Surgical Motion Characterization in Simulated Needle Insertion Procedures

Matthew S. Holden, Tamas Ungi, Derek Sargent, Robert C. McGraw, Gabor Fichtinger
Queen's University, Kingston, Canada

ABSTRACT

PURPOSE: Evaluation of surgical performance in image-guided needle insertions is of emerging interest, to both promote patient safety and improve the efficiency and effectiveness of training. The purpose of this study was to determine if a Markov model-based algorithm can more accurately segment a needle-based surgical procedure into its five constituent tasks than a simple threshold-based algorithm. METHODS: Simulated needle trajectories were generated with known ground truth segmentation by a synthetic procedural data generator, with random noise added to each degree of freedom of motion. The respective learning algorithms were trained, and then tested on different procedures to determine task segmentation accuracy. In the threshold-based algorithm, a change in tasks was detected when the needle crossed a position/velocity threshold. In the Markov model-based algorithm, task segmentation was performed by identifying the sequence of Markov models most likely to have produced the series of observations. RESULTS: For amplitudes of translational noise greater than 0.01mm, the Markov model-based algorithm was significantly more accurate in task segmentation than the threshold-based algorithm (82.3% vs. 49.9%, $p<0.001$ for amplitude 10.0mm). For amplitudes less than 0.01mm, the two algorithms produced insignificantly different results. CONCLUSION: Task segmentation of simulated needle insertion procedures was improved by using a Markov model-based algorithm as opposed to a threshold-based algorithm for procedures involving translational noise.

Keywords: Task segmentation, Lumbar Puncture, Markov Models, Needle trajectories, Simulated procedures

1. PURPOSE

Traditionally, the foundation for training has been expert surgeons supervising trainees during real surgical procedures. Rogers et al. (2000) have shown that feedback in surgical procedure training, such as this process of supervision, is an integral part of the learning process. Supervision by an expert surgeon, however, is time consuming and lacks objectivity and standardization across supervisors (Aggarwal et al. 2007). Computer-assisted training is offered as one option for addressing some of the limitations associated with expert supervision in surgical training.

A specific medical procedure was required, as the system uses the list of workflow steps as procedure-specific input. The lumbar puncture was chosen for our experiments to test the implemented computer-assisted training methods. Lumbar puncture simulators have proven benefits and are popular amongst users (Färber et al. 2009). The methods presented are, however, general to any needle insertion procedure. A system designed to automatically provide medical trainees with immediate feedback should be able to identify workflow steps from the recorded needle trajectory such that it can provide step-by-step instruction and evaluation. Steps of a surgical procedure are referred to as tasks, and determination of which tasks are being performed in a recorded procedure timeline is called task segmentation. This task segmentation can be used to provide users with task-specific feedback and real-time instruction.

The purpose of this study was to determine if an algorithm using Markov models could more accurately segment a simulated lumbar puncture procedure than a simple threshold-based algorithm into its constituent tasks. Through implementing task segmentation in real-time, trainees can be automatically provided with feedback and instruction by the simulation system. Such a system could be used as part of a standardized computer assisted training protocol.

2. METHODS

The lumbar puncture procedure was broken down into five tasks: (1) Find insertion point – Move needle tip to entry point on skin. (2) Find insertion angle – Orient needle in the appropriate angle in the horizontal and vertical planes for insertion. (3) Insertion – Pierce skin and advance needle through the tissues leading to the spinal canal. (4) Verification – Remove needle stylet and check if cerebrospinal fluid flows. (5) Retraction – Remove needle from skin.

For algorithm testing and validation purposes, synthetic procedural data (needle trajectory in each degree of freedom) was generated by inputting a series of points. An entry point and target point defined the needle insertion plan. Intermediate points, dependent on the insertion plan, defined the transitions between tasks. A spline was applied between the points, and noise was added to each degree of freedom independently. This yielded a needle trajectory for each degree of freedom in time. The added noise was taken from a mixed Gaussian distribution with zero mean, and predefined (root-mean-square) amplitude, which was varied to test segmentation accuracy under differing conditions.

