the second task, vein patch insertion, a decreasing trend was observed for both number of movements and time with increasing experience level. Statistical analysis revealed a significant difference in both number of movements and time taken between extremes of experience (BST vs SSpR, unadjusted $p < 0.001$). Correlation coefficients computed for time and movement data with experience was identical to the small bowel anastomosis task, revealing a significant relationship that did not persist when controlling for number of movements, but was significant when controlling for time. None of the path length data could be used to discriminate between experience levels as it was not statistically different (overall $p=0.657$).

The ICSAD was further validated to show that it had good concordance with OSAT scores (52). While the ICSAD is certainly more objective, it provides limited information beyond overall performance to discriminate between different experience levels. In addition, it lacks any ability to identify which components or subtasks of a certain task require additional refinement or practice.

Path length efficiency has been used reliably to discriminate between surgeons of different skill levels in studies that apply motion analysis to laparoscopic procedures. Data obtained from the ICSAD however did not demonstrate a significant difference in this metric with experience level. Datta *et al.* suggest that this may be a consequence of impaired depth perception in laparoscopy and novices are thus forced to make more movements or adjustments to correctly position their instruments during a procedure (51). In addition, they observed a variety of techniques employed during knot tying that would have resulted in variable path lengths while completing the same task amongst surgeons of a similar skill level.

3.4 Methods for motion capture in MIS

Laparoscopic or MIS is touted as the future of surgery. With the advent of Natural Orifice Transluminal Endoscopic Surgery (NOTES) (55), many believe MIS will become the standard of care and supplant open surgical techniques. There is significant evidence for decreased morbidity and shortened recovery time following MIS procedures when compared to their open counterparts for a variety of procedures (56). The practicality of measuring motion data in MIS is made less challenging by the limitations imposed by the technique (57). The small incisions and ports through which instruments are inserted into a body cavity restrict the range of movement to 4 degrees of freedom. In comparison, the human hand has roughly 30 degrees of freedom when wrist and digit motion are combined.(47) For example, the wrist can move linearly along the three directions defined by a 3 axis Cartesian coordinate system in addition to combinations that can be described by yaw, pitch and roll, totalling 6 DOF. Tracking systems and motion analysis technology for MIS based surgery is consequently more advanced and developing more rapidly than technology used in open techniques, likely due to the reduced complexity and relative ease of obtaining motion data for MIS based techniques.

A variety of commercial and research based tracking systems have been developed or applied to laparoscopic procedures. Many of these systems are based on a gimbal mechanism, a set of two or three rings mounted on axes orthogonal to each other (50). The rings provide a stable reference point for movements in three dimensions which is usually taken as the pivot point for active tracking systems. The precision and accuracy of motion data is crucial in MIS as small movements close to the pivot point can result in large movements of the instrument tip, requiring some systems to be recalibrated frequently (50). Chmarra emphasises that there are no established standards for the precision or accuracy of MIS tracking systems, so it is difficult to determine if current methods are accurate enough. Unfortunately, given the bulkiness and active tracking design of these systems, none are appropriate for use in the operating room. Simulation is therefore used with ex-vivo tissue, models or computer generated images to provide tasks in which motion analysis can be applied. Real surgical instruments can be used in some of the systems which accurately reproduce instrument and tissue feel or haptic feedback during a procedure. For virtual reality based simulation some instruments allow for acquisition of the forces applied to an instrument and provide force feedback. Chmarra and colleagues again point out that at the time of their review, feedback systems for virtual reality based simulators needs considerable more work to mimic real word feedback, and that this should be further investigated.

3.5 Motion Analysis Using Descriptive Statistics

Dexterity analysis using descriptive statistics uses motion or force data obtained from tracking systems, such as those described above. Performance metrics can then be generated based on a number of factors. Some of these include economy of motion, repeated motion, velocity of movement, instrument path following, peak forces and tissue damage. The majority of these systems also record temporal metrics such as task completion time and time spent performing each task. With more practice, task completion time decreases as expected but economy of motion decreases much more gradually (58). An inherent weakness of any temporal based metric, however, is that they require the trainee to successfully complete each task. Feedback is therefore only provided at the end of a given procedure, limiting any online or real-time feedback.

An example of a current assessment tool in laparoscopy that utilizes descriptive statistics is the McGill Inanimate System for Training and Evaluation of Laparoscopic Skills (MISTELS) (59). This simulation-based system consists of 5 exercises that all completed in an laparoscopic box trainer and includes a peg transfer, precision cutting, ligation and intra-corporeal suturing task. Each task is scored based on task completion time and precision incurring a time penalty. Reference scores were obtained by taking an average of the performance by a group of chief residents and fellows. The MISTELS system was shown to have both construct and external validity (59). This system was so popular as a new metric of performance that it was integrated into the Fundamentals of Laparoscopic Surgery (FLS) training course administered by the Society of Gastrointestinal and Endoscopic Surgeons (SAGES) (60). However, the FLS assessment process is proctored by an experienced surgeon and requires participants to travel to a limited number of assessment sites in North America for testing. This has consequently limited its widespread use.

Descriptive statistics have also been applied to simulators that utilize computer generated images. Tracking systems are already integrated into these simulators in order to capture user or trainee input. They can also provide real-time feedback more easily by displaying ideal instrument positioning or maneuvers in order to complete a task such as suturing (61). These systems are also more flexible in providing a higher fidelity simulation in order to increase realism. Despite their potential, this technology has yet to mature. Reiley *et al.* reviewed currently available simulators and concluded that they lack overall effectiveness and realism (38). They also reported that these high-fidelity simulators generate inconsistent performance metrics and are unable to differentiate novice trainees from experts. In addition, performance on these simulators showed equivocal correlation with operating room performance based on conflicting studies (62,63).

Overall, given the validity of some of the systems using descriptive statistics to generate performance metrics, this motion analysis method shows promise for the development of more objective global measures of performance. However, these methods have limited feasibility for real-time feedback, especially if temporal metrics are used. A more advanced model is necessary in order to provide more useful and real-time feedback to the trainee.

3.6 Motion Analysis Using Language Models

Arguably the most sophisticated method currently employed for analysing motion data are language models. This method allows for the decomposition of each task into discrete movements that can then be compared between individuals. In order to understand how language models can be applied to surgical movement, it is useful to establish a hierarchy of movements specific to a specific task and procedure. Reiley *et al.* delineated this in Figure 3.3. Each procedure can be subdivided into a series of sequential tasks. For example, during a cholecystectomy, or removal of the gallbladder, two sequential tasks would include dissection of the cystic duct and applying clips to the duct. These tasks can be divided further into well-defined surgical motion units or *surgemes*. An example of a surgeme is the insertion and spreading of a dissecting instrument or the individual hand maneuvers during suture tying.

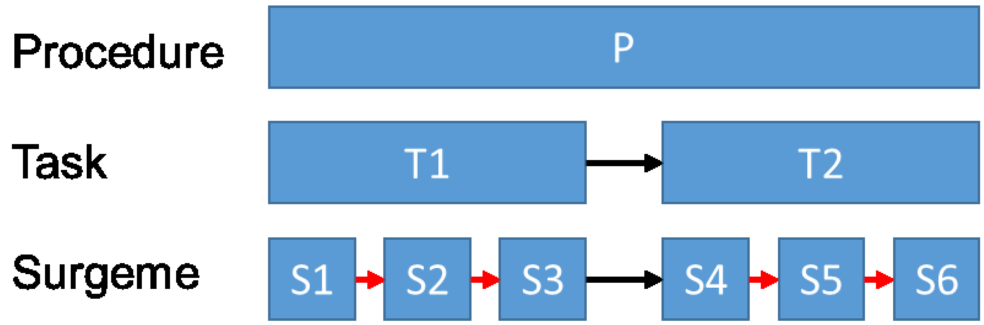

Figure 3.3 Hierarchical decomposition of surgical tasks into three levels from (38). Used with permission.

Surgical dexterity can be assessed at each of these levels using language models. In the first level, the procedure level, Ahmadi *et al.* (38,64) developed an automated analysis of workflow in the operating room using input from 17 sensors. This group was able to successfully model the workflow for a cholecystectomy and used instrument changes to help segment motion data to reliably predict the timing of the next component of the procedure to within 5 seconds. Their method was able to identify, with 92% accuracy, the 14 phases of the operation that their algorithm identified. Analysis of motion data at this level can be used to identify areas for improvement in the workflow for a given surgical team. However, this method lacks the ability to discriminate between surgeons of different skills levels.

Prior to discussing the further decomposition of a surgical procedure into tasks and sub-tasks or surgemes, the concept of statistical modelling using Markov Models (MM) and Hidden Markov Models (HMM) must be introduced. These have been used in a variety of fields including DNA sequence modelling, speech recognition, analysis of facial expression, and human operator modeling for the purpose of transferring human skill to robots (65). An analogy between spoken language and surgery can be used to describe how MM can be used to objectively assess skill level (66). Language and surgery movements share a similar taxonomy and internal etymological structure that allows for the application of mathematical and quantitative models. Skill can thus be assessed by revealing the internal structure of language or movements. The extension of this analogy is that in surgery and language, a procedure can be performed and an idea can be expressed using language in several different ways while retaining the same outcome and meaning respectively (64). In contrast to deterministic models which can exploit a known property of a data source (e.g. a sine wave can be described by amplitude, frequency and phase), these models are based on statistical properties of the data. This requires the assumption that data can be characterised as a parametric random process and that certain parameters for this stochastic process can be computed in a precise manner (67). Rabiner uses the example of a system with 5 distinct states, S_1, S_2, …, S5 (67).

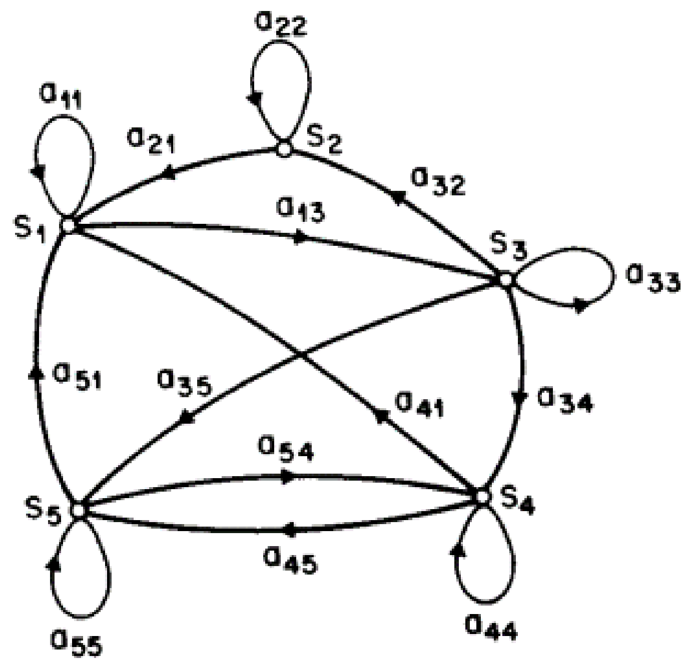

Figure 3.4 A Markov Model (MM) with 5 states with selected state transitions from (67). © 1989 IEEE.

At regular discrete time intervals, the system can change state according to a set of probabilities associated with a given state. The possible state transitions for these 5 example states is depicted in Figure 3.4. A full probabilistic description of each state would require knowledge of all predecessor states, but for a first order MM, this can be simplified to a single predecessor state. These elements compose an observable MM since each state, occurring a specific instant in time corresponds to an observable event. Hidden Markov Models (HMM) are an extension of MM in which the sates are assumed to be hidden or not observable, but observations can be made that depend on the state. This in turn can be used to determine the sequence of states. The advantage of HMM applied to surgical motion data is the ability to automatically segment the data in order to decompose procedures into tasks and surgemes as opposed to manually segmenting and labeling this data.

At the task level, a labour intensive method for analysing a procedure involves manually segmenting and annotating video of a given procedure. The components of the procedure can then be manually compared in order to determine the frequency of each task. For example, MacKenzie *et al.* (68) used this approach to analyse laparoscopic procedures. Video of the procedure as well as the operating room were captured and analysed over 2 years in order to describe the precise timing and location of operating room events, surgical events and eye-hand-tool-tissue interactions. Decomposition of each procedure was accomplished by iterative top-down analyses of surgical procedures and bottom-up analyses of tool motions. The components of each procedure were either serial or parallel. Serial components occurred in a repeatable sequence e.g. isolate gall bladder then remove gall bladder, while parallel components included tasks such as poke and tease tissue. The authors chose 4 tasks that were regularly performed during the procedure including dissection, suturing, anchoring sutures and knotting and cutting sutures. Analysis revealed that knotting a suture had the longest duration of any task, followed by anchor suturing, normal suturing and cutting the suture. This revealed the portion of the procedure that could be targeted to improve efficiency. This form of analysis suggests that trainees should emphasize practice tying intracorporeal sutures in order to improve their overall performance in this procedure. The hierarchical decomposition method, while time consuming and laborious can be applied to measurement or assessment at various levels within the hierarchy. The introduction of a new technique or technology or experience level of a surgeon will manifest in a measurable change in the dynamics of the system. An extension of the task level approach is to study the human machine interface (HMI) which may be further subdivided to model hand and tool interactions as well as tool and tissue interactions. Rosen *et al.* (69) used a force and torque sensor mounted on a laparoscopic grasper as well as video of the tip interacting with the tissue in order to construct a collection of states representing a specific maneuver in the procedure. They recorded five novice surgeons and five expert surgeons performing a laparoscopic cholecystectomy and Nissen fundoplication on a porcine model. Manual frame by frame video analysis was combined with a vector quantization algorithm to define the force and torque signatures associated with 14 different tool/tissue interactions. They found that novices had higher magnitudes of force in tissue manipulation tasks and less in tissue dissection compared to experts. Novices also took more time to complete the procedure and spent more time in an idle state. MM were created for novices and experts and a statistical comparison was used to compare an observed surgeon's MM to the MM for each group. A performance index was created by taking the ratio of statistical similarity of a given surgeon's performance compared to the reference MM. This technique was able to correctly classify 87.5% of surgeons as novices or experts. While this method was able to model the complexity of laparoscopic movements using MM and make comparisons, analysis of the raw data still required the researchers to manually segment transitions between different hand-tool interactions using video of the procedures. This again is very laborious and prevents this technique from providing real-time feedback to trainees.

The final level in which to analyse surgical motion is the subtask or surgeme level. Breaking down surgical motion into modular and reusable motion segments should allow for a method to differentiate novices from experts and provide some insight into how to correct or improve performance. As an extension to their decomposition analysis MacKenzie and colleagues were able to further break down each task into 5 basic motions including reach and orient, grasp and hold/cut, push, pull and release (68). All of these were described by their motion along or about Euclidean space. While these motions were not timed, the frequency of each motion were recorded and when compared in context of their subtask or task provided a quantitative measure for the complexity of a given subtask or task. Greater task complexity resulted in a greater number of tool motions. The authors of this study included a discussion of the many applications this technology could have in improving workflow and the ability to test how a new technology or technique could influence the overall performance of a procedure. At this level of analysis however, they emphasize that this analysis is limited to detecting differences in the functional level of a task and therefore cannot make inferences about the cognitive ability of the surgical team.

The application of MM and HMM to surgical surgemes provides a method for objectively assessing performance as well as providing real-time feedback on the constituent elements responsible for generating a given performance metric. Rosen *et al.* (66) applied MM to modeling MIS procedures using their own Blue DRAGON system, a proprietary system used to acquire kinematic (rotation, translation) and dynamic (force, torque) data from two endoscopic instruments. They recorded motion analysis data for 30 surgeons of varying skill levels, 5 in each year of residency training (R1, R2, …, R5) and 5 expert surgeons who had performed more than 800 laparoscopic procedures. Each participant was required to complete and intra-corporeal knot in a porcine simulation model in addition to 15 predefined tool/tissue and tool/needle-suture interactions. A MM was generated for each surgeon and a learning curve was constructed based on measuring the statistical similarity between trainee and expert MM at each level of training. This was subsequently compared to a subjective assessment similar to the previously described global rating scale, which revealed a significant correlation of 0.86 ($p<0.05$). The MM generated for each participant also provided information regarding the appropriate use of states and state transitions representing different tool/tissue interactions. Time domain analysis could then be used to determine how much time was spent in each state and transition, revealing a similar result found in previous studies regarding the additional time more novice surgeons spend in the idle state. Additional information regarding a quantitative measure of the level of collaboration between the tools provided an additional estimate of skill level, where an experienced surgeon's non-dominant hand will be more active. These models could be run in real-time in order to provide constructive feedback as the procedure was performed, but this was not tested in this study (66).

Chapter 4 – Can Video Gaming Substitute Traditional Simulated Practice for Laparoscopic Novices?

Previous chapters highlighted several theories describing complex motor skill acquisition and recognition. We also discuss the role of simulation for improving surgical skill. This chapter describes a randomized control study that was conducted to investigate the use of video gaming as a training modality for improving spatial perception and ability in simulated laparoscopic procedures.

Introduction

Minimally Invasive Surgery (MIS) procedures demand an exceptional level of coordinated psychomotor activity (70,71). Decomposing the components of even the simplest maneuvers reveals a complex path involving hand-eye coordination, hand-instrument coordination and extrapolation of a 2-dimensional video image into a 3-dimensional working environment. Consequently, hand-eye coordination in MIS procedures is inherently more complicated than that in open procedures (72,73). This makes skill translation from open procedures difficult, and has given rise to an entire industry devoted to simulation based laparoscopic training.

A variety of laparoscopic simulators are currently available ranging from bench-top laparoscopic box trainers to Virtual Reality (VR) simulators that provide haptic feedback (74). However, despite the increasing availability of laparoscopic simulators, their integration into a surgical training curriculum is highly variable among different programs (75). Without protected time, surgical residents in some programs report low utilization due to limited availability or access to simulators outside of regular work hours (76).

In comparison, video gaming is becoming more prevalent and accessible. Current industry data indicates that four out of five American households own a video game device (77). In addition to inducing structural brain plasticity (78) and improving psychomotor ability (79), training based on video gaming is purported to have a greater motivation and acceptance rates among individuals compared to other training modalities. Video gaming mimics much of the constant integration of video images and character or object position with instrument and controller movement. The relationship between video game playing experience and laparoscopic surgical performance was first demonstrated by MIS educators as early as 2007 (80).

More recent studies have attempted to answer how video gaming can improve visuospatial and psychomotor ability in MIS novices. A best evidence topic concluded that there is a positive correlation between video game experience and laparoscopic performance in terms of fewer errors and decreased time to complete a simulated task (81). Further, Schlikum *et. al.* (82) demonstrated in a randomized control trial that systematic exposure to visuospatial intensive gaming can improve performance in simulated laparoscopic tasks. However, there are no previous randomized control trials evaluating the combination of video gaming with traditional practice utilizing a laparoscopic box trainer. In addition, many of the previous studies in this field have been limited by a small sample size.

This study was designed to investigate the combination of visuospatial intensive video gaming with traditional lap box training based on previous evidence of a positive effect of systematic training with the Nintendo Wii (83). We hypothesized that replacing a portion of lap box training with video gaming would result in a similar learning effect and similar performance on a laparoscopic surgical simulator. Undergraduate students were selected as participants to limit any bias introduced by previous surgical or laparoscopic experience and allow for a larger sample size. We chose FLS task performance as our primary outcome measure after exposing four randomized cohorts to a variable amount of video gaming, laparoscopic box training, or a combination of both over a course of six training sessions. In addition, we recorded the best performance of a laparoscopic practice task in each training session to evaluate the cumulative effect of the training program using learning curves.

Methods

Ethics approval was obtained from the University of Alberta Health Research Ethics Board. All participants were recruited from the University of Alberta. Recruitment targets were 20 for each intervention group and 10 for the control group. Blinded block randomization was performed using a randomization scheme generated online using a randomly permuted block method and random block sizes (84). Labels generated for each training group were placed in sealed opaque envelopes based on the randomized list by an individual not involved in recruitment or participant allocation. An overview of the study design and group allocation is delineated in Figure 1. Sixty-four participants completed the training program and final assessment.

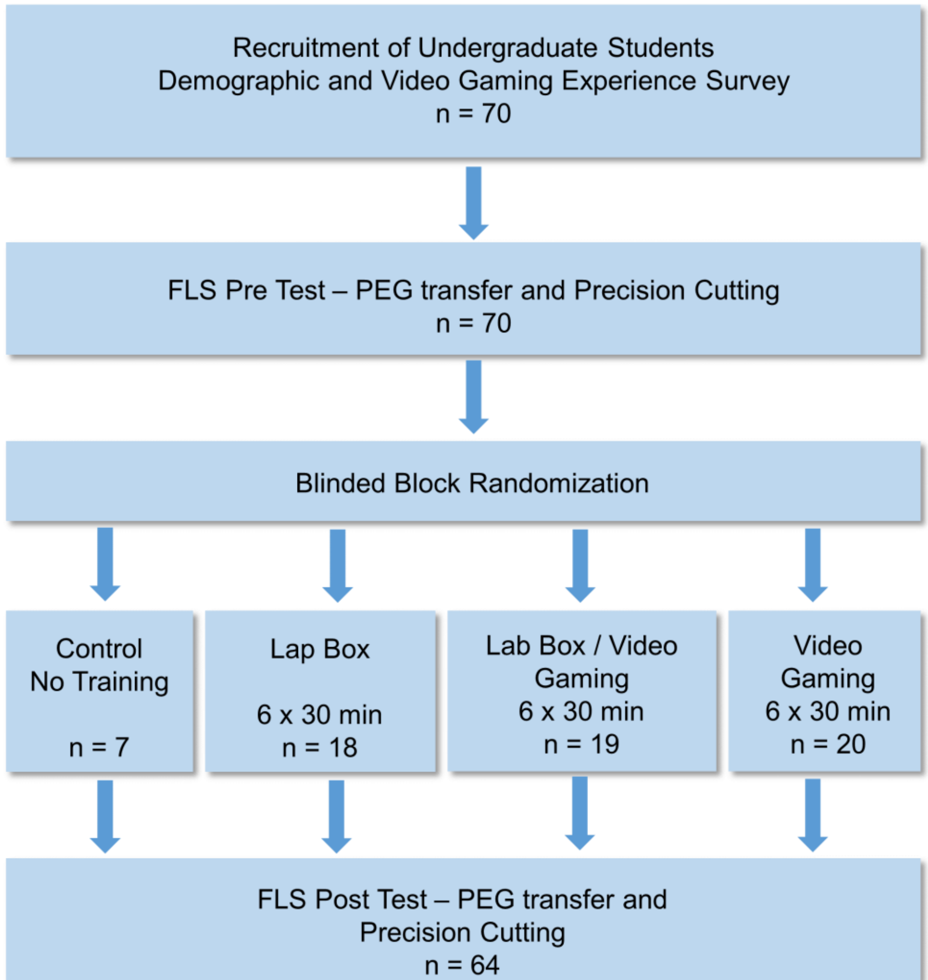

Figure 4.1 Study design consisting of four study arms and the sequence of pre-test, training and post-test evaluation.

Prior to randomization, each participant completed a demographic summary consisting of sex, age, and handedness. Subjects also confirmed whether they met any exclusion criteria, including any prior experience with laparoscopic practice or the video game used in this study. All subjects were required to have normal or corrected to normal vision. Participants reported their previous video game experience by completing a summary of their video game activity during five different age ranges on a seven-point

Likert scale in a similar fashion as Schlickum *et al.* (82). However, we did not match or allocate participants based on previous video gaming experience.

A narrated video was then used to provide standardized instructions regarding the requirements of the two FLS tasks (Figure 4.2a,b). These were adapted from portions of the official FLS course instructional videos for the peg transfer and precision cutting tasks (85). After 5 minutes of practice, each participant completed a single trial of both of these tasks. Both FLS tasks were completed using a Stryker Endoscopy Tower (Stryker Instruments, Kalamazoo, MI) equipped with a High Definition camera and display. We used a 10 mm 30 º laparoscope inserted into a modified lap box to create our assessment simulator. Each participant's total performance time was recorded using a stopwatch as well as number of penalties (e.g. dropping a sleeve).

