

Sensor fusion for laparoscopic surgery skill acquisition

FRASER ANDERSON¹, DANIEL W. BIRCH², PIERRE BOULANGER¹, &
WALTER F. BISCHOF¹

¹Department of Computer Science and ²Department of Surgery, University of Alberta, Edmonton, Alberta, Canada

(Received 15 November 2011; accepted 13 August 2012)

Abstract

Surgical techniques are becoming more complex and require substantial training to master. The development of automated, objective methods to analyze and evaluate surgical skill is necessary to provide trainees with reliable and accurate feedback during their training programs. We present a system to capture, visualize, and analyze the movements of a laparoscopic surgeon for the purposes of skill evaluation. The system records the upper body movement of the surgeon, the position, and orientation of the instruments, and the force and torque applied to the instruments. An empirical study was conducted using the system to record the performances of a number of surgeons with a wide range of skill. The study validated the usefulness of the system, and demonstrated the accuracy of the measurements.

Keywords: *Surgical skill, motion analysis, motor skill, laparoscopy*

Introduction

Surgical techniques are becoming less invasive and more difficult to master. In particular, laparoscopic techniques are more difficult than open surgery for a number of reasons. The instruments are quite long, causing movements made at the handle to be amplified at the tool tip with distorted haptic return. The movements are also reversed due to the lever effect of the trocar: moving the handle to the left moves the tool tip to the right within the surgical field. The surgical instruments are also limited to five degrees of freedom, thus restricting the movements of the surgeon. In addition, the operating field is shown on a monitor that is typically placed high above the patient, making it difficult to watch the surgical field and monitor hand position at the same time. Laparoscopic techniques thus require substantial manual dexterity and a lengthy training process.

Training for laparoscopic surgery is typically performed by having trainees practice on synthetic box trainers (e.g., performing suturing on rubber tissue), live pig models, or more recently, virtual

reality (VR) training systems [1]. Virtual reality trainers hold the most potential for future training methods, as they are not as expensive as porcine models and, unlike synthetic box trainers, do not require the presence of an expert surgeon to evaluate the performance. However, VR trainers are currently unable to provide accurate, meaningful feedback to surgeons, and more work is needed to determine how to best provide an automated analysis of surgical movements.

Accurate skill evaluation is needed not only to provide feedback to surgeons during training, but also to ensure that expert surgeons maintain their high skill level. Current methods of evaluating surgical dexterity are quite subjective, relying on an expert's judgment. The expert typically watches a surgeon perform a number of maneuvers and assigns the surgeon a proficiency score. These scores can vary widely between experts and are thus unreliable. Checklists and scoring sheets, such as the Objective Structured Assessment of Technical Skill (OSATS) rating system [2], represent attempts to remove this subjectivity by providing a standardized scoring sheet for surgical tasks.

Correspondence: Fraser Anderson, Department of Computing Science, 3–63 CSC, University of Alberta, Edmonton, Alberta T6G 2E8, Canada.
E-mail: fraser@ualberta.ca

However, this still requires an expert to assess a surgeon's movement, interpret the scoring sheet, and assign a relevant score.

In recent years, technology has been adopted to address the problem of evaluating surgical dexterity [3]. Motion capture and force-sensing devices allow a computer to record and analyze the surgeon's movements. By relying solely on the data from the sensors, the computer can produce an evaluation of surgical skill that is free from human subjectivity. This method of evaluation has the possibility of being more reliable and accurate than an evaluation by an expert surgeon, but further research is needed before it can replace the expert evaluator. Current research in this area focuses on developing systems to record the movements of the surgeon, and on processing the data to determine how measurements correlate with skill.

In this article, we present a novel recording system that is capable of monitoring the movements of the surgeon, as well as the force and torque applied to the surgical instruments. Our system produces minimal interference with the surgeon's movements, adapts to many contexts, and is designed using several generalizable elements that can be leveraged in other system designs. In addition, our approach enables us to use sensor fusion to compute the total mechanical work done by the surgeon, which we find correlates negatively with surgical skill, and offers unique information not present with other kinematic measures. Our results were validated in an empirical study in which participants performed three separate tasks on a bench-top laparoscopic simulator.

Related work

Several types of recording systems have been used to capture the movements of a surgeon. Electromagnetic (EM) tracking has been used in numerous studies of surgical movement, both outside and within controlled operating room (OR) environments [4–7]. EM tracking systems emit an EM field using orthogonal coils, and use a set of sensors whose position and orientation is determined by the relative strength of the EM field. One of the most popular EM tracking systems for surgical environments is the ICSAD system [8], which uses Polhemus Isotrak II sensors placed on the back of the surgeon's hands. EM sensors are quite small, hence they can be attached to a variety of existing surgical tools as well as to the surgeon. Unfortunately, the magnetic field used by all EM systems is heavily distorted by ferromagnetic objects, often resulting in noisy and distorted measurements in a clinical environment.

Optical motion capture systems track objects in 3D space by locating markers in video streams. Emam et al. used a set of reflective markers placed on the upper body to monitor the motions of the shoulders and elbows of novices and experts and analyze ergonomic factors in laparoscopy [9, 10]. Optical tracking has also been combined with EM tracking to add redundancy and improve the quality of data, as in the study by Hwang et al. [11]. While precise data on the actual tracking accuracy in an OR setting was not provided, the system was able to distinguish between novices and experts using a number of features derived from the measurements.

Surgical motions can also be tracked using instrumented mechanical links attached directly to surgical tools. The joints of the mechanical links are fitted with angle encoders, usually rotational potentiometers or optical encoders. The nature of these systems makes them cumbersome and difficult to use in an OR environment. To date, no systems with mechanical motion capture have been used in operations on humans, but the Blue DRAGON system was used to record movements during laparoscopic operations on animal models [12]. Mechanical motion capture devices most commonly appear as part of a virtual reality training package, such as the LapSim (www.surgical-science.com). The workspace for these virtual trainers is inherently small, instruments do not need to be interchanged, and the mechanical linkages allow for force feedback to the user.

The forces and torques that surgeons generate during procedures have been recorded using strain gauges. These are small electronic sensors that modify their voltage output based on the mechanical strain applied to them. By combining and calibrating several strain gauges, one can design a sensor capable of measuring the forces and torques in multiple axes. Most force and torque sensors are mounted on the shaft of the laparoscopic instrument to capture the dynamics between the surgeon's hand and the tool tip [13–15]. The majority of these sensors are 6-DOF force and torque sensors from ATI Industrial (www.ati-ia.com). In other approaches, force and torque sensors are placed underneath the tissue that is being operated on [16]. This configuration is easier to construct as the laparoscopic instruments do not need to be modified. Such a configuration is, however, impossible to use in the OR as sensors would have to be implanted in the patient. The data recorded from such sensors is a combination of both the left and right instruments, which makes it more difficult to analyze.

