Informatics Approach to Improving Surgical Skills Training

by

Gazi Islam

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved September 2013 by the
Graduate Supervisory Committee:

Baoxin Li, Co-Chair
Jianming Liang, Co-Chair
Valentin Dinu
Robert Greenes
Marshall Smith
Kanav Kahol
Vimla L Patel

ARIZONA STATE UNIVERSITY

December 2013
ABSTRACT

Surgery as a profession requires significant training to improve both clinical decision making and psychomotor proficiency. In the medical knowledge domain, tools have been developed, validated, and accepted for evaluation of surgeons’ competencies. However, assessment of the psychomotor skills still relies on the Halstedian model of apprenticeship, wherein surgeons are observed during residency for judgment of their skills. Although the value of this method of skills assessment cannot be ignored, novel methodologies of objective skills assessment need to be designed, developed, and evaluated that augment the traditional approach. Several sensor-based systems have been developed to measure a user’s skill quantitatively, but use of sensors could interfere with skill execution and thus limit the potential for evaluating real-life surgery. However, having a method to judge skills automatically in real-life conditions should be the ultimate goal, since only with such features that a system would be widely adopted. This research proposes a novel video-based approach for observing surgeons’ hand and surgical tool movements in minimally invasive surgical training exercises as well as during laparoscopic surgery. Because our system does not require surgeons to wear special sensors, it has the distinct advantage over alternatives of offering skills assessment in both learning and real-life environments. The system automatically detects major skill-measuring features from surgical task videos using a computing system composed of a series of computer vision algorithms and provides on-screen real-time performance feedback for more efficient skill learning. Finally, the machine-learning approach is used to develop an observer-independent composite scoring model through objective and quantitative measurement of surgical skills. To increase effectiveness and usability of the developed system, it is integrated with a cloud-based tool, which automatically assesses surgical videos upload to the cloud.
This dissertation is dedicated to my father, Dr Gazi Serajul Islam. It is your shining example that I try to emulate in all that I do. Thank you for everything.

My thanks to...

...my mother, for taking care of me in spite of being world away. Thank you so much for your unconditional love and support.

...my sister and brother-in-law, for your inspiration, guidance and for bringing two little angels (Aria and Saamit) into the world who cheer me up every time I feel inundated with work.

...Nimmy, through the good times and bad, your love, kindness, extensive support and encouragement have been ever-present in this important time of my life. I am eternally grateful.

...Nabeel and Ussala, for both putting me up and for putting up with me whenever I needed a break from work.

...Kanav, for being my mentor, and for your friendship and guidance over the years.
ACKNOWLEDGMENTS

I would like to express my gratitude to Dr. Kanav Kahol for his constant encouragement and belief in me. He has been everything that one could want in a mentor. I am deeply indebted to my co-chair Dr Baoxin Li for providing me funding, logistical support and valuable guidance during the entire project. I am grateful to each member of the committee for their valuable contributions. Dr. Marshall Smith provided extensive access to Banner’s Simulation Education and Training (SimET) Center for research and has frequently offered valuable insights in the clinical aspects of the project. It was a privilege to have Dr. Vimla L Patel on my committee, who provided guidance throughout the duration of the PhD and taught me valuable lessons on how to present scientific work to a diverse audience. I am honored that she has encouraged me to take this work to the next level. Dr. Robert Greenes has provided particularly relevant feedback at every stage. Dr. Valentin Dinu has been a solid resource for the machine learning aspect of this work. Last but not least, my departmental co-chair, Dr. Jianming Liang constantly pushed me to stay on the timeline and provided strategic insights for smooth execution of the entire project.

I also want to thank all my colleagues from the Human Machine Symbiosis Lab, specially, Mithra Vankipuram and Aaron Ashby, who helped me throughout the process of increasing quality and adding value to the product of my efforts through constructive suggestions. Prabal Khanal and Akshay Vankipuram, both brilliant computer scientists, have always helped me with coding and most importantly, patiently taught me how to code.

This work was supported in part by National Science Foundation (Grant No. 0904778). Any opinions and conclusions expressed in this material are mine and do not necessarily reflect the views of NSF.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>vi</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
</tbody>
</table>

## CHAPTER

1. INTRODUCTION ......................................................................................................................... 1
   1.1. Problem Statement ............................................................................................................ 6
   1.2. Research Goal .................................................................................................................... 8
   1.3. Hypotheses ......................................................................................................................... 8
   1.4. Novelties ............................................................................................................................ 9
   1.5. Scope and Limitations ....................................................................................................... 10
   1.6. Contributions ................................................................................................................... 11
   1.7. Structure of the thesis .................................................................................................... 12
   1.8. Publications ....................................................................................................................... 13

2. BACKGROUND ............................................................................................................................ 15
   2.1. Minimally Invasive Surgery ............................................................................................... 15
   2.2. Simulator based training ................................................................................................... 16
   2.3. Skill assessment ................................................................................................................ 19
   2.4. Computer vision ............................................................................................................... 24
   2.5. Machine learning .............................................................................................................. 39
   2.6. Medical education theory ................................................................................................. 43

3. RELATED WORK .......................................................................................................................... 46
   3.1. Simulation based assessment ............................................................................................. 46
   3.2. Preliminary work ................................................................................................................. 52
   3.3. Assessment in Real Surgery ............................................................................................... 57
3. Internet based video assessment ......................................................... 58
4. METHODOLOGY ............................................................................. 60
   4.1. Simulation ................................................................................. 60
   4.2. Performance in actual surgery ................................................... 78
   4.3. Cloud based tool ..................................................................... 90
5. RESULTS .......................................................................................... 98
   5.1. FLS simulator ............................................................................ 98
   5.2. Laparoscopic Cholecystectomy .................................................. 104
6. DISCUSSION .................................................................................... 107
7. CHALLENGES .................................................................................. 112
   7.1. Data collection .......................................................................... 112
   7.2. Data analysis ............................................................................. 113
   7.3. Implementing the cloud based system ....................................... 113
8. FUTURE WORK ................................................................................. 115
   8.1. Implementation in other types of simulations ......................... 115
   8.2. Implementation in other types of surgeries ............................. 115
   8.3. Use of 3D camera .................................................................... 115
   8.4. Complete cloud based solution ................................................. 116
9. CONCLUSION ................................................................................... 118
10. REFERENCES .................................................................................. 120
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Global Assessment Tools</td>
<td>200</td>
</tr>
<tr>
<td>2. Hand movement tracking</td>
<td>47</td>
</tr>
<tr>
<td>3. Tool movement tracking</td>
<td>47</td>
</tr>
<tr>
<td>4. Capturing demographic features</td>
<td>48</td>
</tr>
<tr>
<td>5. Use of pressure sensors</td>
<td>49</td>
</tr>
<tr>
<td>6. Major drawbacks of the other video-based systems</td>
<td>51</td>
</tr>
<tr>
<td>7. T-Test analysis of Mayo Clinic data</td>
<td>54</td>
</tr>
<tr>
<td>8. T-test analysis of Mount Sinai data</td>
<td>56</td>
</tr>
<tr>
<td>9. Gesture Segmentation explain</td>
<td>64</td>
</tr>
<tr>
<td>10. Mean and SD of skill levels in intra-corporeal suturing</td>
<td>74</td>
</tr>
<tr>
<td>11. Sample size calculation</td>
<td>75</td>
</tr>
<tr>
<td>12. Sample size calculation</td>
<td>77</td>
</tr>
<tr>
<td>13. Distribution of subjects</td>
<td>98</td>
</tr>
<tr>
<td>14. ANOVA among expertise levels</td>
<td>101</td>
</tr>
<tr>
<td>15. Random Forest Analysis</td>
<td>101</td>
</tr>
<tr>
<td>16. Error detection</td>
<td>102</td>
</tr>
<tr>
<td>17. Fleiss’es Kappa coefficient between raters</td>
<td>104</td>
</tr>
<tr>
<td>18. Correlation with experts' score</td>
<td>105</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Different FLS exercises.</td>
<td>19</td>
</tr>
<tr>
<td>2. Pixel value of image</td>
<td>25</td>
</tr>
<tr>
<td>3. RGB to grayscale</td>
<td>26</td>
</tr>
<tr>
<td>4. Grayscale to binary</td>
<td>27</td>
</tr>
<tr>
<td>5. RGB to HSV</td>
<td>28</td>
</tr>
<tr>
<td>6. Gaussian blur on 1-D pixel array</td>
<td>30</td>
</tr>
<tr>
<td>7. Smoothing using Gaussian blur</td>
<td>30</td>
</tr>
<tr>
<td>8. Noise filtering</td>
<td>31</td>
</tr>
<tr>
<td>9. Canny's edge detection</td>
<td>33</td>
</tr>
<tr>
<td>10. Detected circle (yellow)</td>
<td>34</td>
</tr>
<tr>
<td>11. Grayscale image and corresponding histogram</td>
<td>35</td>
</tr>
<tr>
<td>12. Histogram</td>
<td>35</td>
</tr>
<tr>
<td>13. Object segmentation</td>
<td>36</td>
</tr>
<tr>
<td>14. Motion history image</td>
<td>37</td>
</tr>
<tr>
<td>15. Gradient of motion</td>
<td>38</td>
</tr>
<tr>
<td>16. Average motion redundancy value for peg transfer exercise</td>
<td>53</td>
</tr>
<tr>
<td>17. Average motion redundancy value for intracorporeal suture exercise</td>
<td>53</td>
</tr>
<tr>
<td>18. Average motion redundancy value for knot tying exercise</td>
<td>54</td>
</tr>
<tr>
<td>19. Average acceleration value of user groups</td>
<td>55</td>
</tr>
<tr>
<td>20. LDA Analysis</td>
<td>56</td>
</tr>
<tr>
<td>21. Mean motion smoothness</td>
<td>57</td>
</tr>
<tr>
<td>22. FLS box with external camera</td>
<td>60</td>
</tr>
<tr>
<td>23. System Architecture</td>
<td>62</td>
</tr>
</tbody>
</table>
49. Review previous scores ................................................................. 95
50. Video rated by experts ............................................................... 96
51. Logout page .............................................................................. 97
52. Jerkiness, Snaps and Tremor scores ......................................... 99
53. Movement efficiency ............................................................... 100
54. Comparison of bimanual dexterity .......................................... 100
55. LDA analysis ........................................................................... 102
56. Learning curve ......................................................................... 103
57. Effect of assessment system on learning .................................. 104
58. Inter-rater reliability of Global scores between rater 1, 2 & 3 .... 105
59. Intubation exercise ................................................................. 115
60. Train back the algorithm ......................................................... 117
Chapter 1

INTRODUCTION

Surgery has existed for thousands of years but was not common until the 1800s. Early surgery was very rare and medicine was the most common way to treat illness. Surgery was avoided, because patients often died as a result. But by 1840 trends had changed: surgery took over as a plausible treatment for certain medical problems. As surgery becomes safer and better, the global health community has recognized that surgical conditions address a significant burden of disease and provide cost-effective interventions. Several new techniques of surgery like minimally invasive surgery (also known as laparoscopic or videoscopic surgery) have allowed surgeons to perform the same procedures as in traditional open surgery, using small incisions (keyhole surgery) instead of large abdominal incisions. Studies have shown major benefits to the patient in terms of reduced post-operative pain, thus resulting in increased comfort, reduced hospital stay, quicker return to normal physical activities and ultimately a quicker return to work. Improved cosmetics and reduced wound complications associated with large scars are also major advantages associated with this technique.

Surgery as a domain however has a significant additional component of psychomotor manipulation and consequent skill base for which training is required. Training in surgery is a critical component of delivering effective patient safety. In fact, due to the demand for greater accountability and patient safety in health care delivery; effective surgical performance measurement is gaining an increasingly high profile. Advances in the medical field with new methods, curricula, and processes, such as the Accreditation Council for Graduate Medical Education competencies or Objective Structured Assessment of Technical Skills, as well as innovative technologies such as web-based
learning and simulation have made surgical skill acquisition more challenging than ever [1].

This is true for both open surgery and laparoscopic surgery, albeit the acquisition of laparoscopic skills requires a longer learning curve than does open surgery [2]. Efficient skill training in minimally invasive surgical (MIS) techniques has become standard in surgical residency programs. To ensure the best surgical performance, systematic simulator training programs are being developed alongside traditional training in hospitals [3]. It is a new and progressive way to intensify surgical resident training and surgical skills learning. Unlike open surgery, MIS is, by its nature, a technique that is very suitable for simulation-based training [4]. The specific psychomotor skills and hand-eye coordination needed for this type of surgery can be trained easily through box-trainers and computer-enhanced simulation trainers.

While simulation training is emerging as an effective means of practicing in a non-threatening environment, evaluation of surgical proficiency in both the traditional training method and simulation based method, needs constant presence of a competent surgeon. This means of evaluating surgical dexterity is highly subjective and does not yield quantitative data [5]. There are checklists that have been developed by surgical educators that guide a human observer with points they should consider for evaluation. Techniques like the OSATS [6] are very human-intensive and depend on the evaluator. This again introduces subjectivity into the process.

An alternative approach lies in considering objective evaluation through informatics-driven analysis of surgical hand motions. Human motion analysis through computation has been an active research area since the decade of the 1960s [7]. From analyzing human gait [8] to analyzing dance [9, 10], human motion analysis has been used in several fields for different application. In the medical world, motion analysis has been
used in rehabilitation, gait analysis, and even in detection of infection patterns through crowd motion analysis. However, the analysis of surgical hand movement has just started emerging as an area of scientific inquiry. Yet, currently there is a lack of accepted informatics solutions to surgical simulation. In order to accurately measure a user’s skill objectively and quantitatively, the system must satisfy the following requirements [11]: (1) the system must possess adequate sensing techniques to monitor the user’s operation; (2) the system must extract relevant features from the sensing data; (3) the system needs a good computational model to represent the skill demonstrated in the operation; and (4) the system must perform equivalent to that of the gold standard. Such a model is essential for accurately measuring the technical competence of the performance.

High-end VR simulators use high-end sensors that map movements into virtual space where they are analyzed by algorithms. This method has some advantages: the reliability of the evaluation process increases with more objective and structured criteria, with further reduction in inter-rater variability through development and the use of automated computerized objective technical skill assessment systems.

But it suffers from practical limitations. Such simulators are often out of reach of residency programs due to their high costs. This is especially true when one considers global health challenges. Also the skill assessment technique cannot be applicable to real surgeries since it will require introducing sensors to extract assessment features from a surgery. Use of sensory devices to track performance is very common in the simulation setting; however introduction of sensors in actual surgery can cause interference in skill execution and can increase the cost of the entire procedure. Also there are practical considerations like hygiene, sanitation and patient safety.
On the other hand, low-end box simulators rely on trainees performing movements which are observed by an examiner and their performance is rated subjectively. This is a challenging task, one that requires evaluation of technical skills of residents by expert surgeons. As mentioned before, this one-to-one apprenticeship model has significant limitations [12]. First, there is a dearth of trained surgeons, especially in the developing world, who can be evaluators. Secondly it requires dedicated time of the surgeon. Hence it is almost impossible to immediately evaluate every surgical training exercise. The trainees are only assessed before moving to the next stage of training. As a result feedback becomes summative and fails to provide guidance on how to improve the individual’s clinical skills [13].

The performance of the trainee in real surgical procedures can be assessed by maintenance of a detailed procedure log, direct observation by an expert, or review of a videotape of the procedure. However, the operative log only reflects the number of procedures performed and not the quality of the performance, and end-of-rotation ratings are not procedure-specific and are subject to rating errors such as selective recall and faculty bias. As a result, even with the trained surgeons, surgical skill measurement remains subjective [14]. Hence, there is a need for an affordable, complete, medical augmented reality performance evaluation system with automatic tracking, which can also be transferable to real surgeries. One method to do this would be through video analysis wherein cameras are employed in the OR to track surgeons’ movements and analyze them to generate a proficiency rating.

