Biomedical Paper

Task Decomposition of Laparoscopic Surgery for Objective Evaluation of Surgical Residents' Learning Curve Using Hidden Markov Model

Jacob Rosen, Ph.D., Massimiliano Solazzo, M.D., Blake Hannaford, Ph.D., and Mika Sinanan, M.D., Ph.D.,
Departments of Electrical Engineering (J.R., B.H.) and Surgery (M.So., M.Si.), University of Washington, Seattle, Washington

ABSTRACT **Objective: Evaluation of the laparoscopic surgical skills of surgical residents is usually a subjective process carried out in the operating room by senior surgeons. The two hypotheses of the current study were: (1) haptic information and tool/tissue interactions (types and transitions) performed in laparoscopic surgery are skill-dependent, and (2) statistical models (Hidden Markov Models—HMMs) incorporating these data are capable of objectively evaluating laparoscopic surgical skills.**

Materials and Methods: Eight subjects (six residents—two first-year (R1), two third-year (R3), and two fifth-year (R5)—and two expert laparoscopic surgeons) performed laparoscopic cholecystectomy on pigs using an instrumented grasper equipped with force/torque (F/T) sensors at the hand/tool interface, and F/T data was synchronized with video of the operative maneuvers. Fourteen types of tool/tissue (T/T) interactions, each associated with unique F/T signatures, were defined from frame-by-frame video analysis. HMMs for each subject and step of the operation were compared to evaluate the statistical distance between expert surgeons and residents with different skill levels.

Results: The statistical distances between HMMs representing expert surgeons and residents were significantly different ($\alpha < 0.05$). Major differences occurred in: (1) F/T magnitudes; (2) type of T/T interactions and transitions between them; and (3) time intervals for each T/T interaction and overall completion time. The greatest difference in performance was between R1 (junior trainee) and R3 (midlevel trainee). Smaller changes were seen as expertise increased beyond the R3 level.

Conclusion: HMMs incorporating haptic and visual information provide an objective tool for evaluating surgical skills. Objective evidence for a "learning curve" suggests that surgical residents acquire a major portion of their laparoscopic skill between year 1 and year 3 of training. Comp Aid Surg 7:49–61 (2002). ©2002 Wiley-Liss, Inc.

Key words: **hidden Markov Model, laparoscopic surgery, surgical skill evaluation, minimally invasive surgery (MIS)**
Key links: **http://rcs.ee.washington.edu/brl/, http://depts.washington.edu/cves/**

Received 22 September 2000; accepted 12 November 2001.

Address correspondence/reprint requests to: Dr. Jacob Rosen, Department of Electrical Engineering, Box 352500, University of Washington, Seattle, WA 98195; Telephone: 206-616-4936; Fax: 206-221-5264; E-mail: rosen@u.washington.edu.

Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/igs.10026

INTRODUCTION

The technical performance of surgery in general, and minimally invasive surgery (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 F/T used include the nature of the tissue being manipulated, the goal of the manipulation, the type of 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.

Although the acquisition of laparoscopic technical skill and the assessment of performance are paramount for surgical resident training, the current method of choice for evaluating surgical skill is a subjective evaluation of performance in the operating room or review of a videotape of the surgery. In the last 5 years, use of surgical simulation for training and evaluating surgical skills has become a subject of ongoing active research.

Simulators for evaluating surgical skills and testing of dexterity in MIS can be roughly divided into three categories: (1) training boxes including physical objects or latex organ packs; (2) virtual reality (VR) simulators including graphical representation of virtual objects or virtual anatomy; and (3) VR simulators with a force-feedback device (haptic display) for simulating forces and torques generated as a result of interaction between the virtual objects or organs and the surgical tools.

One of the most used surgical simulators is the laparoscopic trainer box covered by an opaque membrane through which different trocars are placed at different working angles. The trainee is required to complete several structured laparoscopic tasks that are scored for both precision and speed of performance. Several studies[1–5] have found the laparoscopic training box to be a valuable teaching tool for training and evaluation of basic laparoscopic skills. Using the simulator as a basis for evaluation, performance improved over the course of residency training and correlated with postgraduate year.

A VR simulator for laparoscopic surgery models the movements needed to perform MIS and can generate a score for various aspects of psychomotor skill. A typical example of such a simulator is the MIST-VR system. The MIST-VR uses two laparoscopic instruments mounted on a frame with position sensors that provide instrument movement data that is translated into interactive real-time graphics on a PC. Targets appear randomly within the operating volume according to the skill task, and can be grasped and manipulated with the instruments. Accuracy and errors during performance of the tasks and completion time are logged. Studies performed using the MIST-VR simulator[6,7] concluded that it can objectively assess a number of desirable qualities in laparoscopic surgery, and can distinguish between experienced and novice surgeons.

The use of VR models for teaching complex surgical skills while simulating realistic human/tool and tool/tissue interaction has been a long-term goal of numerous investigators.[8–12] Although haptic devices that provide force feedback to the surgical tool while interacting with the virtual tissue/organ are commercially available (for review, see ref. 13), simulating a realistic force feedback based on biomechanical models of soft tissue is still the subject of active research.[14] The complexity of these biomechanical models is due to the viscoelasticity and nonlinear characteristics of soft tissues. Moreover, the F/T data measured *in vivo*[15–18] are crucial for designing and evaluating haptic force-feedback telerobotic systems[19–22] and VR simulators.

The methodology developed in the current study was based on Hidden Markov Models (HMMs). HMMs were extensively developed in the area of speech recognition.[23–26] Based on the theory developed for speech recognition, HMMs have become useful statistical tools in the fields of human operator modeling in general, and robotics in particular. HMMs were applied for studying teleoperation,[27–29] human manipulation actions,[30] human skills evaluation for the purpose of transferring human skill to robots,[31–33] and manufacturing applications.[34,35] Gesture recognition with HMMs has also received increasing attention from the rehabilitation technology community (see ref. 36 for review). The models are also being applied to the recognition of facial expressions from video images.[37] Moreover, HMMs may well prove useful in many other emerging applications beyond human–computer interfaces, such as DNA and protein modeling,[38] fault diagnosis in nuclear power plants,[39] and detection of pulsar signals.[40] These applications suggest that HMMs have high potential to provide better models of the human operator in complex interactive tasks with machines.

The goal of this study was to define the learning curve of MIS based on new quantitative knowledge of the F/T applied by surgeons on their instruments, and the types of tool/tissue interactions used during the course of MIS surgery. This goal was pursued through several steps: (1) developing instrumented endoscopic tools that contain embedded sensors capable of measuring and recording F/T information; (2) creating a database of F/T signals acquired during actual operating conditions on experimental animals; (3) performing a task decomposition in terms of tool/tissue interactions in MIS based on video analysis; and (4) developing statistical models (HMMs) for evaluating an objective laparoscopic skill level.

