Objective Laparoscopic Skills Assessments of Surgical Residents Using Hidden Markov Models Based on Haptic Information and Tool/Tissue Interactions

Jacob Rosen[+,*], Massimiliano Solazzo[++], Blake Hannaford[+,**], Mika Sinanan[++,***]

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

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

ABSTRACT

Laparoscopic surgical skills evaluation of surgery residents is usually a subjective process, carried out in the operating room by senior surgeons. By its nature, this process is performed using fuzzy criteria. The objective of the current study was to develop and assess an objective laparoscopic surgical skill scale using Hidden Markov Models (HMM) based on haptic information, tool/tissue interactions and visual task decomposition. *Methods:* Eight subjects (six surgical trainees: first year surgical residents 2xR1, third year surgical residents 2xR3 fifth year surgical residents 2xR5; and two expert laparoscopic surgeons: 2xES) performed laparoscopic cholecystectomy following a specific 7 steps protocol on a pig. An instrumented laparoscopic grasper equipped with a three-axis force/torque sensor located at the proximal end with an additional force sensor located on the handle, was used to measure the forces and torques. The hand/tool interface force/torque data was synchronized with a video of the tool operative maneuvers. A synthesis of frame-by-frame video analysis was used to define 14 different types of tool/tissue interactions, each one associated with unique force/torque (F/T) signatures. HMMs were developed for each subject representing the surgical skills by defining the various tool/tissue interactions as states and the associated F/T signatures as observations. The statistical distance between the HMMs representing residents at different levels of their training and the HMMs of expert surgeons were calculated in order to generate a learning curve of selected steps during laparoscopic cholecystectomy. *Results:* Comparison of HMM's between groups showed significant differences between all skill levels, supporting the objective definition of a learning curve. The major differences between skill levels were: (i) magnitudes of F/T applied (ii) types of tool/tissue interactions used and the transition between them and (iii) time intervals spent in each tool/tissue interaction and the overall completion time. The objective HMM analysis showed that the greatest difference in performance was between R1 and R3 groups and then decreased as the level of expertise increased, suggesting that significant laparoscopic surgical capability develops between the first and the third years of their residency training. The power of the methodology using HMM for objective surgical skill assessment arises from the fact that it compiles enormous amount of data regarding different aspects of surgical skill into a very compact model that can be translated into a single number representing the distance from expert performance. Moreover, the methodology is not limited to *in-vivo* condition as demonstrated in the current study. It can be extended to other modalities such as measuring performance in surgical simulators and robotic systems.

1. Introduction

One of the paramount issues in surgical education is the evaluation of surgical skill. An accurate means of assessing surgical skill would allow surgical educators to evaluate the effectiveness of skills training, monitoring progress and learning curves of students and residents along the course of their study. Skill evaluation in surgery in general, and laparoscopic surgery in particular, is currently a ***subjective*** process, carried out in the operating room or performed off-line using a video tape by expert surgeons grading the performance of the student. By its nature, this process is performed using fuzzy criteria.

Surgical skills are accessible for analysis in three different environments : (1) open or minimally invasive surgery (MIS) utilizing traditional surgical tools and equipment, (2) a robotic system using a master/slave setup, and (3) a simulator utilizing a haptic device that generates force feedback in addition to a virtual reality graphic representation of the surgical scene. All of these systems have a human-machine interface. Through this interface, visual, kinematic, and dynamic information is flowing back and forth between the surgeon and the environment. The aim of the current research was to develop methodology to acquire and analyze information at the human/tool interface in order to quantitatively and objectively evaluate surgical skill and learning curves of MIS. The power of the proposed methodology is that it can be incorporated into any of the three environments.

The methodology developed in the current study was based on the Hidden Markov Modeling (HMM). HMMs were extensively developed in the area of speech recognition (for mathematical review see [1]). Based on the theory developed for speech recognition HMMs have become useful statistical tools in the fields of robotics, teleoperation [2, 3, 4], human manipulation actions, manufacturing, gesture recognition. They are also being applied to the recognition of facial expressions from video images, DNA and protein modeling, nuclear power plants, and detection of pulsar signals. These applications suggest that the HMMs have high potential to provide better models of the human operator in complex interactive tasks with machines.