Since a camera is incorporated in MIS training and in real surgeries, video data of the surgical scene are available. The video data can be further analyzed with computer vision analysis without the use of any external sensors. Computer vision analysis (herein referred to simply as “computer vision”) has achieved many technological breakthroughs
in the last few decades. It has been successfully used for object detection, tracking, motion detection and analysis, etc. Gaming platforms like the Microsoft Kinect [15] use computer vision to track entire body movements. A variety of computer vision applications that have been invented in the past few years can be applied in clinical as well as other domains of biomedical informatics to solve many problems [16, 17]. Surgical videos can be explicitly analyzed by a state-of-the-art computer vision algorithm to extract features that correspond to the measurements used standards in surgical performance assessment. These features can be further analyzed with machine learning algorithms. Machine learning algorithms in conjunction with computer vision have been employed to solve several motion analysis problems [7]. In medicine, gait analysis and rehabilitation are examples.

Using computer vision and machine learning algorithms, every surgery task can be assessed and on-screen feedback can be provided during the task. Formative assessment helps students via feedback to identify problems and improve their technique and skills during the tasks [18]. Since the feedback is immediate, our project will use such an approach to seek to increase a) training efficiency by reducing the training workflow cycle time (accomplished by shrinking the feedback response time) and b) training effectiveness by reducing the time it takes to accomplish a desired level of proficiency (improvements). This technique is also expected to be used to assess a surgeon’s performance in an actual surgical operation.

The incorporation of cameras inside the body, which is a feature of minimally invasive surgery, also offers the opportunity to record common surgical procedures. The recorded video can be effectively used for later assessment and trainee self-review. It allows trainers to assess trainee performance whenever convenient, thus increasing
efficiency and limiting “examiner burnout” from fatigue and loss of concentration [19]. Observing surgical procedure video also benefits trainees in self-assessment [20].

Efficient distribution of the videos over a cloud-based video sharing system can be an effective tool [21]. Based on this idea, the video assessment tool developed in this project can be integrated with the sharing tool. Trainees can upload videos and get immediate evaluation scores. Experts conducting assessment on these uploaded videos can also be recorded. These assessment scores can increase the training data for further training of the machine learning algorithm. Although data-driven machine learning is often viewed as an autonomous process, it is increasingly apparent that the application of machine learning to real-world problems requires substantial input from a domain expert before satisfactory solutions can be derived [22].

1.1. Problem Statement

Surgery requires many hours of systematic practice with a simulator to acquire psychomotor skills. Timely feedback on each simulated surgical session can facilitate surgical skills acquisition more efficiently and effectively than having no or delayed feedback [1]. The current evaluation system used by the American Board of Surgery requires constant presence during an assessment or post-hoc analysis on a recorded surgery by a competent surgeon or trainer to subjectively assess the surgical dexterity of the trainee. This assessment is accomplished by providing a composite score, which lacks inter-rater reliability. Rudimentary video and sensor-based systems have been developed to capture a user’s motion and other important features which can be later analyzed objectively and quantitatively to correlate with the skill level [11, 23, 24]. However, the present systems have several drawbacks.

- Most of the skill assessment systems available compel surgeons to wear sensors to track different features of the surgical procedure that can be correlated with the skill
execution. However, integration of wearable sensors interferes with surgical skill execution [11].

- Existing assessment techniques are system-specific, and thus cannot be transferred to different simulators. Most of the skill assessment techniques available are designed for one particular simulator. Since they track different features in different exercises, each of these techniques is designed for one designated task only.

- Since skill assessment can be performed only in simulations, the system may not be transferable to assess a real-life surgery. Not only, as noted above, does integration of sensors causes interference with skill execution, but also the sensors have to be sterilized to be used in real surgery. Moreover the sensors are expensive [25] and can increase the cost of the entire surgical procedure.

- Existing video-based systems that use only external cameras to capture movement of hands while performing surgical exercises are subject to inconsistent lighting, background noise and low-resolution video images [11]. Furthermore, due to the use of long instruments in MIS, it is almost impossible to derive tool movement from hand movement video alone. Other video-based approaches track surgical tool-tip movement, but the analysis is solely dependent on the travel-length measurement of the tool-tip. Moreover no work has been done that combines features from both hand and surgical tool movement.

- Current approaches require the assessment system and capture system to be co-located. However an approach where residencies can upload videos to a remote analysis system would be very helpful, since it would reduce costs and offer a separation between capture and analysis which in the case of examinations can improve the objectivity of the scores.
The focus of the proposed research is to develop a robust system that is free from the above mentioned drawbacks. The work proposes to develop a sensor-free system to record and analyze surgical proficiency. The designed system will be independent of any simulator and will be capable of providing surgical proficiency ratings on any simulator. It builds in the capacity to be transferred to real-world environments, which can measure transfer of skills from simulated to real environments and also analyze real-world surgery simply. The proposed system will be robust by using computer vision techniques to handle light and angle variations.

1.2. Research Goal
This doctoral dissertation proposes an efficient, effective and economic video-based skill assessment technique for minimally invasive surgery training, coupled with real-time feedback assessment that is based on surgeons’ hand and surgical tool movement analyses. It also proposes to apply the technology in assessing performance of real-life laparoscopic cholecystectomy surgery. Finally it proposes work towards development of a cloud-based platform for easy evaluation and sharing of the surgical videos.

1.3. Hypotheses
The proposed system uses a computer vision optical flow algorithm to analyze a surgeon’s hand movement and surgical tool-video in real time, and provide dynamic feedback scores to assess surgical training performance on three key dimensions: smoothness, proficiency and preciseness. A composite score is computed and compared to the gold standard which is determined using the expert surgeons’ rating. The hypotheses are:

- H1: The skill assessment provided by the proposed automatic quantitative scoring system is equivalent to that of the gold standard.
• H2: The immediate feedback system will increase training effectiveness by reducing the time it takes to attain a desired level of proficiency.

1.4. Novelties

• Observer-independent model: The proposed system replaces the need for presence of an expert rater while also providing real-time feedback.

• No sensor integration: The proposed video-based system does not use any wearable sensors; therefore, it is free from interference caused by sensors.

• Interoperability: Motion smoothness and movement efficiency are two of the most important features that can be used to assess surgical skill. These features are commonly applied as a measure of surgical performance in many simulation exercises as well as in evaluation of real surgery. There are several sensor-based approaches to extract motion smoothness and economy of movement. However, none of them are interoperable, because they are specific to a single simulation task, i.e., one particular system is not operable for different surgical exercises. The proposed system extracts the above-mentioned features, and thus can be used in three types of surgical exercises that use box trainer: Pegboard transfer, intra-corporeal suturing, intra-corporeal knot-tying and pattern cutting.

• Effective skill training: The on-screen feedback on motion smoothness, movement efficiency, elapsed time and number of committed errors is expected to help the user to reach the required level of proficiency more quickly than the traditional methods.

• Advantage over other video-based system: Existing video-based systems that use external cameras to capture surgeons’ hand movements are subject to inconsistent lighting, background noise and low-resolution video image quality limitations. However, the proposed system combines the hand movement analysis with surgical
tool movement analysis utilizing consistent, lighted, noise-free tool video that is readily available from the box-trainers. Also the novel use of a robust optical flow analysis algorithm is expected to provide better result than existing systems.

- Transferable to assess real surgery: Extracted movement features can be tracked in real-life surgery. Since the system is free from sensors, it will not induce any extra cost to surgery and will not cause interference in skill execution. The surgical tool video is a constantly available source of information. A very clear view of the surgeon’s hand movements is also captured by using an external camera on the IV stand. Both of these videos are analyzed to provide a composite score for assessing real-life surgery.

- Cloud-based automatic evaluation system: There are existing internet-based systems where recorded videos can be shared for quicker distribution of the videos to the expert for later assessment; however, the scoring remains subjective. The proposed system develops a unique cloud-based tool through which the trainee can upload a video which can be evaluated immediately and automatically using the developed surgical assessment tool. The platform also serves as a video sharing tool, since the experts can access and assess the uploaded videos as well. Their ratings can be used as training data for further refinements of the skill assessment algorithm.

1.5. Scope and Limitations

- The proposed skill assessment system works for three of the exercises in the Fundamentals of Laparoscopic Surgery (FLS) box trainer: pegboard transfer, intra-corporeal suturing, intra-corporeal knot-tying, shape-cutting. These exercises were chosen in consultations with surgeons who pointed to these as being the most
important skills with which to judge overall proficiency. Also, they are the basis of many complex surgical tasks.

- The surgical skill assessment system assesses skills for laparoscopic cholecystectomy. Cholecystectomy or gall bladder removal is a common procedure representing a significant portion of a general surgeon’s workload. Procedures of this type are generally performed by second year residents and require a combination of psychomotor skill and decision making that provides an adequate basis for judgment of skills in a variety of domains.

- The proposed system is designed to provide feedback of an individual’s performance; it is not designed to provide feedback on group performance.

- Since the 3-dimentional tool-view of the operating field is translated to a 2-dimentional video-screen image, it loses movement information on one axis. However, as the work will show, it is adequate to capture most features of surgical hand movements and in the future can be extended to 3d motion analysis.

1.6. Contributions

- A robust system that will analyze both hand and tool movement and combine them with other features (elapsed time, committed errors) to assess surgical proficiency.

- The on-screen real-time performance feedback is expected to facilitate surgical skill learning efficiency.

- A system that can be deployed to assess actual surgeries. Until now almost no work has been done to transfer this work into an actual surgical setting.

- Internet-based tool for automatic assessment of surgical videos.
1.7. Structure of the thesis

Chapter 2 provides a brief description of the challenges of minimally invasive surgery and the surgical training tools. Then it discusses different surgical skill assessment systems, especially video-based systems since that is the key focus of this research. Finally the chapter has brief description of computer vision and machine learning algorithms and tools that were extensively used in the research.

Chapter 3 reviews other research work in the area of surgical skill assessment. It also reviews preliminary findings of this research from the pilot study, where we used computer vision tools to extract different features from laparoscopic surgical training. It shows which features are most relevant in assessing surgical skill and how they can be transferred to assess real surgery. It also explores the possibility of the feedback system as an efficient learning tool for laparoscopic surgery training.

Chapter 4 describes the methodologies of three different parts of this research work. The first and second sections describe the system architecture, data collection technique, analysis techniques and extracted features, scoring mechanism and the experimental design of the psychomotor skill evaluation of MIS training and laparoscopic cholecystectomy, respectively. The third section describes the structure of the cloud-based tool for automatic assessment of uploaded videos.

Chapter 5 presents the results and analysis of the research. It shows how successfully the system extracts the skill-defining features from the surgical videos. Also it shows the significance of the formative and summative feedback in increasing training effectiveness, by reducing the time it takes to attain a desired level of proficiency.
1.8. Publications

Much of the core materials (Chapter 3-6) are based on publications in journals, and presentation and inclusion in proceedings of major conferences in the field of biomedical informatics, user interface design, and surgery. Most of these papers were joint work with Dr Kanav Kahol and Dr Baoxin Li. Other co-authors are called out where applicable.

The prior work in Chapter 3 is based on publication in three major conferences and the Journal of Surgical Education:


All these work were summarized and published as two book chapters:

• Islam, G., Kahol, K. et al. (2013). “Application of Computer Vision Algorithm in Surgical Skill Assessment”. In Broadband Communications and Sensing, Melbourne, Australia.

The methodology and part of the results or Chapter 4 and 5 were published in two major conferences and showcased in AMIA.


Finally the result and discussion of Chapter 5 and 6 were sent to two major journals. Results of the simulator trainings were to be sent to the Journal of Biomedical Informatics and results of the assessment of actual laparoscopic procedures were sent to American Journal of Surgery.


Chapter 2

BACKGROUND

Over 100 years ago, Dr. William Halsted created the first surgical residency training program in the United States. His training paradigm was extremely simple: “see one, do one, teach one” [26, 27]. However, recent technological advances have changed the way in which some surgeries are performed. Surgery has become more demanding and with the reduced training times as deemed by the American Board of surgery, there is a need for new innovative methods to teach surgeons more skills in reduced times. One methodology that has recently emerged as a viable alternative to apprenticeship model of training is simulation. Medical simulation allows surgeons to practice their skills in safe environment. The simulation systems allow multiple practice sessions, objectively measure skills and allow for remote training. This has opened the opportunity for revisiting Halsted’s paradigm in search for improved ways of training surgeons. It is now generally accepted that these skills must initially be learned in training laboratories prior to entering the operating theater [28]. While in general simulation, applies to all kinds of surgeries, its major impact has been in minimally invasive surgery.

2.1. Minimally Invasive Surgery

Minimally invasive surgery is now gaining acceptance as the main surgical procedure because of its known advantages over open surgery. Among MIS procedures, laparoscopic cholecystectomy (LC), a surgical method of removal of infected or inflamed gallbladder with gallstones, is a gold standard technique used for removal of more than 98% of gallbladders [29]. In spite of factors like age or obesity, the benefits of LC over open surgery were clearly evident as it causes less discomfort, shorter hospital stays with quicker recovery time, reduced postoperative pain, smaller scars [30]. Unlike open
surgery, MIS uses keyhole incisions in human body to operate in constrained environments [28]. This type of surgery requires high levels of psychomotor proficiency and significant training [31]. However, Mastery of laparoscopy skills is difficult. Major challenges in MIS are-

- Translation of the 2-dimentional image of the operating field from the video screen into a 3-dimentional mental image [32, 33].
- Learning to operate using long instruments.
- Mastering ambidexterity and eye-hand coordination.

It requires many hours of practice and a certain amount of innate psychomotor skills. A study shows that LC leads to 158 serious complications as against 23 in open cholecystectomy. Serious complications that occur with LC are common bile duct injury, major vessel laceration, haemorrhage, bile leak, bowel perforation, trocar injury, liver injury and stray electrosurgical burns [34]. As a result correct execution and hence assessment during training of surgical performance is necessary as it has been demonstrated to result in fewer errors during operating room performance [35]. This is where simulation training has emerged as an effective tool.

2.2. Simulator based training

Surgical technical progress in residency training is expected to improve with increased training and the repetition of certain procedures [4]. Transfer of training occurs when one learns a simple skill and adapts that skill to various complex skills [36]. Practicing with simple tasks like pegboard transfer or suturing help to acquire skills that can be transferred to actual surgery. Many commercially available simulators are being used in surgical residency program to teach MIS. Simulation-based surgical training can be categorized into three groups [37-39]:

16
• Physical reality: A traditional box simulator provides haptic feedback; however it requires constant presence of an expert surgeon to subjectively measure the progress of the trainee. The FLS Laparoscopic Trainer Box [40] is a widely used physical reality simulator for surgical residents and practicing surgeons that facilitates the development of psychomotor skills and dexterity required during the performance of basic laparoscopic surgery.

• Augmented reality: Augmented reality surgical simulators provide both realistic haptic feedback and objective assessment after the performance. ProMIS [41], computer-enhanced laparoscopic training system (CELTS), Blue Dragon [23, 24, 42] and LTS3e are some widely used augmented reality simulators. These augmented reality simulators use a variety of monitors, cameras, and complex computing equipment, which must be of high quality but also affordable.

• Virtual reality: VR simulators provide explanations of the tasks to be practiced and objective assessment of the performance; however they lack realistic haptic feedback. Muresan et al. showed in research that for novice trainees, the efficacy of VR training is questionable. In contrast, physical and augmented training methods had benefits in terms of time, quality, and perceived workload [43]

To simulate the similar experience MIS, a simulator box developed by the Fundamentals of Laparoscopic Surgery (FLS) is one of the widely used simulators that is endorsed by the American College of Surgeons (ACS) and establishes a standard set of didactic information and manual skills serving as a basic curriculum to guide surgical residents, fellows and practicing surgeons in the performance of basic laparoscopic surgery [40]. The training is composed of a Laparoscopic Trainer Box which consists of a number of non-procedure specific simulation exercises incorporating most of the psychomotor skills necessary for basic laparoscopic surgery. It is now mandatory for surgeons to pass
the basic test of FLS before they become board certified surgeons. Hence, we employed FLS simulator to develop our system.

FLS box offers a number of exercises. However for this study, three particular exercises were considered where each of exercise tests hand-eye coordination, ambidexterity and depth perception (Figure 1) [40]:

- Peg transfer: The peg transfer exercise requires to lift the six objects with a grasper/dissector with the non-dominant hand, transfer the object midair to the dominant hand and then place each object on a peg on the opposite side of the board (in any order). There is no importance placed on the color of the objects or the order in which they are placed. Once all six objects have been transferred, the process is reversed. The exercise is timed and a penalty is assessed for any peg dropped out of the reach of the tool.