MATERIALS AND METHODS

Subjects and Protocol

Eight subjects [six general surgery residents—two first-year residents (R1), two third-year residents (R3), and two fifth-year residents (R5)—and two expert laparoscopic surgeons (ES)] each completed the experimental protocol, which consisted of two phases. During the first phase, subjects watched a 45-min video of the surgical procedure guided by a senior surgeon, to demonstrate the technique of the procedure. This study was not intended to test knowledge of the procedure, but whether the subjects could technically perform it. In the second phase, each subject performed a laparoscopic cholecystectomy on a pig using a standardized seven-step procedure. All surgical procedures and animal care were reviewed and approved by the Animal Care Committee of the University of Washington and the Animal Use Review Division of the US Army Veterinary Corps.

Data from three 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 and a curved dissector (Fig. 1c).

Experimental System Setup

During the laparoscopic procedures, data were acquired from two sources: (1) force/torque (F/T) data measured at the human/tool interface, and (2) visual information of the tool tip interacting with the tissues. The two sources of information were synchronized in time and recorded simultaneously for off-line analysis.

In MIS, the action/reaction forces and torques being applied at the interface between the surgeon's hand and the tool (tool/hand interface) are the sum of forces and torques at three different interfaces: (1) tool tip/tissue of internal organ, (2) port/abdominal wall, and (3) port/tool shaft, in addition to the gravitational forces and inertial forces and torque. Two sets of sensors measured the F/T at the interface between the surgeon's hand and the endoscopic grasper handle (Fig. 1a). The first sensor was a three-axis force/torque sensor (F/T Mini, ATI, Gamer, NC) that was mounted in 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) with a frequency response spectrum of 250 Hz (3 dB). Due to the location of the F/T sensor, it measured the forces and torques applied at the hand/tool interface, which were equal to the sum of all the forces and torques applied at the various interfaces mentioned previously. However, with the current setup, it was impossible to measure the contribution of each separate interface. The sensor orientation was such that the X and Z axes formed a plane parallel to the contact surfaces of the tool's internal jaws, and the Y and Z axes defined a plane perpendicular to that surface (Fig. 1b). A second force sensor (FR1010, Futek, Irvine, CA) was mounted on the endoscopic grasper handle to permit the measurement of grasping force (F_g) applied by the surgeon's fingers.

The grasper's mechanism had a structural feature, like most of the commercially available endoscopic graspers, enabling the surgeon to change the orientation of the tool tip relative to the tissue position without changing the handle orientation. This was achieved by rotating the entire shaft of the grasper relative to the handle, including the tool tip and its rod inside the shaft, using a knob located at the proximal end of the shaft. The importance of this setup from the engineering perspective was that the alignment of the tool-tip origin/coordinate system relative to the F/T sensor origin/coordinate system attached to the outer tube remained unchanged, because the outer tube and the tool tip are linked mechanically.

The F/T data were integrated with the laparoscopic camera view of instrument activity (Fig. 2). The seven channels of F/T data (F_x, F_y, F_z, T_x, T_y, T_z, F_g) were sampled at 30 Hz using a laptop computer with a PCMCIA 12-bit A/D card (DAQ-Card 1200, National Instruments, Austin, TX) (Fig 2a). Preliminary measurements showed that 99% of the force/torque signals' energy (PSD) was included in the 0–10-Hz frequency bandwidth. A LabView (National Instruments) application was

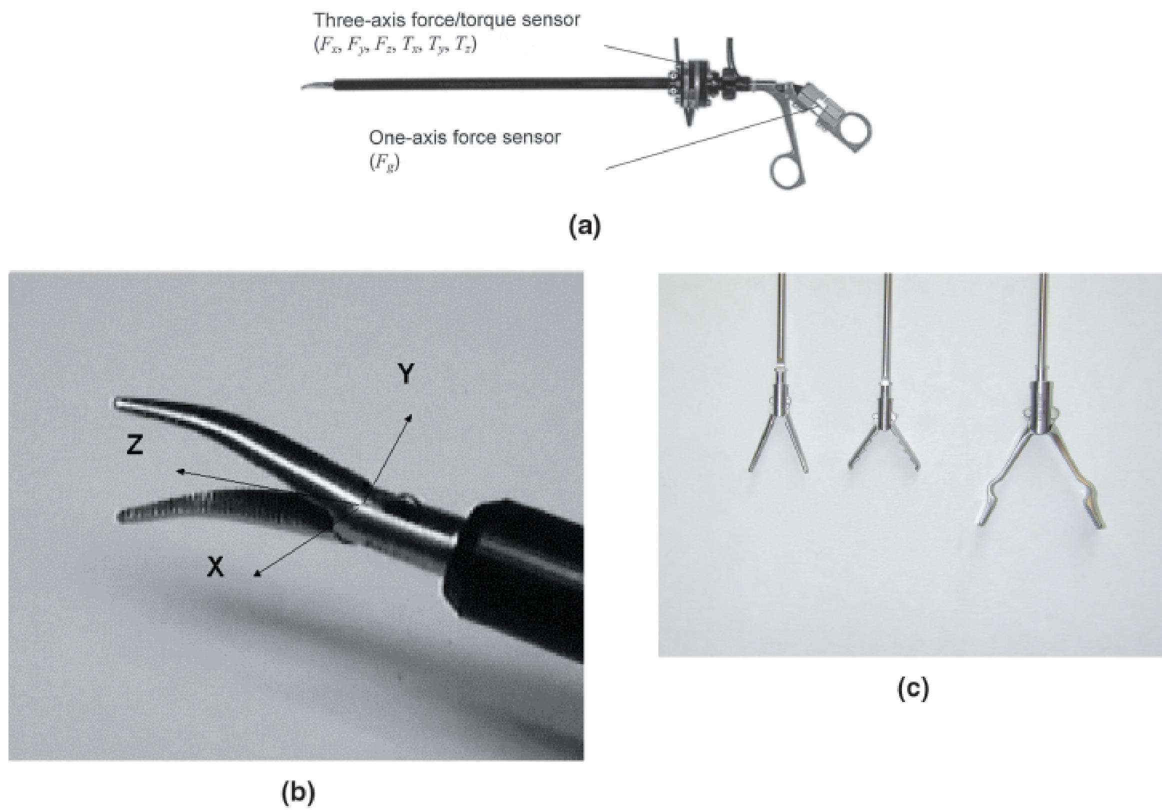


Fig. 1. The instrumented endoscopic grasper. (a) The grasper with the three-axis force/torque sensor implemented on the outer tube and a force sensor located on the instrument handle. (b) The tool tip and XYZ frame aligned with the three-axis force/torque sensor. (c) Tool tips used in the surgical procedure: (from left to right) atraumatic grasper, Babcock grasper, and curved dissector. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

developed with a graphical user interface for acquiring and visualizing the F/T data in real time during an actual operation. The video signal from the endoscopic camera that monitored the grasper's tip interacting with the internal organs or tissues was integrated with the F/T data using a video mixer in a picture-in-picture mode (PIP), allowing correlation of F/T data with instrument activity. The integrated interface was recorded during the operation for off-line frame-by-frame analysis (Fig. 2b).

