Brian Gu CIS II Paper Critical Review

Robotic Path Planning for Surgeon Skill Evaluation in Minimally-Invasive Sinus Surgery

Paper Selection and Relevance

The goal of our project is to determine whether using the REMS robot to teach sinus surgery is superior to conventional free hand teaching methods. Data from our study will consist of tracked tool paths inside a cadaver head sinus using an optical tracker to track the tool. A lingering question we have had as a team is what method we will use to evaluate the data and determine which subjects have learned sinus surgery better than other subjects. The paper I have chosen offers a possible solution to this problem. This paper develops methods to quantify surgical skill using machine learning techniques and computed optimal paths within the sinus. Their methods are compared to current metrics of surgical skill, which use a team of expert surgeons to evaluate the skill of a person performing sinus surgery. If our subjects could be rated on their skill using these methods, this would be a great help to our project.

Paper Introduction

The authors seek to answer two questions in their paper. Can comparison with an optimal surgical path be used to establish a measure for skill evaluation? Does quality of motion, independent of path, provide additional information? Current evaluation of surgical skill is not automated. In order to evaluate skill, a team of expert surgeons watches a resident perform sinus surgery and rates his performance on a scale of 0-3, 0 being a novice and 3 being an expert. This test is called the OSATS. While this is accepted as the ground truth for surgical skill, the authors seek to create an computable and automatic method of skill evaluation. The authors develop two metrics on their own, called Surgical Path Planning (SPP) and Descriptive Curve Coding (DCC), as comparable alternatives to the OSATS. They compare the accuracy of these two methods to a previous method using hidden Markov Models (HMM).

Paper Experiment and Methods

The study the authors do involves 20 subjects (13 novices and 7 experts) performing Functional Endoscopic Sinus Surgery (FESS). Essentially, this involves the subjects touching three targets in the sinus of a cadaver head, namely the right maxillary sinus, the right Eustachian tube, and the left Lamina Papyracea. The paths of every subject were tracked using an electromagnetic tracker and registered to the coordinate system of the cadaver head CT. Each trial was given an OSATS score by surgeons watching the trials being performed. In order to evaluate the recorded paths, the team of authors developed two metrics called SPP and DCC. SPP computes an optimal tool path given the CT scans of the cadaver and compares the path of the subject to the computed path. A support vector machine takes in all the trial paths and grades them on a scale of 0-3 based on how close they are to the computed path. To compute the path, SPP first models the tool as a 200mm length tool, 2mm radius, with 5 degrees of freedom. They then use Probabilistic Road Mapping to find the XYZ coordinates of tip of the robot and Gradient Descent to modify the path to avoid collision with anatomy in the nose by changing the rotation of the tool. The following equations are used to apply Gradient Descent.

$$\mathcal{F}(H^t) = \sum_{j=\{skin,softtissue,bones\}} \gamma_j * Collision_j(H^t)$$

$$H^{t+1} = H^t_\alpha = H^t * H^t_{\alpha_\varphi} * H^t_{\alpha_\theta} * H^t_{\alpha_x} * H^t_{\alpha_y} * H^t_{\alpha_z}$$

Each subject’s path was compared to the computed path. This comparison was held in a feature vector containing the subject’s correlation with the computed path as well as the mean and standard deviation from the path. A Support Vector Machine (SVM) was trained to classify the paths from 0-3 (same as the OSATS).

The second metric used was DCC. DCC allows for the quantification of how smooth a subject’s path was. Essentially, the authors compute a coordinate system in the frame of an imaginary observer that travels along a given path. The algorithm computes the amount of change in this observer’s coordinate system. The algorithm uses codes that describe the change in the coordinate every timestep. Shown below are the codes that correspond to changes in the coordinate system as well as the equations that compute these codes. Cs corresponds to the codes on the left.

