A Holistic Data Acquisition Framework for Robotic Surgical Skill Assessment

Students: Scott Pourshalchi, Giacomo Taylor
Mentors: Dr. Jeremy Brown, Dr. Anand Malpani

Table of Contents

Introduction

Robot-assisted minimally invasive surgery (RAMIS) is quickly becoming the prescribed method of treatment for many different routine and non-routine surgical procedures. There is a need to ensure that all robotic surgeons have a minimal level of skill proficiency before they operate on patients. (Pradarelli, et al.) Current methods of skill assessment rely almost exclusively on structured human grading which can be subjective, tedious, time consuming, and cost ineffective as raters are practicing physicians. (Curry M, Malpani A, Li R, et al.)

There exists some previous work investigating the relationship between robot dynamics and operator skill. Integrating force sensors and accelerometers into surgical skill assessments, such as peg transfer and similar tasks, allows categorization of surgical skill with reasonable accuracy. By analyzing time-series data of contact forces between surgical tools and the workspace, as well as high-frequency vibration patterns, it is possible to extract statistical measures that can be used in both regression and classification machine learning algorithms. This approach has been shown the ability to accurately predict Global Evaluative Assessment of Robotic Skills (G.E.A.R.S.) ratings with high interrater correlation scores. (Brown et al) Having these outputs to agree with the G.E.A.R.S metrics is desirable, as they represent the current standard of judging surgical skill. (Goh, Goldfarb, et al.) In a similar vein, work has also been done using the kinematics of robot end effectors to predict robotic surgical skill. (Malpani, et al) By integrating these two data modalities of kinematic and force data, we hope to work towards creating a system capable of automated, objective, and time-effective surgical skill assessment.

Project Goal

The goal of this project is to develop an intelligent system that can collect data and provide the ability to objectively assess surgical skill using performance data about how surgeons move their hands and connected instruments, as well as how the instruments interact with the surgical workspace.

Technical Approach

The first phase of the project will involve the development of a hardware & software platform using ROS (Python) that collects synchronous streams of motion and physical interaction data (forces and vibrations). This will combine the two previously developed surgical skill assessment platforms created by Dr. Brown and Dr. Malpani.

Figure 1: A system overview

Our system will interface directly with the da Vinci via ethernet connection and collect kinematic data of both the instrument arms and master manipulators. Simultaneously, the computer will receive serial data from the smart task board microcontroller, describing the forces exerted on the environment during the skill task. Our ROS platform will need to combine the two sources of data, accounting for the different frequencies of output. A future iteration of our system will also grab video frames from the da Vinci camera and synchronize this footage to the time-series data collection. This video collection will not only allow for a G.E.A.R.S. assessment retroactively (providing ground truth classification labels for algorithm training), but also to help validate the synchronization of data collection.

As part of our system validation, we created a data visualization application in MATLAB. Seen in figure 2 below, the application allows the user to pick from a selection of time series data to plot together. This allows for easy comparison of different features at any given time. The user can use the sliders to plot the data over a specific region of time. Additionally, the debug plot button can be used to ensure the systems were synchronized over the duration of the trial.