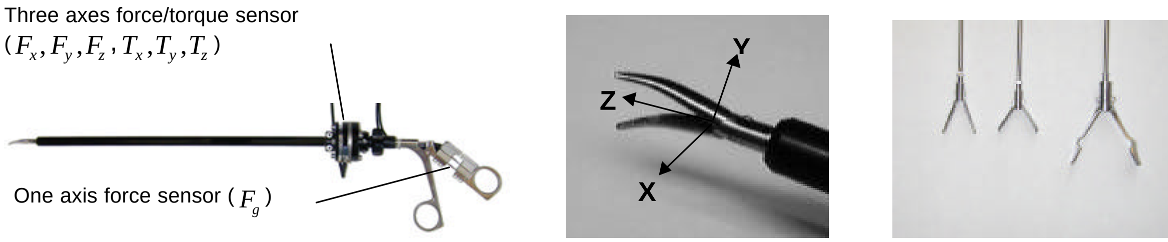
2. Materials and Methods

2.1 Subjects and Protocol

Eight subjects (six general surgery residents: first year residents - 2xR1, third year residents 2xR3, fifth year residents - 2xR5 and two expert, attending laparoscopic surgeons - 2xES) each completed the experimental protocol. . The protocol consisted of two phases. During the first phase, subjects watched a 45-minute video of the surgical procedure guided by a senior surgeon to standardize the technique of the procedure into 7 steps for purposes of the study.. Following this introduction in the second phase, each subject performed a laparoscopic cholecystectomy on a pig using using the force/torque sensing instrument. All surgical procedures and animal care were reviewed and approved by the Animal Care Committee of the University of Washington. Based on pilot data analysis, force/torque data from 3 steps of the laparoscopic cholecystectomy (positioning of the gallbladder - LC-1, exposure of the cystic duct - LC-2, and dissection of the gallbladder - LC-3) were recorded. During these steps the instrumented endoscopic tool was used with an atraumatic grasper, a Babcock grasper, and a curved dissector (Fig. 1c).

2.2 Experimental System Setup

During each procedure, information was collected from two sources: (*i*) force/torque data measured at the human/tool interface and (*ii*) visual information of the tool tip interacting with the tissues. The two sources of information were synchronized in time, displayed in real time using graphical user interface, and acquired simultaneously at a sampling rate of 30 Hz for off-line analysis. Two sets of sensors measured the F/T at the interface between the surgeons’ hand and the endoscopic grasper handle (Fig 1a). The first sensor was a three-axis force/torque sensor (ATI-Mini model) which was mounted into the outer tube (proximal end) of a standard reusable 10-mm endoscopic grasper (Storz). The sensor was capable of simultaneously measuring the three components of force (F_x, F_y, F_z) and three components of torque (T_x, T_y, T_z) in a Cartesian frame (Fig. 1b). A second force sensor (Futek - FR1010) was mounted to the endoscopic grasper handle to permit the measurement of grasping force (F_g) applied by the surgeon's fingers on the instrument. For a detailed description of the system see [5, 6, 7]



(a) (b) (c)

Figure 1: The instrumented endoscopic grasper: (a) The grasper with the three axis force/torque sensor implemented on the outer tube and a force sensor located on the instrument handle (b) The tool tip and X,Y,Z frame aligned with the three axis force/torque sensor (c) Tool tips used in the surgical procedure (from left to right): Atraumatic Grasper, Babcock grasper, Curved dissector.

2.3 Data Analysis

Two types of analysis were performed on the raw data: (*i*) Video Analysis, encoding the tool-tip/tissue interactions into states; and (*ii*) Hidden Markov Modeling (HMM), for modeling. The performance of surgeons at different level of their training (R1, R3, R5, ES) was then compared.