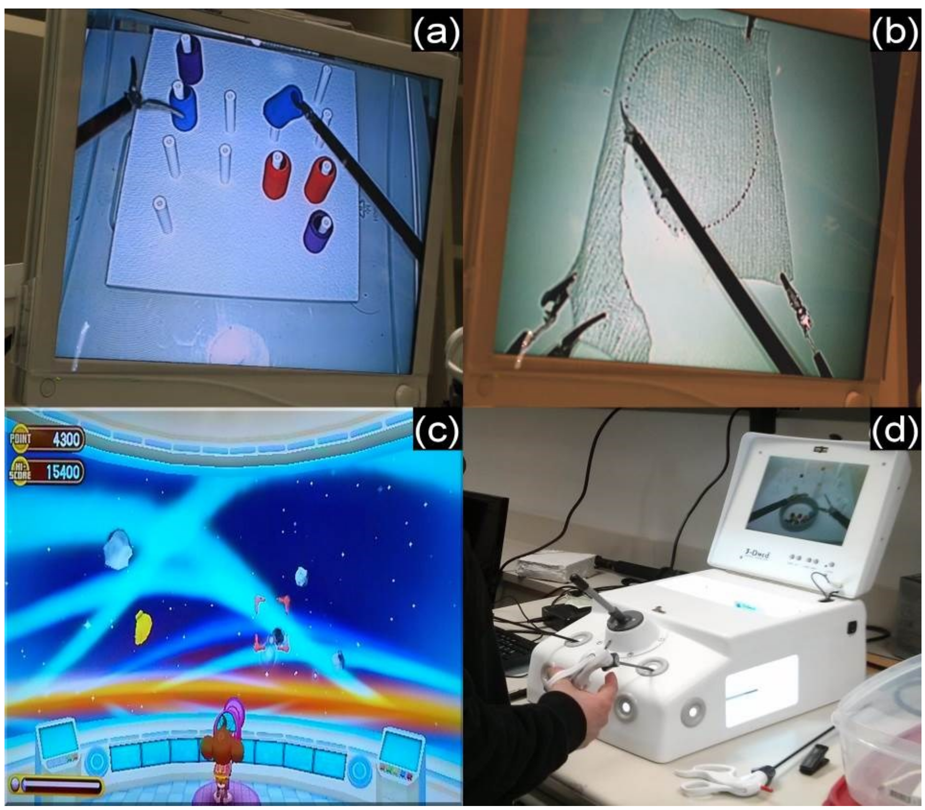

Figure 4.2 FLS tasks displayed on high definition monitor including (a) Peg transfer and (b) Precision cutting. Training tasks including (c) Super Monkey Ball 2 video game and (d) Pea on Peg laparoscopic box exercise.

Each participant was then asked to view a second standardized narrated video outlining the procedures and tasks required for each of the training arms. The first thirty-minute training session was then completed and an additional five sessions were booked using an online scheduling service. All training sessions had to be completed within 2 weeks and each session had to be at least one day apart.

In the control group, no additional practice or training was permitted. These participants were asked to return to complete a second FLS assessment within 2 weeks.
In the lap box training group, subjects were given 30 minutes to practice with a selection of (3-Dmed, Franklin, OH) laparoscopic skill building tasks including 'Pea on a Peg', 'Loop and Wire,' and 'Wire Chaser.' These practice tasks have previously been validated and are commercially available (86). Each participant was asked to complete at least two 'Pea on a Peg' tasks each session in order to monitor performance improvement (Figure 4.2d). The best 'Pea on a Peg' score from each thirty-minute session was recorded along with any penalties (e.g. dropped peas).

The video gaming cohort was exposed to thirty minutes of Super Monkey Ball 2 for Nintendo Wii (Nintendo Co Ltd, Tokyo, Japan). Participants were asked to complete at least two attempts of the 'Asteroid Crash' and 'Dangerous Route' mini-games each session (Figure 4.2c). They were also free to complete the 'adventure mode' or additional attempts of the same mini-games.

The fourth randomized cohort was exposed to a training program consisting of both video gaming and laparoscopic box training for fifteen minutes each. The best mini-game and 'Pea on Peg' score for each session were recorded.

A mixed repeated measures analysis of variance (ANOVA) was conducted to determine if there were statistically significant differences in FLS task performance prior to and following the standardized training program. Post hoc analysis with a Bonferroni adjustment was used to compare performance outcomes in the two FLS tasks utilized in this study across the four training groups. A second two way repeated measures ANOVA was performed to analyze the learning curves or performance improvement over time of the laparoscopic training tasks.

Results

Demographic Comparison

Sixty-four University students were recruited. Six students only attended the initial session and were not included in subsequent analysis. Statistical comparison of each of the training and control groups showed that our sample was homogenous with respect to all of the characteristics that we compared including previous gaming experience (Table 4.1). Gaming experience was generated from the self-reported gaming frequency of participants obtained during enrollment. For example, each of the different gaming frequencies on the Likert-type scale were assigned a score from 0 to 6 and the cumulative sum of gaming frequency in each age segment was totaled. For example, a participant who reported playing video games every day from childhood to present day would have the maximum gaming experience score of 24.

Table 4.1: Demographic information for randomized participants in each training cohort.

Characteristic	Training Group				
	Control	Lap Box	Video Gaming	Lap Box + Video Gaming	p value
Age	n (%)	n (%)	n (%)	n (%)	
15-19	1 (14)	2 (11)	2 (10)	2 (11)	0.74
20-24	6 (86)	14 (78)	16 (80)	12 (63)	
25-30	0	2 (11)	2 (10)	5 (26)	
Sex					
Male	3 (43)	8 (44)	12 (60)	12 (63)	0.68
Female	4 (57)	10 (56)	8 (40)	7 (37)	
Handedness					
Left	0	1 (6)	2 (10)	1 (5)	0.80
Right	7 (100)	17 (94)	18 (90)	18 (95)	
Gaming Experience					
0-10	6 (86)	10 (56)	10 (50)	11 (58)	0.19
11-19	0	5 (28)	9 (45)	8 (42)	
20+	1 (14)	3 (17)	1 (5)	0	

FLS Task Performance

There were no outliers in the data, as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box and no studentized residuals greater than 3.

Peg transfer and precision cutting task performance time was normally distributed for all interventions at all time points, as assessed by Shapiro-Wilk's test ($p > 0.05$). There was homogeneity of variances and covariances, as assessed by Levene's test of homogeneity of varience ($p > 0.05$) and Box's test of equality of covariance matrices ($p = 0.67$), respectively.

There was a statistically significant interaction between the training groups and time on performance of the peg transfer FLS task, $F = 14.29$, $p < 0.001$, and the precision cutting task performance, $F = 14.91$, $p < 0.001$.

Pre- and post-training performance are depicted in Figure 4.3 and Figure 4.4. The mean task time and confidence intervals are reported in Table 4.2 along with p values for comparison of each cohort to the control group.

Performance of the initial peg transfer task was not statistically different between any of the randomized groups prior to starting the training program, $F = 0.063$, $p = 0.98$. Comparison of peg transfer performance following training showed a statistically significant improvement in task time in the lap box ($M = -135$ s, $SE = 15$ s, $p < 0.001$) and combined training group ($M = -76$ s, $SE = 15$ s, $p < 0.001$) compared to the control group. However, there was no difference in the final peg transfer performance in the video gaming cohort compared to the control group ($M = -15$ s, SE 15 s, $p = 0.75$).

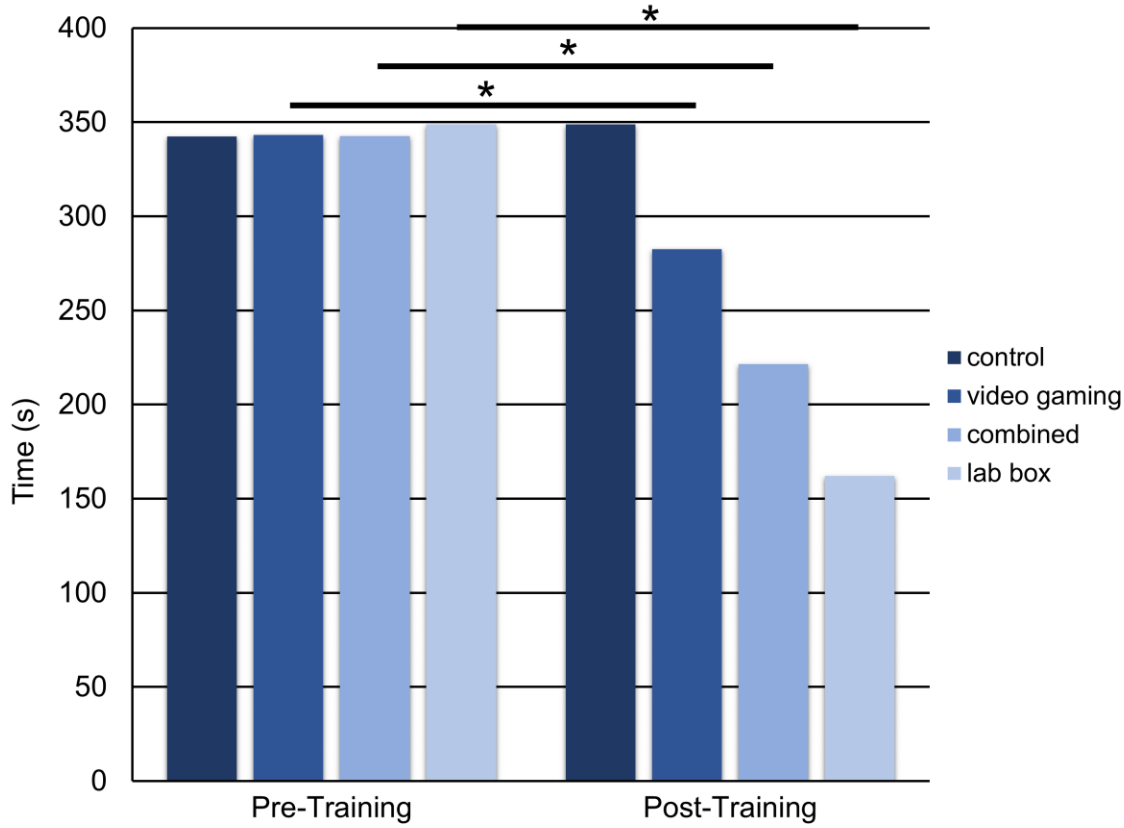

Figure 4.3 Peg transfer task performance pre- and post-training by training group.
* denotes statistical significance ($p < 0.05$).

Performance of the initial precision cutting task was statistically different between the randomized groups prior to starting the training program, F = 3.306, p = 0.03. Comparison of precision cutting performance following training showed a statistically significant improvement in task time in the lap box (M = -172 s, SE = 26 s, $p < 0.001$), combined training group (M = -144 s, SE = 19 s, $p < 0.001$), and video gaming cohort (M = -103 s, SE 18.8 s, p = 0.003). compared to the control group.

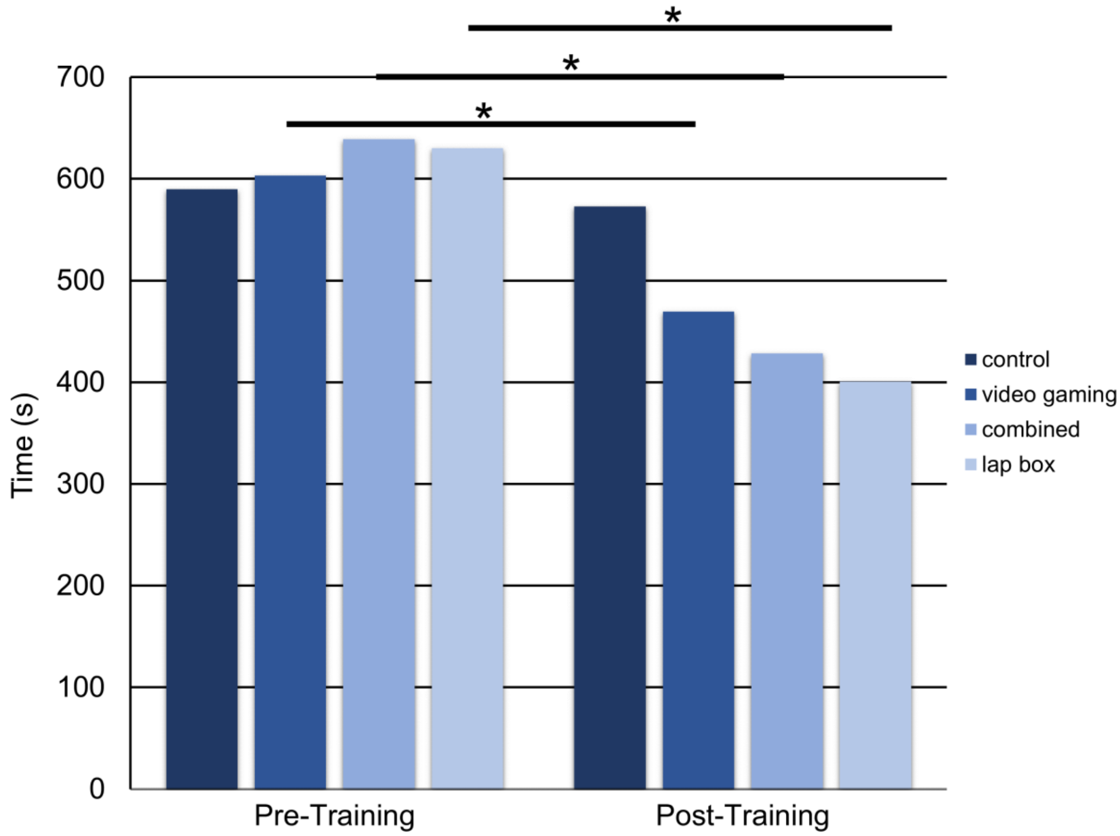

Figure 4.4 Precision cutting task performance pre- and post-training by training group. * denotes statistical significance ($p < 0.05$).

Table 4.2: Post-Training FLS Task Performance

Cohort	Peg Transfer			Precision Cutting		
	Mean ± SE (s)	95% CI (s)	p value	Mean ± SE (s)	95% CI (s)	p value
Video Gaming	282.6 ± 7.5	267.5 – 297.6	0.75	469.6 ± 12.9	443.7 – 495.5	0.003
Combined	221.4 ± 7.7	206.0 – 236.9	< 0.001	428.3 ± 13.3	401.8 – 454.9	< 0.001
Lab Box	162 ± 7.9	146.1 – 177.9	< 0.001	400.7 ± 13.6	373.4 – 428.0	< 0.001
Control	297.4 ± 12.7	271.9 – 322.9		572.7 ± 21.9	528.9 – 616.5	

Pea on Peg (POP) Performance

We compared the best POP time achieved for each subject at each training session in both the lab box exclusive and combined training groups.

There was one extreme outlier in the data which had a studentized residual value of >3 in several instances. This participant's performance actually worsened over the first 3 training sessions. As a result, this participant's data was removed from subsequent analysis.

Inspection of the pea on peg performance data showed moderate positive skewness for each time point. A square root transformation was thus applied to normalize the data prior to subsequent analysis. The learning curves for POP performance are delineated in Figure 4.5.

Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, so the Greenhouse-Geisser correction was used, $\chi^2(2) = 0.035$, $p < 0.001$. There was no statistically significant two-way interaction between treatment and time, $F = 2.357$, $p = 0.736$. This was also apparent in the plot of POP time over training session for both the lap box exclusive and combination cohorts (Figure 5).

The main effect of laparoscopy practice time per session, that is 30 minutes versus 15 minutes, showed a statistically significant difference in the best pea on peg time between both cohorts, $F= 4.739$, $p = 0.045$, with a mean difference of -2.44 (95% CI, -4.82 to -0.64) $s^{1/2}$.

The main effect of practice over multiple sessions showed that there was also a statistically significant difference in task time between each practice session, $F= 65.641$, $p < 0.001$. Multiple pairwise comparisons demonstrated a mean difference in time between each session ranging from -5.1 to -11.2 $s^{1/2}$. The improvement was statistically significant between each training session ($p < 0.05$) except for performance between session 2 and 3 ($p = 0.078$) as well as session 3 and 4 ($p = 0.084$).

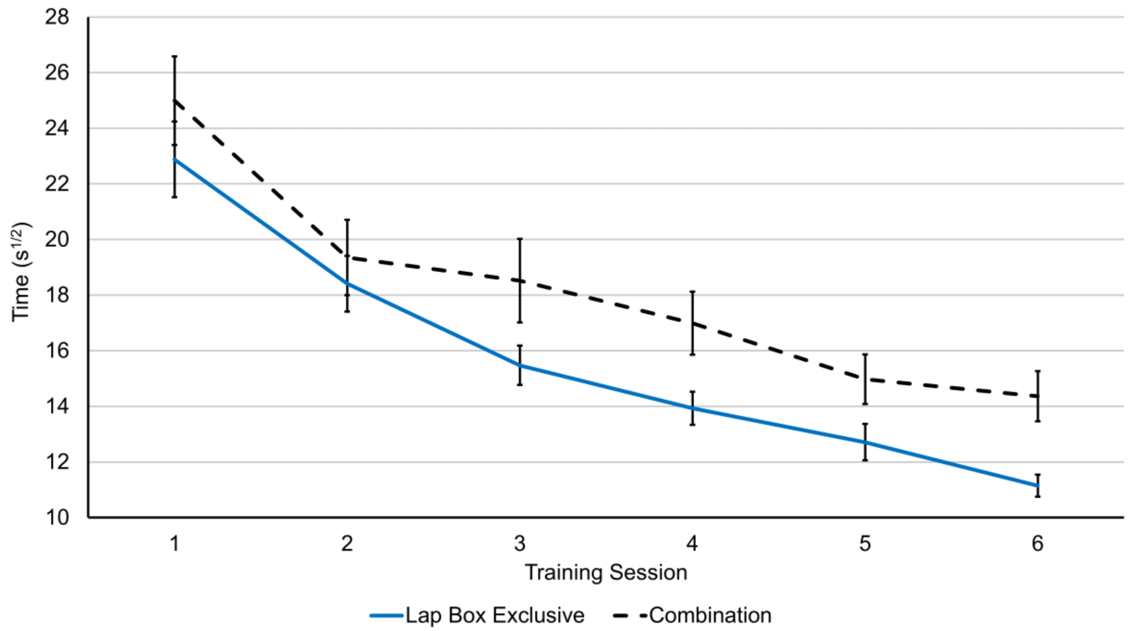

Figure 4.5 Pea on Peg learning curves for laparoscopic exclusive and combined training groups.

By the sixth training session participants in the combination group (M = 14.36 $s^{1/2}$, SE = 0.9 $s^{1/2}$) required 28.8% more time compared to individuals in the laparoscopic exclusive group (M = 11.15 $s^{1/2}$, SE = 0.4 $s^{1/2}$). However, the parallel nature of these learning curves, lack of an interaction and marginal statistical difference ($p = 0.045$) suggest a similar rate of skill acquisition.

Discussion

This controlled laboratory study demonstrated that lap box training was superior to video gaming for improving simulated laparoscopic performance when both were integrated into a structured training program. As in previous studies, we were able to show a positive correlation between video game practice and performance on a laparoscopic simulator (81,82). However, an equivalent amount of time spent video gaming did not result in the same performance improvement as practicing exclusively with a lap box trainer. If both were made available to trainees, our data suggests that practice with a lap box should be prioritized.

Interestingly, in the combined cohort where subjects spent 50% of their time practicing with the lab box and 50% gaming, the learning curves (Figure 5) suggested a similar rate of skill improvement. By the sixth training session, individuals in the lap box exclusive group had double the amount of practice (3 hours vs 1.5 hours) but only outperformed the combined training group by 28.8%. This suggests that video gaming may have contributed to the performance improvement in these individuals. Alternatively, individuals in the lap box group might have been approaching a plateau in performance, where more practice would only result in modest performance improvement. Given that the subjects we recruited in this study were complete novices, as well as significant higher POP scores for some individuals, we do not feel that performance had plateaued over the course of the training program. A lengthier training program would have to be completed in order to confirm this.

Comparison of the magnitude of improvement in the two outcome measures, Peg Transfer and Precision Cutting, suggests that lap box practice and video gaming are more effective for improving the performance of tasks that test spatial perception and bimanual coordination. Despite participants in the control and video gaming cohort demonstrating better initial precision cutting performance, both the lap box and combined cohorts were able to overcome this advantage. While we cannot explain why the control and video gaming cohort exhibited superior initial performance, the effectiveness of lap box practice is certainly demonstrated here.

The video game we selected, Super Monkey Ball 2, is visuospatial intensive and has previously been tested for its effect on laparoscopic performance (83). In addition, the controller for the Wii platform is unique and requires the subject to move their whole hand and controller over a range of six degrees of freedom. This more closely approximates the range of movement in laparoscopy compared to other video game controllers that employ a joystick. Some investigators have further modified the Wii controller to more closely resemble a laparoscopic grasper and combined these with serious games that mimic reaching, grasping and bimanual maneuvers (87). Devices such as these begin to blur the lines between video gaming and virtual reality laparoscopic simulators. The purpose of this study however was to evaluate the effectiveness of training with traditional video gaming we could expect to find in a trainee's home.

This study has several limitations. First, non-surgical trainees were used as participants. We felt it was more important to compare the training effects on complete laparoscopic novices than compare learning effects on residents. Previous research has also suggested that simulated laparoscopy is most beneficial to MIS novices (88). While this allowed us to recruit many more participants, we cannot guarantee that our sample is representative of the baseline ability and technical aptitude of surgical trainees. This represents a selection bias.

As in all studies utilizing FLS performance as an outcome measure, some may question the transferability to operating room performance. There is growing evidence that improved simulator performance does correlate with improved intraoperative performance (89,90). While we acknowledge that FLS task performance might not be a perfect predictor, we were more interested in demonstrating the effect of our training program on learning effects. We believe that the FLS tasks are a reasonable surrogate measure of laparoscopic performance in this context.

Finally, the laparoscopic and video gaming practice completed by each subject was largely self directed. Any individuals involved in data collection were encouraged not to coach or provide feedback to the students during practice. Consequently, we did not test more complex FLS tasks such as suturing. Due to anticipated differences in the quality and quantity of teaching between individuals, we felt it was reasonable to allow self-directed practice. This was also the reason for providing instruction for the FLS tasks in the form of a standardized video.

Chapter 5 – Development of a Marker-less Motion Detection System for Hand Movement in Open Surgical Tasks

This chapter describes a complete study focused on comparing technical ability between surgeons with different levels of experience with the aid of a relatively new optical tracking technique.

Introduction

As a result of technological advances in optical tracking, hand motion data can be acquired with high precision, which has potential benefits for many applications including surgical training. Surgical procedures require a mastery of both technical and judgment skills. The evaluation of technical skill is often carried out informally by senior or board certified surgeons. However, despite attempts to develop standardized evaluation rubrics, such as the Objective Structured Assessment of Technical Skill (OSATS) examination (91), using other surgeons as human evaluators often results in a significant amount of inherent bias and subjectivity.

We propose using computer-assisted motion analysis, which is able to measure and decipher patterns in an individual's motor movement, and provide a more objective and reproducible method for evaluating dexterity. While computer-assisted motion analysis for surgical training has been developed over a number of years, its application to open surgical movement analysis is still in its infancy due to the high number of degrees of freedom associated with hand and finger motion. In recent years, smart sensors have been increasingly utilized to capture motion data for analysis, e.g. accelerometers used to analyze motion state (92) and smartphone acceleration sensors is used to classify physical activities (93). However, these systems have yet to be applied to the evaluation of hand motion performance in open surgery.

This chapter describes a smart sensor setup utilizing a Leap Motion Controller (Leap Motion, San Francisco, CA), for acquiring surgical hand motion without using markers or other devices placed on the hand. Despite the complexity of the human movement, motion analysis can identify patterns in motor movements specific to a particular stage of learning and provides insight into strategies for improving performance. Although this study focuses on surgical tasks, our motion acquisition and analysis technique can be applied to other applications which require comparison of hand motions or gestures.

Objective Analysis of Hand Movements in Surgical Tasks

Evaluating surgical hand movements is challenging given the operating room environment and multidimensional motion data. Our analysis involves determining a quantitative statistical distance (similarity) between experts and novices. Rosen *et al.* (94) addressed this problem in laparoscopic surgery by using a discrete Markov Model (MM). They decomposed a complex surgical task in order to select low-level elements that can be associated with quantifiable and measurable parameters. Murphy and colleagues further demonstrated the utility of the MM language model for this purpose (95). These results indicate that a stochastic approach might describe the surgical process better than a deterministic approach based on validation via comparison with traditional expert performance (96). Previous studies applied on laparoscopic surgery have had the advantage of a decreased complexity of movement due to more restricted DOF (97). In contrast, open surgery involves greater DOF and more bimanual (two-handed) maneuvers. As a result, additional parameters such as orientation and position of individual hand components, i.e. finger segments, are necessary to determine the current state of a Markov Model. To address this parameter complexity issue in open surgery, we explore the Hidden Markov Model (HMM) and introduce a cluster based analysis method to detect discrete sets of highly concentrated or clustered information in the sample data for each predetermined Markov state, leading to a significant reduction in the data complexity. We achieve data reduction in three steps:

1. Creating a subset of the data associated with each state common to all subjects;
2. Using K-means vector quantization algorithm (98) to identify a number of centers associated with each state;
3. Encoding the raw motion data of the surgical tasks based on these clusters in order to convert the multidimensional data into 1-dimensional vectors with finite symbols.

Processing the data in this fashion effectively generates a discrete HMM for each individual. Once the HMMs are defined to characterize subjects with specific skill levels, it is possible to compute the statistical difference between individual's performance based on their hand motion. By comparing trainees to expert level performance, objective criterion can be generated for evaluating user dexterity.

In a Markov Model (MM), each state has an associated physical meaning, but in Hidden Markov Model (HMM) some of the states are abstract and not related to a specific physical interaction (96). HMM supports a more compact model topology that allows the system to model surgical motion in a group of subjects with mixed abilities. By applying an objective evaluation approach making use of the HMM topology, we anticipate an alternative method that eliminates the inherent bias in subjective evaluation. HMM has been used in a similar fashion for activity recognition for personal health applications (99). Here, we apply HMM to differentiate expert and novice performance in open surgery training. Readers interested in more detail can refer to Rabiners' review of HMMs (100).