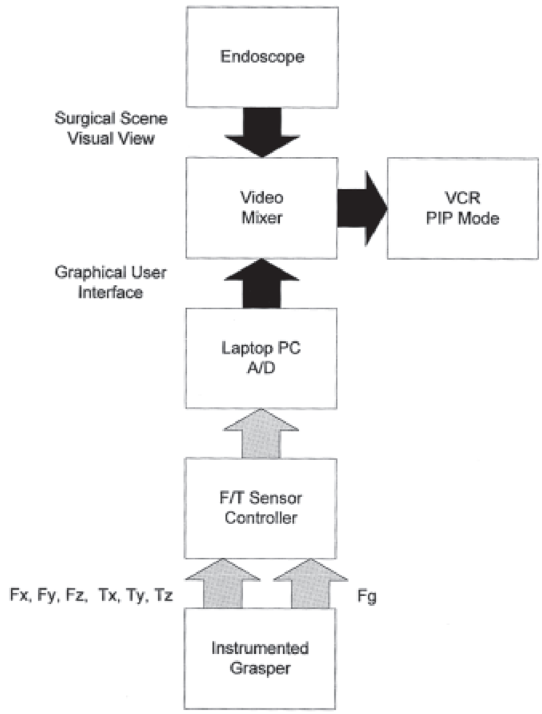
Data Analysis

Two types of analysis were performed on the raw data: (1) video analysis, encoding the tool-tip/tissue interaction into states; and (2) Hidden Markov Modeling, for modeling and comparing the performance of surgeons at different levels of their training (i.e., R1, R3, R5, or ES).

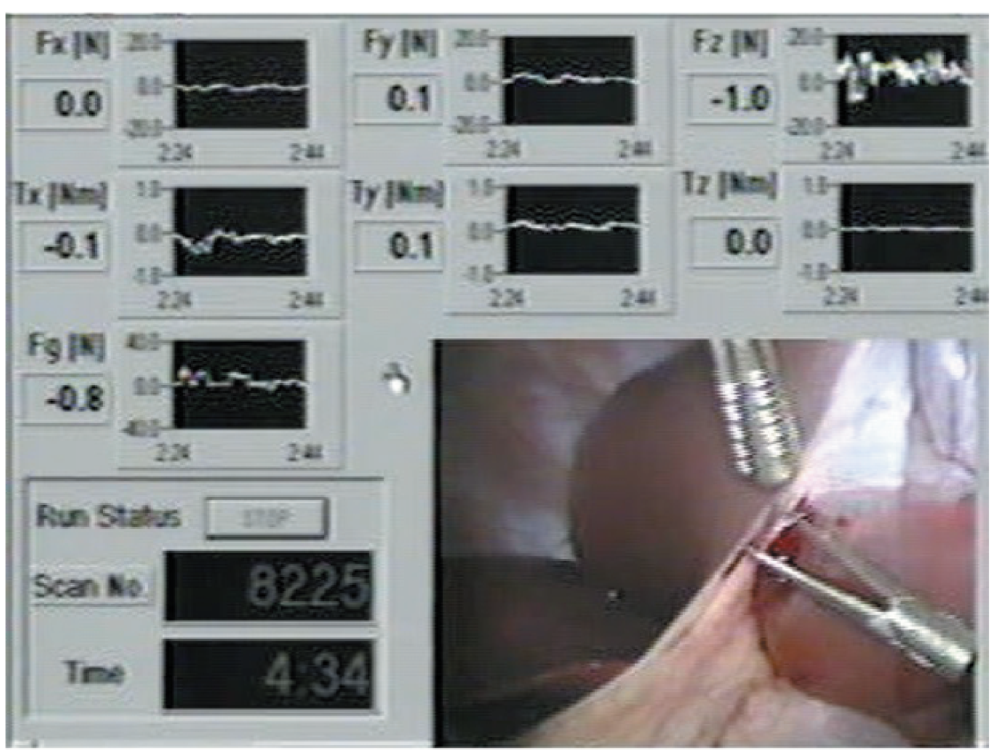
Video Analysis

The video analysis was performed by two expert surgeons encoding the video of each step of the surgical procedure frame by frame (NTSC, 30 frames per second). The encoding process used a codebook of 14 different discrete tool maneuvers in which the endoscopic tool was interacting with the tissue (Table 2). Each identified surgical tool/tissue interaction had a unique F/T pattern. For example, in the laparoscopic cholecystectomy, isolation of the cystic duct and artery (LC-2) involves performing repeated pushing and spreading (PS-SP—see Table 1) maneuvers, which in turn require pushing forces, mainly along the Z axis (F_z), and spreading forces (F_g) on the handle.

These 14 states can be grouped into three broader types based on the number of movements performed simultaneously. Fundamental maneuvers were defined as type I, and included the idle state (moving the tool in space without touching any structures within the insufflated abdomen). The forces and torques used in the idle state mainly represented the interaction of the port with the abdominal wall, in addition to gravitational and inertial forces. In the grasping and spreading states, compression and tension were applied to the tissue



(a)



(b)

Fig. 2. Experimental setup. (a) Block diagram of the experimental setup integrating the force/torque data and the view from the endoscopic camera. (b) Real-time user interface of force/torque information synchronized with the endoscopic view of the procedure using picture-in-picture mode. (Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com)

Table 1. Definition of Tool/Tissue Interactions and the Corresponding Directions of Forces and Torques Applied During MIS

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	±	±	−	±	±		−

by closing/opening the grasper handle. In the pushing state, compression was applied to the tissue by moving the tool along the Z axis. For sweeping, the tool was placed in one position while rotating around the X and Y axes (port frame). Type II and type III states were defined as combinations of two or three states defined by type I (Table 2).

Hidden Markov Model (HMM)

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

The rationale for developing a methodology for objective evaluation of surgical skills based on HMMs and F/T measurements at the human/tool interface came from our previous study[41] showing that F/T applied at that interface varied according to skill levels and the tasks being performed. High F/T magnitudes defined by the absolute value of the force-torque vectors

$$|\vec{F}| = \sqrt{F_x^2 + F_y^2 + F_z^2} \qquad \text{(A)}$$

$$|\vec{T}| = \sqrt{T_x^2 + T_y^2 + T_z^2} \qquad \text{(B)}$$

were applied by novice surgeons (NS) compared to expert surgeons (ES), while performing tissue manipulation. This might be a result of insufficient dexterity on the part of the NS that might have 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 in an effort to avoid irreversible tissue damage. As a result, the NS had to perform more repetitions of the dissection movements to tear the tissue, substantially decreasing the efficiency of the MIS procedure.

