Robotic Surgical Procedure Identification

Jason Driver and Kyle Lindgren
University of Washington Department of Electrical Engineering

INTRODUCTION

In our world of moving objects, there is a need for machines to be able to identify motions in the world such that intelligent decisions can be made. In the new field of robotic surgery, there have been many innovations within the last decade, and while the field is advancing, there are still limitations to overcome. Studying human motion is important for robotics. An example of three reasons identification is important are: gain insight into how humans learn complicated motion skills, how to objectively assess tasks using electronic systems, and how to automate complex human motions so that machines can perform those tasks at the level of our most skilled professionals. Our project is designing and implementing a surgical procedure identifier. This vision software tool will take video input of procedures performed by a surgical robot like the Da Vinci, Intuitive Surgical, Inc. and identify the procedure as being either suturing or knot-tying. The corresponding paper from Johns Hopkins University is entitled "*The JHU-ISI Gesture and Skill Assessment Working Set (JIGSAWS): A Surgical Activity Dataset for Human Motion Modeling*" by Gao et al.

BACKGROUND

The Computational Interaction and Robotics Laboratory at Johns Hopkins University has a dataset called JIGSAWS. JIGSAWS is a surgical activity dataset open to the public (http://cirl.lcsr.jhu.edu/jigsaws) for human motion modeling, and was collected by Johns Hopkins University and Intuitive Surgical Inc. The da Vinci Surgical System was operated by eight separate surgeons with varying levels of skill, who performed five repetitions of two simple surgery tasks: suturing and knot-tying. Two surgeons had over 100 hours of robotic surgery experience, four surgeons had less than 10 hours, and 2 surgeons had between 10 and 100 hours of experience. The video data was recorded using a static endoscopic camera at 30 Hz, 640 x 480 resolution. The JIGSAWS dataset has been used in several other studies for surgical activity recognition and skill assessment. The Computational Interaction and Robotics Laboratory had previously used a multiple kernel learning framework with linear dynamical system and bag-of-features to analyze the video data, but this was not released at the time of our project to compare results [1]. Previous work on the dataset used three different approaches: LDS for modeling video clips corresponding to gestures, bag-of-features approach on spatio-temporal features extracted from videos, and multiple kernel learning to combine the LDS and BoF approaches. With our project, we explore a Hidden Markov Model approach.

METHODS

Feature detection steps

- Create the video data set using the Johns Hopkins University JIGSAW dataset (640 x 480, 30 Hz., static endoscopic camera, 2 minute average)
- Input Video, identify and label corners
- Extract x-position and y-position of the two most pronounced corner clusters
- Input cluster x-y data into Hidden Markov Model
- Classify x-y data motion

Harris Corner Detector Method

The Harris corner detector looks at the first derivatives in the x and y direction, and the second derivative in the x and y directions to find a corner response. The two strong eigenvalue responses indicate a corner, while only one strong eigenvalue would be an edge, and no strong eigenvalues would not be a corner or edge.

Clustering Method

subclust(x,radii) uses subtractive clustering

- Selects the data point with the highest potential to be the first cluster center
- Removes all data points in the vicinity of the first cluster center (as determined by radii), in order to determine the next data cluster and its center location
- Iterates on this process until all of the data is within radii of a cluster center

Hidden Markov Model

Hidden Markov Model visualization with states represented as X#, observations y#, state transition probabilities a##, and emission probabilities b## shown right in Figure 3 [2].

Steps:

- Train using clustered motion data of robot end effectors
- Using state probability and emission probability data, classify unseen videos as video of knot tying or suturing surgical procedure