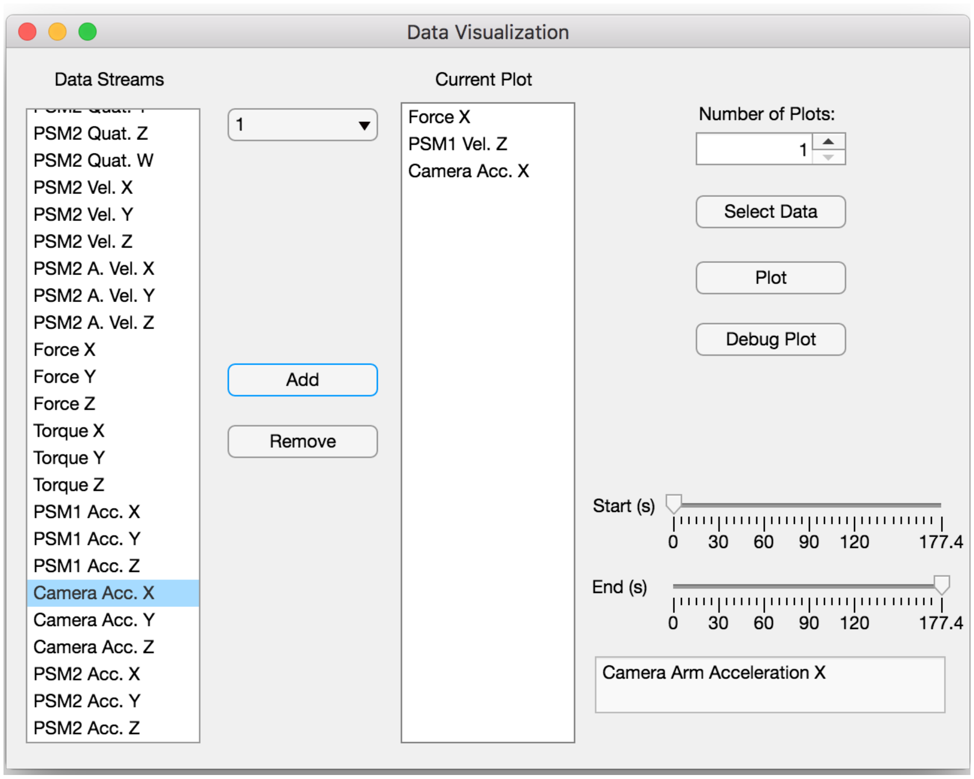


Figure 2: VisaualizeData2.mlapp, a data visualization tool

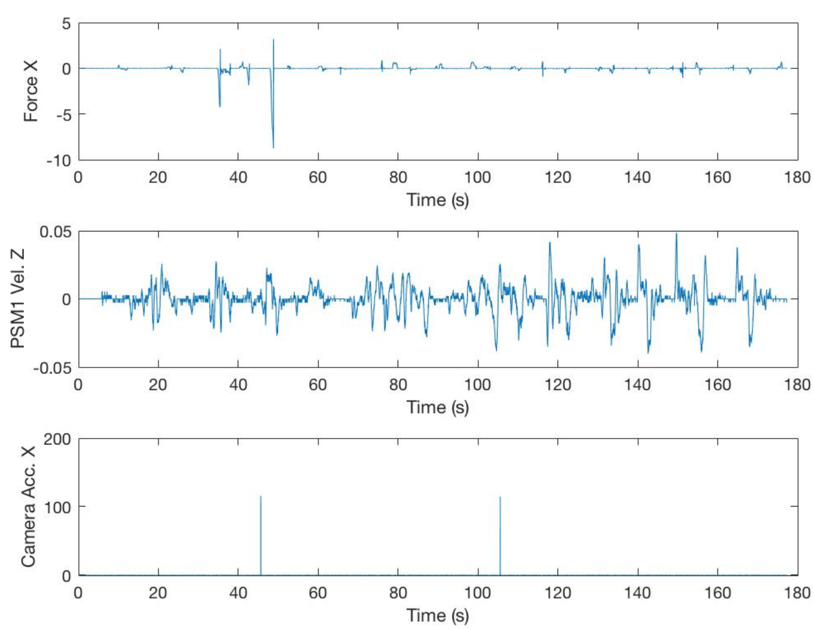


Figure 3: Example program output

Once the system has been validated, we collected pilot data. This will include members of the LCSR group who have experience with the da Vinci system, as well as subjects unfamiliar with it. This will provide a baseline for skilled and unskilled data. This pilot data can then be analyzed in an attempt to discover salient features. Previous work has shown that discrete features can be generated from the time series signals to be used for classification or regression based machine learning. These include mean, standard deviation, minimum, maximum, range, RMS, TSS, and time integral. (J. D. Brown et al.) The provided python module, calc_stats.py, takes the data of a particular trial and calculates several of these metrics

The final phase of this project consists of preparing for large-scale data collection. This beings by examining the pilot data for features or metrics that are statistically different between the skilled and unskilled users. We have determined some features for which the difference between the novice and expert groups are statistically significant. These suggestions for features will hopefully guide future work to determine the best ML approaches for discerning skilled versus unskilled surgeons.

A future goal for this endeavor is to transport our system, comprising of the PC, smart task board, and clip-on sensors to the JHU Minimally Invasive Surgical Teaching Innovation Center, interface with the da Vinci system and monitor setup there to collect data on a large scale from many clinicians of varying skill levels. With this larger set of data, one will be able to more exhaustively evaluate the system's ability to objectively classify surgical skill.