Each laparoscopic surgical step could be decomposed into a series of finite *states* defined by the way the surgeon interacts with the tissues (Table 1). The surgeon could move from one state to the other or remain in the same state for a certain amount of time. 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,* were composed of seven component vectors of data (F_x, F_y, F_z, T_x, T_y, T_z, F_g). The F/Ts were continuous streams of data distributed normally, each state being defined by seven normal distributions functions chartered by a mean and a standard deviation [$N_i(\mu, \sigma)$ $i = 1 \ldots 7$)]. Combining the seven-element vector into a joint multivariable distribution function $f(O)$ was done using Equation (1):

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

where O is the F/T observation vector, μ is the

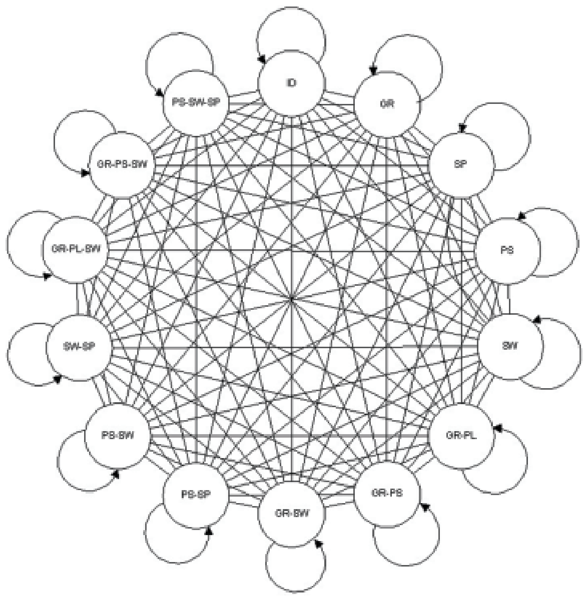


Fig. 3. HMM architecture defined by a 14 fully connected-state diagram (arrowheads of all lines connecting two states were omitted to simplify the drawing).

mean vector, Σ is the covariance matrix, and N is the observation vector size.

The state diagram (Fig. 3) describes the decomposed process of a typical 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 Figure 3.

The HMM is termed "hidden" due to the fact that tool/tissue interactions—the states—are hidden, and the only observed signals are the F/T data. Although the state could be decomposed manually using a frame-by-frame video analysis, this is time consuming and unnecessary, because 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 an HMM (λ):[23] (1) the number of states in the model N; (2) the state transition probability distribution matrix A; (3) the observation symbol probability distribution matrix B; and (4) the initial state distribution vector π. The HMM is then defined by the compact notation (2)

$$\lambda = (A, B, \pi) \tag{2}$$

Given the HMM architecture, there are three basic problems of interest:[23]

1. The evaluation problem: computing the probability (P) of the observation sequence, given the model (λ) and the observation sequence (O).

$$\text{Given: } \begin{Bmatrix} \lambda = (A,B,\pi) \\ O = o_1, o_1, \ldots, o_T \end{Bmatrix} \tag{3}$$

$$\text{Compute: } \{P(O \mid \lambda).$$

2. Uncovering the hidden states: computing the corresponding hidden state sequence (Q), given the observation sequence (O) and the model (l).

$$\text{Given: } \begin{Bmatrix} \lambda = (A,B,\pi) \\ O = o_1, o_2, \ldots, o_T \end{Bmatrix} \tag{4}$$

$$\text{Compute: } \{Q = q_1, q_2, \ldots, q_T$$

3. The training problem: adjusting the model parameters (A, B, π) to maximize the probability (P) of the observation sequence (O).

$$\text{Given: } \{\lambda = (A, B, \pi)$$

$$\text{Adjust: } \{A, B, \pi$$

$$\text{Maximize: } \{P(O \mid \lambda) \tag{5}$$

Using the given HMM architecture (Fig. 3), HMMs were developed for each surgeon performing each step of the surgical procedure (eight 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 those of the the expert surgeons (ES). Given two HMMs λ_1 and λ_2, the statistical distances between them, $D(\lambda_1, \lambda_2)$ and $D(\lambda_2, \lambda_1)$, were defined by Equation (6):

$$\begin{aligned} D(\lambda_1,\lambda_2) &= \frac{1}{T_{O_2}}[\log P(O_2|\lambda_1) - \log P(O_2|\lambda_2)] \\ D(\lambda_2,\lambda_1) &= \frac{1}{T_{O_1}}[\log P(O_1|\lambda_1) - \log P(O_1|\lambda_2)] \end{aligned} \tag{6}$$

$D(\lambda_1, \lambda_2)$ is a measure of how well model λ_1 matches observations generated by model λ_2 relative to how well model λ_2 matches observations generated by itself, whereas T_{O_1} and T_{O_2} stand for the time duration of the observation vectors O_1 and O_2, respectively. Because $D(\lambda_1, \lambda_2)$ and $D(\lambda_2, \lambda_1)$

are nonsymmetrical, the natural expression of the symmetrical statistical distance version is defined by Equation (7):

$$D_s(\lambda_1,\lambda_2) = \frac{D(\lambda_1,\lambda_2) + D(\lambda_2,\lambda_1)}{2} \tag{7}$$

To scale the statistical distance between each of the two subjects defined by the index $j = 1, 2$, and associated with three group levels defined by the index $i = 1, 3, 5$ (R1, R3, R5) and the two subjects of the expert surgeons defined by the index $k = 1, 2$ [ES(1), ES(2)], for each surgical procedure, the statistical distance between a certain subject j in group i and a certain subject k in the expert group [$D_S(\lambda_{\mathrm{Ri}}, \lambda_{\mathrm{ESj}})$] was normalized with respect to the distance between the two expert subjects [$D_S(\lambda_{\mathrm{ES1}}, \lambda_{\mathrm{ES2}})$] [see Equation (8)].

