Hidden Markov Models of Minimally Invasive Surgery

Jacob Rosen[+], Ph.D., Christina Richards[*], M.D.,
Blake Hannaford[+,++], Ph.D., Mika Sinanan[*,**], M.D., Ph.D.

[+] Department of Electrical Engineering, Box 352500
[*] Department of Surgery, Box 356410
University of Washington
Seattle, WA, 98195, USA

E-mail / URL: [++] blake@ee.washington.edu, http://rcs.ee.washington.edu/brl/
[**] mssurg@u.washington.edu, http://depts.washington.edu/cves/

Abstract

A crucial process in surgical education is to evaluate the level of surgical skills. For laparoscopic surgery, skill evaluation is traditionally preformed subjectively by experts grading a video of a procedure performed by a student. By its nature, this process is preformed using fuzzy criteria. The objective of the current study was to develop and assess a skill scale using Discrete Hidden Markov Models (DHMM). Ten surgeons (5 Novice Surgeons - NS; 5 Expert Surgeons - ES) performed a cholecystectomy and Nissen fundoplication in a porcine model. An instrumented laparoscopic grasper equipped with a three-axis force/torque sensor was used to measure the forces/torques at the hand/tool interface synchronized with a video of the tool operative maneuvers. A synthesis of frame-by-frame video analysis and a vector quantization algorithm, defined force/torque signatures for 14 types of tool/tissue interactions. From each step of the surgical procedures, two DHMM were developed representing the performance of 3 surgeons randomly selected from the 5 in the ES and NS groups. The data obtained by the remaining 2 surgeons in each group were used for evaluating the performance scale. The final result was a surgical performance index which represented a ratio of statistical similarity between the examined surgeon's DHMM and the DHMM of NS and ES. The difference between the performance index value, for a surgeon under study, and the NS/ES boundary, was considered to indicate the level of expertise in the surgeon's own group. Using this index, 87.5% of the surgical procedures were correctly classified into the NS and ES groups. The 12.5% of the procedures that were misclassified were preformed by the ES and classified as NS. However, in these cases the performance index values were very close to the NS/ES boundary. Preliminary data suggest that a performance index based on DHMM and force/torque signatures provides an objective means of distinguishing NS from ES. In addition this methodology can be further applied to evaluate haptic virtual reality surgical simulators for improving realism in surgical education.

1. Introduction

The technical performance of surgery in general, and in minimally invasive surgeries (MIS) in particular, requires the application of forces and torques (F/T) on and between tissues to achieve specific goals. Parameters that determine the magnitude of the forces and torques used include the nature and the goal of the tissue being manipulated, the type of the instrument being used, and the skill of the surgeons. One of the more difficult tasks in MIS education is to teach the optimal application of instrument F/T necessary to conduct an operation. The use of virtual reality models for teaching complex surgical skills while

simulating realistic human/tool and tool/tissue interaction has been a long-term goal of numerous investigators [1,2,3,4,5]. Moreover, the F/T information is crucial in designing force-feedback telerobotic systems. The goal of this study is to create new quantitative knowledge of the F/T applied by surgeons on their instruments during MIS. This goal is pursued through several steps: (*i*) developing instrumented endoscopic tools which contain embedded sensors capable of recording F/T information (*ii*) creating a database of F/T signals acquired during actual operating conditions on experimental animals, (*iii*) developing statistical models to evaluate an objective skill level.

2. Methods

2.1 Experimental System and Clinical Protocols

Two types of information were acquired while performing MIS on pigs: (*i*) F/T data measured at the human/tool interface and (*ii*) visual information of the tool tip interacting with the tissues [6]. A standard, reusable 10 mm endoscopic grasper tool with interchangeable tool tips (Storz) was modified by dividing the outer shaft at its proximal end and interposing a three orthogonal axes F/T sensor (ATI) - ($F_x, F_y, F_z, T_x, T_y, T_z$). In addition, a force sensor was added to the instrument handle to measure grasping/spreading forces applied by the surgeon (F_g). The seven channels of F/T data were sampled at 30 Hz. In addition, a LabView (National Instruments) application was developed incorporating a Graphical User Interface (GUI) for acquiring and visualizing the F/T data in real-time. The visual view from the endoscopic camera monitoring the movement of the grasper while interacting with the internal organs/tissues was integrated with the F/T at the human/tool interface using a video mixer in a picture-in-picture mode and synchronized with time. The integrated interface was recorded during the surgical procedures for off-line analysis.