From the recorded data, movement measures are computed and their relation to skill is analyzed. Measures of movement quantity tend to be correlated negatively with surgical skill. For example, the

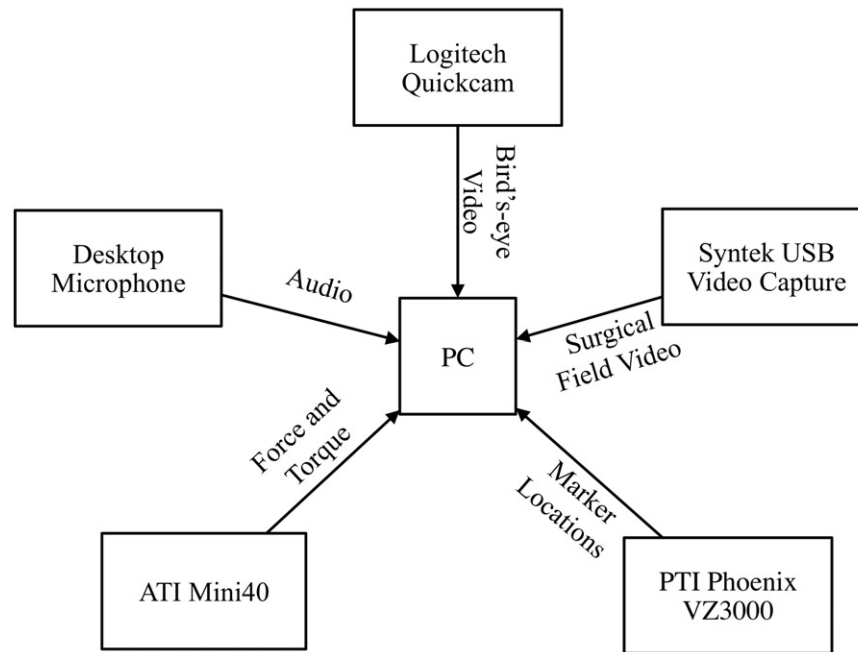


Figure 1. Overview of system components.

total time taken to complete a task is often the most reliable discriminator of surgical skill. Intuitively, the total time to completion is higher in novices, as they make more errors, perform more inefficient movements, and are generally more hesitant than experts. Total time has been shown to correlate negatively with expertise in virtual reality, box trainers, and in the OR [8, 17]. The total path length of the instruments' motion has also been used in surgical skill quantification. This measure is highly correlated with total time, as a longer completion time tends to involve longer movements of the instrument. Similar measures, such as the number of movements, or the amount of movement along the main axis of the shaft, have also been found to correlate negatively with surgical skill [8, 18].

Measures of movement quality have also been analyzed, but do not tend to correlate with surgical skill as strongly as measures of movement quantity. Some studies have found that experts tend to have smoother motions than novices [18, 19]. This could be caused by novices making hesitant movements, having shaky hands, or by a number of other factors. Another formulation of motion smoothness is defined as the "number of changes in velocity" [20], and has also been found to correlate with surgical skill. Dubrowski et al. [16] found that experts apply significantly higher average forces than novices. A higher average force may be explained by the fact that experts make contact with the tissue and thus apply force for a greater proportion of the time, as it has been shown that novices spend more

time in an idle state [21]. Trejos et al. found that novices apply a higher average force, but this claim was substantiated only by a plot of a single stitch from a novice and an expert [22]. These findings are in contradiction to the study by Hwang et al. that found no significant difference between experts and novices in the mean force recorded from a 3-axis sensor mounted in line with the surgical instrument [11]. Attempts to discriminate skill levels using global measures of force have met with little success.

Materials and methods

Recording system

The capture system (Figure 1) records data from an active motion capture system, force and torque sensors, video streams, and an audio microphone. The central component of the system consists of a Windows XP PC with 4 GB of RAM and a quad-core CPU. A multi-threaded C++ program interfaces with all the devices and sensors. A separate thread is created to record each video stream, the motion data, the force data and audio. This allows the large quantity of data, to be processed efficiently on the multi-core machine without introducing significant lag in the system.

Synchronization of the data across the threads is achieved through the use of the Windows system time as a global clock. This is a system-level clock available to all threads and processes. The clock has

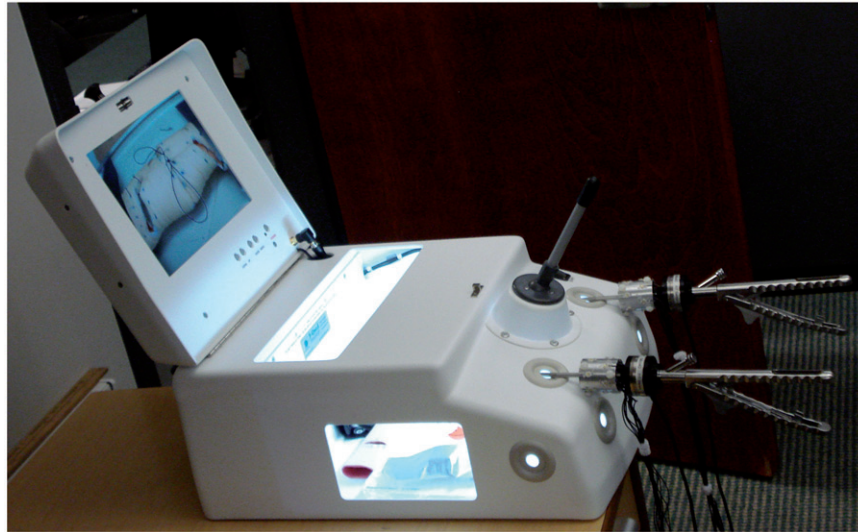


Figure 2. Endo-trainer from 3D-Med used in the studies.

a resolution of 15 ms, which is not sufficient for directly time-stamping all the force measurements, but is sufficiently fine for interpolation of the values for intermediate samples. As the recorded physical movements are relatively slow (a few centimeters per second), this synchronization method is sufficient for an accurate reconstruction of movement.

The configuration of the system is such that it can be easily adapted to nearly any task by using the instrumented needle drivers and deploying the motion capture system in the surgical environment. In the evaluation of our system, an Endo-trainer from 3D-Med (www.3-dmed.com) (Figure 2) was used as the surgical environment. This trainer simulates a laparoscopic surgery environment with a small movable camera to simulate a laparoscope, and rubber holes that simulate the trocars used in real laparoscopy. Synthetic tissue or other items are placed inside the Endo-trainer to be manipulated by the surgeon using laparoscopic instruments. This system is typically used for training residents in laparoscopic skills, as it provides a low-cost, low-risk environment for practicing tool manipulation and suturing techniques. The only modification to the Endo-trainer was the placement of a video splitter on the camera output to allow for video recording.

Motion tracking. A Visualeyez II VZ3000 optical motion tracker from PTI Phoenix (www.ptiphoenix.com) captures the surgeon's movements. The system uses three CCD cameras arranged linearly on a tripod to locate the position of each of the infrared markers. The VZ3000 is able to uniquely identify up to 64 points at 20 Hz by sequentially flashing each marker so that the cameras only

capture a single marker in each frame. The markers are tracked with a precision of 0.7 mm RMSE. The tracking volume is defined by a horizontal and vertical angle of 45° , extending out approximately 7 m. Though the tracking volume can be increased through the use of multiple camera units, a single unit is preferred for portability to an OR environment where space is limited. The VZ3000 software provides access to the motion data through the C++ API provided by PTI Phoenix. After each data frame is retrieved through the API, the marker positions are written to a flat text file along with the system time-stamp.

