1-1-2016

Computational Modeling Approaches For Task Analysis In Robotic-Assisted Surgery

Mahtab Jahanbani Fard
Wayne State University,

COMPUTATIONAL MODELING APPROACHES FOR TASK ANALYSIS IN ROBOTIC-ASSISTED SURGERY

by

MAHTAB JAHANBANI FARD

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2016

MAJOR: INDUSTRIAL ENGINEERING

Approved By:

_____

Advisor   Date

_____

_____

_____

DEDICATION

To my parents and my husband for their endless encouragement and support.

ACKNOWLEDGEMENTS

I owe a debt of gratitude to the many remarkable people who directly or indirectly have helped me with my PhD. I would not be at this stage of my life without their help and support.

First and foremost, I would like to express the deepest gratitude to my advisor Dr. R. Darin Ellis, who has been a tremendous mentor for me. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist. You have inspired me to work harder and aim higher in my graduate studies and beyond. My special appreciation also goes to Dr. Ratna Babu Chinnam, for his brilliant comments, invaluable advice and novel ideas in my research. My dissertation could not be in this level without your support and time. I would also like to thank Dr. Abhilash Pandya and Dr. Brady King from department of Electrical and Computer Engineering for their insightful comments and encouragement. My sincere thanks also goes to Dr. Alper Murat for his great feedback and suggestions. My gratitude extends to Dr. Michael Klein from department of Surgery for his significant contribution at different stages of my research.

Most importantly, none of this would have been possible without the love and patience of my family. Words cannot express how grateful I am to my parents for their endless encouragement, support and all of the sacrifices that they have made on my behalf. I would not have made it this far without them. My heartfelt regard goes to my parents-in-laws for their love and prayer. I also want to thank my brother, Mehrdad, who has always been by my side to help me. I would also like to thank all of my friends who supported me in writing, and incented me to strive towards my goal.

Last but not least, I would like to express my sincere appreciation to my beloved husband, Sattar, who spent sleepless nights with me and was always my support in all the up and down moments during this journey. There are no words to convey how much I love you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

The operating room is a main unit in a hospital where surgical operations are performed. It is a challenging work environment that requires intense cooperation and coordination between a wide range of people and departments. Surgery is continuously subject to technological and medical innovation such as introduction of new imaging technologies, advanced surgical tools, navigation and patient monitoring systems. The purpose of these advances is to improve patient treatment while they transform complicity to daily routine. Recently, some of these challenges have been addressed by introducing technological innovations such as Robotic Minimally Invasive Surgery (RMIS) [1, 2]. RMIS provides additional advantages over conventional laparoscopic surgery, including an increase in dexterity [3] and precision [4]. Robotic-assisted surgery promises to improve patient treatment by enabling shorter hospital stays, shortening recovery time and reducing the risk of infection [5]. Current robotic surgery systems operate in a master-slave mode where relies exclusively on direct surgeon input [6]. A real "*robotic*" surgical system should intelligently understand current task and generate the necessary movement.

It is clear that to develop such an autonomous systems, a more rigorous model of surgical procedures is needed. Surgical tasks need to be modeled and quantified to make them amenable for further study. Goal-oriented human motion and human language are analogous as both of them consist of a low-level elements that, when combined in meaningful sequences, result in an emergent meaning or higher-level task. Hence, techniques that have effectively been applied in the analysis of human speech and language are natural candidates to apply to surgical motion modeling. Consequently, the "*Language of Surgery*" has been defined as a systematic description of surgical activities and rules for decomposing surgical tasks [1]. More specifically, the language of surgical motion includes describing particular activities that are performed by surgeons with their instruments or hands to accomplish a planned surgical objective.

It is worth mentioning that robotic-assisted technology has an important impact on surgical skill assessment and training. Traditionally, surgical skills have been assessed in the operating room by direct supervision and feedback of expert surgeons [7]. However, this method has been criticized for being too subjective and not representing the actual level of skills [8]. New techniques, such as robotic minimally invasive surgery, require new skills, which have different learning curves and require different training methods. These developments have resulted in an increased interest in objective and automated surgical skill assessment. Although different objective based skill assessment methods has been proposed, there is still huge potential for improvement by introducing a framework which being less time-consuming, required less human interaction and be more robust.

The key step for advancing research in robotic-assisted surgery and autonomous skill assessment is to develop techniques that are capable of accurately recognizing fundamental surgical tasks and gestures more intelligently. Surgical gestures need to be quantified to make them amenable for further study. Current systems like *da Vinci* (Intuitive Surgical, Sunnyvale, CA) [9] record motion and video data, enabling development of computational models to recognize and analyze surgical performance through data-driven approaches. Recent advances in data mining and machine learning research for uncovering concealed patterns in huge datasets, like kinematic and video data, offer the possibility to better understand, analyze and model surgical gestures from a system point of view [10]. While machine learning techniques have been used extensively in other fields because of their advantages over traditional statistical methods such as robustness, better prediction ability and higher tolerance to violations of assumptions [11], but it is only recently that these methods have been considered to analyze RMIS tasks. In this dissertation, we shed light on some of the barriers and challenges toward autonomous robotic-assisted surgery and skill assessment by developing novel methods based on computational intelligence techniques.

1.1   Dissertation Contributions

The main contributions of this dissertation in the areas of surgical task analysis, gesture recognition and skill assessment can be summarized as follows:

- Develop a new framework to recognize and classify three important surgical tasks including suturing, needle passing and knot tying based on Cartesian trajectory data captured using *da Vinci* robotic surgery system by [1]. The proposed framework, namely DTW-kNN, integrates time series similarity technique using dynamic time warping (DTW) with the k-nearest neighbor (kNN) classification method and provides high accuracy results. Our proposed method is simple, straightforward and accurate due to minimal preprocessing requirements. Therefore, it has the potential to be applied for autonomous control systems in robotic-assisted surgical devices (Chapter 2).

- Propose an unsupervised gesture segmentation and recognition method, namely UGSR, which has the ability to segment and recognize minimally invasive robotic surgery trajectory paths automatically. We also extend the UGSR method using soft boundary segmentation to address some of the challenges that exist in the surgical motion segmentation. The proposed algorithm can effectively model gradual transitions between surgical activities (Chapter 3).

- Develop a predictive framework for surgical skill assessment to automatically evaluate the performance of surgeons in robotic-assisted surgery. Our classification framework is based on the Global Movement Features (GMFs) that are extracted from kinematic data captured during RMIS tasks. The proposed method addresses some of the limitations in previous work and gives more insight about the underlying patterns of surgical skill levels (Chapter 4).

1.2   Definitions

In this section, we provide the definitions for the terms in the field of robotic-assisted surgery that are used throughout this dissertation.

Surgical Procedure: a medical procedure involving an incision with instruments which consists of several phases to repair damage or arrest disease in a living body, such as a colectomy procedure.

Surgical Phase: the major types of events occurring during surgery. Each procedure is composed of a list of phases.

Surgical Task: a sequence of activities used to achieve a surgical objective, such as suturing, knot tying, etc.

Surgical Gesture (Surgeme): a well-defined surgical motion unit, such as "push needle to the tissue" in suturing.

Based on the definitions, the surgical procudure hierarchy for colectomy surgery is shown in Figure 1.1.
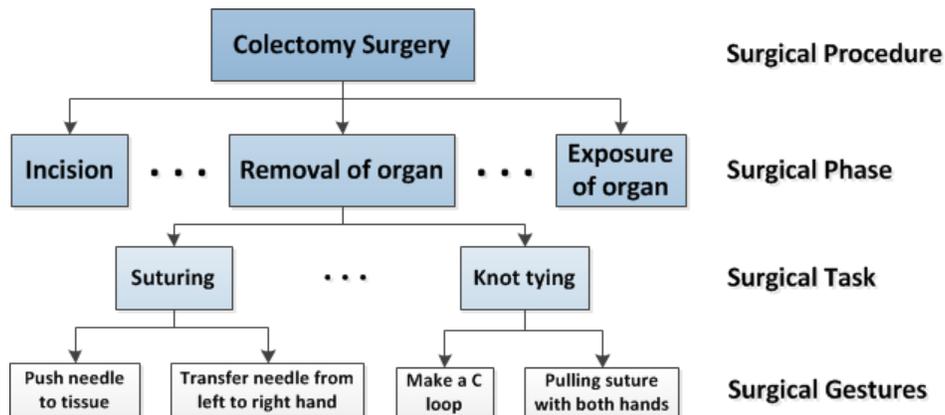
Figure 1.1: Example of surgical procedure hierarchy for colectomy suregry.

1.3 Data Source

In this dissertation, we are using real robotic surgery data, namely JHU-ISI[1] Gesture and Skill Assessment Working Set (JIGSAWS), presented in [1]. This is comprised of data for different surgical tasks including suturing, needle passing and knot tying (Figure 1.2), performed by eight right-handed surgeons with different skill levels (expert, intermediate and novice). The three surgical tasks are defined as follow:

- **Suturing (SU)**: the surgeon picks up needle then proceeds to the incision and passes through tissue. Then after the needle pass, the surgeon extracts the needle out of the tissue.

- **Needle-Passing (NP)**: the surgeon picks up the needle and passes it through four small metal hoops from right to left.

- **Knot-Tying (KT)**: the surgeon picks up one end of a suture tied to flexible tube attached at its ends to the surface of the bench-top model, and ties a single loop knot.

Each user performed around 5 trails of the task. For each of the task, we analyze kinematic data captured using the API of the *da Vinci* at 30 fps. Data consists of 39 trials of SU, 36 trials of KT, and 28 trials of NP. The data for the remaining trials (1 for SU, 12 for NP, and 4 for KT) were unusable because of corrupted data recordings. The data includes 76 motion variables which consist of 19 features for each robotic arms, left and right master side, and the left and right patient side which include 3 Cartesian positions, a rotation matrix consist of 9 variables, 3 linear velocities, 3 angular velocities and a gripper angle.

It is worth mentioning that, although the dataset that we used in this dissertation is very small, this does not impose any limitation to the model that have been developed in this study. However, a larger dataset could improve accuracy of the results and gives more

---

[1]Johns Hopkins University Information Security Institute

| Suturing | Needle Passing | Knot Tying |

Figure 1.2: Snapshot of the three fundamental robotic-assisted minimally invasive surgery tasks.

ability to generalize the conclusion that have been made.

1.4   Dissertation Organization

The rest of this dissertation is organized as follows. In Chapter 2, we present a classification framework to recognize robotic-assisted surgical tasks and gestures using Cartesian trajectory data. In Chapter 3, we propose a novel unsupervised gesture segmentation method to recognize gestures automatically in the surgical task. In Chapter 4, we present a new objective-based framework for automatic surgical skill assessment and dexterity evaluation during robotic surgery. Finally, we summarize our results and present possible directions for future research in Chapter 5.

CHAPTER 2: ROBOTIC SURGERY TASK CLASSIFICATION

2.1   Introduction

The operating room, where surgical procedures are performed, is a challenging work environment. Like other areas, surgery is continuously subject to technological innovations including the introduction of robotic surgical devices [12]. 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 [5]. The ultimate goal of RMIS is to program the surgical robot to perform certain difficult or complex control in an autonomous manner. There is, however, no technical roadmap to a fully autonomous surgical system at the present time [13, 14]. Current RMIS systems operate in a master-slave mode, relying exclusively on direct surgeon input [15]. For example, camera control in current RMIS platforms is an additional task under the direct control of the surgeon. This can distract the surgeon from the smooth flow of the operation, and certainly adds time to the procedure. As an additional complication, the surgeon's judgment is limited by a single video feed returned from the remote environment. Poor camera placement, narrow field of view and other camera properties can significantly impair the surgeon's perception of the environment, inviting increased cognitive workload and the possibility of disorientation. Therefore, to reduce the workload and improve view field of surgeon, an automatic camera control system is desired.

However, it is quite clear that to develop any automatic control system, for camera control or other robotic surgery tasks, a more detailed comprehension of the surgical procedures is needed [2]. The feasibility of the current robotic surgery systems to record quantitative motion and video data enables the opportunity of creating descriptive mathematical models to recognize, classify and analyze surgical tasks. Recent advances in machine learning research for uncovering concealed patterns in huge data sets, like kinematic and video data, offer a possibility to better understand surgical procedures from a system point of view. Surgical

tasks and at a more granular level, surgical gestures, referred to as surgemes [16], need to be quantified to make them amenable for further study in an automatic camera control system.

In recent years, recognizing and understanding a surgical procedure at different levels of granularity has been a focus of research [17, 18]. However, the existing methods suffer from needing a huge amount of preprocessing, parameter tuning and complexity, which make them impractical for online recognition of surgical activities in a real procedure. For any autonomous control system, we need a robust, accurate and fast method with minimum preprocessing steps to recognize surgical tasks and gestures intelligently. In this chapter, we describe the work we have done to address these challenges by developing a task and gesture recognition framework that works directly on raw data captured during robotic surgery. Our framework is based on similarity measure between temporal patterns of robot arms' tool tips trajectory data. In this work, we focus on three important RMIS tasks: knot tying, needle passing and suturing. These are all part of a fundamentals of laparoscopic surgery (FLS) skills training program [19]. The results shows the feasibility of robust real-time surgical task recognition with minimum data preprocessing that will support future research in the area of robotic camera positioning.

2.2 Background and Related Work

Surgical task and gesture recognition has been a subject of research for decades [16]. At the higher level of surgical process modeling, statistical models have been proposed using recorded force and motion data [20], surgical tool usage [21] and video data [22] to classify surgery phases. Most existing work has addressed the recognition of activities using different techniques such as neural networks and Hidden Markov Models (HMM) [23, 24]. At the lower level, effort has been applied to detect surgical gestures using HMM [25–28] or extension of HMM such as Gaussian HMM with Linear Discriminant Analysis (LDA) [29, 30], Gaussian mixture model (GMM) [31] and sparse HMM [32] to identify and classify surgical gesture based on kinematic data. Robotic-assisted surgical devices are generally equipped with

cameras that record the entire procedure. Therefore, surgical video analysis for both higher [21, 33] and lower [34, 35] grain size of analysis have been used. Recently, [36] used both kinematic and video data to classify surgemes in robotic surgery. 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 they lack robustness due to the requirements of parameter estimation and tuning for high dimensional data [32], all of which make them inapplicable for automatic control in robotic surgery. [2] and [10] give more details about works that have been done for robotic-assisted surgical task and gesture recognition.

To get one step closer to an autonomous RMIS system, we need to develop quantitative techniques that can be used as a framework to differentiate important tasks and motions during surgical procedures. One of the important applications of such a task recognition framework is in automatic camera control system. Although there have been several attempts to develop a system to move the camera automatically using image processing approach [37–40], they fail to address the important fact that different surgical tasks and motions may require different camera behaviors [14]. Systems that simply follow the tools are not sufficient [41, 42]. In the current FDA-approved system, *da Vinci* surgical platform (Intuitive Surgical, Sunnyvale, CA, USA) [9], many interface parameters are adjusted once and remain at the same level throughout the operation. Some others such as motion scaling and zooming level of camera are adjusted manually. Here the surgeon needs to pause the procedure to move the camera or change the zoom level which cause many interruptions. Thus, any automated camera control system should be adapted based on the task that is currently being performed. This shows the necessity of developing robust and accurate task recognition models.

2.3 Surgical Task and Gesture Classification Methods

In this study, we proposed a task and gesture recognition framework based on temporal kinematic signals, captured during operations. Before introducing our framework, we briefly

explain Hidden Markov model (HMM) which has been used widely in the area of gesture recognition.
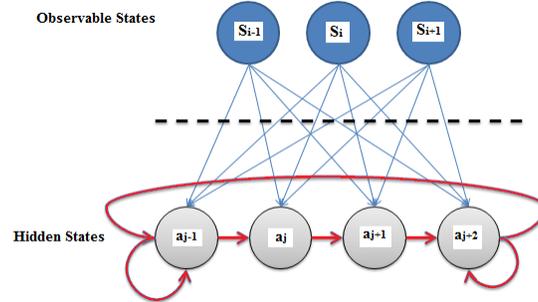
Figure 2.1: Illustration of Hidden Markov Model.

2.3.1   Hidden Markov Model (HMM)

HMM is a statistical Markov model which assumes that the systme under study, has a Markov process with unobserved (hidden) states [45]. Hidden Markov models have found application in many areas, especially in temporal pattern recognition such as speech, handwriting, gesture recognition [46]. A first-order HMM is a tuple $M = \{S, A, p, q\}$ where, S is the set of states in the process, A is the set of actions that can be observed, p is the transition probability function, where $p(s_t|s_{t-1})$ signifies the probability of transition from state s_{t-1} to state s_t, and q is the action observation probability function, where $p(a_t|s_t)$ denotes the probability of observing action a at time t given state s_t (see Figure 2.1). To estimate the current state of a process, we calculate the probability of being at some state s_t after observing a sequence of actions $\{a_1, ..., a_t\}$, which we can write as $p(s_t) = p(s_t|a_1, ..., a_t)$. This estimation is performed iteratively, using the well-known Viterbi algorithm [47]. Doing this for every possible state gives us a probability distribution over the entire state space. The BaumWelch algorithm have been developed to find the maximum likelihood estimate of the parameters of a hidden Markov model given a set of observed feature vectors using the well known expectation maximization (EM) algorithm [47]. The HMM method has been applied to surgical gesture recognition [25] and some other extensions of it such as Gaussian mixture

model (GMM) [31] and sparse HMM [32] to identify and classify surgical gesture based on kinematic data. However, the main drawbacks of these methods are being time consuming, requiring significant human interaction and preprocessing, and lacking robustness due to the requirements of parameter estimation and tuning for high dimensional data [32].

2.3.2 DTW-kNN Classification

To address the challenges that explained in previous sections, we develop a task and gesture recognition framework based on temporal kinematic signals, captured during operations. It classifies robotic-assistant surgical tasks and recognize surgical gestures by integrating temporal sequence similarity measure techniques such as Dynamic Time Warping (DTW) [43] with well-known k-nearest neighbor (kNN) classification method [44]. The DTW-kNN has not been applied for this problem and yields good results. The proposed framework is fast, accurate and robust, all of which makes it applicable for any adaptive control system, such as for the camera, in robotic surgery.

Our proposed framework consists of two key components. The first component measures the similarity between different surgical tasks and gestures. We analyze motion data from a robotic surgery device to extract multivariate time series datasets. In this study, we employ Dynamic Time Warping to measure the similarity between multidimensional temporal trajectory data. The second component is the classification algorithm, which is based on the k-nearest neighbor approach. The combination of these two steps results in a robust classification framework for RMIS data (Figure 2.2). In the following sections, each component will be discussed in detail.

Dynamic Time Warping (DTW) Similarity Measure

The choice of method for measuring (dis)similarity is a critical step in achieving valid classification results. In order to have a meaningful comparison, each time series needs to normalized to have a mean of zero and a standard deviation of one before measuring the distance [48]. Shaped-based similarity measure techniques are one of the well-developed
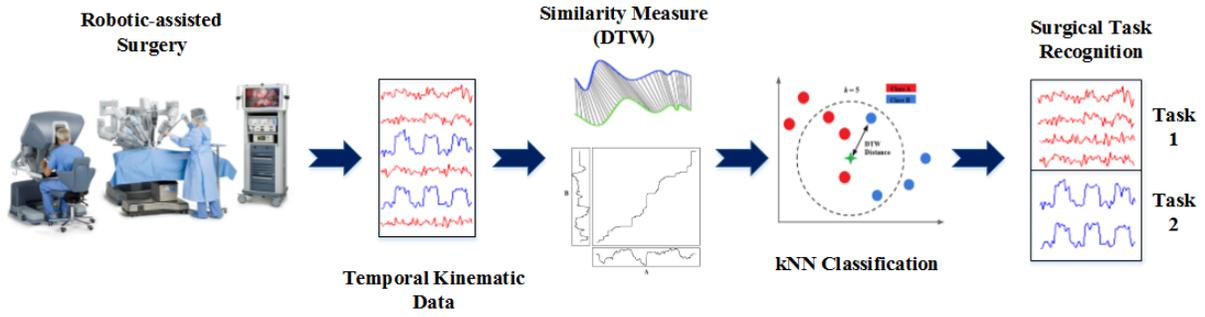
Figure 2.2: Proposed framework consists of two steps: similarity measurement between temporal sequence data and classification using k-nearest neighbor method.

methods in this area where determine the similarity of the overall shape of two time series by directly comparing their individual point values [49]. It is in contrast with feature-based and structure-based methods where first features need to be extracted in order to find higher-level structures of the series.