As a control to compare the algorithm using Markov models against, a simple, intuitive threshold-based algorithm was implemented. For this algorithm, each task was defined by position and velocity thresholds. If the position or velocity of the needle crossed a threshold, this corresponded to a transition in tasks. The optimal threshold values cannot be determined a priori, due to the random nature of the added noise. To determine each value in order to produce the most accurate segmentation given the ground truth segmentation of the training data, an iterative optimization scheme was applied. Each parameter was independently varied over a series of values and the segmentation was tested for optimality. The optimal parameter value was chosen, and this process was repeated for each parameter until a convergence criterion was met.

In the threshold based algorithm, each task was defined in terms of several position and velocity thresholds, each dependent on a single threshold parameter.

- Task 1. Needle-tip outside skin; needle-tip away from entry point or no rotation in needle.
- Task 2. Needle-tip outside skin; needle-tip near entry point; needle rotating.
- Task 3. Needle-tip in flesh; needle-tip moving towards target along entry-target line.
- Task 4. Needle-tip in flesh; no translational motion of needle.
- Task 5. Needle-tip in flesh; needle-tip moving away from target along entry-target line.

The purpose of the Markov model-based algorithm is to map a sequence of observations of needle motion to a sequence of discrete task labels. These task labels are derived by mapping the vector of degrees of freedom in time to the set of tasks. Mapping was done by a k-means clustering algorithm to convert the continuous observations into discrete observations. Additionally, some data processing steps are applied: an orthogonal transformation is used to process continuous data; Markov models are used to process discrete data.

In the literature, Markov models have been used in two different ways for task segmentation: (1) Tasks are described by Markov models, which govern transitions between observations (Yang et al. 1994); (2) Procedures are described by Markov models, which govern transitions between tasks (Liu et al. 2001). For the purposes here, the first type of Markov model will be called Task Markov models, and the second type will be called Procedure Markov models.

The Markov model-based algorithm developed in this study combined the two approaches. A diagram of the workflow is provided in Figure 1.

1. Observations (each degree of freedom at each time step of the tracked needle trajectory) were recorded and concatenated with previous time steps. The length of this observation record must be chosen to capture the entire motion of the needle, but not include any previous motions.
2. An orthogonal Legendre transformation was applied to the concatenated observations for each degree of freedom in order to extract feature information from each sequence of observation. This technique of feature

extraction is adapted from the handwriting recognition literature (Golubitsky & Watt 2008), in which it has been used with great success. This analysis transforms a time history of the needle trajectory as a function of time into a vector space, in which each dimension corresponds to a feature of the motion.

3. The transformed observations were separated into clusters, using the k-means algorithm. To increase the likelihood of the clustering algorithm attaining a global optimum, the fast global k-means algorithm was used (Likas et al. 2003).
4. The sequence of Task Markov models most likely to have produced the sequence of clusters was determined using the Viterbi algorithm. The transitions between Task Markov Models were governed by a Procedure Markov model. This step combines the two approaches for using Markov Models found in the literature.
5. The Task Markov model at end of the sequence was identified as the task being performed, indicating the task segmentation for the time step.

