Student: Giacomo Taylor

Mentors: Dr. Jeremy Brown, Dr. Anand Malpani, Dr. Gina Adrales

Course: 601.446 Computer Integrated Surgery II, Spring 2018

Seminar Paper Review

Citation:

R. Kumar, A. Jog, A. Malpani, B. Vagvolgyi, D. Yuh, H. Nguyen, G. Hager, C. Chen, "Assessing system operation skills in robotic surgery trainees," in The International Journal of Medical Robotics and Computer Assisted Surgery, 2012

Project Recap:

The goal of this project is to develop an intelligent system that can objectively assess robotic surgical skill using performance data about how surgeons move their hands, connected instruments, and how the instruments interact with the surgical workspace. This will be accomplished by building a hardware and software platform that collects motion data from da Vinci and physical interaction data (forces on task board and accelerations of tool). This will combine two previously developed surgical skill assessment platforms. Our platform will then be used to collect pilot data from users of various robotic surgical skill levels. We will search for statistical patterns in the data with the goal of determining the most suitable classification algorithms. Finally, we hope to expand on an existing IRB proposal allowing us to begin work on large scale data collection and data processing in preparation for continuing work on this through machine learning techniques

Paper Selection:

This paper was selected to be reviewed as it was coauthored by one of our mentors. Furthermore, the experimental setup of the paper is very similar to the end application of our system, making it very relevant. Overall the paper summarizes work done by Dr. Anand Malpani on a previous project using only the Da Vinci kinematics to predict robotic surgical skills. This paper also complements the other paper reviewed by a teammate very well, as it describes the other of two systems that are being integrated together in this project. The paper also provides suggestions for ML features and techniques to be used, in addition to using a similar data collection methodology to this project's.

Problem Statement and Key Results:

This paper outlines one of the major problems with the increase in minimally invasive robotic surgeries. Namely, evaluation of robotic surgical skill requires a time-intensive review process where a senior surgeon must observe a trainee and evaluate their skills visually. Not only does this practice costly, it also introduces a significant degree of subjectivity to the evaluation. The increased prevalence of this kind of surgery requires research into objective ways of measuring robotic surgical skill. The key result of this paper is that kinematic data from these robotic setups can be collected (without significant workflow disruption), and can accurately differentiate between skilled and unskilled surgeons.

Significance:

These findings have significant impacts on the medical community. Firstly, implementation of automated skill assessment reduces the need for human raters to assess basic psychomotor skill development. This will save time, money, and may provide a more accurate assessment of skill. Additionally, this paper is one of the first published to demonstrate automatic skill assessment for robotic minimally invasive surgery via purely kinematic data. The fact that this data can be easily collected without disrupting the surgical workflow opens a new realm of possibility for surgical evaluation. Finally, this kind of data evaluation can allow for real-time feedback for a trainee during training, making it possible to provide more specific and tailored feedback, hopefully improving learning curves for this kind of surgical training.

Background:

Training with a clinical robot is the standard for training surgeons in robotic minimally invasive surgery. This method is preferred over virtual reality training because the process is closer to actual surgery than what is currently possible to simulate through VR. However, as mentioned skill evaluation is often subjective, tedious, time consuming, and cost ineffective. Other work published demonstrates the use of surgical workspace interaction forces to classify surgical skill. Yet these methods require the use of additional force sensors that disrupt operating room workflow. These restrictions mean that these methods can only be used for somewhat synthetic scenarios that may not accurately mimic actual surgical procedures.

Methodology:

This paper collects kinematic data from the Da Vinci SI surgical system. All the data collected is available through the device's streaming API. The data is collected by a workstation that can be connected through an ethernet connection to the Da Vinci and does not otherwise affect the setup of the machine. This self-containment of the system is a major advantage.

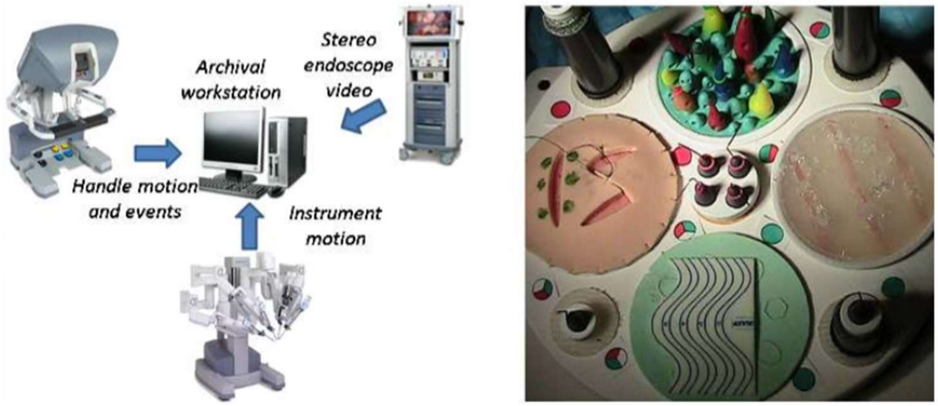

Figure 1. Information flow for the JHU/VISR archival system for the da Vinci system (left), and the benchmarking task pod (right). A demonstration pod (not used for benchmarking), the dissection pod, transection pod and the suturing pod (clockwise, respectively). The posts for the manipulation task are in the center and on the periphery

During a series of surgical skill benchmark procedures (a dissection task, transection task and a suturing task), the system captures all 334 variables that contains the Cartesian positions, velocities, joint angles, joint velocities, torque data, and discrete event data of the system. Furthermore, the endoscope video feed is captured as well, allowing for offline ground truth scoring to be calculated by the authors' expert clinical collaborators.