Procedure	Step	Description	Tool Type	Hand	Video F/T
Laparoscopic Cholecystectomy	LC-1	Positioning Gall Bladder	Atraumatic Grasper	L	+
	LC-2	Exposure of Cyctic Duct	Curved Dissector	R	+
	LC-2*	Divide of Cyctic Duct	Scissors	R	-
	LC-3	Dissection of Gall Bladder Fossae	Curved Dissector	R	+
	LC-4	Exposure of Cystic Artery	Curved Dissector	R	-
	LC-4*	Dividing Artery	Scissors	R	-
Laparoscopic Nissen Fundoplication	LNF-1	Dissect Right Crus	Surgiwand	R	-
	LNF-2	Dissect Left Crus	Surgiwand	R	-
	LNF-3	Dissect Esophagus / Blunt	Curved Dissector	R	+
	LNF-4	Placing a Wrap Around the Esophagus	Babcock Grasper	R	+
	LNF-5	Suture Wrap/ Intracorporeal Knot Tying With Needle Holder	Curved Dissector	R	+
	LNF-6	Coronal Sutures/ Intracorporeal Knot Tying Endostitch	Endostitch	R	-

Table 1: Definitions of surgical procedure steps and tool tip types. (Shaded steps performed, but no F/T were recorded.)

Ten surgeons (five novice surgeons –NS; and five experienced surgeons - ES) performed laparoscopic cholecystectomy and laparoscopic Nissen fundoplication in a porcine model (pig). Protocols for anesthetic management, euthanasia, and survival procedures were reviewed and approved by the Animal Care Committee at the University of Washington and the Animal Use Review Division of the U.S. Army Veterinary Corps.

Each surgical procedure was divided into steps (Table 1). Although all the steps were performed in each procedure, data were recorded only when the grasper was used with the following tool tips: atraumatic grasper, curved dissector, Babcock grasper.

2.2 Data Analysis

Three types of analysis were performed on the raw data: (*i*) Video Analysis (VA), encoding the tool-tip/tissue interaction into states; (*ii*) Vector Quantization (VQ), identifying cluster centers (signatures) in the F/T data; and (*iii*) Hidden Markov Modeling (HMM), comparing the performances of ES and NS.

A frame-by-frame VA, preformed by two expert surgeons first on a preliminary data set and later on the entire database, was used to identify 14 different tool maneuvers in which the endoscopic tool was interacting with the tissue while performing MIS (Table 2). Each surgical maneuver had a unique F/T pattern. For example, in the laparoscopic cholecystectomy, isolation of the cystic duct and artery involves performing repeated pushing and spreading maneuvers which in turn requires pushing forces mainly along the Z axis (F_z) and spreading forces (F_g) on the handle.

Type	State Name	State Acronym	Force / Torque						
			Fx	Fy	Fz	Tx	Ty	Tz	Fg
I	Idle	ID	*	*	*	*	*	*	*
	Grasping	GR							+
	Spreading	SP							-
	Pushing	PS			-				
	Sweeping	SW	+/-	+/-		+/-	+/-		
II	Grasping - Pulling	GR-PL			+				+
	Grasping - Pushing	GR-PS			-				+
	Grasping - Sweeping	GR-SW	+/-	+/-		+/-	+/-		+
	Pushing - Spreading	PS-SP			-				-
	Pushing - Sweeping	PS-SW	+/-	+/-	-	+/-	+/-		
	Sweeping - Spreading	SW-SP	+/-	+/-		+/-	+/-		-
III	Grasping - Pulling - Sweeping	GR-PL-SW	+/-	+/-	+	+/-	+/-		+
	Grasping -Pushing - Sweeping	GR-PS-SW	+/-	+/-	-	+/-	+/-		+
	Pushing - Sweeping - Spreading	PS-SW-SP	+/-	+/-	-	+/-	+/-		-

Table 2: Definition of tool/tissue interactions and the corresponding directions of forces and torques applied in cholecystectomy and Nissen fundoplication during MIS.

