2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Joints-Space Metrics for Automatic Robotic Surgical Gestures Classification

Marco Bombieri, Diego Dall'Alba, Sanat Ramesh, Giovanni Menegozzo, Caitlin Schneider and Paolo Fiorini

Abstract— **Automated surgical gestures classification and recognition are important precursors for achieving the goal of objective evaluation of surgical skills. Many works have been done to discover and validate metrics based on the motion of instruments that can be used as features for automatic classification of surgical gestures. In this work, we present a series of angular metrics that can be used together with Cartesian-based metrics to better describe different surgical gestures. These metrics can be calculated both in Cartesian and joint space, and they are used in this work as features for automatic classification of surgical gestures. To evaluate the proposed metrics, we introduce a novel surgical dataset that contains both Cartesian and joint spaces data acquired with da Vinci Research Kit (dVRK) while a single expert operator is performing 40 subsequent suturing exercises. The obtained results confirm that the application of metrics in the joint space improves the accuracy of automatic gesture classification.**

I. Introduction

The use of Surgical Robotic Systems (SRSs) is growing worldwide, with over 6 million procedures performed on 5000 da Vinci robots [1]. This large number of procedures does not even consider other robotic platforms in use. The superior characteristics of SRS (e.g. enhanced stereo view and improved dexterity) enable a better outcome for the patients [2]. In fact, there is a strong connection between a surgeons performance (i.e. technical skills) and patient outcome. Numerous studies show that the degree of learning of a given surgical technique by the surgeon directly impacts the patients post-operation health status: low-quality surgical treatments increase the risk of serious complications [2][3]. It is, therefore, crucial to have effective and systematic training techniques in robotic surgery to improve the quality of the intervention. We define surgical competence as the skill level required to safely perform a surgical procedure [3].

The concept of surgical competence encompasses both technical and non-technical skills. Nowadays, there are numerous curricula related to training in robotic surgery to improve technical skills, however, no standardized training protocol has been defined and validated [4]. The state of the art in the evaluation of surgical skills today consists of direct or indirect observation of the surgical scene by an experienced surgeon; some problems arise [5]: first of all, the evaluation may vary depending on the examiner, secondly, the evaluation is limited by what the observer can see and by his attention span. Finally, the observation process requires a considerable amount of time and resources. Therefore, it is important to find quantitative metrics that can describe the surgical performance in details, without requiring the presence of an expert surgeon [6]. Automated surgical activity classification and recognition are important precursors for achieving the goal of objective evaluation of surgical skills [3].

Work has been completed to discover and validate metrics based on the motion of instruments that can be used as features for automatic classification of surgical gestures [6]. Most of the motion metrics have been derived from the one developed for standard minimally invasive surgery, thus they are based only on motion analysis performed in Cartesian space (e.g. end-effector velocities and accelerations or total distance traveled [7]). Some research groups have recently introduced metrics based on orientation information [6], gripping forces, and interaction forces with the environment [8]. However, the computation of these metrics requires access to low-level SRS kinematic data, which has been difficult to obtain in the past due to manufacturer trade secret and user/patient privacy. This fact has limited the development and exploitation of such advanced metrics. This fact is also confirmed by the limited number of public datasets developed for benchmarking surgical skills assessment methods, which is limited to JIGSAWS [9]. This situation is rapidly changing thanks to the introduction of SRS research platforms, such as da Vinci Research Kit (dVRK) [10] and Raven II [11]. These systems enable straightforward acquisition of low-level kinematics and status variables.

With this work, we present a series of orientation-based metrics that can be used together with the more traditional Cartesian-based metrics to have an objective assessment of how a surgical gesture should be performed. These metrics can be calculated both in Cartesian space and in joint space and they are used in this work as input features to an automatic classification algorithm. Since existing open source datasets only present information in the Cartesian space, with this work, we also introduce a new surgical dataset that contains information both in Cartesian and in joints spaces.