The study comprised of collecting data from residents of differing skill levels from four different institutions (Johns Hopkins, Children's Hospital, Boston, Stanford/VA Hospitals, and University of Pennsylvania). All participants had varying levels of surgical skill, but typically no prior experience with robotic surgical skill.

The data collected from the API contains 334 dimensions each sampled at 50Hz. The sampled data was used for training and validation of supervised machine learning algorithms, specifically classifiers using kernel SVMs. The ground truth data provided to the SVMs comes from expert collaborators who used the recorded footage to assign OSATS[1] scores to the benchmark tasks performed by the subjects. The label of "skill" was a binary classification, where an OSATS score greater than 13 was determined to be "expert," and less than 10 was designated "trainee." The authors of the study decided to focus on three major categories of human-machine interactions to judge surgical skill:

- Master workspace management
- Camera field of view adjustment
- Instrument safety (field of view considerations)

To this end, two seconds of master manipulator kinematic data was collected during clutch operations of the Da Vinci. Similarly, 0.5 seconds of manipulator kinematic data was

[1] Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg 1997; 84(2): 273–278

saved for camera manipulation events. For each clutch or camera manipulation, the sampled data was post processed to extract feature vectors containing Cartesian pose, Cartesian velocity, and gripper angle. These features were concatenated to create fixed-length vectors that were analyzed with no dimensionality reduction methods applied. Each feature vector was assigned a ground truth expert or trainee label using the expert OSATS scoring. The binary SVMs were then trained with a polynomial kernel on these feature vectors. Varying subsets of the experimental data was used as a training dataset for training the classifier, and the remainder for validation

Results:

In regards to the master workspace adjustments, the authors report classifier performance ranging from 91.75% to 95.7% correct classifications for varying amounts of held-out data. The authors also performed 10-fold cross-validation on the experimental dataset to assess statistical significance. The SVM classifier was tested by performing 10 different tests, each with approximately 10% of randomly selected data for testing, and the rest for training. Using a quadratic kernel, the classifier correctly classified 91.75% of the observations using 50% of the data, rising to 92% with training on 90% of the data. It is worth noting that different kernels using higher order polynomials did not significantly improve these results.

For the camera manipulation analysis, the leave-one-out validation with 10% of the data experienced promising results of approximately 87% accuracy and precision, and recall of 100%. The classifier correctly classified 88.16% of the data, with a 100% recall during the 10-fold cross-validation. Again, the authors report that increasing the dimensionality of the SVM kernel does not improve its performance.

The authors include great visualizations of a sample of the unsafe motion data collection, however, state that analysis of this data is ongoing and do not report on it further.

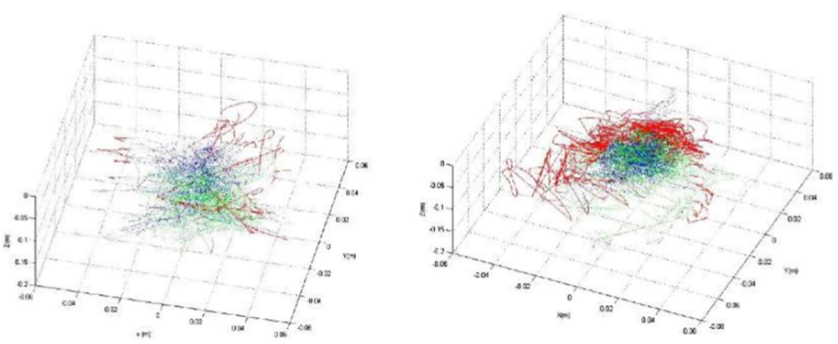

Figure 4. Expert (left) and trainee (right) instrument Cartesian trajectories for left (blue) and right (green) instruments during a training task. The red portions represent where an instrument was manipulated unsafely outside the field of view of the endoscopic camera

Pros and Cons:

Perhaps one of the best key points of this paper is the fact that the proposed procedure of data collection it outlines does not interfere with the current surgical workflow. This means that it can comprise the groundwork for data collection during many robotic surgery interventions. This possibility shows that a more objective evaluation of robotic surgical performance can be implemented at a large scale without interrupting current procedure.

On a similar note, another advantage of this paper is that the benchmark tasks being performed closely mimic portions of actual procedures. While these suturing, transection, etc. tasks may not entirely capture the skills needed for robotic surgery, they provide an exercise that is closer to actual surgery than simple peg transfer or ring walk experiments. The authors also visualize much of the data in a very intuitive and effective manner, which can be especially helpful in machine learning applications, where the meaning of features can often be obscured.

On the other hand, while the visualization display the data nicely, analysis of an entire third of their collected data (unsafe motions and collisions) is ongoing and unfortunately omitted. It would have been very nice and convincing to see a discussion of this data as it would provide a lot of validation to the authors' overall conclusion.

Furthermore, the OSATS specifications the authors use would have benefitted from more explanation; the authors state that they use 6 categories of the OSATS Global Rating Scale of Operative Performance, however the paper they reference describes 7 categories and the authors do not indicate which they omit. Additionally, the expert/trainee distinction they choose based on the OSATS scoring seems somewhat arbitrary and could have been explained.

Conclusion:

Overall this paper is a great resource for this project. It explains the much of the motivation, hardware, data processing, and software related to the project. This paper will continue to be a useful for us as we continue work on our project.