To define a HMM model, we use the notation $\lambda = (A, B, \pi)$, where A is the state transition probability matrix, which contains the probability of transition from state Si to Sj. Assuming we have N states, the elements in matrix A are denoted by Eqn. 5.1.

$$a_{ij} = P\left[q_t = S_j \middle| q_{t-1} = S_i\right]$$
$$1 \le i, j \le N, a_{ij} \ge 0, \sum_{j=1}^{n} a_{ij} = 1 \qquad (5.1)$$

B is the observation probability matrix, which describes the probability of one state Sj , generating one observation v_k at time t and has elements $b_{j(k)}$ defined by the following Eqn. 5.2.

$$b_{j(k)} = P\left[v_k \text{ at time } t \middle| q_t = S_j\right]$$
$$1 \le j \le N, 1 \le k \le M \qquad (5.2)$$

Finally, π is the initial state probability distribution as defined by Eqn. 5.3.

$$\pi_t = P[q_1 = S_t] 1 \leq t \leq N \qquad (5.3)$$

Based on the above computational model, we hypothesized that:

1. Descriptive statistics applied to hand motion data captured by smart sensor devices would be able to differentiate between novice and expert performance.

2. Based on the previous application of HMM to laparoscopy (101), a HMM applied to our tracked data would have at least 80% discriminatory ability to differentiate between expert and novice performance based on normalized statistical distance to an expert model in open surgery.

Methods

We constructed an open surgery simulator for capturing hand motion data with the Leap Motion controller. This consisted of an acrylic box with an adjustable system for suspending a monofilament at a consistent position (15 cm) above the Leap sensor. The experimental apparatus is shown in Figure 5.1. Surgical suture (2-0 Silk or 2-0 polygalactin) was then tied to the nylon monofilament by each participant. Real time video of the tying task was captured using a Canon 40D Camera in “Liveview” mode. The Leap sensor was placed cross to the middle bottom, as shown; with the nylon monofilament towards the rear part of the box to ensure that the participant's hands were centered over the Leap device during the suture tying task. The Leap controller communicated with a Personal Computer (PC) via a Universal Serial Bus (USB). Data was captured using custom software developed in C++.

Figure 5.2 depicts our Graphical User Interface (GUI), which illustrates a 3D hand model on the left panel. The right panel displays the corresponding raw video image captured by the camera. An overview of the motion acquisition pipeline is shown in Figure 5.3. In our system, two streams of motion data are captured. The first is recorded at a fixed frequency simultaneously with video from the camera at 60Hz. The second stream contains all the possible motion data captured by the Leap Controller during the recording period (typically greater than 100Hz). Both data streams are saved with timestamp information to ensure temporal coherence.

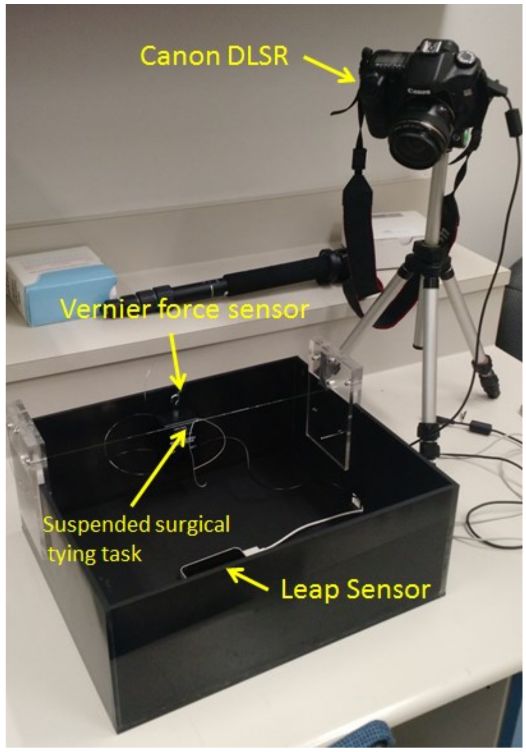

Figure 5.1: Experimental setup. Open surgery simulator with Leap Controller and Canon DSLR for real-time motion and video capture respectively.

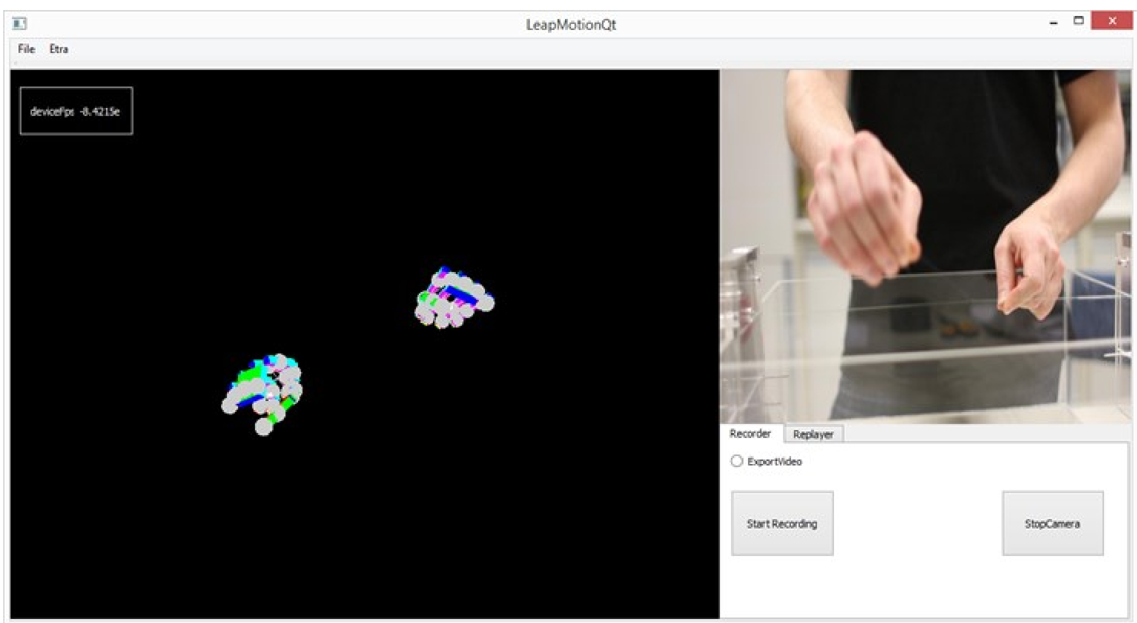

Figure 5.2: Screenshot of the GUI for our motion capture application: (Left) 3D hand model with Cartesian position data, and (Right) video image captured by the digital camera.

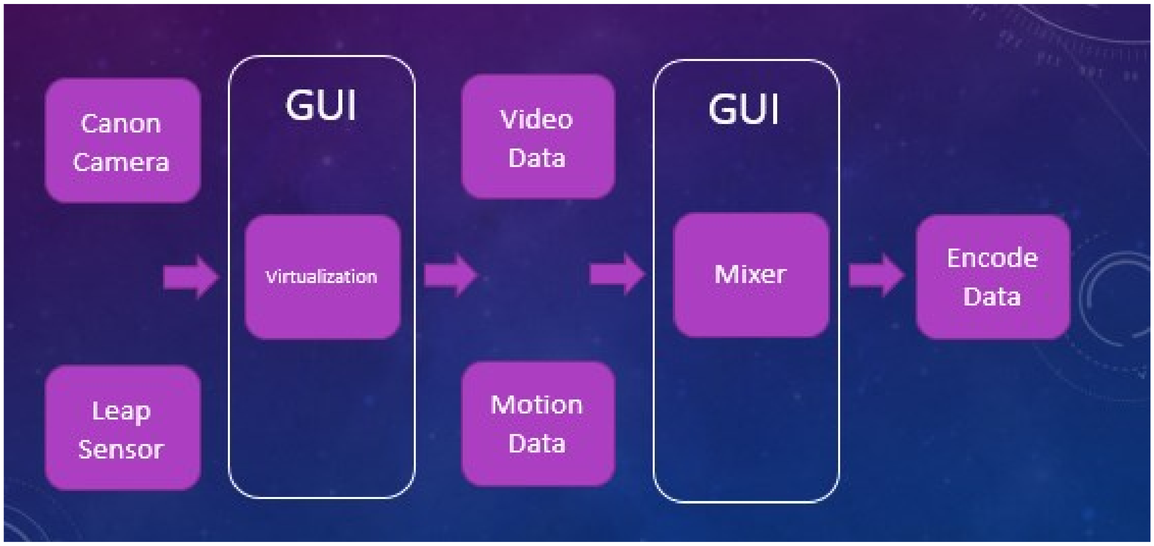

Figure 5.3: An overview of our motion data acquisition pipeline.

We implemented an interpolation function to ensure that the captured data was recorded at a constant frequency of 60 Hz. After interpolation, we applied a two-pass Butterworth filter algorithm to smooth the data and eliminate high frequency noise in the motion tracking data. Two experiments were carried out based on motion data obtained with the Leap Controller. In the first, participants were asked to perform a sequence of one- and two-handed surgical ties to place a total of five square knots in the training box. Motion data was captured from both procedures and analyzed using traditional metrics: path length, number of movements and total time.

In the second experiment, which was designed to test our objective evaluation algorithm, we chose a simple non-surgical procedure involving an object transfer task using a single hand. This required a participant to lift an object, transfer it from Point A to Point B, and release the object prior to transferring the object back to Point A and again to point B (3 transfers total). A total of nine participants were asked to perform the transfer task. The first six participants were asked to perform the task as efficiently and smoothly as possible, emulating expert movement. An additional three participants (controls) were asked to perform the task with more hesitation, including idling between movements to emulate novice behavior.

Results

Descriptive Statistics from Tangential Velocity Analysis

Following interpolation and filtering of pilot motion data obtained from two subjects, tangential velocity analysis was applied to calculate the number of hand movements based on changes in velocity. In addition, the cumulative distance travelled by each hand was calculated to determine the path length. Total time to complete each task was also compared. The number of movements was calculated by using a peak finding algorithm with a threshold set to the mean tangential velocity of all movements. The results are shown in Table 1. As expected, and similar to the metrics obtained with the ICSAD(102) novices required more movements and time to complete each task. Figure 4 depicts the tangential velocity curves and mean velocity threshold (right hand) of novices and experts in the two-hand tying task. Here, novices demonstrate slower hand movement and have decreased peak velocities compared to experts. This pilot experiment demonstrates that the marker-less sensor-based system is comparable to the ICSAD system.

Table 5.1: Surgical knot tying tasks: ICSAD metrics obtained using Leap sensor.

Task	Experience Level	Path Length (m)	Number of Movements	Task time (s)
One hand tie	Novice (n = 1)	2.23	43	49
	Expert (n = 1)	1.11	24	22
Two hand tie	Novice (n = 1)	3.08	40	65
	Expert (n = 1)	3.05	26	36

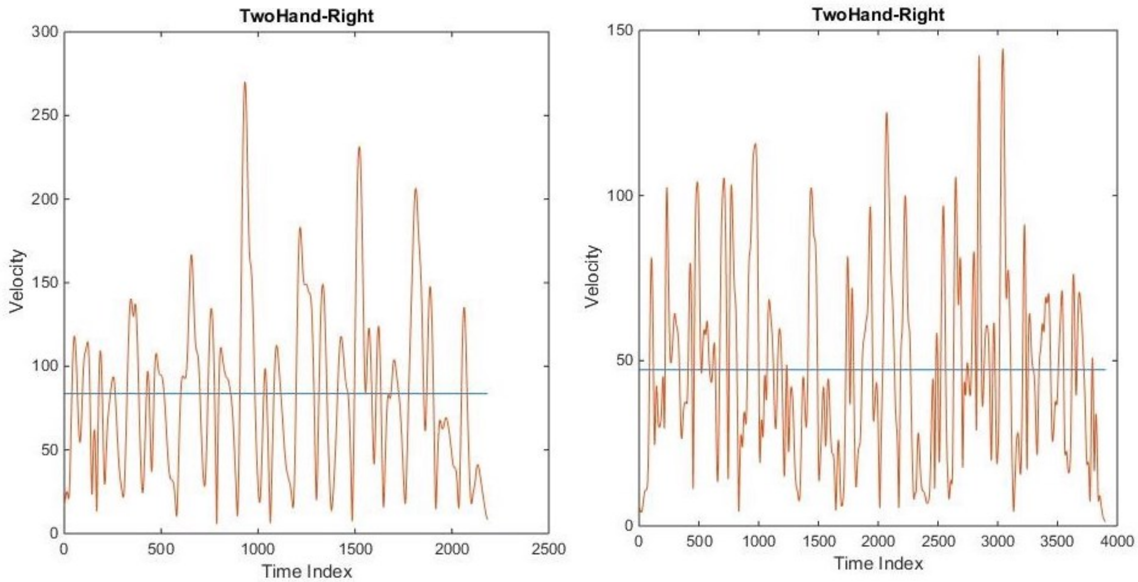

Figure 5.4 Tangential velocity curves and mean velocity threshold of the right hand motion when performing the two-hand tying task (Expert – Left; Novice – Right).

Hidden Markov Model Analysis

Motion capture of the entire hand of each participant during the hand tying tasks produced inconsistent data due to significant optical occlusion by the fingers during complex movement. Consequently, participants completed a non-surgical transfer task where six gestures (states) could be defined including: Idle, Dropping, Grasping, Evaluating, Translating and Releasing. These states were each described using palm velocity (*vx*; *vy*) along the x-axis and y-axis, and the velocity (*vs*) of the change in the distance between the thumb tip and middle finger as depicted in Figure 5. These observable parameters were selected because they undergo significant changes during the predefined gestures.

Following hand motion and video capture, we applied the interpolation function and Butterworth filter to reduce noise in the raw data (Figure 5.6a) and generated a new data set (Figure 5.6b) with a frequency of 60 Hz.

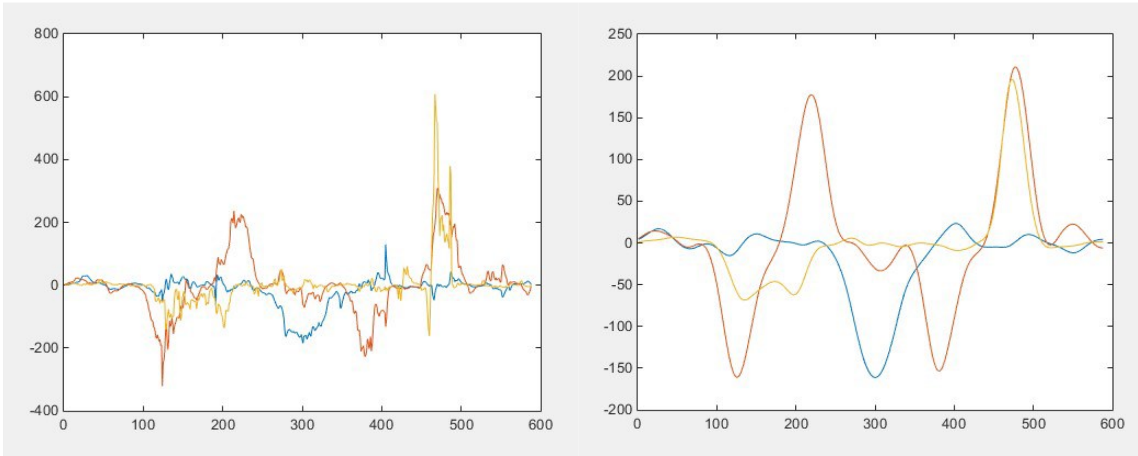

Figure 5.5 Raw velocity data over time (vx-Blue,vy-Yellow,vs-Orange) (a) and filtered/interpolated data (b).

The synchronized video and motion data were reviewed, and an evaluator identified the Frame IDs corresponding to the start and end of each gesture. This manual segmentation of the motion data served as the ground truth for movement classification, permitting the generation of the observation distribution matrix B, the basis for the HMM representative of expert performance. We then used the training function to train the model and initialize all the elements in each row of the state transition probability matrix A.

Initially, the classification was done using frame-by-frame video analysis by an expert; but once initialized, the HMM was able to update matrix A through training algorithms. After initializing Matrix A, we generated Matrix B which stores the observation probabilities. Each one of the six states was associated with a unique set of velocities (*vx; vy; vs*). To simplify the theoretical and computational load of the modeling process, we applied a data reduction process to translate the multidimensional data to a one-dimensional observation vector sequence. As part of this process, we applied a K-means vector quantization algorithm (98).This allowed us to transform the continuous three-dimensional vectors into one-dimensional vectors of 60 observation symbols (10 symbols for each of the 6 states). After applying the K-means algorithms to the motion data representing expert performance, we were able to identify 60 clusters of associated motion data (velocities). The cluster centers were used as the observation symbols for encoding all the multi-dimensional motion data to one-dimensional sequences corresponding to the 60 clusters. To achieve this, we calculated each frame's Euclidean Distance from each of the cluster centers and chose the minimum as the observation symbol for the current frame. We were then able to initialize the observation distribution matrix B corresponding to our expert-encoded observation symbol sequence and the video analysis table using Eqn. 5.4. The initial state probability distribution can be defined based on the assumption that all tasks start in the idle state.

$$B_{jk} = \frac{\#\, frames\ in\ state\ S_j\ \&\ using\ observation\ symbol\ v_k}{\#\, frames\ in\ state\ S_j}$$

$$1 \le j \le N, 1 \le k \le M \qquad (5.4)$$

The above process describes the initialization of all the parameters necessary to define the HMM, $\lambda = (A, B, \pi)$ for our expert reference model. Next, we used additional sets of motion data captured from the expert(s) to train the model using a Baum-Welch algorithm.(103) This step further optimized the parameters in the model, making it more reliable and accurate to describe expert performance. Figure 5.6 and 5.7 show the color-coded results of the optimized matrix A and B following training.

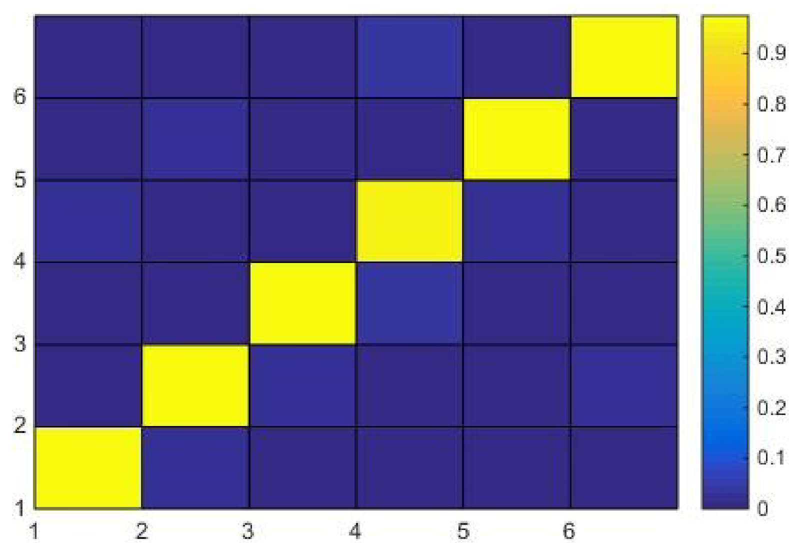

Figure 5.6 Graphical representation of optimized state transition probability matrix A for an expert.

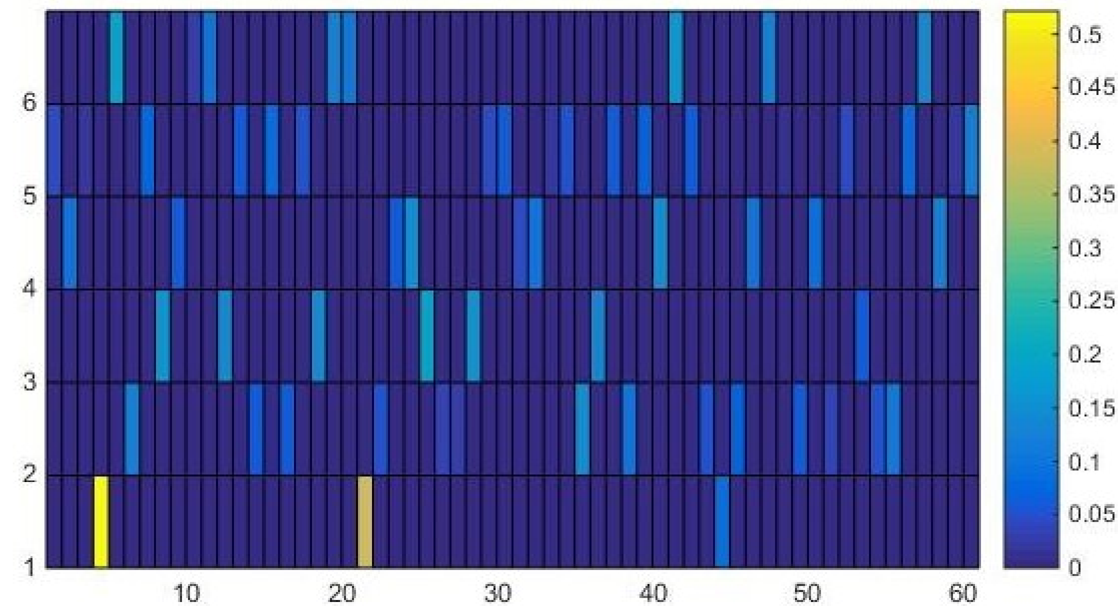

Figure 5.7 Graphical representation of optimized observation distribution matrix B for an expert.

HMM comparison for performance evaluation

After obtaining the trained models corresponding to the expert performance λ_E we applied an evaluation function to compare test behaviors to our expert performance. This function provides the probability that a performance described by a given observation symbol sequences is generated by the expert group model λ_E. The values generated by the evaluation function were normalized as previously demonstrated(104) in order to compare motion data of different durations. Eqn. 5.5 can be used to determine the probability of an observation sequence matching the expert model, where $\hat{O}_i^E$ is the observation sequence generated by λ_E and composed of i observations. Finally, in Eqn. 5.6, S is the normalized distance of a given observation sequence from the normalized set of observations generated by the expert model. This value can be used to objectively assess performance based on different observation sequences obtained from hand motion data. Lower S scores (E4, E5 and E6) imply that the performance closely approximates the expert model. Novices (N1, N2 and N3) have a larger normalized distance.

$$P_N = (O, \lambda_E) = \frac{1}{n} \sum_{t=1}^{n} [logP\left(\hat{O}_t^E \middle| \lambda_E\right] = 1 \qquad (5.5)$$

$$S\,(O,\lambda_E) = \frac{|logP(O|\lambda_E) - P_N(O|\lambda_E)|}{\frac{1}{n}\sum_{t=1}^{n}|logP\left(\hat{O}_t^E\middle|\lambda_E\right) - P_N(O|\lambda_E)|} \quad (5.6)$$

In order to evaluate the effectiveness of HMM for hand motion data classification, we collected motion data from the object transfer task described previously. From a total of nine participants, six experts were in Group E1: (E1;E2;E3) and Group E2: (E4;E5;E6). Three novices were included in Group N: (N1;N2;N3). Training of the expert reference model utilized motion data from Group E1. The evaluation function was then applied to the remaining subjects in Group E2 and Group N. Visual inspection of the velocity profiles in Figure 5.8 already suggested a difference in the motion data between experts and novices. The novice velocities were more discreet when compared to the amount of overlap in the expert trials. Next, we applied the HMM decoding function to identify the hidden states and compared the decoded gestures. We then calculated the normalized distance (SE4; SE5; SE6; SN1; SN2; SN3) for each novice and expert performance. The comparison of normalized distance is shown in Figure 5.9. We performed a statistical comparison between the three replicates using Students t-test and found that the distance S between the expert and novice groups was statistically significant as supported by the p-value of 0:007 in Table 5.2.