The positions and orientations of the needle drivers are computed from the locations of the markers fixed to each tool. Twelve markers are attached to an aluminum collar (32 mm in diameter, 48 mm in height) (Figure 3) in two equally spaced rings. Each of the two rings contains six markers, spaced 11 mm apart; the rings are 16 mm apart. The collar is attached to the shaft of the needle driver using plastic screws. This marker configuration allows the instrument to be reliably tracked as it is moved through the tracking volume, as only three markers need to be visible in a given frame to accurately determine the tool tip location using the template-matching algorithm described below.

Force and torque data. The force and torque applied to the needle driver are recorded using the Mini40 sensor from ATI Industrial. This sensor contains six strain gauges that respond to the force and torque applied to the sensor. The sensor is mounted in line with the shaft of the needle driver to

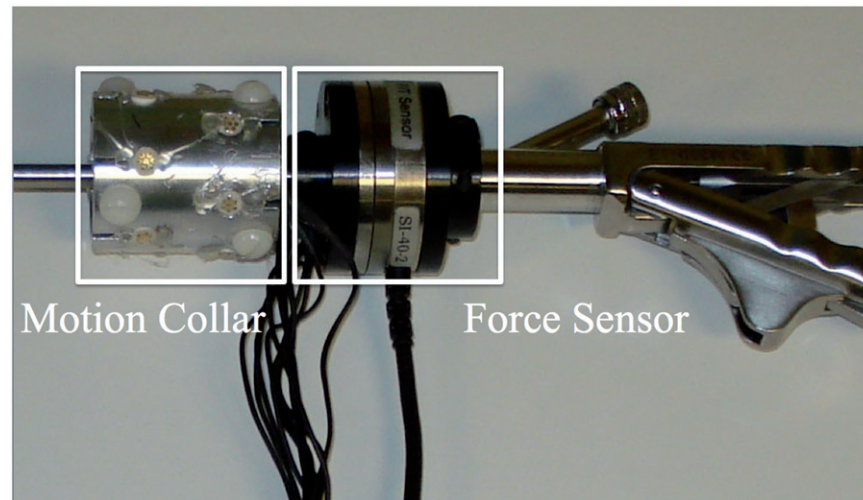


Figure 3. Aluminum collar with markers and force sensor mounted on needle driver.

capture the forces and torques between the surgeon's hand and the tissue being manipulated. The sensor provides accurate force sensing in the range of ± 35 N in the x and y axes, and ± 106 N in the z axis, with less than 0.3 N error in all axes. The torque is calibrated to a range of ± 1.5 Nm in all axes, with less than 0.008 Nm error in all axes. The z axis is aligned with the instrument shaft; the x and y axes are perpendicular to this axis and to each other.

The data from the Mini40 is logged using the NI-PCI 6224 data acquisition card from National Instruments (www.ni.com), at a rate of 1000 Hz. Using the NI-DAQmx C interface provided, the data is read into the CPU 100 samples at a time. To provide accurate synchronization, the force recordings are time-stamped using a linear interpolation algorithm. While a higher-order model could be used to interpolate the time-stamps, the linear model was sufficient to reconstruct and accurately synchronize the movements.

Video and audio recording. Two separate video streams are captured to provide a reference for high-level analysis. Videos are captured using the OpenCV library (opencv.willowgarage.com), encoded using the DivX encoder (www.divx.com). In addition, a flat text file is created that stores the time-stamp for each frame to provide a means to synchronize the video with the motion and force data. Both video streams are recorded and processed on the same thread in the CPU.

One video stream is recorded using a standard web camera, the Logitech QuickCam Orbit MP (www.logitech.com). This camera records video at 30 frames per second with a resolution of

320×240 pixels. The camera focuses on the surgeon's upper body and hands. This information is useful for the expert surgeon in order to evaluate the performance of the trainee surgeon, and can also be used as a reference when comparing the movements to the recorded trajectories. The second video stream records the view from the laparoscopic camera located in the Endo-trainer. This is accomplished by splitting the signal from the laparoscopic camera and routing it to a Syntek STK 1135 USB capture card (www.stk.com.tw). The use of the splitter allows the video to be captured by the PC while still being displayed on the screen of the Endo-trainer. The USB capture card generates video at a resolution of 640×480 pixels at 30 frames per second.

Audio of the recording events is captured using a standard desktop microphone and the LiveInCode program (www.liveincode.rm.pp.ru). The audio is not significant to the automated analysis of the surgical performance, but is captured to provide context for the recorded data.

Data processing

Various data processing steps are required before the recorded movements can be used to assess surgical skill. First, occlusions of the instrument collar need to be compensated for. Once all the missing data are removed, the motion of the instruments can be reconstructed. To locate the instrument in each frame, a template of the instrument is built prior to the surgical performance. This template associates the location of the markers on the collar with the tool-tip position. Next, the coordinate systems of each system need to

```

for  $i = ws/2$  to  $N - ws/2$  do
   $window = p(i - ws/2 : i + ws/2)[1 : 3]$ 
   $error \leftarrow 0$ 
  for  $k = 1$  to  $3$  do
     $s \leftarrow \text{sort}(p[i - ws/2 : i + ws/2][k])$ 
    if  $!(s[\lfloor ws/2 \rfloor - 1] < (p(i)[k] < s[\lceil ws/2 \rceil + 1]))$  then
       $error \leftarrow error + 1$ 
    end if
  end for
  if  $error == 2$  then
     $p(i)[1 : 3] \leftarrow NULL$ 
  end if
end for

```

Figure 4. Median discard filter

be unified, so that the position measurements and force and torque measurements can be analyzed in the same coordinate system.

Processing incomplete data. Missing data from the optical tracker stems from two sources. First, large spikes in position occasionally appear in the data when the wired connection to the active markers is temporarily disrupted. These errors manifest themselves as large changes in the reported position of the marker, sustained over one to five frames. To remove this erroneous data, a modified version of a median filter with a window size (ws) of 19 samples is used. In contrast to a standard median filter, the median discard operation does not replace the target sample with the median of the surrounding samples; it simply discards the sample if it is not near the median of the window for all three axes. Pseudocode for this operation is found in Figure 4.

The second type of incomplete data stems from occasional occlusions that occur during the surgical performance. In these cases, less than three markers are visible, three being needed to compute the proper correspondence to determine the tool-tip position, and the frames are therefore marked as having no data. To generate a continuous signal for analysis (i.e., one without missing data), data for missing positions were imputed using linear interpolation between the nearest visible points. For this operation, and other operations using linear interpolation, a second-order model could be used, but considering the slow speed of the movements, linear interpolation has been shown to be sufficient.

Instrument template tracking. Prior to recording surgical performances, a template is created that relates the positions of the markers on the collar to

the position of the tool tip. The resulting template has a total of 13 positions: 12 for the collar and one for the tool tip. In order to track the tool tip relative to the collar when building the template, an additional marker is required on the tip. The instrument is then moved around the tracking volume and rotated to ensure that all markers are made visible to the camera tracking unit. During this process, all position data are recorded for off-line processing.

To construct the template from the recorded data, the template is first initialized with the frame that has the most visible markers. Then, additional markers are added from other frames by registering corresponding points in the template with those in the frame. This registration is similar to an Iterative Closest Point approach [23], but since the motion capture system identifies each marker uniquely, the correspondences are known, which makes the registration problem simpler. To register each frame, our method finds a rotation \mathbf{R} and translation \mathbf{t} that minimizes the error function ϵ . Each of the N markers in the frame or template has a position (\mathbf{fp}_i or \mathbf{tp}_i):

$$\epsilon = \frac{1}{N} \sum_{i=1}^N \|\mathbf{tp}_i - \mathbf{R} \cdot \mathbf{fp}_i - \mathbf{t}\|^2 \quad (1)$$

The optimal transform (\mathbf{R} , \mathbf{t}) is found using the least squares quaternion method described in [23]. Once the transform is found, it is used to map the marker location into the template space. A pseudocode description of our template-matching algorithm can be seen in Figure 5.