The 14 tool/tissue interactions can be divided into three types based on the number of movements performed simultaneously. The fundamental maneuvers were defined as type I. During the idle state the tool was moving in space without touching any internal organ, and the forces and torques developed in this state represent mainly the interaction with the trocar and the abdominal wall, in addition to the gravitational and inertial forces. In the grasping and spreading states, compression and tension were being applied on the tissue by closing/opening the grasper's handle. In the pushing state compression was applied on the tissue by moving the tool along the Z axis. In sweeping, the tool was placed in one position

while rotating it around the X and Y axes (trocar frame). The rest of the states in groups II and III were combinations of the fundamental states of group I.

In order to define typical forces and torques associated with these 14 tool/tissue interactions, all the surgeons (NS and ES) performed each of the tool/tissue manipulation separately while interacting with different tissues in the abdominal cavity. This subsection of the database was used by the VQ algorithm to encode the multi-dimensional (7D) F/T data ($F_x, F_y, F_z, T_x, T_y, T_z, F_g$) into discrete cluster centers (signatures). The VQ algorithm (K-mean) is based on minimization of the sum of squared distances from all points in the cluster domain ($\overline{X}$) to the cluster center ($\overline{Z}_j$), *i.e.*,

$$min \sum_{X \in S_j(k)} (\overline{X} - \overline{Z}_j) \qquad (1)$$

where $S_j(k)$ is the cluster domain for the cluster center ($\overline{Z}_j$) at the *k*'th iteration.

Based on the defined cluster centers the entire database, including all surgical procedures (Table 1), was encoded into a discrete 1D vector of symbols using the criteria defined by Eq. (1).

The data reduction achieved by using the VQ analysis was the first phase towards a discrete HMM. Further statistical analysis of F/T clusters distribution between NS and ES showed that the F/T magnitude was one of the major differences between the two groups. Moreover, a previous study [6], showed that both the tool/tissue states, the state transitions, and the time spent in each state were different between the two groups and between the surgical procedures. Based on this finding, a three-state HMM architecture (Fig. 1) was used as a generic form applicable to all the surgical procedures. The F/T cluster centers in each one of the tool/tissue interactions (Table 2) were divided into three levels according to their F/T magnitudes (Low (L), Medium (M), and High (H)). All of the tool/tissue interactions at each level were then lumped into three states (L, M, and H) of the HMM.

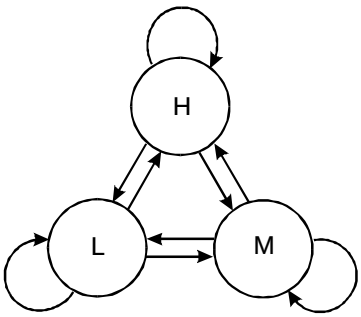

Fig. 1:Three-state, fully connected HMM based on the F/T clustering magnitudes.

Using the given HMM architecture and the encoded F/T signals, two HMMs were trained for each surgical procedure, representing the performances of three subjects out of five in each group (ES and NS). These were referred as the NS-HMM (l_{NS}), and the ES-HMM (l_{ES}). The HMMs developed based on the data of the two other subjects in each group (l_i) and their observation vectors (O_i) were used to evaluate the skill scale based on the statistical similarity to the main groups. Two statistical similarity factors were defined:

$$NSF = log(\mathbf{P}(O_i \mid l_{NS}))/log(\mathbf{P}(O_i \mid l_i)); \quad ESF = log(\mathbf{P}(O_i \mid l_{ES}))/log(\mathbf{P}(O_i \mid l_i)) \quad (2)$$

The *NSF* defined what was the statistical similarity between the performances of the subject under study and the NS group, whereas the *ESF* indicated the statistical similarity relative to the ES group. Using these factors, an absolute NS would be represented by *NSF*=1 and *ESF*=0, and vice versa for an absolute ES.