For $i = 1, 3, 5$, $j = 1, 2$, and $k = 1, 2$,

$$\bar{D}_S(\lambda_{Ri(j)},\lambda_{ES(k)}) = \frac{D(\lambda_{Ri(j)}, \lambda_{ES(k)})}{D_S(\lambda_{ES1},\lambda_{ES2})} \tag{8}$$

The practical meaning of the normalized statistical distance ($D_s(\lambda_{Ri(j)}, \lambda_{ES(k)})$) is to define quantitatively how far each subject is from being an expert surgeon. Because each group included two subjects, calculating all the combinations of the statistical distances between subjects of the *ES* group and subjects of any of the *Ri* group provides four values: $n = 4$ ($\bar{D}_S(\lambda_{Ri(1)}, \lambda_{ES(1)})$, $\bar{D}_S(\lambda_{Ri(1)}, \lambda_{\mathrm{ES}(2)})$, $\bar{D}_S(\lambda_{\mathrm{Ri}(2)}, \lambda_{\mathrm{ES}(1)})$, $\bar{D}_S(\lambda_{\mathrm{Ri}(2)}, \lambda_{\mathrm{ES}(2)})$). The average and the standard deviation of these four values were calculated to provide a single statistical measurement between two skill-level groups. The average of the normalized statistical distance between two groups was defined by Equation (9):

For $i = 1, 3, 5$, and $n = 4$,

$$\bar{D}_S(\lambda_{Ri},\lambda_{ES}) = \frac{\sum_{j=1,k=1}^{j=2,k=2} \bar{D}_S(\lambda_{Ri(j)}, \lambda_{ES(k)})}{n} \tag{9}$$

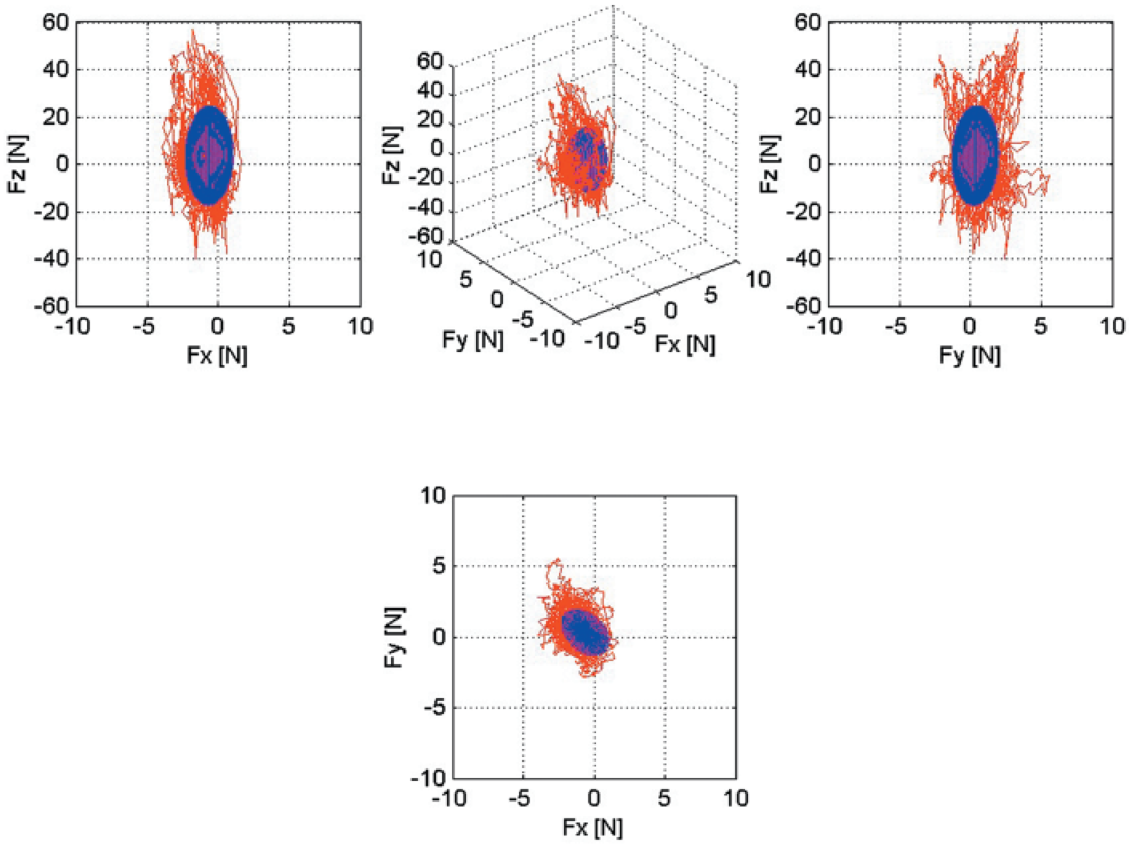
RESULTS

Typical raw data of forces and torques were plotted in a 3D space, showing the loads developed at the sensor location while the gallbladder fossae were dissected in 482 s by an expert surgeon during laparoscopic cholecystectomy (Fig. 4). The forces and torques measured by the F/T ATI sensor can be described as vectors with an origin at the center of the sensor and the coordinate system aligned with the tool coordinate system (Fig. 1b). These vectors are constantly changing both their magnitudes and orientations as a result of the F/T applied by the surgeon's hand on the tool while interacting with the tissues. The F/T vectors can be depicted as arrows attached to the origin, changing their lengths and orientations as a function of time. Figure 4 shows the trace of the tips of these vectors (arrows) as they change during MIS. The forces and torques are represented by a 3D plot in addition to the three orthogonal planes. The ellipsoid defines a region including 95% of the F/T samples.

