

Discovering Discriminative and Interpretable Patterns for Surgical Motion Analysis

Germain Forestier^{1,2}, François Petitjean², Pavel Senin³,
Fabien Despinoy⁴, and Pierre Jannin⁴

¹ MIPS, University of Haute-Alsace, Mulhouse, France

² Faculty of Information Technology, Monash University, Melbourne, Australia

³ Los Alamos National Laboratory, Los Alamos, NM, 87545, USA

⁴ INSERM MediCIS, Unit U1099 LTSI, University of Rennes 1, Rennes, France

Abstract. The analysis of surgical motion has received a growing interest with the development of devices allowing their automatic capture. In this context, the use of advanced surgical training systems make an automated assessment of surgical trainee possible. Automatic and quantitative evaluation of surgical skills is a very important step in improving surgical patient care. In this paper, we present a novel approach for the discovery and ranking of discriminative and interpretable patterns of surgical practice from recordings of surgical motions. A pattern is defined as a series of actions or events in the kinematic data that together are distinctive of a specific gesture or skill level. Our approach is based on the discretization of the continuous kinematic data into strings which are then processed to form bags of words. This step allows us to apply discriminative pattern mining technique based on the word occurrence frequency. We show that the patterns identified by the proposed technique can be used to accurately classify individual gestures and skill levels. We also present how the patterns provide a detailed feedback on the trainee skill assessment. Experimental evaluation performed on the publicly available JIGSAWS dataset shows that the proposed approach successfully classifies gestures and skill levels.

Keywords: surgical motion analysis, skill assessment, pattern mining

1 Introduction

In recent years, analysis of surgical motion has received a growing interest following the development of devices enabling automated capture of surgeon motions such as tracking, robotic and training systems. Surgical training programs now often include surgical simulators which are equipped with sensors for automatic surgical motions recording [1,2]. The ability to collect surgical motion data brings unprecedented opportunities for automated objective analysis and assessment of surgical trainees progression. The main goal of this effort is to support surgeons in technical skills acquisition, as these are shown to correlate with a reduction of patient complications [3]. Hence, automated evaluation of surgical skill level is an important step in surgical patient care improvement.

This article tackles the issue of identifying discriminative and interpretable patterns of surgical practice from recordings of surgical motions. We define a *pattern* a series of actions or events in the kinematic data that together are distinctive of a specific gesture or a skill level. We show, that by using these patterns, we can reach beyond the simple classification of observed surgeons into categories (*e.g.*, expert, novice) by providing a quantitative evidence-supported feedback to the trainee as per where he or she can improve. The proposed approach, based on SAX-VSM algorithm [4], considers surgical motion as continuous multi-dimensional time-series and starts by discretizing them into sequence of letters (*i.e.*, strings) using Symbolic Aggregate approxXimation (SAX) [5]. In turn, SAX sequences are decomposed into subsequences of few consecutive letters via sliding window. The relative frequencies of these subsequences, *i.e.*, the number of times they appear in a given sequence or in a set of sequences, are then used to identify discriminative patterns that characterize specific surgical motion. To discover the patterns, we rely on the Vector Space Model (VSM) [6] which has been originally proposed as an algebraic model for representing collection of text documents. The identified discriminative patterns are then used to perform classification by identifying them in to-be-classified recordings. Furthermore, by highlighting discriminative patterns in the visualization of original motion data, we are able to provide an intuitive visual explanation about *why* a specific skill assessment is provided. We evaluated our method on the kinematic data from the JHU-ISI Gesture and Skill Assessment Dataset (JIGSAWS) [7] that is currently the largest publicly accessible database for surgical gesture analysis. Our experiments have shown that the proposed method accurately classifies gestures and skill levels. The main contributions of this paper are:

- A framework for identifying discriminative and interpretable patterns in surgical activity motion based on SAX [5] and VSM [4].
- Experimental evaluation highlighting the relevance of the proposed method for gestures classification and skill assessment.
- A visualization technique enabling self-assessment of trainee skills.