3. Results

The 5D cluster centers representing the F/T for each type of tool/tissue interactions (Table 2) were plotted in Fig. 2. Comparing the entire cluster centers revealed that there is at least one dimension in which each cluster center was different from every other one. Using these cluster centers for encoding the F/T database and analyzing the F/T cluster center distributions of NS and ES with a non-parametric method showed statistically significant differences ($p<0.05$) between these groups (Fig 3). Higher F/T magnitudes were applied by the NS compared to the ES in steps that involved tissue manipulations (LC-1, LNF-4) and vice versa in steps in which tissue dissection and suturing were preformed (LC-2, LC-3, LNF-3).

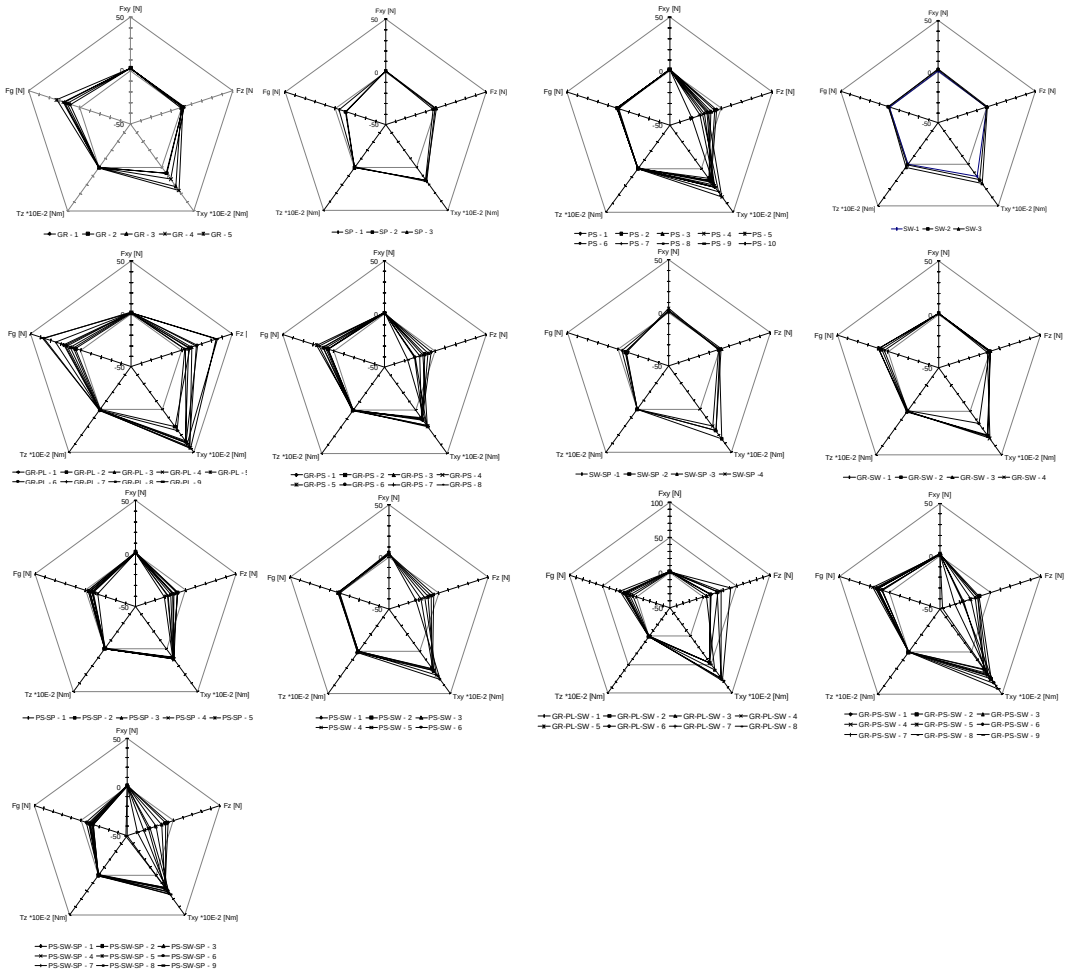

Fig 2: Codebook representing the F/T at the tool/surgeon-hand interface for different tool/tissue interaction. (Each graph shows a different interaction type. The idle state is not shown. Each vertex represents the magnitude of the components of the 5D F/T vector, and each line represents a different cluster center.)

