Surgical Skills Analysis with Virtual Reality

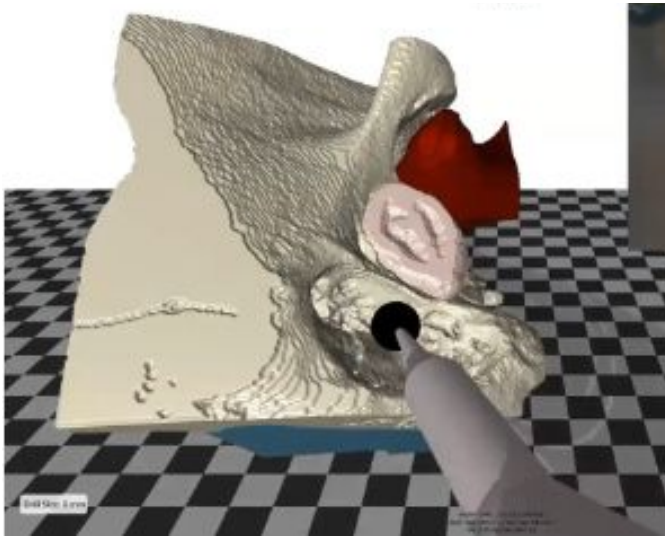
Seminar Presentation

Team members: Aditya Khandeshi, Liza Naydanova, Alexandra Szewc
Mentors: Max Li, Adnan Munawar, Dr. Creighton, Dr. Unberath, Dr. Taylor

03.10.2022
Computer Integrated Surgery II

Project Summary

- Evaluation of technical skill in surgical trainees
- Current Methods of Evaluation for OHNS
  - Objective Structured Assessments of Technical Skills (OSATS)
  - Society for Improving Medical Professional Learning platform
- Limitations of Current Methods
- **Goal:** Develop a convolutional neural network to provide an objective, technical assessment of skill of surgeons performing mastoidectomy
- Plan
  - AMBF Simulator for Temporal Bone Drilling
  - Collect data from attending surgeons and residents
  - Extract important features from AMBF data



Paper Selections

Modeling surgical technical skill using expert assessment for automated computer rating

Azari et al. (2019)

- **Summary:** A computer vision algorithm was developed and used to predict expert performance ratings from surgeon hand motions for tying and suturing tasks.
- **Relevance:** Paper identified features of hand motion that can be used to build kinematic models
    - Jerk and spatiotemporal curvature can be used as features in our model

Developing Effective Automated Feedback in Temporal Bone Surgery Simulation

Wijewickrema et al. (2015)

- **Summary:** Medical students evaluated an automated feedback system for facilitating skill acquisition in VR surgical simulations.
- **Relevance:** Paper identifies informative features for real-time feedback for the cortical mastoidectomy procedure.
    - Drill stroke force and distance to anatomical structures can be features in our model