One of the simplest ways to estimate the distance between two sequences is the Euclidean distance. However, despite the simplicity and efficiency of Euclidean distance, which makes it the most popular distance measure, it has a major drawback. It requires both input sequences to be of the same length, and it is sensitive to distortions, e.g. shifting, noise, and outliers. If, for instance, two time series are indistinguishable, however slightly out of phase with one another, then the Euclidean distance measure will give an extremely poor similarity measure. 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 [43]. Figure 2.3 shows an intuitive representation of DTW versus Euclidean distance.

Given two m-dimensional time series $\mathbf{A} = (\mathbf{a_1}, \mathbf{a_2}, ..., \mathbf{a_m})$ and $\mathbf{B} = (\mathbf{b_1}, \mathbf{b_2}, ..., \mathbf{b_m})$ where $\mathbf{A}$ and $\mathbf{B}$ have $p \times m$ and $q \times m$ dimension, respectively and $p \neq q$. These two sequences can be arranged as $p \times q$ matrix like the sides of a grid (Figure 2.4) in which the distance between every possible combination of time instances $\mathbf{a_i}$ and $\mathbf{b_j}$ is stored. To find the best match between two sequences, a path through the grid that minimizes the overall distance between them is needed. Both sequences start on the bottom left of the grid. To estimate
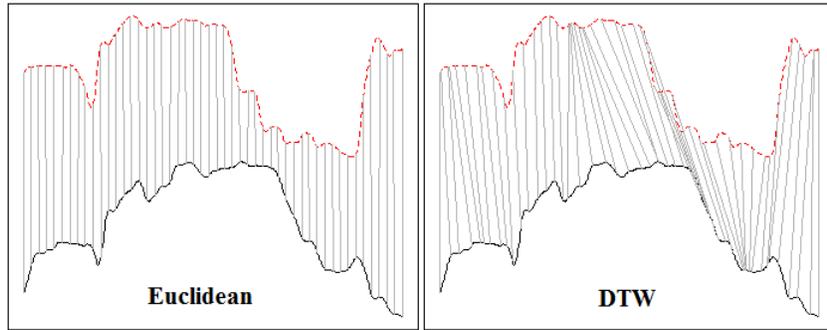
Figure 2.3: Comparison between Euclidean and DTW distance measure of two time series data for knot tying task.

the distance between two time series any L_n norm can be used as

$$D_{L_n}(\mathbf{a_i}, \mathbf{b_j}) = \left(\sum_{l=1}^{m} \left(a(i,l) - b(j,l) \right)^n \right)^{1/n} \tag{2.1}$$

where for multidimensional DTW, we use the well-known Euclidean distance measure by giving $n{=}2$ in Eq. (2.1) to find a distance between two m-dimensional sequences.
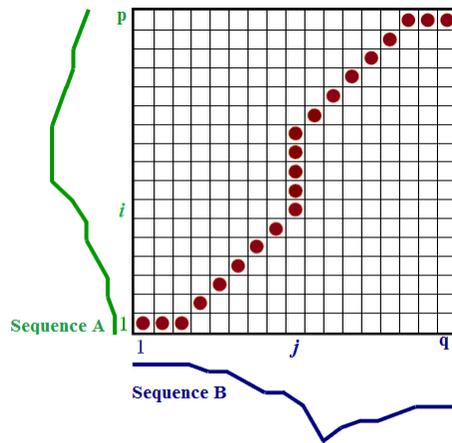
Figure 2.4: Illustrative example for temporal sequence alignment using DTW.

The power of the DTW algorithm is that rather than exploring every conceivable path, the grid keeps track of the cost of the best path. Thus, DTW distance can be formulated as a dynamic programming problem. Using a dynamic programming approach, the warp path must either be incremented by one unit or stay at the same i-axis or j-axis (Figure 2.5).

Therefore, one can formulate it as recurrence of cumulative distance, defined as

$$DTW(i,j) = d(\mathbf{a_i}, \mathbf{b_j}) + \min\left\{DTW(i, j-1), DTW(i-1, j), DTW(i-1, j-1)\right\} \quad (2.2)$$

where $d(\mathbf{a_i}, \mathbf{b_j})$ can be calculated using Eq. (2.1).
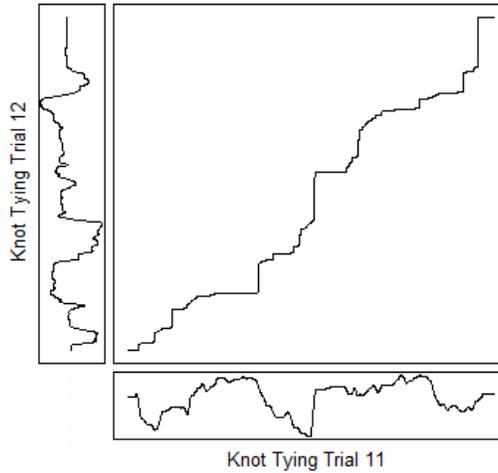
Figure 2.5: Time series alignment using warping matrix with the minimum distance warp path of two trial for knot tying task.

k-Nearest Neighbor Classification

The k-Nearest Neighbors algorithm (kNN) is a supervised distance-based classification method. It has a simple principle behind it, which is to predict the label for the new point based on the closest distance to a predefined number (k) of training samples. Despite its simplicity, k-Nearest Neighbors has been very successful in classification problems [50]. kNN classifier is an instance-based learning where instead of constructing a general model, it simply stores instances of training data. During the classification phase, the majority vote of the k-nearest neighbor for each point is computed. Thus, the label for the query point is assigned based on the most representatives within the nearest neighbors of the points. Figure 2.6 illustrates the kNN algorithm for k=5.
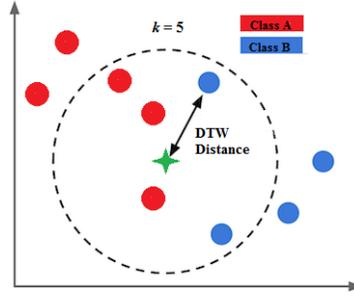
Figure 2.6: Illustrative example for 5-NN classification.

2.4 Experimental Result

In this section, we will describe the details of implementation and performance evaluation for the proposed method. As mentioned in Section 1.3, we apply our method on data presented in [1] which are comprised of data for different surgical tasks including suturing, needle passing and knot tying. Then, the performance results of the proposed classification framework will be presented.

2.4.1 Performance Evaluation

In order to compare the accuracy of the proposed task and gesture recognition framework, we used two cross-validation schemes to account for the structure of the dataset as suggested in [1]:

Leave One Supertrial Out (LOSO): Supertrial i is defined as the set consisting of the ith trial from all subjects for a given surgical task. In the LOSO setup cross-validation, five folds with each fold comprising of data from one of the five supertrials is created. The LOSO setup can be used to evaluate the robustness of a method for repeating task by leaving out the ith trial for all subjects.

Leave One User Out (LOUO): In this cross validation setup, eight folds, each one consisting of data from one of the eight subjects is created. The LOUO setup can be used to evaluate the robustness of a method when a subject is not previously seen in the training data.

The surgical activity annotation is also provided manually by [1]. In this regards, a surgeme can be characterized as a unit of purposeful surgical action which has a distinguishable and meaningful outcome. Table 2.1 lists the surgical gestures and their description for all the three tasks.

Table 2.1: Description of gestures in three surgical tasks: 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			✓

The performance of the different task recognition methods was determined by classification accuracy, which expressed in terms of percentage of subjects in the test set that are classified correctly.

2.4.2 Surgical Task Classification

For the three RMIS tasks, suturing, needle passing and knot tying, the DTW similarity measures the pairwise distance between the three Cartesian position of tool tips of patient side arm of robot for both right and left hands. Then, kNN classification method was applied to recognize different tasks. We test different value for k and our preliminary results shows that the accuracy of proposed model is robust to the values of k in the range between 3 to 7 while the best result achieved for k=5. Table 2.2 shows the overall task recognition accuracy of the three different scenarios of using only right hand, left hand or both on (x, y, z) Cartesian data. The result shows that the proposed DTW-kNN method achieved a

top accuracy of 95.5% and 88% for LOSO and LOUO respectively, when we use both hands Cartesian data.

Table 2.2: Comparison between accuracy of proposed framework for right, left and both hands movement path using LOSO and LOUO validation schema (and their standard deviation).

	Right Hand	Left Hand	Both Hands
LOSO	92.33% ± 0.08%	91.58%±0.09	**95.50%±0.01%**
LOUO	86.35%±0.56%	84.21%±0.57	**88.73%±0.41%**

The performance details of the DTW-kNN framework for each task compared with HMM are presented in Table 2.3. The results show that the proposed framework outperformed state-of-the-art HMM method. It also shows that for LOSO, 100% of suturing, 97% of knot tying and 89% of needle passing are correctly classified. For LOUO validation schema, the correctly classified suturing is 84.6%, needle passing is 85.7% and knot tying is 95.8%.

Table 2.3: Comparison between accuracy for each task using HMM and proposed DTW-kNN for LOSO and LOUO validation schema.

	LOSO		LOUO	
	HMM	**DTW-kNN**	**HMM**	**DTW-kNN**
Suturing	96.4%	**100.0%**	80.7%	**84.6%**
Needle passing	83.5%	**89.3%**	80.8%	**85.7%**
Knot tying	**97.3%**	97.2%	90.9%	**95.8%**

2.4.3   Real-time Task Recognition

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 on the first x% of the total time series signals for each task and evaluated the performance of the method. Figure 2.7 shows the result for the two different validation methods. In LOSO, having only 1% of complete temporal sequence, the model was able to recognize knot tying by 86% accuracy. In addition, we need 9 seconds for needle passing and 6 seconds for suturing to be able to recognize these two tasks with around 88% accuracy in LOSO.
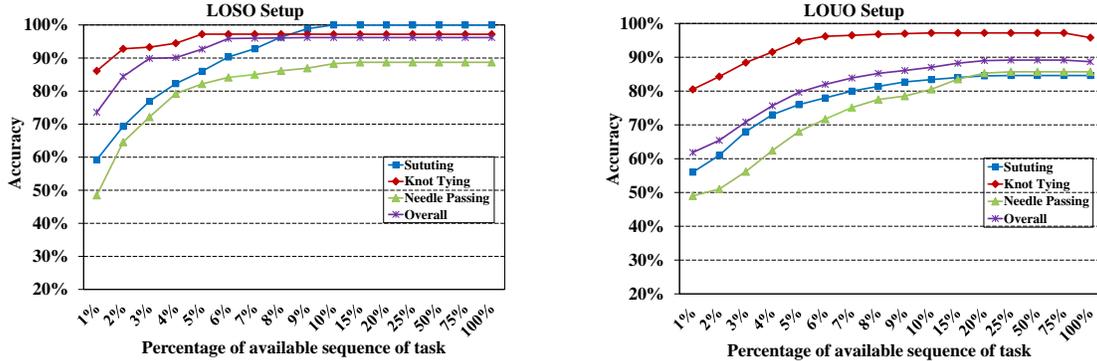
Figure 2.7: Accuracy of task recognition based on different percentage of total temporal sequence of task for LOSO and LOUO validation schema.

2.4.4 Surgical Gesture Recognition

We also applied the DTW-kNN method at smaller granular level to recognize surgical gestures in each task separately. Since gestures are more difficult to recognize compared to tasks, for this experiment, we used all 38 features of the patient-side robot manipulators which enable us to include more information in the model. In Table 2.4 we compared the results of DTW-kNN with Sparse Hidden Markov Model (SHMM) and Linear Dynamic System (LDS) as presented in [36].

Table 2.4: Comparison between accuracy of surgical gesture classification results of SHMM, LDS and proposed DTW-kNN for each task using LOSO and LOUO validation schemas.

	LOSO			LOUO		
	SHMM	**LDS**	**DTW-kNN**	**SHMM**	**LDS**	**DTW-kNN**
Suturing	79.4%	**87.3%**	86.9%	60.8%	74.6%	**82.0%**
Needle passing	76.4%	78.8%	**79.9%**	45.3%	67.3%	**70.1%**
Knot tying	86.8%	85.1%	**89.3%**	72.0%	78.9%	**80.1%**

From Table 2.4 our proposed method can recognize the gestures in suturing with 86.9%, in needle passing with 79.9% and in knot tying with 89.3% accuracy for LOSO. On the other hand, for LOUO the accuracy decreased to 82%, 70.1% and 80.1% for suturing, needle passing and knot tying respectively. However, the results show that DTW-kNN method outperformed other state-of-the-art models. Table 2.5 summarizes the detail performance

of DTW-kNN framework for each gesture separately. For instance, results show that in suturing 86% of gesture G3, pushing needle through tissue, and in needle passing 90% of the same gesture are classified correctly.

Table 2.5: Classification accuracy of each surgical gesture by applying DTW-kNN based on two validation schema, LOSO and LOUO.

	LOSO			LOUO		
	Suturing	Needle Passing	Knot Tying	Suturing	Needle Passing	Knot Tying
G1	95%	87%	85%	94%	86%	78%
G2	90%	79%	-	89%	77%	-
G3	86%	90%	-	86%	89%	-
G4	84%	75%	-	79%	73%	-
G5	79%	71%	-	75%	68%	-
G6	91%	78%	-	88%	76%	-
G8	46%	38%	-	33%	18%	-
G11	92%	79%	95%	90%	76%	91%
G12	-	-	79%	-	-	62%
G13	-	-	89%	-	-	83%
G14	-	-	91%	-	-	84%
G15	-	-	88%	-	-	85%

2.5 Discussion

From the results presented in section 2.4, we observe that classification accuracy decrease when we switch from LOSO validation schema to LOUO. It should be mentioned that the LOUO result provides an insight into the ability of the algorithms to generalize and recognize task performed by surgeon that were unseen during the training phase. For example, from Table 2.3 switching from the LOSO to LOUO, the performance of suturing decreases around 15%, needle passing drops 4% and knot tying drops 1%. Such a difference among tasks suggests that needle passing and suturing are possibly performed in a less similar way between surgeons. We can conclude that different surgeons have their own style to perform suturing and needle passing, while knot tying is performed more consistently.

One important thing to remark upon is the potential of the proposed model to be used for real-time task recognition. For example in knot tying, having only 1% of complete

temporal sequence, the model was able to recognize the task by 86% accuracy. In our dataset, the average time for knot tying, needle passing and suturing are 57, 110 and 120 seconds respectively which means that knot tying can be recognized in less than a second. However, for needle passing and suturing, which are more complex and time consuming tasks, we need slightly more information. From Figure 2.7, although for LOUO validation schema, we need more data to have an accurate task recognition but, this issue can be resolve by increasing user variability in the dataset. Therefore, results imply that to identify the task accurately we do not need to have the complete temporal sequence of trajectory path. It suggests the potential of incorporating the proposed method in real-time camera control. For example, when the surgeon starts suturing, which on average may take 120 seconds, the algorithm can recognize it in first 10 seconds with 90% accuracy. Then, the camera can automatically switch to the predefined mode for suturing which was defined based on the best schema for different tasks such as position of the camera, movement and zooming level [14].

We also implement our proposed method for a more granular level where the model classifies gestures in each surgical tasks. From Table 2.5, 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 G8 (orienting needle) has the lowest accuracy. The reason is that in general, G8 is not part of suturing or needle passing task. For example, when surgeon is not able to finish a gesture such as pushing needle in to the tissue, (s)he may perform to orient needle and does the gesture again. Therefore, it is difficult for the model to recognize it correctly because it is observed in only few numbers of trials in the data.

It is worth noting that the proposed DTW-kNN framework is fast because it builds a classifier directly using raw kinematic trajectory data with minimal preprocessing. The time required for calculating the DTW distance in offline mode is a few minutes and the

classification phase during actual surgery takes only few milliseconds. This stands in bold contrast with the current state-of-the-art surgical task and gesture recognition algorithms, which need a few hours for processing video and kinematic data to build a model as accurate as our proposed framework.

2.6 Conclusion

In this chapter, we proposed a task and gesture recognition framework, namely DTW-kNN, based on a dynamic time warping similarity measure of motion trajectory data and k-nearest neighbor classification method. We 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 to the state-of-the-art methods in the area of surgical gesture recognition. One of the potential applications of such a framework is for autonomous control. For example, to have an 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. Therefore, the task recognition framework presented in this study can lay the groundwork towards development of autonomous robot behaviors during RMIS. Though motivated by application in 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.

CHAPTER 3: SURGICAL GESTURE SEGMENTATION

3.1 Introduction

Research on human gesture recognition has attracted a great deal of attention in the past decades [52, 53]. Different techniques has been proposed to identify human gestures using both motion kinematic and video data [54–56]. Gesture recognition finds applications in varied domains including human computer interaction, autonomous vehicles and medical systems [57, 58]. In medical systems, surgical gesture recognition is very important. Recently, with the current trend toward robot assistants and the execution of autonomous tasks in minimally invasive surgery, robotic surgery gesture recognition has been of growing interest for the past few years [16].

In the preceding chapter, we used kinematic data to classify surgical gestures based on manually annotated data and build an automatic gesture classification framework. However, the proposed method relies on predefined gestures which are given by expert surgeons. Thus, the question that arises is, what if the start and end frames of gestures are not given. In other words, "could we segment the temporal sequence of gestures when no predefined labels are provided?" This is a valid question since providing such a faultless input is often unrealistic. Human annotations can be time-consuming and prone to error by missing segments or applying segmentation criteria inconsistently across a dataset [59]. On the other hand, surgery is a multi-step procedure which consists of complex continuous activities. Surgical tasks may contain superfluous, repeated actions and temporal variation. Hence, deployment of gesture recognition methods in real-world application is a very challenging problem to tackle.

Advances in machine learning and computer vision in one hand and surgical device technology improvement on the other hand can be exploited to address this challenge [10]. Despite the success that have been made in this area [60, 61], there are still opportunities for further enhancements. Such a model should have the capability of capturing all the charac-

teristics of robotic surgery and automatically segment the continuous sequence of gestures. In this chapter, we go one step further where the input data is an unlabeled temporal sequence of surgical tasks. Therefore, the gesture recognition problem will change to temporal segmentation of a surgical task where we want to determine when each gesture starts and when it ends. Within the context of the proposed framework, this study makes additional contributions by developing temporal soft boundary gesture recognition framework which segment surgical activities in an unsupervised manner.

3.2   Background and Related Work

Most of the previous work in the area of surgical gestures recognition is based on supervised methods either using labeled gestures [25, 28, 36] or pre-characterized vocabulary of primitives [30, 60–62], which in both cases requires an expert surgeon to manually label surgical data. These methods have two main drawbacks. First, they are very time-consuming and subjective, which may cause non-robust results. Second, surgery is a complex procedures with a lot of variation between surgeons doing same task. Therefore, creating a comprehensive dictionary that covers all the surgical actions is not practical. Not to mention that for these methods, human judgement is still involved, with its inherent subjectivity [63].

To address the drawbacks of previous works, a number of research groups have attempted to propose unsupervised segmentation techniques where the criteria is learned directly from data [64, 65]. In [66], authors proposed unsupervised trajectory segmentation method which is based on kinematic features such as curvature or torsion of surgical movement path. More recently, trajectory clustering algorithms have been also developed to segment surgical task in an unsupervised fashion. [63] proposed Transition State Clustering (TSC), which segments a set of surgical trajectories by detecting and clustering transitions between linear dynamic regimes. They extended their work by applying deep learning method on video data and proved that combining kinematic and video data, improved the segmentation accuracy [67].

