

Laparoscopic Skill Classification using the Two-Third Power Law and the Isogony Principle

Anna French¹
Timothy M. Kowalewski¹, PhD

¹Department of Mechanical Engineering
University of Minnesota

1 Background

Surgical skill evaluation is a field that attempts to improve patient outcomes by accurately assessing surgeon proficiency. An important application of the information gathered from skill evaluation is providing feedback to the surgeon on their performance. The most commonly utilized methods for judging skill all depend on some type of human intervention. Expert panels are considered the gold standard for skill evaluation, but are cost prohibitive and often take weeks or months to deliver scores. The Fundamentals of Laparoscopic Surgery (FLS) is a widely adopted surgical training regime. Its scoring method is based on task time and number of task-specific errors, which currently requires a human proctor to calculate. This scoring method requires prior information on the distribution of scores among skill levels, which creates a problem any time a new training module or technique is introduced. These scores are not normally provided while training for the FLS skills test, and [1] has shown that FLS scoring does not lend any additional information over sorting skill levels based on task time. Crowd sourced methods such as those in [2] have also been used to provide feedback and have shown concordance with patient outcomes, however it still takes a few hours to generate scores after a training session.

It is desired to find an assessment method that can deliver a score immediately following a training module (or even in real time) and depends neither on human intervention nor on task-specific probability distributions. It is hypothesized that isogony-based surgical tool motion analysis discerns surgical skill level independent of task time.

2 Methods

2.1 Data Set

This study used tool motion data gathered from the EDGE (Electronic Data Generation and Evaluation) study [3]. This dataset contains 295 different samples of surgeons at varying skill levels interacting with a dry-lab surgical training environment performing 108 peg transfer (PegTx), 63 suturing and 124 circle cutting tasks.

Each sample is composed of a video recording of the training module (30 Hz), the cartesian space laparoscopic tool motion data corresponding to each video frame (30 Hz), features

such as task name, and an FLS skill rating. The tool tip position and velocity measurements from the tool motion data were used to calculate our features of interest for evaluating skill. The FLS score was mapped in [1] to a ternary ranking of subjects as novice, intermediate or expert. The trials utilized for this experiment include 157 FLS novices, 71 FLS intermediates and 67 experts.

2.2 Analysis Methods and Algorithms

The time-agnostic velocity gain factor γ has shown promising results. In [5], this feature was used along with the two-third power law and the Isogony Principle to relate pencil tip velocity to the radius of curvature of 2D shapes sketched by a subject as

$$v(t) = \gamma(t)k(t)^{1/3}$$

where v is the velocity of the tool tip and k is the euclidean curvature (i.e. the instantaneous radius of curvature of the tool path.)

The mean and standard deviation of γ for the left and right hand over the course of the training run was taken to create the 4 features for each trial $\sigma(\gamma_L)$, $\mu(\gamma_L)$, $\sigma(\gamma_R)$ and $\mu(\gamma_R)$. These features were used to train a state vector machine (SVM) to predict the skill level of each trial.

The accuracy of the model was evaluated based on its agreement with the FLS classification for the trial, i.e. whether each model correctly classified the trial's FLS score grouping. Trials were grouped as either novice, intermediate, or expert. The FLS score groupings are used as the ground truth data.

To test the expected accuracy of an SVM which uses $\sigma(\gamma_L)$, $\mu(\gamma_L)$, $\sigma(\gamma_R)$ and/or $\mu(\gamma_R)$ as features, a 10-fold cross validation was performed. Cross validation helps ensure that the data used for testing was not included in the model training, and thus did not bias the model. The model's score grouping prediction was compared to the FLS score grouping from the EDGE data set. Each fold of the 10-fold cross validation samples trials evenly from the three skill levels. In addition, the 10-fold cross validation was performed 10 times (creating new 10-fold sets each time) in order to average out any fluctuations in accuracy due to the particular samples chosen for each 10-fold. This means, 100 different models comprised of 100 different partitions of the data were generated, and the statistics for the accuracies of each model are communicated in the box plot figures. In this study, we were only testing discrimination between novice and expert skill levels.

In each of the box plots, the 25th and 75th percentiles are displayed as the box boundaries while the median is the central line in each box. The whiskers mark the most extreme non-outlier points, and the + points are outliers.

3 Results

Task specific models were generated as a basis for comparison to other task-specific methods. In Figure 1, the classification

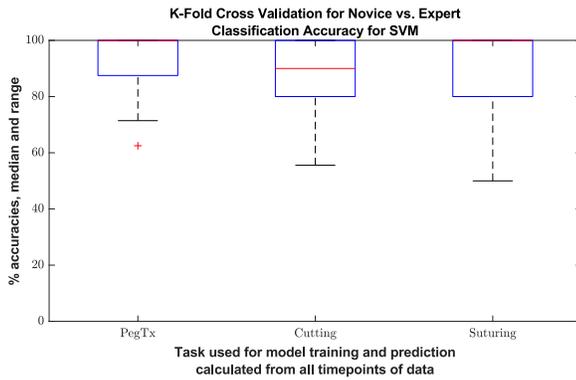


FIGURE 1: Prediction accuracy results of two-class classification, task specific trained model.