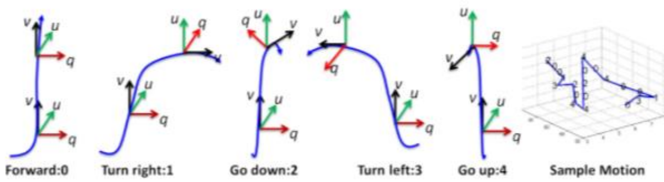


Fig. 3. DCC vocabulary and changes in direction of the attached coordinate system

$$v^s = (\|dr/ds\|^{-1})\frac{dr}{ds}$$

$$u^s = u^{s-1} \text{ if } c^{s-1} = 0, \qquad \text{otherwise } \; u^s = v^{s-1}$$

$$q^s = (\|u^s \times v^s\|^{-1})(u^s \times v^s)$$

$$c^s = \underset{\mathrm{i}}{\operatorname{argmax}}\{v^s.[\mathrm{x_i}]\}\ , \qquad \mathrm{x} = [v^{s-1}, q^{s-1}, u^{s-1}, -q^{s-1}, -u^{s-1}]$$

The output feature vector of DCC is the histogram of the codes from a given path. DCC is especially a powerful tool in this sense because the data is easily extractable and can be easily attached the SVM for SPP. This augmented SVM can output its own OSATS score given all the data from SPP as well as DCC.

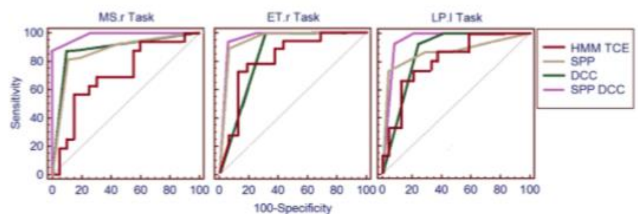
Paper Results and Conclusions

All 20 subject's tool paths were evaluated by HMM, SPP, and SPP+DCC models. Additionally, the output of these models were compared to the ground truth (OSATS score). Accuracy of these models was evaluated using a simple error rate of each trial, calculated as $Error\ rate = \left|S_{OSATS} - \hat{S}\right|/l$

Where S hat is the model score, S_{OSATS} is the actual score and l is a normalizing factor. The mean of this score was subtracted from 1 to calculate the Sim. $Sim = 1 - \mu_{\mathrm{OCER}}$ Higher Sim values means higher accuracy while lower Sim values mean lower accuracy for each model.

Below is the table of results presented in the paper.

			Skill Evaluation Method			
			HMM TCE [4]	SPP	DCC	SPP + DCC
Endoscopic Task	MS.r	Sim	61.24%	83.60%	80.46%	**87.30%**
		OCER	0.38±0.31	0.16±0.30	0.19±0.31	**0.12±0.24**
		AUC	0.70	0.87	0.88	**0.98**
		95% CI	0.53 to 0.84	0.71 to 0.96	0.73 to 0.96	**0.87 to 1.00**
	ET.r	Sim	**79.35%**	**95.19%**	**92.15%**	**97.64%**
		OCER	0.20±0.28	0.04±0.13	0.07±0.20	**0.02±0.07**
		AUC	0.81	0.95	0.84	**0.96**
		95% CI	0.65 to 0.93	0.81 to 0.99	0.67 to 0.94	**0.833 to 0.99**
	LP.l	Sim	75.38%	**85.29%**	**88.37%**	**94.70%**
		OCER	0.24±0.26	0.14±0.26	0.11±0.18	**0.05±0.12**
		AUC	0.81	0.86	0.86	**0.95**
		95% CI	0.65 to 0.91	0.71 to 0.95	0.71 to 0.95	**0.83 to 0.99**



Using Sim as the main metric for accuracy, we can see that SPP+DCC far and above was the most accurate and determining surgical skill. The conclusion that can be drawn from this is that smoothness of the path taken inside the sinus as well as how closely the path sticks to a computed optimal path (which ends up being very close to the path expert surgeons take) are very good metrics for determining the skill of a surgeon.

Assessment of Paper

I felt this paper provided some very thorough insight into the evaluation of surgical skill. This will prove to be incredibly important to the evaluation of our own data. Indeed, our own data will essentially be recorded paths of subjects in the sinus of a cadaver that must be evaluated against a predefined optimal path. While there are definitely many easier options to evaluate our data, this paper provides very mathematically rigorous methods that would allow us to make stronger inferences from our results.

Additionally, while this paper was essentially a machine learning paper, I felt it was very well written and even a novice as myself could keep up with the flow of logic in the paper, even if some of the details were not completely available to me.

A few things that I felt the paper was lacking was perhaps some more elaboration on the statistics and the results. The authors put up some very nice ROC curves at the end of the paper, but did not elaborate on exactly what the parameters of the ROC curves were. Finally, the paper assumed the reader was very familiar with a previous paper written by the same authors, the paper about the Hidden Markov Model that was used as a baseline comparison for the new models they created in this paper.