- Intracorporeal suture: This suturing task to place a suture precisely through two marks on a Penrose drain, that has been slit along its long axis, and then tie an intracorporeal knot. The knot must have one double throw and 2 single throws. A penalty is applied if the drain is avulsed from the block to which it is secured by double-sided adhesive tape.

- Shape cutting: This cutting exercise requires cutting out a circle from a square piece of gauze. One hand should be used to provide traction on the gauze using the grasper and to place the gauze at the best possible angle to the cutting hand. A penalty is assessed for any deviation from the line demarcating the circle.
Figure 1: Different FLS exercises.

All these tests form the basis of many complex technical procedures in surgery. They form adequate basis for basic skills analysis and assessment system.

2.3. Skill assessment

Technical skills may be measured by maintenance of a detailed procedural log, direct observation by an expert, or review of a videotape of the procedure. However, the operative log only reflects the number of procedures performed and not the quality of the performance, and end-of-rotation ratings are not procedure specific and subject to rating errors such as selective recall and faculty bias. As a result, analysis of procedural logs and subjective observation are neither reliable nor valid [14]. Videotaping of the surgery is quite useful as the monitor is usually connected to a videocassette recorder and little additional effort is required to record the operation for review at a later time [41]. In addition, the videos can be uploaded to a social networking and video sharing websites for assessment of laparoscopic skills [21].

2.3.1 Subjective Skill Assessment

To access the surgical performance, different scoring systems are available. The objective structured assessment of technical skill (OSATS) was developed for generic evaluation of surgical performance in open surgery through use of a global rating scale [6]. Global Operative Assessment of Laparoscopic Skills (GOALS) scoring system was developed as
an equivalent for MIS by identifying the essential steps required to successfully complete a LC [44]. Each surgery is evaluated in a standardized fashion by observing the videotape of the cholecystectomy [41]. Operative performance rating system (OPRS) is another system that is designed to provide objective operative performance ratings [45]. It is useful in tracking resident development throughout postgraduate training and offers a structured means of certifying operative skills [45-47]. Each OPR contains procedure-specific and generic items on 5-point Likert scales with anchors at the 1, 3, and 5 positions, and a global assessment of overall performance [14]. The Global Assessment Tools are briefly presented in Table 1.

Table 1: Global Assessment Tools

<table>
<thead>
<tr>
<th>OSATS</th>
<th>GOALS[44, 48]</th>
<th>OPRS[14]</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Access &amp; port insertion</td>
<td>• Time</td>
<td>• Instrument handling</td>
</tr>
<tr>
<td>• Gall bladder retraction</td>
<td>• No of hand movement</td>
<td>• Respect for tissue</td>
</tr>
<tr>
<td>• Cystic duct dissection</td>
<td>• Path length</td>
<td>• Cystic duct dissection</td>
</tr>
<tr>
<td>• Cystic duct clipping &amp; transection</td>
<td>• Depth perception</td>
<td>• Efficiency</td>
</tr>
<tr>
<td></td>
<td>• Bimanual dexterity</td>
<td>• Tissue handling</td>
</tr>
<tr>
<td></td>
<td>• Efficiency</td>
<td>• Autonomy</td>
</tr>
</tbody>
</table>
2.3.2 Automatic assessment

2.3.2.1 VR simulators and sensors

Several studies have been done to address the issue of evaluating the user’s performance in the simulator-based training systems. High end virtual reality (VR) simulators use sensors that map movements into virtual space where they are analyzed by algorithms [52-54]. This method offers a degree of objectivity but suffers from practical limitations since such simulators introduce sensor mechanisms which interfere with the surgeons’ movements [55] and introduce adaptations that are not seen in real surgical environment. Moreover, there are human constraints, such as thinking of a surgical simulator as a videogame, which has no didactic value [56]. Prior experience with videogames can also be a handicap when facing virtual simulators [57]. It can even happen that such systems will create a false sense of security, built on the development of incorrect habits while getting used to a virtual environment [56]. The key then lies in developing an objective proficiency measurement system for surgery that can be (a) reliable, (b) replicable in real surgery and (c) affordable. In order to accurately measure a user’s skill objectively and quantitatively, the system must satisfy the following
requirements: (1) the system must possess adequate sensing techniques to monitor the user’s operation; (2) the system must extract relevant features from the sensing data; and (3) the system needs a good computational model to represent the skill demonstrated in the operation [11].

2.3.2.2 Video based system

During the MIS procedure, the camera and the surgical tools are introduced through the small incisions which allow access to the inside of the patient. The camera transmits an image of the organs inside the abdomen onto a television monitor. The surgeon is not able to see directly into the patient like open surgery, instead the video camera becomes a surgeon’s eyes in laparoscopy surgery and the surgeon uses the image from the video camera positioned inside the patient’s body to perform the procedure. This video can be analyzed with state-of-the-art computer vision algorithm (discussed in 2.4) to extract skill measuring features and then further analyzed by machine learning algorithms (discussed in 2.5) to construct a comprehensive skill assessment technique.

2.3.2.3 Formative vs summative assessment

Either formative or summative assessment could be used to help identify competency, ‘the skills, understanding and professional values of an individual ready for beginning independent dental or allied oral health care practice’ [18]. Formatively, it could be carried out through a longitudinal evaluation of performance over the year or it could be performed in a summative manner by a ‘snapshot’, structured clinical objective test of an individual’s level of competence.

Feedback is an informed, non-evaluative, objective appraisal of performance intended to improve clinical skills. It plays an important role in formative assessment and education as a whole. It is a key step in the acquisition of clinical skills [13]. Feedback highlights the differences between intended and actual result and provides impetus for change in the
student’s future performance of the same activity. It should be formative and without prejudice. It is in this respect that it differs from evaluation, which is summative. Feedback is a difficult part of clinical teaching [58]. If carried out properly it does, however, help the individual to determine, where they are in their skills development: beginner; novice; competent; expert [18].

Summative assessment is more comprehensive in nature and is undertaken towards the end of a program of teaching to ensure a student has reached a set of defined goals and objectives. Used with feedback, it too can provide guidance on how to improve the individual’s skills, in this case, however, the feedback is more of a debrief. Summative assessment is commonly used as an exam to the next stage of the course.

2.3.2.4 Offline assessment

The incorporation of cameras inside the box offers the opportunity to record common surgical procedures. In fact, in real LC, the camera view of the surgical procedure is recorded for later assessment and trainee self-review. This method of assessment has been proven highly effective in improving technical skills acquisition and self-assessment [19]. Use of appraisal using videotaping in surgery for assessing technical skills has been topical, with a few studies highlighting its value. One study showed blinded and independently rated videotapes of senior surgical residents performing laparoscopic surgery maintained very high inter-rater reliability despite assessors being permitted to use the “fast-forward” button [59].

Indirect assessment of videotaped procedures using raters located remotely from the procedure offers a number of advantages [19].

- Trainers may assess whenever convenient, thus increasing efficiency and limiting “examiner burnout” from fatigue and loss of concentration.
The indirect observation of surgical procedures, rather than those within a formal assessment structure, may capture actual practice, thus eliminating the Hawthorne Effect, in which improved performance is observed as a direct result of the subject knowingly being assessed.

De-identified videos offer anonymity, which decreases the chance of biased opinion. Reliability may be further improved by using multiple assessors.

Observing surgical procedure video also benefits trainees in self-assessment [20, 60]. This method of assessment has been proven highly effective in improving technical skills acquisition and self-assessment [19]. The recorded videos can be shared using internet-based tool for quicker distribution of the videos to the expert for assessment [21]. Thus an internet based video sharing tool is also proposed where recorded videos can be uploaded and assessed automatically by the developed algorithm.

2.4. Computer vision

Computer vision is the science of endowing computers or other machines with vision, or the ability to see. When we (human) look at an image, our brain divides the vision signal into many channels that stream different kinds of information into the brain. However, computer looks into an image as a grid of numbers which are usually the pixel values (Figure 2). Moreover any given number within can poses rather large noise component or distortion and so by itself gives very little information. The task of computer vision is then becomes to turn this noisy grid of numbers into a perception. Thus Computer vision is the transformation of pixel data from an image or video to any decision or any new representation. Application of computer vision decision can vary from edge detection, segment finding or shape detection where new representation can be converting color space or filtering an image to reduce noise.
Figure 2: Pixel value of image

2.4.1 Color space transformations
In the RGB model, images are represented by three components, one for each primary color – red, green and blue. However for many application in computer vision color image is converted into grayscale image for reduction of complexity and noise [61]. In an 8 bit grayscale image, every pixel has intensity from 0 to 255 where 0 denotes black and 255 is white (255 is $2^8-1$ where 8 represents number of bits). Figure 3 shows an example of RGB to grayscale image transformation.

Transformations within RGB space like adding/removing the alpha channel, reversing the channel order, conversion to/from 16-bit RGB color (R5:G6:B5 or R5:G5:B5), as well as conversion to/from grayscale using:

$$RGB[A] \text{ to Gray: } \leftarrow 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B$$

and

$$Gray \text{ to } RGB[A]: R \leftarrow Y, G \leftarrow Y, B \leftarrow Y, A \leftarrow 0$$

![Figure 3: RGB to grayscale](image)

Binary images are often produced by thresholding a grayscale or color image, in order to separate an object in the image from the background. Binary images are images whose pixels have only two possible intensity values - 0 for black, and either 1 or 255 for white. The color of the object (usually white) is referred to as the foreground color. The rest
(usually black) is referred to as the background color. However, depending on the image which is to be thresholded, this polarity might be inverted, in which case the object is displayed with 0 and the background is with a non-zero value. Figure 4 shows an example of grayscale to binary image transformation.

The function transforms a grayscale image to a binary image according to the formulas:

$$\text{dst}(x, y) = \begin{cases} \text{maxValue} & \text{if src}(x, y) > T(x, y) \\ 0 & \text{otherwise} \end{cases}$$

And inverse binary transformation is

$$\text{dst}(x, y) = \begin{cases} 0 & \text{if src}(x, y) > T(x, y) \\ \text{maxValue} & \text{otherwise} \end{cases}$$

where $T(x, y)$ is a threshold calculated individually for each pixel.

Figure 4: Grayscale to binary

Since human eye is strongly perceptive to red, green and blue color, RGB model is mostly used in hardware oriented application such as display monitors. However, color image is not simply formed with these three primary colors, rather human visual system characterizes color image by its brightness and chromaticity as well. Brightness is subjective measure of luminous intensity, and chromaticity is defined by hue and
saturation. These characteristics are defined in HSV (hue, saturation & value) color model. Hue is a color attribute and represents a dominant color. Saturation is an expression of the relative purity or the degree to which a pure color is diluted by white light. Value defines how light or dark a color is. The HSV color space can capture the distinct image features better than RGB color space [62, 63]. 3 channels hue (H), saturation (S) and value (V) in HSV color model is defined as [64]:

\[
H \left\{ \begin{array}{ll}
60(G - B)/S & \text{if } V = R \\
120 + 60(B - R)/S & \text{if } V = G \\
240 + 60(R - G)/S & \text{if } V = B \\
\end{array} \right.
\]

if \( H < 0 \) then \( H \leftarrow H + 360 \)

On output \( 0 \leq V \leq 1, 0 \leq S \leq 1, 0 \leq H \leq 360 \)

\[
S \left\{ \begin{array}{ll}
\frac{V - \min(R, G, B)}{V} & \text{if } V \neq 0 \\
0 & \text{otherwise} \\
\end{array} \right.
\]

\[
V \leftarrow \max(R, G, B)
\]

Figure 5 shows the HSV transformation of an RGB image.

Figure 5: RGB to HSV

2.4.2 Smoothing
Smoothing, also called blurring, is a simple and frequently used image processing operation. There are many reasons for smoothing, but it is usually done to reduce noise or camera artifacts. Smoothing is also important when we wish to reduce the resolution of an image in a principled way. The simplest filter is the mean filter where the gray-level pixel value is replaced by the mean value of the surrounding pixels. Similarly the median filter replaces each pixel by the median or “middle” pixel (as opposed to the mean pixel) value in a square neighborhood around the center pixel. Another useful filter is called the Gaussian filter where each pixel weight is inversely proportional to the distance from the central pixel (Figure 6) [65]. The two dimensional Gaussian filter is defined as [17, 66]:

\[
G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}
\]

where \(x\) is the distance from the origin in the horizontal axis, \(y\) is the distance from the origin in the vertical axis, and \(\sigma\) is the standard deviation of the Gaussian distribution and controls the mask size. Gaussian filtering is done by convolving each point in the input array with a Gaussian kernel and then summing to produce the output array. This is probably the most useful though not the fastest [16]. Figure 7 shows the smoothing effect using Gaussian blur.
Advanced morphological transformation [16] performs wide variety of contexts such as removing noise, isolating individual elements, and joining disparate elements in an image. Morphology can also be used to find intensity bumps or holes in an image and to find image gradients. The basic morphological transformations are called dilation and erosion. Dilation is a convolution of some image (or region of an image), with some
kernel with the anchor point at the center to compute the local maximum operator. Erosion is the converse operation which computes a local minimum over the area of the kernel. In general, whereas dilation expands region, erosion reduces region. Moreover, dilation will tend to smooth concavities and erosion will tend to smooth away protrusions. In morphological transformation dilate and erode can be used together:

\[
dst = \text{open}(src, element) = \text{dilate(ero}de(src, element), element)\]

Where,

\[
erode(x, y) = \min_{(x', y') \in \text{kernel}} src(x + x', y + y')
\]

\[
dilate(x, y) = \max_{(x', y') \in \text{kernel}} src(x + x', y + y')
\]

Figure 8 shows the output of morphological transformation in detecting surgical tool.

![Figure 8: Noise filtering](image)

2.4.3 Edge detection

Edge detection is one of the most important areas in lower level computer vision because the success of higher level processing relies heavily on good edges. An image contains an enormous amount of data much of which is relevant. To reduce some of the data, object is separated from background and entities like significant edges are identified. There are several edge detection techniques available that can be classified in three categories. The first category is the gradient operator. This is the simplest and includes 3x3 or 5x5 masks.
of Prewit, Robert, Sobel and Laplacian operators [16]. In the second category the operators are based on a kind of surface fitting model which includes Hueckel, Hartly and Haralick’s facet model approach [65]. Finally the third and the most popular category is employing the derivatives of Gaussian. Gaussian applies weighted averaging of the gray level image to remove noise hence works really well for real scenes. However these operators use fairly large masks and therefor computationally expensive [17].

One of the most popular edge detection techniques is Canny’s edge detector which uses the sum of four complex exponentials and can be approximated by the first derivative of a Gaussian [67]. One of the differences between the Canny algorithm and the simpler, Laplace-based algorithm is that, in the Canny algorithm, the first derivatives are computed in x and y and then combined into four directional derivatives. The points where these directional derivatives are local maxima are then candidates for assembling into edges. However, the most significant new dimension to the Canny algorithm is that it tries to assemble the individual edge candidate pixels into contours. These contours are formed by applying a hysteresis threshold to the pixels. This means that there are two thresholds, an upper and a lower. If a pixel has a gradient larger than the upper threshold, then it is accepted as an edge pixel; if a pixel is below the lower threshold, it is rejected. If the pixel’s gradient is between the thresholds, then it will be accepted only if it is connected to a pixel that is above the high threshold. Canny recommended a ratio of high:low threshold between 2:1 and 3:1. Figure 9 shows an example of Canny’s edge detection.
Computer vision has several techniques for representing shapes and boundaries in a segmented image which can be a useful tool for object recognition. The Hough transform is a method for finding lines, circles, or other simple forms in an image. The simplest case of Hough transform is the linear transform for detecting straight lines. In the image space, the straight line in given by

\[ y = mx + c \]

where \( m \) and \( c \) are the slope and \( y \)-intercept of the line. The equation can be re-written as

\[ c = (-x)m + y \]

This is an equation in the \( c \)-\( m \) space where the point \( x_0, y_0 \) is a straight line with slope \(-x\) and intercept \( y\). If we have several edges in \( x-y \) space \((x_1, y_1), (x_2, y_2), \text{etc.}\), they can be represented by different straight line in the \( c \)-\( m \) space. The lines will intercept at a point in the \( c \)-\( m \) space which is the estimated slope and intercept of a line in the \( x-y \) space. However, vertical lines pose a problem because of their infinite slope. Thus another parameterization of the line is used in polar coordinate [68]:

\[ p = x \cos \theta + y \sin \theta \]

where \( \theta \) is the angle between \( x \)-axis and a line \( p \), drawn perpendicular from the origin to the line being detected.
The Hough circle transform works in a manner roughly analogous to the Hough line transforms. The equation of a circle centered at \((x_0, y_0)\) with radius \(r\) is given by:

\[
(x - x_0)^2 + (y - y_0)^2 = r^2
\]

Since there are three unknowns \(x_0, y_0, r\), therefore the parameter space is also three dimensional. Figure 10 shows the detected circle from the edges using Hough transform.

