Seminar Presentation:
A Holistic Data Acquisition Framework for Robotic Surgical Skill Assessment

Student: Scott Pourshalchi
Mentors: Dr. Jeremy Brown,
Dr. Anand Malpani

"USING CONTACT FORCES AND ROBOT ARM ACCELERATIONS TO AUTOMATICALLY RATE SURGEON SKILL AT PEG TRANSFER"

J. D. Brown, et. al

Laboratory for Computational Sensing + Robotics

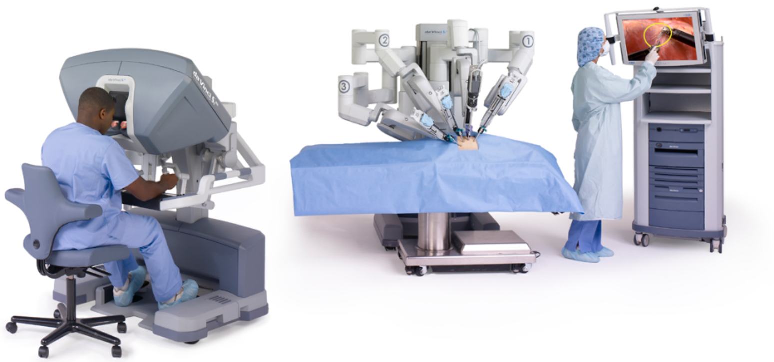
PROJECT RECAP

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.

- Develop a hardware + 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.

- Collect pilot data from users of various robotic surgical skill levels

- Search for patterns in data to prepare for machine learning applications



PAPER SELECTION

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.

UNIVERSITY of PENNSYLVANIA

PROBLEM STATEMENT AND KEY RESULTS

- Problem: Current methods of skill assessment for robotic surgery rely almost exclusively on structured human grading (G.E.A.R.S.) which can be subjective, tedious, time consuming, cost ineffective (raters are practicing physicians).

- Finding: A surgeons skill at robotic peg transfer can be reliably rated via regression using features gathered from force, acceleration, and time sensors external to the robot.

SIGNIFICANCE

- Reduces need for human raters to assess basic psychomotor skill development (save time, money, objectivity)
- Improved trainee learning due to real-time feedback on skill
- First study to demonstrate automatic skill assessment for robotic minimally invasive surgery via physical interaction information (external to robot).

BACKGROUND

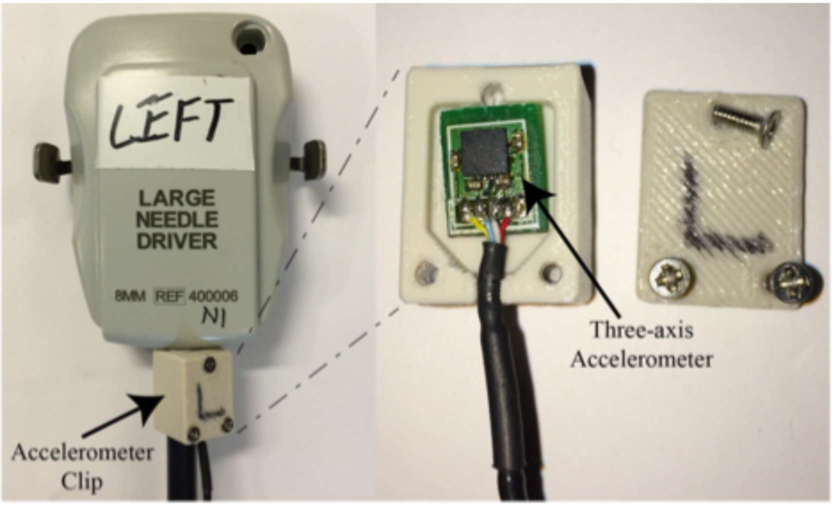
- Training with a clinical robot is the standard for training surgeons in robotic minimally invasive surgery.
- Previous work was published that used robot kinematics to assess skill during training and actual surgical procedures