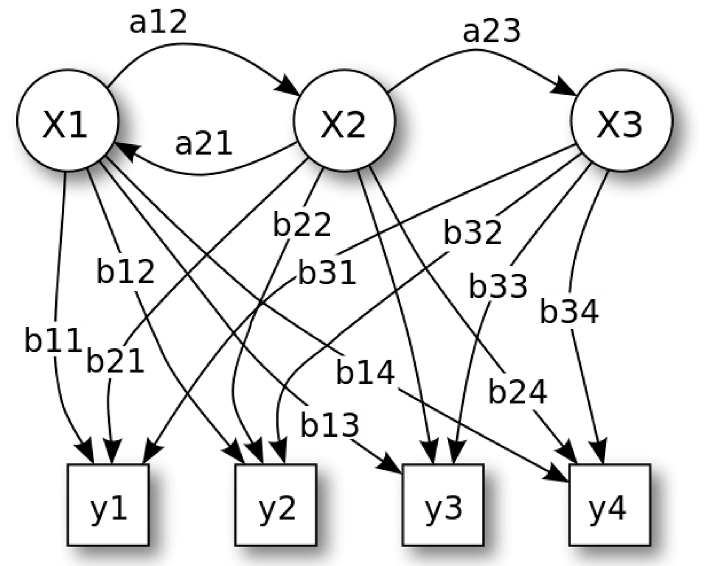

Fig. 3. Hidden Markov Model visualization

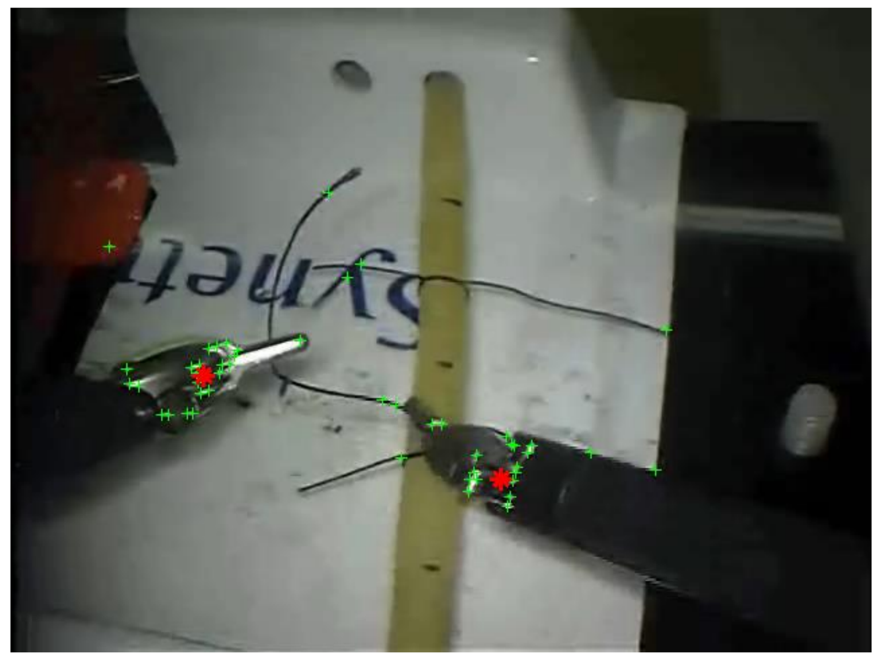

Fig. 1. Harris corner response (green) with clusters (red) on a sample knot-tying frame.

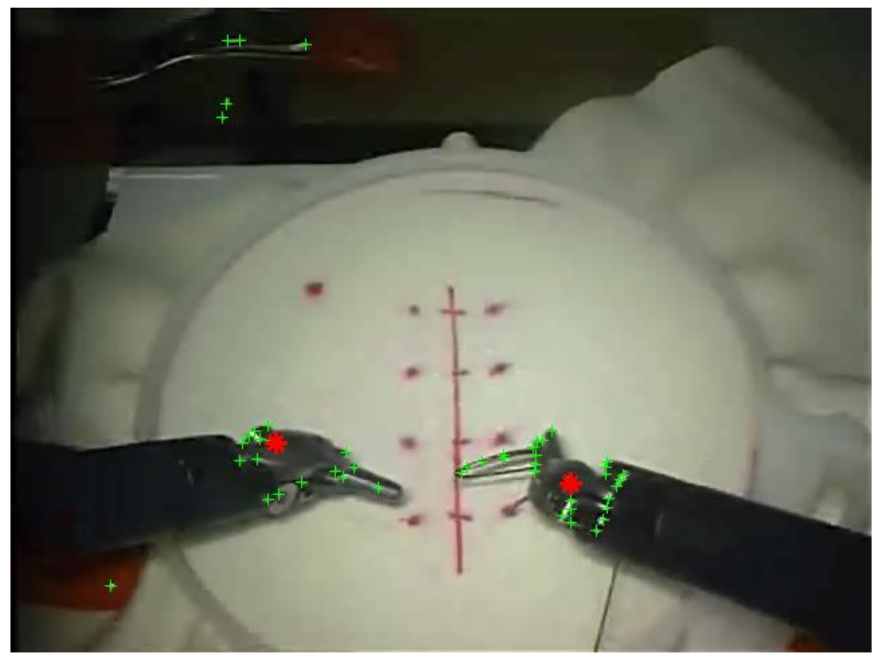

Fig. 2. Harris corner response (green) with clusters (red) on a sample suturing frame.