Results

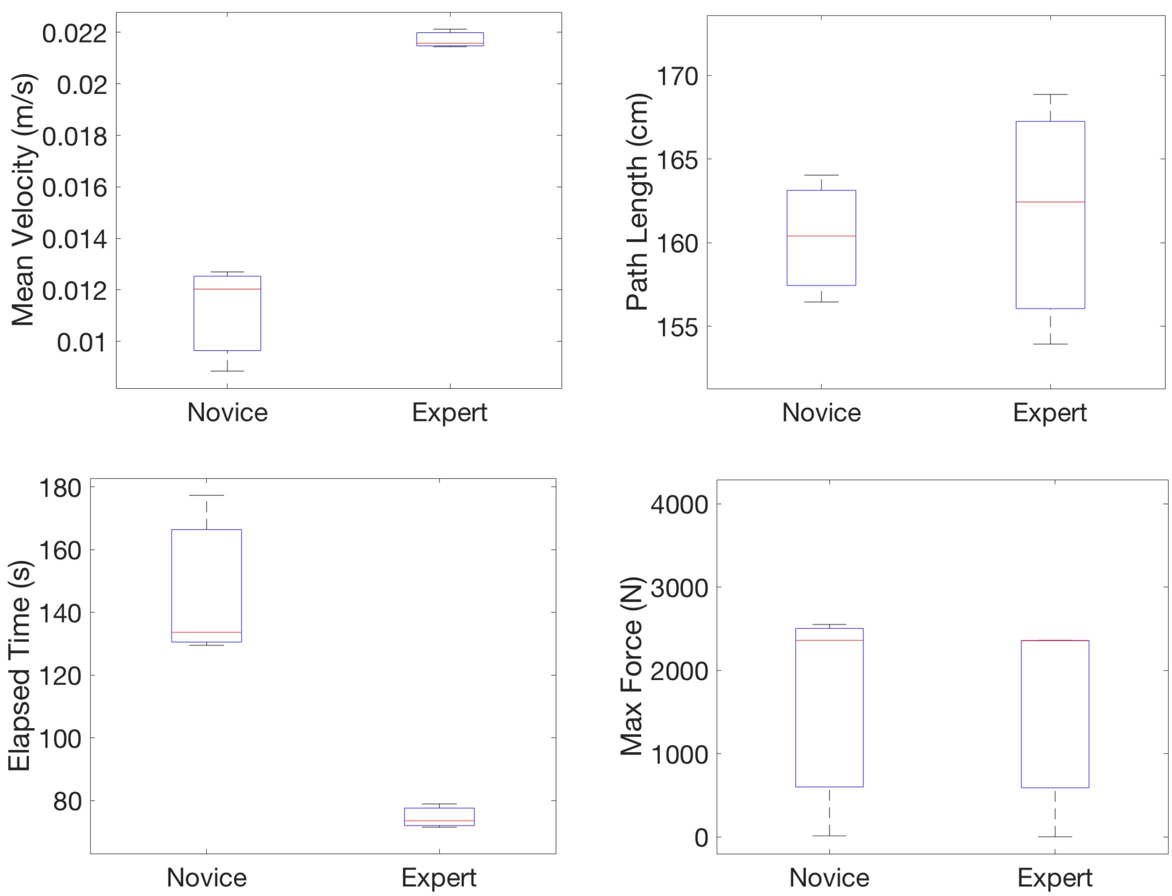
The pilot data collected consists of time series data of two LCSR collaborators, one experienced using the da Vinci and one who had never used the da Vinci previously. Each subject performed the peg transfer task 3 times. During a trial, the subject picked up each of the six blocks from pegs on the left side of the workspace, transferred the block to the opposite gripper tool, and then placed the block on a peg on the right side of the workspace. This process is then reversed so all blocks return to their original pegs after conclusion of the trial.

Each of the five graphs below shows the statistics for a metric calculated over all the trials, grouped by the "novice" or "expert" designation of the subject. The red midline of each box represents the median of the metric over all trials, while the blue box surrounds the 25th to 75th percentile of the data. Lastly, the upper and lowermost horizontal bars outside of each box represent the maximum and minimum values of the data respectively.