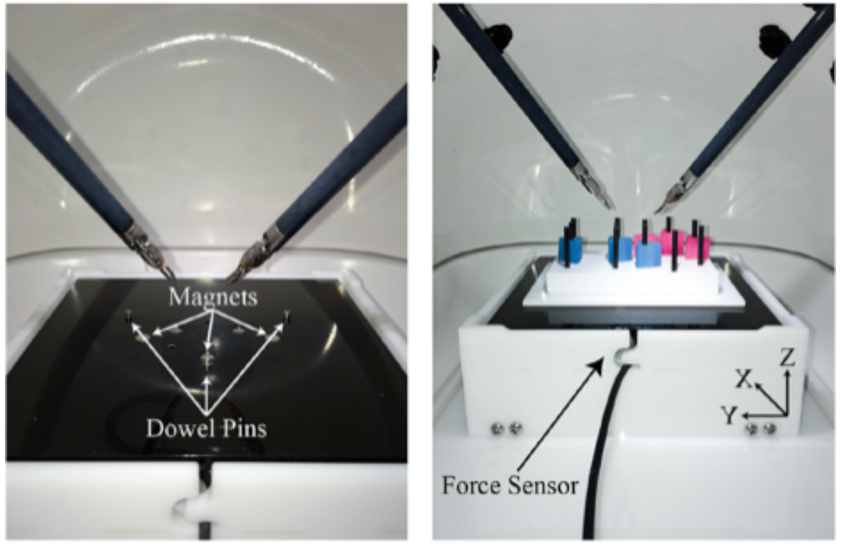
- Kinematics based methods cannot account for potential *master-slave misalignments* due to sensor error or unmeasured quantities such as *compliance and mechanical wear.*
- Few papers have measured the physical interaction between the robot and the environment when analyzing trainee skill development.
- Previous work by same authors showed that the root mean square of high frequency vibrations of both the robotic tools and forces exerted on the task materials are greater for novices than experts.[1]

1. 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.

HARDWARE

- Hardware:
  - High bandwidth 3 axis accelerometer clips for the two primary robotic arms and endoscope
  - Smart Task Board with a three axis force sensor
  - Records at sample rate of 3 kHz
  - Records video at via s-video connection
  - Data collection controlled through Python script



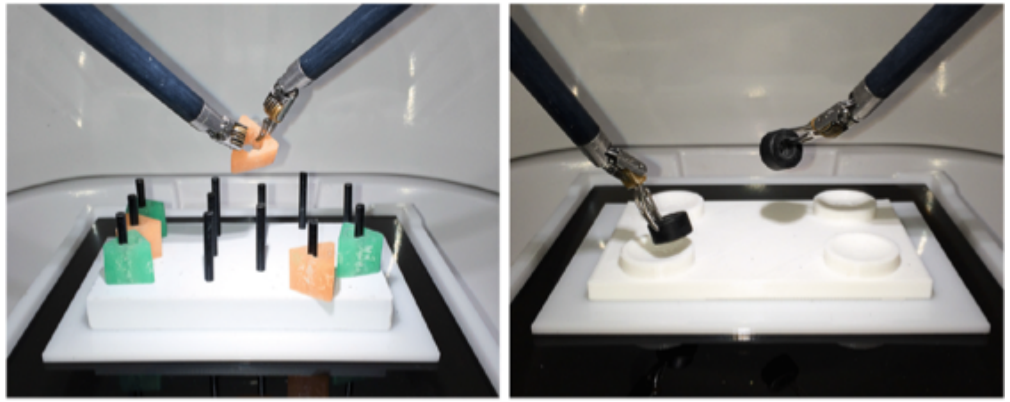


PARTICIPANTS

- Participants (N=38):
  - Obtained participants of various skill levels in robotic surgery (Table 1)
    - Self reported familiarity & number of robotic cases
  - Each participant was allowed to warm up with a practice task (image b), then completed 3 trials of peg transfer (image a), and finally completed a demographic questionnaire.

TABLE I
PARTICIPANT DEMOGRAPHICS

Handedness	Left	Right	Ambidextrous	
	3	32	3	
Familiaritywith Robot	None	Limited	Moderate	Extensive
	13	10	6	9
# Robotic	None	1–100	101–500	501 +
Cases	22	6	7	3



(a) (b)

G.E.A.R.S. RATING

- Rating performed by 2 surgeons with previous experience rating
- 1– 5 Scale on:
  - Depth Perception
  - Bimanual Dexterity
  - Efficiency
  - Force Sensitivity
  - Robotic Control
- Interrater reliability
  - Raters given time to “calibrate”
  - Each rater given set of 10 diverse videos and time to discuss ratings
  - Interrater reliability of ratings assessed using intra-class correlation coefficient (ICC)
  - 0.6 was chosen as the minimum acceptable ICC for “good” reliability.