Video analysis was performed by two expert surgeons , reviewing thevideo of each surgical procedure step, frame by frame (NTSC - 30 frames per second). The encoding process used a library of 14 different discrete tool maneuvers in which the endoscopic tool was interacting with the tissue in a unique F/T pattern (Table 1). . For example, in laparoscopic cholecystectomy, isolation of the cystic duct and artery (LC-2) involves performing repeated pushing and spreading (PS-SP - Table 1) maneuvers which in turn requires pushing forces mainly along the Z axis (F_z) and spreading forces (F_g) on the handle that form a characteristic pattern or signature. These 14 states can be grouped into three broader types (*I*, *II*, *III*) based on the number of movements performed simultaneously. Type *I* are fundamental maneuvers that include the idle state (moving the tool in space without touching any structures within the insufflated abdomen). The forces and torques used in idle state represent mainly the interaction of the trocar with the abdominal wall plus smaller gravitational and inertial forces. In the grasping and spreading states, compression and tension are applied to tissue by closing/opening the grasper handle. In the pushing state, compression is applied to tissue by moving the tool along the Z axis. For sweeping, the tool is placed in one position while rotating around the X and Y axes (trocar frame). Type *II* and type *III* states are defined as combinations of two or three Type I states (Table 2).

During the second step of the data analysis, Hidden Markov Models (HMM) and the methodology for evaluating surgical skill in laparoscopic surgery were develop. HMMs were selected for modeling the surgical procedure because their generic architecture fitted very well the nature of laparoscopic surgery task assessment. Moreover, the HMM mathematical formulation provided a very compact form that statistically summarized relatively complex tasks such as individual steps of a laparoscopic surgery procedure.

Type	State Name	State Acronym	Fx	Fy	Fz	Tx	Ty	Tz	Fg
			Force / Torque Pattern						
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 1: Definition of tool/tissue interactions and the corresponding directions of forces and torques applied during MIS.

Each laparoscopic surgical step could be decomposed into a series of finite *states* defined by the way the surgeon is interacting with the tissues (Table 1). The surgeon could move from one state to another or stay in the same. Once the surgeon was interacting with the tissue in a specific state, a certain F/T signature was applied by the surgeon through the surgical tool to the tissue. These F/T signatures, each defined as an *observation*, was composed of seven components vector of data ($F_x, F_y, F_z, T_x, T_y, T_z, F_g$). Since the F/T were continues stream of data distributed normally, each state could be defined by seven normal distributions functions chartered by a mean and a standard deviation ($N_i(m, s) \quad i = 1...7$). Combining the 7 elements vector into joint multivariable distribution function $f(O)$ was done by using Eq. 1.

$$f(O) = \frac{1}{(\sqrt{2p})^N |\Sigma|^{1/2}} e^{-(O-m)' \Sigma^{-1}(O-m)/2} \qquad (1)$$

where: O is the F/T observation vector; m is the mean vector; Σ is the covariance matrix, and N is the observation vector size.

An example of the state analysis is given in Figure 2. The diagram describes the process of deconstructing a laparoscopic surgical procedure step. Circles in this diagram represented states and lines represented transitions between states. The F/T data - observation signals were not included in Fig. 2. The HMM is termed “hidden” due to the fact that tool/tissue interactions - the states - not included in the analysis and the only observed signals are the F/T data. Although any procedure step could be decomposed manually using a frame-by-frame video analysis, this is time consuming and unnecessary since the data can also be evaluated mathematically by the HMM once its parameters are optimized.

From the mathematical perspective, four elements should be defined in order to specify a HMM (l) [23]: (*i*) the number of states in the model – *N*, (*ii*) the state transition probability distribution matrix – *A*, (*iii*) the observation symbol probability distribution matrix – *B*, and (*iv*) the initial state distribution vector– p . The HMM is then defined by the compact notation (7)

$$l = (A, B, p) \qquad (2)$$

Given the HMM architecture there are three basic problems of interest [1]: (*i*) The evaluation problem – Computing the probability (P) of the observation sequence given the model (l) and the observation sequence (O).

$$\text{Given: } \begin{cases} l = (A, B, p) \\ O = o_1, o_1, ..., o_T \end{cases} \qquad \text{Compute: } \{ P(O \mid l) \qquad (3)$$

(*ii*) Uncover the hidden states – Computing the corresponding hidden state sequence (Q), given the observation sequence (O) and the model (l).

$$\text{Given: } \begin{cases} l = (A, B, p) \\ O = o_1, o_2, ..., o_T \end{cases} \qquad \text{Compute: } \{ Q = q_1, q_2, ..., q_T \qquad (4)$$

(*iii*) The training problem – Adjusting the model parameters (A, B, p) to maximize the probability (P) of the observation sequence (O).