![Figure 10: Detected circle (yellow)](image)

2.4.5 Region segmentation

Region segmentation is an complementary approach to the edge detection. In edge detection, an image is segmented by identifying the boundaries of object where in region based approach regions are segmented that are occupied by the object. An example of simple segmentation is converting a grayscale image to a binary image. A histogram which is a distribution of the gray levels can be used to determine the threshold used in binary images. A histogram graphs number of pixels in an image with particular value. Figure 11 shows the presence of 4 pixel levels in the corresponding pixel values of a grayscale image [65].
When RGB image is converted to HSV colorspace, the hue channel contains the color intensity information. At each location \((x, y)\) the intensity value of each pixel of hue channel is recorded and scaled to find out the hue histogram. Figure 12 shows histogram of the selected small rectangle where two dominant colors that represent the color of the objects.

When tracking, binary thresholding of a hue plane of each input video frame using this pre-computed histogram. Threshold is used to suppress weak colors. It may also make
sense to suppress pixels with non-sufficient color saturation and too dark or too bright pixels. Finally a connected component analysis is performed to find the largest component thus reducing noises. In Figure 13, the detected components with dominant colors are detected.

![Image](image.png)

Figure 13: Object segmentation

2.4.6 Motion detection

In a video, sequence of images is taken at a fixed time interval. The motion of objects in 3-D induces 2-D motion in the image plane. The motion is called optical flow. There are several methods for calculating optical flow. By applying frame-to-frame differencing technique to find object silhouette and motion history images (MHI), the dynamic part of the scene can be captured. Frame-to-frame differentiating function calculates absolute difference between two arrays of consecutive frames [16].

\[
dst = |src1 - src2|
\]

The function extracts templates by thresholding frame differences and then updates the motion history image by passing the resulting silhouette.

\[
mhi(x, y) = \begin{cases} 
\text{timestamp} & \text{if silhouette}(x, y) \neq 0 \\
0 & \text{if silhouette}(x, y) = 0 \text{ and } \text{mhi} < (\text{timestamp} - \text{duration}) \\
\text{mhi}(x, y) & \text{otherwise}
\end{cases}
\]
That is, MHI pixels where motion occurs are set to the current timestamp, while the pixels where motion happened far ago are cleared (Figure 14) [65].

Figure 14: Motion history image

The gradients of the resulting motion history image are taken to produce a mask of valid gradients. First the derivatives $Dx$ and $Dy$ of MHI are calculated to find the gradient orientation as:

$$
\text{orientation}(x, y) = \tan^{-1}\frac{Dy(x,y)}{Dx(x,y)}
$$

After that, mask is filled to indicate where the orientation is valid. For the local motion segments, small segmentation areas are first rejected and then the orientation is calculated using regions of interest (ROIs) that bound the local motions; it then calculates where valid motion within the local ROIs was actually found. Any such motion area that is too small is rejected. Figure 15 show the gradient of motion [65].
2.4.7 Limitations of Computer Vision

Surgical scene is three dimensional, which is any point in space is specified by three coordinates – x, y and z. On the other hand, surgical scenes videos are in 2-D plane, i.e. two coordinates are used to represents any point in the scene –x and y. Hence one dimension is lost in the process. However the use of stereo depth camera can be able to solve this issue.

2.4.8 OpenCV

The computer vision open source library (OpenCV) [16], developed by Intel Corporation, consists of a series of algorithms for motion analysis that has been developed in the last two decades. OpenCV is aimed at providing the basic tools needed to solve computer vision problems. In some cases, high-level functionalities in the library will be sufficient to solve the more complex problems in computer vision. Even when this is not the case, the basic components in the library are complete enough to enable creation of a complete solution of almost any computer vision problems.
2.5. Machine learning

Machine learning is defined as the field of study that gives computers the ability to learn without being explicitly programmed [69]. A computer program is said to learn from experience $E$, with respect to some task $T$, and some performance measure $P$, if its performance on $T$ as measured by $P$ improves with experience $E$ [70]. There are several different types of learning algorithms. The main two types are supervised learning and unsupervised learning. In supervised learning, we teach the computer how to do something, whereas in unsupervised learning we let the computer learn by itself. Regression problem in one type of supervised learning where the goal is to predict a continuous valued output. Where in the classification problem, the goal is to predict a discrete value output.

2.5.1 Linear regression

In supervised learning, a data set called a training set $(x)$ is used to predict the output variable $(y)$. In linear regression it is assumed that output $y$ is a straight linear function of $x$. A linear regression is called univariate linear regression if only one variable is used for predicting all outputs. And in linear regression when there are multiple features, is called multivariate linear regression.

Training dataset $x$ is the input variables often also called the features. And the output variable $y$ is the target variable which are predicted. The training set is used to feed a learning algorithm which outputs a function called hypothesis. Hypothesis is a function that maps from $x$'s to $y$'s that is it takes the value of $x$ and it tries to output the estimated value of $y$ for the corresponding $x$. For training set of $n$ number of input features, the hypothesis is defined as:

$$h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$
Here $\theta_0, \theta_1, \theta_2, ..., \theta_n$ are called the parameters of the function. The idea is to choose $\theta_0, ..., \theta_n$ so that $h_\theta(x)$ is close to $y$ for the training dataset $(x_0, y_0), (x_1, y_1), ..., (x_n, y_n)$. Sum of squared error are commonly used for linear regression problem which is called cost function:

$$J(\theta_0, \theta_1, \theta_2, ..., \theta_n) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

For each $\theta$ it repeatedly update the value of partial derivative of the cost function. It is called gradient decent technique.

$$\text{Repeat}\left\{ \theta_j := \theta_j - \alpha \frac{\delta}{\delta \theta_j} J(\theta_0, \theta_1, \theta_2, ..., \theta_n) \right\}$$

(simultaneously update for $\theta_j$ for $j = 0, ..., n$)

Where $\alpha$ is the learning rate which controls the step of the gradient decent. Now by substituting

$$\text{Repeat}\left\{ \theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} \right\}$$

(simultaneously update for $\theta_j$ for $j = 0, ..., n$)

Gradient decent is an iterative process. Normal equation is another method to solving for $\theta$ analytically. If training data set has $m$ example and $n$ features then $x$ is $m \times n$ dimentional vector, then normal equation can be used to find $\theta$ which is:

$$\theta = (X^T X)^{-1} X^T y$$

Normal equation does not require an intirative operation thus does not require to choose a step size $\alpha$. However it works slower if the training dataset too large where gradient decent works better.

2.5.2 Logistic regression
Most common type of supervised learning problem is called the classification problem where we predict discrete-valued outputs such as a zero-one valued discrete output. For logistic regression the hypothesis is defined as:

\[ h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}} \]

It is call the sigmoid function or simply the logistic function. So the cost function would look like:

\[
J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_\theta(x^{(i)}))
\]

Gradient decent can be applied in the case of logistic regression too where it repeatedly updates the value of \( \theta \) in the partial derivative of the cost function.

\[
\text{Repeat } \left\{ \theta_j := \theta_j - \alpha \frac{\delta}{\delta \theta_j} J(\theta) \right\} \quad (\text{simultaneously update for } \theta_j)
\]

However this process of logistic regression works for only binary output that is for two discrete classes. In case of multi-class classification problems, logistic regression can still work where the technique is called one-versus-all classification. In this technique there are same numbers of classifiers as the output. So whichever classifier gives the highest probability, the value of \( y \) then predicted to be that value. That is how the logistic regression classifier can work on multi-class classification problems as well.

2.5.3 ANOVA, PCA & LDA

Analysis of variance (ANOVA) is used to test for significant difference between means. It produces same result as t-test when there are more than two variables. In such situation, it is better to use ANOVA instead of multiple two sample t-test as it reduces the chance of committing error.
Principle component Analysis involves a mathematical procedure that transforms a number of correlated variables into a number of uncorrelated variables called principal components. It works mainly by discovering or reducing the dimensionality of the data set and by identifying new meaningful underlying variables. The mathematical technique used in PCA is called eigen analysis.

Linear discriminant analysis involves statistical analysis to find a linear combination of features which results in separation or characterization of two or more classes of objects. LDA is closely related to ANOVA by expressing one dependent variable as a linear combination of other measurements and to PCA by looking for the linear combinations of variables which best fit the data. All these there statistical analyses are useful tool to classification and clustering problems hence heavily used in this research.

2.5.4 Random Forest

Random forest is an ensemble classifier that uses many decision tree models to solve classification (categorical variables) or regression (continuous variables) problems [71]. Decision tree is s predictive model that uses a set of binary rules applied to calculate a target value. Rules in random forests are developed using computer software available. Random forest provides excellent accuracy among current algorithms. It can handle thousands of input variables without variable deletion and provides estimates of what variables are important in the classification. Also it generates an internal unbiased estimate of the generalization error as the forest building progresses. These capabilities of Random Forest can be extended to unlabeled data, leading to unsupervised clustering, data views and outlier detection [71].

2.5.5 WEKA

The Waikato Environment for Knowledge Analysis (WEKA) is developed by the University of Waikato, New Zealand as a solution to the perceived need for a unified
workbench that would allow researchers easy access to state-of-the-art techniques in machine learning [72]. Weka has incorporated several standard machine learning techniques into a software "workbench". With it, a researcher is able to use machine learning to derive useful knowledge from databases that are far too large to be analyzed by hand. It has a fully Java-based version and due to its graphical user interfaces it is extremely use-friendly. Also it is free software available under the GNU General Public License.

Weka supports several standard data mining tasks, more specifically, data preprocessing, clustering, classification, regression, visualization, and feature selection. Weka was heavily used in this project since it provides the solution to the major machine learning analysis required for this dissertation.

2.6. Medical education theory

Acquisition of skilled performance and the organization of knowledge using a range of methods, including experimental, quasi-experimental, and naturalistic methods has been a subject of interest to understanding how such theories provide a rationale for instructional design and learning is the nature and development of expertise and adaptive expertise [73]. Bloom’s taxonomy table can be used to classify the instructional and learning activities used to achieve the objectives and the assessments needed to evaluate students’ progress in achieving the objectives. In the revised Bloom’s taxonomies, all educational objectives can be classified according to the Knowledge and Cognitive Process dimensions [74]. Understanding of the fundamental principles underlying human learning and performance can greatly benefit the processes of technology-supported decision-making and problem-solving; the comprehension of information to deliver Internet-based health care; and the design and implementation of collaborative tools in our increasingly interconnected health care environment [75].
2.6.1 Why assessment is necessary

By looking into the nature of expertise and its development, a theory of effective instructional intervention can be designed to support curricular transformation. Study of expertise has influenced the domain of health education several ways: (1) it has formed the basis of expert technology-based systems to aid clinical practice, (2) it has provided a more comprehensive and detailed picture of clinical reasoning in medicine than the evaluation techniques commonly used in medical education, and (3) it has helped us to develop cognitive-based criteria for setting competency levels for professional training [76].

2.6.2 Different types of expertise

According to Vimla et al. [77-79] expertise research was conducted using two approaches: (1) absolute expertise, which studies “exceptional” experts in their respective domain of expertise and how their performance separates them from most individuals; and (2) relative expertise, which compares the performance of novices to experts where experts are assumed to have qualitatively different competence and performance than the novices, based on certain measures such as number of years of schooling, training, or experience in the domain.

Vimla et al [79] also defined four general levels of expertise that reflect a continuum from a beginner to highly-experienced professional. These include (1) novice, an individual who has only everyday knowledge of a domain or one who has the prerequisite knowledge assumed by the domain, e.g., first year medical student; (2) intermediate, an individual who is above the beginner level but below the sub-expert level, e.g., second year medical student; (3) sub-expert, an individual with generic knowledge but inadequate specialized knowledge of the domain, e.g., senior medical resident; and (4)
expert, an individual with a specialized knowledge of the domain, e.g., attending physician.

2.6.3 Simulation-based training

Simulation-based training has been mainly targeted towards psychomotor skill acquisition and orientation purposes, it is now increasingly being employed for fine tuning cognitive functions such as attention, decision-making and memory, providing environments for embodied learning. Specially the gaming-based simulations are very immersive and engaging which provide an interactive reward-based mechanism for learning and objective evaluation. Kahol and Smith [80, 81] have developed a generic methodology to develop simulation exercises or employ off the shelf simulation exercises for training surgeons which showed the positive effect of gaming on skills. Well-designed simulated learning environment provide “affordances” that are perceptually obvious to the user, making human interactions with objects virtually effortless [82]. Moreover by integrating real-time performance feedback in simulator can make simulations more engaging and interesting. However, in simulation-based training most learning are context-bound, where surgical knowledge is taught in relation to specific clinical problems to ensure their integration. However, this integration is often so context-dependent that its transfer to other situations is difficult [83]. Hence, performance measurement in actual surgical situation is also very important for understanding the cognitive science of human learning.
Chapter 3

RELATED WORK

3.1. Simulation based assessment

Most of the simulators use time as a metric for measurement of surgical proficiency and rewards faster performance with higher performance score [84]. Other measures of proficiency include kinematics of the laparoscopic tools. Tracking hand and surgical tool movement is one of the most important features in assessing surgical performance. Many sensor-based systems have been developed for accurate tracking of a surgeon’s hand or surgical tool movement. Pressure sensors were embedded in several simulators to observe the force applied by the surgeon’s while performing a surgical task [25, 85, 86]. A surgeon’s respiration rate [87] was also examined for any correlation useful to skill assessment. Some systems use generic measures such as smoothness of the movements as a measure of proficiency. The WKS system [85-87] measures force and movement of the dummy skin in a suture/ligature training system to evaluate performance. By using wireless sensor glove and body sensor network (BSN) technology [88], hand gesture data can be captured and analyzed with a Hidden Markov Model (HMM) for surgical skill assessment. Several systems have been developed to measure performance in actual surgery. The Wasada Bioinstrumentation (WB) system [89] uses a series of sensors to track head, arm and hand movement as well as several physiological parameters to analyze a surgeon’s performance during laparoscopic surgery. Sadahiro et al. used a force platform to measure fluctuations of an operator’s center of pressure (COP) [90] to estimate the skill level in the operating room. Most of these systems need multiple wearable sensors, which could interfere with an operator’s skill execution. Also the
sensors need to be sterile to be used in actual surgery and can make the entire surgery very expensive. All these work are summarized in the following tables-

Table 2: Hand movement tracking

<table>
<thead>
<tr>
<th>Author</th>
<th>Surgical Task</th>
<th>Technique</th>
<th>Drawback</th>
</tr>
</thead>
</table>
| Rosen et al. [23, 24, 42, 91] | Extracorporeal knot | • Captures hand movement data using sensor integrated gloves  
• Uses HMM to compare movement sequence | • Task specific  
• Sensor interference  
• Not transferable to real surgery |
| Saggio et al.   | Tissue dissection, extracorporeal suture, knot tying | • Captures hand movement data using sensor integrated gloves  
• Uses HMM to compare movement sequence | • Sensor interference  
• Not transferable to real surgery |
| King et al. [88] | Tissue dissection | • Captures hand movement data using sensor integrated gloves  
• Uses HMM to compare movement sequence | • Task specific  
• Sensor interference  
• Not transferable to real surgery |

Table 3: Tool movement tracking

<table>
<thead>
<tr>
<th>Author</th>
<th>Surgical Task</th>
<th>Technique</th>
<th>Drawback</th>
</tr>
</thead>
</table>