There are few shortcomings in this existing work. First, in order to leverage video data in the model, significant pre-processing is needed. Therefore, most state-of-the-art methods are inapplicable for real-time gesture segmentation of robotic surgery. Second, these methods are very sensitive to the amount of training data and need many parameters to be specified and tuned during the training process [32]. Finally, and perhaps the most importantly, previous works do not model one of the important aspects of human activities, i.e., continuity. Human activities exhibit some inherently continuous properties with a gradual transitions that need to be considered during model building [68, 69].

3.3 Unsupervised Gesture Segmentation and Recognition Method

In this section, we introduce our proposed **Soft** boundary **U**nsupervised **G**esture **S**egmentation and **R**ecognition method, namely **Soft-UGSR**, a novel soft boundary framework to segment and recognize surgical gestures automatically. Figure 3.1 summarizes the propose framework. It has the ability to segment and recognize robotic surgery trajectory path in an unsupervised manner. Soft-UGSR approach is based on bottom-up segmentation technique where it starts from the finest possible segments of data and merged until some criterion is met. In order to find hidden structure of surgical gestures, local probabilistic principal component analysis (PPCA) models used to measure the homogeneity of the segments. We extent fuzzy temporal clustering approach [70] by introducing new compatibility criteria to capture repeated and superfluous action in robotic surgery. To avoid unnecessary preprocessing and assumption, our work focuses on segmentation of kinematic trajectory data of robotic surgery. Within context of the proposed framework, this study makes several contributions. First, we apply new compatibility criteria during cluster merging process to capture superfluous motion, which is inseparable part of surgical task, by measuring dynamic time warping (DTW) distance between two consecutive segments. Second, soft boundary clustering approach models the gradual transition between surgical activities. Third, contrary to previous works, Soft-UGSR uses only kinematic data. This reduces the computation time
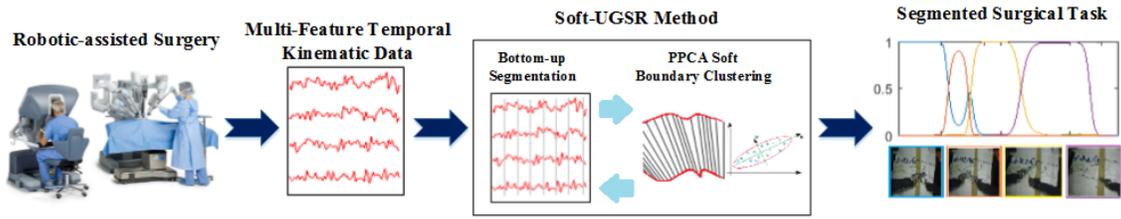
Figure 3.1: Illustration of Soft-UGSR algorithm. Raw kinematic data capture from robotic surgery devices. The proposed method first used bottom up approach to segment data in to some initial cluster. Then Soft boundary approach is applied to find the final segments. This step does recursively until some criteria is met.

of algorithm significantly and make the proposed method applicable for real-time segmentation of surgical tasks. Lastly, we introduce new evaluation metrics which give more insight about performance of segmentation methods. The proposed method could offer an advanced quantitative evaluation of surgical gestures which can be applied to accurately detect and understand surgical gestures without any human intervention. To our knowledge, this problem has not been addressed in literature as comprehensive as the one we presented in this study.

3.3.1 Problem Formulation

We begin by defining the problem and notations for temporal sequence segmentation method. Table 3.1 describes the notations used in this paper. Let us define $\mathbf{T}$ as a $N \times p$ matrix of temporal data, where N refers to length of sequence and p trajectory features. $\mathbf{T}$ consists of n non-overlapping and consecutive segments, $S_{\mathbf{T}}$. Each S_i has a starting point s_i and ending points e_i where

$$S_{\mathbf{T}} = \{S_i(s_i, e_i), 1 \leq i \leq n\} \tag{3.1}$$

We want to divide a given $\mathbf{T}$ into a desired n segments or in other words, find (s_i, e_i) for each S_i. This temporal sequence segmentation introduces a new clustering problem where the points in each cluster must have similar features and also come from successive time points. This problem can be transformed into an optimization problem where the cost function $C(S_i)$, which defines as a distance between data points in i^{th} segment, should be

Table 3.1: Notations used in this chapter

Name	Description
N	length of data (number of frame in a task)
p	number of trajectory features
q	number of selected principle components
$\mathbf{T}$	$N \times p$ matrix of data
$\mathbf{x_j}$	$1 \times p$ matrix of jth data
$\mathbf{v_i}$	$1 \times p$ matrix of center of ith segment
n	number of segment in a data
$S_i(s_i, e_i)$	i^{th} segment start at s_i and end at e_i
$C(S_\mathbf{T})$	segmentation cost function
$\beta_i(t_j)$	membership score of jth data point to the ith segment
$A_i(t_j)$	Gaussian membership of jth data point to the ith segment
Λ_i	matrix of eigenvalues for segment i
$\mathbf{U_i}$	matrix of eigenvectors for segment i
$\mathbf{F_i}$	covariance matrix of segment i
$\mathbf{z_j}$	spatio-temporal representation of jth data equal to $[\mathbf{t_j}, \mathbf{x_j}]$
$\eta_\mathbf{i}$	center of ith segment
μ_{ij}	degree of membership for j^{th} to belong to i^{th} segment

minimized.

$$C\left(S_\mathbf{T}\right) = \sum_{i=1}^{n} C\left(S_i\right) \tag{3.2}$$

The intuitive approach to solve Eq. (3.2) is recursive dynamic programming, but it is computationally expensive, especially for high-dimensional data. Thus, this optimization problem should be solved heuristically. Three methods of sliding window, top-down and bottom-up are widely used in time series segmentation [71]. In this work, we applied bottom-up greedy approach which has been shown to be an efficient and accurate technique. It starts from the finest possible segments of data and merges them until some criterion is met. We will explain this method in more detail in section 3.3.4.

The similarity criterion can be defined in many ways. One way is based on distance between two or more data points belonging to the same cluster. Hence, the cost function in Eq. (3.2) can be written as

$$C\left(S_{\mathbf{T}}\right) = \sum_{i=1}^{n}\sum_{j \in S_i}\beta_i(t_j)dis(\mathbf{x_j}, \mathbf{v_i}) \tag{3.3}$$

where $dis(\mathbf{x_j}, \mathbf{v_i})$ defines as distance between j^{th} data point $\mathbf{x_j}$ and center of ith segment $\mathbf{v_i}$ and $\beta_i(t_j) = 0, 1$ stands for the hard boundary of the jth data point in the ith segment. $\beta_i(t_j)$ is 1 if $s_{i-1} < j \le s_i$ and 0 otherwise. Different methods can be used to define the distance function in Eq. (3.3). In this study, we use PCA-based method. The idea of using PCA method is rooted in its ability to reveal the internal structure of the data in a way that best explains the variance of the data.

3.3.2 Probabilistic PCA Distance Measure

Principal component analysis (PCA) is central in the study of multivariate data [72]. It uses an orthogonal transformation to project high dimensional correlated variables into a linearly uncorrelated variables called principal components. This mapping uses only the first few q nonzero eigenvalues λ_is and the corresponding eigenvectors $\mathbf{U}_i$ of ith cluster's covariance matrix as $\mathbf{F}_i = \mathbf{U}_i\Lambda_i\mathbf{U}_i^T$. Thus, the q-dimensional reduced representation of jth data point $\mathbf{x_j}$ in i^{th} cluster is $\mathbf{y_{i,j}} = \mathbf{W}_i^T(\mathbf{x_j})$, where $\mathbf{W}_i = \mathbf{U}_{i,q}\Lambda_{i,q}^{1/2}$ is the weight matrix contains the first q principal orthogonal axes.

One important limitation of PCA is that it is a non-parametric method which cannot provide us with the probability of each data point $\mathbf{x_j}$ belong to i^{th} cluster. Also for PCA, the covariance matrix needs to be calculated during each iteration. This is very computationally intensive, especially for large scale data. To address these limitations, we applied probabilistic PCA (PPCA) [73]. The log-likelihood of observing the data under PPCA model is defined as

$$\mathcal{L} = \sum_{j=1}^{N}\ln\left(p(\mathbf{x_j}|\eta_i)\right) = -\frac{N}{2}\left\{n\ln(2\pi) + \ln(\det(\mathbf{A_i})) + tr(\mathbf{A_i}^{-1}\mathbf{F_i})\right\} \tag{3.4}$$

where $\mathbf{A_i}$ is

$$\mathbf{A_i} = \sigma_{i,x}^2\mathbf{I} + \mathbf{U_i}\lambda_i\mathbf{U_i}^T \tag{3.5}$$
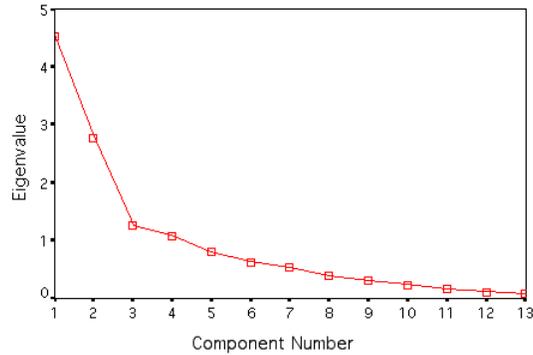
Figure 3.2: Illustrative example for screeplot of randomly generated data.

which is the modified covariance matrix of the ith cluster. The expectation maximization (EM) algorithm is used to solve this problem.

The significant challenge in using PCA is choosing the number of components. In order to address this challenge, one can retain components that cumulatively explain a certain percentage of variation. This can be done by analyzing the eigenvalues of the covariance matrices of the initial segments which can be shown in a scree plot (see Figure 3.2). A scree plot displays the eigenvalues associated with components in descending order versus the number of the components. Therefore, the ordered eigenvalues according to their contribution to the variance of data can be used to choose the right number of principal components in the PPCA algorithm.

3.3.3 Soft Boundary Clustering Algorithm

The cost function in Eq. (3.3) is based on a crisp boundary where data is divided into distinct clusters and each data point can only belong to exactly one of them [74]. However, in many applications, this is not practical where data points can belong to more than one cluster, i. e. each point can be associated to each cluster through a set of membership levels [75]. For example in gesture and motion segmentation, the changes of variables are usually vague and are not focused on any particular motion. In other words, the interval between two consecutive gestures, namely transition motion, can be defined through soft boundaries between segments. In this study, following [70], each data point has a membership level for

belonging to a segment. Hence, $\beta_i(t_j)$ in Eq. (3.3) is the membership score of jth data point in the ith segment. Using Gaussian functions, $\beta_i(t_j)$ can be defined as

$$\beta_i(t_j) = \frac{A_i(t_j)}{\sum_{k=1}^{n} A_k(t_j)} \tag{3.6}$$

where

$$A_i(t_j) = \exp\left(-\frac{(t_j - v_{i,t})^2}{2\sigma_{i,t}^2}\right) \tag{3.7}$$

The algorithm, which is similar to the modified Gath-Geva (GG) clustering [76], assumes that data can be effectively modeled as a mixture of multivariate Gaussian distributions (including time as a variable), so it minimizes the sum of the weighted squared distances between the spatio-temporal data as $\mathbf{z_j} = [\mathbf{t_j}, \mathbf{x_j}]$ and the η_i cluster center as

$$
\begin{aligned}
\underset{\eta_i:i=1,\ldots,n}{\text{minimize}} \quad & \sum_{i=1}^{n}\sum_{j=1}^{N} (\mu_{ij})^m \, dis(\mathbf{z_j}, \eta_\mathbf{i}) \\
\text{subject to} \quad & \sum_{i=1}^{n} \mu_{ij} = 1, \ \forall j \\
& 0 \le \mu_{ij} \le 1, \ \forall i, j
\end{aligned}
\tag{3.8}
$$

where μ_{ij} represents the degree of membership of the observation $\mathbf{z_j} = [\mathbf{t_j}, \mathbf{x_j}]$ in the ith cluster and $m \in [1, \infty)$ is a weighting exponent that determines the fuzziness of the resulting clusters and we choose $m=2$ as suggested by literature [70].

The GG clustering algorithm can be interpreted in a probabilistic framework, since the $dis(\mathbf{z_j}, \eta_\mathbf{i})$ distance is inversely proportional to the probability that the $\mathbf{z_j}$ data point belongs to the ith cluster, $p(\mathbf{z_j}|\eta_i)$. The data are assumed to be normally distributed random variables with expected value $\mathbf{v_i}$ and covariance matrix $\mathbf{F_i}$. Thus, the GG clustering algorithm is equivalent to the identification of a mixture of Gaussians that model the $p(\mathbf{z_j}|\eta)$ probability density function expanded in a sum over the n clusters

$$p(\mathbf{z_j}|\eta) = \sum_{i=1}^{n} p(\mathbf{z_j}|\eta_i)p(\eta_i) \tag{3.9}$$

Different distance norms can be utilized, such as the one that proposed in section 3.3.2. Therefore, with the independency assumption between the time variable and the x_j variables, the distance measure for the ith cluster is based on

$$dis(\mathbf{z_j}, \eta_\mathbf{i}) = \frac{1}{p(\mathbf{z_j}|\eta_i)} \tag{3.10}$$

where

$$p(\mathbf{z_j}|\eta_i) = \frac{1}{\sqrt{2\pi\sigma_{it}^2}} \exp\left(-\frac{1}{2}\frac{(t_j - v_i)^2}{\sigma_{it}^2}\right) \times \frac{\alpha_i}{(2\pi)^{\frac{r}{2}}\sqrt{\det(\mathbf{A_i})}} \exp\left(-\frac{1}{2}(\mathbf{x_j} - \mathbf{v_i})^T \mathbf{A_i}^{-1}(\mathbf{x_j} - \mathbf{v_i})\right) \tag{3.11}$$

which consists of two terms. The first one is the distance between the jth data point and the center of the ith segment in time where the center and the standard deviation of the Gaussian function are

$$v_i = \frac{\sum_{j=1}^{N}(\mu_{ij})^m t_j}{\sum_{j=1}^{N}(\mu_{ij})^m}, \quad \sigma_i^2 = \frac{\sum_{j=1}^{N}(\mu_{ij})^m(t_j - v_j)^2}{\sum_{j=1}^{N}(\mu_{ij})^m} \tag{3.12}$$

The second term represents the distance between the cluster and the data in the feature space where α_i represents the *a priori* probability of the cluster i and $\mathbf{v_i}$ as

$$\alpha_i = \frac{1}{N} \sum_{j=1}^{N} \mu_{ij}, \quad \mathbf{v_i} = \frac{\sum_{j=1}^{N}(\mu_{ij})^m \mathbf{x_j}}{\sum_{j=1}^{N}(\mu_{ij})^m} \tag{3.13}$$

which means the coordinate of the ith cluster center in the feature space and r is the rank of $\mathbf{A_i}$ distance norm corresponding to the ith cluster and can be calculated from Eq. (3.5) in section 3.3.2.

Algorithm 1 *PPCA-based Soft Boundary Clustering*

Require: $\mathbf{T}$, ϵ, initial membership matrix $\mathbf{M}$

Output: Membership matrix $\mathbf{M}$

Repeat $l = 1, 2, ...$

1: Calculate the center of each cluster η_i by Eqs. (3.12),(3.13)

2: Calculate $\mathbf{A_i}$ by Eq. (3.5)

3: Compute the $dis(\mathbf{z_j}, \eta_{\mathbf{i}})$ using Eq. (3.11)

4: Update the membership matrix $\mathbf{M}$ as

$$\mu_{ij}^{l} = \frac{1}{\sum_{k=1}^{n} (dis(\mathbf{z_j}, \eta_{\mathbf{i}})/dis(\mathbf{z_j}, \eta_{\mathbf{k}}))^{2/m-1}}, i \in [1, n], j \in [1, N]$$

until $||\mathbf{M^l} - \mathbf{M}^{l-1}|| < \epsilon$

3.3.4   Unsupervised Segmentation

As mentioned in introduction, we aim to build an unsupervised segmentation model which needs no labeled data as an input to the model. Thus, the usefulness of the proposed clustering algorithm is when the parameters of model can be determined in an unsupervised manner. One of the challenging parameters to be determined is the number of segments. For this purpose, we applied bottom-up technique that has been extensively used in literature for unsupervised segmentation.

The bottom-up algorithm begins by creating the finest possible approximation of the time series, so that $N/2$ segments are used to approximate the N-length temporal sequence. Next, the cost of merging each pair of adjacent segments is calculated, and the algorithm begins to iteratively merge the lowest cost pair until a stopping criteria is met. When the pair of adjacent segments i and $i+1$ are merged, the cost of merging the new segment with its right neighbor must be calculated. In addition, the cost of merging the $i-1$ segments with its new larger neighbor must be recalculated. By applying the bottom-up approach the algorithm selects the number of segments automatically. For this purpose, compaction rules and cluster merging algorithm need to be defined.

Cluster Compaction Rules

The compatibility of two clusters can be measured through the similarity between their PCA results [77]. Consider two segments of data, S_a and S_b. Let the PCA models for each of them consist of q principle components. The similarity between these subspaces is defined based on the sum of the squares of the cosines of the angles between each principal component

$$S_{PCA}(S_a, S_b) = \frac{1}{q} \sum_{i=1}^{q} \sum_{j=1}^{q} \cos^2 \theta(i,j) = \frac{1}{q} tr(\mathbf{U_{i,q}}^T \mathbf{U_{j,q}} \mathbf{U_{j,q}}^T \mathbf{U_{i,q}}^T) \qquad (3.14)$$

The $\mathbf{U_{i,q}}$ and $\mathbf{U_{j,q}}$ subspaces contain the q most important principal components that account for most of the variance in their corresponding data. Thus, $S_{PCA}(a,b)$ can be considered as a measure of similarity between the segment S_a and S_b.

It should be noted that the purpose of the segmentation is also to detect changes in the mean of the variables. Therefore, the distance among the cluster centres need to be taken into account as

$$dis(\mathbf{v_a}, \mathbf{v_b}) = ||\mathbf{v_a} - \mathbf{v_b}|| \qquad (3.15)$$

In the context of gesture and motion studies, repeating an action in order to achieve the goal is commonplace. However, many movements are spurious and should not be considered as separate gestures in the segmentation phase. To find this heuristically, we set the third criteria as a DTW distance between consecutive segments to be less than a predefined threshold. DTW distance can be calculated using Eq. (2.2). We know that if the consecutive segments are repeating actions, they occurred in the same location. Thus, DTW distance can be a good indicator for this behaviour.

All the three compatibility rules can be formulated as follows

$$r_{a,b}(1) = S_{PCA}(S_a, S_b)$$
$$r_{a,b}(2) = dis(\mathbf{v_a}, \mathbf{v_b}) \qquad (3.16)$$
$$r_{a,b}(3) = D_{DTW}(S_a, S_b)$$
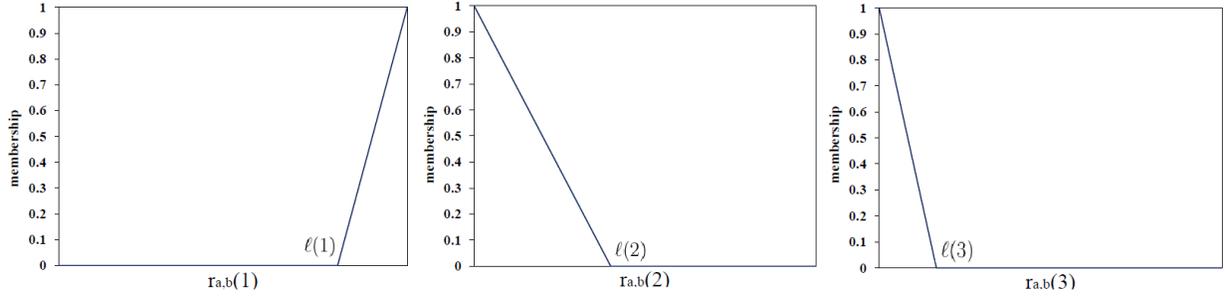
Figure 3.3: Membership functions for three cluster compaction rules.

