

HMM Assessment of Quality of Movement Trajectory in Laparoscopic Surgery

Julian J.H. Leong, Marios Nicolaou, Louis Atallah, George P. Mylonas,
Ara W. Darzi, and Guang-Zhong Yang

Royal Society/Wolfson Medical Image Computing Laboratory & Department of Biosurgery
and Surgical Technology, Imperial College London, London, United Kingdom
{j.leong, m.nicolaou, l.atallah, george.mylonas,
a.darzi, g.z.yang}@imperial.ac.uk

Abstract. Laparoscopic surgery poses many different constraints to the operating surgeon, this has resulted in a slow uptake of advanced laparoscopic procedures. Traditional approaches to the assessment of surgical performance rely on prior classification of a cohort of surgeons' technical skills for validation, which may introduce subjective bias to the outcome. In this study, Hidden Markov Models (HMMs) are used to learn surgical maneuvers from 11 subjects with mixed abilities. By using the leave-one-out method, the HMMs are trained without prior clustering subjects into different skills levels, and the output likelihood indicates the similarity of a particular subject's motion trajectories to the group. The experimental results demonstrate the strength of the method in ranking the quality of trajectories of the subjects, highlighting its value in minimizing the subjective bias in skills assessment for minimally invasive surgery.

1 Introduction

Minimally Invasive Surgery (MIS) was first introduced over 20 years ago [1-2], however, many surgeons are still restricted to performing relatively simple procedures, for example laparoscopic cholecystectomy and diagnostic arthroscopy. The uptake of advanced MIS procedures, such as laparoscopic colectomy for cancer [3], is still very slow in many countries. These procedures are mainly reserved for tertiary referral centers, performed by highly sub-specialized surgeons.

The constraints in the MIS environment have been well documented; the lack of 3D vision, limited haptic feedback and the fulcrum effect restrict the variety of surgery performed. It is, however, quite clear that some surgeons are superior in performing these tasks than others. This has motivated extensive research into the objective assessment of surgical skills. The methodology has now evolved from subjective qualitative assessment by the trainers, knowledge assessment using post-graduate examinations, to objective quantitative approaches using time or movement parameters. Quantitative methods for assessing surgical dexterity have been widely validated for a number of open and laparoscopic procedures [4]. The validation for these methods, however, relies upon prior definition of expertise, and this classification is mostly based on the assumption that experience equals technical excellence.

The purpose of this study is to examine the use of Hidden Markov Models (HMMs) based on view invariant trajectory representations for assessing complex MIS tasks. In order to increase the complexity of hand-eye coordination, view rotation tasks have been introduced in this study. This effect has been studied previously and the ability to handle mental rotation tasks has been suggested to be indicative of the innate ability in mastering laparoscopy [5-6]. We demonstrate that with the proposed method, a probabilistic framework can be formulated to allow the observation of trajectories without prior, arbitrary classification of the subjects' abilities, thus minimizing subjective bias in MIS skills assessment.

2 Modeling Instrument Motion Trajectory

2.1 Instrument Tip Tracking and Calibration

In order to obtain the trajectory of the instrument tips in Euclidean space and not interfere with the experimental task, a tracking device was attached rigidly to the handles of the laparoscopic instruments. For accurate positioning of the tracking device, a Polaris (Northern Digital Inc, Ontario) 6-DOF infrared (IR) tracker was used. The Polaris tracker is able to track a number of passive, active, wired and wireless IR tools in real time simultaneously. Data interfacing was achieved through RS-232 and the provided tracking accuracy is 0.35 mm RMS at a sampling rate of 60Hz. The offset of the instrument tips from the IR markers was calculated using the Pivot function in the NDI ToolViewer Software version 3.02.01.

The accuracy of the instrument tip tracking arrangement was further assessed by mounting the laparoscopic instrument on a Stäubli RX60L robotic arm with repeatability accuracy of +/- 0.02mm and six degrees of freedom. Ten points for calibration were set up using the robotic arm to manipulate the instrument whilst the IR markers were being tracked.