2 Background

Surgical motion analysis is mainly based on kinematic data recorded by surgical robot [8,9] and video data [10,11,12]. Kinematic data usually include multiple attributes such as the position of robot’s tools, rotations, and velocities. From such data, significant amount of work has been devoted to the segmentation of surgical tasks into more detailed gestures [13,14,15]. Segmenting surgical motion into gestures makes it possible to obtain a finer description of surgical task leading to more detailed feedback on skill assessment [16,17]. Previous work concerned with gesture segmentation using kinematic and video data uses Hidden Markov Models [18,19], Conditional Random Fields [20] and Linear Dynamical Systems [11]. Main drawback of these approaches is the difficulty for the trainee to understand the output and to use it as a feedback to improve performance. In contrast, our approach seeks not only to identify that a surgical motion has been

performed by a novice surgeon, but also to explain *why* it has been classified as such. This step is critical in justifying the reasons why the trainee is still considered as a novice and to help him or her to focus on the specific steps that require improvement.

3 Method

3.1 Symbolic Aggregate approXimation (SAX)

We propose to use Symbolic Aggregate approXimation (SAX) [5] to discretize the input time series [21]. For time series T of length n , SAX obtains a lower-dimensional representation by first performing a z -normalization then dividing the time series into s equal-sized segments. Next, for each segment, SAX computes a mean value and maps it to a symbol according to a pre-defined set of breakpoints dividing the data space into α equiprobable regions, where α is the user specified alphabet size. While dimensionality reduction is a desirable feature for exploring global patterns, the high compression ratio (n/s) significantly affects performance in cases where localized phenomena are of interest. Thus, for the local pattern discovery, SAX is typically applied to a set of subsequences that represent local features – a technique called subsequence discretization [22] which is implemented via a sliding window. Note that other time-series discretization approaches could have been used at this step [23].

3.2 Bag of words representation of kinematic data

Following the approach proposed in [4], a sliding window technique is used to convert a time series T of length n into the set of m SAX words, where $m = (n - l_s) + 1$ and l_s the sliding window length. A sliding window of length l_s is applied across the time series T and the overlapping extracted subsequences are converted into SAX words and then put in a collection. This collection is a *bag of words* representation of the original time series T .

In the case of kinematic data, this process is performed independently for each dimension of the data (*e.g.*, x coordinate, y coordinate, etc.). All features are normalized on a per-trial per-feature basis. Each word extracted in each dimension of the data is postfixed with the name of the dimension (*e.g.* x , y , etc.). We assume that depending of the gesture or the skill level to classify, different kinematic features can be relevant. Note, that this methodology can be used regardless of the available kinematic data (*e.g.* number of features, etc.). Figure 1 illustrates the conversion of kinematic data for one trial into a bag of words using SAX.

3.3 Vector Space Model (VSM)

We rely on the original definition of vector space model as it is known in Information Retrieval (IR) [6,4]. The *tf*idf* weight for a term t is defined as a product

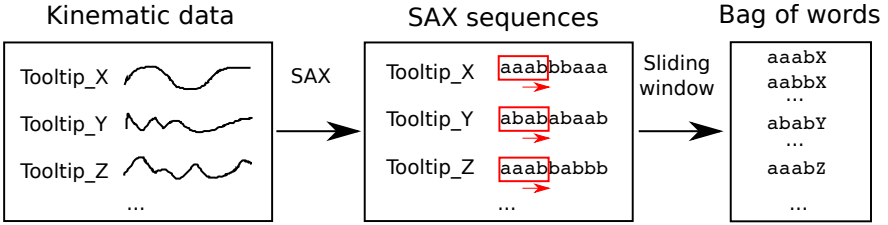


Fig. 1: Conversion of kinematic data for one trial into a bag of words using SAX [5] and a sliding window of size 4 (in red).

of two factors: term frequency (tf) and inverse document frequency (idf). The first factor corresponds to logarithmically scaled term frequency [24].

$$tf_{t,d} = \begin{cases} \log(1 + f_{t,d}), & \text{if } f_{t,d} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where t is the term, d is a bag of words (a document in IR terms), and $f_{t,d}$ is the frequency of t in d . The inverse document frequency [24] is defined as

$$idf_{t,D} = \log \frac{|D|}{|d \in D : t \in d|} = \log \frac{N}{df_t} \quad (2)$$

where N is the cardinality of a corpus D (the total number of classes) and the denominator df_t is the number of bags where the term t appears. Then, $tf*idf$ weight value for a term t in the bag d of a corpus D is defined as

$$tf*idf(t, d, D) = tf_{t,d} \times idf_{t,D} = \log(1 + f_{t,d}) \times \log \frac{N}{df_t} \quad (3)$$

for all cases where $f_{t,d} > 0$ and $df_t > 0$, or zero otherwise.