Table 5.2: Normalized distance for novice and expert participants

Experience Level	Statistical Distance	p value
Novice (n=3)	7.1±0.80	0.007
Expert (n=3)	2.01±0.06	

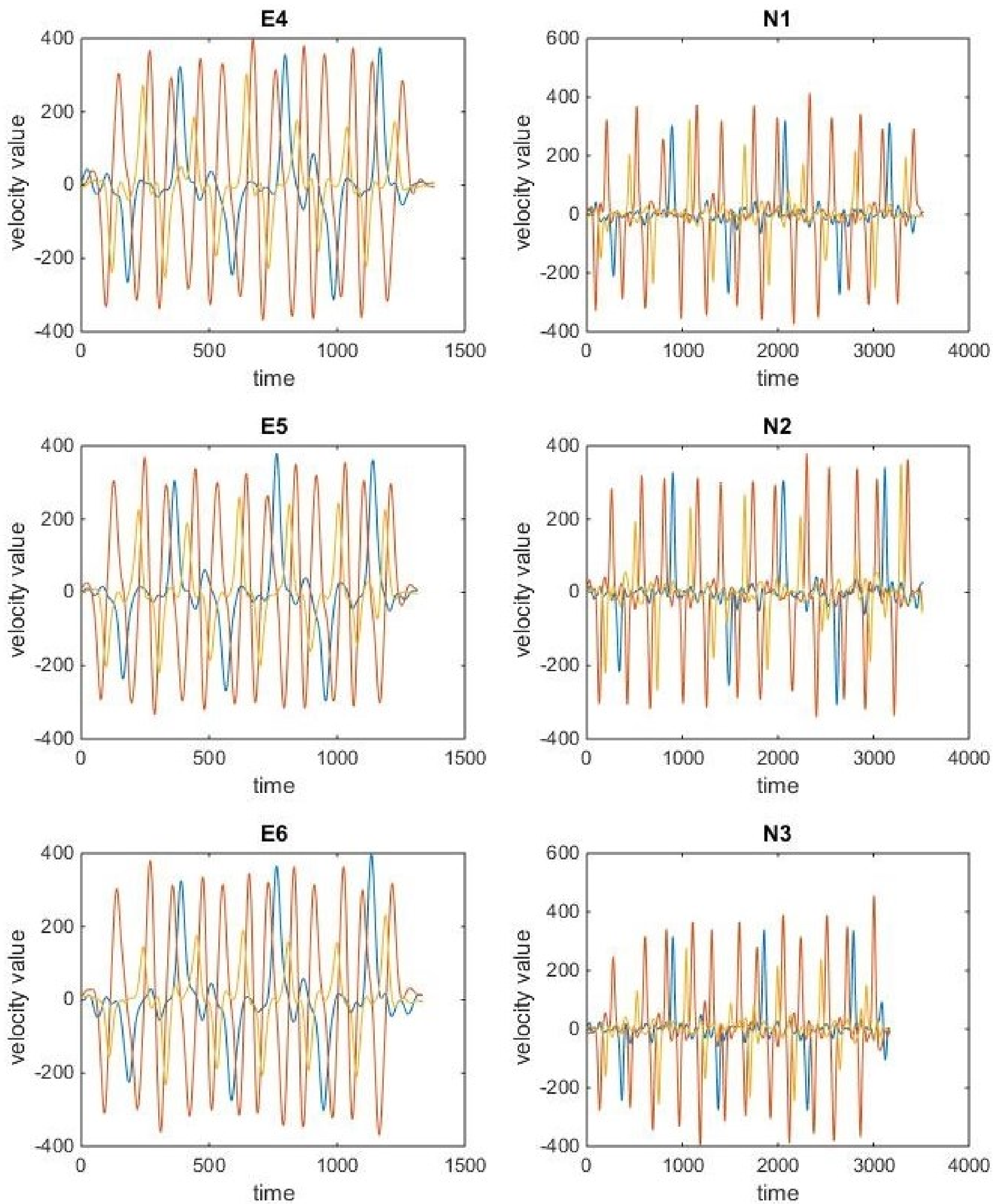

Figure 5.8 Velocity profiles for novices (Ni) and experts (Ei) performing the transfer task.

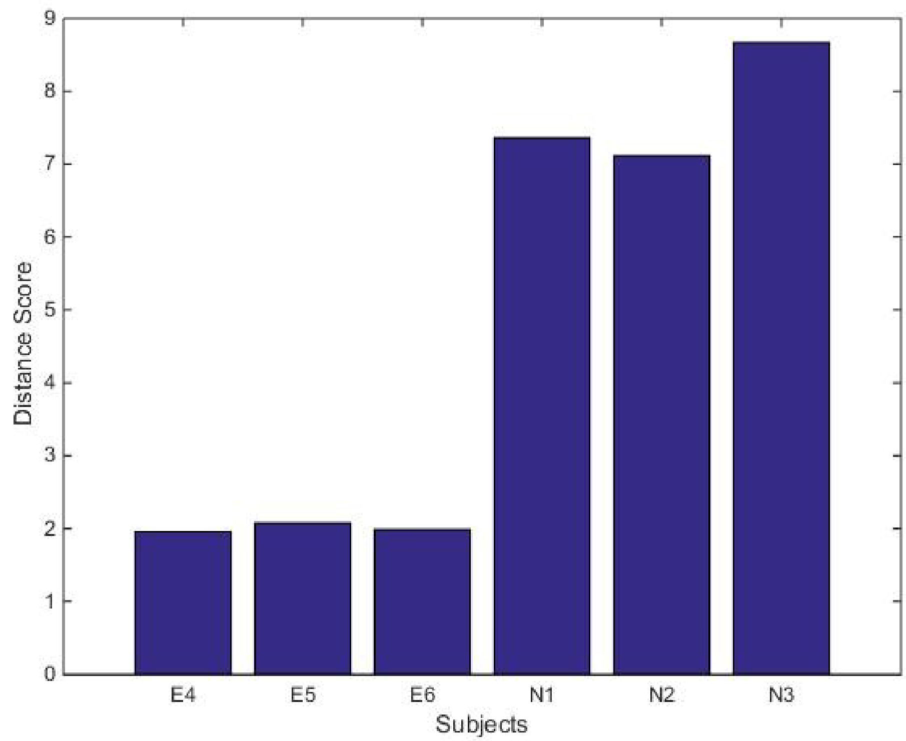

Figure 5.9 Normalized statistical distance between novice (Ni) and expert (Ei) test participants.

Discussion

This study demonstrated that hand motion data obtained from smart sensors can be used to evaluate individual performance at different dexterity levels. Our computer assisted system can measure performance differences between experts and novices using both low level descriptive statistics and a Hidden Markov Model. To the best of our knowledge, this is the first application of a marker-less tracking system for objectively measuring surgical dexterity. While some video based methods have been demonstrated in the past, our method provides much richer information including hand position in 3D space and supports much higher degrees of freedom. In addition, our algorithm is sensor-platform independent.

Surgical movement can be enormously complex and variable based on procedure and patient specific factors. With a large enough library of gestures or states, we may eventually be able to fully classify all of the maneuvers in a particular operation. Additional work is required to improve the accuracy and reliability of the acquisition systems obtaining the motion data on which to apply the analysis presented here. We believe the feedback based on a granular analysis at the subtask or surgeme(105) level holds the key to improving the objectivity and usefulness of this technique as both an assessment and training tool.

Another limitation of a HMM system is that unique maneuvers may be classified as deviating from expert behaviour and thus misclassified. However, we believe that surgery is still comprised of enough rudimentary surgemes on which to base a comparison. This has yet to be investigated.

From a training perspective, while the normalized statistical distance between models presented here does little to improve upon the utility of a rating generated by an expert evaluator, a more in depth evaluation of each particular gesture can provide insight into particular maneuvers that might need additional practice. For example, a potential energy comparison of segmented laparoscopic gestures has been used to differentiate between novice and expert behaviour (25). The ability to generate these comparisons in real time may pave the way to surgical simulators that can identify and provide valuable feedback to surgical trainees.

Finally, improving the redundancy of information in optical-based motion capture systems is known to improve accuracy and accommodate for occlusion. Future work is planned to explore the use of a multi-leap system for acquiring hand motion data.

Chapter 6 – Development of a Surgical Hand Kinematic and 3D Eye Tracking System

6.1 System Overview

In order to capture a detailed kinematic data set for subsequent motion and language model analysis, we designed and constructed a prototype open surgery hand motion and eye tracking capture system. This involved the integration of multiple systems including a novel 3D gaze tracking platform, a glove mounted electromagnetic finger and hand tracking system, force instrumented forceps to measure force application, optical tracking of work surfaces, as well as high definition video and audio. A flow diagram of the system and its components is provided in Figure 6.1. Figure 6.2 shows the arrangement of the eye and hand tracking systems.

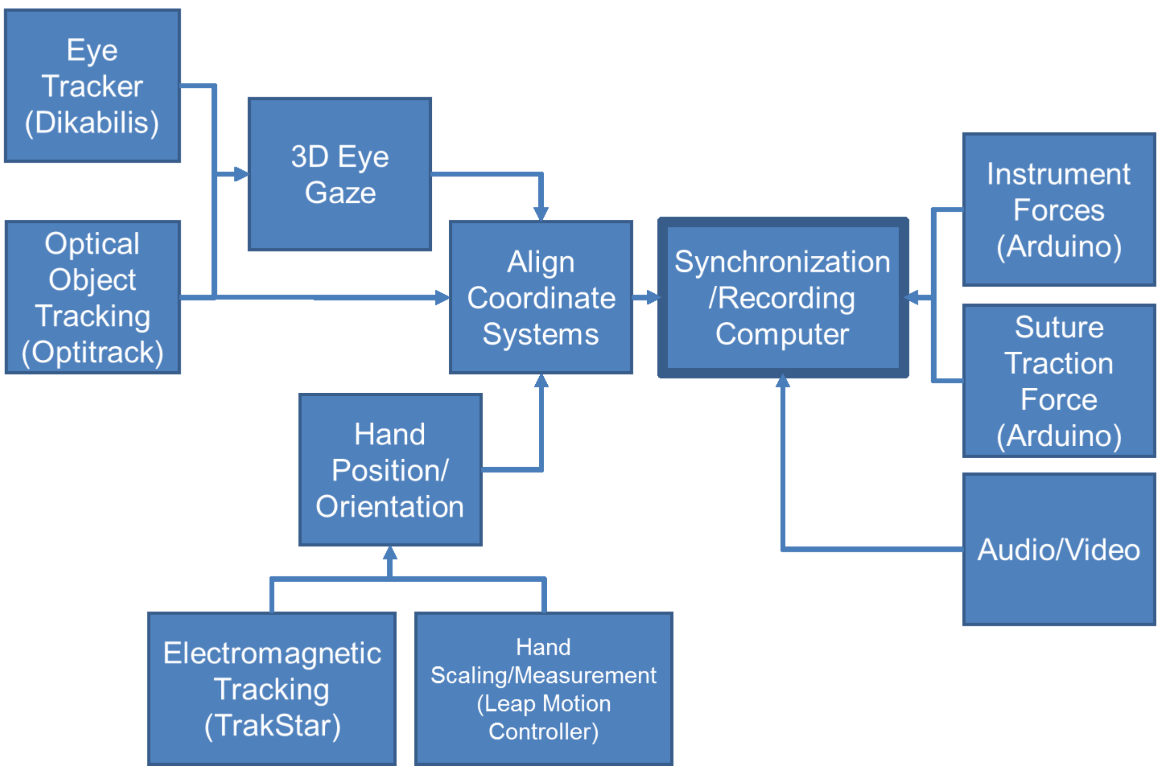

Figure 6.1: Schematic overview of the multi-channel kinematic capture system.

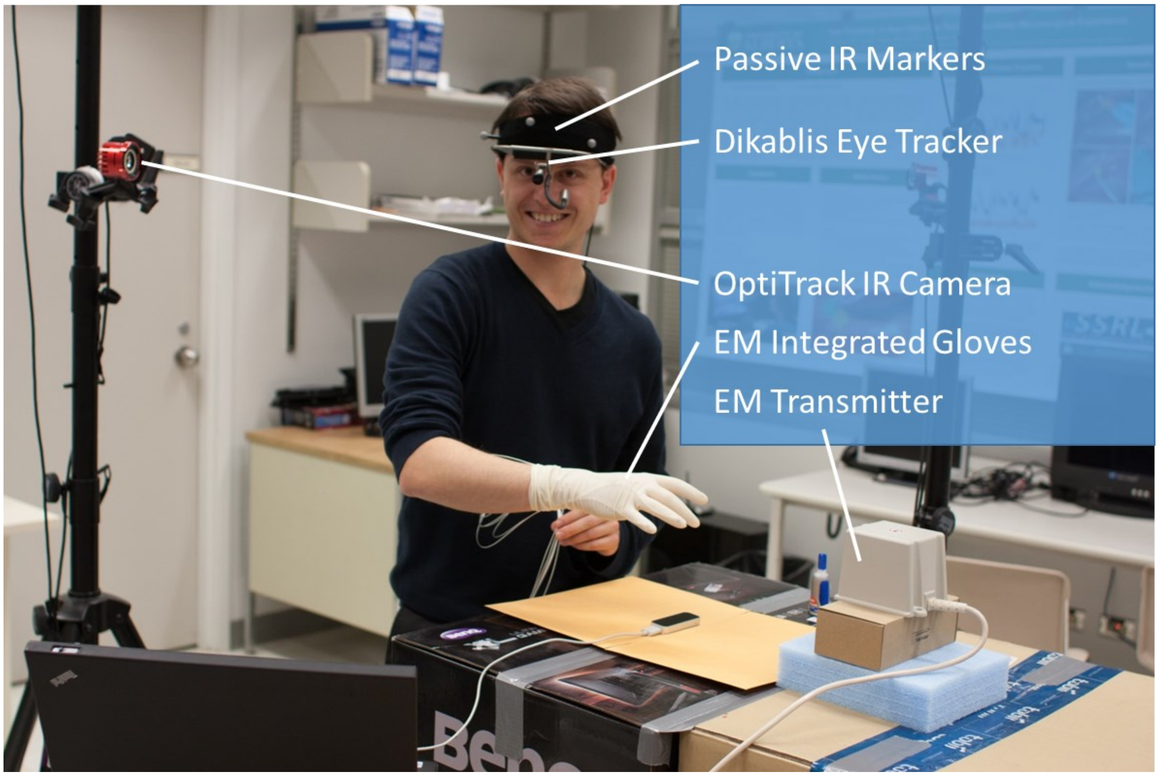

Figure 6.2: Arrangement of optical tracking (OptiTrack), eye tracking (Dikablis) and hand tracking (TrakSTAR) systems.

Each component of the apparatus is described in the following subsections, including design requirements and assembly of the initial prototype. Chapters 7 and 8 describe the validation experiments that were carried out in order to demonstrate the accuracy of the system.

6.2 3D Gaze Tracking

While eye tracking can easily be applied to the 2-dimensional images displayed on a monitor, tracking eye gaze in a 3-dimensional environment is significantly more challenging. This section details the development of a technique for tracking eye gaze in 3-dimensions by combining head and pupil tracking.

Previous studies have demonstrated a similar technology, utilizing optical tracking head tracking with a monocular eye tracker (106). However, by combining a more robust calibration algorithm, it is likely that the system presented here is more accurate. Data acquired from two commercially available technologies including an OptiTrack

(NaturalPoint, Inc., Corvallis, OR) passive infrared motion capture system and Dikablis (Ergoneers GmbH, Manching, Germany) eye tracker were used to generate the 3D gaze vector. In order to track head position and direction, three infrared makers were placed in an asymmetric triangular configuration on the edges of the head mounted eye tracker (Figure 6.2). Following calibration of both the Optitrack system using the proprietary Motive version 1.6 (Figure 6.3) and Dikablis Recorder version 2.5 software, both data streams were streamed to custom developed recording software using NAT-Net and TCP/IP network protocols. The raw motion and eye tracking data was saved using a system generated timestamp. This ensures temporal coherence for subsequent analysis. Additional details regarding the algorithm used to generate the 3D gaze vector are provided in Chapter 7.

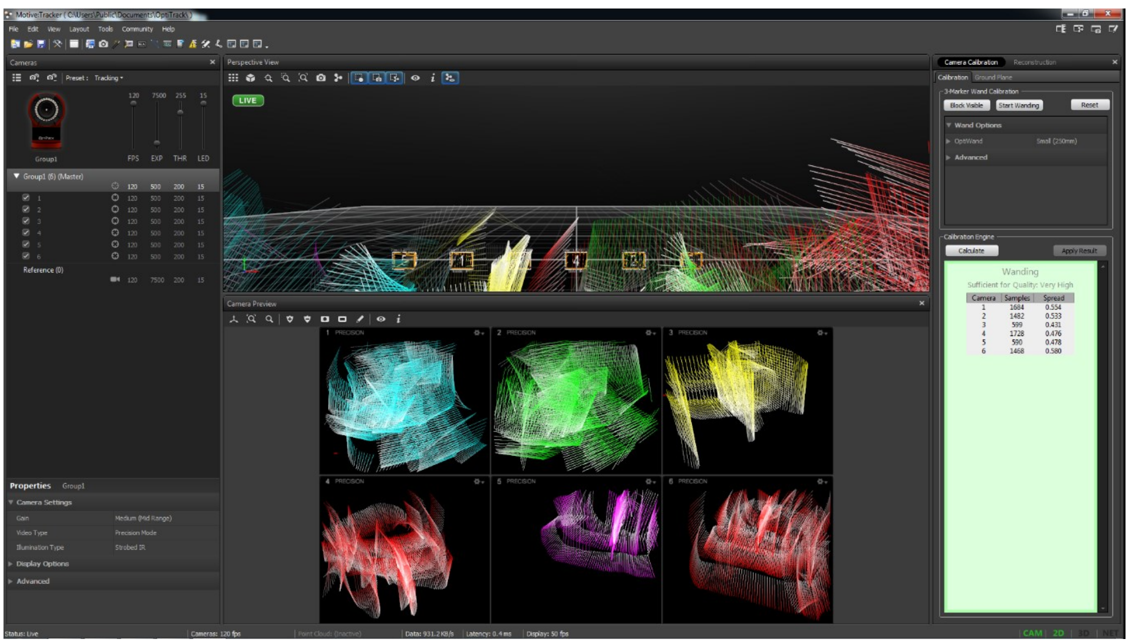

Figure 6.3: Motive software following “wanding” calibration procedure.

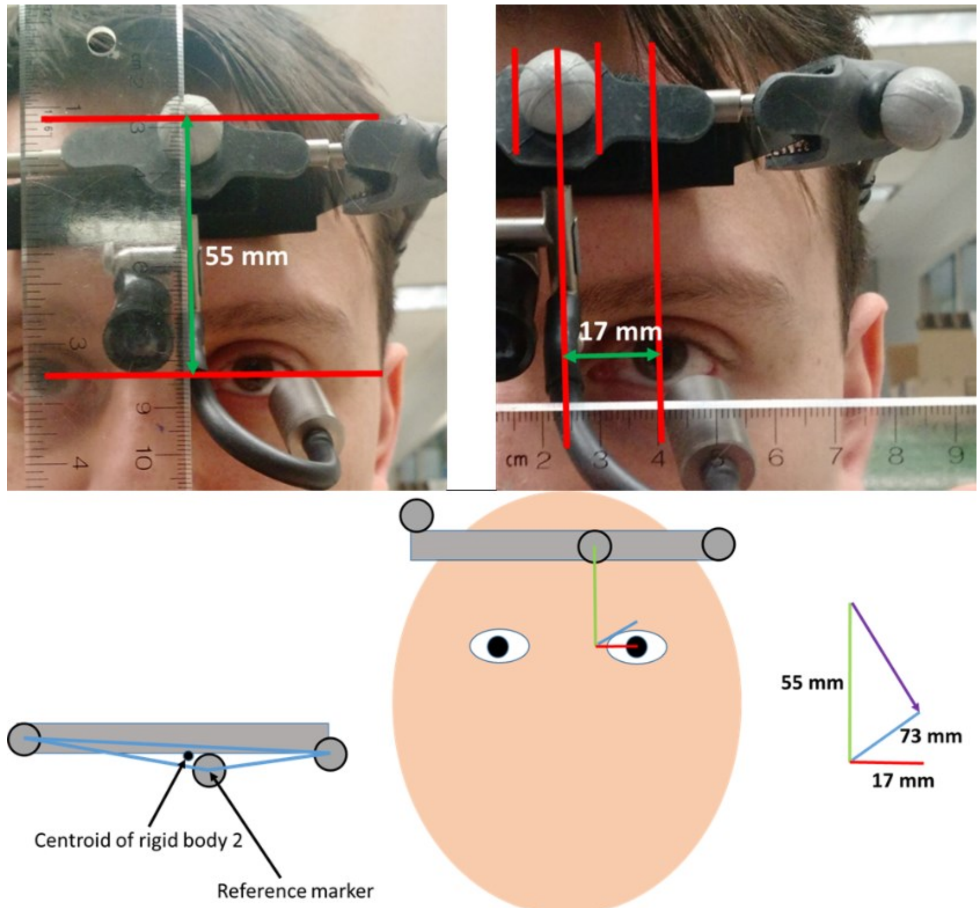

Figure 6.4 Schematic showing positioning of eye tracker and OptiTrack IR markers on Dikablis eye tracker and measurement of the pupil offset with respect to the reference marker.

Initially the pupil position mapped onto the field camera was used to generate the gaze vector. This required either automatic or manual calibration of the Dikablis using four points arranged in a rectangle as described in the Dikablis manual. Manual calibration using this method was used in the empirical study described in Chapter 7. However, the raw pupil position as obtained from the Dikablis eye camera can also be used. This is the basis for subsequent improvements to the eye tracking system and eliminates the 2D calibration step using a rectangle. The intensity of the infrared LED, position of the IR eye camera and threshold brightness for pupil detection must still be optimized for each subject.

6.3 Electromagnetic Hand and Finger Tracking

As described previously in Chapter 4, the ICSAD was the first validated instrument for analysing hand motion during traditional open surgical procedures. However, it only consisted of two electromagnetic trackers positioned on the dorsum of each hand. To perform a more sophisticated analysis utilizing computer modeling, a more detailed description of hand and finger motion is desired. This would permit the analysis of the kinesthetic elements of hand motion during a complex bimanual task such as suturing or hand-tying in surgery. The optical capture technique described in Chapter 5 utilizing a Leap Motion Controller suffered from significant occlusion during complex surgical tasks. While the technology will continue to improve, a more robust technique for the reliable determination of hand and finger position and orientation was desired. We selected electromagnetic tracking as the best alternative. The design requirements for the electromagnetic glove system included the following:

- Accurate determination of hand and finger position and orientation
- Minimal impediment to natural hand and finger movement
- Minimal reduction in tactile feedback

The TrakSTAR electromagnetic system was selected for the glove system. This system allows for the multiplexing of multiple EM sensors with a single transmitter. Small 1.5 mm outer diameter sensors minimize the bulkiness of the final glove system. In order to protect and maintain the orientation of each sensor, encapsulation in a rubber or silicone material was desirable. A schematic of the encapsulated sensor design is shown in Figure 6.5.

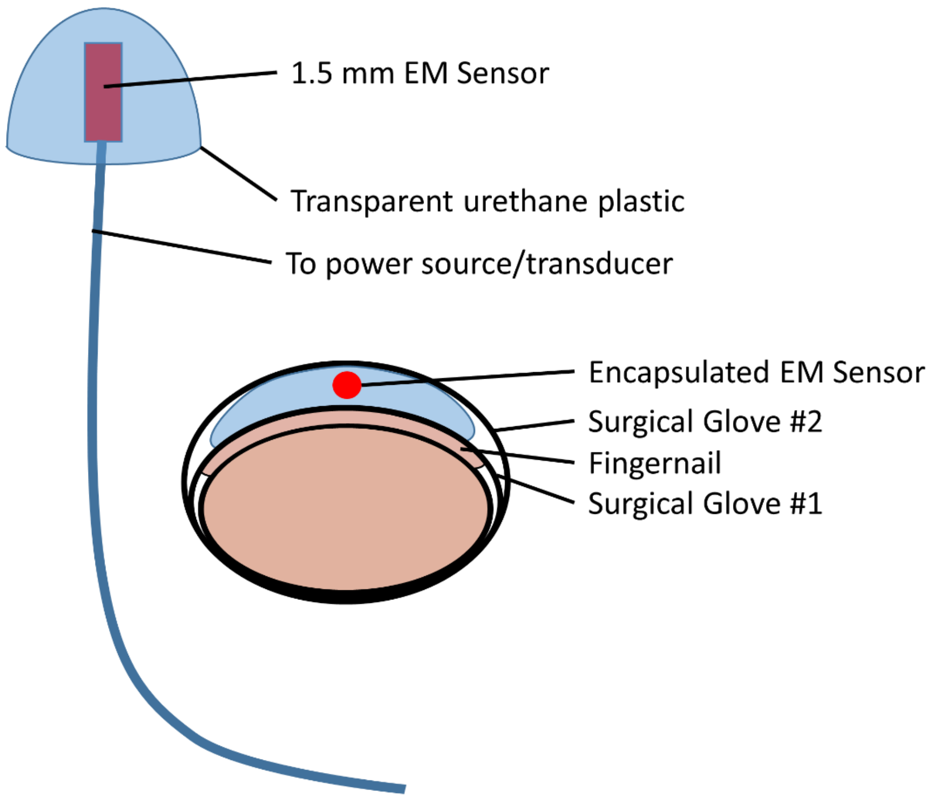

Figure 6.5 Encapsulated sensor design and positioning for each finger.

TinkerCAD (Autodesk, San Rafael, CA), a web-based HTML5/WebGL computer aided design tool was used to design the encapsulation mold for the EM sensors. Each well was designed to be 2.8 mm deep and 12 mm wide to minimize the amount of material necessary for encapsulation and allow fixation on small and large fingernails. The mold was 3D printed on a 3D printer at high resolution (0.02 mm) and sealed with XTC-3D (Smooth-On Inc., Macungie, PA). Two encapsulation materials were tested including Encapso-K and Clearflex-30 (Smooth-On Inc., Macungie, PA). Both are pliable with a Shore A hardness of 30. Encapsulation with Encapso-K was successful but the material was too friable in repeated testing. Clear-flex 30 was subsequently used, but required additional preparation including vacuum degassing which was accomplished with a commercially available Foodsaver V2040 vacuum sealer and canister (Sunbeam, Boca Raton, FL).

Following encapsulation and approximately 72-96 h hardening time, the encapsulated sensors were removed from the 3D mold and excess material was trimmed from the edges. A total of twelve Model 130 EM sensors (NDI, Waterloo, ON) were encapsulated in order to instrument each finger and the dorsum of each hand. After donning a pair of Encore surgical gloves (Ansell, Iselin, NJ), 5 sensors were fixed to dorsal aspect of each finger over the nailbed using a small amount of cyanoacrylate. A second pair of Encore gloves were then placed over the first pair of gloves and sensors. The cables for each sensor were positioned to run along the medial or lateral aspect of each finger in order to facilitate normal finger flexion. The sixth sensor on each hand was placed over the dorsal aspect of the midpoint of the third metacarpal.