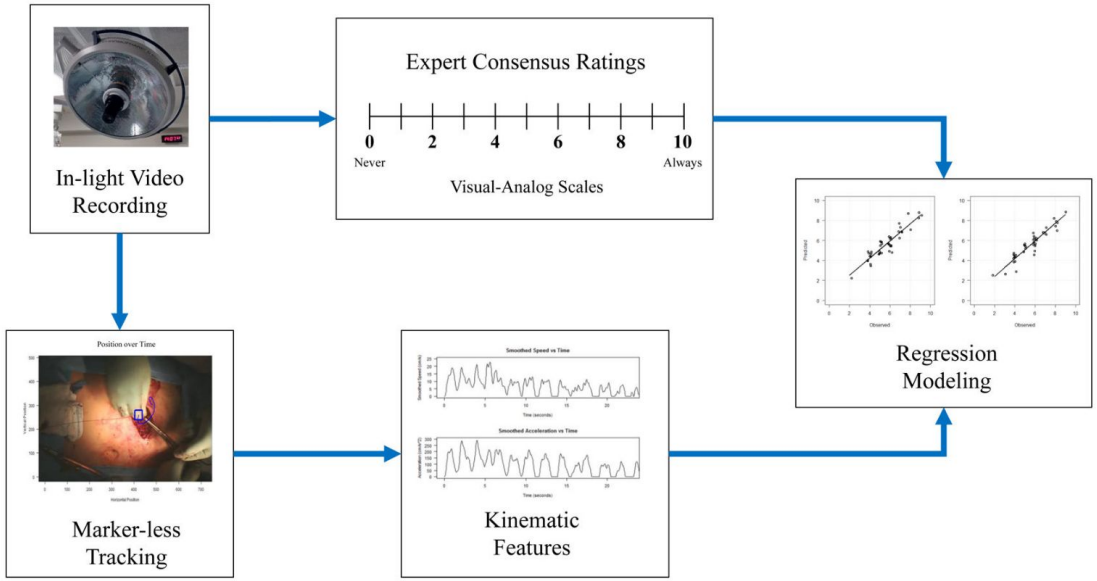
Problem Summary and Background

- Correctly implementing OSATS is time consuming and resource intensive
- Motion capture and tracking methods have been aiming to fix this problem
    - Methods often depend on space consuming systems and markers or sensors on surgeons' hands
- Authors used video motion capture of surgeon's hands
    - Non-invasive and scalable means of observing surgical motion
- Builds on previous work in developing system and kinematic models

Azari et al. (2019)

Methods

- Participants
  - 9 surgeons doing 16 surgical cases
    - 6 attendings, 3 residents
- Video Selection
  - Video segments where hands were visible >5 seconds were labeled
- Rating Scales
  - Subjective visual analog scale
    - Motion economy, fluidity of motion, and tissue handling
- Motion Tracking
  - Custom software to track ROI
- Calibration
  - Visible measurements of hands used to calibrate each clip



Azari et al. (2019)

Methods: Kinematic Features

TABLE 1.

Features Used for Modeling the Consensus Rating Scales

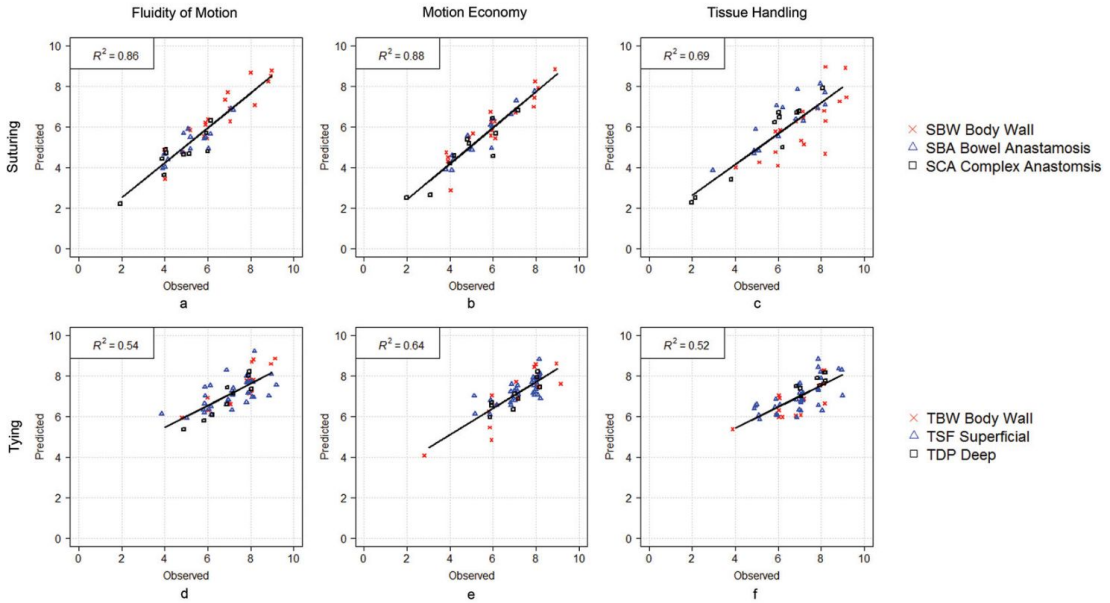
Summary Kinematics	Mean, median and max of speed, acceleration, curvature and jerk
RMS Speed & Acceleration	Root mean-squared transformation of speed; acceleration (respectively)
Moving Averages	Simple, exponential and moving averages applied to original and smoothed speed, acceleration and curvature signals
Peak Counts	Number of peaks, applied to pure and filtered speed, acceleration, curvature and all moving averages at thresholds at intervals of 20% from 0% to 100% of maximum signal amplitude and intervals of 20 mm/s for speed (up to 100) and 200 mm/s^2 for acceleration (up to 1000)
Peak Frequencies	Number of peaks divided by length of video clip (s) for each type of peak count category
Peak Variation	Coefficient of variation in peak arrivals (standard deviation / mean) for each peak count category
Idle time (%)	Percent of time spent below speed threshold (set every 5 mm/s up to 50 mm/s)
Working area	Mean variance in x-y distance calibrated positions over all recorded frames
Path Density(Avg, Med, Sd, Max)	Ratio of recurrent (range 0-7) recorded x-y positions to total recorded x-y positions
Speed Density(Avg, Med, Sd, Max)	Ratio of instantaneous speed in recurrent (range 0-7) recorded x-y positions to total recorded x-y positions
Curvature Density(Avg, Med, Sd, Max)	Ratio of instantaneous curvature values in recurrent (range 0-7) recorded x-y positions to total recorded x-y positions