Calibration of coordinate systems. There are three separate coordinate systems that need to be unified in order to provide a complete analysis of the data: those for the tracking system, template, and force

```

Template  $\leftarrow \{\}$ 
Unseen  $\leftarrow AllMarkers$ 
initFrame  $\leftarrow findFrame(max(isVisible).markers)$ 
add(Template, initFrame)
remove(Unseen, initFrame)
while Unseen  $\neq \{\}$  do
  f = findFrame(max(inTemplate) && unseen > 0)
  [R, t]  $\leftarrow findCorrespondence(Template, f)$ 
  for marker in f do
    if isUnseen(marker) then
      push(Template, R marker.p + t)
      remove(Unseen, marker)
    end if
  end for
end while

```

Figure 5. Template construction algorithm.

sensor. The tracking coordinate system is the native reference frame of the motion capture system; it has its origin at the center of the camera unit and measures the 3D position in centimeters. The template coordinate system is similar to the tracking coordinate system, but the location of the origin varies with the data used to construct the template. In this coordinate system, only the relative positions between the markers are important. The force sensor reference frame is centered around the tip of the instrument with forces measured in three dimensions in Newtons (N). The torque around these axes is measured in Newton-meters (Nm).

The common coordinate system is selected as the tracking coordinate system. To unify all the coordinate systems, the template of the instrument is first aligned with the force and torque coordinate system following template construction. To transform the forces, torques and positions from the template space into the tracker coordinate system, a rigid body transform is computed for each frame and applied to the recorded data.

(1) *Alignment of template with force and torque.* To align the force and torque with the template coordinate system, only the orientations of the axes are considered since the measurements use different units. As the force reported in the z axis is already calibrated to respond to force along the instrument's shaft, only the x and y axes need to be aligned. Aligning the two coordinate systems involves translating the tool-tip position to the origin and aligning the z axis with the shaft of the instrument. Alignment of the z axis is achieved by rotating the instrument about the origin such that

the centroid of all of the markers on the collar is in line with the positive z axis, i.e., the x and y components of the centroid are zero.

To complete the alignment, the instrument template needs to be rotated around the z axis to line up the x and y axes. The angle of rotation is found by securing the instrument parallel to the ground, recording the force measurements in this position, and then hanging a weight on the tool tip, as depicted in Figure 6. Subtracting the initial measurement from the measurements while the force is applied gives the direction of the force in the force reference frame (\mathbf{f}_{fr}). The instrument, as well as the position of the weight, is motion captured during this procedure, providing the direction of the force applied in the tracker space (\mathbf{f}_{trk}). The dot product of \mathbf{f}_{trk} and \mathbf{f}_{fr} represents the angular offset between the force sensor and instrument template. By rotating the template around the z axis by the resulting angle, the axes of the instrument coordinate system become aligned with the force coordinate system.

(2) *Alignment of template with tracker.* After the template has been constructed, it can be used to locate the position of the tool tip and the orientation of the instrument in every frame. The application of the template to compute the tool-tip position uses a process similar to the template construction. On every frame, correspondences between the visible instrument markers and the markers in the force-aligned template are found. A rigid body transform $T_{\text{ins} \rightarrow \text{trk}} = [\mathbf{R}_{\text{ins} \rightarrow \text{trk}} \cdot \mathbf{p} + \mathbf{t}_{\text{ins} \rightarrow \text{trk}}]$ that minimizes the distance between these corresponding markers is computed using the same least-squares



Figure 6. Needle driver, string, and weight used during the calibration procedure to align the template and force coordinate systems. Orange ellipses indicate the location of optical markers.

method presented earlier. If fewer than three markers are visible in a given frame, it is marked as having no data and an interpolation operation is used.

To compute the tool-tip position in the tracker coordinate system, the transform $T_{ins \rightarrow trk}$ is applied to the tool-tip position in the force-aligned template. To compute the force in the tracker coordinate system only, $\mathbf{R}_{ins \rightarrow trk}$ is applied to the force and torque recordings. The translation is omitted, as the force and position are aligned only in the orientation of their reference axes, not in the scale or units of the axes. The resulting rotation matrix ($\mathbf{R}_{ins \rightarrow trk}$) can also be used to analyze the orientation of the instrument as well. A sample of the resulting tool-tip trajectory, with the instrument axis overlaid to indicate orientation, is shown in Figure 7.

Position and force sensor fusion

The unification of the instrument kinematics and dynamics in our system allows the computation of the measure of mechanical work. This is a novel measure which is used to estimate the movement economy of the surgeon. Additionally, we provide the computation of several other measures which have been used to analyze surgical skill, as the details of their computation are often left ambiguous. These measures are used as a comparison for the newly proposed mechanical work measure.

Mechanical work. The measure of mechanical work combines the force and position information to estimate the energy used by the surgeon during the tasks. It should be noted that this measure of mechanical work does not represent all energy expended by the surgeon, as only the work measured at the instrument is considered, not the kinetic energy of movement through free space, or other body movements. This measure is meant to capture the movement economy of a surgeon as well as the care with which the tissue is handled, as

peaks in the energy used may damage delicate tissue.

Studies in motor learning have demonstrated that experts in various motor skills show decreased energy use stemming from more economical movement [24]. A study by Sparrow and Newell [25] shows that physiological measures of energy use, such as heart rate and oxygen consumption, tend to decrease as a person becomes more proficient in a complex motor skill. Additionally, Heise and Cornwell [26] demonstrated that mechanical efficiency increased as participants developed skill in a ball-throwing task. Our approach to assessing movement economy was to measure the mechanical work at the tool tip, rather than use somewhat intrusive sensors to measure heart rate and respiration rate to estimate energy consumption.

Mechanical work was computed from the force and tool-tip position value in the tracker coordinate system. For each position measurement $\mathbf{p}(t)$, the corresponding force measurement $\mathbf{F}(t)$ was located by finding the force measurement with the nearest time-stamp to t . Before computation, the position data was smoothed with a Gaussian filter with a standard deviation of $\sigma = 20$ mm to remove measurement noise. The change in energy between sample points was computed as

$$E(t) = (\mathbf{f}(t) - \mathbf{f}(t-1)) \cdot (\mathbf{p}(t) - \mathbf{p}(t-1)). \quad (2)$$

The position and force measurements from a randomly sampled segment are shown in Figure 8, along with the resulting computed work values. A similar process can be applied to the torque and instrument orientation to analyze the work being done to rotate the instrument, such as during needle insertions that involve a strong rotational component around the instrument shaft.