In general, the three rules should be evaluated for each pair of cluster during merging process. However, since surgical tasks are temporal sequence of motion, the criteria will be only checked for the successor and predecessor segments. This reduces the computation time significantly.

Decision Making Algorithm

The compatibility rules measure various aspects of the similarity of the clusters. Therefore, the overall cluster compatibility should be obtained through an aggregation procedure such as the fuzzy decision making method [78]. The decision goals for each rule must be defined using a fuzzy set. In this work, we consider the triangular membership function as shown in Figure 3.3. The important parameters of the membership functions are the limits of their support, characterized by the $\ell(1)$, $\ell(2)$ and $\ell(3)$ knot points which are given by averaging compatibilities according to

$$\ell(1) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=n+1}^{n} r_{a,b}(1) \tag{3.17}$$

$$\ell(2) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=n+1}^{n} r_{a,b}(2) \tag{3.18}$$

$$\ell(3) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=n+1}^{n} r_{a,b}(3) \tag{3.19}$$

The $\mu_{a,b}(1)$, $\mu_{a,b}(2)$ and $\mu_{a,b}(3)$ membership is obtained by evaluating the membership functions with the value $r_{a,b}(1)$, $r_{a,b}(2)$ and $r_{a,b}(3)$. The overall cluster compatibility is determined by aggregating the three rules as

$$O_{a,b} = \left[\frac{\left(\mu_{a,b}(1)\right)^2 + \left(\mu_{a,b}(2)\right)^2 + \left(\mu_{a,b}(3)\right)^2}{3} \right]^{1/2} \tag{3.20}$$

Given the $\mathbf{O}$ for a pair of cluster, the decision should be made based on comparing element of this matrix with some predefined threshold τ.

Cluster Merging

In each iteration of the bottom-up approach, two adjacent clusters are merged. Thus, we need to compute the new initial parameters of the cluster from the merged clusters directly. This can be done in several ways [79]. The method developed in [71] is applied in this work. The mean, $\mathbf{v_{i*}}$, of the resulting cluster is computed from the individual cluster means

$$\mathbf{v_{i*}} = \frac{N_i}{N_{i*}}\mathbf{v_i} + \frac{N_{i+1}}{N_{i*}}\mathbf{v_{i+1}} \tag{3.21}$$

where $N_i = \sum_{j=1}^{N} \mu_{ij}$, $N_{i+1} = \sum_{j=1}^{N} \mu_{i+1,j}$ and $N_{i*} = N_i + N_{i+1}$. The new covariance matrix , F_{i*}, is calculated from the old covariance matrices F_i and F_{i+1} as

$$\mathbf{F_{i*}} = \frac{N_i - 1}{N_{i*-1}}\mathbf{F_i} + \frac{N_{i+1} - 1}{N_{i*} - 1}\mathbf{F_{i+1}} + \frac{N_i N_{i+1}}{N_{i*}(N_{i*} - 1)}[(\mathbf{v_i} - \mathbf{v_{i+1}})(\mathbf{v_i} - \mathbf{v_{i+1}})^T] \tag{3.22}$$

Based on the obtained results the $p(\mathbf{x_i}|\eta_{i*})$ probabilities can be easily computed. Algorithm 2 shows the unsupervised bottom-up segmentation.

3.4 Performance Evaluation Metrics

One of the most important parts of any model development is validation. In the context of temporal sequence segmentation, designing a good metric to measure segmentation quality is a challenging problem. While the criteria of a good segmentation are often application-

---

Algorithm 2 *Soft-UGSR Segmentation*

Require: temporal data $\mathbf{T}$, N, τ
Output: Segmented sequence of gestures
Step 1: *Create initial segments*
1: **for** $i = 1 : 2 : N$
2:    $\mathbf{S_i} = \mathbf{T}[i : i + 1]$
3: **end**
4: **while** $\mathbf{O} > \tau$ do
Step 2: *Find fuzzy membership matrix* $\mathbf{M}$
5: Apply Algorithm 1
Step 3: *Find cost of merging each pair of segments*
6: **for** $i = 1 : length(\mathbf{S}) - 1$
7:    Calculate $\mathbf{O}_{i,i+1}$ by Eq. (3.20)
8: **end**
Step 4: *Merge compatible segments*
9: $index = \max(\mathbf{O}_{i,i+1})$
10: $\mathbf{S}(index) = merge(\mathbf{S}(index), \mathbf{S}(index + 1))$
11: **delete**($\mathbf{S}(index + 1)$)
12: **update** cluster center by Eqs.(3.21), (3.22)

---

dependent and hard to explicitly define, for many applications the difference between a favorable and inferior segmentation is noticeable. In general, we can divide evaluation methods in to two categories: supervised metrics and unsupervised metrics.

3.4.1 Supervised Metrics

For supervised performance evaluation, we evaluate our proposed segmentation algorithm by comparing the segmented sequence against a manually-segmented reference. In the context of robotic-assisted surgery, the ground-truth is provided by expert surgeons. Thus, the supervised evaluation metrics can be defined as follows (see Figure 3.4),

- *Recall*: the fraction of correctly estimated points over total segment points in ground-truth.

$$Recall = \frac{TP}{TP + FN} = \frac{S_i \cap S_{ref}}{S_{ref}}$$

In the context of segmentation performance evaluation, recall also refers to Seg_{match}.
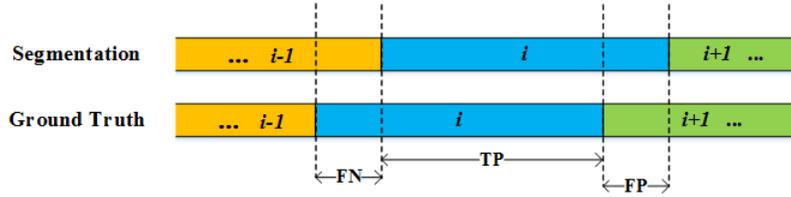
Figure 3.4: Illustrative example to explain supervised evaluation metrics. In this figure TP refers to true positive, FN false negative and FP false positive.

- *Precision*: the fraction of correctly estimated segment points over total segment points by segmentation method.

$$Precision = \frac{TP}{TP + FP} = \frac{S_i \cap S_{ref}}{S_i}$$

- *F-score*: the harmonic mean of precision and recall as

$$F - score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

- *Seg-score*: the size of the similarity between segmentation result and ground truth divided by the union of both as

$$Seg - score = \frac{TP}{TP + FN + FP} = \frac{S_i \cap S_{ref}}{S_i \cup S_{ref}}$$

It should also be mentioned that for the soft boundary segmentation method (Soft-UGSR) we can applied the above supervised metric by adapting area under the membership function in the formulas. The importance of supervised evaluation metrics is that the direct comparison between a segmented sequence and a reference is possible, which is believed to provide a finer resolution of evaluation.

3.4.2 Unsupervised Metrics

Supervised methods give the ability to have a direct comparison between a segmented time sequence and a reference but they require a ground truth. A manually-created reference for time sequence in many applications such as robotic-assisted surgery is intrinsically

subjective, tedious and time-consuming. Therefore, we need to evaluate the performance of the segmentation method when there is no reference time sequence. For this purpose, we applied unsupervised evaluation metrics that do not require a reference, but instead, evaluate a segmented gesture sequence based on how well it matches a broad set of characteristics of surgical gestures. Considering segmentation as a clustering problem, we can quantify the unsupervised metrics using

- *Davies-Bouldin Index (DBI)*: as a ratio of the within-cluster to the between-cluster separation defined as

$$DBI = 1 - \frac{1}{n} \sum_{i=1}^{n} \max_{j \neq i} \left(\frac{\sigma_i + \sigma_j}{dis(\mathbf{v}_i, \mathbf{v}_j)} \right) \quad DBI \in [0, 1]$$

  where n is the number of clusters, $\mathbf{v}_i$ is the centroid of cluster i, σ_i is the average distance of all elements in cluster i to centroid $\mathbf{v}_i$, and $dis(\mathbf{v}_i, \mathbf{v}_j)$ is the distance between two centroids. Algorithms that produce clusters with low within-cluster distances and high between-cluster distances will have a higher DB index. Hence, we can conclude that the higher the DBI the better segmentation.

- *Silhouette Index (SI)*: evaluate the cluster by comparing the average distance within a cluster with the average distance to the points in the nearest cluster as

$$SI(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

  where $a(i)$ is the average dissimilarity of data point i with all other data within the same cluster and $b(i)$ is the minimum average dissimilarity of i to any other cluster, of which i is not a member. The SI is in the range of -1 to 1 and for ease of comparison we rescale it to be in $[0, 1]$. The average $SI(i)$ over all data of a cluster measures the tightness of the cluster segment. Therefore, points with a high SI value are considered

well-clustered.

These metrics can be used to evaluate the Soft-UGSR method by adapting the membership level in the formula [80]. The ability to work without reference sequence of gesture allows unsupervised evaluation to operate over a wide range of conditions. This property also makes unsupervised evaluation uniquely suitable for automatic control of online segmentation in real-time systems, where a wide variety of time sequence trajectory data, whose contents are not known beforehand, need to be processed. Thus, it is very important to evaluate the segmentation methods based on unsupervised metrics.

3.5   Experimental Results

In this section, we report the experimental results of surgical gesture segmentation and recognition using the proposed UGSR method on real robotic-assisted surgery data presented in [1]. We use 38 motion variables from kinematic data captured using the API of the *da Vinci*. The data have manually annotated labels for each trial that show the start and end frame of each gesture and it can be used as a ground truth for validation. We evaluate the proposed method using both supervised and unsupervised metrics and discuss the presented result.

3.5.1   Preprocessing and Parameters Setting

We build our segmentation method directly on kinematic trajectory data to avoid unnecessary data preprocessing. We choose fuzzy membership number $m=2$ as suggested in the literature [70]. Another parameter which should be defined is termination tolerance for PPCA-based soft boundary clustering. If ϵ is chosen to be a very small number, the computation time will be increase and if it is a large number the boundary between segments are not distinguishable. Therefore we choose $\epsilon=0.0001$. Also number of segments is directly depend on the pre-defined threshold for the compatibility matrix. If $\tau=0$, it means we should merge segments even if they are not really compatible and it result in few segments. On the other hand, when $\tau=1$ we end up with many segments because this threshold is very tight.

Hence, our preliminary result suggest that τ should be in the range of 0.65 to 0.8.

As explained in the section 3.3, one of the important initial parameter that needs to be determined is number of principle components (PCs) for the PPCA distance measure. For this purpose we use screeplot which shows in Figure 3.5. It can be observed that suturing and needle passing need almost 10 PCs to explain approximately 99% of variation in data while for knot tying, 5 PCs is sufficient to cover same amount of variation. This can explain the complexity of suturing and needle passing tasks, while knot tying is less complex and fewer eigenvectors can cover the variation in the data.

3.5.2 Surgical Task Gesture Segmentation Result

We illustrate the result of implementing the proposed Soft-UGSR method along with ground truth for suturing, needle passing and knot tying in Figures 3.6, 3.7 and 3.8 respectively. From the results, we observed that each gesture is encoded by a soft boundaries. When a current gesture is going to transfer to a new gesture, the fuzzy membership function $\beta(t)$ of current gesture decreases and the new gesture's membership score increases. Considering the time dimension, each gesture obtains its maximum membership score at the center of a segment. The results also show that gestures with longer event duration generally obtains a more confident segmentation result with higher membership score while for short events, the score decreases.

3.5.3 Performance Evaluation of Proposed Method

As explained in 3.4, we explored two sets of metrics to evaluate segmentation results: supervised and unsupervised. Table 3.2 show the results of proposed methods (UGSR and Soft-UGSR) for each task. For supervised metrics, we compare the segmentation result of proposed method with ground truth that are manually annotated by expert surgeons. From Table 3.2 we observe that the UGSR algorithm can segment the suturing task with an F-score of 72.8% and a segmentation score of 64.8%. The performance improves once we applied Soft-UGSR to 75.8% and 67.4%. The F-score for needle passing is 64% and for knot tying
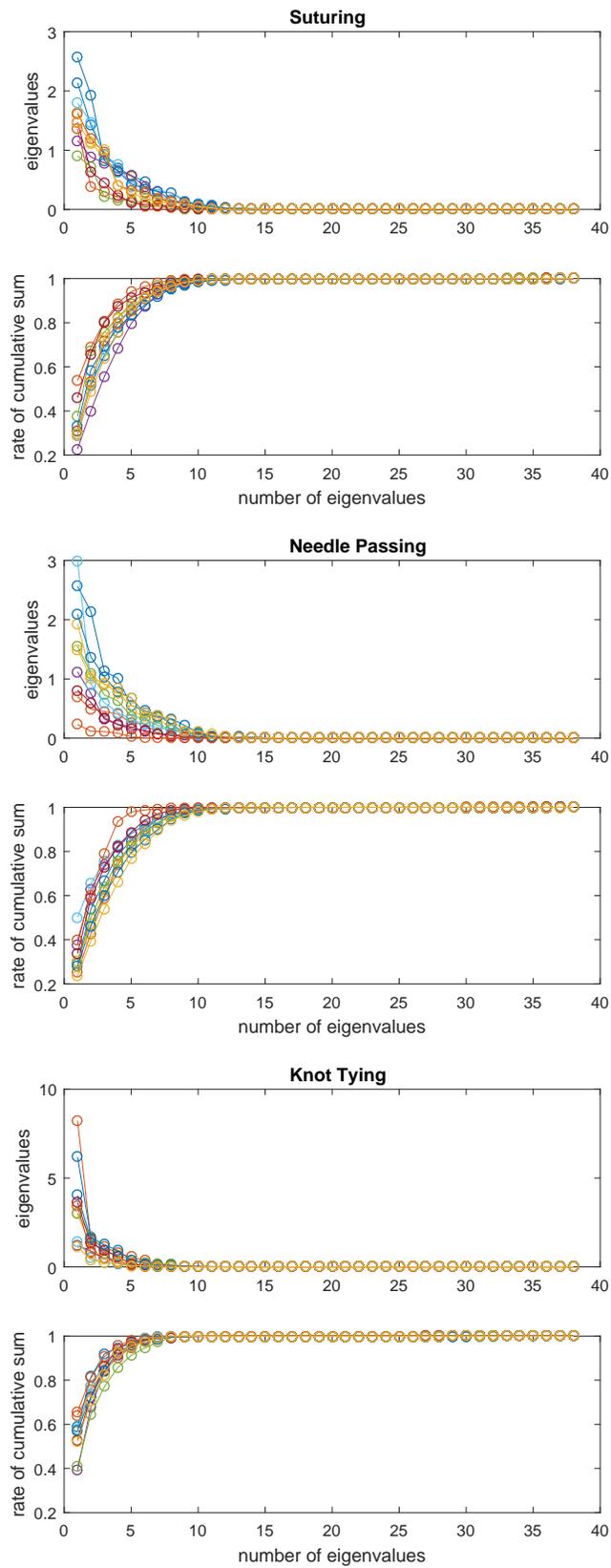
Figure 3.5: Screeplot for three tasks: Suturing, Needle Passing and Knot Tying. This is used to find number of principle components for the PPCA method.
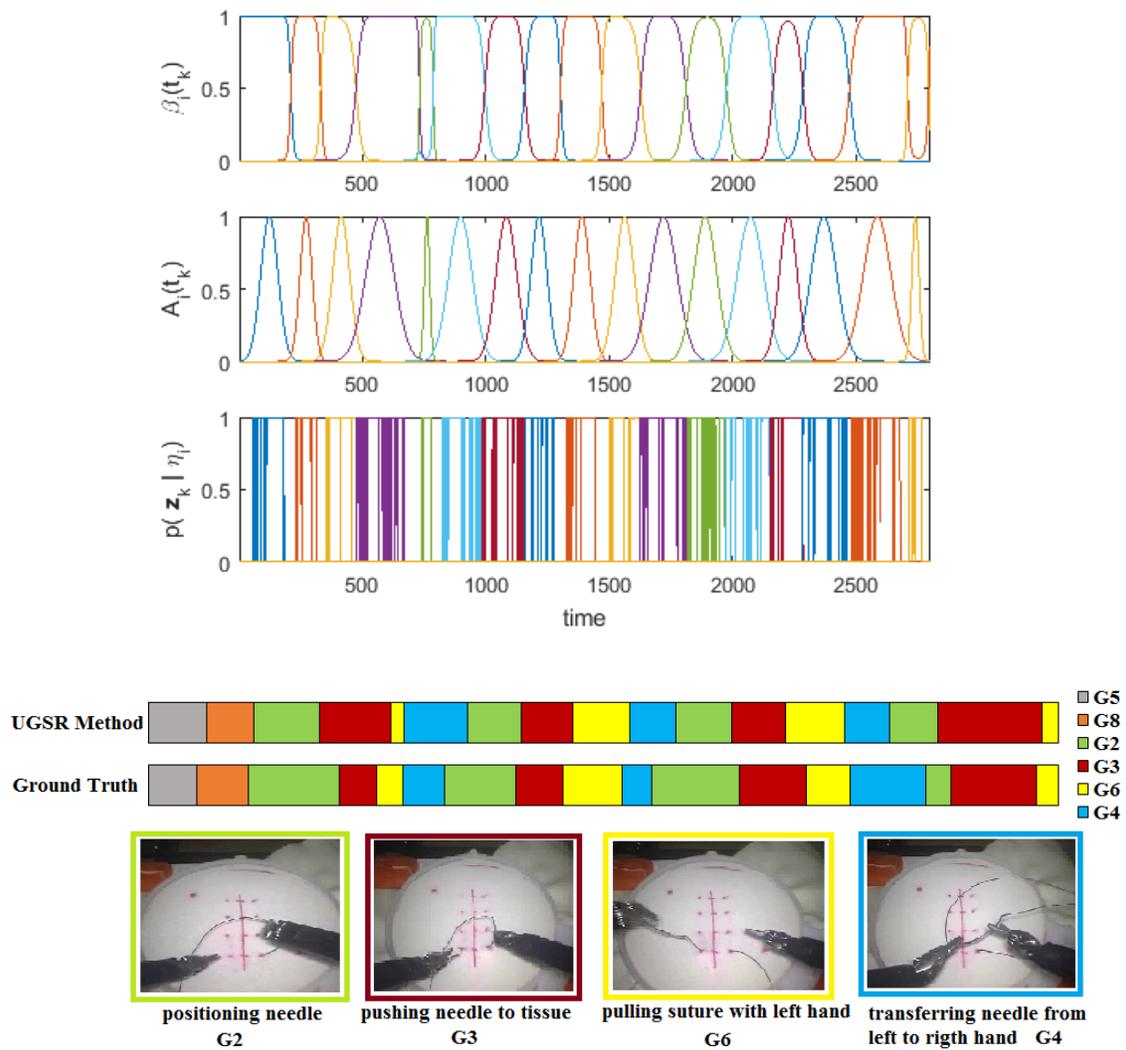
Figure 3.6: Soft boundary segmentation, event-level activity recognition results and comparisons with ground truth for a trial of the suturing task.
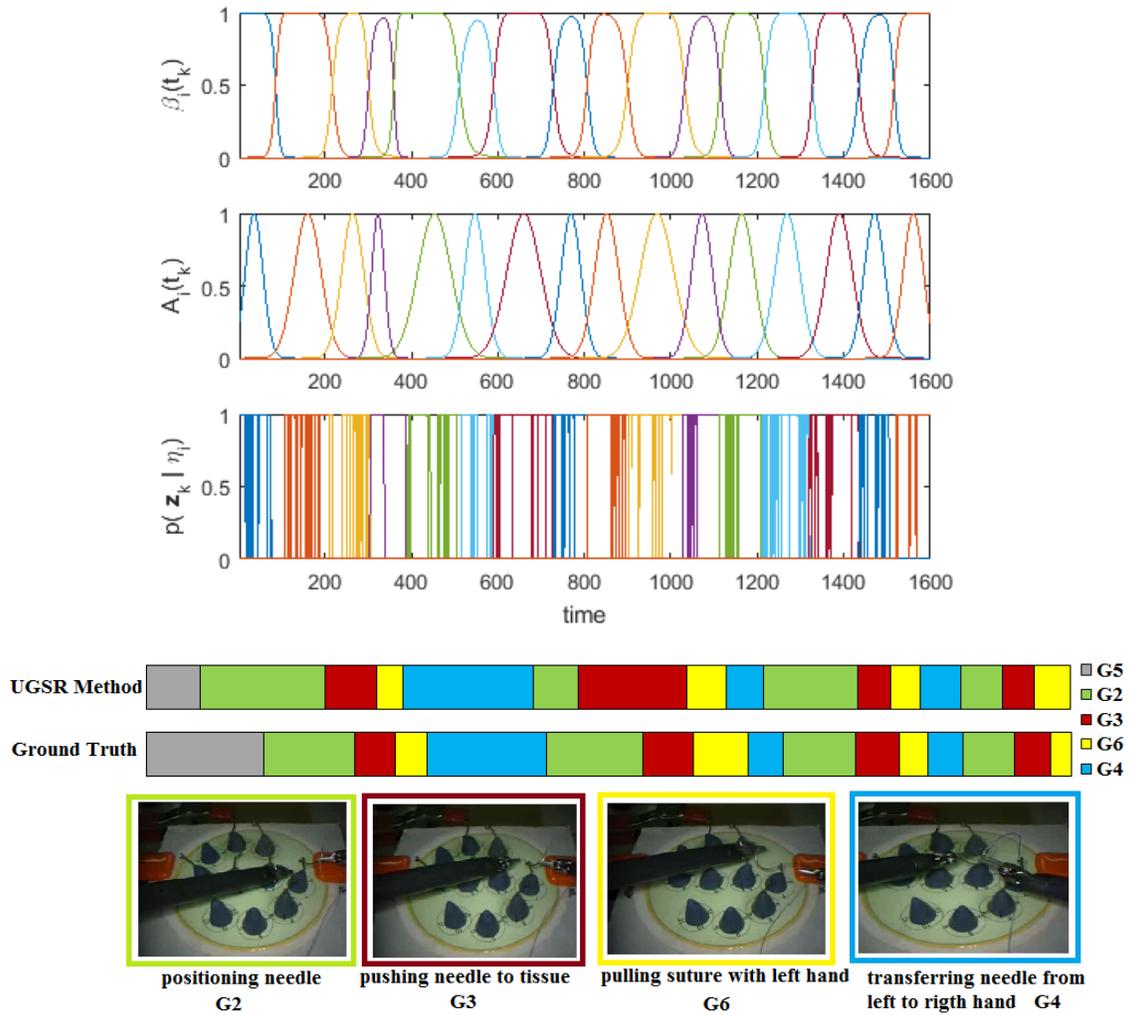
Figure 3.7: Soft boundary segmentation, event-level activity recognition results and comparisons with ground truth for a trial of the needle passing task.
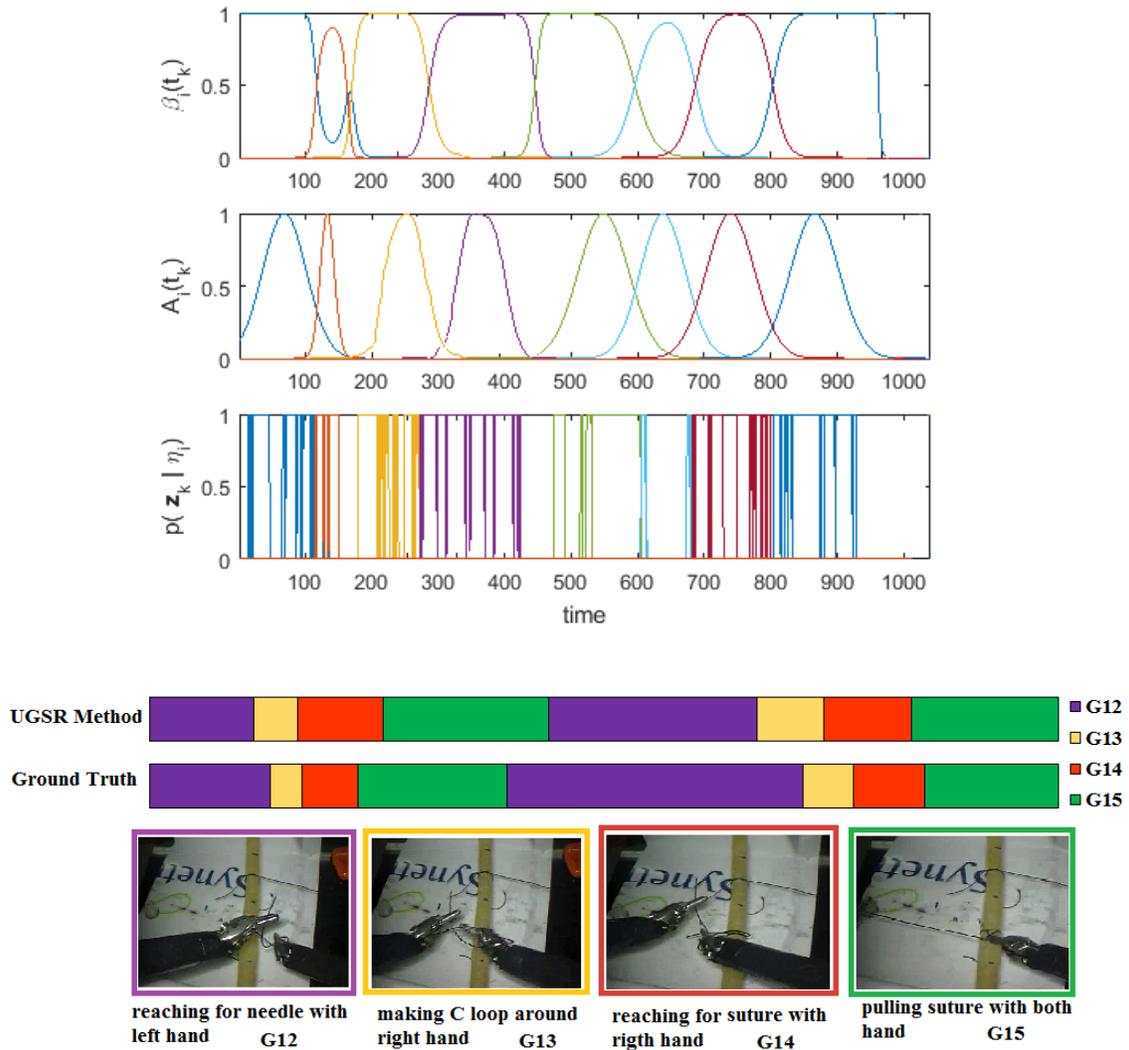
Figure 3.8: Soft boundary segmentation, event-level activity recognition results and comparisons with ground truth for a trial of the knot tying task.