2.2 Subjects and Experimental Setup

Eleven subjects were recruited for the study (9 medical student and 2 practicing surgeons, 2 subjects were left-handed). None of the medical students had previous simulated or real laparoscopic experience. They were then randomized into two groups. All subjects were consented prior to the study, and were given a short introduction of laparoscopic surgery and the instruments used. Both groups were required to perform a laparoscopic task in a box trainer, after familiarization with the instruments and environment. A Karl Storz laparoscopic stack with an S1 camera head, Xenon Nova light source, and Hopkins II 0° endoscope was used for the experiment, along with two laparoscopic graspers with IR tracking devices rigidly attached.

For both groups, the first task was to touch two standardized points (A and B) on a simulated plastic small bowel model as illustrated in Fig. 1. Each point was attached to an in-house designed touch sensitive circuit switch to mark the beginning and end of each trajectory. Each time the circuit was completed, an alarm will indicate a successful contact. The subjects were then asked to touch alternatively the points A and B with the left instrument 10 times, and then the right instrument, and this step was repeated. The subjects were then required to perform a complex task: first, lift a

simulated omental flap using the left instrument to grasp the left corner, and touch a third point (C) hidden underneath the flap with the right. This was repeated 3 times. A total of 36 trajectories were obtained between the two points. For the second task, Group 1 were required to repeat the first task with the laparoscopic camera rotated 90 degrees counter clockwise and Group 2 with the camera rotated 90 degrees clockwise.



Fig. 1. (a) Experiment setup showing the arrangement of the IR sensors in relationship to the laparoscope tools, and the plastic small bowel model with a simulated omental flap with and without the camera view being rotated (b-c)

For the analysis of the trajectories, the experiments were divided into: (1) left instrument motion from point A to B; (2) left instrument motion from point B to A; (3) right instrument motion from point A to B; (4) right instrument motion from point B to A; (5) left instrument motion in the complex task; (6) right instrument motion to touch point C.

2.3 View Invariant Representation of 3D Trajectories

Prior to HMM analysis, the 3D trajectories were mapped to a view invariant representation based on the Centroid Distance Function (CDF) [7]. The instrument tip positions from IR tracking after offset correction were modeled as a parametric curve:

$$r[t] = \{x[t], y[t], z[t]\} \quad (1)$$

$$t = 0 \dots N - 1$$

CDF is a feature that is affine invariant, and is also widely used in image retrieval applications [7]. In essence, CDF describes a time series of the distance of each point in the trajectory from the centroid of the trajectory, and therefore after scale normalization it is rotational and translational invariant. This significantly simplifies the subsequent HMM classification by foregoing cumbersome preprocessing steps. Fig. 2 shows the CDF projection of the instrument trajectories for a surgeon and a novice, respectively.

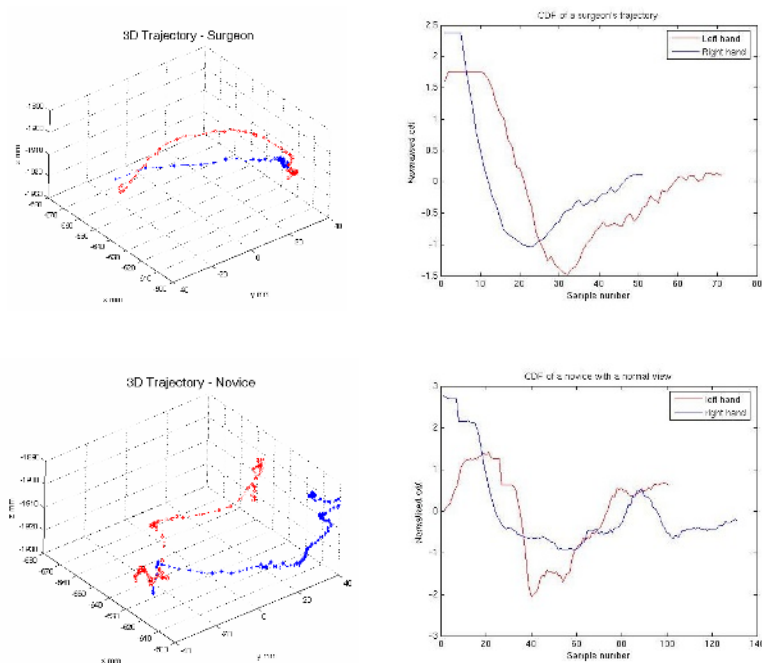


Fig. 2. Left column: 3D trajectories of surgeon and novice, blue shows the right hand and red shows the left. Right column: CDF representation of the same trajectories