The total mechanical work measured should reflect the efficiency of the surgeon's movements, not only with respect to minimizing the path length, but also to the application of force. As with the other

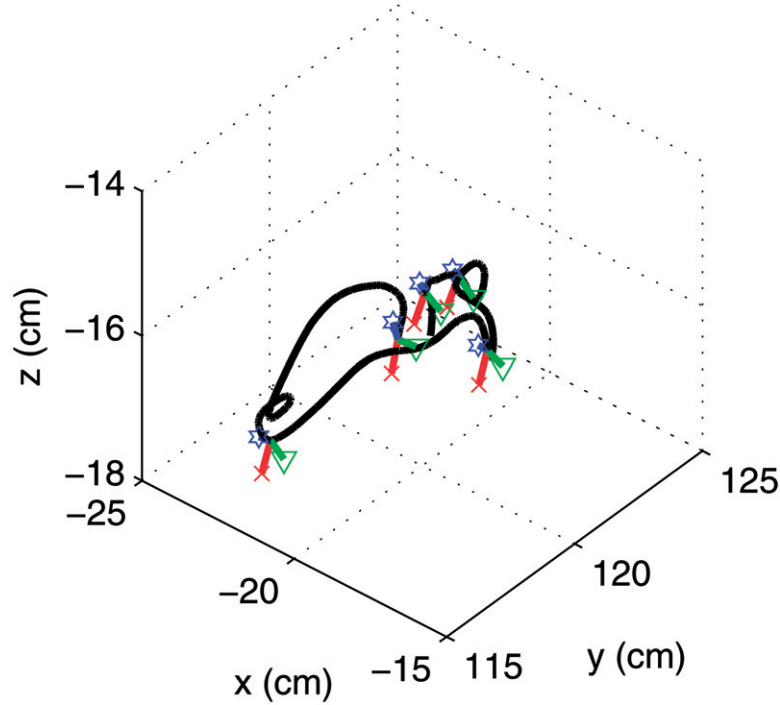


Figure 7. Sample trajectory recorded from the system, with instrument orientation axes overlaid at constant time intervals. The black line represents the 3D trajectory taken by the instrument, while the red, green and blue lines (terminated by an \times , triangle and star, respectively) represent the local x , y , and z axes of the instrument.

quantitative measures, total mechanical work tends to increase as task duration increases. The calculation for this measure is as follows:

$$W = \sum_t^T (\mathbf{f}(t) - \mathbf{f}(t-1)) \cdot (\mathbf{p}(t) - \mathbf{p}(t-1)). \quad (3)$$

Additionally, we examine the use of energy as a movement quality parameter. Similar to previous approaches for force information, we compute the mean and maximum value of the work signal:

$$\overline{W} = \frac{W}{T}, \quad (4)$$

and

$$W_{\max} = \max(W). \quad (5)$$

Global skill indicators

(1) *Total time.* The time required to complete a task has been shown to be a reliable measure of skill. Novices take more time than experts as they are more hesitant, less efficient, and are forced to repeat more movements due to errors. This metric is applicable to nearly all surgical tasks, and can be computed without sophisticated measuring equipment. A drawback of this measure is that it is a very general assessment of skill, giving no feedback on how a surgeon can improve other than to “go faster”.

(2) *Cumulative path length.* The cumulative path length represents the distance that a tool tip travels through space as the task is performed. This measure can reflect the movement efficiency of the surgeon, as shorter paths indicate more economical movements. It can also reflect mistakes that are made, as repeating a number of movements substantially increases the path length. The cumulative path length for each task is independently computed from the positions of each tool tip as the cumulative sum of the Euclidean distances between sample points:

$$l = \sum_t \|\mathbf{p}(t) - \mathbf{p}(t-1)\|. \quad (6)$$

(3) *Global motion smoothness.* Motion smoothness is derived from the curvature of the trajectory over time, which is a 1D signal that describes how the trajectory changes direction in space over time. The curvature value, κ , is used to provide a global measure of motion smoothness and is calculated as the natural logarithm of the median curvature signals for each trajectory. The logarithm function is used to normalize the unbounded values of κ :

$$\kappa = \ln \text{Median} \left(\frac{\left\| \frac{\partial \mathbf{p}(s)}{\partial s} \times \frac{\partial^2 \mathbf{p}(s)}{\partial s^2} \right\|}{\left\| \frac{\partial \mathbf{p}(s)}{\partial s} \right\|^3} \right). \quad (7)$$

Since the trajectory is initially parameterized by time, a stationary tool tip will have a first derivative

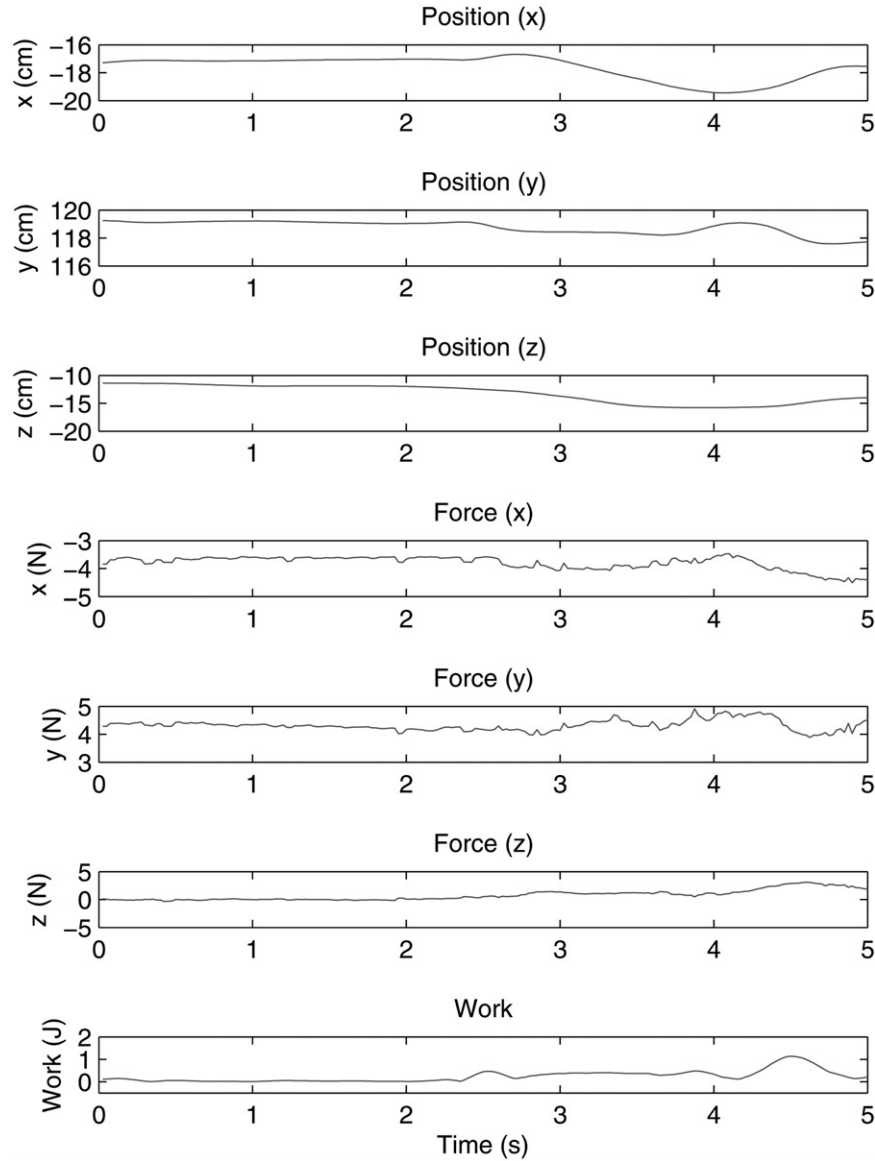


Figure 8. Sampled position, force, and resulting work signal from a random segment of a recorded procedure.

of zero, resulting in a division by zero operation in the curvature calculation. To avoid this problem, the path can be parameterized by the cumulative arc length, s , allowing trajectories of identical paths to be represented identically, even if executed at different speeds. This new representation of the trajectory, $p(s)$, can be calculated with the following mapping from time to cumulative arc length:

$$s = \sum_0^t \sqrt{\left(\frac{dx(t)}{dt}\right)^2 + \left(\frac{dy(t)}{dt}\right)^2 + \left(\frac{dz(t)}{dt}\right)^2} dt. \quad (8)$$

This parameterization produces a non-uniform sampling in the arc-length domain, so uniform sampling is required for the derivative

calculations. The resulting arc-length domain signals are then re-sampled at a uniform interval of 5 mm using linear interpolation.