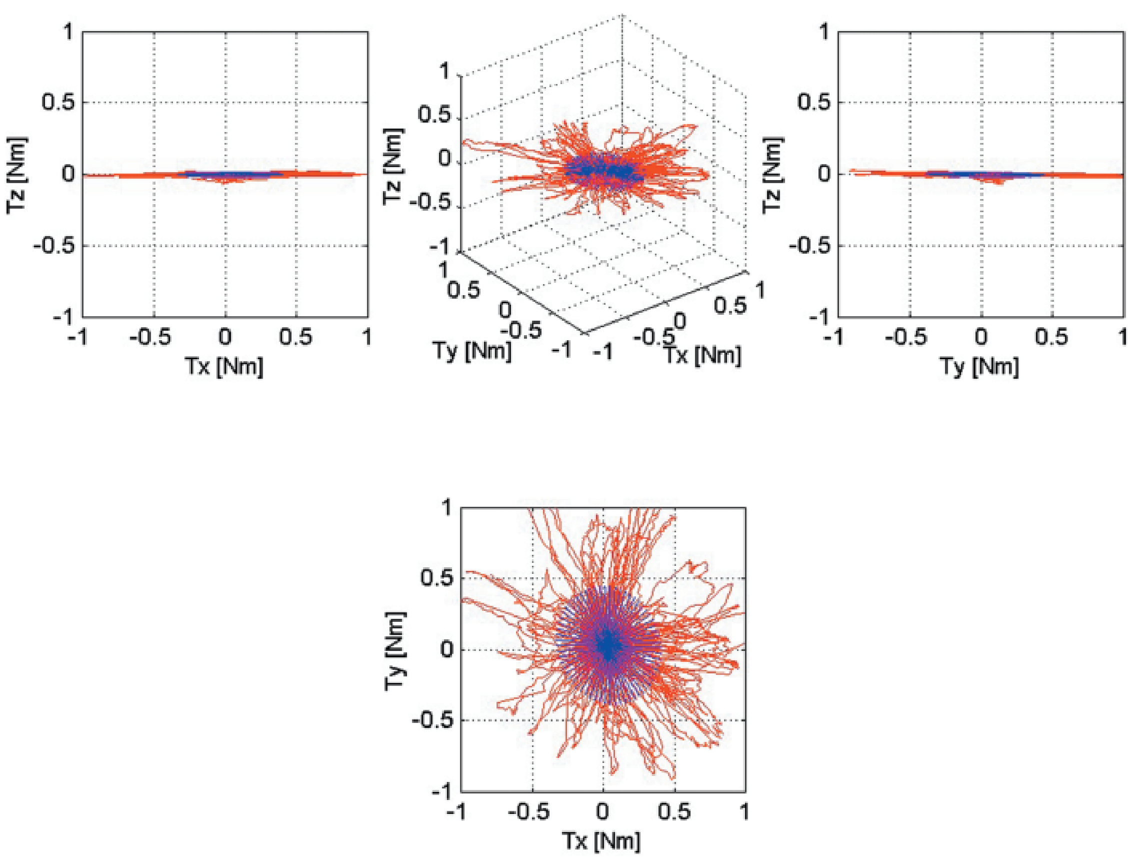
The forces along the Z axis (in/out of the port) were higher compared to the forces in the XY plane. On the other hand, torques developed by rotating the tool around the Z axis were extremely low compared to the torques generated while rotating the tool along the X and Y axes and sweeping the tissue or performing lateral retraction. Similar trends in terms of the F/T magnitude ratios between the X, Y, and Z axes were found in the data measured in other steps of the MIS procedures. These raw data demonstrated the complexity of the surgical task. Deeper understanding of this task is gained by decomposing it to its prime elements, as demonstrated by the video analysis.

Frame-by-frame analysis of videotapes of the surgical procedures incorporating the visual view of the tool/tissue interaction and graphs of the F/T at the tool/hand interface allowed definition of the primary tool/tissue interactions in the MIS procedure, and of the direction of forces and torques associated with them (Table 1). Once these tool/tissue interaction archetypes were defined, each step of the surgical procedure could be manually decomposed into a list of tool/tissue interactions. This list was further transformed into a more compact diagram (as shown in Fig. 5) defining a typical tool/tissue transition for a surgical procedure. The tool/tissue transition diagram (state diagram) depicted in Figure 5 represents the surgical step in which the gallbladder fossa was dissected.

The idle state is the only state connected to all the others in the state transition diagram (Fig. 5). This state, in which no tool/tissue interaction was performed, was mainly used by both expert and novice surgeons to move from one operative state to the other. However, the expert surgeons used the idle state only as a transition state, while the novices spent significant amounts of time in this state planning the next tool/tissue interaction. Another major difference between surgeons from different skill groups was related to the tool/tissue interac-



(a)



(b)

Fig. 4. Forces (a) and torques (b) measured at the human/tool interface while dissecting the gallbladder fossa during laparoscopic cholecystectomy. For the definitions of the X, Y, and Z directions, see Figure 1b. (Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com)

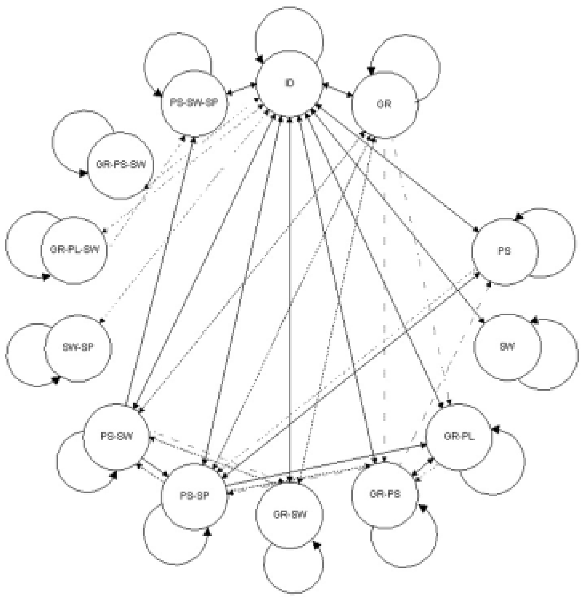


Fig. 5. The state diagram based on a frame-by-frame video analysis of the dissection of the gallbladder fossa during laparoscopic cholecystectomy. Each circle represents a different state characterized by a tool/tissue interaction, and the arrows represent transitions between states. In some cases, the surgeon stays within the same state: this is depicted by an arrow to the same state. (Dashed line: states and transitions performed by ES; dotted line: states and transitions performed by R1 representing a novice surgeon; solid line: states and transitions performed both by R1 and ES).

tion and tool/tissue transitions used by these two groups. Surgeons took different paths to reach the same goal (Fig. 5). Each group utilized states and transitions not used by the other group. Figure 5, for example, was essentially constructed from two separate models representing the expert surgeon's model (ES) (solid line and dashed line) and the novice surgeon's model (R1) (solid line and dotted line). For the purpose of evaluating surgical skills using the HMM, the model representing different groups must share the same architecture. This requirement led to generalized model architecture, as described previously in Figure 3.

The final and most profound analysis included the HMM analysis. HMMs were developed for each one of the eight subjects (R1, R3, R5, ES) that performed the three steps of the laparoscopic cholecystectomy (LC-1, LC-2, LC-3). The normalized statistical distances [$\overline{D}_s$—see Equation (8)] between ES and R1, R3, and R5 (Fig. 6) were plotted in Figure 7. The strength of this methodology is that it brings together different aspects of the surgical procedure into a single number that practically indicates how far the surgical performance of the subject under study is from that of an expert performing the same surgical task. In this way, $\overline{D}_s$ provides an objective criterion for evaluating surgical performance in MIS.

The objective laparoscopic surgical skill learning curve showed significant differences between all skill levels (Fig. 7). The value of the normalized statistical distance ($\overline{D}_s$) between various skill-level groups (R1, R3, and R5) and the ES converged exponentially to a value of 1 as the level of expertise increased. However, the highest gradient was between R1 and R3. This results indicate that surgical residents acquire a major portion of their laparoscopic surgical capabilities between the first and third years of their residency training. Calculating the $\overline{D}_s$ values for LC-1 (not plotted in Fig. 7) showed no significant difference between the groups. The practical meaning of this result is that LC-1 does not include sufficient haptic information to use this data for differentiating between groups with different skill levels. On the other hand, LC-2 and LC-3 do provide such information, as shown in Figure 7.