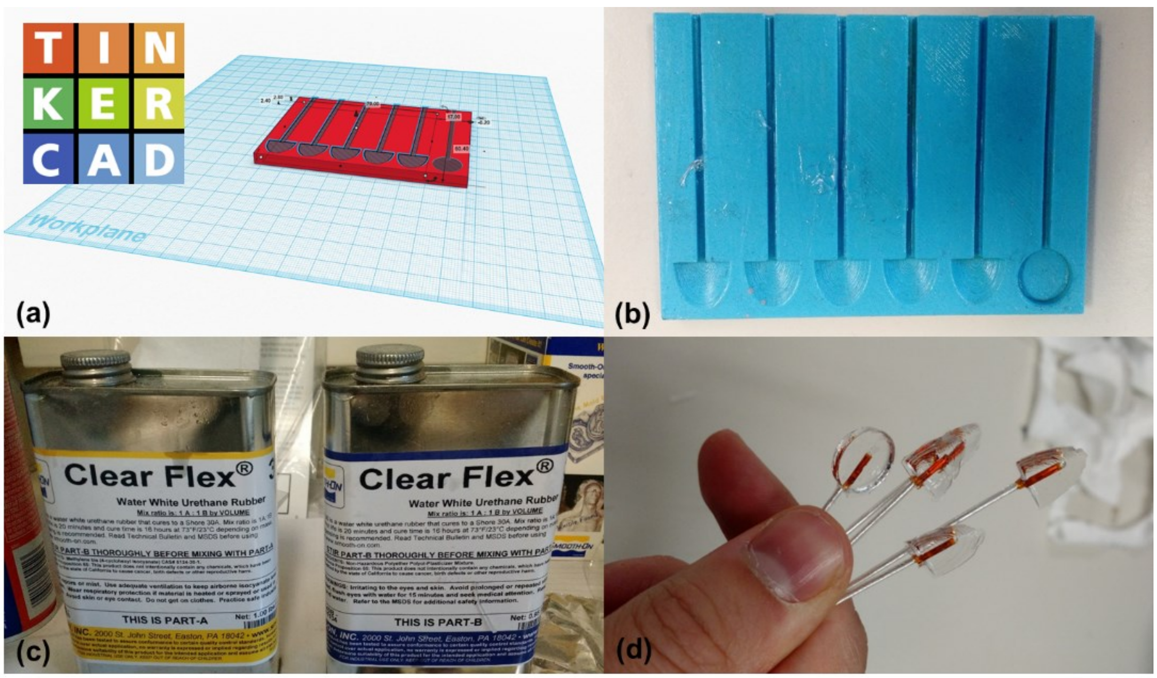

Figure 6.6 Design and fabrication of the EM sensor encapsulation material including 3D design in TinkerCAD software (a), 3D printed mould (b), Clear-Flex-30 urethane rubber used for encapsulation (c) and final encapsulated EM sensors (d).

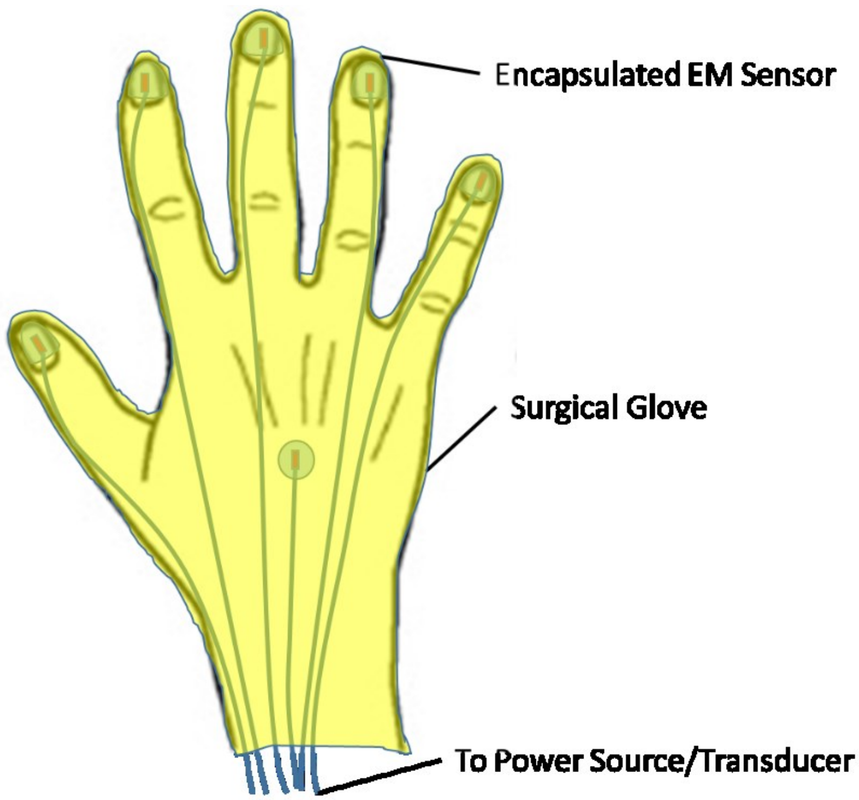

Figure 6.7 Sketch of EM sensor placement on each hand. Encapsulated sensors are placed on dorsal aspect of each finger (fingernail) and a sixth at the midpoint of the third metacarpal.

All twelve sensors were connected to an NDI 3D Guidance TrakSTAR system consisting of 3 units operating in multi-unit sync (MUS) mode with a single transmitter magnet. This system connected to the main recording/synchronization computer via a USB interface. The synchronization and recording software utilized the application program interface (API) to initialize and capture each sensors position and orientation at a frequency of 60 Hz. While the TrakSTAR system permits the polling of sensor position at much higher frequencies (>100 Hz), testing showed that reducing the frequency resulted in less noise and interference.

Spatial coherence was achieved by calibrating the EM sensors using an L-Frame of optical IR markers registered to the OptiTrack MoCap system. This ensured that the EM data and OptiTrack data shared the same coordinate system. A diagram of the different orientations of the coordinate systems for each instrument and software are shown in figure

6.8. All the systems were right handed, necessitating rotational transformations to achieve spatial coherence.

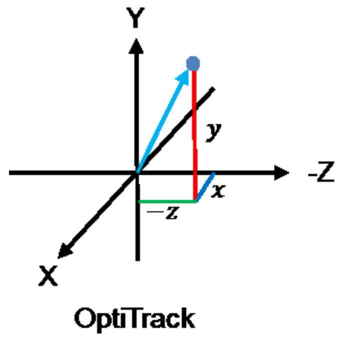

Figure 6.8 Comparison of coordinate systems used by each instrument (OptiTrak, TrakSTAR) and MATLAB visualization.

In order to both visualize and analyse the recorded hand and eye tracking data, we implemented a parsing script in MatLab to reorient all of the data into a shared coordinate system. For example, the position of a tracked object or surface in Optitrack was rotated 90 degrees clockwise about the X axis according to Eqn. 6.1.

$$P = (x_{MatLab}, y_{MatLab}, z_{MatLab}) = (x_{OptiTrack}, -z_{OptiTrack}, y_{OptiTrack}) \quad (6.1)$$

6.4 Instrument Force Measurement

Similar to the force and torque measurements measured in analogous laparoscopic studies, a method for measuring the amount of force applied to various surgical instruments was desired. A variety of electronic sensors were reviewed for this purpose. However, as with the EM glove system, minimizing the bulkiness of the sensor and approximating the normal tactile feedback while using the instrument was desired. Force sensitive resistors (FSRs) were selected for instrumenting our surgical instruments. These are inexpensive piezoelectric sensors that can be used to determine forces applied to the active area of the sensor by measuring resistance across the sensor, where larger forces result in decreased resistance. Two sensors were fixed to a pair of Adson forceps at the typical position for the thumb and forefinger. Figure 6.9 shows a cartoon of the sensor design. A 1.5 mm thick silicon dielectric/interface layer was positioned over the active area of the FSR and a thin layer of transparent 0.5 mm PVC plastic was placed over this to create the final sensor sandwich. An Arduino microcontroller was used to interface with the FSRs and provide a means for recording the resistance measurements and translate these into force in Newtons following a polynomial calibration procedure. A schematic of the Arduino circuit and final circuit are delineated in Figure 6.10. A set of standard weights were used to calibrate the sensor over a range of 150 g to 500 g, with 6 replicate recordings for each weight. The second degree polynomial calibration curve for two of the sensors is shown in Figure 6.11.

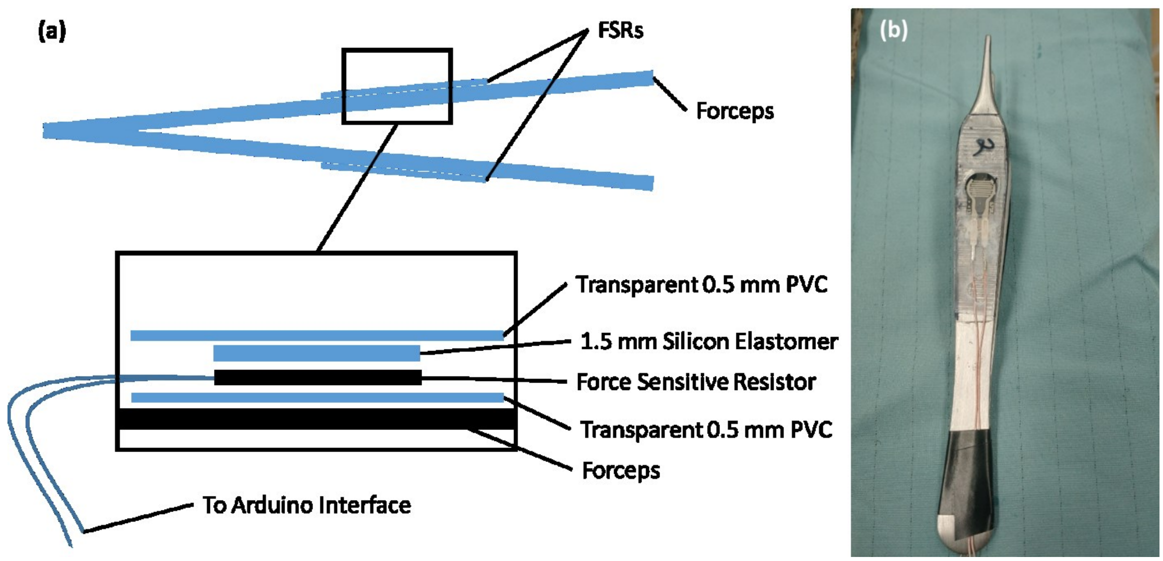

Figure 6.9 Schematic diagram of force sensitive resistor mounted on surgical forceps (a) and photo of prototype FSR mounted on forceps (b).

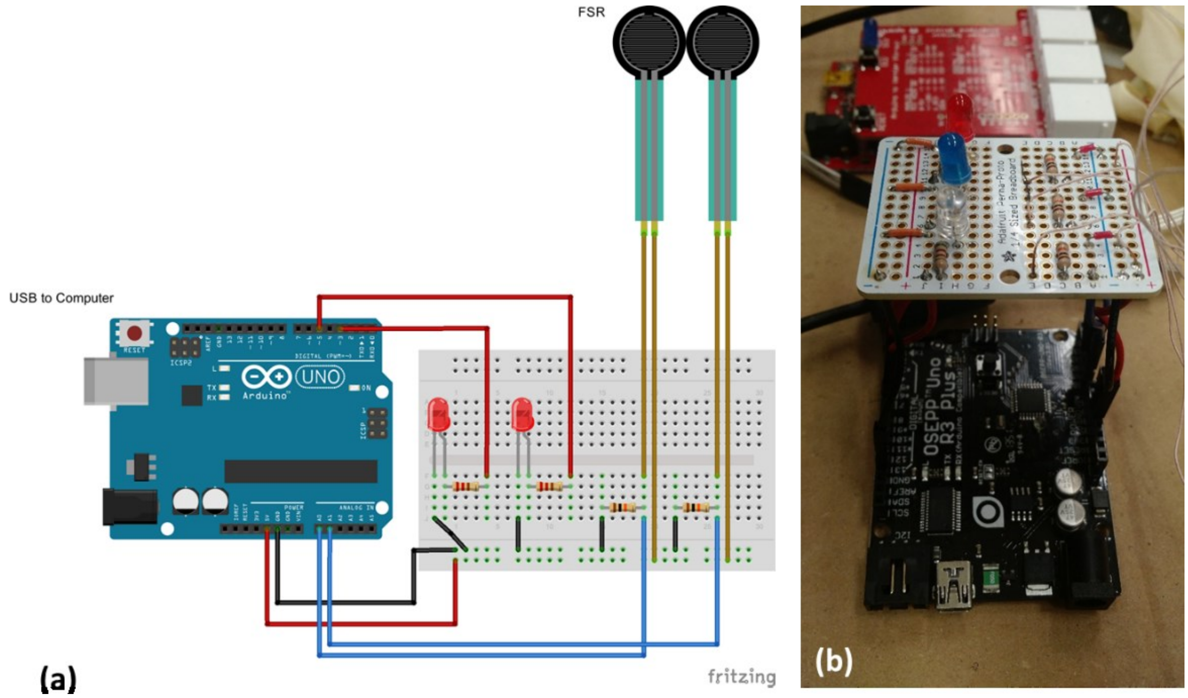

Figure 6.10 Schematic diagram of Arduino circuit (a) and final assembled circuit (b).

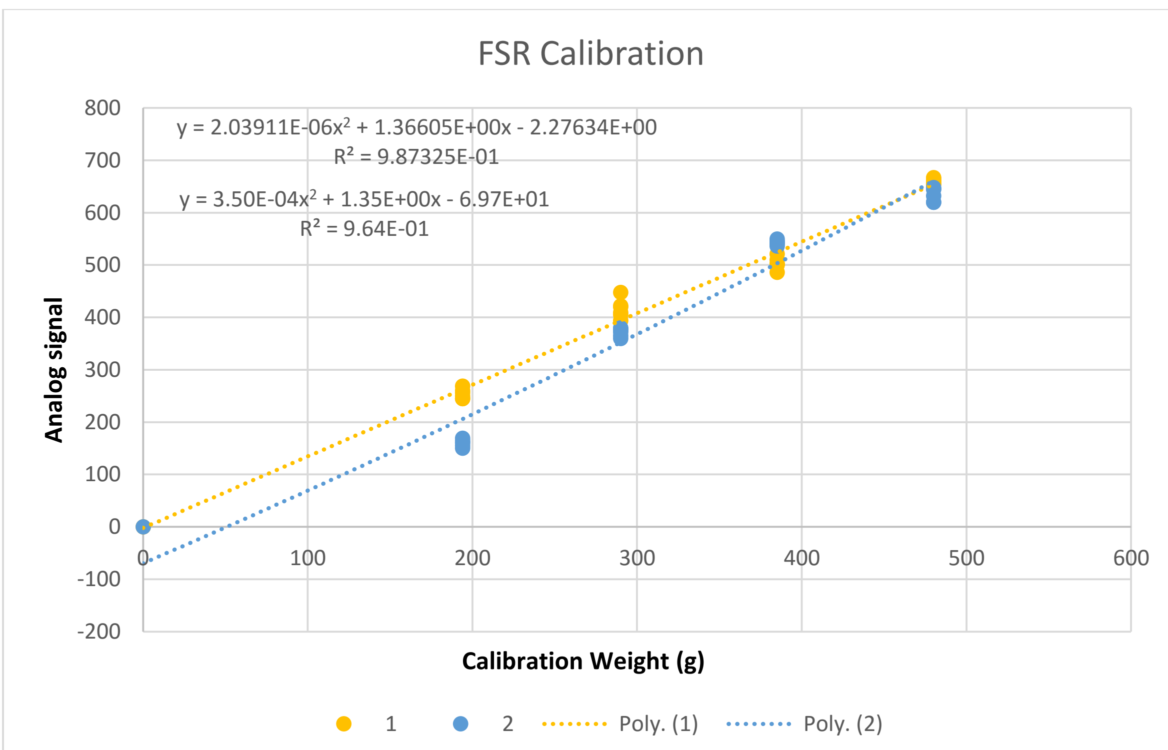

Figure 6.11 Second degree polynomial calibration curve for FSR 1 and 2.

6.5 Suture Traction Force Measurement

Many surgical procedures require the delicate application of tension to a suture while performing hand ties, e.g. ligation of small vessels. Measurement of the magnitude of this tension force or forces applied to the tissue being ligated has been shown to vary between surgeons with varying levels of experience (107). Additionally, providing feedback of forces the forces applied to a simulated model can also improve the speed of attaining proficiency in surgical knot tying (108). A surgical knot tying simulator was constructed based on the apparatus previously employed in Chapter 5. In a similar fashion as demonstrated by Hsu *et al.* (107), a dual range force sensor was attached to a nylon monofilament under tension. However, an additional 'pulley' in the form of a non-ferromagnetic aluminum rod was introduced to translate any forces applied to the monofilament into the vertical axis, ensuring alignment with the axis of the force sensor. This approach is likely more robust than that presented by Hsu *et al.*, as forces applied orthogonally to the force sensor without this pulley would not be detected. Figure 6.12 shows a labelled picture of the knot tying simulator.

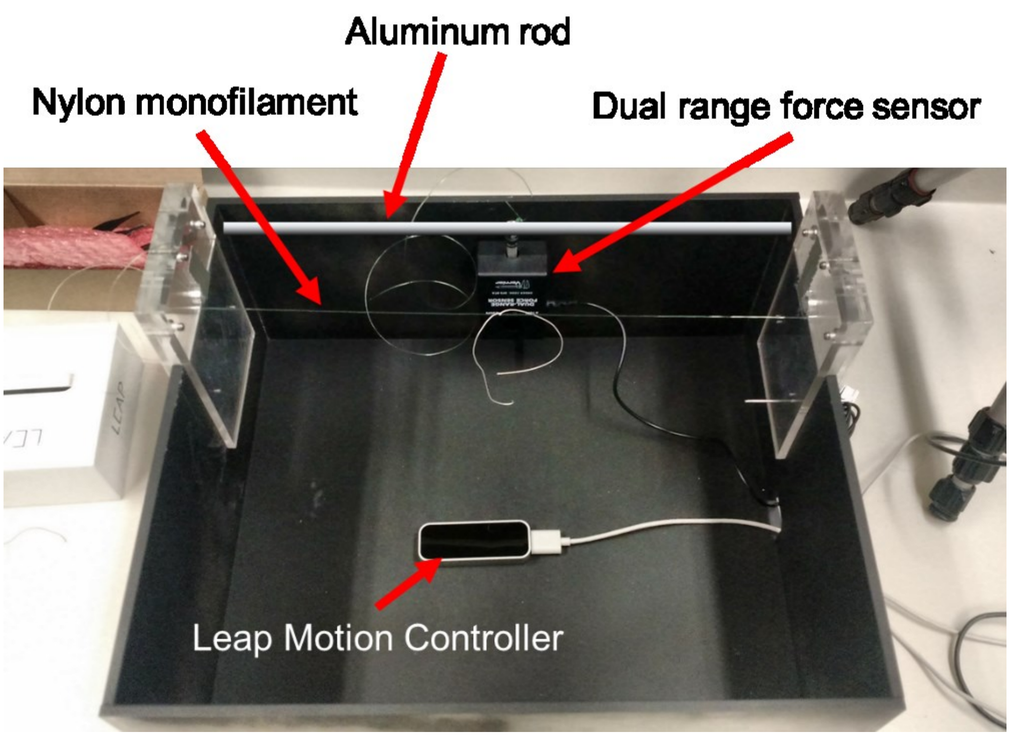

Figure 6.12 Knot tying simulator box with integrated dual range force sensor.

6.6 Validation Studies

To assess the accuracy of each individual component and combined components, we completed two studies to demonstrate the accuracy of 3D gaze tracking and combined 3D gaze and hand tracking. The methodology and results of these experiments are described in Chapters 7 and 8. With respect to 3D gaze tracking, other researchers have described a similar technology, the VICON-Eye Tracking Visualizer, but the accuracy and precision of this technology has yet to be reported (106).

Chapter 7 – Real-time Generation and Recording of a 3-Dimensional Gaze Vector from Synchronized Eye and Head Tracking

Introduction

Traditional eye tracking platforms utilize a camera or cameras directed towards the user's eyes in order to capture pupil position. Video is simultaneously obtained in a point-of-view fashion to acquire an image of the environment (109). Following calibration to align gaze position with an array of points typically at the same focal distance then allows the experimenter to monitor gaze position by tracking pupil position relative to the point of view video. Recently, new techniques (110-112) have been developed for assessing the focal point of an individual's vision with data obtained from a binocular eye tracker. This allows for a determination of gaze fixation in three dimensions by estimating gaze depth. Despite reasonable accuracy, these systems require a reliable determination of pupil position for both eyes in order to determine convergence. In addition, these techniques are sensitive to changes in head position. More recently, investigators were able to combine optical head position data and eye tracking data to determine a 3D gaze vector. This system allows the individual wearing the device to move freely and interact with objects in a complex 3D environment (106).

The Schack group (106) provided an excellent summary of previous systems that combined head and eye tracking technologies, along with their limitations. Many of these systems require the users' head to remain stationary or require a bulky binocular camera system that occludes a significant portion of an individual's field of view. Alternatively, desktop eye trackers or those integrated in monitors limit the area of tracking and can also be sensitive to head movement (113). The VICON-Eye Tracking Visualizer overcame these limitations by integrating monocular eye tracking with head tracking (114). This allows the subject wearing the device more freedom to interact with their environment compared to traditional 2D eye trackers that overlay gaze position on a 2D video frame. While the ability to measure fixations using this system was demonstrated, the accuracy and precision of the system was not reported.

Using a similar strategy, a method for determining gaze position in three dimensions from data obtained using a monocular eye tracker was developed. Two dimensional gaze data was combined with head tracking data to compute a 3D gaze vector.

Similar to the VICON-Eye Tracking Visualizer, this system allows an individual to move their head freely. The intended application for this system is the acquisition of 3D gaze information during open surgical tasks.

To validate the 3D gaze system described in Chapter 6 and determine the accuracy of 3D gaze determination, two experiments were conducted to measure gaze position when interacting with both a physical and virtual marker. An OptiTrack MoCap system (NaturalPoint, Inc., Corvallis, OR) was used to determine both head position and the position of both the physical markers and monitor for displaying the virtual markers.

Two measures of gaze accuracy were compared across a group of volunteer subjects – Euclidean distance from the marker to the gaze vector and gaze angle. We selected a viewing distance that approximates the average distance an individual will experience when interacting with a bimanual task while standing with their forearms parallel to the ground. This is the usual case for complex bimanual tasks such as surgery.

Methods

2D Eye tracking

Eye pupil position was captured using the Dikablis Essential Monocular Eye Tracker (Ergoneers GmbH, Manching, Germany). This consists of a head mounted eye tracker with a single infrared camera positioned just below the left eye (eye cam) and a camera mounted near the centre of the eye tracker positioned just above the subject's nose (field cam). A dedicated computer system was used to run the Dikablis Recording software provided with this eye tracker. This software automatically provides functionality for streaming all gaze data over a local area network using TCP/IP. Prior to streaming the 2D gaze tracking data to our recording software the Dikablis system is calibration at a particular distance. We performed manual calibration of the eye tracker at three distances corresponding to the three virtual array distances of 55 cm, 70 cm and 85 cm. This calibration involves having a subject fixate on the four corners of a rectangle. We used the corners of the monitor used to display the virtual array to accomplish this.

Head and Object Tracking

Optical tracking was accomplished using an array of six Flex-13 OptiTrack cameras. Motive version 1.6 software was used to calibrate the OptiTrack camera array and stream tracking data to our recording software using a TCP/IP protocol on the same computer (local loopback). Following calibration with a 250mm OptiTrack wand, we selected and tracked the position and orientation of the subject's head, the monitor used to display the virtual array and a gaze calibration triangle we refer to as the tracking object. In order to accurately measure head position and movement, we placed three IR makers on the rigid portion of the Dikablis eye tracker in an asymmetric triangular configuration. Similarly, three IR markers were placed on the corners of a widescreen monitor to form an asymmetric triangle. These markers were selected in the Motive software and labeled as rigid bodies in order to track their position in real-time.

3D Gaze vector generation

We implemented a custom software solution in C++ for synchronizing the 2D eye tracking data with the OptiTrack rigid body position data. Here we provide a description of the algorithm we employed for determining a 3D gaze vector via interpolation of a library of known calibration gaze vectors. The algorithm is broken up into two parts. Once the 3D gaze vector is determined it is saved, along with the synchronized and timestamped OptiTrack position data into a compressed XML file for offline analysis. Each 3D gaze vector is saved as a pair of two coordinates representing the origin and a second point on the gaze vector.

Algorithm Part 1

Here we obtain a sample of know gaze vectors. This is done during the calibration process when the subject is asked to fixate on the centroid of the tracking object in a variety of positions. During this step we are able to determine the gaze vector by calculating the vector between the subject's head and the tracking object. This vector is normalized so that the distance between the head and the tracking object is not a factor. We then apply the inverse transformation of the head to the vector. This ensures that the position and orientation of the head do not affect the calibration. Finally, we save the vector and the associated 2D coordinates of the current FOV position streamed from the Dikablis recording software. For demonstrating the accuracy of our system, at least 10 calibration positions spanning the subjects FOV were obtained.