The derivatives of the instrument trajectory are calculated and low-pass filtered in a single operation. The filtered derivatives of the trajectories $\frac{\partial p(s)}{\partial s}$ and $\frac{\partial^2 p(s)}{\partial s^2}$ are calculated by convolving each of the dimensions (x , y and z) with the first and second derivatives (respectively) of a Gaussian kernel with a standard deviation of $\sigma = 45$ mm, as shown in Equations (9) and (10):

$$\frac{\partial p(s)}{\partial s} = p(s) * x \cdot e^{-\frac{s^2}{2\sigma^2}}, \quad (9)$$

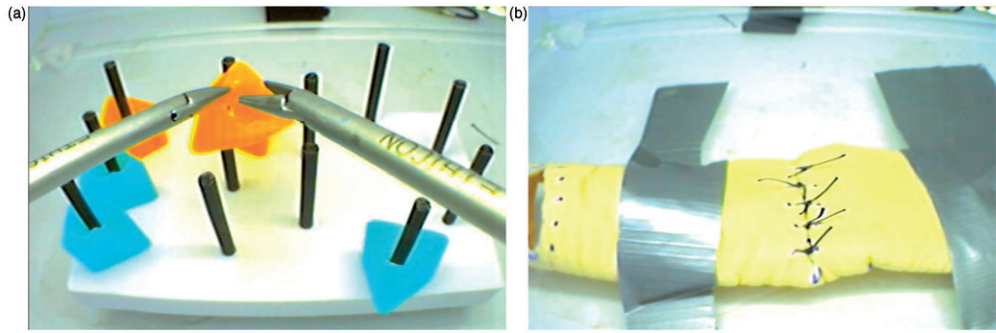


Figure 9. View from the laparoscopic camera for (a) the FLS pegboard task and (b) the simple interrupted suturing task.

$$\frac{\partial^2 \mathbf{p}(s)}{\partial s^2} = \mathbf{p}(s) * \frac{\mathbf{p}(s)^2 - \sigma^2}{\sigma^4} \cdot e^{\frac{-\mathbf{p}(s)^2}{2\sigma^2}}. \quad (10)$$

(4) *Global force features.* The mean and peak values of the applied force have previously been used in the surgical skill evaluation analysis. It is believed that the mean and peak force reflect the delicacy with which the tissue is being handled. However, this can be confounded by novices applying higher forces then spending time in the idle state, thinking about their next movement. In our analysis, the mean and peak force are computed as the mean and maximum values of the magnitude of the force. Computation is identical to the method used to compute mean and maximum of the work signal in Equations (4) and (5).

Experimental design

An empirical study was conducted to validate the reliability and effectiveness of the system. Ten subjects participated in the experiment, which consisted of a series of laparoscopic tasks traditionally used in laparoscopic surgery. All subjects were right-handed males, and all received monetary compensation for their participation. Participants represented a wide range of skill levels, including one staff surgeon, two surgical fellows, and seven general surgery residents from all five years of the residency program.

The first task performed was the pegboard task from the Fundamental Laparoscopic Skills (FLS) program (Figure 9a). This task consists of moving a number of colored rubber collars from one side of the board to the other using laparoscopic instruments. This is a standardized task used to evaluate general laparoscopic dexterity, and all participants were familiar with this task.

Task 2 required participants to place a series of five simple interrupted sutures in a piece of synthetic bowel tissue. Participants were asked to close

a small hole that had been cut in the bowel by inserting five simple interrupted sutures at marked points around the hole. The marked points were 5 mm from the edge of the hole and 1 cm away from the neighboring holes. A template was used to mark the points and the hole. All participants performed one warm-up suture before completing the five test sutures. Figure 9b shows the synthetic tissue with the markings before and after the simple interrupted sutures were inserted. This task is more difficult than the FLS pegboard, and all subjects were familiar with this task.

Finally, Task 3 required participants to perform a continuous running suture. The continuous running suture is used to close a hole using a single thread and is quite difficult for a number of reasons. The longer thread (30 cm instead of 15 cm) is more difficult to manipulate laparoscopically, and participants must also plan their stitch more carefully, ensuring that there is enough thread left at the end to complete the final knot. This was the most difficult task performed by participants, and some of the novice participants reviewed a video of the task being performed prior to beginning the trial.

Results

Correlations were calculated between the measures and the expert-assessed ranking of the participant. An expert surgeon (D.W.B.) with many years of OR and teaching experience assessed the performance of the participants following the experiments. The ranking was based solely on the videos, and no information that could identify the participants was shown to the expert. A positive but not significant correlation was found to exist between the expert's rank and the participant's training level ($\rho = 0.23$, $p = 0.44$). Correlations between expert ranking and skill measure were computed using Spearman's rank correlation coefficient for non-parametric data.

Table I. Significance of correlations between each measure and each task. (1: fundamental laparoscopic skills; 2: simple interrupted suture; 3: continuous running suture) for non-dominant (N) and dominant hand (D). * indicates significant correlation ($p < 0.05$); – indicates non-significant or no correlation.

	1N	1D	2N	2D	3N	3D
Total time	*	*	*	*	*	*
Path length	–	*	*	*	–	*
Total energy	*	–	*	–	–	–
Curvature	–	–	–	–	–	–
F/T mean, peak	–	–	–	–	–	–
Work mean, peak	–	–	–	–	–	–

A summary of the significance of the various measures can be found in Table I.

Completion time correlated significantly with expert-assessed skill ($p < 0.05$) on all three tasks. The strongest correlation was with the simple interrupted sutures ($\rho = -0.84$, $p = 0.01$), the weakest with the FLS tasks ($\rho = -0.74$, $p = 0.02$). These results are consistent with other studies of surgical skill evaluation. The strong correlation suggests that the expert evaluations of the participants are accurate, and that skill on the simple interrupted suture task is indicative of skill on the other tasks.

A significant correlation was found between cumulative path length of the instrument in the dominant hand and expert-assessed skill for all tasks ($p < 0.05$). The path length of the instrument in the non-dominant hand was found to be significant only for the simple interrupted suturing task, not for the FLS pegboard or the continuous running suture. This is consistent with other studies (e.g., that by Cristancho et al. [6]) that have shown that measures of skill computed from the dominant hand are generally more discriminatory. The strongest correlation was found in the path length of the instrument in the dominant hand during the simple interrupted sutures ($\rho = -0.84$, $p < 0.01$).