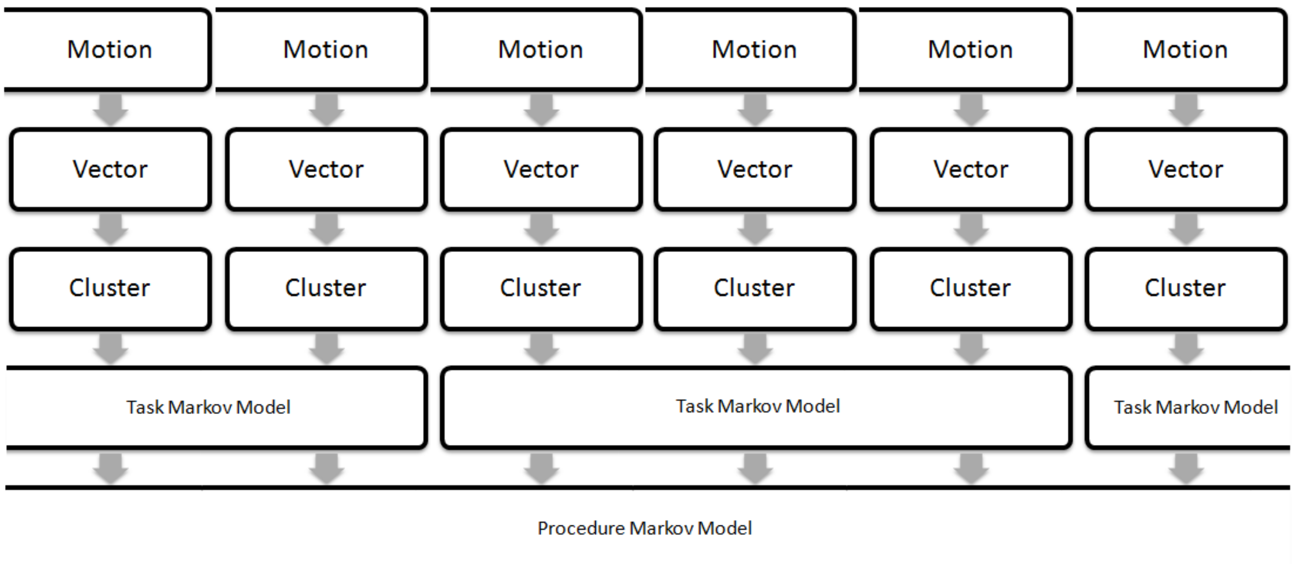

Figure 1. Illustration of workflow used in the Markov-Model based algorithm for task segmentation. The Task Markov Model is used to determine the task segmentation of the lumbar puncture procedure.

To compare the accuracy of the threshold-based algorithm with the Markov model-based algorithm, the following procedure was used:

1. Synthetic data with predefined noise amplitude was generated for training (training data set) via the synthetic data generator, using the same amplitude of noise for all procedures.
2. The segmentation algorithms were trained using the training data set (parameters were optimized for threshold-based algorithm; Markov models were trained for Markov model-based algorithm).
3. Synthetic data with predefined noise amplitude was generated for testing (testing data set) via the synthetic data generator, using the same amplitude of noise as for the training data set.
4. Task segmentation of the testing data was performed using both algorithms, and was compared to the known ground truth segmentation to determine accuracy.

Because the general workflow of the lumbar puncture procedure is well-known, two additional matrices of task transition data were used to increase the accuracy of the task segmentation. First, a matrix of transitions indicated physically impossible task transitions was used to exclude impossible task segmentations. Second, a matrix of likely task transitions was used to ensure that all task transitions that were likely (due to the nature of the procedure) had a non-zero probability of occurring. Because these matrices use only information from the workflow of the procedure, they can be readily produced for any needle-based procedure.

Additionally, a matrix indicating which task must be performed next in the procedural workflow was used to automatically provide instruction as the procedure was being analyzed. This instruction was based upon the task being performed and all of the previous tasks that had been performed.

Each training and test procedure consisted of each of the tasks performed once, in order. The length and position used for each task in the synthetic procedures was calculated from an average of experimental data of lumbar puncture procedures performed by medical trainees on a phantom. Prior to adding noise, total procedure time for generated data was 30.7s, with (root-mean-square) translational amplitude of motion 70.9mm. Accuracies were averaged over a series of ten segmentations. The difference in accuracy between the two algorithms is more indicative of task segmentation performance than the exact accuracy, as the synthetic procedural data generated is not directly comparable to the procedural data used in other studies.

Segmentation accuracy was determined as the percentage of time steps at which the automated segmentation matched the ground truth segmentation. The accuracies of the threshold-based algorithm and the Markov model-based algorithm were compared using a two-tailed sample t-test with significance level α=0.01, using the Bonferroni correction to adjust significance level for repeated statistical tests. The segmentation accuracies for each amplitude of noise were compared independently of the segmentation accuracies at other levels of noise.

