The International Journal of **Medical Robotics** and **Computer Assisted Surgery**

THE INTERNATIONAL JOURNAL OF MEDICAL ROBOTICS AND COMPUTER ASSISTED SURGERY
Int J Med Robotics Comput Assist Surg 2017; 13: e1766.
Published online 19 August 2016 in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/rcs.1766

ORIGINAL ARTICLE

Distance-based time series classification approach for task recognition with application in surgical robot autonomy

Mahtab J. Fard[1]*
Abhilash K. Pandya[2]
Ratna B. Chinnam[1]
Michael D. Klein[3]
R. Darin Ellis[1]

[1]*Department of Industrial and Systems Engineering, Wayne State University, Detroit, MI, USA*

[2]*Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI, USA*

[3]*Department of Surgery, Wayne State University School of Medicine and Pediatric Surgery, Children's Hospital of Michigan, Detroit, MI, USA*

*Correspondence to: Mahtab J. Fard, Department of Industrial and Systems Engineering, Wayne State University, 4815 Fourth St., Detroit, MI 48202, USA.
E-mail: fard@wayne.edu

Accepted: 12 July 2016

Abstract

Background Robotic-assisted surgery allows surgeons to perform many types of complex operations with greater precision than is possible with conventional surgery. Despite these advantages, in current systems, a surgeon should communicate with the device directly and manually. To allow the robot to adjust parameters such as camera position, the system needs to know automatically what task the surgeon is performing.

Methods A distance-based time series classification framework has been developed which measures dynamic time warping distance between temporal trajectory data of robot arms and classifies surgical tasks and gestures using a k-nearest neighbor algorithm.

Results Results on real robotic surgery data show that the proposed framework outperformed state-of-the-art methods by up to 9% across three tasks and by 8% across gestures.

Conclusion The proposed framework is robust and accurate. Therefore, it can be used to develop adaptive control systems that will be more responsive to surgeons' needs by identifying next movements of the surgeon. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords task and gesture recognition; robotic surgery; automatic camera control; time series classification; dynamic time warping; k-nearest neighbor; distance-based classification

Introduction

Surgery is continuously subject to technological innovations including the introduction of robotic surgical devices (1). Advances in robotic minimally invasive surgery (RMIS) have the potential to improve patient outcomes by shorter hospital stays, quicker recovery time and less chance of infection (2). The ultimate goal of RMIS is to program the surgical robot to perform certain difficult or complex surgery in an autonomous manner. However, there is no technical roadmap to a fully autonomous surgical system at the present time (3,4). Current RMIS systems operate in a master–slave mode, relying exclusively on direct surgeon input (5). For example, camera control in current

RMIS platforms is an additional task under direct control of the surgeon. In the current FDA-approved system, the da Vinci surgical platform (Intuitive Surgical, Sunnyvale, CA, USA) (6), many interface parameters are set once and remain at the same level throughout the operation while different surgical tasks and motions may require different camera behaviors (4). For instance, it has been shown (4) that a wide view is desirable for the looping phase while the view should be narrow when grabbing the free end of the suture in a knot tying task. Hence, the surgeon must stop the procedure to move the camera or change the zoom level. This can distract the surgeon from the smooth flow of the operation, and certainly adds time to the procedure. Therefore, to reduce the workload and improve the surgeon's field of view, an automatic camera control system is desired.

It is, however, quite clear that to develop any automatic control system, a more detailed comprehension of the surgical procedures is needed (7). On the one hand, the feasibility of current robotic surgery systems to record quantitative motion and video data motivates the development of descriptive mathematical models to recognize and analyze surgical tasks. On the other hand, recent advances in machine learning research for uncovering concealed patterns in huge data sets, like kinematic and video data, offer the possibility to better understand surgical procedures from a system point of view. Surgical tasks and at a more granular level, surgical gestures need to be quantified to make them amenable to further study in autonomous surgical system (8).

To answer this query, we develop a distance-based time series classification method by integrating a dynamic time warping (DTW) (9) distance measure with k-nearest neighbor (kNN) classification method (10) to recognize and classify surgical tasks and gestures. Figure 1 summarizes the proposed classification framework. We evaluate the performance of our proposed method on real robotic surgery data where we focus on three important RMIS tasks: knot tying, needle passing and suturing, which are all part of a fundamentals of laparoscopic surgery (FLS) skills training program (11). Results show that the DTW-kNN framework is fast, accurate and robust, all of which makes it applicable for any adaptive control system in robotic surgery.

The rest of the paper is organized as follows. In the following section we offer background knowledge and related works in two domains: surgical task and gesture recognition techniques and time series classification methods. Then we explain the experimental data and introduce our DTW-kNN framework. The results are provided and discussed in the 4th and 5th sections. Conclusions and future work are discussed in the final section.

