Automatic Recognition of Surgical Motions Using Statistical Modeling for Capturing Variability

Carol E. Reiley *(creiley@jhu.edu)*, Henry C. Lin (*hcl@cs.jhu.edu)*, Balakrishnan Varadarajan *(balaji.iitm1@gmail.edu)*, Sanjeev Khudanpur (*khudanpur@jhu.edu*), David D. Yuh *(dyuh@csurg.jhmi.jhu.edu),* Gregory D. Hager (*hager@cs.jhu.edu*)
Department of Computer Science, Department of Electrical and Computer Engineering and Division of Cardiac Surgery,
The Johns Hopkins University

Problem

Currently, training and evaluation of technical surgical skills is a time consuming, unstructured, subjective assessment process in which an expert surgeon watches a resident or fellow perform a set of specified tasks and provides verbal feedback. An alternative is to develop automated, objective training and evaluation techniques. 3D virtual reality (VR) training systems have been proposed for this purpose; however, it has been shown that VR technology is not advanced enough to simulate a live surgery [1, 2]. Another emerging system is procedural modeling [3], which incorporates force/torque sensors on laparoscopic instruments for ex-vivo benchtop models or simulated surgeries. However, as the number of subjects increases, it becomes a greater challenge to build robust models to account for various skill level and styles. In our work, we examine the use of user-adaptive statistical models of motion as a basis for developing automated skill evaluation.

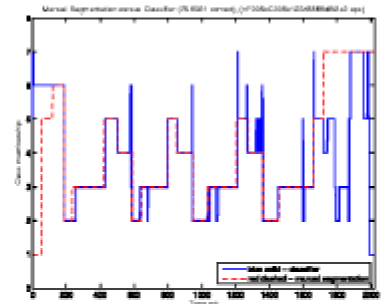
Methods

In our previous work, we automatically recognized and segmented one operator's motion during surgical tasks by identifying surgemes, elementary surgical motions, using both motion and stereo visualization. This was used to evaluate the surgical task of suturing against benchtop models [4]. In this paper, we report initial results on validating these benchtop models on a more diviser pool of users, and on data from live surgeries. Our machine learning methodology for recognizing surgical motions includes 3 components:

1. Data Collection: The orientation, position, and velocities of the robot joints and endoscope are recorded from robot kinematics at 10Hz. The stereo video data is recorded at a resolution of 640x480 at 30fps in synchrony with the kinematic data.
2. Feature Extraction: This is performed in three steps: 1) local feature extraction with temporal information, 2) feature normalization to normalize the data, and 3) linear discriminant analysis (LDA) to project the highly redundant features into a lower dimension.
3. Classification/Modeling: Maximum Likelihood Linear Regression (MLLR), a user specific modeling technique, was used to train a classifier from manually labeled data. We applied a leave one user out paradigm to test the accuracy of statistical modeling.

Results

Our study comprised 35 trials from seven subjects performing surgical suturing task on benchtop models using phantom tissue. Validation experiments were done using da Vinci surgeons and nonsurgeons on the robot-



assisted system. We applied the recognition and segmentation technique of various statistical methods including Gaussian Mixture Models, 3-state Hidden Markov Models, and supervised and unsupervised Maximum Likelihood Linear Regression (MLLR) to test the robustness of the motion recognition algorithm of a variety of users. Success was defined by comparing the accuracy of the automatically labeled data with frame by frame manually labeled data. This shows an improvement using user specific models like MLLR to account for larger data sets.

	LDA	GMM	3-state HMM	Supervised MLLR	Unsupervised MLLR
0	68.91%	67.9%	66.8%	70.4%	69.8%
1	64.09%	63.2%	64.6%	68.6%	66.5%
2	59.95%	60.4%	59.4%	61.2%	62.3%
3	67.52%	70.6%	72.8%	75.6%	75.4%
4	63.95%	67.5%	66.7%	69.3%	69.1%
5	76.82%	72.7%	71.2%	75.9%	73.1%
6	69.27%	70.2%	71.9%	75.7%	76.2%
Average	**67.21%**	**67.49%**	**67.62%**	**70.94%**	**70.34%**

We have also begun to assess the surgical motion similarity between these benchtop models and live surgery. Preliminary results of showed that the recognition algorithm learned from the benchtop model had on average slow recognition rates of 20% for suturing, 21% for needle passing, and 17% for knot tying when tested against three trials of live surgical models.

Conclusion and Discussion

Despite a (subjectively) high user variability on our new data sets, statistical modeling shows surprising robustness. With additional data mined from video, we believe much better results are possible in the very near future.

References

[1] A.G. Gallager, R.M. Satava. Virtual reality as a metric for the assessment of laparoscopic psychomotor skills. Surgical Endoscopy, 16(2): 1746-1752, December 2002.

[2] S. Waxberg, K. Goodell, D. Avgerinos, S. Schwaitzberg, C.G.L. Cao. Evaluation of physical versus virtual surgical simulators. Human Factors and Ergonomics Society, pp.1675-1679, 2004.

[3] J. Rosen, J.D. Brown, L. Chang, M. Sinanan, B. Hannaford, 'Generalized Approach for Modeling Minimally Invasive Surgery as a Stochastic Process Using a Discrete Markov Model,' IEEE Transactions on Biomedical Engineering, 53(1), pp. 399-413, March 2006.

[4] H.C. Lin, I. Shafran, D.D. Yuh, G.D. Hager. Towards Automatic Skill Evaluation: Detection and Segmentation of Robot Assisted Surgical Motions. Computer Aided Surgery, 11(5): 220-230, September 2006.