References

[1] Yixin Gao, S. Swaroop Vedula, Carol E. Reiley, Narges Ahmidi, Balakrishnan Varadarajan, Henry C. Lin, Lingling Tao, Luca Zappella, Benjamín Béjar, David D. Yuh, Chi Chiung Grace Chen, René Vidal, Sanjeev Khudanpur and Gregory D. Hager, *The JHU-ISI Gesture and Skill Assessment Working Set (JIGSAWS): A Surgical Activity Dataset for Human Motion Modeling*, In Modeling and Monitoring of Computer Assisted Interventions (M2CAI) – MICCAI Workshop, 2014

[2] "Hidden Markov model," Wikipedia, the free encyclopedia. 20-Apr-2016.

RESULTS

A confusion matrix detailing our results is shown below in Figure 4. A confusion matrix with a strong diagonal magnitude indicates good results. The choice to use an averaging filter on the coordinate space can be seen in Figure 5 below to limit the effect of outliers for each tracked cluster and to reduce noise.

State 1	Predicted Class 1	Predictive Class 2	Predictive Unidentified Class
Actual Class 1	4	2	1
Actual Class 2	3	5	0

Fig. 4. Confusion matrix with knot-tying as state 1 and suturing as state 2.

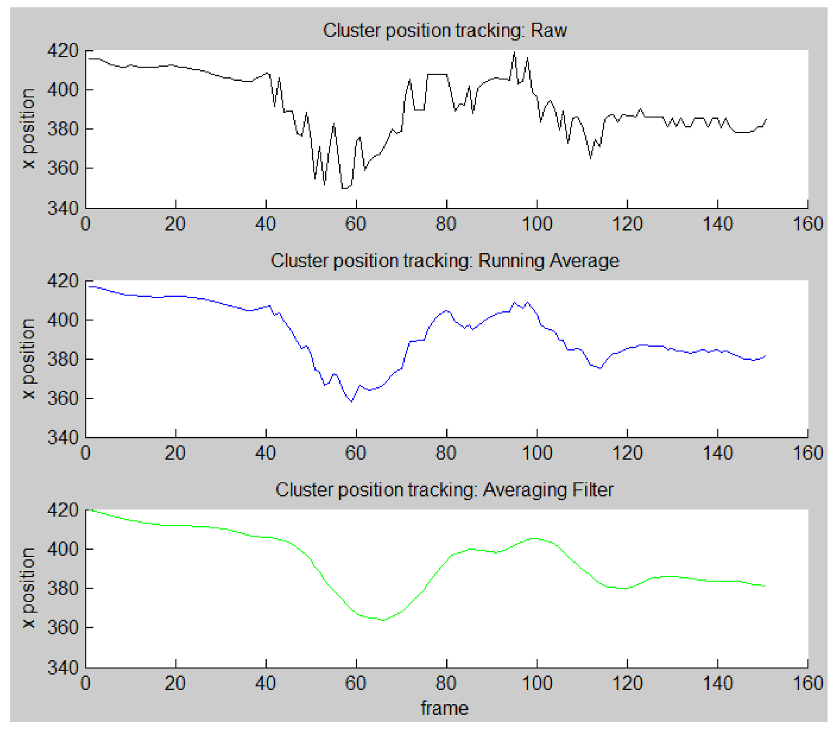

Fig. 5. Cluster position tracking: Raw, Running Average, and Filtered.

DISCUSSION

In this work we have shown object detection, motion tracking, and motion identification can be achieved with limited data and established computer vision techniques. Obtaining corners using the Harris corner method and then clustering these using subtractive clustering allowed us to identify with reasonable accuracy the two robotic end effectors in each frame. The x and y coordinates of our found end effectors in each video frame are used to track their trajectories through the videos. These trajectories are then used to train our Hidden Markov Model which produces the state transition probabilities and emission probabilities used to predict the state (procedure) of unseen data.

Our results are significant as they show success in procedure identification using a limited set of training data and only video input. Robustness of the methods chosen are evident by the variation of surgeon skill in the procedure data and below average video recording quality.

CONCLUSIONS

The robotic surgery classifier using single point camera image data works using a Hidden Markov Model but there are limitations in accuracy and precision. In the future, we plan on developing a system that would rate different surgery environments to know which feature detection method would be best suited for that particular video environment. This would greatly help reduce unwanted features before processing the data through the learning model.