Background

In recent years, understanding and recognizing surgical procedures at different levels of granularity has been the focus of research (12–14). Surgical procedures can generally be broken down to four main levels, from higher to lower: phases, steps, tasks and gestures (motions) (7). At the higher level, statistical models have been proposed using recorded force and motion data (15), surgical tools usage (16) and video data (17) to classify surgical phases. At a more granular level, effort has been applied to identify and classify surgical gestures based on kinematic and video data using techniques such as Linear Discriminant Analysis (LDA) (18,19), Linear Dynamic Systems (LDS) (20,21), hidden Markov models (HMM) (22–25) and extensions of HMM (26,27). These techniques are categorized as feature-based time series classification methods (28) where the important features need to be extracted from temporal sequence of surgical tasks using techniques such as Fourier transformation (22,25). Therefore, performance of feature-based methods depends strongly on the quality of extracted features. Despite the fact that these methods have the ability to find the underlying structure of RMIS tasks, they suffer from common drawbacks. They are time consuming, require significant human interaction and preprocessing, and lack robustness due to the requirements of parameter estimation and tuning for high dimensional data (27). These make them

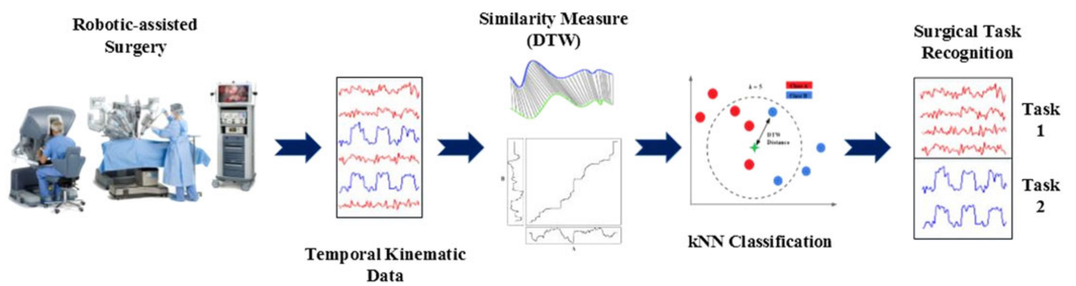

Figure 1. Proposed framework consisting of two steps: similarity measurement between temporal sequence of surgical task and classification using k-nearest neighbor method

DOI: 10.1002/rcs

impractical for automatic control system in robotic surgery where a robust, fast and accurate classifying method is needed.

We address these challenges by developing a distance-based (also known as shape-based) time series classification framework (29). The proposed method does not need any hand-crafted features, instead it works directly on raw kinematic data captured from tool tip position during robotic surgery. The well-known distance-based classifier is a k-Nearest Neighbors (kNN) algorithm (10) which has been empirically proven to be very accurate, efficient and difficult to beat in the time series classification domain (28,30). For this purpose, distance between two temporal sequences needs to be carefully defined to reflect the similarity of data. Among many distance measurement techniques (30), Dynamic Time Warping (DTW) (9) is the most popular for time series data. It has been shown to be the best similarity measurement in many domains (30–33). Contrary to Euclidean methods, where the point to point distance between two sequences is calculated (34), DTW can align time series with different length and distortion to measure distance accurately. Thus, the combination of kNN and DTW could result in a robust classification framework with high accuracy and minimum data preprocessing, all of which make the applicable of the proposed framework feasible for any adaptive control system, such as a camera, in robotic surgery.

Materials and Methods

We apply our proposed method on real robotic surgical data presented in (35). The data comprise of eight surgeons who performed around five trials of different surgical tasks (Figure 2). For each trial, we analyze temporal kinematic data captured using the API of the da Vinci at 30 Hz. Data consist of 19 features for each left and right patient-side robotic arm (38 features in total): 3 Cartesian positions, a rotation matrix consisting of 9 variables, 3 linear velocities, 3 angular velocities and a gripper angle. For task classification, we use only Cartesian position data (x, y, z) of both hands (6 variables) while for surgical gestures recognition, which is more challenging compared with task recognition, all 38 variables are used. The start and end time for each gesture is also provided in the dataset. Table 1 lists gestures and their descriptions for all three tasks (35). It is worth mentioning that although from Table 1, gesture labels are the same for suturing and needle passing, actual gesture content at the atomic sub-task level varies with task context. As an example, 'pushing needle through tissue' in suturing means passing needle through a hole in a suture box from up to down while this gesture in needle passing means passing needle through a metal hole from left to right (Figure 2).