To reiterate, the main contributions of this paper are two: the definition of joints space metrics and the introduction of a new public annotated dataset. This allows you to objectively describe how a surgical procedure should be performed and will allow in the future to evaluate the degree of learning of

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 742671 "ARS") and under the Marie Sklodowska-Curie (grant agreement No. 813782 "ATLAS")

M. Bombieri, D. Dall'Alba, S. Ramesh, G. Menegozzo and P. Fiorini are with the Department of Computer Science, University of Verona, Verona, Italy. C. Schneider is with Electrical and Computer Engineering Department, University of British Columbia, Vancouver, Canada. Corresponding author: paolo.fiorini@univr.it

a surgical procedure by a novice. The experimental results show that the application of metrics in the joint space significantly improves the results of the automatic classification compared to those obtained by applying the metrics to the Cartesian space only, as described by [6].

II. Method

A. Metrics

In this work, we assumed that each trial is already divided into different temporal segments corresponding to different surgical actions or gestures. This segmentation could be manually performed by an expert user or automatically extracted with data processing methods [3] [12]. The main metrics considered in this work are calculated for each trial and each segment. We have extended the metrics introduced in [6] to consider their application to the joint space. In the following formulas, we will show how to calculate metrics on a generic joint and how to extend them to multiple joints simultaneously. In more detail, the considered metrics in Cartesian space are [6]:
Task Time which is defined as:

$$T = t_{I_{i+1}} - t_{I_i} \tag{1}$$

where t_{I_i} and $t_{I_{i+1}}$ are the timestamps corresponding to the beginning of segment number i and $i+1$ respectively.

Total distance travelled between two successive frames is defined as:

$$\Delta d_{i,i+1} = \|[\Delta x_i, \Delta y_i, \Delta z_i]\| \tag{2}$$

where Δx_i, Δy_i, Δz_i are the differences between frame i and $i+1$ with respect to x, y, z positions, respectively and $\|\cdot\|$ is the Euclidean norm.
Using $\Delta d_{j,j+1}$ we can define *Path Length* as:

$$P = \sum_{j=1}^{N-1} \Delta d_{j,j+1} \tag{3}$$

where N is the number of samples of which the surgical sub-action is composed.

In the joint space, we introduce the *Angular Displacement joint* (ADIJ) metric as:

$$ADIJ_k = \sum_{i=1}^{N-1} |\Delta\Theta_{i+1,i}| \tag{4}$$

The angle $\Delta\Theta_{i+1,i}$ represents the orientation change between pairs of consecutive sampled angles for the joint k.

Another metric defined in the joint space is the *Time Angular Displacement joint* (TADJ):

$$TADJ_k = \frac{1}{T} \sum_{i=1}^{N-1} |\Delta\Theta_{i,i+1}| \tag{5}$$

where T is the duration of the surgical action and k is the number of joint under analysis.
We also define the metric *Rate Of Change Joint* (ROCJ):

$$ROCJ_k = \frac{1}{N-1} \sum_{i=1}^{N-1} \omega_i \tag{6}$$

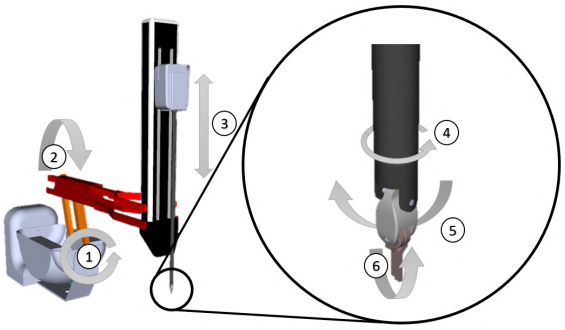

Fig. 1: Schematic representation of the considered robotic manipulator with joints configuration: on the left Base unit, on the right Instrument unit

where ω_i is the angular speed of the joint number k in the frame i and N is the number of samples.