is 73.8% using Soft-UGSR.

Table 3.2: Comparison between performance evaluation of different metrics for each task using proposed segmentation methods.

| Task | Methods | Supervised Metrics | | | | Unsupervised Metrics | |
		Recall	Precision	F-score	Seg-score	DPI	SI
Suturing	UGSR	0.744	0.712	0.728	0.648	0.729	0.647
	Soft-UGSR	0.782	0.735	0.758	0.674	0.706	0.623
Needle Passing	UGSR	0.698	0.679	0.688	0.616	0.698	0.617
	Soft-UGSR	0.712	0.726	0.719	0.640	0.655	0.605
Knot Tying	UGSR	0.794	0.762	0.778	0.693	0.767	0.713
	Soft-UGSR	0.833	0.812	0.822	0.738	0.727	0.687

We also provide the result of performance evaluation for F-score and Seg-score at a more granular level. Tables 3.3, 3.4 and 3.5 show the results of segmentation methods for each gesture in suturing, needle passing and knot tying, respectively. From Table 3.3, we observe that four important gestures, G2, G3, G4 and G6, in suturing and needle passing and all four gestures in knot tying are segmented with relatively high scores.

Table 3.3: F-score and Seg-score for each gesture in suturing task using proposed segmentation methods.

		G2	G3	G4	G5	G6	G8
UGSR	F-score	0.821	0.798	0.777	0.483	0.803	0.473
	Seg-score	0.741	0.715	0.692	0.499	0.718	0.489
Soft-UGSR	F-score	0.847	0.834	0.810	0.576	0.855	0.486
	Seg-score	0.770	0.753	0.725	0.543	0.775	0.498

3.6   Discussion

The results suggest that the proposed method can extract useful information from temporal trajectory data of robotic-assisted surgical tasks. This information can be used to estimate start and end time points of the surgical activities. We observed from results in Table 3.2 that the proposed method has better overall performance for knot tying compared

Table 3.4: F-score and Seg-score for each gesture in needle passing task using proposed segmentation methods.

		G2	G3	G4	G5	G6	G8
UGSR	F-score	0.761	0.766	0.764	0.488	0.755	0.485
	Seg-score	0.696	0.682	0.679	0.497	0.673	0.493
Soft-UGSR	F-score	0.836	0.796	0.795	0.507	0.786	0.504
	Seg-score	0.760	0.711	0.709	0.507	0.702	0.502

Table 3.5: F-score and Seg-score for each gesture in knot tying task using proposed segmentation methods.

		G12	G13	G14	G15
UGSR	F-score	0.792	0.673	0.785	0.836
	Seg-score	0.727	0.605	0.708	0.758
Soft-UGSR	F-score	0.818	0.713	0.825	0.878
	Seg-score	0.751	0.636	0.750	0.809

to suturing and needle passing. This suggests that knot tying is less complex and different surgeons perform this task in a more similar way than suturing and needle passing.

The relatively low performance rate of unsupervised metrics in the Soft-UGSR method compared to UGSR is a consequence of soft boundary segmentation. It is quite clear that soft boundary segmentation provides more insight about continuous aspect of activities and measure the inconsistency in the form of fuzzy membership score. But on the other hand, it reduces the tightness and compactness index of segments.

Furthermore from Tables 3.3-3.5, the final matching evaluation between ground truth (manually annotated by expert surgeons) and the proposed method shows that an accurate recognition of most important surgical gestures can be achieved using the proposed method. For instance, we observe that for suturing and needle passing, gestures such as pushing needle through tissue (G3), pulling suture (G6) and transferring the needle from left to right hand (G4) were appropriately recognized. Conversely, some other gestures such as orienting needle (G8) or moving to center with grip in hand (G5) generated poor results, partly due

to the their short duration and specific characteristics. In other words, these gestures are not part of the task and they are rarely used by a surgeon as a corrective action in order to finish the task. Consequently, proposed segmentation method cannot segment and recognize them correctly.

We should note that although there is a need to define some parameters of the algorithm before its application, it is possible to apply the method for high-dimension data such as temporal sequence of RMIS task even if almost nothing is known about the structure of the sequences in advance. Hence, the proposed algorithm does not rely on any prior information about the surgical task and therefore, extension of the UGSR-based methods to other tasks is straightforward. This is in contrast with previous works that are very sensitive to the amount of training data and parameter tuning.

It is worth mentioning that most of the prior work for gesture segmentation relies on video data [67]. Processing video data in order to make them amenable for any further analysis, is very time consuming. It may take upto few hours depend on length of video [36]. Thus, most of the state-of-the-art methods are inapplicable for real-time gesture segmentation in robotic surgery. However, Soft-UGSR method takes few minutes for suturing and needle passing and only few seconds for knot tying to segment these three surgical tasks with relatively high accuracy. Therefore, from the result in this study we conclude that, in contrary to prior belief, kinematic data offers relevant information for low-level recognition of surgical gestures. This information can be used directly to segment surgical tasks into meaningful gestures. Hence, key advantages of our proposed approach compared to prior methods are requiring minimal preprocessing and providing high segmentation accuracy.

3.7   Conclusion

Surgical gesture segmentation and recognition has many applications including autonomous control of robotic surgery or online haptic or visual feedback to junior surgeon during training. Robotic-assisted surgical tasks are multi-step procedures that consist of complex con-

tinuous activities. These characteristics of RMIS tasks should be appropriately models. To answer this query, we proposed a new approach to automatically segment and recognize surgical gestures. We have also shown the potential of applying soft boundary segmentation method to effectively model surgical activities gradual transitions and segment continuous gestures. The proposed method could offer an advanced quantitative evaluation of surgical gestures that can be applied to accurately detect and understand gestures in robotic surgery without any human intervention. The results of this study, suggest a number of important directions for future work. First, we plan to apply UGSR method at skill level to see the impact of surgeon dexterity on gesture recognition. It is also interesting to demonstrate the algorithm for online gesture segmentation in robotic surgical procedures that consist of different tasks and have more complexity. Finally, though motivated by application in robotic-assisted surgery, the proposed method is also applicable to various other domains such as robotic assembly lines, autonomous vehicle or generally in any system that a human and robot are interacting.

CHAPTER 4: ROBOTIC SURGERY SKILL ASSESSMENT

4.1 Introduction

Despite advances in computer systems and simulation methods today surgical training is still based on direct observation involving expert monitoring [81]. These methods are limited by a lack of consistency, reliability and efficiency due to the subjective nature of experts' observation [82]. Clinical competence of practicing surgeons will always be a matter of public concern [81]. In the last 20 years objective assessment methods have been developed, such as Objective Structured Assessment of Technical Skills (OSATS) [83]. Using OSATS, an expert surgeon gives scores to surgical trainees based on predefined criteria such as flow of surgery, motion time and final product by observing the surgery in person or watching the recorded video of the operation. This technique, however, requires the constant presence and attention of the expert.

Surgical training is undergoing a paradigm shift. More emphasis is being placed on the development of technical skills earlier in training, and all training is becoming competency-based. In this model the trainee advances not based on time spent and observer subjective evaluation, but on acquisition of competence in various skills. Hence, a new method of surgical skill assessment is required to ensure that surgeons have adequate skill level to be allowed to operate freely on patients. Among many possible approaches, those that provide noninvasive monitoring of expert surgeon and have the ability to automatically evaluate trainee's skill are of increased interest. One field of surgery where such techniques can be developed is robotic surgery, as here all movements are already digitalized and therefore easily susceptible to analysis. Once validated in this environment, one can expect these methods to be generalized to all minimally invasive surgery and even open surgery by tracking tool movements.

In spite of the fact that these technologies introduce new challenges for training and evaluation of junior surgeons dexterity, but on the other hand, they open new opportunities

for automated objective skill assessment that was not available before [2]. Robotic surgical systems such as *da Vinci* (Intuitive Surgical, Sunnyvale, CA) [9] record motion and video data, enabling development of computational models to analyze surgical skills through data-driven approaches. However, elaborating such models has always lagged behind. It is, however, quite clear that to develop any framework that automatically evaluates surgical skills, a more rigorous model of surgical procedures is needed [6].

Recent advances in data mining and machine learning research have a huge impact on ongoing clinical decision support system studies [84]. The ability of machine learning method to uncover concealed patterns in huge dataset, like kinematic and video data, offer the possibility to better understand and model surgical data [10]. These models can then frame the premise for creating objective measure of surgical skill levels and consequently evaluate surgeon dexterity. Hence, the key step is to extract meaningful features from quantitative motion data that explains the underlying signatures of surgeon's dexterity. For this purpose, we extend the current motion trajectory parameters such as velocity or distance traveled by introducing new features to quantify variability and complexity of the surgical flow. We believe these features provide more detailed analysis of skill level modality in RMIS. The key differentiation of the proposed method from existing work is that we introduce an automated skill assessment framework based on machine learning methods that has the ability to find complex patterns in surgeon's dexterity. The proposed framework has the important characteristics of the objective skill evaluation such as repeatability, stability and clinical relevance.

4.2 Background and Related Work

While surgical skills are learned in the operating room under the direct supervision of expert surgeons [7], the dexterity component of these skills can be learned in currently available simulations. Evaluation and measurement of dexterity has traditionally been conducted by an expert observer via direct observation and judgment. This type of evaluation is time-

consuming for the expert and can be unreliable [8]. A range of structure-based techniques such as Objective Structured Assessment of Technical Skills (OSATS) [83], Global Assessment of Laparoscopy Skills (GOALS) [85] and Global Rating Score (GRS) [86] have been introduced and validated. Using these evaluation methods, the trainees perform standardized surgical tasks while an expert surgeon evaluates their performance. The evaluations are based on checklist grading and emphasize different aspects of dexterity, such as time to motion, efficiency, tissue handling, overall performance, etc. These methods are still subjective, and require the presence of an expert surgeon. With the advent of minimally invasive surgery (MIS) and robotic-assisted minimally invasive surgery (RMIS), the need for automated objective surgical assessment methods is even more pressing [87]. Although these new technologies offer some advantages, on the other hand they require long and difficult training and pose new challenges for surgical training [88].

A number of researchers developed skill assessment methods by decomposing a surgical tasks into pre-defined surgical gestures [16, 89]. Most existing work in this area uses statistical approaches such as Hidden Markov Model (HMM) [25, 30, 32, 90] and descriptive curve coding (DCC) [91, 92]. Although these methods have the ability to find the underlying structure of MIS/RMIS tasks, they are context-based and suffer from requiring a large number of training samples and complex parameter tuning, causing in a lack of robustness in the results [32].

Most researches in objective surgical skill assessment has been focusing entirely on global motion features (GMF) because of their simplicity in implementation and interpretation [93]. Metrics such as operation time, speed, number of hand movements [94], force and torque signatures [20, 95], path length and motion smoothness [93, 96] have been widely used to identify the relation between the global features and surgical tools movement pattern of expert and novice during Laparoscopic surgery [97].

Although previous work built the foundation of objective surgical skill assessment, the

current state-of-the-art has a few shortcomings. First, they mostly focus on descriptive statistical methods to show the dependency of surgical skill level and GFMs. However, these measures alone are not an adequate proficiency measurement. More advanced techniques such as data mining and machine learning algorithms need to be applied [98]. While machine learning techniques have been used extensively in other fields because of their advantages over traditional statistical methods such as robustness, better prediction ability and higher tolerance to violations of assumptions assumptions (e. c. normality or undependability of data) [11], but it is only recently that these methods have been considered to analyze RMIS tasks [10]. Thus, developing quantitative classification techniques that can automatically and accurately evaluate surgical skills needs to be investigated. In this paper, we address the limitation of previous methods by introducing a skill assessment framework to automatically classify surgeons based on their expertise.

4.3   Proposed Surgical Skill Evaluation Framework

Here, we develop a predictive framework for objective skill assessment based on the trajectory movement of the surgical robot arms. For this, we quantify surgical task by extracting global movement features (GMFs) from the raw motion data for two fundamentals of laparoscopic surgery (FLS) tasks, knot tying and suturing. We apply principle component analysis (PCA) to reduce a set of possible correlated features to a smaller set of uncorrelated synthetic features. Based on the newly developed features, different classifiers, including k-nearest neighbor, logistic regression and support vector machines have been applied. Figure 4.1 summarizes our proposed skill evaluation framework. The classifier with the high accuracy can be used to automatically predict the skill level of surgeon for each task.
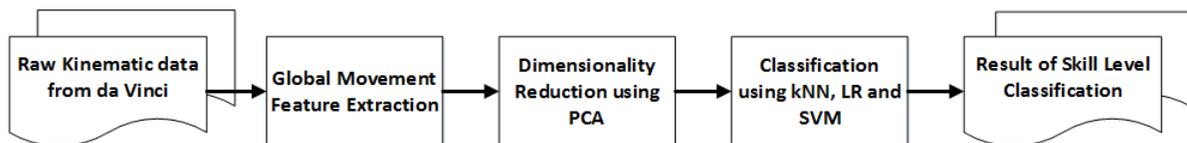
Figure 4.1: Illustration of proposed skill evaluation framework for robotic-assisted surgery.