Figure 2. Snapshot of the three fundamental RMIS tasks (35)

The aim of this work is to classify robotic surgery task and gesture based on pre-labeled kinematic data. The proposed classification framework consists of two key components: (1) measuring similarity between different surgical tasks and gesture; and (2) classification based on the k-nearest neighbor algorithm (Figure 1). In the following sections, each component will be discussed in detail.

Surgical task similarity measure

The choice of method for measuring (dis)similarity is a critical step in achieving valid classification results and in the context of time series data, different similarity measure have been developed (30). Our framework is based on similarity of the overall shape of two temporal sequences by directly comparing their individual point

Table 1 Gesture description for suturing, needle passing and knot tying

Index	Gesture description	Suturing	Needle passing	Knot tying
G1	Reaching for needle with right hand	✓	✓	✓
G2	Positioning needle	✓	✓	
G3	Pushing needle through tissue	✓	✓	
G4	Transferring needle from left to right	✓	✓	
G5	Moving to center with needle in grip	✓	✓	
G6	Pulling suture with left hand	✓	✓	
G8	Orienting needle	✓	✓	
G11	Dropping suture at end and moving to end points	✓	✓	✓
G12	Reaching for needle with left hand			✓
G13	Making C loop around right hand			✓
G14	Reaching for suture with right hand			✓
G15	Pulling suture with both hands			✓

values (36). To have a meaningful comparison, each temporal sequence of a surgical task needs to be normalized to have a mean of zero and a standard deviation of one.

One of the simplest ways to measure similarity between two sequences is the Euclidean distance (34). However, despite the simplicity and efficiency of this method, which makes it the most popular distance measure, it requires both input sequences to have the same length. In addition, Euclidean distance is sensitive to distortions, e.g. shifting, noise, and outliers. If, for instance, two time series are similar, however slightly out of phase with one another, then the Euclidean distance will give an extremely poor similarity measure (Figure 3). In order to handle this problem, warping distances such as Dynamic Time Warping (DTW) have been proposed to search for the best alignment between two time series (9). Figure 3 shows an intuitive representation of DTW versus Euclidean distance. From the figure, point i from time series A is aligned to the same point in time series B in Euclidean distance measurement. While for DTW, a nonlinear alignment of these two time series produce a more intuitive similarity measure, where point i is aligned to point $i+1$.

Consider two time series $\boldsymbol{A} = \boldsymbol{a}_{1\times m}, \ \ldots, \boldsymbol{a}_{i\times m}, \ldots, \boldsymbol{a}_{p\times m}$ and $\boldsymbol{B} = \boldsymbol{b}_{1\times m}, \ \ldots, \boldsymbol{b}_{i\times m}, \ldots, \boldsymbol{b}_{q\times m}$ with $p\times m$ and $q\times m$ dimension, respectively, where p and q refer to length of sequences and m represent the number of features. Two sequences can be arranged as a $p\times q$ matrix of the sides of a grid in which the distance between every possible combination of time instances $\boldsymbol{a}_i$ and $\boldsymbol{b}_j$ is stored (Figure 4). To find the best match between two sequences, a path through the grid that minimizes the overall distance is needed. This path can be efficiently found using dynamic programming (9) as follows

$$d(\boldsymbol{a}_i, \boldsymbol{b}_j) = d_{EUC}(\boldsymbol{a}_i, \ \boldsymbol{b}_j) + min\begin{cases} d(\boldsymbol{a}_{i-1}, \boldsymbol{b}_j) \\ d(\boldsymbol{a}_i, \boldsymbol{b}_{j-1}) \\ d(\boldsymbol{a}_{i-1}, \boldsymbol{b}_{j-1}) \end{cases} \tag{1}$$

where $d_{EUC}(\boldsymbol{a}_i, \ \boldsymbol{b}_j)$ is the Euclidean distance between the ith point of sequence A and jth point of sequence B which can be calculated as

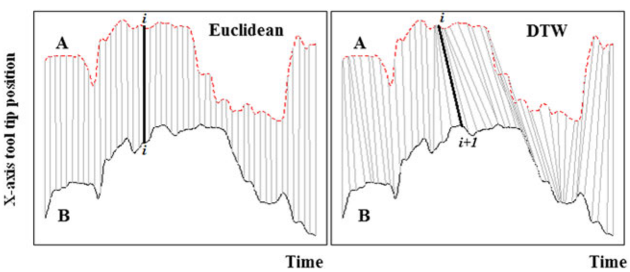

Figure 3. Comparison between Euclidean distance and DTW of X-axis da Vinci tool tip position for two time series sample data