In addition, we assume that joint data contain also information about current joint effort (e.g joint motor current, joint force or torque) as a scalar value τ_i. Based on this information we introduce the *mean effort* (MEJ) which represents how the joint k interacts with the environment:

$$MEJ_k = \frac{1}{N-1} \sum_{i=1}^{N-1} \tau_i \tag{7}$$

The structure of the da Vinci slave robotic manipulator used to do experimental validation can be divided into two parts: the *Base Unit* i.e. the first three joints of the robot and the *Instrument Unit* i.e. the last three joints of the robot, see Fig. 1. It is possible to extend these formulas in order to consider the average value of metrics on all joints of the base/instrument part of the robot. Let M_k be a metric calculated on a general joint k (M_k can be *ADIJ*, *TADJ*, *ROCJ* or *MEJ*). The value of this metric on joints of the base unit is expressible as:

$$M_{base\_unit} = \frac{1}{3} \sum_{k=1}^{3} M_k \tag{8}$$

while the value of this metric on joints of the instrument unit is expressible as:

$$M_{inst\_unit} = \frac{1}{3} \sum_{k=4}^{6} M_k \tag{9}$$

The main problem is that publicly available and freely usable surgical datasets do not contain information on joints space. With this work, we introduce a new surgical dataset that contains information both in the Cartesian space and in the joints space. This dataset will be presented in detail in the next section.

B. Dataset

We introduce a publicly dataset called YEAST (Yet Another Surgical Training Dataset). The acquired dataset consists of 42 trials of the same suturing task performed by a single expert user. The expert is right hand dominant and she has more than 50 hours of experience with da Vinci surgical robotic system. Trials can be divided into two macro-classes: the first 20 using a different phantom position than the last 20

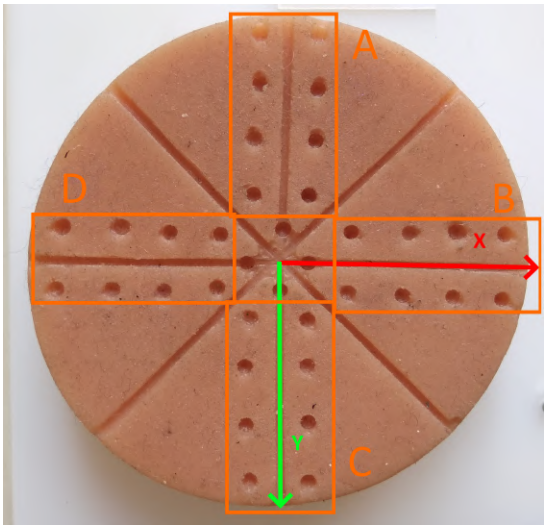

Fig. 2: Axes of the phantom (z-axis outward). Orange boxes represent the different sections of the phantom used for performing the different trials.

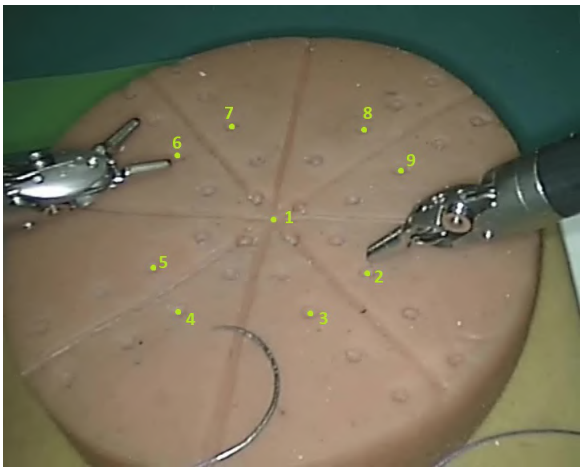

Fig. 3: Right camera endoscope image extracted from trial 22 showing the rotated position of the phantom. Numbered points indicate ordered fiducials used for spatial calibration.

in which the phantom is turned 45 degrees clockwise. Trials 21 and 42 do not contain suturing task but a procedure useful for spatial calibration as explained later in this paper. Figure 2 shows the training phantom used for the experimental validation and the common reference system defined. Each trial is executed in a different section of the phantom as represented in Figure 4 by letters A, B, C and D. The trials are executed following lettering ordering (clock-wise ordering starting from vertical section). Each trial consists in a 3 pass suturing task executed with a 1/2 circle suture needle following reference points present in the phantom.