3. RESULTS AND DISCUSSION

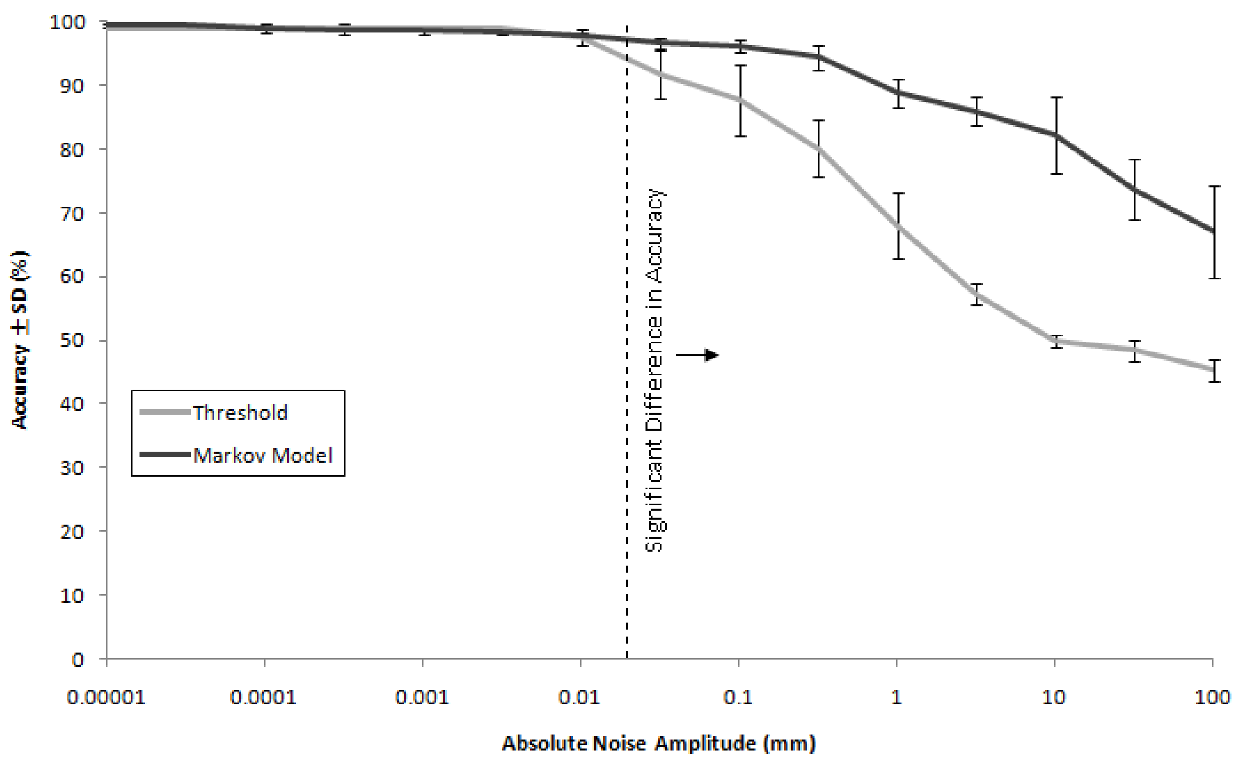

Figure 2. Task segmentation accuracy as a function of translational noise amplitude for the threshold-based algorithm (light) and the Markov model-based algorithm (dark). Area to the right of dashed line indicates significant difference (α=0.01) in accuracy between the two algorithms.

The segmentation accuracy was insignificantly different for small amplitudes of translational noise; however, for large amplitudes of translational noise (all amplitudes greater than or equal to 0.0316mm), the Markov model-based algorithm yielded significantly better results (67.1% vs. 45.3%, $p<0.001$ for maximum amplitude 100mm). The greatest difference in accuracy was for noise amplitude of 10.0mm (82.3% vs. 49.9%, $p<0.001$). In scenarios with no noise, the Markov model-based algorithm slightly outperformed the threshold-based algorithm (99.7% vs. 99.0%, $p<0.001$). Figure 2 illustrates the task segmentation accuracy of the two algorithms as a function of translational noise amplitude.

For rotational noise, the threshold algorithm was more accurate on average by 3.6%. Overall, rotational noise had little effect on either segmentation; the threshold algorithm exhibited no change in accuracy due to rotational noise, while the Markov Model algorithm only exhibited a decrease in accuracy by 2.6% between the minimum and maximum amplitudes of noise.