47
<table>
<thead>
<tr>
<th>Authors</th>
<th>Task/Tool Description</th>
<th>Features</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aizuddin et al. [25]</td>
<td>Suture/ligature</td>
<td>• Sensors integrated in the tool to measure applied force</td>
<td>• No significance analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Task specific</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Not transferable to real surgery</td>
</tr>
<tr>
<td>Solis et al. [85-87]</td>
<td>Suture/ligature</td>
<td>• Sensors integrated in the tool to measure applied force</td>
<td>• Task specific</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Not transferable to real surgery</td>
</tr>
<tr>
<td>Megali et al. [89, 92]</td>
<td>Touching the dissector nibs</td>
<td>• Captures tool movement data using sensor integrated tools</td>
<td>• No significance analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Uses HMM to compare movement sequence</td>
<td>• Task specific</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Sensor interference</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Not transferable to real surgery</td>
</tr>
<tr>
<td>Dosis et al. [5, 52]</td>
<td>General laparoscopic task</td>
<td>• Captures tool movement data using sensor integrated tools</td>
<td>• Sensor interference</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Assess skill by merging motion data with completion time</td>
<td>• Not transferable to real surgery</td>
</tr>
<tr>
<td>Wolpert et al. [93]</td>
<td>Touching the dissector nibs</td>
<td>• Captures tool movement data using sensor integrated tools</td>
<td>• Sensor interference</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Assess skill by merging motion data with completion time</td>
<td>• Not transferable to real surgery</td>
</tr>
</tbody>
</table>
Table 4: Capturing demographic features

<table>
<thead>
<tr>
<th>Author</th>
<th>Surgical Task</th>
<th>Technique</th>
<th>Drawback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sadahiro et al.</td>
<td>Pegboard transfer, cutting, suturing</td>
<td>• Uses force platform to captures surgeon’s center of pressure (COP) value</td>
<td>• Uses only surgeon’s COP value for analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Requires surgeons to stand steadily at a single point</td>
</tr>
<tr>
<td>King et al. [88]</td>
<td>5 laparoscopic gestures</td>
<td>• Captures hand movement data using sensor integrated gloves</td>
<td>• Task specific</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Uses probe on surgeons body to measure respiration rate</td>
<td>• Sensor interference</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Not transferable to real surgery</td>
</tr>
</tbody>
</table>

Table 5: Use of pressure sensors

<table>
<thead>
<tr>
<th>Author</th>
<th>Surgical Task</th>
<th>Technique</th>
<th>Drawback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zeeca et al.</td>
<td>Suture/ligature</td>
<td>• Uses pressure sensor in the dummy skin to measure pressure</td>
<td>• Task specific</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Uses motion sensors on the dummy skin to measure movement</td>
<td>• Not transferable to real surgery</td>
</tr>
</tbody>
</table>
| Mackel et al. [91] | e-pelvis examination | • Sensors integrated in the dummy skin to measure applied force | • Task specific  
| | | | • Not transferable to real surgery |
| Oshima et al. [25, 87] | Suture/ligature | • Sensors integrated in the dummy skin to measure applied force  
| | | • Assess skill by merging motion data with completion time | • No significance analysis  
| | | | • Task specific  
| | | | • Sensor interference |

Video-based approach is free from any wearable sensor. It uses video camera to track long sequence of hand/tool movement and then tries to extract different skill assessing features using computer-vision techniques. Because of its non-contact nature of the tracking technique, there is a minimal interference with the skill execution. Chen et al. proposed a computer-vision based approach to track the finger-tips and hand while tying a surgical knot [11]. It uses special markers on the gloves to record the hand movement from a very close distance. Finally, a HMM is used to to find the log-likelihood of the probability of the observation sequence to evaluate the performance for the given task. Lacey et al. proposed another video-based approach for observing continuous, long sequence of surgical tool movements in minimally invasive surgical (MIS) training, and then modeling and evaluating the psychomotor skill demonstrated in the observation [29]. Video data of the tool movement was obtained from a MIS simulator called ProMIS [18]. A computer vision algorithm was used to detect the tip of the surgical tool and calculate total distance travelled. Finally, the smoothness value was derived from the path length value of the surgical tool.
Apart from its obvious advantages over sensor-based system, both of these video-based systems have some drawbacks. They are presented in the following table-

Table 6: Major drawbacks of the other video-based systems

<table>
<thead>
<tr>
<th>Author</th>
<th>Surgical Task</th>
<th>Technique</th>
<th>Drawback</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>- Uses HMM to compare hand gestures</td>
<td>- Requires very close vision of the hand</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- No significance analysis</td>
</tr>
<tr>
<td>Lacey et al. [94]</td>
<td>General laparoscopic task using ProMIS</td>
<td>- Uses marker to capture tool-tip movement using external camera</td>
<td>- Analysis solely dependent on tool movement smoothness value</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Derives smoothness value from the path-length value of the tool</td>
<td>- Requires markers on the tool-tip</td>
</tr>
</tbody>
</table>

Different algorithms are used to assess surgical skill from all the features captured using either sensors or video-camera. All these algorithms can be classified into 3 major categories- regression models, data mining algorithms and statistical analyses. Regression models use data as parameters and develop equations to provide a scoring mechanism for the surgical tasks [25, 85, 86, 90, 94-96]. When multiple dimensional data are observed, various data mining tools are used to reduce the dimensionality of the
data so that the clusters can be observed in a two or three dimensional plane. This makes it easier to differentiate among the different skill levels [11, 23, 24, 42, 55, 88, 91, 92, 97]. However sometimes calculating simple statistical mean and variance could be sufficient to detect significant differences in skill levels [5, 87, 89, 93]. Despite its obvious importance in surgical teaching programs, the assessment of surgical skills has been studied infrequently and inadequately. Munz et al. performed objective evaluation of the performance on a series of simulated tasks such as open abdominal, laparoscopic, and intubation skills between a small group of junior and senior trainees. The results have shown no significant differences between participants for all procedural tasks regardless of grouping, level of training, or total number of years in training [98]. Feldman et al. compared objective assessment of technical skills and subjective in-training evaluations of surgical residents. However the experimental result was statistically insignificant [99].

3.2. Preliminary work

In the preliminary work of the proposed research, a significant difference in tool motion redundancy was observed between residents of 3 post-graduate years [100, 101] . Participants from Mayo Clinic, Rochester, MN performed peg transfer and intra-corporeal suturing using an FLS box [40] and free-hand knot-tying exercise using suture and pencils. Surgical tool tremor value for peg-board transfer and intra-corporeal suturing was extracted using computer vision algorithm and then normalized. The value of redundant tool motion was also derived from the tool tremor values and is presented in the following bar-chart. For the knot-tying exercise, the user’s hand was detected and redundant movement was analyzed. The classification method consists of two stages: training and testing. In the training phase, the classification method learns the distribution of the extracted motion features for the experienced, intermediate, and
novice categories. In the testing phase, the motion analysis data of a new resident are compared with the previously learned distributions of each group (experienced, intermediate, novice). The classification method then estimates the group to which the resident most likely belongs. For all three surgical tasks, a significant difference in motion redundancy was observed [100-102]. Motion redundancy is shown as the percentage of absolute motion in the following figures (Figure 16, 17 & 18).

![Peg Transfer](image)

Figure 16: Average motion redundancy value for peg transfer exercise

![Intracorporeal Suture](image)

Figure 17: Average motion redundancy value for intracorporeal suture exercise
Finally a two-way t-test was performed on the data (Table 7). Although there was no significant difference between PGY-1 and PGY-2 groups, significant differences in the measure of tool movement redundancy were observed between both the groups and PGY-3 [103]. This is consistent with surgical skills education in the residencies wherein third year residents show quantum increase in their skills due to more exposure to surgical procedures as compared to PGY1 and PGY2 residents [104].

Table 7: T-Test analysis of Mayo Clinic data

<table>
<thead>
<tr>
<th>ttest</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGY-1 : PGY-2</td>
<td>0.1427</td>
</tr>
<tr>
<td>PGY-2 : PGY-3</td>
<td>0.0492</td>
</tr>
<tr>
<td>PGY-1 : PGY-3</td>
<td>0.0297</td>
</tr>
</tbody>
</table>

A data set from Mount Sinai Hospital, New York, of a simulated intubation exercise has 14 users- 3 attending surgeons, 3 surgical residents (2 PGY-2s and 1 PGY-1) and 8 medical students. Hand movement for the entire dataset was analyzed and the average motion redundancy value for each group of users is shown in Figure 19. The amount of
movement acceleration significantly increases as the expertise level goes from attending to surgical resident to medical student. Moreover it is noticeable that the acceleration value is higher for the “difficult” level than the “easy” level of the same task [105].

![Intubation Graph]

Figure 19: Average acceleration value of user groups

Linear discriminant analysis (LDA) is a commonly used statistical technique for data classification [106]. In this study, we investigate whether a classification method based on LDA can be used in laparoscopy to recognize residents’ level of experience objectively (i.e., as experienced, intermediate, or novice) based on his or her psychomotor skills. The LDA analysis of the acceleration data is shown in Figure 20. User’s data roughly conform to 3 clusters. One of the surgical residents fall into expert’s cluster whereas another falls into novice’s cluster.
Figure 20: LDA Analysis

T-test analysis of the 3 user groups was also performed using Matlab. The t-test returns a very significant p-value \((p = 0.0032)\) between attending and medical students. Since the performances of resident fall into both expert and novice cluster, the p-value is comparatively higher (thus less significant) than for the other 2 groups.

Table 8: T-test analysis of Mount Sinai data

<table>
<thead>
<tr>
<th>Test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attending : Med Student</td>
<td>0.0032</td>
</tr>
<tr>
<td>Attending : Resident</td>
<td>0.0202</td>
</tr>
<tr>
<td>Resident : Med Student</td>
<td>0.0705</td>
</tr>
</tbody>
</table>

To find the effect of providing real-time feedback, a pilot study has been done where a total of 13 surgical residents’ (8 as control and 5 as test) pegboard transfer exercise tool video data have been analyzed. For each of the subject, motion smoothness values have
been extracted and mean motion smoothness value for both of the groups is plotted in Figure 21.

![Figure 21: Mean motion smoothness](image)

**3.3. Assessment in Real Surgery**

Although surgical assessment techniques follow structured scoring features, the ratings remain subjective and can vary widely from person to person. This inter-rater reliability issue provides the need for a computer assisted objecting assessment technique that can detect important features in a surgery and provide automatic objective scoring. To do so, successfully tracking of anatomical structures and surgical instruments are important. Few works have been done in this area. McKenna et al developed first markerless tracking system for tracking surgical instruments in monocular endoscopic videos. This method uses pixel’s color value as a discriminatory cue for whether or not it is occupied by a surgical instrument [107]. However the tracking performs poorly in real surgical videos and the method has not been used surgical performance assessment.
Akbari et al. describe an image processing method for artery detection [108]. The method detects the artery by sensing the change in movement of tissues over the artery based on the artery’s pulsatile movement. These changes are detected from images captured at systolic and diastolic times. However, requires additional information about the systolic and diastolic times. In real-time implementation, this would mean the BP monitor has to be synchronized with the system to detect if the image was captured at systolic or diastolic time [108]. While this is not directly relevant to our problem area, the detection of movements part does apply as a subsection of this research and clues from this algorithms can be used to look at micro-movement of surgeons.

Djaghoul et al analyze different segmentation and tracking methods and presented a new multi-modal method for automatic detection and markerless tracking of the gallbladder during the LC intervention[109]. This was primarily done to assess surgical movement of gall bladder and also assess patient safety. The idea was to also register images of gallbladder and sense its size as a measure of its infection. Although the method determines a coarse representation of the gallbladder in laparoscopic images, no work toward surgical proficiency assessment has been done.

Lahane et al used series of computer vision algorithms to track surgical instruments and coarse location of the cystic artery. It indicates the possibility of an injury to the cystic artery by automatically detecting the proximity of the surgical instruments with respect to the cystic artery[30]. The work has been done on simulated surgical videos and the study does not return any assessment score other than detecting proximity of the instruments to the cystic artery.

3.4. Internet based video assessment

Use of videotaping in surgery for assessing technical skills has been topical, with a few studies highlighting its value [19]. Some studies show the effectiveness of sharing video
using internet-based tool for quicker distribution of the videos to the expert for assessment [21]. Overall this is an underexplored area. However with the increase in internet penetration and cloud computing, this is an important area of exploration. American Board of Surgery and surgical residencies annually spends significant resources in supporting the evaluation program which is human intensive. For FLS too, there is a need to support evaluators program. Globally, an internet based system if validated could allow unprecedented access for surgeons and residents worldwide to world class evaluations. The key idea would be to use this system not to replace the current system but to augment it. For example, such a system can be employed for formative feedback.
Chapter 4

METHODOLOGY

4.1. Simulation

4.1.1. System architecture

The proposed system was tested with the Fundamentals of Laparoscopic Surgery trainer box. It analyzes the video image rendered from the box and two external cameras and finally displays feedback/output on the screen along with the tool video (Figure 22 & 23).

The components are:

- **Camera**: The FLS box contains a camera inside the box. It renders composite (RCA) to the display screen. However for further analysis of that video, a video converter (Pinnacle Dazzle) has been used in this project to convert the video from analog to digital format. The camera is calibrated as the exercise is centered to camera’s field of view. Two HD webcams (Logitech C270) are used to capture each of the hands.
movement. The user is asked to hold the surgical tool while touching two markers inside the box. This hand position is taken as initial position and the cameras are positioned to capture hand centered in the camera’s field of views. Videos from two external webcams are already rendered in digital format.

• Login: All the users were given user names and corresponding passwords. They were required to login at the beginning of each session. Each exercise video and performance score was stored against user’s login information. In storing the training data gathered during the training sessions, the system strictly followed HIPAA regulations.

• Processing system: The digital video format was sent to the PC through the USB port. Microsoft Visual Studio 2008 was used to write code in C/C++ to capture this video and then analyzed it using a set of computer vision and machine learning algorithms. Finally the video was sent to the display screen using serial or HDMI port or directly shown on the laptop screen.
4.1.2. Object segmentation

The FLS activities require surgeon to move certain objects. The object movements, their jerkiness and placement can inform decisions about surgeons’ proficiency. For example, gauze that is cut by the surgeon provides valuable information about the surgeons’ proficiency. The cut shape, its size and duration to execute that cut are all important cues about surgical proficiency. The system required hence an efficient object segmentation algorithm. The project tests a number of computer vision algorithms to analyze motion.

Color image is converted from RGB color space to Hue-saturation-value (HSV) color space and then further split into single channel Hue image. To isolate marker of the tool and hands, histogram of the Hue channel is calculated. Color of the tool-marker is red.
and the gloves are purple whose corresponding histogram value is found and binary thresholding applied to detect the tool and hand respectively. In this case, if the pixel value in the hue channel matches the histogram color value, it changes to 255 i.e. white. The rest of the pixels turn 0 i.e. black.

Finally to maximize the elimination of noise, advanced morphological transformation (discussed in Chapter 2) is used. Figure 24 and 25 show the tool and hand detection respectively.

Figure 24: Tool detection

Figure 25: Hand detection

4.1.3. Motion detection
Motion and its features as mentioned before is critical to surgical proficiency. After successful detection of the hand/tool, optical flow shows the gradient of resultant movement (Figure 26).

Both left and right hand and tool gradient values are recorded in Cartesian coordinate and degree. Coordinates are used to find the position of the tool analyzed to extract smoothness, extra movement and movement efficiency. Tremor is calculated by analyzing both the position and direction of movement. Angular values of the direction of hand movement are interpreted into 5 gestures – left, right, up, down and stationary for further extraction of features like perception of depth and hand movement efficiency. Table 9 summarizes the classification of gestures according to their angular direction of movement.