2.4 Hidden Markov Modeling

HMMs are finite state stochastic machines that allow dynamic time warping for modeling time series data. HMMs have been used to classify movement trajectories, however, segmentation was necessary to avoid violating the Markovian property which assumes independence of a current state from past states given the previous one [8]. Each trajectory was regarded as one independent action, and the CDF for the trajectory was used as input signals to the HMM, hence the notion was made that each trajectory adhere to the Markovian assumption. In this study, HMM was used to learn each trajectory of a given experiment and view rotation. The leave-one-out method was used to train the HMM from all subjects excluding the test subject's data. The trained HMM was then used to calculate the log likelihood of the test subject and indicate similarities or differences to the learned model.

An HMM can be described by three model parameters representing the relationship between the hidden states (h) and the observed data (x). These parameters are:

$$(\pi_i, a_{ij}, p(x/h)) \quad (2)$$

where π_i is the initial state probability, a_{ij} the transition matrix between the hidden states and $p(x/h)$ the probability of generating an observation given the hidden state.

In this study, a fixed number of states were used. However, in order to have a more flexible model, the observation probabilities were modeled by a Gaussian Mixture Model (GMM). Thus $p(x/h)$ can be defined as:

$$p(x/h) = \sum_{m=1}^M c_m \frac{1}{(2\pi)^{P/2} |\Sigma_m|^{1/2}} \exp\left(-\frac{(x-\mu_m)^T \Sigma_m^{-1} (x-\mu_m)}{2}\right) \quad (3)$$

where c_m is the mixing parameter, μ_m and Σ_m are the mean and covariance matrix of the component m of the GMM.

The K means algorithm was used to initialize the parameters of the observation GMM (mainly the means μ_m). Two versions of the K means algorithm were implemented to compare performance, using Euclidean distance and Derivative Dynamic Time Warping [9] to calculate the similarity between two trajectories. The second option can take signal ‘warping’ into consideration when finding cluster centers, and the length of the ‘warping’ was then used to normalize the distance between the two trajectories. As the resulting cluster centers were identical with both methods, Euclidean distance was used as it was more computational efficient. The Expectation Maximization (EM) algorithm was used to calculate the maximum likelihood of the parameters of the HMM, namely the means and covariances of the components of the GMM and the parameters π_i and a_{ij} .

2.5 Categorized Observational Score

The video of all the tasks were scored by two independent observers who were blinded to the identity of the subjects, the scoring system used was a modified version of the Objective Structured Assessment of Technical Skills (OSATS) global rating scale. This is a widely validated score developed by Martin *et al*, using 8 categories each with a Likert scale of 1-5 anchored by descriptors [10]. The modification was necessary as 3 of the categories did not apply to this particular experiment (suture handling, knowledge of procedure and quality of final product).

3 Results

To demonstrate how the effect of view rotation increases the complexity of the tasks, Fig. 3(a) illustrates the average time increased for a subject to complete a trajectory when the view is rotated. Fig. 3(b) demonstrates the OSATS global rating scale (with good inter-observer agreement) decreases with view rotation, however this does not completely correlate with the time measurement. For example, subject 2 in the rotated task had the third lowest score in OSATS, but had the fifth longest time. OSATS is an assessment of the quality of technique, whereas time is a crude metric for performance. There is a significant correlation between mean (unfiltered) path length and time taken ($r = 0.916$ $p < 0.001$). Interestingly, one subject improved slightly in the rotated task (subject 7).