$$\text{Given: } \{ l = (A, B, p) \text{; Adjust: } \{ A, B, p \text{ ; Maximize: } \{ P(O \mid l) \qquad (5)$$

Using the given HMM architecture (Fig. 3a), HMMs were trained for each surgeon performing each step of the surgical procedure (8 HMM models, one for each surgeon performing one surgical procedure step). The skill level of each subject (R1, R3, R5) was evaluated based on the statistical distance between his/her HMMs and the expert surgeons (ES)

Given two HMMs l_1 and l_2 the statistical distances between them $D(l_1, l_2)$ and $D(l_2, l_1)$ were defined by Eq. 6

$$D(l_1, l_2) = \frac{1}{T_{O_2}}[\log P(O_2 \mid l_1) - \log P(O_2 \mid l_2)] \text{ ; } D(l_2, l_1) = \frac{1}{T_{O_1}}[\log P(O_1 \mid l_1) - \log P(O_1 \mid l_2)] \qquad (6)$$

$D(l_1, l_2)$ is a measure of how well model l_1 matches observations generated by model l_2 relative to how well model l_2 matches observations generated by itself. Since $D(l_1, l_2)$ and $D(l_2, l_1)$ are nonsymmetrical, The natural expression of the symmetrical version is defined by Eq. 7.

$$D_S(l_1, l_2) = \frac{D(l_1, l_2) + D(l_2, l_1)}{2} \qquad (7)$$

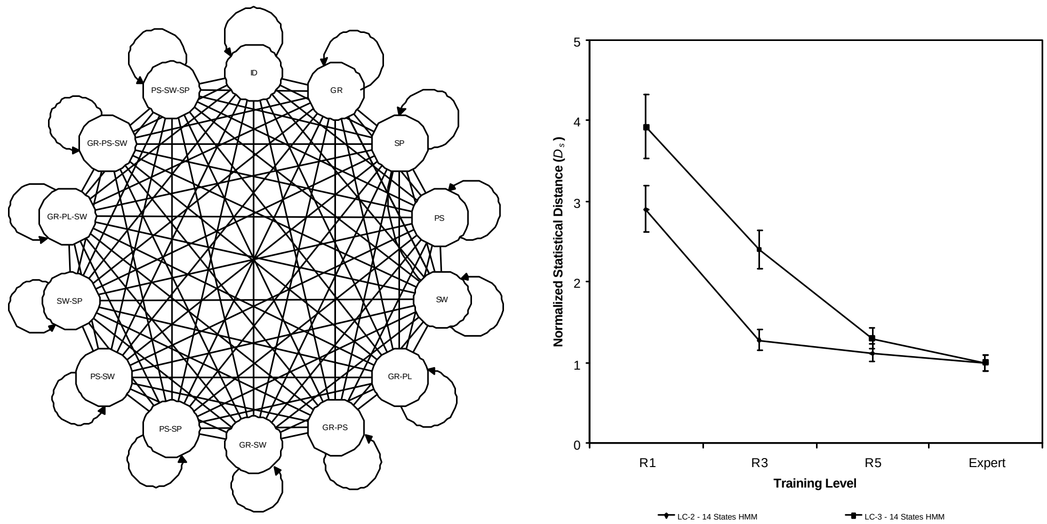
In order to scale the statistical distance between the various groups (R1, R3, R5) and the expert surgeons (ES), for each surgical procedure the statistical distance between a certain group and the expert group ($D_S(l_{Ri}, l_{ESi})$) was normalized with respect to the distance between the two experts ($D_S(l_{ES1}, l_{ES2})$) - Eq. 8.

$$\overline{D}_S(l_{Ri}, l_{ESi}) = \frac{D(l_{Ri}, l_{ESi})}{D_S(l_{ES1}, l_{ES2})} \qquad (8)$$

The practical meaning of the normalized statistical distance ($\overline{D}_S(l_{Ri}, l_{ESi})$) is how far each subject is from performing like the sampled expert surgeons.