Once all frequencies are computed, the term frequency matrix becomes the term weight matrix, whose columns are used as *class term weight* vectors to perform classification using Cosine similarity. For two vectors \mathbf{a} and \mathbf{b} , the Cosine similarity is based on their inner product and defined as

$$\text{similarity}(\mathbf{a}, \mathbf{b}) = \cos(\theta) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \cdot \|\mathbf{b}\|} \quad (4)$$

3.4 Training and classifying kinematic data

The training step starts by transforming the kinematic data into SAX representation using two parameters: the size of the sliding window l_s , and the size of the alphabet α . Then, the algorithm builds a corpus of N bags corresponding to the subsequences extracted from the N classes of kinematic data, i.e. same skill level or same gesture depending on the application. The $tf*idf$ weighting is then applied to create N real-valued weight vectors of equal length, representing the different class of kinematic data.



Fig. 2: Snapshots of the three surgical tasks in the JIGSAWS dataset (from left to right): suturing, knot-tying, needle-passing [7].

In order to classify an unlabeled kinematic data, the method transforms it into a terms frequency vector using exactly the same sliding window and SAX parameters used for the training part. It computes the cosine similarity measure (Eq. 4) between this term frequency vector and the N *tf-idf* weight vectors representing the training classes. The unlabeled kinematic data is assigned to the class whose vector yields the maximal cosine similarity value.

4 Experimental evaluation

The JIGSAWS dataset [7] includes 8 subjects with 3 different skill levels (novice, intermediate and expert) performing 3–5 trials of three tasks (suturing, knot tying, and needle passing). The Figure 2 illustrates the three tasks. Each trial lasts about 2 minutes and is represented by the kinematic data of both master and slave manipulators of the da Vinci robotic surgical system recorded at a constant rate of 30 Hz. Kinematic data consists of 76 motion variables including positions and velocities of both master and slave manipulators. All trials in the JIGSAWS dataset were manually segmented into 15 surgical gestures. Video of the trials are also available and are synchronized with the kinematic data. A detailed description of the dataset is available in [25].

Our training step first transforms the kinematic data time series into SAX representation configured by two parameters: the sliding window length (l_s) and SAX alphabet size (α). The number of segments per window was kept equal to the length of the window which means that every point of the time series was transformed into a letter. This choice was made to allow us to map back the patterns on the original time series. Parameters l_s and α were optimized using cross-validation on the training data. As they can differ for each specific classification problem, their values are provided along with the experimental results.

4.1 Gesture classification

We considered the gesture boundaries to be known and we used the kinematic data alone. We present results for two cross-validation configurations provided

Table 1: Gesture classification performance, assuming known boundaries and using kinematic data only.

Method	Metric	Leave-one-supertrial-out			Leave-one-user-out		
		Suturing	Needle Passing	Knot Tying	Suturing	Needle Passing	Knot Tying
	(l_s, α)	(8,19)	(13,18)	(15,7)	(8,19)	(14,18)	(10,12)
<i>Proposed</i>	Micro	93.69	81.08	92.45	88.27	75.29	89.76
	Macro	79.95	74.67	89.78	68.77	67.54	82.29
LDS [25]	Micro	84.61	59.76	81.67	73.64	47.96	71.42
LDS [25]	Macro	63.87	46.55	74.51	51.75	32.59	63.99
HMM [25]	Micro	92.56	75.68	89.76	80.83	66.22	78.44
HMM [25]	Macro	79.66	72.36	87.29	65.03	62.70	72.68

with the JIGSAWS data [7]. In the first configuration – leave one supertrial out (LOSO) – for each iteration of cross-validation (five in total), one trial of each subject was left out for the test and the remaining trials were used for training. In the second configuration – leave one user out (LOUO) – for each iteration of the cross-validation (eight in total), all the trials belonging to a particular subject were left out for the test. These are the standard benchmark configurations provided in [7]. We report micro (average of total correct predictions across all classes) and macro (average of true positive rates for each class) performance results as defined in [25].