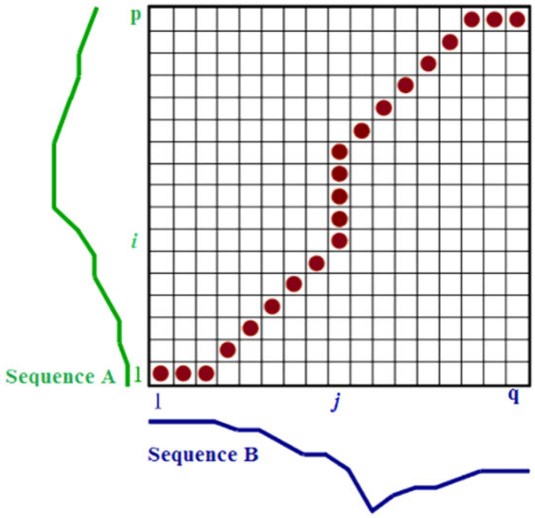

Figure 4. Illustrative example for temporal sequence alignment using DTW

$$d_{EUC}(\boldsymbol{a}_i, \ \boldsymbol{b}_j) = \sqrt{\sum_{l=1}^{m}(a_{i,l} - b_{j,l})^2} \tag{2}$$

Therefore overall Dynamic Time Warping distance between two sequences is

$$DTW(\boldsymbol{A}, \boldsymbol{B}) = d(\boldsymbol{a}_p, \boldsymbol{b}_q) \tag{3}$$

DTW has been shown to be an appropriate choice for time series classification problems of high dimensions (32), hence, it can be used for real-time surgical task and gesture recognition.

Surgical tasks classification

After measuring the distance between each pair of sequences in the dataset, the subsequent step is classification based on their distance. We use one of the most common distance-based classification method called k-nearest neighbors in our framework. kNN is a non-parametric method, which means it does not make any assumptions about the underlying data distribution. In addition, kNN does not have an explicit training phase or in other words, it has low training burden (lazy learner). During the classification phase, the majority vote of the k closest distance neighbors for each point is computed. Then, the label for the query point is assigned based on the most representatives within the nearest neighbors of the point. Figure 5 illustrates the kNN algorithm for k = 5.

For kNN, the only parameter that needs to be provided is k. In general, a small value of k means that any noise present in the data will have an influence on the result;

DOI: 10.1002/rcs

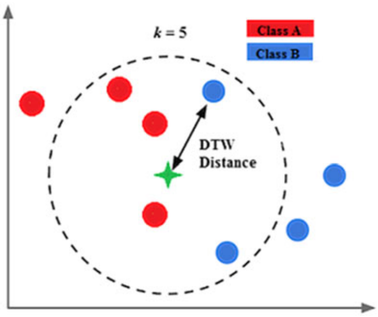

Figure 5. Illustrative example for 5-NN classification. Consider a training set consisting of two classes (A red dots and B blue dots) with four instances apiece. Suppose we want to classify an unlabeled observation, indicated by the green star. The class for new data can be assigned by a majority vote of the k nearest neighbors based on their distance to the green star. For the case k = 5 (dashed circle line), three neighbors are of Class A and two are of Class B, so we classify the unlabeled observation as a member of A

however, a large value for k allows samples of other classes to be included in the neighborhood of the test data, resulting in poor classification and high computational expenses. In order to find the best value for k, we try different values in the range 1 to 10 and the best classification accuracy is reported in the result section. The balance of simplicity on one hand and accuracy on the other hand make the kNN the best candidate for our time series RMIS task and gesture classification framework.

Performance evaluation

The accuracy of the proposed DTW-kNN framework was compared with that of different state-of-the-art methods. To classify the surgical task followed by (25), we applied HMM on features extracted from a short time Fourier transform of the Cartesian position variable of both hands in the kinematic data. For surgical gesture recognition we compared the performance of the proposed DTW-kNN framework with a sparse hidden Markov Model (27) and a Linear Dynamic System (21). The result of our proposed classification framework is directly comparable with these state-of-the-art methods since we applied the same dataset as explained in (35).

We used two model validation schema suggested by (35). The first is leave-one-super-trial-out (LOSO), where one trial for each of the surgeons is left out for testing. The second is leave-one-user-out (LOUO), where we leave all the trials from one surgeon out for testing. While the first evaluates the robustness of the method across repeating tasks by leaving out one trial for all surgeons, the second schema evaluates the robustness of the method when a surgeon was not previously seen in the training data. The performance of the different task recognition methods was determined by mean over all iteration classification accuracy, expressed in terms of percentage of subjects in the test set that are classified correctly.

Results

The performance evaluation of the proposed framework will be presented in this section.