3. Results

The data analysis demonstrated several phenomena. First, expert and novice surgeons took different paths to reach the same goal. Each group utilized states and transitions not used by the other group. Secondly, studying the median completion time of the novice surgeon group and the expert surgeon group showed a significant difference between these groups ($p<0.05$). The surgical procedure's completion time was longer for the R1 by a factor of 1.5 to 4.8 when compared to the ES. The difference between R1 and ES was more profound in steps requiring higher dexterity and more complex skills compared to steps where a specific organ was placed in a specific position (e.g. positioning of the gallbladder). The main factor contributing to the significant difference in the completion times between R1 and ES was the time spent in the *idle* state. The R1 spent significantly more time in the idle state compare to the ES.



(a) (b)

Figure 3: HMM Analysis **- (a)** HMM architecture defined by 14 fully connected state diagram (arrow heads of all the lines connected two states were omitted for simplifying the drawing) **(b)** The performance or "learning" curve of surgical residents while performing MIS **-** Normalized statistical distance between two different HMM architectures (continuous 14 state model, and discrete 4 state model) representing the performance of surgical residents(R1, R3, R5) at different year of training compared to experts surgeons

HMMs were developed foreach of the 8 subjects (and statistical distances were normalized ($\overline{D}_S$ - Eq. 8) between ES and R1 ,R3 ,R5 (Fig. 3b). The objective laparoscopic surgical skill learning curve showed significant differences between all skill levels (Fig. 7). The $\overline{D}_S$ value converged to a value of one exponentially as expertise increased although the highest gradient was between R1 and R3. This results suggest that surgical residents acquire a major portion of their laparoscopic surgical capabilities between the first and the third years of training. Calculating the $\overline{D}_S$ values for LC-1 (not plotted in Fig. 7) showed no significant difference between the groups. This is correlated with the F/T magnitude differentiation analysis between the R1 and ES. The practical meaning of that result is that a simple surgical maneuver such as LC-1 may not include sufficient haptic information to differentiate skill levels. On the other hand, more complex steps such as LC-2 and LC-3 do provide such information (Fig. 3b).

4. Discussion

Surgery with minimally invasive techniques 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 evaluating technicalperformance.. The power of this methodology is that it brings together thousands of observations through different aspect of a surgical procedure into a single, objectively-derived number. This number represents the probability that surgical performance for the subject under study approximates that of an expert.

The preliminary results expressed in this study suggest that HMMs derived from carefully standardized surgical tasks should allow objective quantification of skill based on the statistical distance between HMMs. Moreover, this methodology may be useful to determine if a surgical trainee's technical performance matches his or her peers.

Another facet of the HMM methodology for objective surgical skill assessment arises from the fact that it is not limited to the *in-vivo* conditions demonstrated in the current studybut could be extended to other modalities such as surgical simulators and robotic systems.

Acknowledgments

This work was supported by a major grant from United States Surgical, a Division of Tyco-Healthcare, Inc. to the University of Washington, USSC Center for Videoendoscopic Surgery.

References

[1] Rabiner L.R., a Tutorial on hidden Markov models and selected application in speech recognition, Proceedings of the IEEE, Vol. 77, No. 2, Feb 1989.

[2] B. Hannaford, P. Lee, "Hidden Markov Model Analysis of Force/Torque Information in Telemanipulation," Proceedings 1st International Symposium on Experimental Robotics, Montreal, June 1989.

[3] B. Hannaford, P. Lee, "Hidden Markov Model of Force Torque Information in Telemanipulation," International Journal of Robotics Research, vol. 10, no. 5, pp. 528-539, 1991.

[4] B. Hannaford, P. Lee, "Multi-Dimensional Hidden Markov Model of Telemanipulation Tasks with Varying Outcomes," Proceedings IEEE Intl. Conf. Systems Man and Cybernetics, Los Angeles, CA, Nov. 1990.

[5] Rosen J., 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, Studies in Health Technology and Informatics - Medicine Meets Virtual Reality, Vol. 62, pp. 290-296, January 1999.

[6] Rosen J., C. Richards, B. Hannaford, M. Sinanan, Hidden Markov Models of Minimally Invasive Surgery, Studies in Health Technology and Informatics - Medicine Meets Virtual Reality, Vol. 70 pp. 279-285, January 2000.

[7] Richards C., J. Rosen, B. Hannaford, M. MacFarlane, C. Pellegrini, M. Sinanan, Skills Evaluation in Minimally Invasive Surgery Using Force/Torque Signatures, Surgical Endoscopy