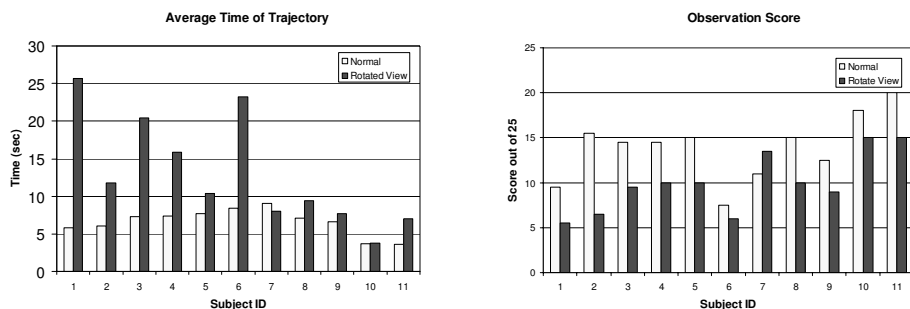


Fig. 3. (a) The average time for the trajectories of each subject. (b) The modified OSATS score for all the subjects involved in the experiments.

CDF representations of motion trajectories are illustrated in Fig. 4(a). This figure shows that the surgeon's trajectory generally lies closer to the mean of all the CDF than the novice's trajectory in this particular experiment. However, the average CDF trajectory has a distinctively different shape to either of the sample trajectories.

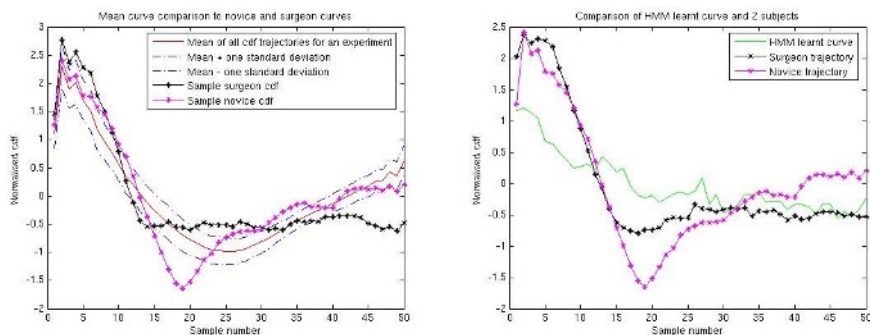


Fig. 4. (a) The mean of CDF trajectory for one of the experiments with standard deviation. An example of a practicing surgeon's trajectory is shown in black and a novice in pink. (b) HMM learnt curve shown in green.

In Fig. 4(b), the HMM learnt curve is plotted instead, this is the average of ten observations after training the HMM. The set of trajectories showed a large variability with effects of time warping. By calculating the likelihood to the HMM, a measure of similarity of a certain trajectory to the whole group can be obtained.

The parameters of the HMM including the number of Gaussians in the observation GMM, as well as the number of the hidden states of the HMM were selected experimentally. Parameters that lead to the least variation in the values of the test data likelihoods were selected as the parameters that can provide good data representation for this dataset. The number of hidden nodes was selected as 4 with a mixture of 2 Gaussians for the GMM.

The log likelihoods of the subjects' trajectories belonging to the training data sets, using the leave-one-out method are illustrated in Figs. 5(a) and (b). Lower values

indicate the subjects' trajectories are more likely to match the models learnt by the HMMs. The log likelihood values are negative. The effect of view rotation accentuates the difference between the subjects as shown in Fig. 5(b).

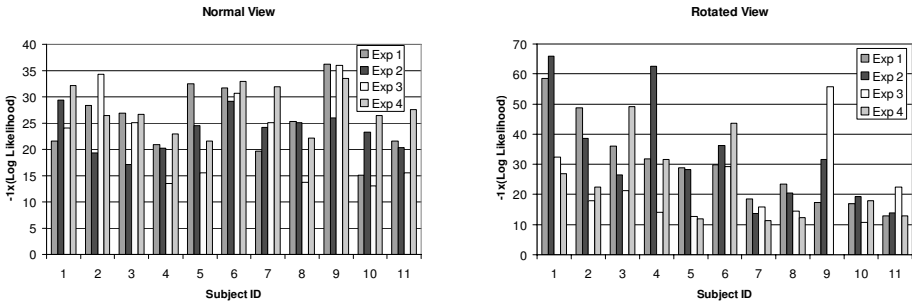


Fig. 5. (a) The negative log likelihood of the subject in each experiment to belong to the group in the normal view orientation and (b) rotated

Finally, the log likelihood values were ranked in each experiment independently, rank number 1 represents the most similar and vice versa. The mean rank of experiment 1-4 for each subject was calculated in both normal and rotated view in Fig 6(a). In order to compare this with the most validated surgical rating scale, this rank is plotted against the ranking in the OSATS score in Fig. 6(b). There was a very significant correlation between them ($r=0.93$ $p < 0.001$).