Table 9: Gesture Segmentation explain

<table>
<thead>
<tr>
<th>Angle</th>
<th>Direction</th>
</tr>
</thead>
</table>

Figure 26: Motion Detection

4.1.4. Feature extraction
4.1.4.1. Motion smoothness (tool video)

Smooth steady motion is one of the most important features for assessing surgical skill [110]. From the coordinates of the tool position, Euclidean distance between every frame is calculated. If the hand movement is $x$, then hand motion, acceleration and jerkiness can be found from the following equations -

$$\text{Hand motion} = \frac{dx}{dt}$$

$$\text{Hand motion acceleration} = \frac{d^2x}{dt^2}$$

$$\text{hand motion jerkiness} = \frac{d^3x}{dt^3}$$

Velocity, acceleration, jerk and snap of the tool movements which are the 1\textsuperscript{st}, 2\textsuperscript{nd}, 3\textsuperscript{rd} and 4\textsuperscript{th} derivatives for movement of 5 frames respectively are derived and summed up. Deceleration and negative jerk and snaps features are derived from it; however, deceleration and its higher derivatives are ignored since they have no association with smooth steady motion. Positive jerk value for the entire duration of exercise is calculated to find the counter feature i.e. motion smoothness [111].

4.1.4.2. Snaps (tool video)

Extra movement or snaps is another important factor that may be used to determine surgical proficiency. It is calculated by the 4\textsuperscript{th} derivative of hand movement [111]. In the preliminary finding of the proposed work, novices showed significantly higher number of

<table>
<thead>
<tr>
<th>Angle Range</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$45^\circ \leq \leq 135^\circ$</td>
<td>Up</td>
</tr>
<tr>
<td>$135^\circ \leq \leq 225^\circ$</td>
<td>Left</td>
</tr>
<tr>
<td>$225^\circ \leq \leq 315^\circ$</td>
<td>Down</td>
</tr>
<tr>
<td>$45^\circ \geq \geq 315^\circ$</td>
<td>Right</td>
</tr>
<tr>
<td>0</td>
<td>No movement</td>
</tr>
</tbody>
</table>
snaps than the experts. Total value of positive snaps is calculated to find this movement feature in every exercise.

\[ Snaps = \frac{d^4x}{dt^4} \]

4.1.4.3. Movement efficiency (tool video)

On surgical simulators, measures of economy of tool movement have been shown to be reliable, valid, and objective measures of technical competence [112]. Once the Euclidean distance between tool positions and the starting frame tool position is calculated for each frame, the values are averaged to find the tool movement efficiency. Figure 27 shows the combined hand movement trajectory for 5 experts and 5 novices (all right handed)-

- Green (left) and purple (right) dots represent 5 experts’ trajectory of tool motion.
- Blue (left) and red (right) dots represent 5 novices’ trajectory of tool motion.
- Novices have higher variance of movements than experts – poor movement efficiency.
- Novices show denser dominant hand movement than non-dominant hand – poor bimanual dexterity.
4.1.4.4. Tremor (tool video)

Tremor is an involuntary, roughly sinusoidal component inherent in normal human hand motion. It has been found to consist of a “mechanical-reflex” component which is thought to originate from the central nervous system and has a frequency range of 8 – 12 Hz \[110\]. Imprecision in laparoscopic surgery due to tremor has long been a concern. Surgical tool tremor is calculated by the motions that changes direction in 2 frames and as small as 5 pixels value. It is characterized by the following function-

```c
{
    {
        {
            tremor = true;
        }
    }
}
```

Figure 27: Trajectory of tool movement
else tremor = false;

}

4.1.4.5. Depth perception (hand video)
Repetitive motion towards the direction of the tool is conserved as up-down motion and accounted for the perception of depth. This extra hand movement is made mostly by the inexperienced residents due- 1) Translation of the 2-dimentional image of the operating field from the video screen into a 3-dimentional mental image [31]. 2) Learning to operate using long instruments and 3) Mastering ambidexterity and eye-hand coordination. This motion is calculated by the summing total number of hand motion in up and down direction.

4.1.4.6. Bimanual dexterity (hand video)
It is a measurement of how well the surgeon is able to optimize the use of both hands [ref]. If the resident ignores the non-dominant hand then the resident probably has not mastered bimanual dexterity. By calculating percentage of idle time compare to the total time to complete the exercise, bimanual dexterity is calculated.

4.1.5. Detecting Error

All the above-mentioned extracted features are not quantitatively perceivable to human. However each exercise has some task specific errors that are recorded by an expert observer. It requires the expert observer to be present during the exercise or to go over the video of the exercise to detect the errors. This research also proposes automatic and accurate detection of task-specific errors from the tool video. Error varies from task-to-task. There are different algorithms to monitor and record errors in peg transfer, intracorporeal suturing and shape cutting exercises.
4.1.5.1. Peg transfer

The program detects all the objects and counts the total number of it by counting the total number of pixels inside the blue rectangle (Figure 28). If at the end of the exercise not all the objects are inside the rectangle then it returns the number of object missing as error.

Figure 28: Drop detection

4.1.5.2. Intracorporeal suturing

During intrecorporeal suturing, if the painrose drain is pulled excessively from its original location then it is considered as tissue damage [113]. The program detects the painrose drain and automatically calculates deviation from its original position during the exercise (Figure 29).
4.1.5.3. Shape cutting

In shape cutting exercise, after the exercise the user keep the piece under the camera and upon pressing ‘spacebar’, it automatically detects the inside and outside cuts. The program saves a snapshot and converts from RGB to grayscale image. Then it performs Gaussian blur filter to smooth the image and then Canny’s filter to find edges. Finally the Hough transformation finds the actual circle and 2 binary images were produced for the contents outside and inside the circle respectively (all these steps are explained in Chapter 2). Figure 30 shows the result of the detection. By counting the number of white pixels, the program automatically finds the outside and inside imperfection and provides the combined precision of cut score.
4.1.6. Multivariable Linear Regression

The computer vision algorithm extracts six features from each of the laparoscopic cholecystectomy videos and its corresponding hand movement videos. However, each of the videos is rated by expert surgeons in three categories: smoothness, efficiency and precision. Three of the extracted features are grouped into a single feature the experts rated. Multivariate linear regression models are used to find the relationship between extracted features and experts’ scores.

\[
h_a(\text{smooth}) = a_0 + a_1(\text{jerkinesss}) + a_2(\text{snaps}) + a_3(\text{tremor})
\]

\[
h_b(\text{efficiency})
= b_0 + b_1(\text{movement efficiency}) + b_2(\text{depth perception})
+ b_3(\text{bimanual dexterity})
\]

Precision score is solely dependent on the time to complete exercise and committed errors. After discussing with expert surgeons, equation for measuring error for each of the surgery is given below-
\[
error_{\text{peg transfer}} = \frac{1}{5} \times ((\text{completion time (if } > 48\text{)} - 48\text{)sec} + (\text{no. of drops} \times 25\text{)sec})
\]

\[
error_{\text{Intracorporeal suturing}} = \frac{1}{12} \times ((\text{completion time (if } > 112\text{)} - 112\text{)sec} + \text{tissue damage})
\]

\[
error_{\text{Shape cutting}} = \frac{1}{10} \times ((\text{completion time (if } > 98\text{)} - 98\text{)sec} + \text{cut imperfection})
\]

*48s, 112s and 98s are the average time for experts to complete the exercises respectively.

### 4.1.7. Final scores

Once the exercise is complete, the system calculates scores for each feature and will display it on the screen along with a composite score. Features from hand and tool video are normalized to 100. Finally the following scoring scheme is derived for each of the exercises to calculate the composite score:

\[
\text{score}_{\text{peg transfer}} = \frac{h_a(\text{smooth}) + h_b(\text{efficiency})}{2} - error_{\text{peg transfer}}
\]

\[
\text{score}_{\text{Intracorporeal suturing}} = \frac{h_a(\text{smooth}) + h_b(\text{efficiency})}{2} - error_{\text{Intracorporeal suturing}}
\]

\[
\text{score}_{\text{Shape cutting}} = \frac{h_a(\text{smooth}) + h_b(\text{efficiency})}{2} - error_{\text{Shape cutting}}
\]

### 4.1.8. Study Design

We designed and developed a novel, video-based and automatic surgical skill assessment system. Now, to evaluate the performance of the system, a proper design of experiment is required that can prove the proposed hypotheses.
4.1.8.1. Automated score vs. manual score

The proposed surgical skill assessment technique is expected to be credible enough to replace the traditional system where an expert rater rates a surgical exercise subjectively. Videos of surgeon’s hand and surgical tool movements of four surgical exercises will be collected to be scored by the system. Each of these videos will be scored by 3 expert raters. The average score for each video would be considered its gold standard. Next the proposed system will be used to obtain the score for each video and the difference between the gold standard score and system suggested score will be calculated. Finally the mean of this difference will be used to perform a paired t-test to test the null hypothesis that the two scores are equivalent.
To find the required sample size, a priori power calculations have been done assuming the value of alpha and the power that is expected to achieve. To find the effect size, result of the intra-corporeal suturing has been chosen from the pilot study (Figure 32). It demonstrates lowest difference in motion smoothness between the three skill levels. Mean and standard deviation of the three expertise levels are shown in Table 10. Along with these values alpha and power have been assumed to be 0.05 and 0.8 respectively and the sample size has been calculated for each pair of the expertise levels using G*Power tool. The result is presented in Table 11. Since the study shows the difference between experts and novices, video data will be collected from 10 experts and 10 novices for the evaluation of the system.

Table 10: Mean and SD of skill levels in intra-corporeal suturing

<table>
<thead>
<tr>
<th></th>
<th>Novice</th>
<th>Intermediate</th>
<th>Expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>37</td>
<td>31</td>
<td>24</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>10</td>
<td>6</td>
<td>5</td>
</tr>
</tbody>
</table>
Figure 32: Tool motion smoothness value for intra-corporeal suturing

Table 11: Sample size calculation

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.8</td>
<td>68</td>
<td>15</td>
<td>10</td>
</tr>
</tbody>
</table>

Experiment 1- Stratification of expertise

To design the experiment, power of 80 with a two tailed alpha of 0.05 were chosen which gave 8 subjects in each group. However we chose to collect data from at least 10 subjects in each expertise group – expert, intermediate and novice. An expert is defined as a surgeon with at least 5 years of experience after board certification. For each user data was collected for 4 pegboard transfer, 4 inter-corporeal suturing and 4 shape cutting exercises. 2 exercises for each session were taken as warm-up and not included in the analysis. The last two were included as the scores. The de-identified videos were rated by an independent panel of 3 expert raters (board certified surgeons). The videos were
presented in a randomized fashion and differently to all raters. Inter rater reliability was established by Fleiss’ Kappa and Pearson’s correlation coefficient. Videos with low inter rater reliability were discarded. The board ratings of experts, intermediate and novices were stratified. If outliers were detected in any group they were removed from the analysis (removes anybody in the group at least 2 standard deviations away from mean scores). Stratified scores of experts, intermediate and novices were grouped for all exercises and compared separately for each exercise using ANOVA. Automated score was generated by the algorithm. The correlation between experts’ average ratings and the automated score was found. A high correlation will show that the algorithm can generate ratings consistent with expert analysis.

4.1.8.2. Training efficacy

The proposed study involves one control group and one test group. No feedback is provided to the control group. However the test group is provided real-time feedback on motion smoothness, movement efficiency, number of committed errors and elapsed time. Each subject from both of the groups repeats a single surgical exercise in a three weeks period. For each of the subject, performance score for each parameter is calculated and added up to form a composite score. A curve of the composite score is plotted over time using the Matlab program which is called the learning curve. The learning curve for the test group is expected to be steeper than that of the control group.

To find the sample size, a priori power analysis has been done assuming the value of alpha and the power that is expected to achieve. The effect size has been calculated from the pilot study using the following equation:

Effect size = Difference in means / Common standard deviation
From the pilot study, mean and standard deviation of slopes of the controlled group and test group have been used to find the effect size using the same G*Power tool. Finally a sample size of 20 subjects for each group has been found for a power of 0.8.

Table 12: Sample size calculation

<table>
<thead>
<tr>
<th>Mean</th>
<th>Common Standard Deviation</th>
<th>Alpha</th>
<th>Power</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Group (8)</td>
<td>0.074</td>
<td>0.03</td>
<td>0.05</td>
<td>20</td>
</tr>
<tr>
<td>Test Group (5)</td>
<td>0.097</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Experiment 2 – Formative Feedback vs. Summative Feedback

According to the design, at least 20 novices data were collected in each group – control and experimental. No real time feedback or assessment score were provided to the control group. Where the experimental group received performance feedback and detailed assessment score at the end of each exercise. Novices from each group were asked to perform 20 iterations of pegboard transfer and shape cutting exercises in 5 sessions. 1 exercise for each session is taken as warm-up and is not included in the analysis. The last 2 are included as the scores. The de-identified videos were rated by an independent panel of 3 expert raters (board certified surgeons). The videos were presented in a randomized fashion and differently to all raters. Inter rater reliability was established by Fleiss’ Kappa and Pearson’s correlation coefficient. Videos with low inter rater reliability were discarded. Automated score was generated by the algorithm. ANOVA was performed between control group and experimental group for the performance. A p<0.05 is taken as a statistically significant difference.

4.1.9. Venue and Participants
The prototype has been built and video data is being collected from the simulation centers of several hospitals in the nation-

- Mayo Clinic Hospital, Phoenix, AZ
- Banner Health Good Samaritan Hospital, Phoenix, AZ
- Mayo Clinic, Rochester, MN
- Mount Sinai Hospital, New York, NY

Data for 3 surgical training tasks is collected- peg transfer, intracorporeal suturing and shape cutting. The participants are postgraduate year 1st to 5th (PGY1-PGY5) surgical residents and attending surgeons. The subjects are asked to perform pegboard transfer and the tool movement video data will be analyzed to compare the learning curve.

4.2. Performance in actual surgery
4.2.1. Data collection

We have collected video data of LC performed by different years of PG residents. Surgical scene video (tool video) is readily available as it is recorded for assessment purpose. To capture surgeon’s hand movement (hand video), a HD camcorder is mounted on the IV
stand from where it has a direct view to the surgeon’s hand and incision ports (Figure 33). Finally these two videos are synchronized and combined in a single video file to be assessed by experts and developed system.

4.2.2. Expert’s evaluation

Each Cholecystectomy (gallbladder removals) video is divided into 3 clips:

- Inferior surface of the right lobe of the liver is retracted upwards by retractor.
- Cystic Duct and Artery are dissected first and divided: The cystic duct and the cystic artery are identified, artery is clipped with 3 tiny titanium clips and both artery and duct are cut
- Gallbladder is removed

These are identified by surgeons as the three phases of the surgery.

Each clip of the de-identified videos is rated by an independent panel of 3 expert raters (board certified surgeons).

We developed a 3-item global rating scale that are scored using a 5-point Likert scale where “1” represents the lowest level of performance, and “5” is considered ideal performance. The clips are rated in terms of smoothness, efficiency and preciseness. The videos are presented in a randomized fashion and differently to all raters. Inter rater reliability is established by Fleiss’ Kappa.
4.2.3. Computer vision tool

The following two tools and hand videos are analyzed to detect objects and find movement features. A series of computer vision functions are used to successfully segment objects and extract motion features [16].

4.2.3.1. Object segmentation
Color image is converted from RGB color space to Hue-saturation-value (HSV) color space (Figure 35). The HSV image of the tool video is split into 2 images with single channels- Hue and Saturation. Hue channel is used to isolate gallbladder from the scene and Saturation channel is for detecting surgical instruments.

![Figure 35: RGB to HSV](image)

Due to the absence of color, its saturation value is quite low for the surgical tool-tips. Inverse binary threshold is used to convert the lower threshold value of the surgical tool to white. However the resulting frames contain large amount of noise resulting from other objects with lower saturation value [114] (Figure 36).

![Figure 36: HSV to Saturation](image)

In the tool video, surgical instruments are the major moving part in the scene. By applying frame-to-frame differencing technique to find object silhouette and motion history images (MHI), the dynamic part of the scene can be captured. Frame-to-frame
differentiating function calculates absolute difference between two arrays of consecutive frames. The function extracts templates by thresholding frame differences and then updates the motion history image by passing the resulting silhouette. That is, MHI pixels where motion occurs are set to the current timestamp, while the pixels where motion happened far ago are cleared. Figure 37 shows the MHI.