Surgical task recognition

For the three RMIS tasks, suturing, needle passing and knot tying, the DTW measures the pairwise distance between three Cartesian position of tool tips of patient side arm of robot for both right and left hands. Then, the kNN classification method was applied to recognize different tasks based on the DTW distance measurement. We test different values for k and our preliminary results show that the accuracy of the proposed model is robust to the values of k in the range 3 to 7 and we report the result achieved for k = 5. The performance of the DTW-kNN framework for each task is compared with HMM in Table 2. The results show that the proposed method outperformed HMM by 6% on average across the three tasks. It also shows that for LOSO, 100% of suturing, 89.3% of needle passing and 97.2% of knot tying are correctly classified. For LOUO validation schema, the correctly classified suturing is 87.6%, needle passing is 85.7% and knot tying is 95.8%.

Real-time task classification

To check whether the proposed method has the potential to be used for real-time task recognition, we ran an experiment where the complete temporal sequence of the task was not used. Instead, we applied our model to the first x% of the total time series signals for each task and evaluated the performance of the method. Figure 6 shows the result for the two different validation methods. In LOSO, having 5% of the complete temporal sequence, the model was able to recognize knot tying at 96.2%, needle passing at 82.4% and suturing at 86.1% accuracy. In addition, for LOUO we need longer temporal sequences to be able to recognize these tasks with the same accuracy as we get at 5% in LOSO.

Surgical gesture recognition

We also applied the DTW-kNN method at a more granular level to recognize surgical gestures for each task

DOI: 10.1002/rcs

Table 2 Comparison between classification accuracy for each task using HMM and proposed DTW-kNN for LOSO and LOUO model validation (with standard deviation). The best classification performance is highlighted in bold

	LOSO		LOUO	
	HMM	DTW-kNN	HMM	DTW-kNN
Suturing	96.4% ± 4%	**100% ± 0%**	80.7% ± 14%	**87.6% ± 11%**
Needle passing	83.5% ± 8%	**89.3% ± 7%**	80.8% ± 12%	**85.7% ± 9%**
Knot tying	**97.3% ± 3%**	97.2% ± 3%	90.9% ± 9%	**95.8% ± 8%**

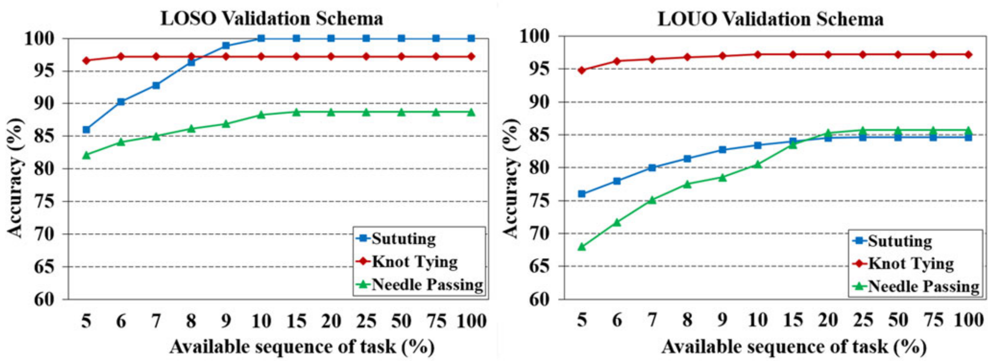

Figure 6. Accuracy of task recognition based on different percentages of total temporal sequence of task for LOSO and LOUO model validation

separately. Different values for k were tested and the best accuracy was achieved for k = 5, while the result is robust for values of k in the range 3 to 9. In Table 3, we compare the results of the sparse Hidden Markov Model (SHMM), Linear Dynamic System (LDS) as presented in (21) and proposed DTW-kNN for each task using LOSO and LOUO model validation. We should note that the standard deviation is provided only for the proposed DTW-kNN since no value is reported for the other two methods in (21). From Table 3 the results show that the DTW-kNN method outperformed other state-of-the-art models. It can recognize gestures in suturing with 86.9%, needle passing with 79.9% and knot tying with 88.3% accuracy for LOSO. On the other hand, for LOUO the accuracy degraded to 80.4%, 70.1% and 85.1% for suturing, needle passing and knot tying respectively. Figure 7 summarizes the performance of the DTW-kNN framework for each gesture. For example in LOSO, gesture G2 (positioning needle) is recognized with 90.1% accuracy for suturing and 79.2% accuracy for needle passing.

Discussion

From the results, we observe that classification accuracy decreases when switching from the LOSO validation schema to LOUO. The LOUO results provide insight into the generalizability of the algorithms to recognize task performed by surgeons that were unseen during the training phase. Table 2 shows that knot tying has less degradation (around 3%) when switching from LOSO to LOUO compared with needle passing and suturing which drop by around 10%. The simplest explanation is that knot tying is very different from the other two tasks and therefore, it can be easily recognized, regardless of the amount of variability that exists between surgeons. However, such a difference among tasks can also suggest that surgeons possibly perform knot tying in a more similar way while suturing and needle passing are performed differently. The higher performance of LOSO compared with LOUO also indicates that a short calibration procedure could be conducted when the surgeon starts using the system for the first time, and this might improve the ability of the algorithm to detect the correct task. It is worth mentioning that from Table 3, DTW-kNN outperformed the state-of-the-art methods by 2% on average across 12 gestures for LOSO and by 7% for LOUO, which shows that our proposed method is more robust compared with them.