Typically, for small amplitudes of noise, incorrect segmentation by the Markov model-based algorithm was due to delay in recognition of a task transition, rather than an incorrect classification of a motion. This was not an issue in the threshold-based algorithm, as the task segmentation was independent of previous observations. This observation is reflective of the fact that tasks are not completely disjoint, task transitions may take place over several time steps and multiple tasks may be performed simultaneously.

Results are consistent with the work of Ahmidi et al. (2010) in demonstrating that Markov models are an effective tool in identifying tasks in procedural records. Although, the task segmentation here differs from the laparoscopic task classification in the work of Ahmidi et al., in which task transitions were explicitly indicated by pausing for several seconds, and thus, the results are not directly comparable.

Further work may involve implementation of data processing steps in Markov model-based algorithm to extract the information from the orthogonally transformed data, and reducing the noise. Such processing steps include principal component analysis and linear discriminant analysis.

Additionally, further work in simulating human noise will test the algorithm under conditions more similar to those encountered in data from real needle-based procedures. Human noise has been successfully modeled in many ways (Tu et al. 2007). The Gaussian noise used here does not reflect the noise actually exhibited by human surgeons; however, testing the Markov model-based algorithm using Gaussian noise verifies its accuracy under non-ideal conditions. Modeling human noise will be useful to further verify the feasibility of the Markov-model based algorithms for task segmentation of real procedures.

4. CONCLUSION

An algorithm using Markov models for task segmentation of a lumbar puncture procedure produced significantly more accurate results than a simple, intuitive algorithm using position and velocity thresholds for translational noise. Additionally, the Markov Model based algorithm has the advantage of greater flexibility than the threshold-based algorithms for implementation on different needle-based procedures.

This study has verified the feasibility of a Markov model-based algorithm for task segmentation of real needle-based surgical procedures. This analysis has potential application for providing instruction and feedback to medical trainees in standardized computer assisted training protocols.

5. ACKNOWLEDGEMENTS

M. S. Holden was supported by the NSERC CREATE Training Program in Bone and Joint Health Technologies. T. Ungi was supported by the Queen’s University-Ministry of Research & Innovation Ontario Post-Doctoral fellowship. G. Fichtinger was funded as a Cancer Ontario Research Chair.

REFERENCES

[1] Aggarwal R., Grantcharov T. P., Darzi A., “Framework for systematic training and assessment of technical skills,” J Am Coll Surg 204(4): 697-705 (2007).
[2] Ahmidi N., Hager G. D., Ishii L., Fichtinger G., Gallia G. L., Ishii M., “Surgical task and skill classification from eye tracking and tool motion in minimally invasive surgery,” Med Image Comput Comput Assist Interv 13(3): 295-302 (2010).
[3] Färber M., Hummel F., Gerloff C., Handels H., “Virtual reality simulator for the training of lumbar punctures,” Methods Inf Med 48(5): 493-501 (2009).
[4] Golubitsky O., Watt S. M., “Online stroke modeling for handwriting recognition,” Proc. 18th Annual International Conference on Computer Science and Software Engineering 1705-7345 (2008).
[5] Likas A., Vlassis N., Verbeek J., “The global k-means clustering algorithm,” Pattern Recognition 36(2): 451-461 (2003).
[6] Liu A., Kaufmann C., Tanaka D., “An Architecture for Simulating Needle-Based Surgical Procedures,” Med Image Comput Comput Assist Interv Proceedings 1137-1144 (2001).
[7] Rogers D. A., Regehr G., Howdieshell T. R., Yeh K. A., Palm E., “The impact of external feedback on computer-assisted learning for surgical technical skill training,” Am J Surg 179(4): 341-343 (2000).
[8] Tu K., Thornburg H., Fulmer M., Spanias A., “Tracking the path shape qualities of human motion,” IEEE Acoustics, Speech and Signal Processing 2: 781-784 (2007).
[9] Yang J., Xu Y., Chen C., “Hidden Markov model Approach to Skill Learning and Its Application to Telerobotics,” IEEE Transactions on Robotics and Automation 10(5): 621-631 (1994).