Table 1 presents the results for gesture classification assuming known boundaries and using kinematic data only. For comparison purposes, we also report state-of-the-art results for Linear Dynamical Systems (LDS) and Hidden Markov Models (HMM) from [25]. The proposed method outperforms both LDS and HMM methods in terms of micro and macro performances for the three tasks and the two cross-validation configurations. These results show that our method accurately identifies patterns that are specific to a gesture motion. One of the interesting features of the proposed method is the ability to use different kinematic data depending of the gesture. As our method computes the frequencies for each component of the kinematic data for each gesture independently, the most discriminative attributes of a given gesture naturally stand out. Furthermore, the *tf*idf* regularization discards the motion patterns that are common to every gesture (*i.e.*, irrelevant for classification as not distinctive of any class).

The LOUO configuration is known to be particularly challenging, because we attempt to classify gestures of a subject without having any of his or her other attempts. The good performance of our approach can be explained by its ability to identify highly discriminative patterns that are the most distinctive of each gesture. These results also indicate that our method generalizes well, as shown by the fact that it can accurately classify gestures from unobserved trainees.

Table 2: Skill classification performance per trial using kinematic data only.

Method	Metric	Leave-one-supertrial-out		
		Suturing	Needle Passing	Knot Tying
	(l_s, α)	(10,9)	(12,13)	(5,14)
<i>Proposed</i>	Micro	89.74	96.30	61.11
	Macro	86.67	95.83	53.33
SHMM [19]	Micro	97.40	96.20	94.40

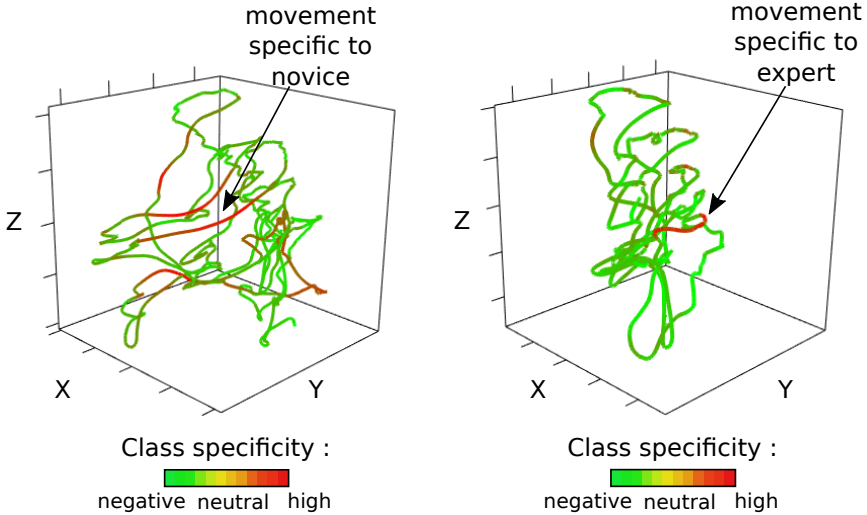
4.2 Skills classification

For skill classification, we performed experiments to identify the skill level (novice, intermediate or expert) at the trial level. In this experiment, we used the leave one trial out (LOSO) cross-validation configuration. Table 2 presents the results for the three tasks and reports micro and macro performances. The results are better for Suturing and Needle Passing tasks than for Knot Tying task. The poor performance on the Knot Tying task can be explained by the minor difference between the Expert and Intermediate subjects for this task (mean GRS is 17.7 and 17.1 for expert and intermediate respectively). We also report the state-of-the-art results from [19] for the Suturing task. The SHMM approach gives better results for the per trial classification configuration as it uses global temporal information, whereas our method is focusing on the local patterns regardless of their location within larger time series. Furthermore, the SHMM approach [19] uses gestures boundaries to learn the temporal model while our method is not using this information.

4.3 Interpretable patterns visualization

Our approach outputs a set of discriminative patterns weighted by the class specificity for each of the input class. These lists of ranked patterns can be studied to better understand what makes each class distinctive. As the use of $tf*idf$ (Eq. (3)) discards patterns that are common to all classes, only patterns having discriminative power remain.