Algorithm Part 2

Following acquisition of the calibration vectors interpolation can be used to find the three-dimensional gaze vector for all possible x and y coordinates of the FOV streamed from the Dikablis. This process is delineated in Figure 7.1. First we solve for the convex hull surrounding the known FOV coordinates. Next, we triangulate the points while maintaining the outer convex hull. Finally, we perform a linear interpolation between the three points of each triangle. Here, each point represents an x and y coordinate of a known FOV with a known three dimensional gaze vector associated with it. During the interpolation step we calculate the corresponding gaze vector for each position in the FOV. This allows us to quickly look up the gaze vector when given the FOV coordinates from the eye tracking software.

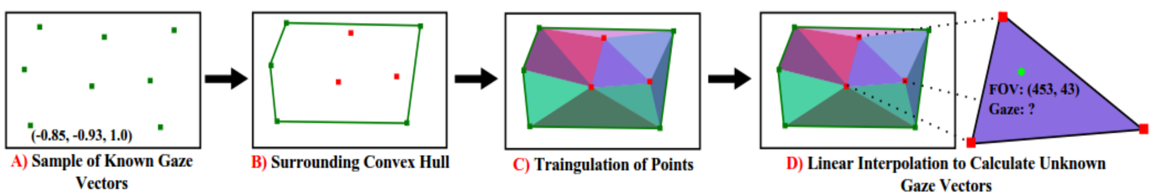

Figure 7.1 The calibration process for 3D gaze tracking. The surrounding rectangle represents the field of view (FOV). Each point represents a given position in the FOV with a known calculated three-dimension gaze vector.

Physical and Virtual Test Arrays

To determine the accuracy of our generated 3D gaze vector we compared the position of this vector to an array of known points. A physical array was constructed by elevating nine 12 mm passive IR reflective markers above a board with wooden dowels arranged in a 3x3 array (Figure 7.2). The height of each marker was randomized and spanned a range of zero to 30 cm above the surface of the board. This back edge of this platform was positioned 70 cm from the subject and the OptiTrack system was used to accurately determine the position of each marker as a point in 3D space.

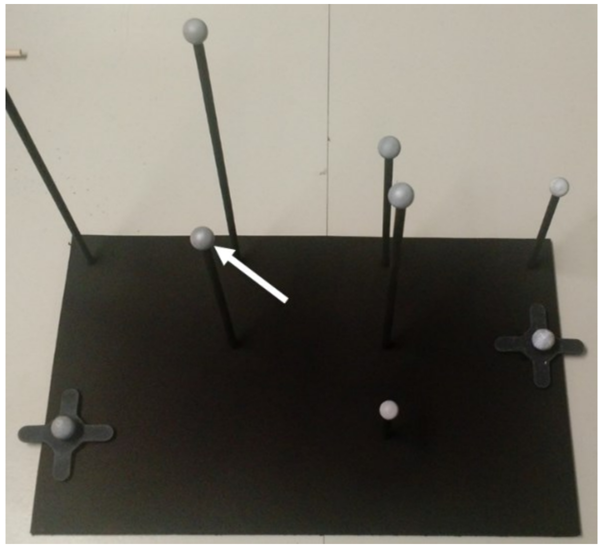

Figure 7.2 Physical array comprised of 9 passive IR markers position at various heights.

To display our virtual array, a MATLAB script was used to display a sequence of 15 circles on a 24” widescreen monitor. These virtual markers were evenly distributed in a 3x5 array. In order to determine the physical position of each virtual marker we used the dot pitch of our monitor (0.38mm) to translate the pixel address of each marker into a physical position measured from pixel (0,0) at the upper left corner of the monitor. The position of pixel (0,0) relative to the three IR markers placed on the bezel of the monitor was also measured. Figure 7.3 shows the positioning of the IR markers with respect to the screen edge.

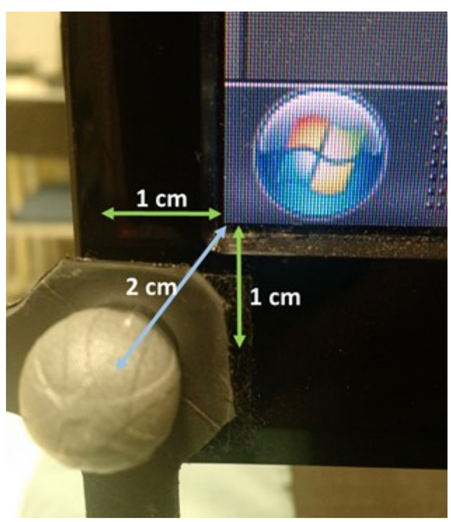

Figure 7.3 Placement of OptiTrack IR marker at corner of LCD monitor and three dimensional translations from centre of marker to pixel (0,0).

Validation Experiments

Institutional approval from the University of Alberta Research Ethics Board was obtained prior to enrolling any subjects. We recruited individuals with normal or corrected normal vision and ability to wear a head mounted eye tracker over the left eye. We excluded any individuals with any health condition affecting vision or eye movement e.g. strabismus, nystagmus, previous globe injury, etc.

Subjects performed all testing from a seated position. We completed the 2D calibration of the Dikablis using the corners of the monitor at an initial distance of 55 cm. Next, the IR markers on the monitor, eye tracker and calibration triangle were selected and registered as rigid bodies in the Motive software. A set of known calibration vectors were then recorded by moving the calibration triangle throughout the subject's field of view. Subjects were instructed to focus their fixation on the centroid of the calibration triangle as it moved to each calibration position.

Each virtual marker was displayed individually for 3000 ms and plotted in a linear sequence from left to right along each row starting at the upper left of the screen. An audio file was used to provide a standardized cue for each subject to fixate on the next marker in the physical array.

Analysis

Prior to determining the gaze distance and angle error from our validation experiments we executed a MATLAB script in order to obtain spatial coherence between the different coordinate systems. The origin of the eye vector was translated to the approximate back of the subject's retina from the centroid of the tracked triangle representing head position. We used the average reported diameter of an adult eye globe in this calculation.(115) The data was visualized using MATLAB 3D graph functions in order to confirm that spatial coherence had been obtained (Figure 7.4 and 7.5). Euclidean distance was determined by solving for the minimal distance between the line represented by the 3D gaze vector and a point in 3D space represented by a given IR marker or point on the virtual array (116). The gaze angle error was determined by comparing the ideal gaze angle represented by a vector originating from the eye to the tracked marker position.

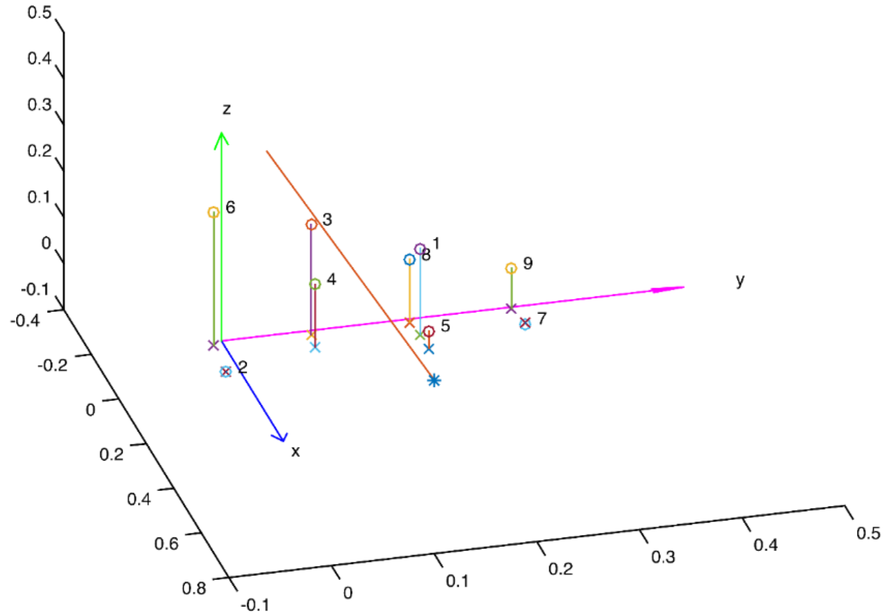

Figure 7.4 MATLAB visualization of the physical IR marker array and gaze vector in line with marker 3, each axis represents distance in meters.

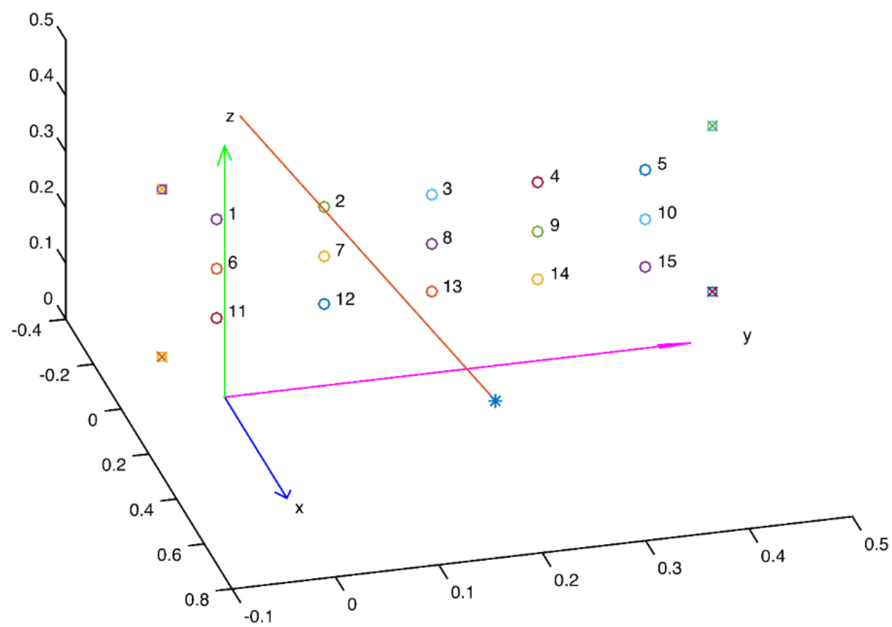

Figure 7.5 MATLAB visualization of the virtual marker positions displayed on the LCD monitor and gaze vector in line with marker 2, each axis represents distance in meters.

Results

MATLAB visualizations of the physical array and virtual arrays are depicted in Figures 7.4 and 7.5. In both, the origin of the gaze vector represented by $*$. The 3D gaze vector is represented as a 1 m long line extending towards the gaze targets represented by $\circ$. Distance and gaze angle error were determined at three distances for each subject using the virtual array and at a single distance for the physical array. Figure 7.6 delineates the Euclidian distance and visual angle error for a single subject. The pooled mean distances and angle error for each physical and virtual marker are summarized in Table 7.1 and Table 7.2 respectively. The overall accuracy for 3D gaze from data pooled for all three viewing distances was 2.78 ± 1.6 cm for the physical array and 2.49 ± 0.95 cm for the virtual array.

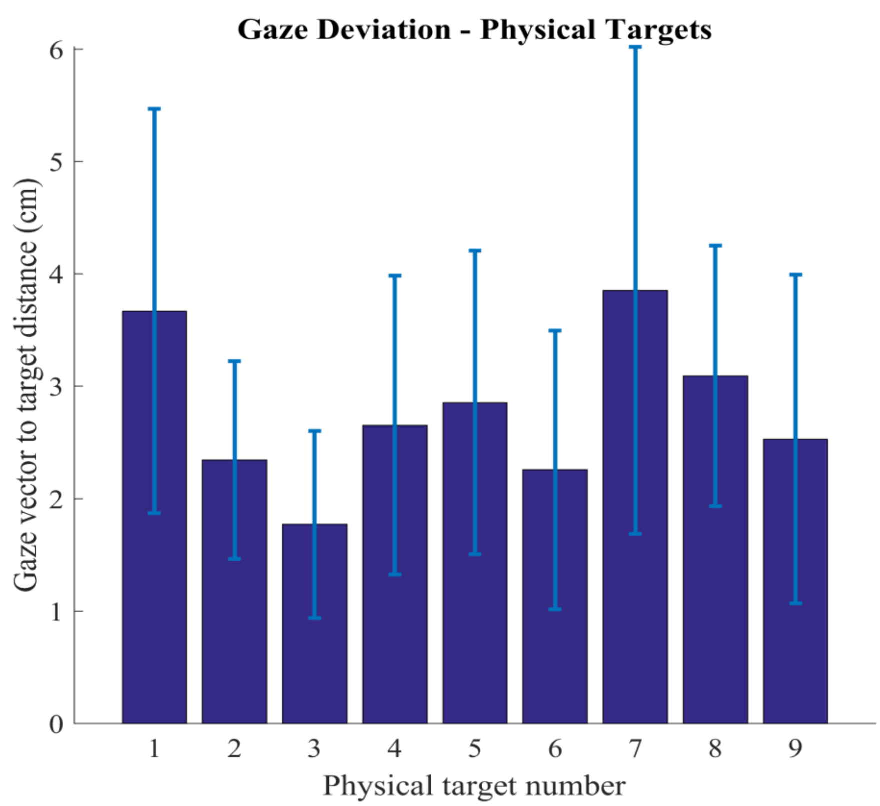

Figure 7.6 Euclidian distance from each of the physical markers for a single subject.

Table 7.1: Mean gaze distance and angle error from the OptiTrack IR marker position in the physical array

Physical Array Marker	Mean Euclidian Distance (cm)	Gaze Angle Error (º)
1	3.5 ± 2.0	3.0 ± 1.7
2	2.3 ± 1.0	2.1 ± 0.9
3	1.8 ± 1.0	1.5 ± 0.8
4	2.7 ± 1.5	2.4 ± 1.3
5	2.9 ± 1.5	2.4 ± 1.4
6	2.3 ± 1.4	2.0 ± 1.2
7	3.9 ± 2.4	3.6 ± 2.3
8	3.1 ± 1.3	2.4 ± 1.0
9	2.5 ± 1.6	2.1 ± 1.4

Table 7.2: Mean gaze and angle error from virtual marker position plotted on LCD monitor

Virtual Array Marker	Mean Euclidian Distance (cm)	Gaze Angle Error (º)
1	1.5 ± 0.2	1.3 ± 0.2
2	1.3 ± 0.4	1.0 ± 0.3
3	2.4 ± 0.7	1.9 ± 0.5
4	3.8 ± 1.8	3.1 ± 1.6
5	2.8 ± 0.2	2.2 ± 0.1
6	2.7 ± 0.7	2.2 ± 0.5
7	1.6 ± 0.5	1.4 ± 0.5
8	1.6 ± 0.3	1.3 ± 0.3
9	2.7 ± 1.3	2.1 ± 0.7
10	2.4 ± 0.1	1.9 ± 0.1
11	3.0 ± 0.5	2.4 ± 0.4
12	3.9 ± 1.4	3.0 ± 1.0
13	2.5 ± 0.8	2.0 ± 0.6
14	2.7 ± 0.7	2.2 ± 0.4
15	2.5 ± 1.9	1.9 ± 0.6

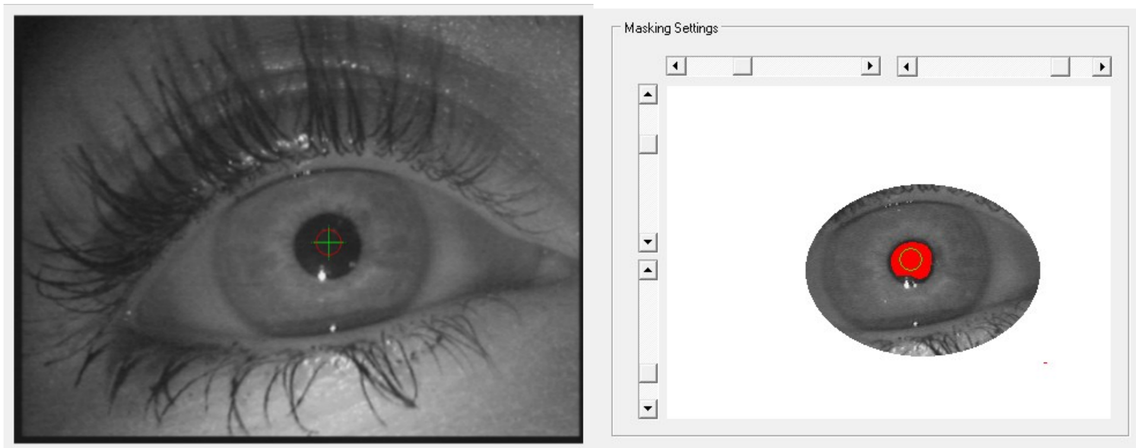

Figure 7.8 Dikablis Recorder software showing pupil detection overlay and reflection of IR LED at the inferior aspect of the subject's pupil.

Discussion

We have successfully demonstrated the development of a 3D gaze tracking system by integrating traditional 2D eye tracking technology with optical head tracking. Our results demonstrate the accuracy of this system for viewing objects at a focal distance of 55-85 cm. Our synchronization and recording software allows us to retain all of the information streamed from the Dikablis recorder regarding 2D fixation, pupil size, and timestamp information. While we did no demonstrate it here, this information could be used to determine fixation duration and other traditional eye tracking metrics.

Inspection of the gaze accuracy data revealed decreased accuracy for both physical and virtual markers near the periphery of the subject's vision. The Dikablis eye tracker we employed utilizes a linear calibration method for both X and Y gaze coordinates. At more extreme gaze angles, where the eye is focused on objects to the extreme left or right for example, there are some torsional movements of the eye (115). Based on the calibration method it is likely that these movements resulted in a less accurate determination of gaze direction. While our method of generating multiple gaze vectors likely compensated for some of this error, we also performed a linear calibration based on a convex hull for both x and y coordinates. Future optimization of this method might involve measuring and correcting for torsional or non-linear gaze deviation at more extreme gaze angles.

A decrease in gaze accuracy for markers below the horizon of some subject's gaze was also measured. This was likely due to an inaccurate determination of the pupils centre by the proprietary Dikablis Recorder software. Despite numerous attempts to optimize the position of the eye tracking camera on the Dikablis, we were sometimes forced to position the camera closer to the individual's gaze horizon. This resulted in a reflection of the IR emitter just below the eye camera in the image of the subject's eye and pupil as in Figure 5.7. Following the initial manual 2D calibration of the eye tracker, there was a discrepancy in the plotted gaze position for targets towards the lower portion of a subject's field of view when this reflection in the pupil was significant. Despite attempts to reduce the intensity of the IR LED intensity, this appeared to reduce the accuracy of the system for some subjects.

Visual inspection of the IR spectrum in the Dikablis Recorder software during pupil detection calibration demonstrated a significant amount of noise. There appeared to be an oscillation in the overall amplitude of the spectrum on the order of 1-2 Hz. The use of an AC power conditioner or DC power source appeared to reduce the intensity of this noise. For all of the accuracy testing in this study we chose to use a DC power source from a lithium polymer battery.

The synchronization and recording system previously described in Chapter 6 allows for the eye tracking acquisition system to be independently developed and optimized as the data is generated on a separate system. This will facilitate future improvements to pupil detection and overall accuracy.

Chapter 8 Validation of a Combined Eye and Hand Tracking System

Introduction

Investigators have previously used eye tracking technology to investigate the relationship between vision and motor movement or planning (117). However, to our knowledge, no previous studies have combined eye and hand tracking in a 3-dimensional environment. By utilizing the kinematic system described in Chapter 6, we sought to demonstrate how gaze overlap with hand position in 3D space could be used to measure eye hand interaction during a complex bimanual task. We chose a rudimentary surgical task, the placement of three simple interrupted sutures in simulated tissue, to measure eye-hand interaction objectively, and investigate the use of these metrics for determining surgical dexterity or experience. We recruited novice medical students with limited previous suturing experience as well as senior surgical resident trainees to determine if portions of the simulated task could be used to predict experience or ability. Once acquired, manual and computer assisted automatic segmentation of the task can be performed to compare the time required for each particular portion of the task, a measure of movement economy. These metrics may form the basis for new objective measures of performance and the generation of procedure specific feedback for surgical trainees. As discussed in previous chapters, examples of segmentation strategies for decomposing surgical maneuvers include Hidden Markov Models (HMM) (96,118) and affine speed (24). This chapter describes the successful collection of synchronized 3D gaze, hand motion and instrument forces during a simulated surgical task. Once a more complete data set is obtained, other measurements including proactive eye gaze – eye movements that precede motor movement (119), will be compared to determine if this is substantially different between novice and experienced trainees. The current results report the accuracy of both the 3D gaze and hand tracking technologies and achievement of both temporally and spatially coherent data.

Methods

Health ethics board approval was obtained from the University of Alberta Health Ethics Board. Senior medical students who had completed their clinical rotation in surgery as well as senior general surgical residents in Post-Graduate Year 3 (PGY3) or greater were recruited. Participants requiring vision aids were requested to wear contact lenses in order to improve eye tracking performance.

The system described in Chapter 6 for obtaining synchronized 3D eye gaze and hand motion was calibrated and each participant was fitted with the eye tracker and surgical gloves with embedded electromagnetic tracking. Adson forceps with integrated force sensors were provided for the suturing task. A Canon digital SLR camera was used to acquire audio and video of each participant performing the simulated task. In order to assist with modeling of each subject's hands, the positions of each MCP, PIP and DIP joint in both hands were marked with a permanent marker and photographed adjacent to a ruler for scale.

Prior to completing the simulated surgical task, each participant was asked to direct their gaze to the center of the starting positions for each instrument and place their right index finger at the same position. This generated the data necessary for validating the accuracy of the synchronized hand and eye tracking system.

Suturing Task

each participant was asked to place three interrupted sutures in a marked location of 3-Dmed (3-Dmed, Franklin, OH) synthetic skin. Instruments including forceps with integrated force sensors, a needle driver, and scissors were provided. Each instrument was placed on the board in a marked and labeled starting position. Figure 8.1 shows a photograph of the suturing simulation work area. Participants were instructed to perform an instrument tie for making their surgical knots and to place each instrument back in the designated starting position when not in use. This resulted in a sequence of grasping the forceps and driver initially, returning the forceps second, and returning the driver before grasping the scissors to complete the task. This forced each participant to decompose the major steps of the suturing task into discrete and easy to identify segments.

Accuracy Validation Task

To determine the accuracy of the eye tracker when positioned over the surgical simulation and of the hand EM tracking system, the center of each of the instrument home positions was taken as a physical target. This is similar to the methodology employed in Chapter 7 where a physical array of IR markers was employed. Video of the accuracy task was used to determine the approximate time corresponding to the mid-point of gaze fixation and finger positioning over a particular target. A 250 ms sample of gaze data was taken around each time point, filtered using MATLABS median filter (*medfilt1*), and the average Euclidean distance and standard deviation was determined. A similar 250 ms sample of position data for the EM tracker corresponding to the index finger of the right hand (RightD2) was used to determine the accuracy of the EM system. The EM data did not require filtering.

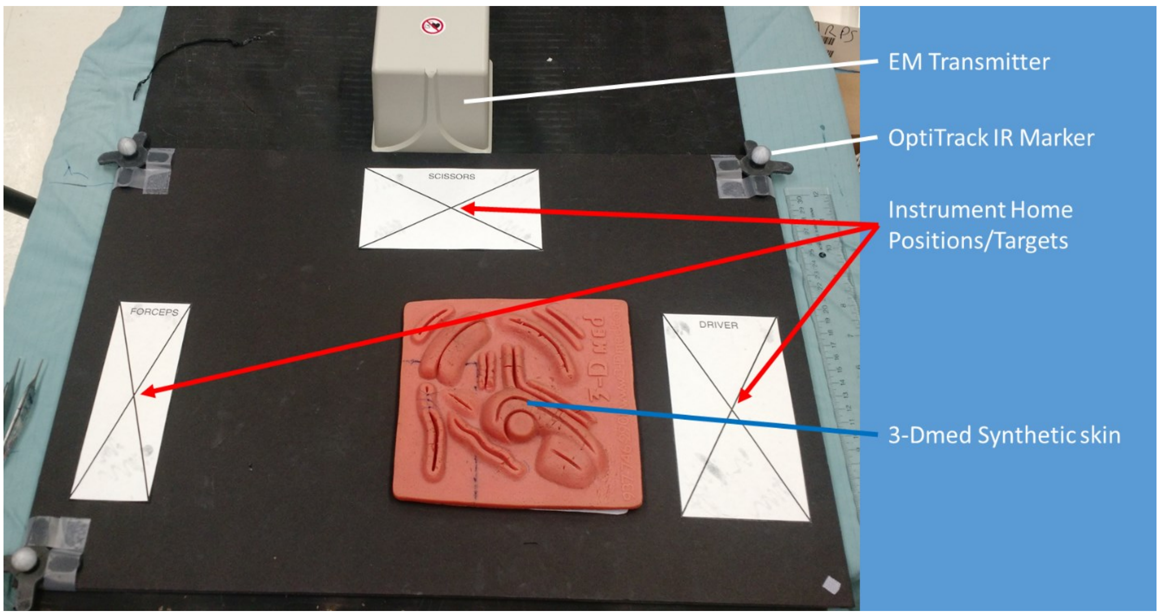

Figure 8.1 Experimental setup showing position of EM transmitter with respect to the simulated surgical task and instrument home positions/targets

Results

Calibration of the OptiTrack MoCap system typically resulted in a triangulation residual mean error of ≤ 0.2 mm as reported by Motive software. Calibration of the Dikablis was completed as described in Chapter 6 and 7 with 25 gaze vectors using a calibration target composed of 3 IR markers arranged in an asymmetric triangle. The TrakSTAR EM tracking system was aligned with the OptiTrack frame of reference by inverting a designated finger and placing it over the three IR markers representing the ground plane. Following capture of the accuracy and suturing task, the synchronized data was saved to file in XML format. A MATLAB script was then used to parse the raw data and load the necessary elements for further analysis into MATLAB arrays.