4.3.1 Preliminary Study on Surgeon's Movement Path

Surgical tasks have different characteristics, such as smoothness, straightness or response orientation, which account for competence while relying only on instrument motion. For instance, studies have shown that the tool motion of an experienced surgeon had more clearly defined features than that of a less experienced surgeon while performing the same task [16]. Figures 4.2 and 4.3 illustrate the Cartesian position plots of an expert and a novice surgeon doing knot tying and four throw suturing on the da Vinci surgical robot.
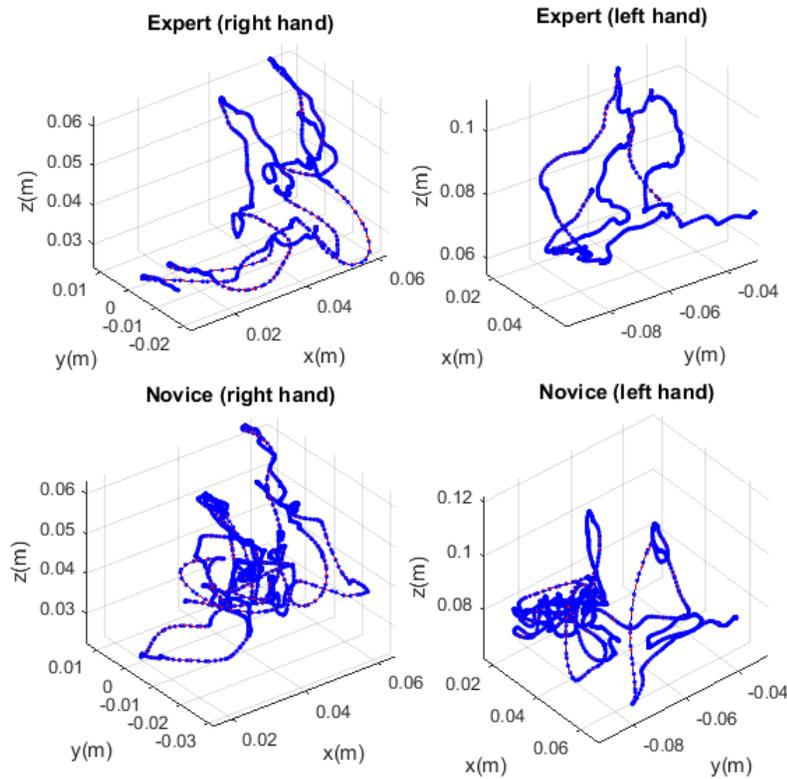
Figure 4.2: Illustration of the 3D Cartesian trajectory path in blue line where red arrows show the direction of movement for an expert and a novice surgeon doing knot tying on the *da Vinci* surgical robot based on data from [1].

In order to have a better understanding about experts' and novices' patterns during RMIS tasks, we compute the pairwise DTW distance (Eq. 2.2) within a group of expert surgeons and compare it with DTW distance between novices and experts. Figure 4.4 presents the

box plot for different RMIS tasks. It shows that expert surgeons do the tasks in a more similar path compared to novices.
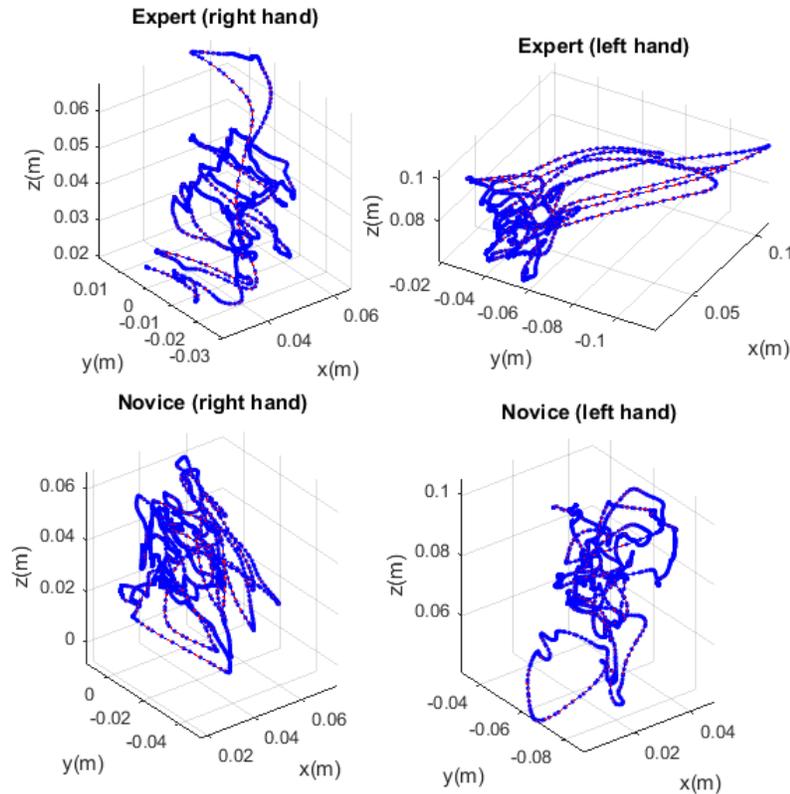
Figure 4.3: Illustration of the 3D Cartesian trajectory path in blue line where red arrows show the direction of movement for an expert and a novice surgeon doing four throw suturing on the *da Vinci* surgical robot based on data from [1].

4.3.2 Global Movement Features (GMFs)

All of the initial studies in 4.3.1 show that there is a significant difference between skill level of surgeon and their movement trajectory. In order to transform these parameters into quantitative metrics, we applied kinematic analysis theory that has been successfully used in previous work to study psychomotor skills [93]. Metrics such as task completion time, length of path, depth perception and velocity can show some aspects of surgeon's dexterity. However, other aspects such as smoothness, curvature, torsion and complexity of the motion need to be quantified. In the following, we explain the six important GMFs from the clinical
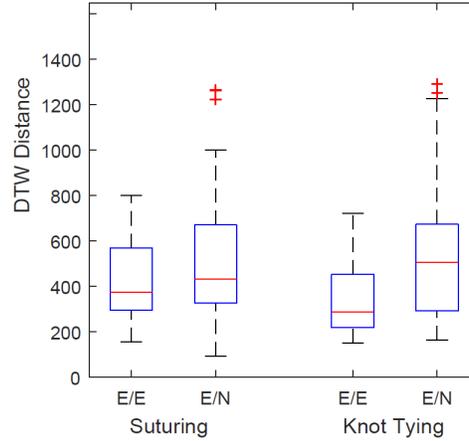
Figure 4.4: Boxplot for DTW distance within expert surgeons (E/E) versus expert and novice (E/N) for suturing and knot tying tasks.

point of view and introduce our new two features that measure the average turning angle and tortuosity of the task.

- **Task Completion Time (TCT):** defined as total time required to complete the task, measured in seconds.

- **Path Length (PL):** is the length of the curve described by the tip of the instrument while performing the task (in cm). We calculate it using sum of all consecutive pairs' Euclidean distance.

- **Depth Perception (DP):** is the total distance traveled by the instrument along its axis (in cm).

- **Speed:** can be defined as the magnitude of velocity and calculated as the rate of position change from previous time step as $dis(p_i, p_{i-1})/\Delta t_i$, where $dis(p_i, p_{(i-1)})$ can be calculated as a Euclidean distance between i^{th} point and of $(i-1)^{th}$ point (in cm/s). Given that the time difference between two consecutive frames in our signal is constant, Δt_i is equal to 1.

- **Motion Smoothness:** is a measure of the rhythmic pattern of acceleration and

deceleration. Smoothness has most often been based on minimizing jerk, the third time derivative of position, which represents a change in acceleration (in cm/s^3).

- **Curvature:** measured the straightness of the path and is calculated at each point by the following equation [99]

$$\kappa_i = \frac{v_i \times a_i}{v_i^3} \tag{4.1}$$

where v_i and a_i are instantaneous velocity and acceleration of the instrument tips respectively, which can be calculated directly by computing the first and second derivatives of the positions of the instrument tips. The curvature measures how fast a curve is changing direction at a given point. For straight movement, the mean of curvature is close to zero while, larger values indicate curved movements.

- **Turning Angle (TA):** is calculated as the direction of the movement with regard to the previous and next time steps. It can be defined as

$$TA_i = \theta_i = Arccos\left(\frac{u_{i-1}.u_{i+1}}{||u_{i-1}||||u_{i+1}||}\right) \tag{4.2}$$

where u_{i-1} is the vector from p_{i-1} to p_i and u_{i+1} is vector from p_i to $p_{(i+1)}$ as shown in Figure 4.5.
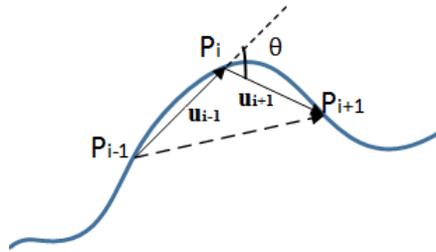
Figure 4.5: Illustration of computing turning angle for movement trajectory features.

- **Tortuosity:** is a property of a curve being twisted or having many turns. It has been used successfully in variety of research such as analyzing animal path [100], evaluating the performance of human robot interaction [101] or distinguishing cognitive impair-

Figure 4.6: Example of measuring tortuosity by walking a pair of dividers of a certain size along the path.

ment through walking behavior [102]. We introduce it as a new metric, to measure the path complexity or in other words, movement path variability in robotic surgical task. Tortuosity can be quantified using the Fractal Dimension (F) [103] where a ratio of a pattern changes with the scale at which it is measured, providing a statistical index for variability. More specifically, the length of the path is measured by walking a pair of dividers of a certain size along the path (see Figure 4.6). If this is done for larger and larger dividers, then the slope of the plot for log (path length) versus log (divider size) is $1 - F$, yielding one overall estimate for F over a range of scales [100]. Figure 4.7 shows the influence of the divider step size on the path length for a suturing task.

Figure 4.7: Example for influence of the divider step size on the path length for a suturing task.

Hence, applying regression to measure the slope can be showed as,

$$\log(D_i) = b + a \log(S_i) \quad i \in [1, 2, ..., n] \tag{4.3}$$

where D_i is the total distance, S_i is the currently employed measuring system, a is the slope of the regression line and n is the number of different scales employed to calculate the total distance of a trajectory path. Thus, F is then calculated as $F = 1 + a$. The Fractal Dimension for movement paths lies between 1 and 2 where F is 1 when the path is straight and 2 when the path is so tortuous that it occupies a whole plane (Brownian motion). As an example in Figure 4.3, the path of an expert's right hand has a tortuosity of 1.14 while the path tortuosity of a novice's right hand is 1.67. In order to make this metric more robust to different measuring scales, the F value is computed twice by starting to measure total distance from two ends of a trajectory [104]. In this study, we use the average of F to measure the complexity of trajectories.

4.3.3 Dimensionality Reduction Using PCA

We extract GMFs for both hands using Cartesian positions of right and left patient-side manipulator end-effectors of *da Vinci* arms. Speed, motion smoothness, turning angle and curvature are temporal features and were calculated for each point in data. Therefore, the descriptive statistics including mean and standard deviation are derived for these features. Finally, a total of 23 global movement features are derived from each trajectory: 12 spatial characteristics of tool tip movement (including length path, depth perception as well as mean and standard deviation of speed and motion smoothness for each hands) and 10 features which show the curvature and torsion of movement (including tortuosity with mean and standard deviation of turning angle and curvature for each hands) and task completion time. Then, we employed principle component analysis (PCA) to reduce a set of possible correlated features to a smaller set of uncorrelated synthetic features [105]. As a result, the original feature set is reduced to 10 principal components, which together contribute 95%

of the original information. For each task the new features and corresponding surgical skill level (expert, novice) are used to train a classifier.

4.3.4 Surgical Skill Classification Methods

Features that are extracted in the previous section are used to quantify the movement pattern of surgeons with different levels dexterity. Our aim is to build a discriminative model to differentiate between surgeons with different levels of expertise while doing RMIS tasks. Surgeons are categorized into two skill levels, expert and novice. Thus, this is a binary classification problem that can be resolved by machine learning algorithms. In particular, we compare three frequently used machine learning techniques, k-nearest neighbor (kNN) [44], Logistic regression [106] and Support Vector Machine [107].

k-Nearest Neighbor (kNN)

The first classifier that we used is k-nearest neighbor. The principle of this technique is to predict the label for the new point based on the closest distance to predefined number (k) of training samples. kNN classifier is an instance-based learning where instead of constructing a general model, it simply stores instances of training data. During the classification phase, the majority of the k nearest neighbors for each point is computed. Thus, the label for the query point is assigned based on the most representatives within the nearest neighbors of the points. We examined different k where the best results were obtained with $k=3$.

Logistic Regression (LR)

One of the most well-established statistical models is Logistic regression where the dependent variable is categorical. In this model, logit transformation of a linear combination of features is used to resolve a binary classification problem. Formally, the logistic regression model can be formalized as

$$p(x) = \frac{1}{1 + e^{(-(\beta_0 + \beta.x))}} \tag{4.4}$$

where β is the coefficient for corresponding x feature and $p(x)$ is the probability of belonging to one of the classes.

Support Vector Machine (SVM)

Support vector machine (SVM) is an important classification method that constructs a hyperplane and tries to maximize the margin that separates two classes of data shown as $2/||\omega||$ (see Figure 4.8).
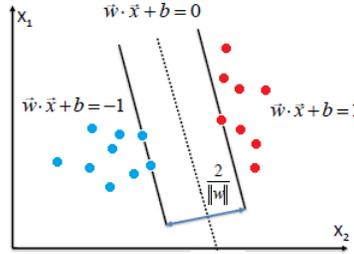
Figure 4.8: Illustration of support vector machine (SVM) method for two features.

The ability to learn a non-linear separable function by mapping data to higher dimension space makes this classifier unbeatable for some classification problems. Linear SVM can be formalized as

$$\text{minimize} \quad \frac{2}{||\omega||}$$
$$\text{subject to} \quad y_i(\omega \cdot x_i + b) \geq 1 \quad \forall i = 1, ..., n$$

(4.5)

where y_i is the class label for i^{th} data. In order to solve the non-linear classification problem, SVM uses a kernel transformation (see Figure 4.9). In this study we applied radial basis function (RBF) which is one of the most popular kernel functions used in SVM [108], defined as

$$K(x_i, x_j) = e^{(-\gamma||x_i - x_j||^2)}$$

(4.6)

where γ controls the width of RBF function. The γ parameters can be seen as the inverse of the radius of influence of samples selected by the model as support vectors. If γ is too large, the radius of the area of influence of the support vectors only includes the support vector itself. While for a very small value of γ, the model is too constrained and cannot capture the complexity or "shape" of the data. The region of influence of any selected support vector would include the whole training set. It is suggested [109] to choose γ as inverse of number

of features. Therefore in this study we set $\gamma=0.1$.

Another important parameter in SVM algorithm is C, the penalty associated to the instances which are either misclassified or violates the maximal margin. Therefore Eq. (4.5) can be rewrite as

$$
\begin{aligned}
\text{minimize} \quad & \frac{2}{||\omega||} + C \sum_{i=1}^{n} \xi_i \\
\text{subject to} \quad & y_i(\omega \cdot \phi(x_i) + b) \geq 1 - \xi_i \quad \forall i = 1, ..., n \\
& \xi_i \geq 1 \quad\quad\quad\quad\quad\quad\quad \forall i = 1, ..., n \\
\text{where} \quad & \phi(x_i)^t \cdot \phi(x_j) = K(x_i, x_j)
\end{aligned}
\tag{4.7}
$$

where ξ_i is is the smallest non-negative number satisfying $y_i(\omega \cdot x_i + b) \geq 1 - \xi_i$ and C is a regularization term, which provides a way to control over-fitting. If C becomes large, it is unattractive to not respect the data at the cost of reducing the geometric margin and on the other hand, when it is small, it is easy to account for some data points with the use of slack variables and to have a fat margin placed so it models the bulk of the data. In this study we set $C=1$.
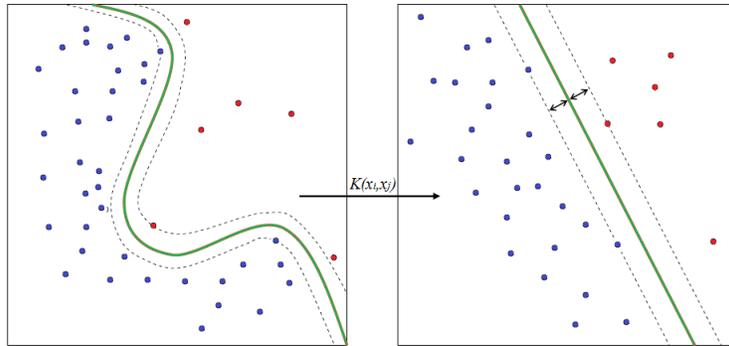
Figure 4.9: Illustration of kernel function in support vector machine (SVM) method for two features.

4.4 Experimental Results

We implement our model on real robotic surgery data presented in [1] as described in detail in first chapter. Data includes manual annotation for surgeons skill based on a global rating score (GRS). Based on the scores, surgeons are divided into two categories of experts and novices.

4.4.1 Performance Evaluation

Classifier validation was conducted using two model validation schemas as suggested in [1]. Leave-one-super-trial-out (LOSO), where one trial for each of the surgeons is left out and leave-one-user-out (LOUO), where all the trials from one surgeon are left out for testing. While the first validation method evaluates the robustness of a method for repeating a task, the second schema evaluates the robustness of a method when a subject is not previously seen in the training data. The performance of the different classification methods was determined by classification accuracy, which is expressed as a percentage of surgeons that their skill level are correctly classified.

4.4.2 GMFs Descriptive Analysis

We start our skill assessment analysis by providing some explorative statistics on GMFs extracted from RMIS motion trajectory data. The box plots for each task performed by novices and experts are shown in Figures 4.10 and 4.11. The plots shows that for some features such as curvature and turning angle in knot tying or tortuosity and smoothness in suturing there is a clear distinction between novices and experts. From Figure 4.10 and 4.11, we observe that all features (except time, path length and tortuosity) are higher for expert surgeons compared to novices.