The list of weighted discriminative patterns can be used to visualize, on a given trial, where are the areas that are specific to the current skill level of the trial. We propose to use a heat map-like visualization technique that provides immediate insight into the layout of the “important” class-characteristic patterns (as described in [4]). Figure 3 shows, for the Suturing task, the two individual 5th trials of subjects B (Novice) and E (Expert), using (x, y, z) coordinates for the right hand. In this figure, we used respectively the $tf*idf$ weights vectors of the 5th fold for the Novice on subject B and for the Expert on subject E. The red areas correspond to specific motions that are correlated with a skill level.



(a) Trial 5 of Suturing task of subject B (novice) using Novice class $tf*idf$ vector weights of 5th fold (best viewed in color). (b) Trial 5 of Suturing task of subject E (expert) using Expert class $tf*idf$ vector weights of 5th fold (best view in color).

Fig. 3: Example of interpretable feedback using a heat-map visualization of subsequence importance to a class identification. The value corresponds to the combination of the $tf*idf$ weights of all patterns which cover the point.

For Subject B (Figure 3a), these areas correspond to motions that were only observed among the novices. By contrast, green areas correspond to motions that are common to all subjects regardless of their skill. This visualization provides a rich information about what makes a specific skill level distinctive and can also be used to provide individual and personalized feedback. As the videos of the trials are also available, this result has to be displayed side-by-side with the videos in order to show to the trainee the movements that are specific. Note that a more detailed analysis could be performed by observing which kinematic data features are specific in these areas or by performing the analysis on a per gesture basis. Visualization (like Figure 3) for all the subject trials for the Suturing task are available on the companion webpage¹.

Note that as the $tf*idf$ weight vectors are computed prior to the classification step, it is possible to display this heat-map visualization in real-time during the trial. We provide a video in the supplementary material attached to this paper that shows the real-time computation of this visualization while a trainee performs an suturing task. We believe that this tool is an interesting addition

¹ <http://germain-forestier.info/src/aime2017/>

to existing learning tools for surgery as it provides a way to obtain a feedback on which parts of an exercise have been used to classify the attempt.

5 Conclusion

In this paper, we presented a new method for discovery of discriminative and interpretable patterns in surgical activity motion. Our method uses SAX to discretize the kinematic data into sequence of letters. A sliding window is then used to build bag of words. Finally, *tf*idf* framework is applied to identify motion class-characteristic patterns. Experiments performed on the JIGSAWS dataset has shown that our method successfully classifies gestures and skill levels. The strong advantage of the proposed technique is the ability to provide a precise quantitative feedback for the classification results. Of course, the evaluation of our visualization approach needs to be performed within curriculum.

Acknowledgement

This work was supported by the Australian Research Council under award DE170100037. This material is based upon work supported by the Air Force Office of Scientific Research, Asian Office of Aerospace Research and Development (AOARD) under award number FA2386-16-1-4023.

References

1. Tsuda, S., Scott, D., Doyle, J., Jones, D.B.: Surgical skills training and simulation. *Current problems in surgery* **46**(4) (2009) 271–370
2. Forestier, G., Petitjean, F., Riffaud, L., Jannin, P.: Optimal sub-sequence matching for the automatic prediction of surgical tasks. In: *AIME 15th Conference on Artificial Intelligence in Medicine*. Volume 9105., Springer (2015) 123–132
3. Dlouhy, B.J., Rao, R.C.: Surgical skill and complication rates after bariatric surgery. *The New England Journal of Medicine* **370**(3) (2014) 285–285
4. Senin, P., Malinchik, S.: SAX-VSM: Interpretable time series classification using SAX and vector space model. In: *International Conference on Data Mining, IEEE* (2013) 1175–1180
5. Lin, J., Keogh, E., Wei, L., Lonardi, S.: Experiencing SAX: a novel symbolic representation of time series. *Data Mining and Knowledge Discovery* **15**(2) (2007)
6. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. *Communications of the ACM* **18**(11) (1975) 613–620
7. Gao, Y., Vedula, S.S., Reiley, C.E., Ahmidi, N., Varadarajan, B., Lin, H.C., Tao, L., Zappella, L., Béjar, B., Yuh, D.D., et al.: JHU-ISI gesture and skill assessment working set (JIGSAWS): A surgical activity dataset for human motion modeling. *Modeling and Monitoring of Computer Assisted Interventions (M2CAI)–MICCAI Workshop* (2014) 1–10
8. Reiley, C.E., Hager, G.D.: Decomposition of robotic surgical tasks: an analysis of subtasks and their correlation to skill. In: *Modeling and Monitoring of Computer Assisted Interventions (M2CAI) – MICCAI Workshop*. (2009)