Another interesting thing to remark upon is the potential of the proposed model to be used for real-time task recognition. For example, for knot tying, having only 5% of the complete temporal sequence, the model was able to recognize the task with 96.2% accuracy. In the dataset, the average time for knot tying, needle passing and suturing are 57, 110 and 120 s, respectively, which means that almost all of the task can be recognize with

DOI: 10.1002/rcs

Table 3 Comparison between accuracy of surgical gesture classification results of SHMM, LDS and proposed DTW-kNN (with the standard deviation) for each task using LOSO and LOUO model validation. The best classification performance is highlighted in bold

	LOSO			LOUO		
	SHMM	LDS	DTW-kNN	SHMM	LDS	DTW-kNN
Suturing	79.40%	**87.30%**	86.9% ± 3%	60.80%	74.60%	**80.4% ± 6%**
Needle passing	76.40%	78.80%	**79.9% ± 4%**	45.30%	67.30%	**70.1% ± 8%**
Knot tying	86.80%	85.10%	**88.3% ± 2%**	72.00%	78.90%	**85.1% ± 3%**

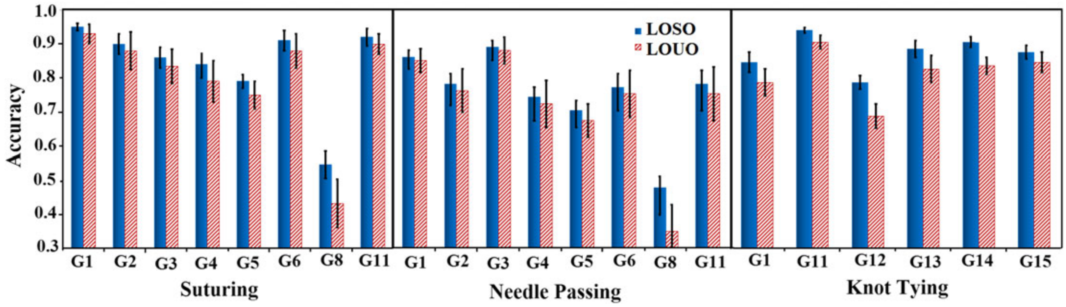

Figure 7. Classification accuracy for each surgical gesture using DTW-kNN and LOSO and LOUO model validation

high accuracy within the first 8 s. In addition, from Figure 6, for the LOUO validation schema more data is needed to have highly accurate task recognition but this issue can be resolved by increasing the training data to include more users with different skill levels and variability. Results also imply that to classify a task accurately we do not need to have the complete time series trajectory data. This suggests the potential of incorporating the proposed method in real-time camera control. For example, when the surgeon starts suturing, the algorithm can recognize it in the first 8 s with 96% accuracy. Then, the camera can automatically switch to the predefined mode for this task, which has been defined based on the best possible schema such as position of the camera or zooming level (4).

We also implement our proposed method at a more granular level to classify surgical gestures. From Figure 7 , it can be observed that some gestures such as G3 (pushing needle through tissue), G4 (transferring needle from left to right) or G6 (pulling suture with left hand) can be recognized with high accuracy while recognizing G8 (orienting needle) has the lowest accuracy. One possibility is that in general, G8 is a redundant gesture and it is not part of the whole suturing or needle passing. For example, when a surgeon is not able to finish a gesture such as pushing the needle into the tissue, (s)he may need to reorient the needle and start the gesture again. Therefore, it is difficult for the model to recognize it correctly because it is like an anomaly gesture.

It is worth noting that the proposed DTW-kNN framework is fast compared with HMM and LDS approaches because it builds a classifier directly using raw kinematic trajectory data with minimal preprocessing. The time required to calculate the DTW distance is a few minutes and the classification phase takes only a few milliseconds. This stands in bold contrast with current state-of-the-art surgical task and gesture recognition algorithms, which need a few hours processing video and kinematic data to build a model as accurate as our proposed framework (21).