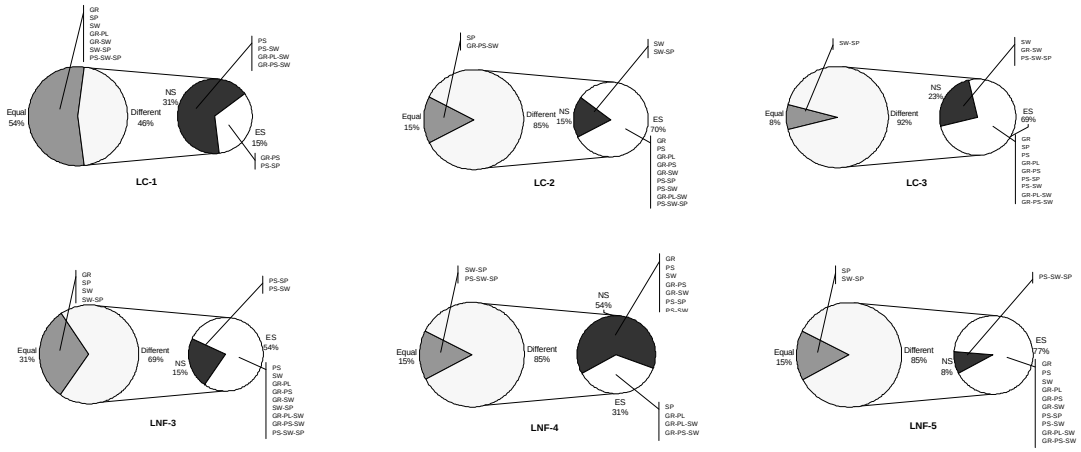

Fig 3: Comparison of high-level F/T magnitudes and the corresponding tool/tissue interactions in which they occur while performing surgical procedures by NS and ES.

The *NSF* and the *ESF* (Eq. 2) for two NS and two ES performing steps of LC and LNF procedures were plotted as a scatterplot in Fig 4. The solid line represents the boundary between the NS (top-left) and the ES (right-bottom); along this line *NSF=ESF*. Except for three cases the HMM was capable of classifying NS and ES correctly into the two groups. Dotted lines represent iso-performances in which all the points on each line have the same squared distance ratio between the points: (0,1) –absolute NS, and (1,0) – absolute ES (Eq. 3).

$$C^2 = \{ d((NSF_i, ESF_i),(0,1)) / d((NSF_i, ESF_i),(1,0)) \}^2 \qquad (3)$$

The performance parameter C^2 mapped the two-dimensional performance domain (Fig 4a) into a one-dimensional performance scale.