Each trial consists of 10 Comma Separated Value (CSV) files that captured raw kinematic data (time, position and orientation) and two videos that reproduce the entire surgical scene captured by two endoscopic cameras. Files contained in the dataset are listed below:

- `(ECM|PSM1|PSM2|MTML|MTMR)_position_Cartesian_current.csv`: they contain temporal information about position and orientation in the Cartesian 3D space of the ECM, PSM1, PSM2, MTML and MTMR respectively [1]
- `(ECM|PSM1|PSM2|MTML|MTMR)_state_joint_current.csv`: they contain temporal information about position, velocity and effort of each joint for ECM, PSM1, PSM2, MTML and MTMR respectively. joints are: *outer-yaw, outer-pitch, insertion, outer-roll, outer-wrist-pitch e outer-wrist-yaw*.

The Cartesian position and the orientation of MTM and PSM are calculated starting from the joint angles using direct kinematics. Velocities in the joint space are computed directly by the dVRK robot controller. Acceleration and jerks are instead calculated with numerical derivation, using a finite difference method. We performed a spatial and temporal calibration of the robot's kinematics to have a single reference system in which all the robot's components are represented.

1) Temporal Calibration: During the surgical task, kinematics of MTM, PSM, and ECM were captured and videos recorded using two cameras (Left and Right): kinematics are gathered with a frequency of 100Hz (every 10 ms) while the video frames are updated at 25Hz (40 ms). The main problem is that the kinematics are not synchronized with respect to the video frames and consequently the description at a specific time t obtained by kinematic analysis is not consistent with the description at the same time t by video analysis: we need to find a way to temporally synchronize the kinematics with video frames; doing this we would know how to map each frame to the corresponding kinematic representation.

To solve the problem, we calculate the initial instant in which a task starts by analyzing both kinematics (obtaining t_{0_kin}) and frames (obtaining t_{0_video}). For kinematics, we calculate the Euclidean distance in 3D space traveled by PSMs between two successive updates with the assumption that without any movement, kinematics return (approximately) the same value: if this distance is greater than a threshold (experimentally estimated) then we have found t_{0_kin}. For videos we did similarity analysis between adjacent frames in order to detect movement; The time corresponding to the first motion, as seen in the video becomes t_{0_video}. For similarity analysis, we explored two alternatives: MSE (Mean Square Error) [13] metric and SSIM (Structural Similarity) [13] metric. In tasks under examination, the first metric provides more accurate results, so it was adopted in the proposed synchronization method. The effectiveness of the MSE metric for motion detection is proven by [14][15][16]. Once we find t_{0_kin} and t_{0_video} we can express the desynchronization between kinematics and video as $\Delta t\_psm = t_{0\_kin} - t_{0\_video}$. At this point, we calculate the initial synchronized time as $t_{0\_synch} = t_{0\_video} + \Delta t\_psm$. A further problem rises up: as mentioned above kinematics sampling rate is greater than frames sampling rate, so t_{0_synch} found is just a fictitious

[1]ECM: Endoscopic Camera Manipulator; PSM: Patient Side Manipulator; MTML: Master Tool Manipulator Left; MTMR: Master Tool Manipulator Right.