4.4.3 Surgical Skill Classification

The results of performing three classification methods, kNN, logistic regression and SVM using spatial features (S), curvature-based features (C) and combination of both based on two model validation schemas, LOSO and LOUO for knot tying and suturing are shown in
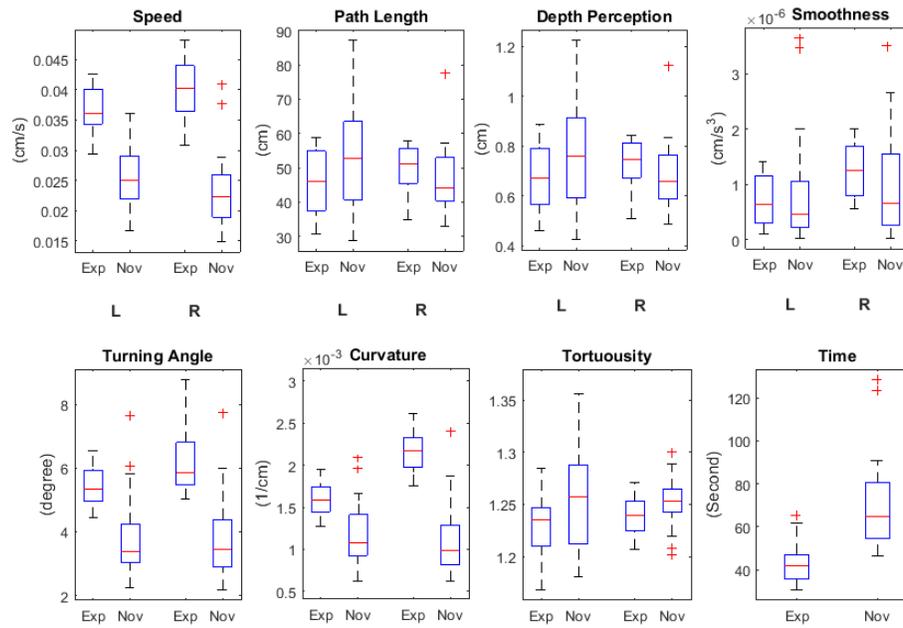
Figure 4.10: Box plots of Exp (experts) and Nov (novices) in knot tying for L (left hand) and R (right hand).
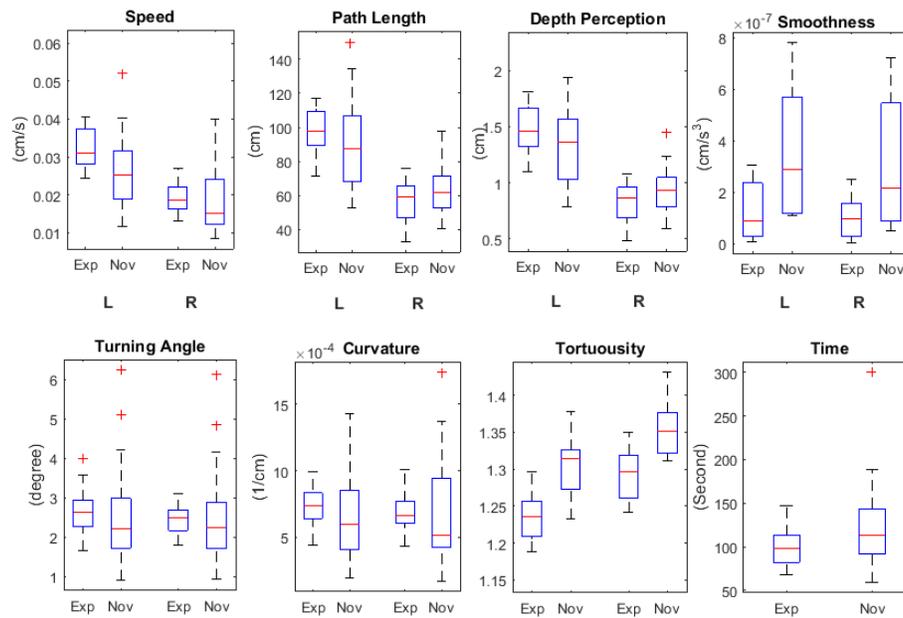
Figure 4.11: Box plots of Exp (experts) and Nov (novices) in suturing for L (left hand) and R (right hand).

Tables 4.1 and 4.2, respectively. The best accuracy was obtained for the combination of all global movement features for both tasks. Table 4.1 shows that for knot tying the highest overall accuracy for LOSO is 83.3% and for LOUO is 77.9%. From Table 4.2, the best overall accuracy that has been achieved for suturing is 90.5% in LOSO and 79.8% in LOUO.

Table 4.1: Accuracy of skill level assessment framework for knot tying using two validation schema (LOSO and LOUO) based on spatial motion features (S), curvature features (C) and combination of both (S+C).

		Novices			Experts			Overall		
		kNN	LR	SVM	kNN	LR	SVM	kNN	LR	SVM
	S	0.722	0.722	0.633	0.778	0.722	0.709	0.750	0.722	0.626
LOSO	C	0.761	0.769	0.691	0.793	0.776	0.713	0.777	0.793	0.673
	S+C	0.753	**0.792**	0.711	**0.864**	0.854	0.777	0.821	**0.823**	0.754
	S	0.657	0.660	0.651	0.662	0.682	0.742	0.660	0.671	0.656
LOUO	C	0.630	0.690	0.751	0.712	0.685	0.799	0.671	0.687	0.747
	S+C	0.695	0.687	**0.753**	0.763	0.716	**0.805**	0.729	0.702	**0.779**

Results also show that logistic regression for LOSO and SVM for LOUO model validation schema provide the best classification performance for both tasks compared to other methods. It should also be noted that for knot tying, 88.9% of experts and 79.2% of novices were classified correctly in LOSO while for LOUO, the classification accuracy reduce to 80.5% and 75.3% for experts and novices, respectively. For suturing, using LOSO as a validation schema, 95.2% of experts and 88.9% of novices are correctly classified. For LOUO, we achieved the accuracy of 81.2% for experts and 74.7% for novices.

4.5 Discussion

The result of descriptive analysis (Figures 4.10 and 4.11) illustrate several important aspects of surgical skill assessment in RMIS. First, we can conclude that, contrary to prior belief [96], not all global features should be seen as cost functions, where a lower value describes a better performance. For instance, we observed that experts have a higher curvature compared to novices in both suturing and knot tying. This can be explained by looking at

Table 4.2: Accuracy of skill level assessment framework for suturing using two validation schema (LOSO and LOUO) based on spatial motion features (S), curvature features (C) and combination of both (S+C).

		Novices			Experts			Overall		
		kNN	LR	SVM	kNN	LR	SVM	kNN	LR	SVM
LOSO	S	0.667	0.722	0.645	0.857	0.847	0.679	0.769	0.790	0.635
	C	0.742	0.832	0.671	0.952	0.877	0.719	0.846	0.855	0.695
	S+C	0.833	**0.889**	0.693	**0.952**	0.905	0.787	0.897	**0.899**	0.754
LOUO	S	0.639	0.669	0.642	0.683	0.730	0.705	0.660	0.697	0.671
	C	0.706	0.679	0.699	0.720	0.771	0.795	0.713	0.725	0.775
	S+C	0.687	0.697	**0.747**	0.750	0.789	**0.812**	0.719	0.744	**0.798**

Figures 4.2 and 4.3. For suturing (Figure 4.3), we can see that expert surgeons make a decisive sharp turn with their left hands which can translate as surgeon's dexterity in working with the left hand. Also, for both knot tying and suturing, the path length for the left hand is greater for experts compared to novices, which may suggest that this pattern is an important property because it gives the surgeon enough room for planning and performing further movement. In addition, although, all the surgeons in this study were right-handed, but the results show that features extracted from the non-dominant hand can be equally, if not more, discriminative than the dominant hand's features. This is in complete agreement with literature in task skill acquisition where dexterity can be assessed base on non-dominant hand performance [93].

From the results shown in Tables 4.1 and 4.2, the classification accuracy improves when a combination of spatial and curvature features are used. This is consistent with previous studies [93, 97] that emphasized, task completion time and distance traveled are insufficient to explain all aspects of surgical assessment. The results from this study clearly suggest that these additional objective measures of robotic surgical performance, such as curvature, turning angle and particularly tortuosity can distinguish expert and novice surgeons. Hence, these features can be applied globally on RMIS tasks as they have the potential to identify additional aspects of surgeon skill level which cannot be quantified by task completion time

and distance traveled alone.

Tables 4.1 and 4.2 show that, for almost all experiments, the best overall accuracy was obtained from logistic regression for LOSO schema while SVM gives the best result for LOUO. It should be noted that the results of LOUO provide an insight into the ability of classification algorithm to evaluate the skill level of a surgeon that was not in the training set. Therefore, we can conclude that finding underlying patterns of different surgeons even with the same skill levels are very challenging.Hence, simple classification methods such as kNN or logistic regression do not have the ability to find the discriminative pattern and more sophisticated method such as SVM with nonlinear kernel (e.g. RBF) is needed to assess the skill level of surgeons who are not previously seen in the training data. In other word, SVM has more generalizability ability in the context of surgical skill evaluation compared to other classification methods.

From the results it can be observed that the overall accuracy of surgical skill classification decreased 6% for knot tying and 11% for suturing when we switched from the LOSO validation schema to LOUO. This suggests that surgeons with the same level of expertise perform knot tying in a more similar way compared to suturing. It is worth mentioning that the overall classification accuracy for suturing is higher than knot tying. We can conclude that skill levels of surgeons are very discriminative in complex tasks such as suturing because these tasks require higher dexterity that make the surgeon's level of expertise more distinguishable.

4.6   Conclusion

In this study, we have described the development of an automated objective skill assessment method based on global movement features extracted from trajectory motion data of the surgical robot arms. The global features we used to build our classifiers are time to complete task, path length, depth perception, speed, smoothness and curvature. We also introduced two new features for the first time in the context of RMIS, turning angle and

tortuosity which turn out to have a good discrimination ability to separate novices from experts. Previous attempts in objective surgical skill assessment mostly used simple statistical methods. However, robotic-assisted surgical tasks have specific complexity which cannot be modeled effectively unless more advanced machine learning methods such as SVM are employed. We found no other studies that were able to distinguish between dexterity levels of surgeons with the degree of accuracy presented in this study.

This study demonstrates the ability of machine learning methods to automatically distinguish between expert and novice performance in robotic-assisted surgical tasks. It is generally accepted that the skill levels of surgeons vary and each surgical task has different levels of complexity. This complexity is not only captured through the global features extracted from trajectory movement data such as tortuosity, but also through more advanced machine learning methods that are needed to model the underlying pattern of surgical skill level. The results presented in this study could form a basis for decision support tools that effectively, objectively and automatically evaluate surgeon's dexterity and provide more personalized skill assessment and online feedback to trainees based on their performance. Furthermore, the proposed method can be applied on a more granular level of tasks in roboticassisted surgery, such as surgical gestures, to provide more insight into the skill level differences of surgeons. Future research could focus on performing more validation studies with a larger number of participants for different surgical tasks which would yield a larger training set with the potential for improving the classification result.

CHAPTER 5: CONCLUSION AND FUTURE RESEARCH

In this chapter, we present a summary and future directions of the presented research that may stem from our work.

5.1 Summary

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. The ultimate goal of these devices is to program the surgical robot to perform certain difficult or complex surgical tasks in an autonomous manner. It is, however, quite clear that to develop any automatic control system, a more detailed comprehension of the surgical procedures is needed. Thus the key step is to develop techniques that are capable of accurately recognizing fundamental surgical tasks and gestures more intelligently. Surgical gestures need to be quantified to make them amenable for further study. Recent advances in data mining and machine learning research for uncovering concealed patterns in huge dataset, like kinematic and video data, offer the possibility to better understand, analyze and model surgical gestures from a system point of view. Current systems like *da Vinci* record motion and video data, enabling development of computational models to recognize and analyze surgical performance through data-driven approaches. In this dissertation, we shed lights on some of the barriers and challenges toward autonomous robotic-assisted surgery and skill assessment by developing novel methods based on computational intelligence techniques.

5.1.1 Robotic Surgery Task and Gesture Classification

Robotic-assisted surgery allows surgeons to perform many types of complex operations with greater precision and flexibility than is possible with conventional surgery. Despite these advantages, in current robotic surgery systems, surgeon communicates with the device directly and manually. To allow the robot to adjust parameters such as camera position, the system needs to know what task the surgeon is currently performing automatically. To address this challenge, we developed a task and gesture recognition framework for three

important surgical tasks including suturing, needle passing and knot tying, based on temporal kinematic signals captured during operations. It classifies robotic-assisted surgical tasks and recognizes surgical gestures by integrating temporal sequence similarity measure techniques such as Dynamic Time Warping (DTW) with the well-known k-nearest neighbor (kNN) classification method. The proposed framework is fast, accurate and robust (even when only a small portion of temporal data is available), all of which makes it applicable for any adaptive control system, such as camera control, in robotic surgery.

5.1.2 Unsupervised Surgical Gesture Segmentation

Surgical gesture segmentation and recognition has many applications including autonomous control of robotic surgery or online haptic or visual feedback to junior surgeons during training. Robotic-assisted surgical tasks are multi-step procedures that consist of complex continuous activities. These characteristics of RMIS should be appropriately modeled. We propose a novel unsupervised gesture segmentation and recognition (UGSR) method. To avoid unnecessary preprocessing and assumption, our work focuses on segmentation of kinematic trajectory data that contains temporal sequences of surgical gestures. We applied soft boundary algorithm (Soft-USGR) to explicitly model the gradual transition between activities in the context of robotic surgery. The proposed method could offer an advanced quantitative evaluation of surgical gestures that can be applied to accurately detect and understand surgemes in robotic surgery without any human intervention.

5.1.3 Robotic-assisted Surgery Skill Assessment

Evaluating surgeon dexterity has predominantly been a subjective task. Developments of objective methods for surgical skill assessment are of increased interest. Recently, with technological advances such as robotic-assisted surgery, new opportunities for objective and automated assessment frameworks have arisen. We developed predictive framework for objective skill assessment method based on trajectory movement of the surgical robot arms. For this purpose, we extend the current motion trajectory parameters such as velocity or

distance traveled by introducing new features such as tortuosity to quantify smoothness, variability and complexity of the surgical flow. Based on the newly developed features, different classifiers, including k-nearest neighbor, logistic regression and support vector machines have been applied. The proposed method addressed some of the limitations in previous work and gives more insight about underlying patterns of surgical skill levels. The experimental results showed the ability of machine learning methods to automatically distinguish between expert and novice performance in robotic-assisted surgical tasks with relatively high accuracy.

5.2   Future Research Direction

We believe that this dissertation will encourage new research work in the area of autonomous robotic surgery. The following are immediate research opportunities and promising directions that can be pursued following our work.

First, the results of robotic surgery task and gesture classification can be used to develop real-time adaptive control systems that will be more responsive to surgeons' needs by identifying and predicting future movements of the surgeon. For example, to have an automatic camera control we need to know what the surgeon is doing in order to predict the next movement and adjust the camera mode. Therefore, we plan to implement the DTW-kNN algorithm on *da Vinci* to recognize task and adjust the camera automatically based on that.

It is also interesting to apply unsupervised gesture segmentation and recognition (UGSR) method at skill level to see the impact of surgeon dexterity on gesture recognition. It would be interesting and fruitful to demonstrate the result for online gesture segmentation in complete robotic surgical procedures that contains different tasks with more complexity.

The results presented for surgical skill assessment could form a basis for decision support tools that effectively, objectively and automatedly evaluate surgeon's dexterity and give more insight about skill level differences. Furthermore, the proposed method can be applied on more granular level of task in robotic-assisted surgery, such as surgical gestures, to provide personalized skill assessment and online feedback to trainees based on their performance.

Future research could focus on performing more validation studies with a larger number of participants for different surgical tasks which would yield a larger training set with the potential for improving the classification results.

Finally, though motivated by application in robotic-assisted surgery, the proposed methods are also applicable to various other domains such as robotic assembly lines, autonomous vehicle or generally in any system that human and robot interact.

REFERENCES

[1] Y. Gao, S. S. Vedula, C. E. Reiley, N. Ahmidi, B. Varadarajan, H. C. Lin, L. Tao, L. Zappella, B. Béjar, D. D. Yuh, et al. 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] F. Lalys, P. Jannin. Surgical process modelling: a review. *International Journal of Computer Assisted Radiology and Surgery*, 9(3):495–511, 2014.

[3] K. Moorthy, Y. Munz, A. Dosis, J. Hernandez, S. Martin, F. Bello, T. Rockall, A. Darzi. Dexterity enhancement with robotic surgery. *Surgical Endoscopy And Other Interventional Techniques*, 18(5):790–795, 2004.

[4] G. Dakin, M. Gagner. Comparison of laparoscopic skills performance between standard instruments and two surgical robotic systems. *Surgical Endoscopy And Other Interventional Techniques*, 17(4):574–579, 2003.

[5] A. Cuschieri. Whither minimal access surgery: tribulations and expectations. *The American Journal of Surgery*, 169(1):9–19, 1995.

[6] A. K. Pandya, L. A. Reisner, B. King, N. Lucas, A. Composto, M. Klein, R. D. Ellis. A review of camera viewpoint automation in robotic and laparoscopic surgery. *Robotics*, 3(3):310–329, 2014.

[7] R. K. Reznick. Teaching and testing technical skills. *The American Journal of Surgery*, 165(3):358–361, 1993.

[8] A. Darzi, S. Smith, N. Taffinder. Assessing operative skill: needs to become more objective. *BMJ: British Medical Journal*, 318(7188):887, 1999.

[9] G. Guthart, J. K. Salisbury Jr. The intuitive TM telesurgery system: Overview and application. In *Robotics and Automation, IEEE International Conference on (ICRA), San Francisco, California, USA*, pp. 618–621. 2000.

[10] Y. Kassahun, B. Yu, A. T. Tibebu, D. Stoyanov, S. Giannarou, J. H. Metzen, E. Vander Poorten. Surgical robotics beyond enhanced dexterity instrumentation: a survey of machine learning techniques and their role in intelligent and autonomous surgical actions. *International Journal of Computer Assisted Radiology and Surgery*, 2015.

[11] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

[12] A. R. Lanfranco, A. E. Castellanos, J. P. Desai, W. C. Meyers. Robotic Surgery. *Annals of Surgery*, 239(1):14–21, 2004.

[13] B. W. King, L. A. Reisner, A. K. Pandya, A. M. Composto, R. D. Ellis, M. D. Klein. Towards an autonomous robot for camera control during laparoscopic surgery. *Journal of Laparoendoscopic & Advanced Surgical Techniques. Part A*, 23(12):1027–1030, 2013.

[14] R. D. Ellis, A. J. Munaco, L. A. Reisner, M. D. Klein, A. M. Composto, A. K. Pandya, B. W. King. Task analysis of laparoscopic camera control schemes. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 2015.

[15] A. K. Pandya, L. A. Reisner, B. W. King, N. Lucas, A. M. Composto, M. D. Klein, R. D. Ellis. A Review of Camera Viewpoint Automation in Robotic and Laparoscopic Surgery. *Robotics*, 3:310–329, 2014.

[16] H. C. Lin, I. Shafran, D. Yuh, G. D. Hager. Towards automatic skill evaluation: Detection and segmentation of robot-assisted surgical motions. *Computer Aided Surgery*, 11(5):220–230, 2006.

[17] L. Bouarfa, P. P. Jonker, J. Dankelman. Discovery of high-level tasks in the operating room. *Journal of Biomedical Informatics*, 44(3):455–62, 2011.

[18] C. E. Reiley, H. C. Lin, D. D. Yuh, G. D. Hager. Review of methods for objective surgical skill evaluation. *Surgical Endoscopy*, 25(2):356–366, 2011.

[19] G. M. Fried, L. S. Feldman, M. C. Vassiliou, S. A. Fraser, D. Stanbridge, G. Ghitulescu, C. G. Andrew. Proving the Value of Simulation in Laparoscopic Surgery. *Annals of Surgery*, 240(3):518–528, 2004.

[20] J. Rosen, B. Hannaford, C. G. Richards, M. N. Sinanan. Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills. *Biomedical Engineering, IEEE Transactions on*, 48(5):579–591, 2001.

[21] T. Blum, N. Padoy, H. Feussner, N. Navab. Modeling and online recognition of surgical phases using Hidden Markov Models. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*, pp. 627–35. 2008.

[22] F. Lalys, L. Riffaud, D. Bouget, P. Jannin. A framework for the recognition of high-level surgical tasks from video images for cataract surgeries. *Biomedical Engineering, IEEE Transactions on*, 59(4):966–76, 2012.

[23] D. Sanchez, M. Tentori, J. Favela. Activity recognition for the smart hospital. *Intelligent Systems, IEEE*, 23(2):50–57, 2008.

[24] J. E. Bardram, A. Doryab, R. M. Jensen, P. M. Lange, K. L. G. Nielsen, S. T. Petersen. Phase recognition during surgical procedures using embedded and body-worn sensors. *2011 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 45–53, 2011.