Contrary to most other measures in the literature and those analyzed in this study, total mechanical work had a greater correlation with skill when analyzing the non-dominant hand, rather than the dominant hand. Both the FLS task and the simple interrupted sutures showed a significant negative correlation ($\rho = -0.79$, -0.68 , $p < 0.05$) between expert-assessed skill and total mechanical work for the non-dominant hand. A negative correlation was also found when analyzing the continuous running suture, as well as the dominant hand on all tasks, but these correlations were not significant. Total mechanical work may not have the same discriminatory power as the other quantitative measures, but it appears to have some unique information not

captured by the path length measure. Further tests with a larger sample size may reveal a stronger relationship.

No correlation was found between expert-assessed skill and the calculated motion smoothness for either hand. These findings are contrary to prior studies that did find a relation between assessed skill and motion smoothness from curvature in the dominant hand for some tasks [19]. Both the mean and maximum of the magnitude of the force and torque values are computed. No correlation was found between expert-assessed skill and the mean or peak force for each hand. Similar results are found in the torque values. The forces applied by the participant can vary widely depending on the particular movements being executed: low forces may be applied when manipulating tissue and high forces when tightening a suture. This variation makes it difficult to extract a reliable global measure that accurately reflects surgical skill. The results demonstrate the difficulty in extracting global measures from the complex force data.

To analyze the applied energy as a global measure of skill quality, the mean and peak of the energy signal were computed. Similar to the motion smoothness and force measurement, no correlation was found between expert-assessed skill and the mean and peak of the energy signals. This does not indicate that the measures of motion quality are of no use, but it is likely that the small sample used in the study was not sufficient to validate these skill measures. Further studies with a large sample including more expert surgeons would be beneficial in assessing the utility of these methods.

Discussion

The presented system offers several advantages over existing systems that are able to record surgical movements. The design of the system minimizes interference with the surgical performance while providing reliable, accurate measurements. In addition, the fusion of force and position measurements allows for new measures that show strong potential in assessing surgical skill. Below we outline some of the successful features of the system, including those elements that are generalizable to other surgical recording systems.

One successful design aspect of our system was the use of a single PC for recording all of the data. With this configuration, all measurements are easily synchronized with a common clock and the set-up and configuration time for a single experiment was minimal. Multi-PC set-ups were considered to reduce the load on a single machine, but the

additional software (and hardware) complexity was not worth the minor improvement in performance. As everything is on a single machine, difficulties in communication over a network, compensation for network lag, and inter-process communication can be avoided, which simplifies the development. The system is capable of effectively combining the recorded data, specifically the kinetic and kinematic information. Calibration procedures for unifying several coordinate systems (force/torque, tracker and template) proved successful in reconstructing the complete movement of the surgeon. These procedures can be easily generalized and applied in other systems to unify the recordings from separate sensors.

In contrast to systems that record movements using rigid mechanical links, our system is portable and can be adapted to numerous situations. The use of optical tracking and the instrument template matching method allows a wide range of instruments to be tracked, not just the instruments that are fixed to the system. Additionally, the surgeon has more freedom to move the instruments naturally, as they are not constrained to a pre-defined workspace. This also enables the system to record in different scenarios with little set-up time, as there is only a single tripod to mount in the workspace and a short calibration procedure to perform. The use of the video splitter ensures that the surgeon's primary task is not interfered with, and that the task can be completely captured. These features are crucial in developing a robust system for recording *in-vivo* measurements.

The use of the aluminum collar provides several benefits over other tracking techniques such as attaching a three-marker rigid body to the instrument. First, the collar is lightweight, and the weight is evenly distributed around the instrument's shaft. This minimizes any interference by the recording system with the surgeon's performance. Second, the distribution of markers within the collar is quite compact, ensuring that markers will be visible to the recording system regardless of the tool's orientation. Finally, the aluminum collar is easily attached and detached from the instrument using plastic screws, allowing it to be readily used on any number of laparoscopic instruments.

The template matching procedure and dense placement of tracking markers proved to be an effective technique for tracking the instrument. This approach allowed us to track the instrument from cameras clustered around a single position, which is an essential feature for developing systems for ORs. In the OR, space is limited and people often surround the patient, with the only available viewpoint being above the patient. This template

matching technique can easily be generalized to other systems, as well as to the tracking of other objects. While not presented in this article, we have successfully used this technique to track a user's wrist by placing several markers around the wrist and building a template on the fly.

In the early development stages of our system, an EM-based tracking system (Polhemus Ultratrak) was used; however, the measurements were found to be too noisy to be useful during simple testing. In contrast, our final system is robust to EM interference and noise, which increases the effective tracking range and the accuracy of the measurements. In a typical OR there is a substantial number of electronic devices and ferrous objects that distort the field used by EM systems. In particular, the electro-cautery tools in use today emit an incredible amount of EM radiation that will severely impact EM-based systems. In contrast, our optical system is robust to EM interference and provides reliable, precise measurements in a much greater workspace than EM-based systems allow.

The effectiveness and usefulness of the system is demonstrated in the empirical study. The measurements provided by the motion capture equipment, as well as by the force and torque sensors, are accurate and reliable enough to be used to analyze skill. Established measures of motion quantity, such as total time and total path length of the dominant hand, were found to correlate negatively with surgical skill for all of the tasks described in the Materials and methods section.

From the fusion of the instrument position and the recorded force, we computed a measure of total energy. This is a measure of movement quantity, and was found to correlate negatively with surgical skill for the non-dominant hand. The correlation is stronger than with motion quality measures, indicating that the measure is fairly robust. However, this is a weaker correlation than the path length and total time, indicating that it may not be as reliable. The correlation of total energy with surgical skill is stronger when analyzing the non-dominant hand, which indicates that the measure is complimentary to existing measures that tend to correlate more strongly on the dominant hand. From conversations with surgeons, some believe that the non-dominant hand begins to take on more and different functions as the skill of the surgeon progresses, which could lead to different patterns of "work".

The measure of work requires both position and force sensors to be used simultaneously, which can be seen as a disadvantage. While this is true, the most widely used application of these measures is in virtual surgical trainers. The majority of such

training systems already have instruments outfitted with these sensors, and thus no additional overhead is required. It is also foreseeable that sensors will become smaller and less intrusive in the future. In recent years, there has been significant progress made in passive and markerless motion capture, which leaves the force sensor as the only significant burden on the surgeon. In many systems, we foresee the utility of the energy measure outweighing any costs associated with collecting the separate data streams.

All measures of motion quality did not correlate with surgical skill. This is not a surprising finding as measures of motion quality tend to be correlated with surgical skill less reliably than measures of motion quantity, and the study was conducted with a small number of participants. However, this does not mean measures of motion quality are of no use: it is possible that they provide unique information that can help discriminate between two subjects that otherwise look similar. For instance, if a novice performs a fast, low-quality stitch, it may be indistinguishable from an expert's fast, high-quality stitch in terms of time or path length. However, the two stitches will likely have very different force and smoothness patterns. Additionally, motion quality measures may help distinguish between similar levels of skill, for instance, among different experts, but our limited data set prevents this analysis.