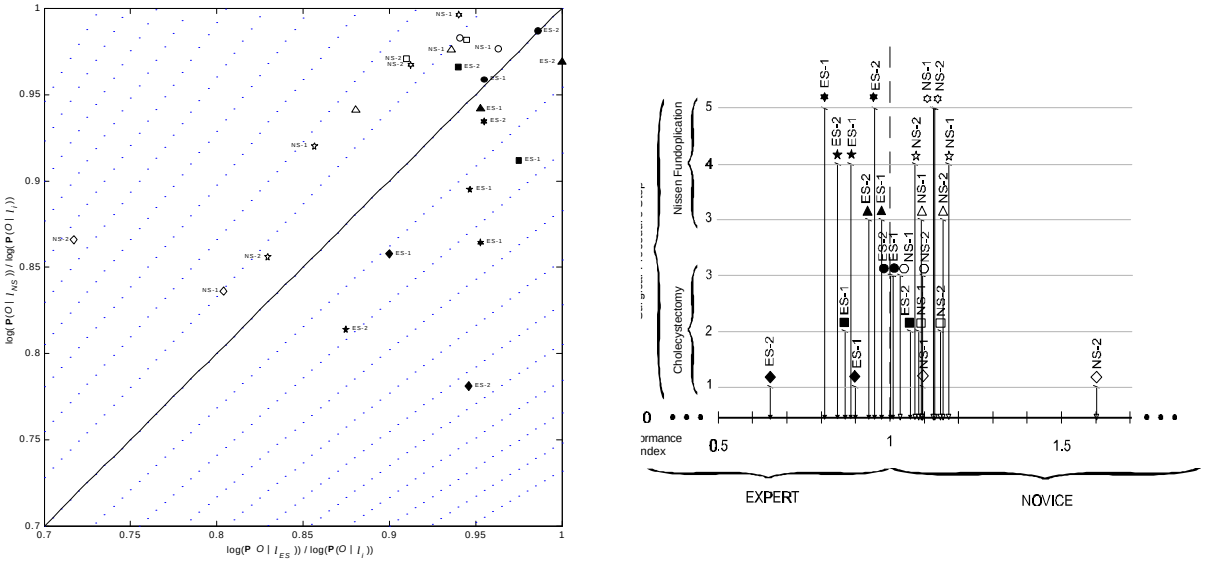

Fig 4: Scales for evaluating a surgeon's performance in MIS: (*a*) Statistical distances map generated by the HMM, and (*b*) The C^2 performance scale.

4. Conclusions

Minimally invasive surgery is a complex task that requires a synthesis between visual and haptic information. Analyzing MIS in terms of these two sources of information is a key step towards developing objective criteria for training surgeons and evaluating the performance of a master/slave robotic system for teleoperation or a haptic device for virtual reality simulations. The magnitude of F/T applied by NS and ES varied based on the task being preformed. High F/T magnitudes were applied by NS compared to ES while performing tissue manipulation. This might be a result of insufficient dexterity of the NS that might a potential for tissue damage. However, low F/T magnitudes were applied by the NS compared to the ES during tissue dissection, which might also indicate excessive caution to avoid irreversible tissue damage. By doing that more repetition of the dissection movements were required to be preformed by the NS in order to tear the tissue, a process which substantially decrease the efficiency of the MIS procedure. Using the F/T information in real-time during the course of learning as a feedback information to the NS may improve the learning curve, reducing soft tissue injury and increase the efficiency during endoscopic surgery.

The force/torque signatures and the HMMs are objective criteria for evaluating skills and performance in MIS. The results suggest that HMMs of surgical procedures allow objective quantification of skill. Using these techniques, 87.5% of the surgical procedures were correctly classified into the NS and ES groups. The 12.5% of the procedures that were misclassified were preformed by the ES and classified as NS. However, in these cases the performance index values were very close to the NS/ES boundary.

The approach outlined in this study could be extended by increasing the size of the database to include more surgical procedures preformed by more surgeons. This information, combined with other feedback data, may be used as a basis to develop teaching techniques for optimizing tool usage in MIS. The novice surgeons could practice these skills outside of the operating room on animal models or by using realistic virtual reality simulators until they had achieved the desired level of competence, and compare themselves to norms established by experienced surgeons.

Acknowledgments

This work was supported by the U.S. Army Medical Research and Material Command under DARPA grant- DMAD17-97-1-725.

References

[1] Satava R., Virtual Reality Surgical Simulator. Surg Endosc (1993) 7: 203-205

[2] Berkely J. S. Weghorst, H. Gladstone, G. Raugi, M. Ganter, Fast Finite Element Modeling for Surgical Simulation, Proceedings, MMVR-99, IOS Press, San Francisco, January 1999

[3] Bro-Nielsen M, J. Tasto, PREOP Endoscopic Simulator: A PC Based Immersive Training System for Bronchoscopy, Proceedings, MMVR-99, IOS, San Francisco, January 1999.

[4] Oppenheimer P., S. Weghorst, M. MacFarlane, M. Sinanan, Immersive Surgical Robotic Interfaces, Proceedings, MMVR-99, IOS, San Francisco, January 1999.

[5] Szekely G. et. al., Virtual Reality Based Surgery Simulation for endoscopic Gynaecology, Proceedings, MMVR-99, IOS, San Francisco, January 1999.

[6] J. Rosen , M. MacFarlane, C. Richards, B. Hannaford, C. Pellegrini, M. Sinanan, 'Surgeon/Endoscopic Tool Force-Torque Signatures In The Evaluation of Surgical Skills During Minimally Invasive Surgery,' Proceedings, MMVR-99, IOS, San Francisco, January 1999.