Conclusion and Future Work

We proposed a task and gesture recognition framework, namely DTW-kNN, which is based on a dynamic time warping distance measure of motion trajectory data obtained from the API of the da Vinci and k-nearest neighbor classification method. The proposed framework outperformed other state-of-the-art methods by 4% to 9% across the three surgical tasks and 2% to 8% across the 12 gestures. We also showed that the combination of these two algorithms turns out to be robust, accurate and fast. These characteristics are a key advantage of our proposed approach compared with the state-of-the-art methods in the area of surgical gesture recognition. One of the potential applications of such a framework is

DOI: 10.1002/rcs

for an autonomous control system. For example, to have automatic camera control we need to know what the surgeon is doing in order to predict the next movement and adjust the camera mode based on that. This cannot be achieved unless we have a good understanding of surgical procedures at a different level of granularity. The task recognition framework presented in this paper can lay the groundwork towards development of autonomous surgical robot behaviors. However, more analysis needs to be done to evaluate the performance of the proposed method in real robotic surgery. Therefore, our next step is to implement the DTW-kNN algorithm on da Vinci to recognize tasks and to adjust the camera automatically.

Future work will focus on the generalizability aspects of the task and gesture recognition model by implementing the proposed framework on a larger dataset consisting of different surgical tasks and more users with different skill levels. In addition, other time series distance measure such as longest common subsequence (LCSS) (37) and classification methods such as linear dynamic systems (LDS) can be applied to potentially improve the accuracy of gesture recognition results. It is worth mentioning that in this work we used kinematic data to classify surgical gestures based on manually annotated data. Therefore, the proposed method relies on predefined gestures that are given by expert surgeons. One interesting research direction of this work is recognition of surgical gestures when no predefined labels are provided (38). Though motivated by application in an autonomous RMIS control system, the proposed algorithm is also applicable to various other domains such as robotic surgical skill assessment and training where real-time feedback to surgeons about their performance always has a high importance.

Author Justifications/Contributions

Mahtab J. Fard: Worked on developing model, analysis of results, and writing the manuscript

Abhilash K. Pandya: Provided subject matter expertise on surgery for model development, and analysis of results and implications for robotic interface design

Ratna B. Chinnam: Worked on development of the model and analysis of results

Michael D. Klein: Provided subject matter expertise on surgery for model development, and analysis of results

R. Darin Ellis: Worked on analysis of results, and writing the manuscript

The manuscript has been read and approved by all the authors.

The paper contains original, unpublished work and is not being submitted for publication to any other journal.

References

1. Lanfranco AR, Castellanos AE, Desai JP, Meyers WC. Robotic surgery. *Ann Surg* 2004; **239**(1): 14–21.
2. Cuschieri A. Whither minimal access surgery: tribulations and expectations. *Am J Surg* 1995; **169**(1): 9–19.
3. King BW, Reisner L a, Pandya AK, *et al.* Towards an autonomous robot for camera control during laparoscopic surgery. *J Laparoendosc Adv Surg Tech A* 2013; **23**(12): 1027–1030.
4. Ellis RD, Munaco AJ, Reisner LA, *et al.* Task analysis of laparoscopic camera control schemes. *Int J Med Robot Comput Assist Surg MRCAS* 2015. doi:10.1002/rcs.1716.
5. Pandya AK, Reisner LA, King BW, *et al.* A review of camera viewpoint automation in robotic and laparoscopic surgery. *Robotics* 2014; **3**: 310–329.
6. Guthart GS, Salisbury JK. The Intuitive telesurgery system: overview and application. Proc 2000 ICRA Millenn Conf IEEE Int Conf Robot Autom Symp Proc 2000; 1: 618–621.
7. Lalys F, Jannin P. Surgical process modelling: a review. *Int J Comput Assist Radiol Surg* 2014; **9**(3): 495–511.
8. Jahanbani Fard M. Computational modeling approaches for task analysis in robotic-assisted surgery; 2016.
9. Berndt DJ, Clifford J. Finding patterns in time series: a dynamic programming approach. Advances in knowledge discovery and data mining. *Am Assoc Artif Intell* 1996: 229–248.
10. Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Trans Inf Theory* 1967; **13**(1): 21–27.
11. Fried GM, Feldman LS, Vassiliou MC, *et al.* Proving the value of simulation in laparoscopic surgery. *Ann Surg* 2004; **240**(3): 518–528.
12. Bouarfa L, Jonker PP, Dankelman J. Discovery of high-level tasks in the operating room. *J Biomed Inform* 2011; **44**(3): 455–462.
13. Reiley CE, Lin HC, Yuh DD, Hager GD. Review of methods for objective surgical skill evaluation. *Surg Endosc* 2011; **25**(2): 356–366.
14. Kassahun Y, Yu B, Tibebu AT, *et al.* Surgical robotics beyond enhanced dexterity instrumentation: a survey of machine learning techniques and their role in intelligent and autonomous surgical actions. *Int J Comput Assist Radiol Surg* 2016; **11**(4): 553–568.
15. Rosen J, Hannaford B, Richards CG, Sinanan MN. Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills. *IEEE Trans Biomed Eng* 2001; **48**(5): 579–591.
16. Blum T, Padoy N, Feussner H, Navab N. Modeling and online recognition of surgical phases using Hidden Markov Models. *Med Image Comput Comput Assist Interv* 2008; **11**(2): 627–635.
17. Lalys F, Riffaud L, Bouget D, Jannin P. A framework for the recognition of high-level surgical tasks from video images for cataract surgeries. *IEEE Trans Biomed Eng* 2012; **59**(4): 966–976.
18. Varadarajan B, Reiley C, Lin H, Khudanpur S, Hager GD. Data-derived models for segmentation with application to surgical assessment and training. *Med Image Comput Comput Assist Interv* 2009; **12**(1): 426–434.
19. Lin HC, Shafran I, Yuh D, Hager GD. Towards automatic skill evaluation: detection and segmentation of robot-assisted surgical motions. *Comput Aided Surg* 2006; **11**(5): 220–230.
20. Haro BB, Zappella L, Vidal R. Surgical gesture classification from video data. Proceeding Int Conf Med Image Comput Comput Interv (MICCAI). 2012; 15: 34–41.
21. Zappella L, Béjar B, Hager GD, Vidal R. Surgical gesture classification from video and kinematic data. *Med Image Anal* 2013; **17**(7): 732–745.
22. Rosen J, Solazzo M, Hannaford B, Sinanan M. Task decomposition of laparoscopic surgery for objective evaluation of surgical residents' learning curve using hidden Markov model. *Comput Aided Surg* 2002; **7**(1): 49–61.