The five metrics analyzed were calculated as follows:

- Mean velocity: the average of each instantaneous velocity reading for the two end effectors sampled at the refresh rate of the da Vinci API

- Path length: the total path of the two end effector tips over the whole trial, found by summing the differential distances between each sample reading
- Elapsed time: the amount of time between the subject beginning and finishing the task
- Max force: the maximum force exerted on the smart taskboard force sensor during the entire trial
- Force integral: the force reading of the smart taskboard sensor summed (integrated) over each instantaneous time to capture the total force exerted



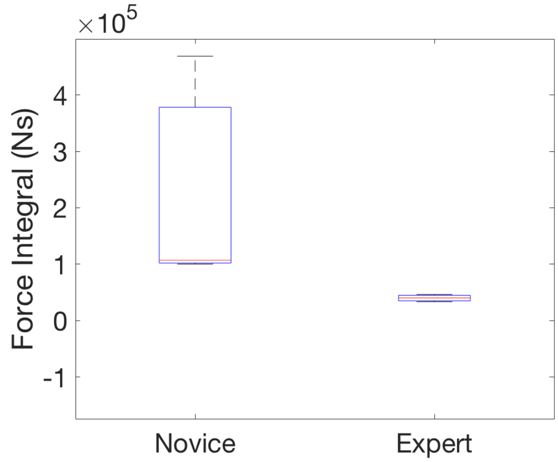


Figure 4: Box plots of novice and expert user metrics

As seen in the figure above, certain metrics distinguish novice users from expert users more clearly. For instance the box plots for force integral, mean velocity, and elapsed time, have no overlap region. In contrast, there is large overlap for the box plots of path length and max force.

Discussion

Our results serve as a proof of concept for this system. We were able to combine two previously existing systems into one, and use this to collect preliminary data from users. This data suggests that there are measurable differences between users are skilled and unskilled related to these metrics. In a study with larger scale data collection, related metrics may prove to be statistically significantly different, and therefore make good choices for features in classification algorithms. This gives good reason to conduct further research into the usage of this data for machine learning application. The metrics we have found to be different between the skilled and unskilled groups are likely to be useful features in a machine learning algorithm that can predict robotic surgical skill.

Management Plan

Students had weekly meetings with mentors Dr. Brown and Dr. Malpani to discuss progress, obstacles, and brainstorm solutions. Students met twice weekly to go over progress made on individual responsibilities. This time was also used to work on tasks that required both students. Source control was handled via an LCSR git repository:

https://git.lcsr.jhu.edu/gtaylo34/davinci-kinematics.

This repository was kept private due to the protected nature of the da Vinci API. Design and documentation of the system (user manual, interface specifications) have been kept in a google doc:

https://docs.google.com/document/d/1rh5adNx-cgUqRqL0r09x8H59MVoyLfF0Mu0UvFmYkdA/edit?usp=sharing.

Responsibilities were divided among the students to best distribute the workload and complement the skills of each student. The table shows the breakdown of responsibilities for this project:

Scott	**Giacomo**
Hardware Setup	Installation of programming environment
Project Documentation	Data acquisition system
Data visualization application	Data analysis post-processing
Oversee data collection	

Final Deliverable Status

	Planned	Accomplished
Minimum	• Functional computer • Program to integrate two systems • User manual/ documentation	• Functional computer • Program to integrate two systems • User manual/ documentation
Expected	• Collect pilot data • Program for data visualization, feature extraction	• Collect pilot data • Program for data visualization, feature extraction
Maximum	• Statistically significant metrics • Machine learning feature suggestions	• Metrics indicating statistical significance • Machine learning feature suggestions

This chart depicts the final status of our deliverables, as updated during the midpoint presentation. For our minimum deliverables, we have created, validated, and documented our program on a computer in the Mock OR. For our expected deliverables, we have successfully collected pilot data and created a program to visualize this data and a program to analyze this data. Finally, for our maximum deliverables we analyzed the data to determine which metrics were significantly different between skilled and unskilled groups of users and suggested these as features for future machine learning usages.

Lessons Learned

Many valuable lessons were learned through this course and project. First and possibly most importantly, check for loose cables. This simple fix, was the solution to a number of large problems in this project. Loose cables kept the da Vinci machine down for almost two weeks. Loose cables also caused issues during data collection, invalidating some trials. Remembering to check that all cables are well connected will save users time and frustration.