Visualization of the raw data was accomplished by reconstructing both the 3D gaze vector and hand model from the eye tracking and EM data respectively. While the finger tips were plotted as points in 3D space, the palm was drawn as a six sided polygon and rotated according to the quaternion rotation data encoded by the EM tracking system for the sixth marker on each hand. This marker corresponds to the dorsal aspect of the midpoint of the third metacarpal.

Figure 8.2 shows the MATLAB 3D visualization of the hands and eye gaze vector with respect to the instrument home positions and accuracy targets. Here the corrected origin of the 3D gaze vector is represented by $*$ and a 1 m line extending towards the fixation point. The intersection of this vector with the ground plane can be used to determine the fixation point if the physical location of the surface is known. For testing the accuracy of the eye gaze vector, the Euclidean distance between a line (the gaze vector) and a point (the marked center of each target) was calculated based on Eqn. 8.1, where $\mathbf{x}_1 = (x_1, y_1, z_1)$ and $\mathbf{x}_2 = (x_2, y_2, z_2)$ are two points on the vector and $\mathbf{x}_0 = (x_0, y_0, z_0)$ is a point in Euclidean space, $\mathbb{R}^3$. Both video or inspection of the distance over time curve can be inspected to determine when the gaze is fixated on a particular target. Figure 8.3 shows the distance between the eye vector and each target over the course of the experiment for a single subject. Figure 8.4 is a graph of gaze deviation for all 3 subjects. The participants were prompted to move from one target to the next in sequence so the distance approaches zero in sequence, as expected. Table 8.1 reports a summary of the 3D gaze accuracy for three subjects.

$$D = \frac{|(\mathbf{x}_0 - \mathbf{x}_1) - (\mathbf{x}_0 - \mathbf{x}_2)|}{|(\mathbf{x}_2 - \mathbf{x}_1)|} \qquad (8.1)$$

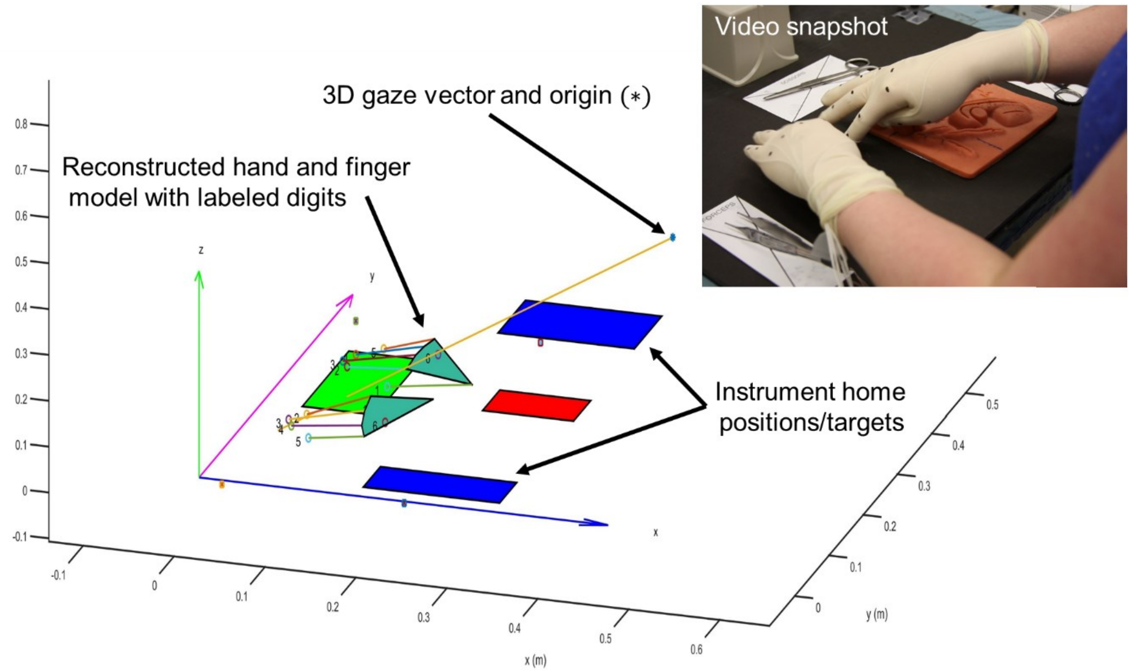

Figure 8.2 MATLAB visualization of the accuracy experiment showing the position of the home instrument positions and surgical task (rectangles), reconstructed hand models and 3D gaze vector.

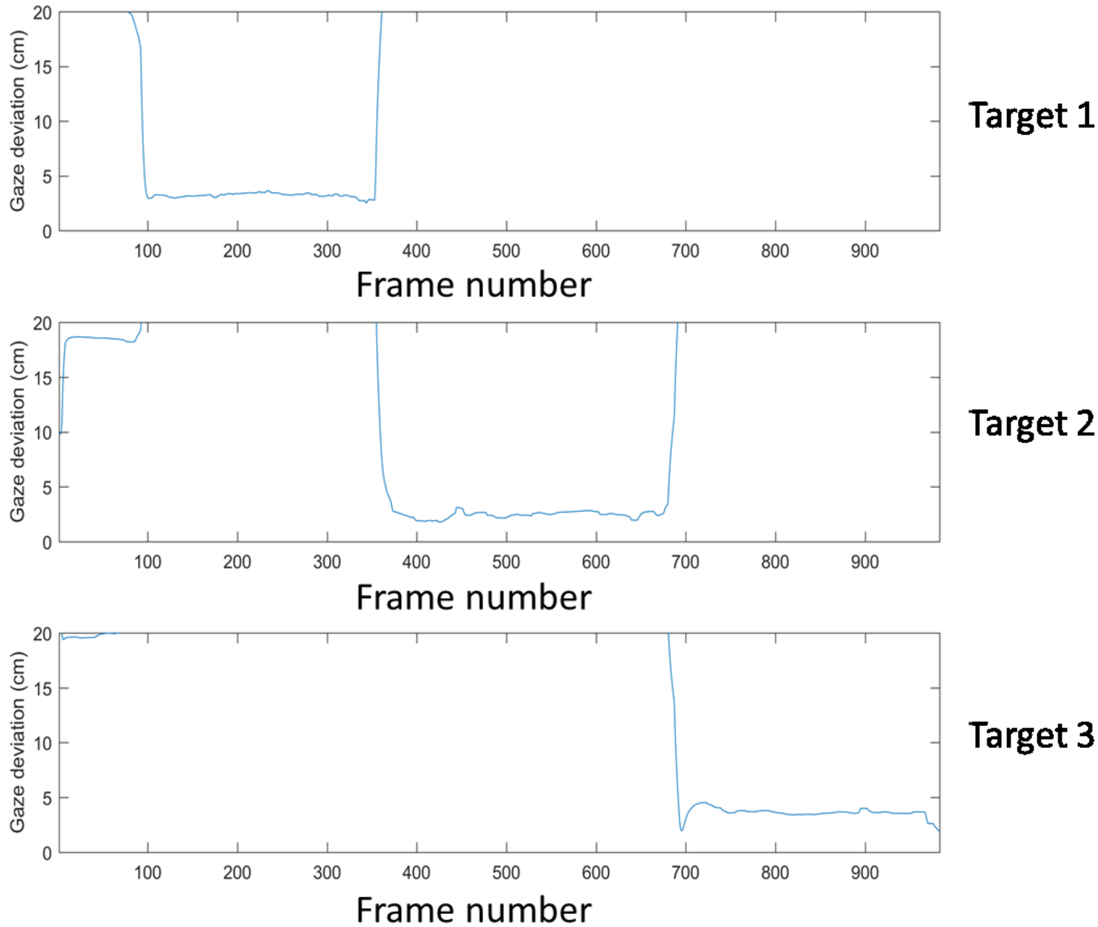

Figure 8.3 Gaze vector distance from three target positions during accuracy test. Participants fixated on targets in order from top to bottom.

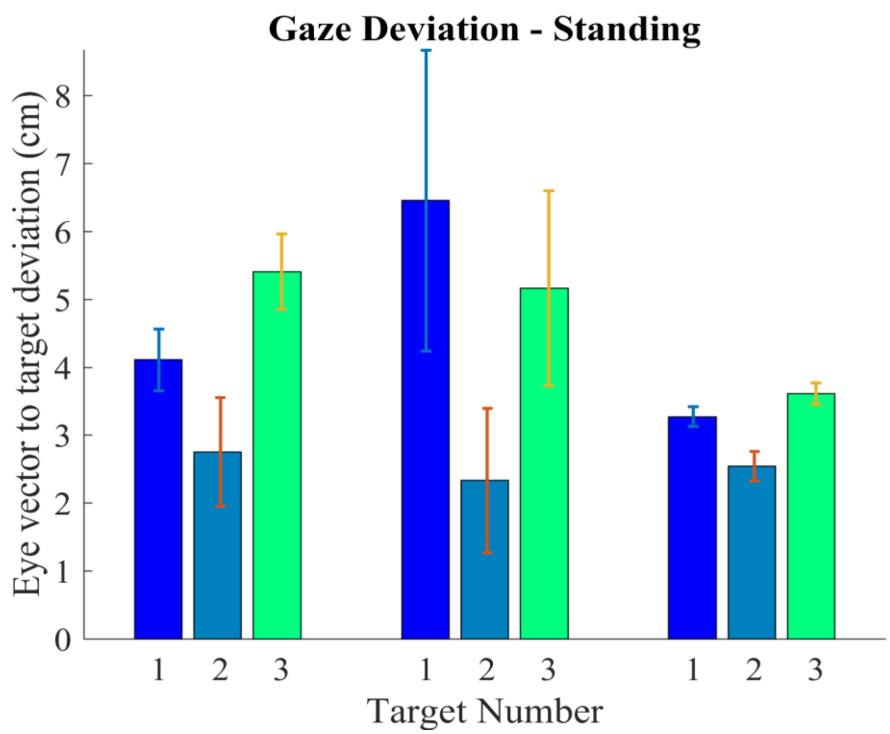

Figure 8.4 3D gaze deviation from target center for 3 subjects.

Table 8.1: Gaze deviation for three test targets (n=3)

Target	Euclidean distance – gaze vector to target ± SD (cm)
1	4.6 ± 1.3
2	2.5 ± 0.8
3	4.7 ± 0.9

In a similar fashion, the EM data for the right second digit (index finger) was evaluated for accuracy. The three dimensional coordinates for this sensor were translated +1 cm in the MATLAB z axis to correct for the placement of the sensor over the fingernail and not the finger pad. The Euclidean distance, D, between two points was then determined as per Eqn. 8.2, where x_n, y_n, z_n are the coordinates of a point in Euclidean space, $\mathbb{R}^3$ (120). Figure 8.5 shows the distance between the right index finger and the three accuracy targets for a single subject. Figure 8.6 delineates the finger position accuracy for all three subjects. A summary of the accuracy data is reported in Table 8.2.

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \qquad (8.2)$$

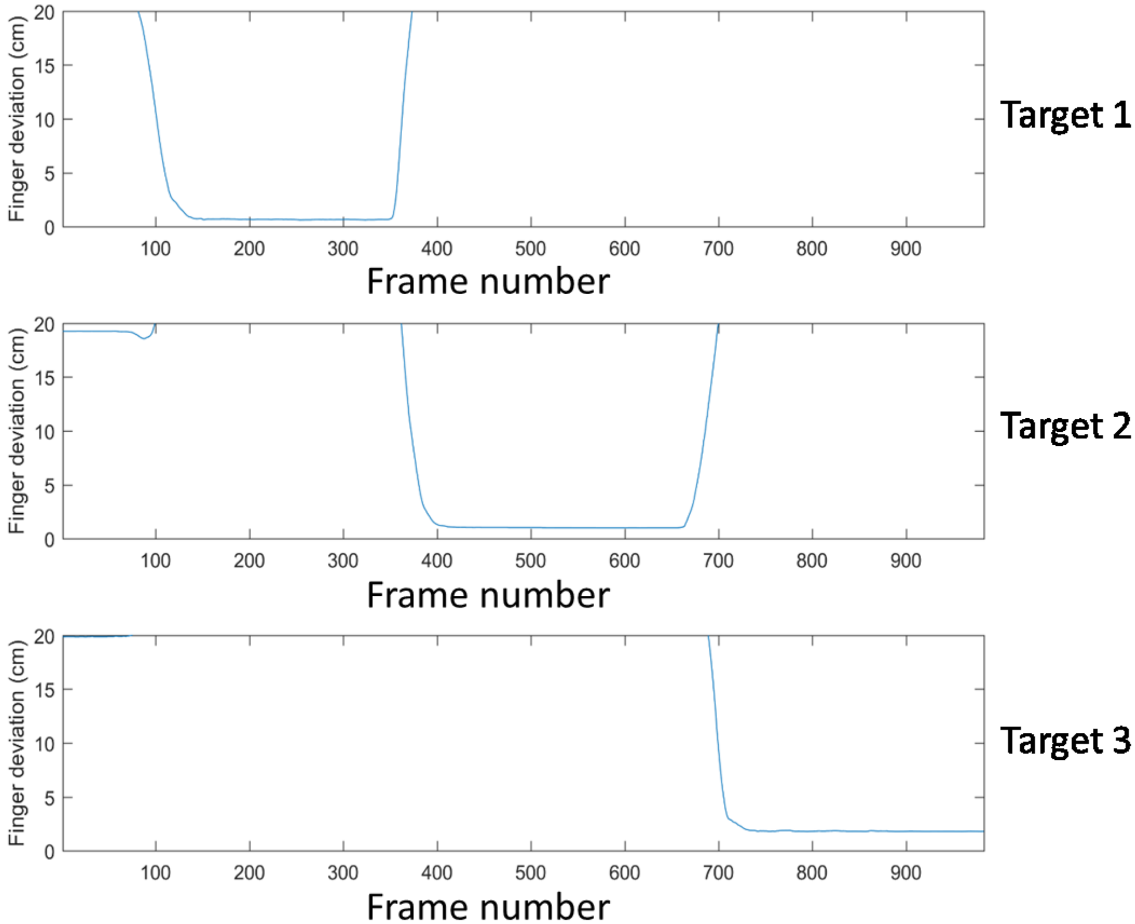

Figure 8.5 Distance from tip of right index finger to three target positions during accuracy test.

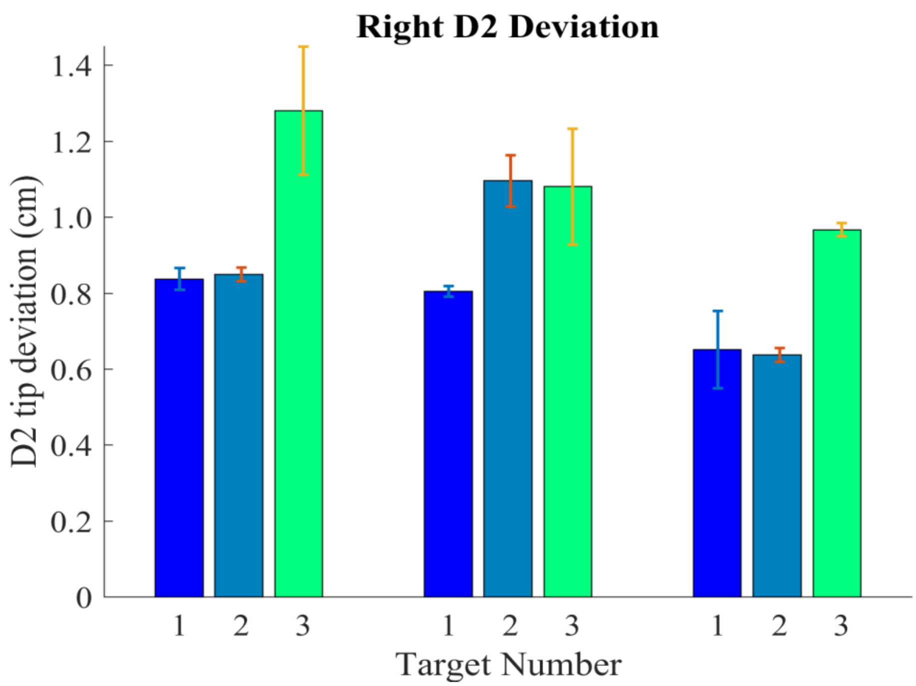

Figure 8.6: Finger position (EM) deviation from target center for 3 subjects.

Table 8.2: Finger position accuracy for three test targets (n=3)

Target	Euclidean distance – Right D2 tip to target ± SD (cm)
1	0.76 ± 0.06
2	0.86 ± 0.04
3	1.11 ± 0.13

Overall, the pooled gaze and EM finger tip deviations were 4.0 ± 1 cm and 0.91 ± 0.09 cm, respectively.

Discussion

We have successfully demonstrated the acquisition of synchronized 3D gaze and hand motion while performing a complex bimanual task. Validation of the 3D gaze and finger position acquisition resulted in an overall accuracy of 4.0 ± 1 cm and 0.91 ± 0.09 cm, respectively.

This system was designed for acquiring kinematic data from individuals performing surgical tasks, but this technique will likely be useful to industry or researchers interested in studying the acquisition of fundamental motor skills or optimizing training strategies for complex bimanual tasks. We utilized many more sensors on each hand compared to the ICSAD in an attempt to record highly detailed motion information necessary for characterising specific hand gestures. We have little doubt that this system will be able to discriminate between surgical trainees with different levels of experience using similar descriptive statistics. However, our aim is to provide additional information regarding specific portions of a maneuver. To analyze the kinesthetic of a surgical procedure, a much richer data set is required in order to determine the variable orientation of each hand and finger during a specific subtask. In addition, our 3D gaze acquisition system can acquire precision visual information regarding gaze behaviour. Previous research (121) has shown that optimizing gaze strategies or *gaze training* can have a demonstrable effect on surgical performance. By far, the most significant advancement presented in this study is the ability to record both hand motion and eye gaze together. Spatially and temporally coherent data will permit the future evaluation of eye hand interaction and coordination. This has not been achieved with the same precision and accuracy as presented here.

While this system can provide highly detailed information regarding an individual's motion and eye gaze behaviour, the calibration process is still quite cumbersome and consists of three calibration maneuvers for generating a 3D gaze vector and two for obtained spatially coherent position and rotation data from the EM system. This calibration process can take upwards of 30 minutes to complete for one subject. Previous studies (106) have criticized similar technologies because of the complexity of the calibration procedure. The monocular eye tracker is the least reliable instrument in this system. While data was obtained for the surgical task described above, visualization and playback revealed significantly aberrant eye tracking. For this reason, our focus shifted to improving the reliability of pupil detection. Chapter 9 provides a description of ongoing attempts to develop an in-house technique for more reliable eye tracking. As previously stated, the modular nature of our system allows us to develop and improve each technology, e.g. eye tracking, independently.

Chapter 9 – Conclusion and Future Research Directions

This thesis provides an overview of some of the latest techniques for assessing technical dexterity in surgery. A solid understanding of the prevailing theories in motor skill acquisition serves as a valuable tool for hypothesis generation and the design of experiments for testing these theories. We presented an innovative system designed to acquire eye, hand and instrument motion and kinematic data for use with computer aided assessment tools we hope to further develop. Our system solves the challenge of collecting this type of data in the open surgical domain. We are optimistic that computer aided analysis previously applied to kinematic data in MIS can be applied to this data set. While we have yet to conclusively demonstrate that a kinesthetic approach can be used to objectively evaluate surgical trainees, focusing on this technique during training appears to have a benefit. Along the same lines, there is strong evidence for the use of laparoscopic simulation devices, those that emulate remote manipulation, for improving laparoscopic skill.

The first empirical study in Chapter 4 showed how video gaming can supplement the current simulation based training of novice surgical trainees. This study represents the largest randomized control trial to date on the effect of video gaming on laparoscopic skill acquisition. Some interesting findings emerged from the repeated measure analysis and learning curves that were constructed for the laparoscopic practice task. The learning effect of even a brief exposure to laparoscopic practice was demonstrated. This is an important factor to be cognisant of in all human performance studies that involve repeated exposure to a particular task. In conclusion, lap box training remains the superior training method for laparoscopic novices. The effect is most pronounced with highly visuospatial intensive tasks exercises such as the FLS peg transfer task. However, video gaming appears to reinforce laparoscopic skill acquisition when combined with lap box practice, as evidenced by parallel learning curves.

Chapter 3 provided an overview of the current methods available for motion analysis in a variety of surgical environments (open, MIS, robotic). Motion analysis can be used to collect kinematic and kinesthetic data related to surgical expertise. Laparoscopic techniques benefit from restricted DOF and well established instrument tracking technology that facilitate the application of computer based analysis methods.

Language models are a method for automatic segmentation of hand motion into surgemes (surgical gestures) that can be compared with an internal standard. Combining data from multiple sources such as hand and finder position, gaze fixation and force data as was done in MIS based studies should be able to generate useful relationships (e.g. hand-tool-tissue interactions) specific to a particular gesture or surgeme. Depending on the reliability of the acquired motion data, a HMM developed for a particular individual can be augmented with additional motion or force data. Despite the increased complexity of hand and instrument movement in open procedures, we successfully applied a language model to a non-surgical transfer task (Chapter 5). Significant occlusion from optical hand motion capture motivated us to pursue a EM hand tracking method.

The surgical hand motion and eye tracking system described in Chapters 6 represents a novel method for acquiring synchronized hand motion and eye tracking data. The combination of head and eye tracking to generate a 3D gaze vector has previously been described, but the empirical study in Chapter 7 is the first to demonstrate the accuracy of our interpolation technique. The achievement of temporally and spatially coherent data acquisition in a single system is significant. This study demonstrated the generation of a 3D gaze vector with an accuracy of 2.5-2.8 cm using both a virtual and physical array of gaze targets. Despite some limitations, this system can be used to acquire eye tracking data while users interact with a 3D environment at a variety of depths of field. Testing of the combined eye and hand capture system with the participant in a standing position demonstrated gaze accuracies of 4.0 ± 1 cm and fingertip accuracies of 0.9 ± 0.09 cm. These figures indicate that the system is able to capture eye gaze and finger position at a reasonable accuracy for future analysis of hand gesture and eye gaze behaviour. These experiments also served to identify challenges with the design and calibration of the equipment. The focus of future development of this system is discussed below.

Future Work

Based on the results presented in Chapters 7 and 8, our focus will center on improving the reliability and robustness of the eye tracker prior to completing data collection for more sophisticated simulated surgical procedures.

As discussed in Chapter 8, the reliability of the monocular eye tracker was highly variable between subjects. Under ideal circumstances, such as those replicated in Chapter

7, the eye tracker could produce impressive results. However, when we transitioned to a standing position over a surgical task, pupil occlusion became significant. The Dikablis Recorder software used in this system utilizes a threshold algorithm for determining pupil position from the IR video positioned just below a subject's left eye. Unfortunately, harsh shadows and the reflection of the IR LED just below the IR camera significantly affected the ability of the software to identify the center of the pupil reliably. However, the Dikablis is not a closed system, and the video from the eye camera can be captured by way of a video capture device independent of the Dikablis software. This has enabled us to record raw video of the eye for multiple subjects, emulating some of the occlusion from eyelashes and different lighting conditions that result in inaccurate pupil detection (Figure 9.1). With the aid of collaborators from the Department of Computer Science, we are implementing a machine learning based method for pupil detection. A recent study (122) has demonstrated the robustness of this technique. We hope this will significantly improve the reliability of the eye tracker and also eliminate some of the calibration steps required in order to generate the 3D gaze vector.

Following the implementation of an improved method for acquiring 3D gaze, the system described in Chapters 6-8 will be applied to surgical trainees. In addition to the rudimentary suturing task described in Chapter 8, a small bowel anastomosis and hand typing task will be completed. The year of training of each resident, their self-reported surgical experience based on number of cardinal procedures and performance on a Mental Rotation Test (MRT-A) (123) will be assessed prior to completion of the simulated tasks. To familiarize junior trainees with a small bowel anastomosis on an animal model, a standardized video has been prepared which provides and overview of the technique (Figure 9.2). High-definition video will be acquired during the simulated tasks, and a blinded expert reviewer will assess each participant based on an adapted OSATS scoring checklist as described in Chapter 4. In addition, the hand motion data acquired from the EM system can be used to generate the ICSAD (51) metrics (path length, number of movements, overall task time) for comparison. These results can then be used to validate any future objective assessment measures derived from the surgical motion and eye tracking system.