TABLE II

FREQUENCIES OF GEARS RATINGS AVERAGED ACROSS RATERS AND ROUNDED, AND FINAL ICC FOR RATED TRIALS

GEARS Domain	Ratings					ICC
	1	2	3	4	5	
Depth Perception	0	14	45	41	10	0.76
Bimanual Dexterity	0	9	41	44	16	0.80
Efficiency	3	14	44	30	19	0.89
Force Sensitivity	0	15	42	43	10	0.74
Robotic Control	2	10	48	42	8	0.80
Overall						0.88

FEATURE EXTRACTION

- Broke down time-series data into a set of discrete features for use in machine learning algorithm
- Acceleration data used to calculate roll (rotation around the shaft) and pitch (shaft angle relative to the horizontal)
  - $Roll\ \phi = \tan^{-1}\left(\frac{a_{fy}}{a_{fz}}\right)$
  - $Pitch\ \theta = \tan^{-1}\frac{-a_{fx}}{\sqrt{a_{fy}^2+a_{fz}^2}}$
- Time Features:
  - Total elapsed time, total active time, square root and log (skill may be non-linear with time)

- Descriptive features:
  - Mean, standard deviation, minimum, maximum, range, Root Mean Square (RMS), Total Sum of Square (TSS), time integral OF…
  - Force directions and magnitude
  - Tool and camera roll and pitch angles, angular velocity, accelerations
  - Product of right and left tool acceleration in each frequency band
  - Product of force magnitude and right/left tool acceleration in each frequency band

MACHINE LEARNING & RESULTS

- 10 learners: Regression and classification for each G.E.A.R.S. domain
- 33 for training set, reserve 4 participants for testing -> approximately 90%|10% standard
- Regression learners computed in MATLAB using LIBSVM library, Glmnet library, and the Statistics and Machine Learning Toolbox.
- Random forest classification learners implemented using TreeBagger function in MATLAB's statistics and machine learning toolbox.
- Training took approximately 30 min for all 5 domains, 30 sec for rating calculation/classification

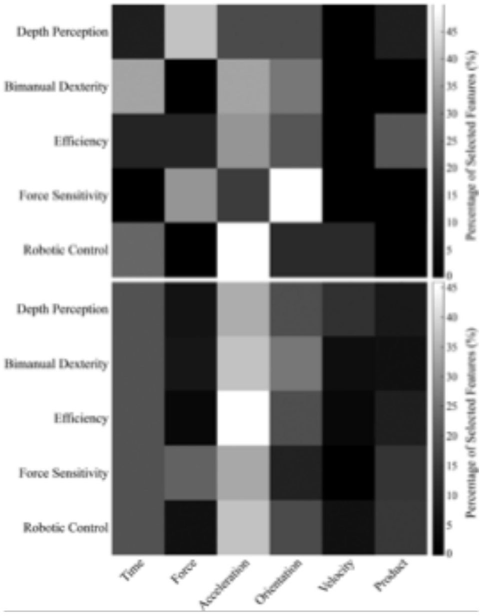


TABLE III
EXACT ACCURACY ACROSS TESTING SETS

GEARS Domain	Regression Learner	Classification Learner
Depth Perception	63.3 ± 9.5%	71.7 ± 9.5%
Bimanual Dexterity	66.7 ± 11.8%	53.3 ± 16.2%
Efficiency	73.3 ± 16.0%	58.3 ± 8.3%
Force Sensitivity	63.3 ± 9.5%	51.7 ± 10.9%
Robotic Control	71.7 ± 12.6%	75.0 ± 15.6%

Values shown are mean ± standard deviation across the five testing sets.