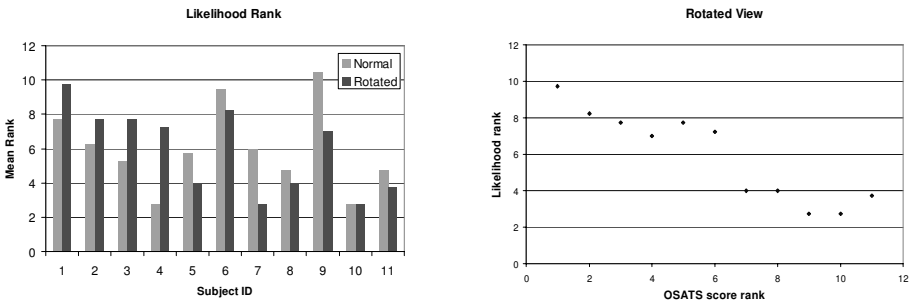


Fig. 6. (a) The mean rank of subjects' likelihood of belonging to the test group, (b) scatter plot of the rank of likelihood generated by the trained HMM against OSATS score ranking in the rotated tasks

4 Discussion and Conclusions

This study has shown that HMM can be used to learn models of surgical motion trajectories even in a group of subjects with mixed abilities without prior classification of technical skills levels. The HMMs calculated the likelihood of the practicing surgeons as the most representative trajectories (subjects 10 and 11 in figure 6), this also

correlated very well with the observation scores, which is a labor intensive method to assess the quality of surgical technique. One of the important considerations of applying the proposed HMM scheme is the feature representation of the trajectories. In general, it should be invariant to affine transformation, as this can cope with trajectories with different starting points, rotations and approaching directions. In this study, we have used CDF as a means of invariant feature representation. Other approaches based on extrema in acceleration measured by high frequency wavelet coefficients are also applicable [11].

References

1. Litynski G.S.: Profiles in laparoscopy: Mouset, Dubois and Perissat: the laparoscopic breakthrough in Europe (1987-1988). *JLS* **3(2)** (1999) 163-167
2. Semm K.: Endoscopic appendectomy. *Endoscopy* **15(2)** (1983) 59-64
3. Harinath G., Shah P.R., Haray P.N., Foster M.E.: Laparoscopic colorectal surgery in Great Britain and Ireland--where are we now? *Colorectal Dis* **7(1)** (2005) 86-89
4. Moorthy K., Munz Y., Sarker S.K., Darzi A.: Objective assessment of technical skills in surgery. *BMJ* **327(7422)** (2003) 1032-1037
5. Jordan J.A., Gallagher A.G., McGuigan J., McClure N.: Randomly alternating image presentation during laparoscopic training leads to faster automation to the "fulcrum effect". *Endoscopy* **15(10)** (2000) 317-321
6. Nicolaou M., James A., Darzi A., Yang G.Z.: Extraction of laparoscopic strategies of novices using eye tracking. In MIPS XI conference (Windermere, UK, 2005)
7. Bashir F., Khokhar A., Schonfeld D.: View-Invariant Trajectory-Based Activity Classification and Recognition. *ACM Multimedia Systems Journal (MMSJ)* (2006)
8. Rabiner L.R.: A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of IEEE* **77(2)** (1989) 267-296
9. Keogh E.J. and Pazzani M.J.: Derivative Dynamic Time Warping. In First SIAM International conference on Data Mining (Chicago, IL, 2001)
10. Martin J.A., 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.* **84(2)** (1997) 273-278
11. Chen W., Chang S.F.: Motion Trajectory matching of Video Objects. *Proceedings of SPIE* **3972** (2000) 544-553