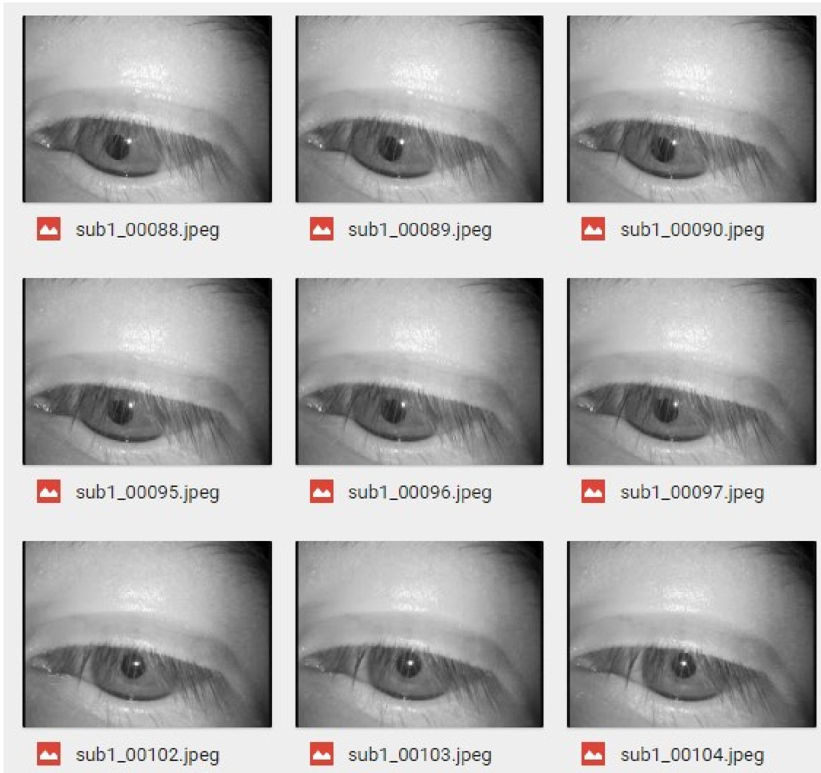

Figure 9.1 Unprocessed video frames captured from the Dikablis eye tracker.

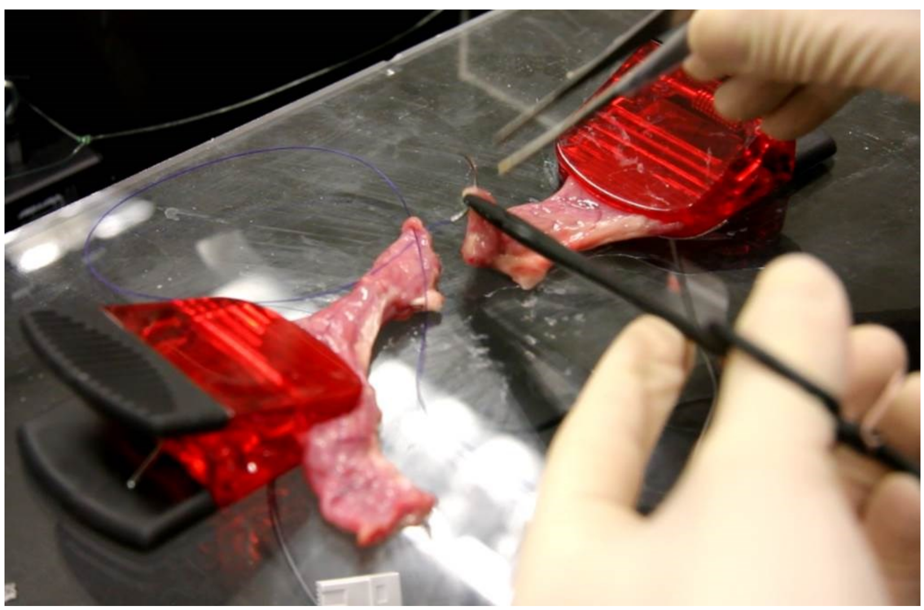

Figure 9.2 Video snapshot of standardized video providing instructions for performing a small bowel anastomosis.

References

(1) Mayberry JC. Residency reform Halsted-style. J Am Coll Surg 2003 Sep;197(3):433-435.

(2) Van Bruwaene S, Schijven MP, Miserez M. Assessment of procedural skills using virtual simulation remains a challenge. J Surg Educ 2014 Sep-Oct;71(5):654-661.

(3) Darzi A, Smith S, Taffinder N. Assessing operative skill. Needs to become more objective. BMJ 1999 Apr 3;318(7188):887-888.

(4) Mullen MG, Salerno EP, Michaels AD, Hedrick TL, Sohn MW, Smith PW, et al. Declining Operative Experience for Junior-Level Residents: Is This an Unintended Consequence of Minimally Invasive Surgery? J Surg Educ 2016 Apr 5.

(5) Kennedy AM, Boyle EM, Traynor O, Walsh T, Hill AD. Video gaming enhances psychomotor skills but not visuospatial and perceptual abilities in surgical trainees. J Surg Educ 2011 Sep-Oct;68(5):414-420.

(6) Reznick RK, MacRae H. Teaching surgical skills--changes in the wind. N Engl J Med 2006 Dec 21;355(25):2664-2669.

(7) Barden CB, Specht MC, McCarter MD, Daly JM, Fahey TJ,3rd. Effects of limited work hours on surgical training. J Am Coll Surg 2002 Oct;195(4):531-538.

(8) Leach DC. A model for GME: shifting from process to outcomes. A progress report from the Accreditation Council for Graduate Medical Education. Med Educ 2004 Jan;38(1):12-14.

(9) Meier AH. Running a surgical education center: from small to large. Surg Clin North Am 2010 Jun;90(3):491-504.

(10) Torkington J, Smith SG, Rees BI, Darzi A. Skill transfer from virtual reality to a real laparoscopic task. Surg Endosc 2001 Oct;15(10):1076-1079.

(11) Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg 2002 Oct;236(4):458-63; discussion 463-4.

(12) Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med 2004 Oct;79(10 Suppl):S70-81.

(13) Fonseca AL, Evans LV, Gusberg RJ. Open surgical simulation in residency training: a review of its status and a case for its incorporation. J Surg Educ 2013 Jan-Feb;70(1):129-137.

(14) Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg 1997 Feb;84(2):273-278.

(15) Doyle JD, Webber EM, Sidhu RS. A universal global rating scale for the evaluation of technical skills in the operating room. Am J Surg 2007 May;193(5):551-5; discussion 555.

(16) Carr S. The Foundation Programme assessment tools: an opportunity to enhance feedback to trainees? Postgrad Med J 2006 Sep;82(971):576-579.

(17) Gofton WT, Dudek NL, Wood TJ, Balaa F, Hamstra SJ. The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE): a tool to assess surgical competence. Acad Med 2012 Oct;87(10):1401-1407.

(18) Memon MA, Brigden D, Subramanya MS, Memon B. Assessing the surgeon's technical skills: analysis of the available tools. Acad Med 2010 May;85(5):869-880.

(19) Adams JA. A closed-loop theory of motor learning. J Mot Behav 1971 Jun;3(2):111-149.

(20) Margolis JF, Christina RW. A test of Schmidt's schema theory of discrete motor skill learning. Res Q Exerc Sport 1981 Dec;52(4):474-483.

(21) Fitts PM, Posner MI. Human performance. 1967.

(22) Kirkman MA. Deliberate practice, domain-specific expertise, and implications for surgical education in current climes. J Surg Educ 2013 May-Jun;70(3):309-317.

(23) Huang E, Chern H, O'Sullivan P, Cook B, McDonald E, Palmer B, et al. A better way to teach knot tying: a randomized controlled trial comparing the kinesthetic and traditional methods. Am J Surg 2014 Oct;208(4):690-694.

(24) Cifuentes J, Boulanger P, Pham MT, Moreau R, Prieto F. Automatic gesture analysis using constant affine velocity. Conf Proc IEEE Eng Med Biol Soc 2014;2014:1826-1829.

(25) J. A. Cifuentes Quintero. Development of a new technique for objective assessment of gestures in mini-invasive surgeryInstitut National des Sciences Appliqu´ees de Lyon - Universidad Nacional de Colombia, 2015; 2015.

(26) Kneebone R. Simulation, safety and surgery. Qual Saf Health Care 2010 Oct;19 Suppl 3:i47-52.

(27) Henry FM. Specificity vs. generality in learning motor skill. Classical studies on physical activity 1968:328-331.

(28) Proteau L, Marteniuk RG, Girouard Y, Dugas C. On the type of information used to control and learn an aiming movement after moderate and extensive training. Human Movement Science 1987;6(2):181-199.

(29) Proteau L. On the specificity of learning and the role of visual information for movement control. 1992.

(30) Rittenhouse N, Sharma B, Sonnadara R, Mihailidis A, Grantcharov T. Design and validation of an assessment tool for open surgical procedures. Surg Endosc 2014;28(3):918-924.

(31) Ericsson KA. The acquisition of expert performance: An introduction to some of the issues. 1996.

(32) Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. Ann Intern Med 2002;137(6):511-520.

(33) Stefanidis D, Sevdalis N, Paige J, Zevin B, Aggarwal R, Grantcharov T, et al. Simulation in surgery: what's needed next? Ann Surg 2015 May;261(5):846-853.

(34) McCluney A, Vassiliou M, Kaneva P, Cao J, Stanbridge D, Feldman L, et al. FLS simulator performance predicts intraoperative laparoscopic skill. Surg Endosc 2007;21(11):1991-1995.

(35) Sturm LP, Windsor JA, Cosman PH, Cregan P, Hewett PJ, Maddern GJ. A systematic review of skills transfer after surgical simulation training. Ann Surg 2008 Aug;248(2):166-179.

(36) Davies J, Khatib M, Bello F. Open surgical simulation—a review. Journal of surgical education 2013;70(5):618-627.

(37) Grober ED, Hamstra SJ, Wanzel KR, Reznick RK, Matsumoto ED, Sidhu RS, et al. Laboratory based training in urological microsurgery with bench model simulators: a randomized controlled trial evaluating the durability of technical skill. J Urol 2004 Jul;172(1):378-381.

(38) Reiley CE, Lin HC, Yuh DD, Hager GD. Review of methods for objective surgical skill evaluation. Surg Endosc 2011;25(2):356-366.

(39) Shah J, Darzi A. Surgical skills assessment: an ongoing debate. BJU Int 2001;88(7):655-660.

(40) Reznick RK. Teaching and testing technical skills. The American journal of surgery 1993;165(3):358-361.

(41) Martin J, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg 1997;84(2):273-278.

(42) Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative “bench station” examination. The American Journal of Surgery 1997;173(3):226-230.

(43) Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Academic Medicine 1998;73(9):993-997.

(44) Doyle JD, Webber EM, Sidhu RS. A universal global rating scale for the evaluation of technical skills in the operating room. The American journal of surgery 2007;193(5):551-555.

(45) Heery E, Noon M. A dictionary of human resource management. : OUP Oxford; 2008.

(46) Bridgewater B, Grayson AD, Jackson M, Brooks N, Grotte GJ, Keenan DJ, et al. Surgeon specific mortality in adult cardiac surgery: comparison between crude and risk stratified data. BMJ 2003 Jul 5;327(7405):13-17.

(47) Rosenbaum DA. Human motor control. : Academic press; 2009.

(48) Stergiou N, Harbourne RT, Cavanaugh JT. Optimal movement variability: a new theoretical perspective for neurologic physical therapy. Journal of Neurologic Physical Therapy 2006;30(3):120-129.

(49) Esquenazi A. Gait analysis in lower-limb amputation and prosthetic rehabilitation. Phys Med Rehabil Clin N Am 2014;25(1):153-167.

(50) Chmarra M, Grimbergen C, Dankelman J. Systems for tracking minimally invasive surgical instruments. Minimally Invasive Therapy & Allied Technologies 2007;16(6):328-340.

(51) Datta V, Mackay S, Mandalia M, Darzi A. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. J Am Coll Surg 2001;193(5):479-485.

(52) Cox M, Irby DM, Reznick RK, MacRae H. Teaching surgical skills—changes in the wind. N Engl J Med 2006;355(25):2664-2669.

(53) Glarner CE, Hu Y, Chen C, Radwin RG, Zhao Q, Craven MW, et al. Quantifying technical skills during open operations using video-based motion analysis. Surgery 2014;156(3):729-734.

(54) Automated video exposure assessment of repetitive motion. Proceedings of the Human Factors and Ergonomics Society Annual Meeting: SAGE Publications; 2011.

(55) Marescaux J, Dallemagne B, Perretta S, Wattiez A, Mutter D, Coumaros D. Surgery without scars: report of transluminal cholecystectomy in a human being. Archives of surgery 2007;142(9):823-826.

(56) Guillou PJ, Quirke P, Thorpe H, Walker J, Jayne DG, Smith AM, et al. Short-term endpoints of conventional versus laparoscopic-assisted surgery in patients with colorectal cancer (MRC CLASICC trial): multicentre, randomised controlled trial. The lancet 2005;365(9472):1718-1726.

(57) Laparoscopic surgery, perceptual limitations and force: A review. First Canadian student conference on biomedical computing; 2006.

(58) Moorthy K, Munz Y, Sarker SK, Darzi A. Objective assessment of technical skills in surgery. Br Med J 2003;327(7422):1032.

(59) Fried GM, Feldman LS, Vassiliou MC, Fraser SA, Stanbridge D, Ghitulescu G, et al. Proving the value of simulation in laparoscopic surgery. Ann Surg 2004;240(3):518-528.

(60) Peters JH, Fried GM, Swanstrom LL, Soper NJ, Sillin LF, Schirmer B, et al. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. Surgery 2004;135(1):21-27.

(61) O'toole RV, Playter RR, Krummel TM, Blank WC, Cornelius NH, Roberts WR, et al. Measuring and developing suturing technique with a virtual reality surgical simulator. J Am Coll Surg 1999;189(1):114-127.

(62) Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg 2002;236(4):458-464.

(63) Chaudhry A, Sutton C, Wood J, Stone R, McCloy R. Learning rate for laparoscopic surgical skills on MIST VR, a virtual reality simulator: quality of human-computer interface. Ann R Coll Surg Engl 1999 Jul;81(4):281-286.

(64) Ahmadi S, Sielhorst T, Stauder R, Horn M, Feussner H, Navab N. Recovery of surgical workflow without explicit models. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006: Springer; 2006. p. 420-428.

(65) Rosen J, Solazzo M, Hannaford B, Sinanan M. Task decomposition of laparoscopic surgery for objective evaluation of surgical residents' learning curve using hidden Markov model. Computer Aided Surgery 2002;7(1):49-61.

(66) Rosen J, Brown JD, Chang L, Sinanan MN, Hannaford B. Generalized approach for modeling minimally invasive surgery as a stochastic process using a discrete Markov model. Biomedical Engineering, IEEE Transactions on 2006;53(3):399-413.

(67) Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 1989;77(2):257-286.

(68) MacKenzie L, Ibbotson J, Cao C, Lomax A. Hierarchical decomposition of laparoscopic surgery: a human factors approach to investigating the operating room environment. Minimally Invasive Therapy & Allied Technologies 2001;10(3):121-127.

(69) Rosen J, Hannaford B, Richards CG, Sinanan MN. Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills. Biomedical Engineering, IEEE Transactions on 2001;48(5):579-591.

(70) Cuschieri A. Visual Displays and Visual Perception in Minimal Access Surgery. Semin Laparosc Surg 1995 Sep;2(3):209-214.

(71) Moorthy K, Munz Y, Adams S, Pandey V, Darzi A. A human factors analysis of technical and team skills among surgical trainees during procedural simulations in a simulated operating theatre. Ann Surg 2005 Nov;242(5):631-639.

(72) Swanstrom L, Zheng B. Spatial orientation and off-axis challenges for NOTES. Gastrointest Endosc Clin N Am 2008 Apr;18(2):315-24; ix.

(73) Martinec DV, Gatta P, Zheng B, Denk PM, Swanstrom LL. The trade-off between flexibility and maneuverability: task performance with articulating laparoscopic instruments. Surg Endosc 2009 Dec;23(12):2697-2701.

(74) Pinzon D, Byrns S, Zheng B. Prevailing Trends in Haptic Feedback Simulation for Minimally Invasive Surgery. Surg Innov 2016 Feb 2.

(75) Clements M, Morrison K, Schenkman NS. Evaluation of Laparoscopic Curricula in American Urology Residency Training: A 5-year update. J Endourol 2016 Jan 5.

(76) Shetty S, Zevin B, Grantcharov TP, Roberts KE, Duffy AJ. Perceptions, training experiences, and preferences of surgical residents toward laparoscopic simulation training: a resident survey. J Surg Educ 2014 Sep-Oct;71(5):727-733.

(77) Entertainment Software Association (ESA). Essential Facts About the Computer and Video Game Industry: 2015 Sales, Demographic and Usage Data. 2016; Available at: http://www.theesa.com/wp-content/uploads/2015/04/ESA-Essential-Facts-2015.pdf. Accessed January 20, 2016.

(78) Kuhn S, Gleich T, Lorenz RC, Lindenberger U, Gallinat J. Playing Super Mario induces structural brain plasticity: gray matter changes resulting from training with a commercial video game. Mol Psychiatry 2014 Feb;19(2):265-271.

(79) Kennedy AM, Boyle EM, Traynor O, Walsh T, Hill AD. Video gaming enhances psychomotor skills but not visuospatial and perceptual abilities in surgical trainees. J Surg Educ 2011 Sep-Oct;68(5):414-420.

(80) Rosser JC,Jr, Lynch PJ, Cuddihy L, Gentile DA, Klonsky J, Merrell R. The impact of video games on training surgeons in the 21st century. Arch Surg 2007 Feb;142(2):181-6; discusssion 186.

(81) Ou Y, McGlone ER, Camm CF, Khan OA. Does playing video games improve laparoscopic skills? Int J Surg 2013;11(5):365-369.

(82) Schlickum MK, Hedman L, Enochsson L, Kjellin A, Fellander-Tsai L. Systematic video game training in surgical novices improves performance in virtual reality endoscopic surgical simulators: a prospective randomized study. World J Surg 2009 Nov;33(11):2360-2367.

(83) Badurdeen S, Abdul-Samad O, Story G, Wilson C, Down S, Harris A. Nintendo Wii video-gaming ability predicts laparoscopic skill. Surg Endosc 2010 Aug;24(8):1824-1828.

(84) Dallal GE. Randomization.com. 2008; Available at: randomization.com. Accessed August 10, 2014.

(85) Peters JH, Fried GM, Swanstrom LL, Soper NJ, Sillin LF, Schirmer B, et al. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. Surgery 2004 Jan;135(1):21-27.

(86) Schreuder HW, van den Berg CB, Hazebroek EJ, Verheijen RH, Schijven MP. Laparoscopic skills training using inexpensive box trainers: which exercises to choose when constructing a validated training course. BJOG 2011 Dec;118(13):1576-1584.

(87) Jalink MB, Goris J, Heineman E, Pierie JP, Ten Cate Hoedemaker HO. Face validity of a Wii U video game for training basic laparoscopic skills. Am J Surg 2015 Jun;209(6):1102-1106.

(88) Gauger PG, Hauge LS, Andreatta PB, Hamstra SJ, Hillard ML, Arble EP, et al. Laparoscopic simulation training with proficiency targets improves practice and performance of novice surgeons. Am J Surg 2010 Jan;199(1):72-80.

(89) Dawe SR, Windsor JA, Broeders JA, Cregan PC, Hewett PJ, Maddern GJ. A systematic review of surgical skills transfer after simulation-based training: laparoscopic cholecystectomy and endoscopy. Ann Surg 2014 Feb;259(2):236-248.

(90) Sroka G, Feldman LS, Vassiliou MC, Kaneva PA, Fayez R, Fried GM. Fundamentals of laparoscopic surgery simulator training to proficiency improves laparoscopic performance in the operating room-a randomized controlled trial. Am J Surg 2010 Jan;199(1):115-120.

(91) Martin J, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg 1997;84(2):273-278.

(92) Jin GH, Lee SB, Lee TS. Context awareness of human motion states using accelerometer. J Med Syst 2008;32(2):93-100.

(93) Arif M, Bilal M, Kattan A, Ahamed SI. Better physical activity classification using smartphone acceleration sensor. J Med Syst 2014;38(9):1-10.

(94) Rosen J, Brown JD, Chang L, Sinanan MN, Hannaford B. Generalized approach for modeling minimally invasive surgery as a stochastic process using a discrete Markov model. Biomedical Engineering, IEEE Transactions on 2006;53(3):399-413.

(95) Automatic motion recognition and skill evaluation for dynamic tasks. Proc. Eurohaptics; 2003.

(96) Leong JJ, Nicolaou M, Atallah L, Mylonas GP, Darzi AW, Yang G. HMM assessment of quality of movement trajectory in laparoscopic surgery. Computer Aided Surgery 2007;12(6):335-346.

(97) Reiley CE, Lin HC, Yuh DD, Hager GD. Review of methods for objective surgical skill evaluation. Surg Endosc 2011;25(2):356-366.

(98) Juang B, Rabiner LR. The segmental K-means algorithm for estimating parameters of hidden Markov models. Acoustics, Speech and Signal Processing, IEEE Transactions on 1990;38(9):1639-1641.

(99) Li Z, Wei Z, Yue Y, Wang H, Jia W, Burke LE, et al. An adaptive hidden markov model for activity recognition based on a wearable multi-sensor device. J Med Syst 2015;39(5):1-10.

(100) Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 1989;77(2):257-286.

(101) Loukas C, Georgiou E. Multivariate autoregressive modeling of hand kinematics for laparoscopic skills assessment of surgical trainees. Biomedical Engineering, IEEE Transactions on 2011;58(11):3289-3297.

(102) Datta V, Mackay S, Mandalia M, Darzi A. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. J Am Coll Surg 2001;193(5):479-485.

(103) Baum LE, Petrie T, Soules G, Weiss N. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. The annals of mathematical statistics 1970;41(1):164-171.

(104) Megali G, Sinigaglia S, Tonet O, Dario P. Modelling and evaluation of surgical performance using hidden Markov models. Biomedical Engineering, IEEE Transactions on 2006;53(10):1911-1919.

(105) Reiley CE, Hager GD. Task versus subtask surgical skill evaluation of robotic minimally invasive surgery. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009: Springer; 2009. p. 435-442.

(106) Automatic analysis of 3D gaze coordinates on scene objects using data from eye-tracking and motion-capture systems. Proceedings of the Symposium on Eye Tracking Research and Applications: ACM; 2012.

(107) Hsu JL, Korndorffer JR, Brown KM. Design of vessel ligation simulator for deliberate practice. J Surg Res 2015;197(2):231-235.

(108) Hsu JL, Korndorffer JR, Brown KM. Force feedback vessel ligation simulator in knot-tying proficiency training. The American Journal of Surgery 2016;211(2):411-415.

(109) Tien T, Pucher PH, Sodergren MH, Sriskandarajah K, Yang GZ, Darzi A. Eye tracking for skills assessment and training: a systematic review. J Surg Res 2014 Sep;191(1):169-178.

(110) Bogdanova R, Boulanger P, Zheng B. Three-Dimensional Eye Tracking in a Surgical Scenario. Surg Innov 2015 Oct;22(5):522-527.

(111) Lanata A, Greco A, Valenza G, Scilingo EP. On the tridimensional estimation of the gaze point by a stereoscopic wearable eye tracker. Conf Proc IEEE Eng Med Biol Soc 2015 Aug;2015:2283-2286.

(112) Abbott WW, Faisal AA. Ultra-low-cost 3D gaze estimation: an intuitive high information throughput compliment to direct brain-machine interfaces. J Neural Eng 2012 Aug;9(4):046016-2560/9/4/046016. Epub 2012 Jul 12.

(113) Duchowski A. Eye tracking methodology: Theory and practice. : Springer Science & Business Media; 2007.

(114) Maycock J, Essig K, Prinzhorn D, Schack T, Ritter H. Probing attention for insights into imitation learning mechanisms.

(115) Riordan-Eva P, Whitcher J. Vaughan & Asbury's general ophthalmology. : Wiley Online Library; 2008.

(116) Weisstein EW. Point-Line Distance--3-Dimensional. From *MathWorld*--A Wolfram Web Resource. Available at: http://mathworld.wolfram.com/Point-LineDistance3-Dimensional.html.

(117) Belardinelli A, Herbort O, Butz MV. Goal-oriented gaze strategies afforded by object interaction. Vision Res 2015;106:47-57.

(118) Hybrid HMM/SVM model for the analysis and segmentation of teleoperation tasks. Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on; 2004.

(119) Khan RS, Tien G, Atkins MS, Zheng B, Panton ON, Meneghetti AT. Analysis of eye gaze: Do novice surgeons look at the same location as expert surgeons during a laparoscopic operation? Surg Endosc 2012;26(12):3536-3540.

(120) Weisstein EW. "Point-Point Distance--3-Dimensional." From *MathWorld*--A Wolfram Web Resource.
. Available at: http://mathworld.wolfram.com/Point-PointDistance3-Dimensional.html.

(121) Wilson MR, Vine SJ, Bright E, Masters RS, Defriend D, McGrath JS. Gaze training enhances laparoscopic technical skill acquisition and multi-tasking performance: a randomized, controlled study. Surg Endosc 2011;25(12):3731-3739.

(122) Fuhl W, Santini T, Kasneci G, Kasneci E. PupilNet: Convolutional Neural Networks for Robust Pupil Detection. arXiv preprint arXiv:1601.04902 2016.

(123) Peters M, Laeng B, Latham K, Jackson M, Zaiyouna R, Richardson C. A redrawn Vandenberg and Kuse mental rotations test-different versions and factors that affect performance. Brain Cogn 1995;28(1):39-58.