Figure 37: RGB to Motion History Image (MHI)

Now, both MHI frames and thresholded saturation frames have primarily surgical tool highlighted. By performing pixel-wise AND operation over both the frames, we can accurately detect the surgical instrument [30].

$$dst(I) = \{src1(I)\} \& \{src2(I)\}$$

Finally to maximize the elimination of noise, advanced morphological transformation is used. Figure 38 shows the result of AND operation between the binary Saturation channel and MHI and the final result after applying morphological transformation.
To isolate gallbladder, histogram of the Hue channel is calculated. Color of the gallbladder is yellowish-green whose corresponding histogram value is found and binary thresholding applied to detect the gallbladder. In this case, if the pixel value in the hue channel matches the histogram color value, it changes to 255 i.e. white. The rest of the pixels turn 0 i.e. black (Figure 39).

In case of the hand video, hands are detected by primarily transforming the RGB image to Hue channel. Motion history images are also constructed from the RGB image. Finally, by applying pixel-wise AND operation and morphological transformation, successful detection of hands is achieved.
4.2.3.2. Motion detection

Once the surgical tools and hands are detected, the gradients of the resulting motion history image are taken to produce a mask of valid gradients. After rejecting the very small motions, resultant local motion is found (Figure 40). If the combined local motion is greater than a threshold value then it finds the global motion of the entire scene which is associated with camera movement.

![Figure 40: Gradient of motion](image)

4.2.4. Feature extraction

Automated score of smoothness, efficiency and preciseness are generated by the algorithm. Each of the features is multidimensional by its nature and it is important to develop a system that captures each of the dimensions. This requires different algorithms for each of the proficiency measures. Figure 41 shows the captured dimensions for each feature:
4.2.4.1. Smoothness

- Smooth movements: Smooth steady motion is one of the most important features for assessing surgical skill [110]. Using the same algorithm as the tool video for FLS simulator, motion features like smoothness, snaps and tremor are calculated by analyzing both the position and direction of movement.

- Depth perception: Hand video that is captured by the digital camcorder attached on the IV stand is analyzed using optical flow algorithm. Repetitive motion towards the direction of the tool is conserved as up-down motion and accounted for the perception of depth. This extra hand movement is made mostly by the inexperienced residents due to- 1) Translation of the 2-dimentional image of the operating field from the video screen into a 3-dimentional mental image [31]. 2) Learning to operate using long instruments and 3) Mastering ambidexterity and eye-hand coordination. A poor depth perception might result activating the cautery before contact with tissues and frequent overshooting to accurate direction of the
instruments in the correct plane [44]. This motion is tracked from the hand video and counted as score to lack of perception of depth.

4.2.4.2. Efficiency

- Bimanual dexterity: It is a measurement of how well the surgeon is able to optimize the use of both hands [44, 48]. If the resident completely ignores the non-dominant hand then the resident probably has not mastered bimanual dexterity. By tracking tool-tip movement, percentage of time with both the tool movements is calculated by the system.

- Frequent change of tool: By observing the surgical videos performed by the expert surgeons, it is found that a surgeon change the instruments 10 times on average to complete the entire procedure [115]. However this number varies from 8 to 25 in case of residents. As the complexity arises in the surgery, number of instrument change increase rapidly. In the hand video, usually hand movements are very subtle, however during the change of instruments, the motion detection algorithm returns a peak value which helps to count the number of times tool is changed (Figure 42).
4.2.4.3. Preciseness

- Presence of blood: Among the complications during LC, hemorrhage is usually more difficult to control than that during open cholecystectomy as blood tends to obscure the operative field [34]. The gallbladder is cut away using either a laser or electrocautery device; both of which use localized heat to minimize bleeding [30]. However clipping and dissecting the cystic artery is crucial. Since the cystic artery supplies oxygenated blood to the gallbladder, uncontrolled bleeding from the cystic artery and its branches is a serious problem that may increase the risk of intraoperative lesions to vital vascular and biliary structures [34]. Color of the gallbladder is yellowish-green which is easily detectable in the scene using the computer vision object segmentation algorithm (Figure 43). However the presence of blood makes the gallbladder undetectable. The algorithm calculates the average number of yellowish-
green pixels in the first 100 frames and last 100 frames; and compares them to detect the presence of blood in the scene.

Figure 43: Ruptured cystic artery causes hemorrhage

- Tissue handling: Laparoscopic instruments are long and inserted through trocars. It causes fulcrum effect and the operator experiences the relative loss of tactile feedback [44]. Adaptation to this situation can be measured by looking into poor grasper control, grasper frequently slips and abrupt movements that is rough with tissue [44]. Using motion detection algorithm it is feasible to track snaps (grasp a tissue and slips) and rough movements which the system automatically counts and averages them over time.

4.2.5. Multivariable Linear Regression

The computer vision algorithm extracts six features from each of the LC videos and its corresponding hand movement videos. However, each of the videos is rated by expert
surgeons in three categories: smoothness, efficiency and precision. Two of the extracted features are grouped into a single feature the experts rated. Multivariate linear regression models are used to find the relationship between extracted features and experts’ scores.

\[
h_a(\text{smooth}) = a_0 + a_1(\text{smooth movement}) + a_2(\text{depth perception})
\]

\[
h_b(\text{efficiency}) = b_0 + b_1(\text{bimanual dexterity}) + b_2(\text{tool change frequency})
\]

\[
h_c(\text{precise}) = c_0 + c_1(\text{presence of blood}) + c_2(\text{snaps})
\]

Where \(a_i, b_i\) and \(c_i\) are the regression parameters. Weka is used to assess the regression parameters and to run the regression model. Entire dataset is normalized between 0 to 5 and then randomly divided into 60% training data, 10% cross-validation data and 30% test data. Cross-validation data is used to select the appropriate regression model.

4.2.6. Study Design

According to the design, at least 30 laparoscopic cholecystectomy videos planned to be analyzed. The de-identified videos were rated by an independent panel of 3 expert raters (board certified surgeons). The videos were presented in a randomized fashion and differently to all raters. Inter rater reliability was established by Fleiss’ Kappa and Pearson’s correlation coefficient. Videos with low inter rater reliability were discarded. Automated score was generated by the algorithm. The correlation between expert average score and the automated score was found. A high correlation would show that the algorithm can generate ratings consistent with expert analysis.

4.3. Cloud based tool

Our study also developed an cloud based tool where user can upload FLS exercise videos and observe immediate assessment score. Users can also track their progress by observing the past scores. All the uploaded videos are de-identified and added to an
online surgical video repository. Expert raters can observe these videos and provide performance score for smoothness, precision and efficiency features. Once a video will be rated by minimum of three expert raters with an acceptable inter-rater reliability, the score of the video will be used as a training data to train back the neural network. Figure 44 shows the sitemap of the cloud based tool.

4.3.1. System architecture

We used EasyPHP to build the webtool. EasyPHP is a software bundle that installs web server services onto the Windows computer and allows quick-and-easy development of PHP and MySQL on a localhost. The package includes an Apache server, a MySQL database, and the PHP extension. The Apache HTTP Server, commonly referred to as Apache, is a web server software program notable for playing a key role in the initial growth of the World Wide Web. MySQL is an open source relational database management system that runs as a server providing multi-user access to a number of databases. The SQL phrase stands for Structured Query Language. PHP is a server-side scripting language designed for web development but also used as a general-purpose programming language. While PHP originally stood for Personal Home Page, it now stands for PHP: Hypertext Preprocessor, a recursive acronym. PHP code is interpreted by a web server with a PHP processor module which generates the resulting web page: PHP commands can be embedded directly into an HTML source document rather than calling an external file to process data. It has also evolved to include a command-line interface capability and can be used in standalone graphical applications.
4.3.1.1. Registration and log in

Log in is the first page of the browser. It provides option to login as a user (surgical resident) or rater (expert surgeon) (Figure 45). For the 1\textsuperscript{st} time users, the page also provides as option “Not registered yet” which will take user to sign up for a user or rater account (Figure 46). All user account information is stored in a secured SQL database.
4.3.1.2. Automatic video evaluation

Once sign in, user can upload videos of FLS training or real LC and get automatic evaluation scores (Figure 47). User chooses the type of the exercise and selects the video file to upload (Figure 48). Depending on the internet connection speed and the size of the video file, uploading video file varies. Once the file is successfully uploaded, it takes only 10s-30s to evaluate the video. The software provides scores for individual features.
as well as the composite score. All these uploaded videos are deidentified (with user’s discretion) and stored in the secured server for expert raters’ evaluation.

Figure 47: Available actions

![Available actions](image)

**Figure 47: Available actions**

Figure 48: Upload video for automatic evaluation

![Upload video](image)

**Figure 48: Upload video for automatic evaluation**

4.3.1.3. Check progress

At any point user can review the scores for previously uploaded video (Figure 49). It can be a useful tool for observing the learning progress.
4.3.1.4. Rating a video

Expert surgeons sign in to the web tool and play any de-identified videos. They are asked to rate the videos in smoothness, efficiency and preciseness category (Figure 50). The tool keeps track of the scores and once a video gets score from 3 expert raters, it removes the video from the video pool that is available to the expert raters.
At every page of the web tool, the user (or the rater) has the option to log out from the tool (Figure 51). It is extremely important in case of shared system for proper and secured use of individual’s information. Any user can accidentally upload a video while
logged on to the previous users account and can cause unwanted changes in the learning progress data.

Figure 51: Logout page
Chapter 5

RESULTS

5.1. FLS simulator

Video data of the users consisting of medical students and different post graduate year (PGY) of surgical residents were collected while practicing peg transfer, intra corporeal suturing and shape cutting exercise using FLS box. Each subject conducted the exercise two consecutive times and videos of the second trial are analyzed. Each exercise data was captured with three synchronous cameras - two cameras for capturing each of the hand movements and one camera for capturing the tool movement. Total 52 sample video were collected and grouped as shown in Table 13.

Table 13: Distribution of subjects

<table>
<thead>
<tr>
<th>Group</th>
<th>Experience Level</th>
<th>No of subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert</td>
<td>PGY5 - PGY4</td>
<td>10</td>
</tr>
<tr>
<td>Intermediate</td>
<td>PGY1 - PGY3</td>
<td>10</td>
</tr>
<tr>
<td>Novice</td>
<td>Medical Students</td>
<td>32</td>
</tr>
</tbody>
</table>

5.1.1. Performance evaluation

Surgical tool and hand videos were analyzed for all the frames. Arrays of position and movement gradient were captured from each frame where each sample consisted of both left and right tools/hands. The angular value of the movement gradient was converted to four motions-left, up, right and down. Several algorithms analyze the displacement and direction of movements and prepared an occurrence matrix for each of the gestures.

Figure 52 shows jerkiness, snaps and tremor scores of three levels of expertise for both left and right surgical tools. The plot clearly shows that experts have less numbers of jerky hence smoother movements than intermediates and novices. Total number of
snaps was also calculated and average snaps among different expertise group are plotted in Figure 52. Tremors are very short movements and very difficult to analyze subjectively. However, the motion detection algorithm is able to detect very subtle movement and the system calculates tremors for both the hands and tools. Figure 52 also shows tremor features where it shows an increment of tremor as the experience level decreases. Due to the use of longer instruments, very subtle hand movement can cause greater tool movement. Hence, more numbers of tremors are noticeable in tool analysis.

![Bar chart showing Jerkiness, Snaps and Tremor scores](figure52.png)

**Figure 52: Jerkiness, Snaps and Tremor scores**

Motion redundancy for both the hand movements was analyzed. More than 65% of the redundant motion was observed in the up-down direction. It results from extra motion in perception of depth. Figure 53 plots motion redundancy in the up-down direction and it appears that as the level of expertise increases, perception of depth increases hence redundant motion in vertical direction reduces. Also the average traveling distance for the tool was calculated and found that experts have about 40% less movement than novices to complete the same exercise.
Percentage of time the non-dominant hand has any motion compared to the dominant hand was calculated to find bimanual dexterity. Both intermediates and novices showed almost 40% less activity in the non-dominant hand than experts group (Figure 54).
Table 14 presents the results of ANOVA among the different skill levels which is very significant for all the six features. p<0.05 was taken as statistically significant.

Table 14: ANOVA among expertise levels

<table>
<thead>
<tr>
<th>Gestures</th>
<th>Expert vs Intermediate</th>
<th>Intermediate vs Novice</th>
<th>Expert vs Novice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tool</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoothness</td>
<td>0.0032</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Snaps</td>
<td>0.0425</td>
<td>0.0325</td>
<td>0.00345</td>
</tr>
<tr>
<td>Movement efficiency</td>
<td>0.0295</td>
<td>0.0002</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Tremor</td>
<td>0.0002</td>
<td>0.0045</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Hand</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depth perception</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Bimanual dexterity</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Random Forest classifier was used for the classification of each gesture data. 66% of the data for both hand and tool gestures were used as training data. Rest 34% of data was unlabeled and used for testing of each gesture [116]. Each gesture shows significantly high true positive rate (Table 15).

Table 15: Random Forest Analysis

<table>
<thead>
<tr>
<th>Gestures</th>
<th>True positive</th>
<th>False positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tool</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoothness</td>
<td>72.20%</td>
<td>17.70%</td>
</tr>
<tr>
<td>Snaps</td>
<td>66.70%</td>
<td>34.80%</td>
</tr>
<tr>
<td>Movement efficiency</td>
<td>60.00%</td>
<td>29.70%</td>
</tr>
<tr>
<td>Tremor</td>
<td>80.00%</td>
<td>4.50%</td>
</tr>
<tr>
<td>Hand</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depth perception</td>
<td>72.20%</td>
<td>16.00%</td>
</tr>
<tr>
<td>Bimanual dexterity</td>
<td>94.40%</td>
<td>1.60%</td>
</tr>
</tbody>
</table>
Motion smoothness, motion redundancy and tremor value for both hand and tool gestures were normalized and the Linear Discriminant Analysis (LDA) was performed [116]. Figure 55 shows the LDA analysis result of both hand and tool gestures where User’s data roughly conform to 3 distinct regions in a 2-dimensional projection space. An intermediate surgeon’s motions tend to not separate as well, indicating less consistent motions. These initial experiments validated the hypothesis that LDA could be used to simplify the original data into a simpler, low-dimensional data set. Also features from the tool movement were more accurate in contrasting the expertise level than that of hand movement.

![Figure 55: LDA analysis](image)

5.1.2. Error detection

For error detection, automatic error score were compared with observer’s score for each of the videos. The system is able to detect drops very accurately. The overall sensitivity of the system was 87% (Table 16).

Table 16: Error detection

<table>
<thead>
<tr>
<th>Exercise</th>
<th>Error</th>
<th>Correlation coefficient</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pegboard transfer</td>
<td>Drop count</td>
<td>0.98</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Intracorporeal suturing</td>
<td>Tissue damage</td>
<td>0.77</td>
<td>0.0008</td>
</tr>
</tbody>
</table>
Shape cutting    | Precision of cut    | 0.86    | <0.0001

5.1.3. Efficient skill learning

The proposed video-based surgical skill assessment technique can provide immediate feedback; hence it was also being tested as a tool for the efficient skill learning technique. 32 medical students (novices) were used as control group i.e. no on-screen feedback or assessment scores were provided. 22 medical students were used as the experimental group where they were provided real-time on-screen feedback and assessment score at the end every trial. All participants in each group have performed the peg transfer and shape cutting exercise at least 16 times each. Finally, the average score of each of the group is plotted in Figure 56 where an improved learning curve is observed for the experimental group. Paired t-test is performed which results a significant difference in performance between the two groups (p<0.0001) [116].

![Figure 56: Learning curve](image-url)
Average performance scores for both the control and experimental groups were calculated. Figure 57 shows significant improvement (11%) in the psychomotor skill learning performance of the experimental group over the control group.

![Figure 57: Effect of assessment system on learning](image)

### 5.2. Laparoscopic Cholecystectomy

#### 5.2.1. Inter-rate reliability

45 de-identified LC videos were sent to 3 expert surgeons for evaluation. Each video is analyzed and rated in terms of smoothness, efficiency and precision. To find the inter-rater reliability, Fleiss’es Kappa coefficient between experts’ score is calculated. Table 17 shows fair to moderate agreement between the rates with significant p-value. Finally an average score of each rater for every video is calculated and plotted in Figure 58 [116].