Conclusions

We have developed a system for recording the movement of a laparoscopic surgeon and the surgical instruments, and the forces and torques applied to the instruments. This system has the potential to be implemented in an OR setting, providing a rich, accurate data set for use in analyzing surgical movements. We have described the system in detail, allowing parts or the whole design to be recreated in another setting, thereby facilitating future research in the area. An empirical validation was performed with the system, demonstrating its effectiveness and the utility of the fusion of force and position information. The measure of total mechanical work is complementary to existing measures, and can be used to improve the accuracy of the assessment of surgical performance. This system and measure can be used to better understand surgical dexterity, conduct further research in the area, and develop new methods of evaluating the surgical technique.

While the developed system and measures were able to discriminate skill at a coarse level, we believe

that to fully analyze complex surgical tasks, sensor-based data is not sufficient and must be linked to high-level knowledge. The link between behavioral data and sensor data in complex human activities is rarely straightforward. There are many sources of behavioral data, ranging from verbal and nonverbal communication between a surgeon and his team to high-level identification of the gesture by highly skilled surgeons. The synchronized visual and audio recording with the tool tracking information will become critical for this analysis and will be the subject of our next investigation.

We are also currently exploring the deployment of this system in an OR environment. Challenges in this environment include more complex variable tasks, as well as logistical issues in the deployment. One of the major challenges involves tracking the instruments without impacting the performance of the surgeon. To that end, passive motion capture markers (that is, reflective markers that do not have wires) and a smaller force and torque sensor will greatly improve the usability of the system. Sterility may be an issue, but many of the smaller components (e.g., the instruments and sensors) can be sterilized using gas sterilization. One challenge that remains without a clear solution is mitigating the impact of occlusion in crowded OR environments, which is currently one of the largest problems in capturing accurate, robust data from *in-vivo* operations.

Declaration of interest: Fraser Anderson is supported by the Canadian Institutes of Health Research; Pierre Boulanger and Walter Bischof are supported by the Natural Science and Engineering Research Council; and Daniel Birch is supported by Alberta Health Services and the University of Alberta.

References

1. Dunkin B, Adrales GL, Apelgren K, Mellinger JD. Surgical simulation: a current review. *Surg Endosc* 2007; 21(3):357–366.
2. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, Brown M. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997;84(2):273–278.
3. Reiley CE, Lin HC, Yuh DD, Hager GD. Review of methods for objective surgical skill evaluation. *Surg Endosc* 2011;25(2):356–366.
4. Feng C, Haniffa H, Rozenblit J, Peng J, Hamilton A, Salkini M. Surgical training and performance assessment using a motion tracking system. *Proceedings of the 2nd European Modeling and Simulation Symposium, Barcelona, Spain, October 2006*. pp 647–652.

5. Dubois P, Thommen Q, Jambon AC. In vivo measurement of surgical gestures. *IEEE Trans Biomed Eng* 2002;49(1):49–54.
6. Cristancho SM, Hodgson AJ, Panton ONM, Meneghetti A, Warnock G, Qayumi K. Intraoperative monitoring of laparoscopic skill development based on quantitative measures. *Surg Endosc* 2009;23(10):2181–2190.
7. Grober ED, Roberts M, Shin EJ, Mahdi M, Bacal V. Intraoperative assessment of technical skills on live patients using economy of hand motion: establishing learning curves of surgical competence. *Am J Surg* 2010;199(1):81–85.
8. Datta V, Mackay S, Mandalia M, Darzi A. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. *J Am Coll Surgeons* 2001;193(5):479–485.
9. Emam TA, Hanna GB, Kimber C, Cuschieri A. Differences between experts and trainees in the motion pattern of the dominant upper limb during intracorporeal endoscopic knotting. *Digest Surg* 2000;17(2):120–123.
10. Emam TA, Hanna GB, Kimber C, Dunkley P, Cuschieri A. Effect of intracorporeal-extracorporeal instrument length ratio on endoscopic task performance and surgeon movements. *Arch Surg Chicago* 2000;135(1):62–65.
11. Hwang H, Lim J, Kinnaird C, Nagy AG, Panton ONM, Hodgson AJ, Qayumi KA. Correlating motor performance with surgical error in laparoscopic cholecystectomy. *Surg Endosc* 2006;20(4):651–655.
12. Rosen J, Brown JD, Chang L, Barreca M, Sinanan M, Hannaford B. The BlueDRAGON – A system for measuring the kinematics and the dynamics of minimally invasive surgical tools in-vivo. In: *Proceedings of the IEEE International Conference on Robotics and Automation*, Washington, DC, USA, May 2002. pp 1876–1881.
13. Lamata de la Orden P, Gomez Aguilera EJ, Lamata Hernandez F, Oltra Pastor A, Sanchez Margallo FM, del Pozo Guerrero F. Understanding perceptual boundaries in laparoscopic surgery. *IEEE Trans Biomed Eng* 2008;55(3):866–873.
14. Richards C, Rosen J, Hannaford B, Pellegrini C, Sinanan M. Skills evaluation in minimally invasive surgery using force/torque signatures. *Surg Endosc* 2000;14(9):791–798.
15. Gunther S, Rosen J, Hannaford B, Sinanan M. The red DRAGON: A multi-modality system for simulation and training in minimally invasive surgery. *Stud Health Technol Inform* 2007;125:149–154.
16. Dubrowski A, Sidhu R, Park J, Carnahan H. Quantification of motion characteristics and forces applied to tissues during suturing. *Am J Surg* 2005;190(1):131–136.
17. Cotin S, Stylopoulos N, Ottensmeyer M, Neumann P, Rattner D, Dawson S. Metrics for laparoscopic skills trainers: The weakest link! In: Dohi T, Kikinis R, editors. *Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2002)*, Tokyo, Japan, September 2002. Part I. *Lecture Notes in Computer Science* 2488. Berlin: Springer; 2002. pp 35–43.
18. Chmarra MK, Kolkman W, Jansen FW, Grimbergen CA, Dankelman J. The influence of experience and camera holding on laparoscopic instrument movements measured with the TrEndo tracking system. *Surg Endosc* 2007;21(11):2069–2075.
19. Judkins TN, Oleynikov D, Stergiou N. Objective evaluation of expert and novice performance during robotic surgical training tasks. *Surg Endosc* 2009;23(3):590–597.
20. Oostema JA, Abdel MP, Gould JC. Time-efficient laparoscopic skills assessment using an augmented-reality simulator. *Surg Endosc* 2008;22(12):2621–2624.
21. Rosen J, Solazzo M, Hannaford B, Sinanan M. Objective laparoscopic skills assessments of surgical residents using hidden markov models based on haptic information and tool/tissue interactions. *Stud Health Technol Inform* 2001; 81:417–423.
22. Trejos AL, Patel RV, Naish MD, Schlachta CM. Design of a sensorized instrument for skills assessment and training in minimally invasive surgery. In: *Proceedings of the 2nd IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob 2008)*, London, Ontario, Canada, October 2008. pp 965–970.
23. Besl PJ, McKay ND. A method for registration of 3-D shapes. *IEEE Trans Pattern Anal Machine Intel* 1992;14(2):239–256.
24. Magill RA. *Motor learning and control: Concepts and applications*. Boston, MA: McGraw-Hill; 2001.
25. Sparrow WA, Newell KM. Metabolic energy expenditure and the regulation of movement economy. *Psychon Bull Rev* 1998;5(2):173–196.
26. Heise GD, Cornwell A. Relative contributions to the net joint moment for a planar multijoint throwing skill: Early and late in practice. *Res Q Exerc Sport* 1997;68(2): 116–124.