23. Dosis A, Belle F, Gillies D, *et al.* Laparoscopic task recognition using hidden markov model. *Stud Health Technol Inform* 2005; **111**: 115–122.
24. Reiley CE, Lin HC, Varadarajan B, *et al.* Automatic recognition of surgical motions using statistical modeling for capturing variability. *Stud Health Technol Inform* 2008; **132**(1): 396–401.
25. Reiley CE, Hager GD. Task versus subtask surgical skill evaluation of robotic minimally invasive surgery. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). 2009; 5761 LNCS(PART 1): 435–442.
26. Leong JJH, Nicolaou M, Atallah L, *et al.* HMM assessment of quality of movement trajectory in laparoscopic surgery. *Comput Aided Surg* 2007; **12**(6): 335–346.
27. Tao L, Elhamifar E. Sparse hidden markov models for surgical gesture classification and skill evaluation. Inf Process Comput Interv Berlin/Heidelberg; 2012; 167–77.
28. Xing Z, Pei J, Keogh E. A brief survey on sequence classification. *ACM SIGKDD Explor Newsl* 2010; **12**(1): 40.
29. Fu TC. A review on time series data mining. *Eng Appl Artif Intell* 2011; **24**(1): 164–181.
30. Ding H, Trajcevski G, Scheuermann P. Querying and mining of time series data: experimental comparison of representations and distance measures. *Proc VLDB Endow* 2008; **1**(2): 1542–1552.
31. Xi X, Keogh E, Shelton C, Wei L, Ratanamahatana CA. Fast time series classification using numerosity reduction. Proceedings of 23rd International Conference on Machine Learning, 2006; 1033–1040.
32. Rakthanmanon T, Campana B, Mueen A, *et al*. Searching and mining trillions of time series subsequences under dynamic time warping. Proc 18th ACM SIGKDD Int Conf Knowl Discov Data Min. 2012; 262–270.
33. Forestier G, Lalys F, Riffaud L, *et al.* Classification of surgical processes using dynamic time warping. *J Biomed Inform Elsevier Inc* 2012; **45**(2): 255–264.
34. Batista GE a P a, Wang X, Keogh EJ. A complexity-invariant distance measure for time series. *SIAM Int Conf Data Min* 2011: 699–710.
35. Gao Y, Vedula SS, Reiley CE, *et al.* JHU-ISI gesture and skill assessment working set (JIGSAWS): a surgical activity dataset for human motion modeling. *Model Monit Comput Assist Interv* 2014: 1–10.
36. Lin JF, Kuli D. Segmenting human motion for automated rehabilitation exercise analysis. *IEEE Trans Neural Syst Rehabil Eng* 2012; **22**(1): 2881–2884.
37. Vlachos M, Kollios G, Gunopulos D. Discovering similar multidimensional trajectories. Proceedings of 18th International Conference on Data Engineering, 2002; 673–684.
38. Fard MJ, Ameri S, Chinnam RB, Ellis RD. Soft boundary approach for unsupervised gesture segmentation in robotic-assisted surgery. *IEEE Robot Autom Lett IEEE* 2017; **2**(1): 171–178. doi:10.1109/LRA.2016.2585303.