**Table 17: Fleiss’es Kappa coefficient between raters**

<table>
<thead>
<tr>
<th>Features</th>
<th>Fleiss’es kappa</th>
<th>p-value</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoothness</td>
<td>0.4349</td>
<td>&lt;0.0001</td>
<td>Moderate agreement</td>
</tr>
<tr>
<td>Efficiency</td>
<td>0.3581</td>
<td>0.0001</td>
<td>Moderate agreement</td>
</tr>
<tr>
<td>------------</td>
<td>--------</td>
<td>--------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Precision</td>
<td>0.4393</td>
<td>&lt;0.0001</td>
<td>Moderate agreement</td>
</tr>
</tbody>
</table>

![Graph showing inter-rater reliability of Global scores between rater 1, 2 & 3](image)

**Figure 58**: Inter-rater reliability of Global scores between rater 1, 2 & 3

### 5.2.2. Correlation with expert’s rating

Training dataset was used to find the regression parameters. Once all the parameters were determined, the test data set was run to find the correlation coefficient between raters’ average score and predicted score. All three equations showed more than 85% correlation and significant p-value (Table 18).

**Table 18**: Correlation with experts’ score

<table>
<thead>
<tr>
<th>Features</th>
<th>Correlation coefficient</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metric</td>
<td>Equation</td>
<td>R²</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Smoothness</td>
<td>$1.07 + 0.76 \times (\textit{smooth movement}) - 0.26 \times (\textit{depth perception})$</td>
<td>0.9291</td>
</tr>
<tr>
<td>Efficiency</td>
<td>$1.66 + 0.64 \times (\textit{bimanual dexterity}) - 0.41 \times (\textit{tool change frequency})$</td>
<td>0.9097</td>
</tr>
<tr>
<td>Precision</td>
<td>$-0.43 + 4.16 \times (\textit{presence of blood}) - 0.99 \times (\textit{snaps})$</td>
<td>0.8576</td>
</tr>
</tbody>
</table>
Chapter 6

DISCUSSION

Several video and sensor-based systems have been developed to capture a user's motion and other important features which can be later analyzed objectively and quantitatively to correlate with the skill level [11]. However, most of the quantitatively skill assessment systems available are sensor-based i.e. sensors are integrated in surgical tool or surgeon's gloves to track different movement features. Although the sensory systems capture motion features quite well, there is not enough work done in combining tool and hand movement data together to assess the surgical proficiency. Other researchers compelled surgeons to wear sensors to monitor the demographic features or body's center of pressure value and observed the correlation with the skill execution. However, these data are not comprehensive enough to successfully assess the skill level. Moreover, skill assessment is done only in simulated training environment. Since, integration of wearable sensors interferes with surgical skill execution [11], none of these works are done towards assessing actual surgical procedure. Moreover the sensors have to be sterilized to be used in real surgery can increase the cost of the entire surgical procedure [25].

Drawbacks of sensor integration bring researchers’ attention towards video based system. Some video-based system use only external cameras to capture movement of hand while performing surgical exercise. It requires camera to have line-of-sight of the hands and subject to inconsistent lighting, background noise and low-resolution video images [11]. Moreover, due to the use of long instruments in MIS, it is almost impossible to derive tool movement from hand movement video. Other video-based approaches track surgical tool-tip; however, the analysis is solely dependent on the travel-length value of the tool-tip. Moreover, no work has been done that combines features from both
hand and surgical tool movement. This is critical in surgery wherein surgeons hand and tool movements are an important component of ensuring patient safety.

Some video-based approaches are done towards accessing surgical videos. Few works have been done in this area. However tracking surgical tools or anatomical structure can be very challenging than that of training environment. In box simulators, surgical tool tip is usually tracked using a colorful marker. Series of computer vision algorithms are used towards markerless surgical tool-tip tracking. Some researchers tracked anatomical structures such as cystic artery and gallbladder by detecting shape, color or movement due to pulse. However none of the studies provide any assessment score other than detecting proximity of the instruments to the unsafe region of anatomical structure. Moreover, all these work are performed on simulated surgical environment since free from camera movement or lighting variations.

Hence, there is a need of complete medical augmented reality performance evaluation system with automatic tracking system which can be employed in assessing surgery in real-life scenarios. In the present study, we aimed to develop a video-based assessment tool for laparoscopic surgery training to assess the generic and specific technical aspects of surgical features. 52 FLS training exercise videos performed by medical students and different postgraduate year residents were analyzed. Several functions were developed that combined a series of computer vision algorithms to accurately track surgical tool-tip, surgeon’s hand and objects in the surgical scene. Developed motion detection function detected the position and direction of surgical tool and surgeons hand movement and used these information to extract a number of psychomotor skill assessing features like jerkiness, snaps, tremor, movement efficiency, perception of depth and bimanual dexterity. For each of the features, ANOVA analysis was done which showed statistically significant difference in variance between the consecutive expertise
groups i.e. novices-vs-intermediates and intermediates–vs-experts. Also when classified into 3 groups for each of these features (using Random Forest classifier), it showed on average 74% correctly classified data according to the expertise level. Each of these videos were de-identified and sent to three exert raters to be rated in three categories: smoothness, efficiency and precision. Scores with only higher inter-rater reliability were considered to design the scoring algorithms. Sets of three of the extracted features were found associated with each of the rated features by the experts. For example smoothness score is associated with the jerkiness, snaps and tremor; efficiency score is associated with movement efficiency, depth perception and bimanual dexterity; and precision is associated with the number of error committed and the time taken to complete the exercise. Multivariate linear regression model was built for each of the categories and regression parameters are found using the tool ‘Weka’. The entire dataset was split in 60% training, 10% cross-validation and 30% testing to find the parameters and test the regression models. Analysis result showed very high correlation (91% average) with experts’ ratings with significant p-value. The system automatically detected the number of errors committed during exercise. Errors are very task specific and used to be detected by continuous observations of the expert observers. However, the system successfully detected 87% of the errors automatically. Thus, it proves the first hypothesis that the skill assessment provided by the proposed automatic quantitative scoring system is equivalent to that of the gold standard.

Although our current findings contrasted an expert group with a novice group; expertise is best viewed as a continuum with a number of levels that result in unique performance characteristics. The development of expertise is marked by specific transitions corresponding to reorganizations of knowledge and non-monotonic (not linear) increases in the learning curve [117]. To observe the effect of the feedback system on
process of development of expertise, we conducted an experiment with 54 medical students where all of them were asked to perform 2 FLS box exercises 16 times each. None of the subjects had any prior experience with the task. 32 of the novices were used as control group where no performance feedback was provided during learning period. They simply followed the guidelines for the tasks. However with the other 22 novices, on-screen real-time feedback and performance score were provided subsequent to each exercise.

The performance curve for the control group of novices who were not provided any feedback during the course of 16 exercise session, showed “intermediate effect” in their learning curve. Intermediate effect is explained by Vimla et al. [118, 119] as “a temporary decline in performance as knowledge is acquired and organized, when a linear increase in performance with the length of training or time on task would be expected”. On the other hand, the performance curve of the experimental group exhibited more steady performance throughout the learning session. Although the rate of learning became saturated after certain number of trials; the learning curve did not show the intermediate effect for the experimental group. Paired t-test showed significant improvement in performance for the experimental group over the control group. Average group score showed 11% better performance for the experimental group over the control group, hence proving the second hypothesis that the immediate feedback system would increase training effectiveness by reducing the time it takes to attain a desired level of proficiency.

The developed algorithms were also used to assess real-life laparoscopic cholecystectomies (LC). To evaluate the performance of the system, 45 LC videos performed by postgraduate residents were analyzed. Although in real-life surgeries, due to the complex and dynamic scenarios, performance depends on knowledge-based solution [120], we carefully chose 45 cases with similar difficulties. Several functions
were developed that combined a series of computer vision algorithms to accurately track surgical tool-tip, surgeon’s hand and gallbladder in an LC video. Motion detection function that was developed for box simulators, extracted the similar psychomotor skill assessing features from the surgical tool movements. Features like perception of depth and frequency of tool switch were extracted from the captured hand movement videos. The system also detected presence of blood in the surgery scene that resulted from faulty dissection of cystic artery by detecting the color of the gallbladder. Each of these de-identified videos was rated by three experts in terms of smoothness, efficiency and precision. Scores with only higher inter-rater reliability were considered to design the scoring algorithms. Smoothness scores were found associated with the extracted smoothness and depth perception; efficiency score is associated with bimanual dexterity and tool switch frequency; and precision is associated with presence of blood in the scene and number of unsafe movements. Multivariate linear regression model was built for each of the categories and regression parameters were found using Weka. The entire dataset was split in 60% training, 10% cross-validation and 30% testing to find the parameters and test the regression models. Analysis result showed very high correlation (90% average) with experts’ ratings with significant p-value which supports the claim of the automatic skill assessment system can be successfully use to assess real-life surgeries.
Chapter 7

CHALLENGES

7.1. Data collection

To establish any algorithm, a lot of data is required to assess its performance. Collection of data in clinical settings requires additional clearance thus making it a more challenging undertaking. Some major data collection related challenges are discussed below:

- **IRB approval:** To collect data from human subject, the first step is to process the IRB approval. The process may take a few months to complete. Furthermore, any modification in the data collection procedure requires IRB to be modified as well, thus making the data collection session lengthier.

- **Scheduling:** For this research, all the subjects were medical students, surgical residents or attending surgeons. It is very difficult to schedule time for data collection with their very busy schedule.

- **Absence of subject:** Residents do not usually want to miss opportunities to observe the trauma cases or other patient related emergencies. When any of these cases arise, they usually skip the data collection session to observe the cases.

- ** Interruption during data collection:** Since the data collection is placed in the simulation center inside the hospital, in case of patient emergencies, residents and attending physicians are often called to the scene. When a pager rings at the time of the data collection or the resident/attending physicians name is announced, usually they have to make the call or leave the session.

- **Camera placement in the OR:** In laparoscopic cholecystectomy, the video of tool movement are recorded for further analysis. However to capture hand movement, an
HD camera is placed on the IV (intravenous) stand for accurate line-of-sight view [26], which requires the presence of the researcher during or before/after the procedure. To enter to the OR, researcher has to follow stringent sanitation procedures.

7.2. Data analysis

Analyzing video data from the box simulator is not very difficult, since they are collected in a controlled and simulated environment. However, analyzing videos of actual laparoscopic cholecystectomies was challenging due to the uncontrolled nature of the data. The challenges are -

- Camera movement: In the FLS box, the camera that captures the tool view is always stationary. But in actual laparoscopic cholecystectomy, the resident often moves the camera to acquire a better viewing angle of the surgical scene. Since the analysis largely depends on motion tracking, in case of camera movement, everything in the scene experiences relative motion. In the system, camera movement was detected as global movement and motion parameters for those particular frames were disregarded. Hence we lost the tool movement data for the frames in which the camera was moved.

- Variation of light: A fiber optic cable connected to a light source is used to light the surgical scene. Unlike the lighting in the FLS box, operators can always move the light for a better view of the scene. Variation of lighting creates bright reflections which makes it very difficult to process object segmentation. The frames where the objects are undetected due to variation in light are discarded.

7.3. Implementing the cloud based system

- Requires the user to capture videos: User can assess any surgical task by uploading the video to the system which will assess the video and provide the score
immediately. In laparoscopic cholecystectomy, the tool view is usually recorded for further analysis. In case of the FLS training box, the tool view is directly fed to the monitor/display and usually not recorded. To record the tool view video, the user has to connect an ADC converter and a computer. However for both cholecystectomy and the FLS training box, capturing hand movement requires the presence of the researcher at the scene.

- Requires high-speed internet: High definition recording of surgical exercise or actual surgery requires lots of memory in a storage device. Specifically, a cholecystectomy procedure can be hour long and the size of the file can be a few hundred megabytes. Uploading a video file of a few hundreds megabytes might be challenging for users with a slower speed of internet access.
Chapter 8

FUTURE WORK

8.1. Implementation in other types of simulations

In the pilot study we analyzed video data of intubation (Figure 59), extracorporeal suture and knot tying. Some of these findings were published in different international conferences [100, 101]. More research can be done in this area and real-time skill assessment system can be developed in an array of different clinical exercises.

Figure 59: Intubation exercise

8.2. Implementation in other types of surgeries

Since the developed system can successfully evaluate performance of laparoscopic cholecystectomy, it can be further developed to analyze other common laparoscopic procedures like salpingo-oophorectomies, partial nephrectomies, etc. Although minimally invasive surgery is gaining more and more popularity, still a large number of open surgeries are performed. By using similar computer vision approach, open surgeries can be evaluated as well.

8.3. Use of 3D camera

Surgical field is a 3-dimensional space. However when this scene is captured with video camera and displayed on a 2-dimensional display, it loses information in one axis. Recent advances in 3D depth cameras such as Microsoft Kinect sensor lets the computer directly sense the third dimension (depth) of the motion and the environment. By using
3D camera, another dimension of the movement information can be captured which will facilitate more accurate feature extraction.

8.4. Complete cloud-based solution
The developed cloud-based system allows users to upload surgical videos and get instant performance scores. However it requires user to capture video data of the exercise. A cloud-based video capturing tool can be integrated with the system that will upload video to the cloud while exercise is in progress. The system will provide real-time on-screen feedback as well as the subsequent assessment score. The expert raters will rate these videos on the cloud. Once a video is rated by at least 3 expert raters, the inter-rater reliability will be calculated. Scores with high inter-rater reliability will be used as a training data and added with the exiting training-data set to train back the algorithm. Hence an automatic black propagation neural network can be integrated with the entire system (Figure 60). More the number of videos will be rated, more robust the assessment tool would be.
Figure 60: Train back the algorithm
Chapter 9

CONCLUSION

This research proposes a video-based surgical skill assessment technique. It uses a computer vision algorithm to analyze the video of surgeon’s hand and surgical tool movement and extract features like surgical tool-movement smoothness, movement efficiency, individual gesture proficiency, task specific errors, etc. Finally by assigning weight to each of the parameters, a composite scoring system is developed for the assessment of surgical skill. Data from different PGY level of surgical residents and expert surgeons is being collected to train and test the algorithm. Since the research analyzes video data of the surgery rather than any wearable sensors, it is cost effective and also overcomes the drawbacks of most of surgical skill assessment techniques presently available.

The proposed video-based surgical skill assessment technique can provide real-time on-screen feedback, hence it is also being tested as a tool for the efficient skill learning technique. After analyzing data from the experiment with on-screen feedback system, the result shows steeper learning curve than the system without the feedback. However more data is being collected for analysis to strengthen the hypothesis.

Therefore, this dissertation promises a cost-effective and robust surgical skill assessment technique which also promotes toward efficient surgical skill learning. However, the key for successful execution of skill assessment is to deploy the algorithms in a real surgical setting. Until now almost no work has been done to transfer this work into an actual surgical setting. Apart from induced interference in surgical execution, the sensor-based systems needs to be completely sterile to be used in actual surgery. Also the extensive use of sensors makes the entire surgery and the training procedure more expensive. Since
the video-based approach is free from the drawbacks of the other systems, this technique can be potentially useful to assess a surgeon’s performance in an actual surgical operation. This research proposes to perform some testing of the system in actual surgery and include results towards the dissertation. However, further analysis and research need to be done in this area.

Objective evaluation remains the holy grail of this line of experimentation and to that effect the ultimate achievement would be acceptance of this tool by boards of surgeries of this as a validated tool. While there are several challenges that need to be addressed to get this validated like large scale multisite trials and repeatability, the work presented here lays the foundation for such experimentation. A huge opportunity lies in addressing the need for objective evaluation of both cognitive skills and psychomotor skills in tandem. We also need to understand how surgical errors evolve and if such a system can help predict an error before it occurs. For example increased snap values could predict an impending mistake. While these remain important questions, the work here takes the first step at developing a comprehensive affordable and scalable approach to surgical proficiency determination. The developed informatics tool can not only present a practical solution but also encourage further line of research and investigation.
REFERENCES


85. Solis, J., et al. *Development of a sensor system towards the acquisition of quantitative information of the training progress of surgical skills.* in


105. Islam, G., K. Kahol, and e. al, Understanding specialized human motion through gesture analysis: Comparisons of expert and novice anesthesiologists performing direct laryngoscopy using motion capture technology in International Anesthesia Research Society Meeting2011: Vancouver, Canada.