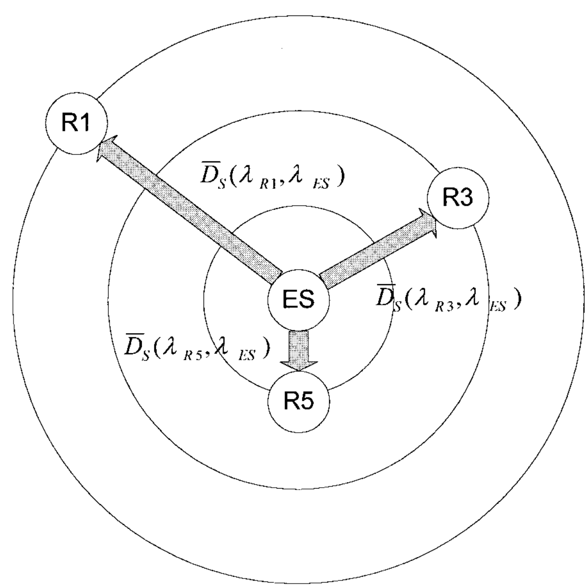


Fig. 6. Schematic representation of the averaged statistical distance $\overline{D}_s$ defining the statistical similarities between the expert group and various groups of residents associated with different skill levels. Note that each group included two subjects, so the statistical distance $\overline{D}_s$ between two selected groups is the average of four distances.

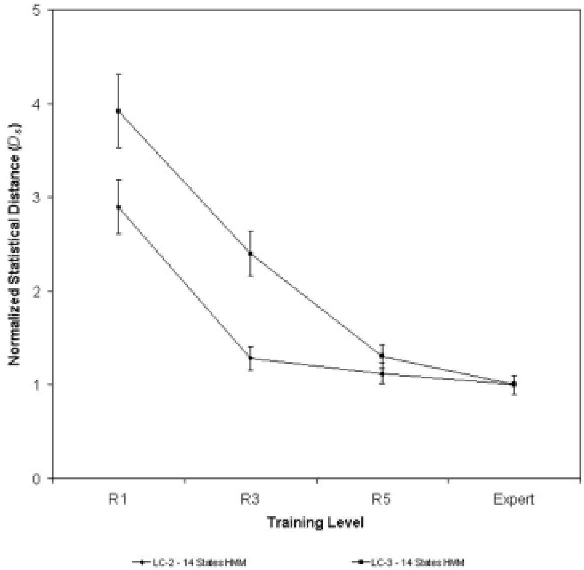


Fig. 7. The learning curve of surgical residence while performing laparoscopic cholecystectomy: normalized statistical distance between surgical residents at different stages of their training (R1, R3, R5) and expert surgeons while performing selected steps of a laparoscopic cholecystectomy (exposure of the cystic duct, LC-2, and dissection of the gallbladder, LC-3). The vertical bars around data points depict the standard deviations from the average statistical distances between the various groups.

DISCUSSION

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 VR simulators incorporating haptic technology and master/slave robotic systems for telesurgery. In addition, using the F/T information in real time during the course of learning to provide feedback to the surgical residents may improve the learning curve, reduce soft-tissue injury, and increase 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 based on the statistical distance between HMMs representing surgical residents at different levels of their training and HMMs representing expert surgeons. Moreover, this methodology can be used to determine if the performance of a student matches his/her training level. The advantage of using this method in terms of data reduction is that it condenses enormous amounts of multidimensional data into a single datum expressed as the normalized statistical distance from an expert performance, or, in other words, the statistical similarity between the subject under study and an expert performer. The strength of the methodology using HMM for objective surgical skill assessment is that it is not limited to the *in vivo* condition as demonstrated in the current study. It can be extended to other modalities, such as surgical simulators and robotic systems for telesurgery. The proposed methodology derives its power from decomposing the surgical task to its prime elements: tool/tissue interactions. These elements are inherent in MIS no matter which modality is being used.

Increasing the size of the database to include more surgical procedures performed by more surgeons could extend the approach outlined in this study. This extension is essential for full validation of the proposed methodology. An intermediate step toward achieving that goal is to determine how many subjects at each level are required to show a statistically significant difference between different skill levels. These requirements initiated a more extensive experimental protocol, currently underway, which includes five subjects in each skill group. This new database will provide a deeper insight into the complex multidisciplinary issue of objective evaluation of surgical skills.

Another possible approach to further exploration of the proposed methodology that avoids *in vivo* experiments on laboratory animals is to use VR simulators incorporating haptic devices. This transformation from an *in vivo* surgical setup to VR simulators will be possible only when these simulators can realistically represent the surgical setup from both the graphic and haptic perspectives. This information, combined with other feedback data, may form the basis of teaching techniques for optimizing tool usage in MIS. The novice surgeons could practice their skills outside the operating room using realistic VR simulators until they have 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, and by a major grant from U.S. Surgical, a division of Tyco, Inc.,

to the University of Washington, Center for Videoendoscopic Surgery.