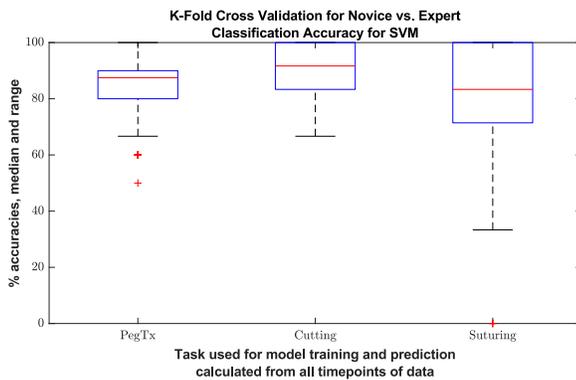


FIGURE 2: Prediction accuracy results of two-class classification, model trained in task-agnostic manner.

accuracies of task-specific models are displayed. To generate this figure, trials were separated based on which task was performed and a model to represent skill for each specific task was generated. A 10-fold cross-validation was done on each of these 3 partitions. This figure shows that for PegTx and Suturing models had a median classification accuracy of 100% per model.

In Figure 2, one model was trained using the full dataset in a task-agnostic manner (using trials across all tasks), and a similar cross validation technique to Figure 1 was performed. This figure shows good prediction accuracy for all three tasks despite the fact that the model was trained task-agnostic.

In Figure 3, five different models were trained using the full dataset in a task-agnostic manner, four of which were trained based on a single feature. The “all” column of this figure uses the same model from Figure 2. This shows the relative strength of each feature in prediction accuracy, where the accuracy of models trained using $\sigma(\gamma_R)$ or $\sigma(\gamma_L)$ alone are close in accuracy to the models trained using all four features.

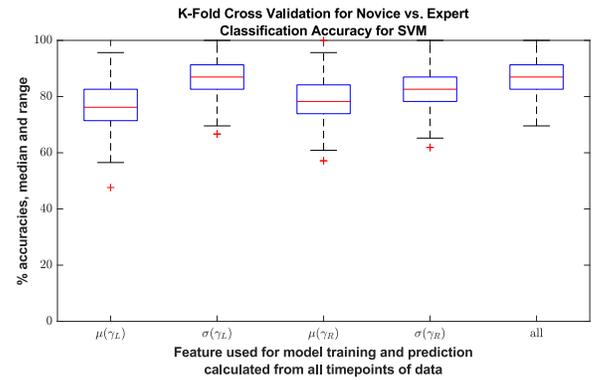


FIGURE 3: Prediction accuracy results of two-class classification, models trained using either single or all four features.

4 Interpretation

It was shown that task and time agnostic isogony-based features can be used to train automated skill evaluation models with good agreement to rough groupings of FLS scores. The ability to automate skill evaluation will allow surgeons in training to obtain feedback faster and more frequently. Further work will incorporate more trials to train more consistent models. These γ features will also be applied to other machine learning algorithms and used for 3-class classification to discern all FLS skill levels.

REFERENCES

- [1] Kowalewski, T. M., 2012. “Real-time quantitative assessment of surgical skill”. PhD thesis, University of Washington.
- [2] Kowalewski, T. M., Comstock, B., Sweet, R., Schaffhausen, C., Menhadji, A., Averch, T., Box, G., Brand, T., Ferrandino, M., Kaouk, J., et al., 2016. “Crowd-sourced assessment of technical skills for validation of basic laparoscopic urologic skills tasks”. *The Journal of urology*, **195**(6), pp. 1859–1865.
- [3] Kowalewski, T. M., White, L. W., Lendvay, T. S., Jiang, I. S., Sweet, R., Wright, A., Hannaford, B., and Sinanan, M. N., 2014. “Beyond task time: automated measurement augments fundamentals of laparoscopic skills methodology”. *Journal of Surgical Research*, **192**(2), pp. 329–338.
- [4] Rana, J., and Kowalewski, T., 2014. “Feasibility of a low-cost instrumented trocar for universal surgical procedure analyses”. *Journal of Medical Devices*, **8**(3), p. 030936.
- [5] Lacquaniti, F., Terzuolo, C., and Viviani, P., 1983. “The law relating the kinematic and figural aspects of drawing movements”. *Acta psychologica*, **54**(1-3), pp. 115–130.