Kinematic variable family names (left) and associated descriptions (right).

Azari et al. (2019)

Results

- **Video Data**
  - 103 video clips, average length of 11.72 seconds
- **Task Expert Rating Scales**
  - Consensus achieved after maximum of 3 iterations
  - Mean=5.91, std=1.62 (suturing)
  - Mean=7.12, std=1.10 (tying)
- **Prediction Models of Expert Ratings**
- **Prediction Model Validity**
  - Validated by comparing error between PRESS and SSE.



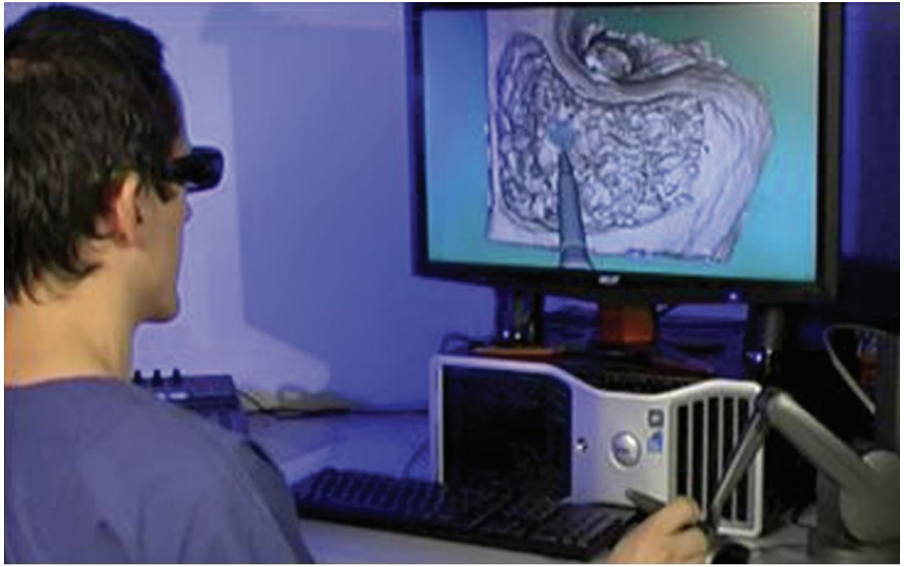
Critical Assessment

- **Importance of paper**
  - Progress towards more objective, reproducible and accessible assessments of surgical skill
  - Prediction models can be used to create automatic feedback system
- **Relevance to project**
  - Defines potentially informative kinematic features
    - Jerk (measure of smoothness)
    - Spatiotemporal curvature (count of discrete movements)
- **Limitations (discussed by authors)**
  - Limited range of scores
  - Surgical context not addressed
  - Procedure outcomes not taken into account
- **Limitations (noted by our group)**
  - Process of obtaining consensus decisions
  - Further information on what features were used for each of the models
  - Calibration process
- **Further Research**
  - Collect videos of specific clinically simulated scenarios
  - Include information about patient and procedure outcomes

Wijewickrema et al. (2015)

Problem Summary and Key Result

- Human expert resources are limited for increasing numbers of surgical trainees
- Existing need for opportunities for repeated, risk-free “deliberate practice”
- Objectives of this study included testing effectiveness, accuracy, and usefulness of feedback system.
    - Key Result: feedback improved drilling behavior of trainees



Melbourne University’s temporal bone surgery simulator.