- Precision > 0.2 -> Indicates performance was better than random chance

TABLE V
RANGE (MEDIAN) OF ICC(2,4) BETWEEN THE THREE RATERS AND EACH LEARNER (REGRESSION AND CLASSIFICATION) FOR THE FIVE RESERVED TESTING SETS

GEARS Domain	Regression Learner	Classification Learner
Depth Perception	0.80–0.88 (0.81)	0.76–0.84 (0.83)
Bimanual Dexterity	0.71–0.91 (0.86)	0.71–0.86 (0.84)
Efficiency	0.84–0.93 (0.88)	0.83–0.91 (0.88)
Force Sensitivity	0.70–0.90 (0.79)	0.66–0.84 (0.76)
Robotic Control	0.66–0.87 (0.79)	0.69–0.86 (0.81)
Overall	0.88–0.93 (0.89)	0.87–0.89 (0.89)

THE GOOD AND THE BAD

PROS:

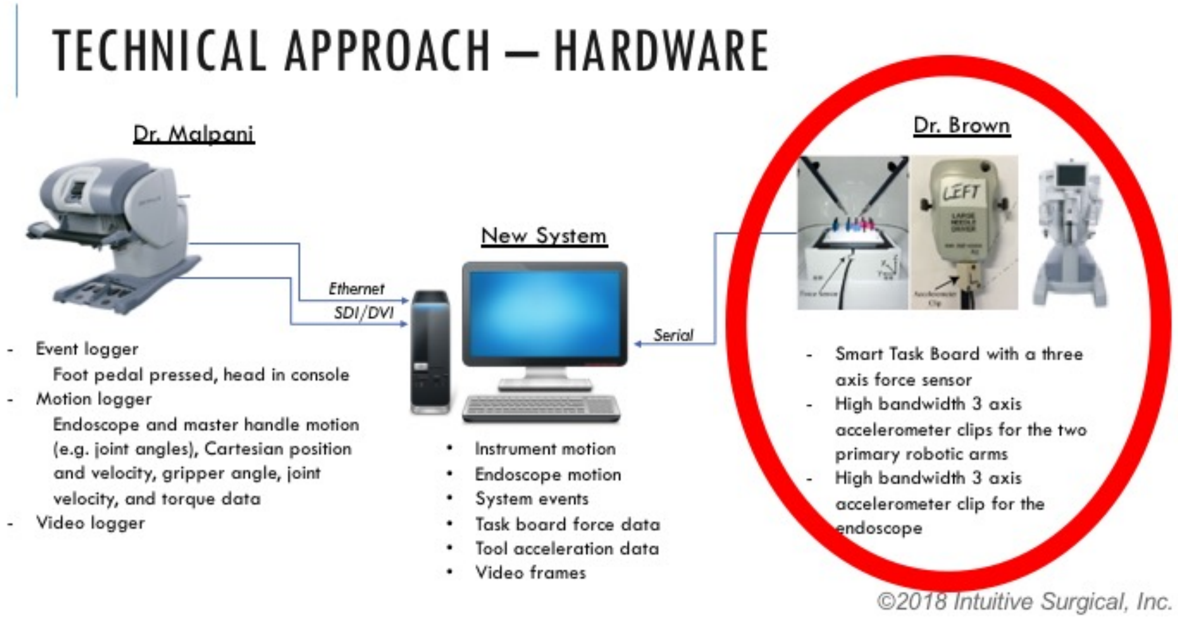
- Online supplements contain machine learning performance analysis without force data which cant be accessed *in vivo*.
- Lengthy discussion section – evaluates validity of results, impact of results, etc.
- Direct comparison of time saved through this approach (110 trials rated in 6 months by human grading, 20 min by regression).
- Approach is much more flexible than kinematics (Doesn’t interfere with robot control or operation, accounts for master-slave misalignment and compliance).

CONS:

- More description of why features are chosen and why machine learning methods are chosen.
- Lowest accuracy G.E.A.R.S. domain was force sensitivity – Suggests possibly not examining the right features.
- Paper demonstrates results for peg transfer – actual surgery is much more complex.
- Unequal representation among skill levels and low number of participants.

RELEVANCE TO PROJECT

- Describes hardware information of the data acquisition system we will use *(minimum deliverable)*
- Describes data preprocessing and important features used in machine learning techniques *(expected deliverable)*
- Describes user study similar to what we will create for IRB proposal *(maximum deliverable)*
- Discussion suggests next steps for project is combination of physical and motion interaction data – relates directly to our project.



CONCLUSION

- Paper was a great resource for this project:
  - Explains motivation, hardware, data processing, software
- We will continue to refer to this paper in the future.