9. Reiley, C.E., Plaku, E., Hager, G.D.: Motion generation of robotic surgical tasks: Learning from expert demonstrations. In: International Conference on Engineering in Medicine and Biology Society, IEEE (2010) 967–970
10. Haro, B.B., Zappella, L., Vidal, R.: Surgical gesture classification from video data. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI. Springer (2012) 34–41
11. Zappella, L., Béjar, B., Hager, G., Vidal, R.: Surgical gesture classification from video and kinematic data. *Medical Image Analysis* **17**(7) (2013) 732–745
12. Zia, A., Sharma, Y., Bettadapura, V., Sarin, E.L., Ploetz, T., Clements, M.A., Essa, I.: Automated video-based assessment of surgical skills for training and evaluation in medical schools. *International journal of computer assisted radiology and surgery* **11**(9) (2016) 1623–1636
13. Reiley, C.E., Hager, G.D.: Task versus subtask surgical skill evaluation of robotic minimally invasive surgery. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009. Springer (2009) 435–442
14. Despinoy, F., Bouget, D., Forestier, G., Penet, C., Zemiti, N., Poignet, P., Jannin, P.: Unsupervised trajectory segmentation for surgical gesture recognition in robotic training. *IEEE Transactions on Biomedical Engineering* (2015)
15. Gao, Y., Vedula, S.S., Lee, G.I., Lee, M.R., Khudanpur, S., Hager, G.D.: Unsupervised surgical data alignment with application to automatic activity annotation. In: International Conference on Robotics and Automation, IEEE (2016) 4158–4163
16. Zhou, Y., Ioannou, I., Wijewickrema, S., Bailey, J., Kennedy, G., O’Leary, S.: Automated segmentation of surgical motion for performance analysis and feedback. In: International Conference on Medical Image Computing and Computer-Assisted Intervention–MICCAI, Springer (2015) 379–386
17. Kowalewski, T.M., White, L.W., Lendvay, T.S., Jiang, I.S., Sweet, R., Wright, A., Hannaford, B., Sinanan, M.N.: Beyond task time: automated measurement augments fundamentals of laparoscopic skills methodology. *Journal of Surgical Research* **192**(2) (2014) 329–338
18. Reiley, C.E., Lin, H.C., Varadarajan, B., Vagvolgyi, B., Khudanpur, S., Yuh, D., Hager, G.: Automatic recognition of surgical motions using statistical modeling for capturing variability. *Studies in Health Technology and Informatics* **132** (2008) 396
19. Tao, L., Elhamifar, E., Khudanpur, S., Hager, G.D., Vidal, R.: Sparse hidden markov models for surgical gesture classification and skill evaluation. In: Information Processing in Computer-Assisted Interventions. Springer (2012) 167–177
20. Tao, L., Zappella, L., Hager, G.D., Vidal, R.: Surgical gesture segmentation and recognition. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI. Springer (2013) 339–346
21. Höppner, F.: Time series abstraction methods—a survey. In: GI Jahrestagung. (2002) 777–786
22. Patel, P., Keogh, E., Lin, J., Lonardi, S.: Mining motifs in massive time series databases. In: International Conference on Data Mining, IEEE (2002) 370–377
23. Moskovitch, R., Shahar, Y.: Classification-driven temporal discretization of multi-variate time series. *Data Mining and Knowledge Discovery* **29**(4) (2015) 871–913
24. Manning, C.D., Raghavan, P., Schütze, H., et al.: Introduction to information retrieval. Volume 1. Cambridge University Press (2008)
25. Ahmidi, N., Tao, L., Sefati, S., Gao, Y., Lea, C., Bejar, B., Zappella, L., Khudanpur, S., Vidal, R., Hager, G.D.: A dataset and benchmarks for segmentation and recognition of gestures in robotic surgery. *IEEE Transactions on Biomedical Engineering* (2017)