Wijewickrema et al. (2015)

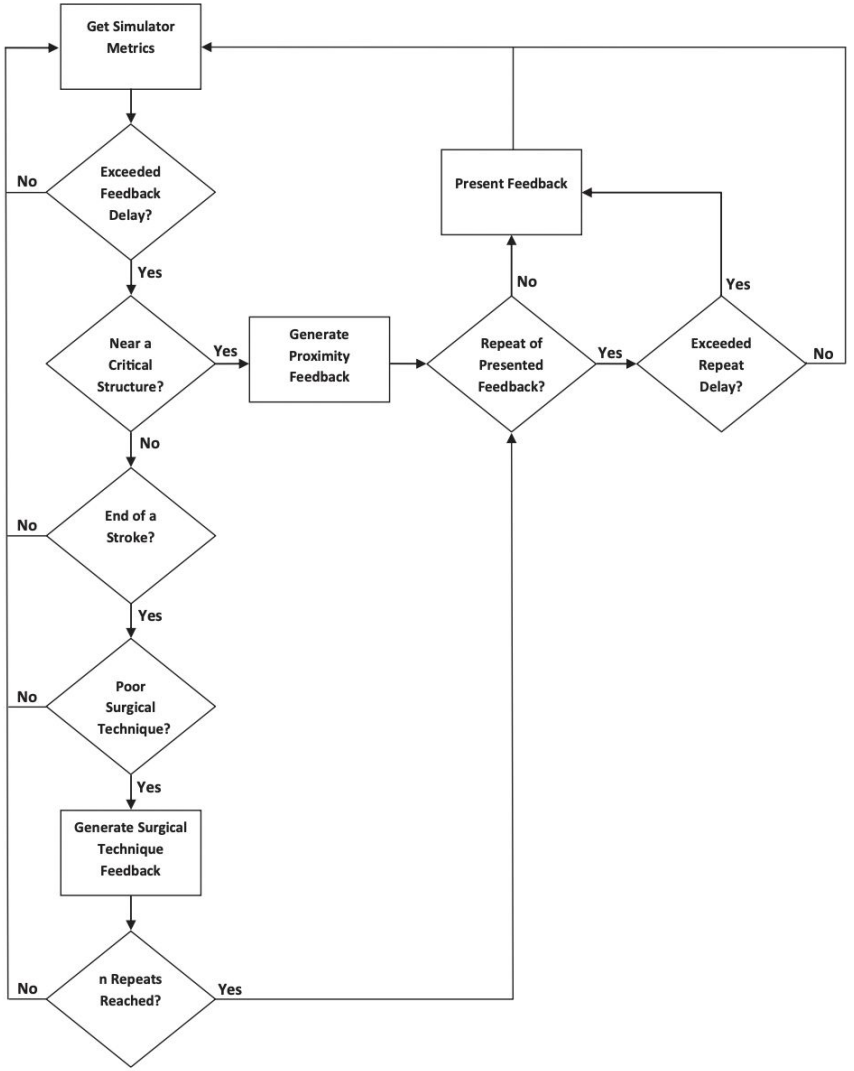
Background & Previous Work

- Previous work towards overcoming need for expert supervision in VR training environments have used end-of-task summative assessment:
  - Feedback not meaningful
  - Univariate analyses based on individual skill metrics
- Aim to develop multivariate feedback to more closely emulate meaningful and nuanced advice that human experts provide during surgical training.

- Important considerations: experience level of trainees vs. skill level used to train classifier.

Methods

- Participants
  - 24 medical students
- Test Platform
  - University of Melbourne VR temporal bone surgery simulator
  - 3D visual of mastoidectomy field
  - 3D haptic feedback device
- Feedback System
  - Classifier to recognize expert and trainee duration, length, speed, acceleration, force, straightness, burr size, magnification, bone removal rate, and distance to anatomical structures of strokes.