[25] J. Rosen, M. Solazzo, B. Hannaford, M. Sinanan. Task decomposition of laparoscopic surgery for objective evaluation of surgical residents' learning curve using hidden markov model. *Computer Aided Surgery*, 7(1):49–61, 2002.

[26] R. Aggarwal, A. Darzi. Laparoscopic task recognition using hidden markov models. *Medicine Meets Virtual Reality 13: The Magical Next Becomes the Medical Now*, 111:115, 2005.

[27] C. E. Reiley, H. C. Lin, B. Varadarajan, B. Vagvolgyi, S. Khudanpur, D. D. Yuh, G. D. Hager. Automatic recognition of surgical motions using statistical modeling for capturing variability. *Studies in Health Technology and Informatics*, 132(1):396–401, 2008.

[28] C. E. Reiley, G. D. Hager. Task versus subtask surgical skill evaluation of robotic minimally invasive surgery. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, pp. 435–442. Springer, 2009.

[29] H. C. Lin, I. Shafran, T. E. Murphy, A. M. Okamura, D. D. Yuh, G. D. Hager. Automatic detection and segmentation of robot-assisted surgical motions. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2005*, pp. 802–810. Springer, 2005.

[30] B. Varadarajan, C. Reiley, H. Lin, S. Khudanpur, G. Hager. Data-derived models for segmentation with application to surgical assessment and training. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, pp. 426–434. Springer, 2009.

[31] J. J. Leong, M. Nicolaou, L. Atallah, G. P. Mylonas, A. W. Darzi, G.-Z. Yang. Hmm assessment of quality of movement trajectory in laparoscopic surgery. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pp. 752–759. Springer, 2006.

[32] L. Tao, E. Elhamifar, S. Khudanpur, G. D. Hager, R. Vidal. Sparse hidden markov models for surgical gesture classification and skill evaluation. In *Information Processing in Computer-Assisted Interventions*, pp. 167–177. Springer, 2012.

[33] N. Padoy, T. Blum, S. A. Ahmadi, H. Feussner, M. O. Berger, N. Navab. Statistical modeling and recognition of surgical workflow. *Medical Image Analysis*, 16(3):632–641, 2012.

[34] S. K. Jun, M. S. Narayanan, P. Agarwal, A. Eddib, P. Singhal, S. Garimella, V. Krovi. Robotic Minimally Invasive Surgical skill assessment based on automated video-analysis motion studies. *IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, pp. 25–31, 2012.

[35] B. B. Haro, L. Zappella, R. Vidal. Surgical gesture classification from video data.

In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pp. 34–41. Springer, 2012.

[36] L. Zappella, B. Béjar, G. Hager, R. Vidal. Surgical gesture classification from video and kinematic data. *Medical Image Analysis*, 17(7):732–45, 2013.

[37] C. Lee, Y. F. Wang, D. R. Uecker, Y. UWang. Image analysis for automated tracking in robot-assisted endoscopic surgery. *Proceedings of 12th International Conference on Pattern Recognition*, 1:88–92, 1994.

[38] D. R. Uecker, C. Lee, Y. F. Wang, Y. Wang. Automated instrument tracking in robotically assisted laparoscopic surgery. *Journal of Image Guided Surgery*, 1(6):308–325, 1995.

[39] G. Q. Wei, G. Hirzinger. Real-Time Visual Servoing for Laparoscopic Surgery. *Engineering in Medicine and Biology Magazine, IEEE*, 16(1):40–45, 1997.

[40] K. Omote, H. Feussner, A. Ungeheuer, K. Arbter, G. q. Wei, J. Ru. Self-Guided Robotic Camera Control for Laparoscopic Surgery Compared with Human Camera Control. *The American Journal of Surgery*, 177(4):321–324, 1999.

[41] M. Kranzfelder, A. Schneider, A. Fiolka, E. Schwan, S. Gillen, D. Wilhelm, R. Schirren, S. Reiser, B. Jensen, H. Feussner. Real-time instrument detection in minimally invasive surgery using radiofrequency identification technology. *The Journal of Surgical Research*, 185(2):704–10, 2013.

[42] O. Weede, A. Bihlmaier, J. Hutzl, B. P. Müller Stich, H. Wörn. Towards Cognitive Medical Robotics in Minimal Invasive Surgery. *Proceedings of Conference on Advances In Robotics - AIR '13*, pp. 1–8, 2013.

[43] D. Bernad. Finding patterns in time series: a dynamic programming approach. *Advances in Knowledge Discovery and Data Mining*, pp. 229–248, 1996.

[44] T. Cover, P. Hart. Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1):21–27, 1967.

[45] L. R. Rabiner, B. H. Juang. An introduction to hidden markov models. *ASSP Magazine, IEEE*, 3(1):4–16, 1986.

[46] J. Yang, Y. Xu. Hidden markov model for gesture recognition. *Technical report*, DTIC Document, 1994.

[47] M. Gales, S. Young. The application of hidden markov models in speech recognition. *Foundations and Trends in Signal Processing*, 1(3):195–304, 2008.

[48] E. Keogh, S. Kasetty. On the need for time series data mining benchmarks: a survey and empirical demonstration. *Data Mining and Knowledge Discovery*, 7(4):349–371, 2003.

[49] J. Lin, S. Williamson, K. Borne, D. DeBarr. Pattern recognition in time series. *Advances in Machine Learning and Data Mining for Astronomy*, 1:617–645, 2012.

[50] N. Bhatia, et al. Survey of nearest neighbor techniques. *International Journal of Computer Science and Information Security*, 8(2):302–305, 2010.

[51] MATLAB. *version 8.5.01 (R2014a)*. The MathWorks Inc., Natick, Massachusetts, 2014.

[52] C. R. Wren, A. Azarbayejani, T. Darrell, A. P. Pentland. Pfinder: Real-time tracking of the human body. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):780–785, 1997.

[53] C. Bregler. Learning and recognizing human dynamics in video sequences. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pp. 568–574. IEEE, 1997.

[54] H.-K. Lee, J. H. Kim. An hmm-based threshold model approach for gesture recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(10):961–973, 1999.

[55] G. K. Cheung, S. Baker, T. Kanade. Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture. In *Computer Vision*

and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, volume 1, pp. I–77. IEEE, 2003.

[56] D. Weinland, R. Ronfard, E. Boyer. A survey of vision-based methods for action representation, segmentation and recognition. *Computer Vision and Image Understanding*, 115(2):224–241, 2011.

[57] S. Mitra, T. Acharya. Gesture recognition: A survey. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(3):311–324, 2007.

[58] D. Del Vecchio, R. M. Murray, P. Perona. Decomposition of human motion into dynamics-based primitives with application to drawing tasks. *Automatica*, 39(12):2085–2098, 2003.

[59] J. Reason. *Human Error*. Cambridge university press, 1990.

[60] S. Calinon, F. D'halluin, E. L. Sauser, D. G. Caldwell, A. G. Billard. Learning and reproduction of gestures by imitation. *Robotics & Automation Magazine, IEEE*, 17(2):44–54, 2010.

[61] L. Tao, L. Zappella, G. D. Hager, R. Vidal. Surgical gesture segmentation and recognition. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, pp. 339–346. Springer, 2013.

[62] S. Niekum, S. Osentoski, G. Konidaris, S. Chitta, B. Marthi, A. G. Barto. Learning grounded finite-state representations from unstructured demonstrations. *The International Journal of Robotics Research*, 34(2):131–157, 2015.

[63] S. Krishnan, A. Garg, S. Patil1, C. Le, G. Hager, P. Abbeel, K. Goldberg. Transition state clustering: Unsupervised surgical trajectory segmentation for robot learning. In *International Symposium of Robotics Research*. Springer STAR, 2015.

[64] T. S. Wang, H. Y. Shum, Y. Q. Xu, N. N. Zheng. Unsupervised analysis of human gestures. In *Advances in Multimedia Information ProcessingPCM 2001*, pp. 174–181. Springer, 2001.

[65] J. Alon, V. Athitsos, Q. Yuan, S. Sclaroff. A unified framework for gesture recognition and spatiotemporal gesture segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(9):1685–1699, 2009.

[66] F. Despinoy, D. Bouget, G. Forestier, C. Penet, N. Zemiti, P. Poignet, P. Jannin. Unsupervised trajectory segmentation for surgical gesture recognition in robotic training. *Biomedical Engineering, IEEE Transactions on*, 99(P):1–1, 2015.

[67] A. Murali, A. Garg, S. Krishnan, F. T. Pokorny, P. Abbeel, T. Darrell, K. Goldberg. TSC-DL: Unsupervised trajectory segmentation of multi-modal surgical demonstrations with deep learning. In *Robotics and Automation, IEEE International Conference on (ICRA), Stockholm, Sweden*. 2016.

[68] B. C. Bedregal, A. Costa, G. P. Dimuro. Fuzzy rule-based hand gesture recognition. In *Artificial Intelligence in Theory and Practice*, pp. 285–294. Springer, 2006.

[69] B. C. Bedregal, G. Dimuro, A. Costa. Hand gesture recognition in an interval fuzzy approach. *Trends in Applied and Computational Mathematics*, 8(1):21–31, 2007.

[70] J. Abonyi, B. Feil, S. Nemeth, P. Arva. Modified gath–geva clustering for fuzzy segmentation of multivariate time-series. *Fuzzy Sets and Systems*, 149(1):39–56, 2005.

[71] E. Keogh, S. Chu, D. Hart, M. Pazzani. An online algorithm for segmenting time series. In *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pp. 289–296. IEEE, 2001.

[72] I. Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.

[73] M. E. Tipping, C. M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622, 1999.

[74] J. C. Bezdek, R. Ehrlich, W. Full. FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2):191–203, 1984.

[75] J. V. de Oliveira, W. Pedrycz, editors. *Advances in fuzzy clustering and its applications*. Wiley Online Library, 2007.

[76] J. Abonyi, R. Babuška, F. Szeifert. Modified gath-geva fuzzy clustering for identification of takagi-sugeno fuzzy models. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 32(5):612–621, 2002.

[77] W. Krzanowski. Between-groups comparison of principal components. *Journal of the American Statistical Association*, 74(367):703–707, 1979.

[78] P. M. Kelly. An algorithm for merging hyperellipsoidal clusters. *Los Alamos National Laboratory, Tech. Rep*, 1994.

[79] S. Mitra, S. K. Pal, P. Mitra. Data mining in soft computing framework: a survey. *Neural Networks, IEEE Transactions on*, 13(1):3–14, 2002.

[80] M. Rawashdeh, A. Ralescu. Fuzzy cluster validity with generalized silhouettes. In *Proceedings of the 23rd Annual Midwest Artificial Intelligence and Cognitive Science Conference, Cincinnati, Ohio, April*, pp. 11–18. Citeseer, 2012.

[81] T. P. Grantcharov, L. Bardram, P. Funch-Jensen, J. Rosenberg. Assessment of technical surgical skills. *The European Journal of Surgery*, 168(3):139–144, 2002.

[82] B. Schout, A. Hendrikx, F. Scheele, B. Bemelmans, A. Scherpbier. Validation and implementation of surgical simulators: a critical review of present, past, and future. *Surgical Endoscopy*, 24(3):536–546, 2010.

[83] J. Martin, G. Regehr, R. Reznick, H. MacRae, J. Murnaghan, C. Hutchison, M. Brown. Objective structured assessment of technical skill (OSATS) for surgical residents. *British Journal of Surgery*, 84(2):273–278, 1997.

[84] S. Dreiseitl, M. Binder. Do physicians value decision support? A look at the effect of decision support systems on physician opinion. *Artificial Intelligence in Medicine*, 33(1):25–30, 2005.

[85] A. A. Gumbs, N. J. Hogle, D. L. Fowler. Evaluation of Resident Laparoscopic Performance Using Global Operative Assessment of Laparoscopic Skills. *Journal of the American College of Surgeons*, 204(2):308–313, 2007.

[86] J. D. Doyle, E. M. Webber, R. S. Sidhu. A universal global rating scale for the evaluation of technical skills in the operating room. *The American journal of surgery*, 193(5):551–555, 2007.

[87] T. N. Judkins, D. Oleynikov, N. Stergiou. Objective evaluation of expert and novice performance during robotic surgical training tasks. *Surgical Endoscopy*, 23(3):590–597, 2009.

[88] J. Marescaux, J. Leroy, M. Gagner, F. Rubino, D. Mutter, M. Vix, S. E. Butner, M. K. Smith. *Nature*, 413(6854):379–80, 2001.

[89] L. MacKenzie, J. Ibbotson, C. Cao, A. Lomax. Hierarchical decomposition of laparoscopic surgery: a human factors approach to investigating the operating room environment. *Minimally Invasive Therapy & Allied Technologies*, 10(3):121–127, 2001.

[90] N. Ahmidi, G. D. Hager, L. Ishii, G. Fichtinger, G. L. Gallia, M. Ishii. Surgical task and skill classification from eye tracking and tool motion in minimally invasive surgery. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*, volume 13, pp. 295–302. 2010.

[91] N. Ahmidi, G. D. Hager, L. Ishii, G. L. Gallia, M. Ishii. Surgical task and skill classification from eye tracking and tool motion in minimally invasive surgery. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, volume 15, pp. 471–478. 2012.

[92] N. Ahmidi, P. Poddar, J. D. Jones, S. S. Vedula, L. Ishii, G. D. Hager, M. Ishii. Automated objective surgical skill assessment in the operating room from unstructured tool motion in septoplasty. *International Journal of Computer Assisted Radiology and Surgery*, 10(6):981–991, 2015.

[93] M. K. Chmarra, S. Klein, J. C. F. De Winter, F. W. Jansen, J. Dankelman. Objective classification of residents based on their psychomotor laparoscopic skills. *Surgical Endoscopy and Other Interventional Techniques*, 24(5):1031–1039, 2010.

[94] V. Datta, S. Mackay, M. Mandalia, A. Darzi. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. *Journal of the American College of Surgeons*, 193(5):479–485, 2001.

[95] Y. Yamauchi, J. Yamashita, O. Morikawa, R. Hashimoto, M. Mochimaru, Y. Fukui, H. Uno, K. Yokoyama. Surgical skill evaluation by force data for endoscopic sinus surgery training system. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2002*, pp. 44–51. Springer, 2002.

[96] S. Cotin, N. Stylopoulos, M. P. Ottensmeyer, P. F. Neumann, D. W. Rattner, S. Dawson. Metrics for Laparoscopic Skills Trainers: The Weakest Link! In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2002*, volume 2488, pp. 35–43. 2002.

[97] T. N. Judkins, D. Oleynikov, N. Stergiou. Objective evaluation of expert and novice performance during robotic surgical training tasks. *Surgical Endoscopy*, 23(3):590–597, 2009.

[98] P. Rhienmora, P. Haddawy, S. Suebnukarn, M. N. Dailey. Intelligent dental training simulator with objective skill assessment and feedback. *Artificial Intelligence in Medicine*, 52(2):115–121, 2011.

[99] E. Abbena, S. Salamon, A. Gray. *Modern Differential Geometry of Curves and Surfaces with Mathematica*. CRC press, 2006.

[100] V. O. Nams. Using animal movement paths to measure response to spatial scale. *Oecologia*, 143(2):179–188, 2005.

[101] R. Voyles, J. Bae, R. Godzdanker. The gestural joystick and the efficacy of the path tortuosity metric for human/robot interaction. *Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*, pp. 91–97, 2008.

[102] W. D. Kearns, V. O. Nams, J. L. Fozard. Tortuosity in Movement Paths Is Related to Cognitive Impairment. *Methods of Information in Medicine*, 49(6):592–598, 2010.

[103] B. Mandelbrot. How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension. *Science*, 156(3775):636–638, 1967.

[104] V. O. Nams. Improving Accuracy and Precision in Estimating Fractal Dimension of Animal movement paths. *Acta Biotheoretica*, 54(1):1–11, 2006.

[105] H. Abdi, L. J. Williams. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):433–459, 2010.

[106] D. G. Kleinbaum, M. Klein. *Logistic Regression.* Statistics for Biology and Health. Springer New York, New York, NY, 2010.

[107] L. Saitta. Suppor Vector Networks. *Machine Learning*, 20:273–297, 1995.

[108] V. Vapnik. *Statistical Learning Theory.* 1998.

[109] C.-C. Chang, C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

ABSTRACT

COMPUTATIONAL MODELING APPROACHES FOR TASK ANALYSIS IN ROBOTIC-ASSISTED SURGERY

by

MAHTAB JAHANBANI FARD

May 2016

Advisor: Dr. R. Darin Ellis

Major: Industrial Engineering

Degree: Doctor of Philosophy

Surgery is continuously subject to technological innovations including the introduction of robotic surgical devices. The ultimate goal is to program the surgical robot to perform certain difficult or complex surgical tasks in an autonomous manner. The feasibility of current robotic surgery systems to record quantitative motion and video data motivates developing descriptive mathematical models to recognize, classify and analyze surgical tasks. Recent advances in machine learning research for uncovering concealed patterns in huge data sets, like kinematic and video data, offer a possibility to better understand surgical procedures from a system point of view. This dissertation focuses on bridging the gap between these two lines of the research by developing computational models for task analysis in robotic-assisted surgery.

The key step for advance study in robotic-assisted surgery and autonomous skill assessment is to develop techniques that are capable of recognizing fundamental surgical tasks intelligently. Surgical tasks and at a more granular level, surgical gestures, need to be quantified to make them amenable for further study. To answer to this query, we introduce a new framework, namely DTW-kNN, to recognize and classify three important surgical tasks including suturing, needle passing and knot tying based on kinematic data captured using *da Vinci* robotic surgery system. Our proposed method needs minimum preprocessing that

results in simple, straightforward and accurate framework which can be applied for any autonomous control system. We also propose an unsupervised gesture segmentation and recognition (UGSR) method which has the ability to automatically segment and recognize temporal sequence of gestures in RMIS task. We also extent our model by applying soft boundary segmentation (Soft-UGSR) to address some of the challenges that exist in the surgical motion segmentation. The proposed algorithm can effectively model gradual transitions between surgical activities.

Additionally, surgical training is undergoing a paradigm shift with more emphasis on the development of technical skills earlier in training. Thus metrics for the skills, especially objective metrics, become crucial. One field of surgery where such techniques can be developed is robotic surgery, as here all movements are already digitalized and therefore easily susceptible to analysis. Robotic surgery requires surgeons to perform a much longer and difficult training process which create numerous new challenges for surgical training. Hence, a new method of surgical skill assessment is required to ensure that surgeons have adequate skill level to be allowed to operate freely on patients. Among many possible approaches, those that provide noninvasive monitoring of expert surgeon and have the ability to automatically evaluate surgeon's skill are of increased interest. Therefore, in this dissertation we develop a predictive framework for surgical skill assessment to automatically evaluate performance of surgeon in RMIS. Our classification framework is based on the Global Movement Features (GMFs) which extracted from kinematic movement data. The proposed method addresses some of the limitations in previous work and gives more insight about underlying patterns of surgical skill levels.

AUTOBIOGRAPHICAL STATEMENT

Mahtab Jahanbani Fard received her B.Sc. degree in Physics and M.Sc. in Industrial Engineering from Isfahan University of Technology. She received her second M.Sc. degree in Computer Science from Wayne State University. She is currently a Ph.D. candidate in Industrial & System Engineering at Wayne State University. Her research interests include human factor, machine learning, data mining, artificial intelligence, robotic surgery, survival analysis and healthcare. During her study at Wayne State University, she made a number of technical presentations at INFORMS, AHEF and IEOM. She has received the best track paper and best graduate paper award for the 2015 IEOM. Her papers have been published and/or are under review in venues such as Artificial Intelligent in Medicine (AIM), IEEE Robotics and Automation Letters (RA-L), Medical Robotics and Computer Assisted Surgery (Int J MRCAS), IEEE Transactions on Knowledge and Data Engineering (TKDE), Knowledge Discovery and Data Mining (KDD). She is a student member of HFES, IEEE, INFORMS, and IIE.