REFERENCES

1. Rosser JC, Rosser LE, Savalgi RS. Objective evaluation of a laparoscopic skill program for residents and senior surgeons. Arch Surg 1998;133:657–661.
2. Rosser JC, Rosser LE, Savalgi RS. Skill acquisition and assessment for laparoscopic surgery. Arch Surg 1997;132:200–204.
3. Derossis AM, Fried GM, Abrahamowicz M, Sigman HH, Barkun JS, Meakins JL. Development of a model for training and evaluation of laparoscopic skills. Am J Surg 1998;175:482–487.
4. Derossis AM, Antoniuk M, Fried GM. Evaluation of laparoscopic skills: a 2-year follow-up during residency training. CJS 1999;42(4):293–296.
5. Melvin WS, Johnson JA, Ellison EC. Laparoscopic skill enhancement. Am J Surg 1996;172:377–379.
6. Gallagher AG, McClure N, McGuigan J, Crothers I, Browning J. Virtual reality training in laparoscopic surgery: a preliminary assessment of minimally invasive surgical trainer virtual reality (MIST VR). Endoscopy 1999;31(4):310–313.
7. Chaudhry A, Sutton C, Wood J, Stone R, McCloy R. Learning rate for laparoscopic surgical skills on MIST VR, a virtual reality simulator: quality of human–computer interface. Ann R Coll Surg Engl 1999;81(4): 281–286.
8. Satava R. Virtual reality surgical simulator. Surg Endosc 1993;7:203–205.
9. Ota D, Loftin B, Saito T, Lea R, Keller J. Virtual reality in surgical education. Comput Biol Med 1995; 25:127–137.
10. Noar M. Endoscopy simulation: a brave new world? Endoscopy 1991;23:147–149.
11. Rosen J. Advanced surgical technologies for plastic and reconstructive surgery. Otolaryngol Clin N Am 1998;31:357–381.
12. Bicchi A, et al., A sensorized minimally invasive surgery tool for detecting tissue elastic properties. Proceedings of the IEEE International Conference on Robotics and Automation, 1996. p 884–888.
13. Chen E, Marcus B. Force feedback for surgical simulators. Proc IEEE 1998;86(3):524–530.
14. Berkley J, Oppenheimer P, Weghorst S, Berg D, Raugi G, Haynor D, Ganter M, Brooking C, Turkiyyah G. Creating fast finite element models from medical images. In: Westwood JD, Hoffman HM, Mogel GT, Robb RA, Stredney D, editors: Medicine Meets Virtual Reality 2000. Envisioning Healing: Interactive Technology and the Patient–Practitioner Dialogue. Studies in Health Technology and Informatics Vol. 70. Amsterdam: IOS Press, 2000. p 26–32.
15. Morimoto A, Foral R, Kuhlman J, Zucker K, Curet M, Bocklage T, MacFarlane T, Kory L. Force sensor for laparoscopic Babcock. In: Morgan KS, Hoffman HM, Stredney D, Weghorst SJ, editors. Medicine Meets Virtual Reality 1997, Global Healthcare Grid Studies in Health Technology and Informatics vol. 39. Amsterdam: IOS Press, 1997. p 354–361.
16. MacFarlane M, Rosen J, Hannaford B, Pellegrini C, Sinanan M. Force feedback grasper helps restore the sense of touch in minimally invasive surgery. J Gastrointestinal Surg 1999;3(3):278–285.
17. Rosen J, MacFarlane M, Richards C, Hannaford B, Sinanan M. Surgeon–tool force/torque signatures—evaluation of surgical skills in minimally invasive surgery. In: Westwood JD, Hoffman HM, Robb RA, Stredney D, editors: Medicine Meets Virtual Reality. The Convergence of Physical and Informational Technologies: Options for a New Era in Medicine. Studies in Health Technology and Informatics Vol. 62. Amsterdam: IOS Press, 1999. p 290–296.
18. Rosen J, Richards C, Hannaford B, Sinanan M. Hidden Markov models of minimally invasive surgery. In: Westwood JD, Hoffman HM, Mogel GT, Robb RA, Stredney D, editors: Medicine Meets Virtual Reality 2000. Envisioning Healing: Interactive Technology and the Patient–Practitioner Dialogue. Studies in Health Technology and Informatics Vol. 70. Amsterdam: IOS Press, 2000. p 279–285.
19. Hannaford B, Trujillo J, Sinanan M, Moreyra M, Rosen J, Brown J, Lueschke R, MacFarlane M. Computerized endoscopic surgical grasper. In: Westwood JD, Hoffman HM, Stredney D, Weghorst SJ, editors: Medicine Meets Virtual Reality. Art, Science, Technology: Healthcare (R)evolution. Studies in Health Technology and Informatics Vol. 50. Amsterdam: IOS Press, 1998. p 265–271.
20. Rosen J, Hannaford B, MacFarlane M, Sinanan M. Force controlled and teleoperated endoscopic grasper for minimally invasive surgery—experimental performance evaluation. IEEE Trans Biomed Eng 1999; 46(10):1212–1221.
21. Buess GF, et al. Robotics in allied technologies. Arch Surg 2000;135:229–235.
22. Howe RD, Matsuoka Y. Robotics for surgery. Annu Rev Biomed Eng 1999;1:211–240.
23. Rabiner LR. A tutorial on hidden Markov models and selected application in speech recognition. Proc IEEE 1989;77(2):257–286.
24. Baker JK. The DRAGON system—an overview. IEEE Trans Acoustics Speech Signal Processing 1975;23(1):24–29.
25. Levinson SE, Rabiner LR, Sondhi MM. An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition. Bell Sys Tech J 1983;62(4):1035–1074.
26. Reddy R. Foundations and grand challenges of artificial intelligence: 1988 AAAI presidential address. AI Magazine, Winter 1988, p 9–21.
27. Hannaford B, Lee L. Hidden Markov model analysis of force/torque information in telemanipulation. Proceedings of the 1st International Symposium on Experimental Robotics, Montreal, June 1989.

28. Hannaford B, Lee P. Hidden Markov model of force torque information in telemanipulation. Int J Robotics Res 1991;10(5):528–539.
29. Hannaford B, Lee P. Multi-dimensional hidden Markov model of telemanipulation tasks with varying outcomes. Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Los Angeles, CA, November 1990.
30. Pook P, Ballard DH. Recognizing teleoperated manipulations. Proceedings of IEEE Robotics and Automation, Atlanta, GA, May 1993. Vol. 2. p 578–585.
31. Nechyba MC, Xu Y. Stochastic similarity for validating human control strategy models. IEEE Trans Robotics Automation 1998;14(3):437–451.
32. Yang J, Xu YS, Chen CS. Human action learning via hidden Markov model. IEEE Transactions on Systems, Man and Cybernetics, Part A (Systems and Humans) 1997;27(1):34–44.
33. Itabashi K, Hirana K, Suzuki T, Okuma S, Fujiwara F. Modeling and realization of the peg-in-hole task based on hidden Markov model. Proceedings of the IEEE International Conference on Robotics and Automation, Leuven, Belgium, May 1998. p 1142.
34. Hannaford B. Hidden Markov model analysis of manufacturing process information. Proceedings of IROS 1991, Osaka, Japan, November 1991.
35. McCarragher BJ, Hovland G, Sikka P, Aigner P, Austin D. Hybrid dynamic modeling and control of constrained manipulation systems. IEEE Robotics and Automation Magazine 1997;4(2).
36. Wachsmuth I, Frohlich M, editors. Gesture and Sign Language in Human–Computer Interaction. International Gesture Workshop Proceedings. Berlin: Springer, 1998.
37. Lien JJ, Kanade T, Cohn JF, Ching CL. Automated facial expression recognition based on FACS action units. Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 14–16 April 1998. p 390–395.
38. Baldi P, Brunak S. Bioinformatics. New York: MIT Press, 1998.
39. Kwon K. Application of HMM to accident diagnosis in nuclear power plants. Proceedings of Topical Meeting on Computer-Based Human Support Systems. American Nuclear Society, Philadelphia, PA, June 1995.
40. Freed P. Detecting pulsars with hidden Markov models. Third International Symposium on Signal Processing and Applications. Vol. 1. p 179–183.
41. Rosen J, Hannaford B, Richards C, Sinanan M. Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills, IEEE Trans Biomed Eng 2001;48(5) p. 579–591.