Many lessons were learned with regards to management. When creating a timeline for the project, plan to encounter setbacks. Allow yourself time to tackle unforeseen problems. Solutions hardly work as intended the first time. Additionally, once you create a timeline, stick to it. Update the timeline if needed. Referring back to the timeline can help keep the project on schedule.

Finally, determine what tasks have dependencies and what can be done in parallel. This will keep time from being wasted waiting for certain tasks to be completed. A project of this size will undoubtedly encounter many obstacles; being able to maximize productivity during these setbacks is crucial to staying on track. The easiest way to do this is to ensure you are making use of the full team.

Future Work

Preparations have been made to pass on this work to future student researchers, potentially for summer 2018. However, the current student researchers may continue this work in Fall 2018. This project saw the creation of a framework for multimodal data collection. Pilot data was collected to demonstrate basic differences between skilled and unskilled users. This needs to be explored more fully. A large scale study must be conducted to collect much more thorough data. This would include getting clinicians of varying skill levels to participate. With more data, the project many be expanded to apply machine learning algorithms to determine which features can be used to accurately predict the skill of a user.

In future machine learning algorithms, possible choices for skill label can include surgical appointment, number of robotic surgical procedures completed, or a G.E.A.R.S. rating for each trial. To obtain G.E.A.R.S. rating for each trial video capture would be necessary. This is one technical modification that should be included in future work. Additionally, the data visualization software should be modified to include video in the visualization so users can see what moments in the trial correspond to what features in the data.

Acknowledgements

This project could not have been made possible without the help of our wonderful mentors. Special thanks to the participants in our study. Additionally, thank you to Anton Deguet for help setting up the programming environment and general debugging. Code made possible thanks to creators of the jhu-ciist and jhu-saw libraries.

References

Curry, Martin, et al. "Objective Assessment in Residency-Based Training for Transoral Robotic Surgery." The Laryngoscope, vol. 122, no. 10, 2012, pp. 2184–2192., doi:10.1002/lary.23369.

E. D. Gomez et al., "Objective assessment of robotic surgical skill using instrument contact vibrations", Surgical Endoscopy Interventional Techn., vol. 30, pp. 1419-1431, 2015.

Goh, Alvin C., et al. "Global Evaluative Assessment of Robotic Skills: Validation of a Clinical Assessment Tool to Measure Robotic Surgical Skills." The Journal of Urology , vol. 187, no. 1, 2012, pp. 247–252., doi:10.1016/j.juro.2011.09.032.

J. D. Brown, C. E. O'Brien, S. C. Leung, K. R. Dumon, D. I. Lee and K. J. Kuchenbecker, "Using Contact Forces and Robot Arm Accelerations to Automatically Rate Surgeon Skill at Peg Transfer," in IEEE Transactions on Biomedical Engineering, vol. 64, no. 9, pp. 2263-2275, Sept. 2017.

K. Bark et al., "Surgical instrument vibrations are a construct-valid measure of technical skill in robotic peg transfer and suturing tasks", Proc. Hamlyn Symp. Med. Robot., pp. 50-51, 2012.

Pradarelli, Jason C., Darrell A. Campbell, and Justin B. Dimick. "Hospital Credentialing and Privileging of Surgeons: A Potential Safety Blind Spot." JAMA 313, no. 13 (April 7, 2015): 1313–14.

Rajesh Kumar, Amod Jog, Balazs Vagvolgyi, Hiep Nguyen, Gregory Hager, Chi Chiung Grace Chen, David Yuh, "Objective measures for longitudinal assessment of robotic surgery training," The Journal of Thoracic and Cardiovascular Surgery, Volume 143, Issue 3, 2012, Pages 528-534, ISSN 0022-5223

Shackelford, Stacy, Mark Bowyer. "Modern Metrics for Evaluating Surgical Technical Skills." Current Surgery Reports, vol. 5, no. 10, 2017, doi:10.1007/s40137-017-0187-0.

Appendices

All code can be found at:
https://git.lcsr.jhu.edu/gtaylo34/davinci-kinematics

Documentation/user manual can be found at:
https://docs.google.com/document/d/1rh5adNx-cgUqRqL0r09x8H59MVoyLfF0Mu0UvFmYkdA/edit?usp=sharing