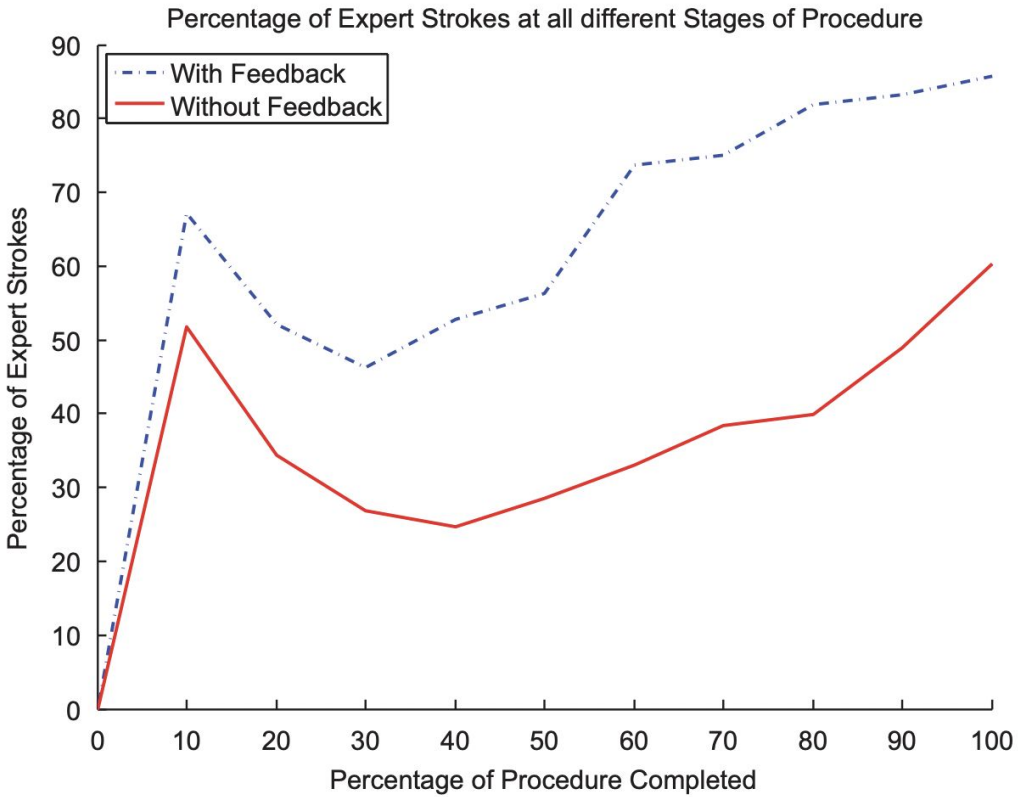


Wijewickrema et al. (2015)

Wijewickrema et al. (2015)

Results & Significance

- **Feedback Effectiveness**
  - Elevated rate of expert strokes in the feedback group, with $\chi 2(1) = 14.450$, $P < .001$
  - Only bone removal rate varied
  - No difference in structural component damage
- **Feedback Accuracy**
  - Feedback was provided by the system 88.6% of the time
  - Accurate in 84.2% of these instances
- **Feedback Usefulness**
  - 8/12 participants paid attention to feedback
  - 5/12 trainees found contradictory or incorrect feedback
  - 5/12 trainees found feedback unclear



Critical Assessment

- **Importance of paper**
  - Novel attempt to demonstrate how multidimensional surgical skill feedback can be integrated into procedure simulations in real-time
  - Identified significant skill metrics that can be adapted into feedback
- **Relevance to project**
  - Defines potentially informative features
    - Average drill stroke force
    - Average drill distance to anatomical structures
- **Limitations (discussed by authors)**
  - Unclear or contradictory feedback
  - Inexperience of medical students
- **Limitations (noted by our group)**
  - No significant difference in damage to key anatomical structures or metrics
  - No classifier details provided
- **Further Research**
  - Improve replicability of system by providing specifications
  - Repeat study with surgical residents

Bibliography

Azari, D. P., Frasier, L. L., Quamme, S., Greenberg, C. C., Pugh, C. M., Greenberg, J. A., & Radwin, R. G. (2019). Modeling Surgical Technical Skill Using Expert Assessment for Automated Computer Rating. *Annals of surgery*, *269*(3), 574–581. https://doi.org/10.1097/SLA.0000000000002478

Wijewickrema, S., Piromchai, P., Zhou, Y., Ioannou, I., Bailey, J., Kennedy, G., & O'Leary, S. (2015). Developing effective automated feedback in temporal bone surgery simulation. *Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery*, *152*(6), 1082–1088. https://doi.org/10.1177/0194599815570880