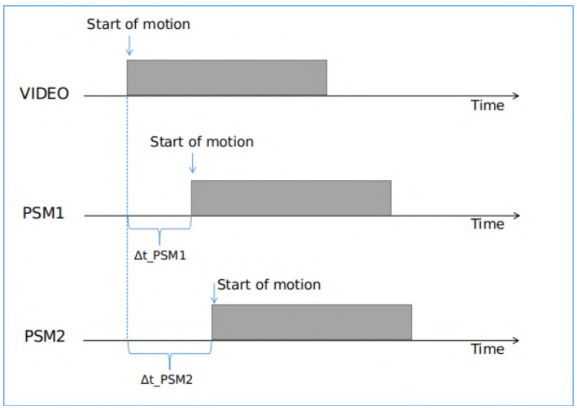

Fig. 4: Diagram representing the temporal calibration problem with different data streams and temporal offset.

timestamp that will not exactly match any real timestamp: we decided to associate it with the nearest (but not future) timestamp. The Figure 4 shows an illustration of the temporal calibration problem.

2) Spatial Calibration: PSM1 and PSM2 have independent spatial reference frames; we would like to rotate and to translate frames in order to uniform them to a unique reference frame named World. After that, we want to project the World into the camera's reference space in order to have a uniform spatial view of the scene. We choose as World the phantom on which tasks are executed; on it, we select 9 fiducials (see Figure 3) and for each, we measure the Cartesian position. The purpose is to find the best rotation and the best translation between two sets of 3D points which allow us to align them. In our application, we use a Euclidean transformation as it preserves shape and size. In the previous literature, there have been several proposed solutions for this problem: the one we adopted is taken from [8] and [17].

Adopting the procedure reported in the article above, all PSMs have been mapped to World; the remaining step is to map World in the camera space. This problem is known in the literature as *Perspective-n-Point problem* [18]: we want to find the pose of an object having a calibrated camera, locations of n 3D points of an object and the corresponding 2D projections of the same points. We solved this problem using ad hoc functions provided by the OpenCV library.

The dataset and related detailed documentation are available at *gitlab.com/altairLab/yeast-dataset.git*

C. Sub-actions

The main surgical activity was sub-divided into elementary sub-actions following information in the literature (JIGSAWS convention [9]) to which extensions were made to consider the use of both hands. The table I summarizes the elementary sub-operations (in bold our extension to JIGSAWS nomenclature).

D. Automatic classification

To further investigate the ability of the previously described metrics to discriminate gestures we trained a model classifier that, given a set of metrics, automatically recognise

TABLE I: Labels used for gesture annotation in the proposed dataset. Bold font indicate additional label introduced with respect to original JIGSAWS convention.

Gesture	Description
G1	Reaching for needle with right hand
G2	Positioning needle with right hand
G3	Pushing needle through tissue with right hand
G4	Transferring needle from left to right
G5	Moving to center with needle in grip
G6	Pulling suture with left hand
G7	Pulling suture with right hand
G8	Orienting needle with right hand
G12	Reaching for needle with left hand
G15	Pulling suture with both hands
G20	Positioning needle with left hand
G21	Pushing needle through tissue with left hand
G22	Transferring needle from right to left
G23	Orienting needle with left hand

the gestures. We selected Random Forest Classifier (RFC) as implemented from the Scikit-learn library for the following reasons [19] [20]:

- RFC shows the highest performance when applied on the general dataset without hyper tuning parameters that were outside the scope of this article [21];
- RFC has improved explainability since it is possible to characterize the Mean Decrease Impurity (MDI) to state variables importance [22];

1) Random Forest Classifier formalization: as stated in [23] a random forest is an ensemble of T axis-parallel decision trees that are trained independently. In RFC, each non-leaf node is associated with a split function $f(x;\theta)$

$$f(x;\theta) = \begin{cases} 1 \; if \; x(\theta_1) < \theta_2 \\ 0 \; otherwise \end{cases}$$

where $\theta_1 \in \{1,2,...,d\}$ is the selected feature and $\theta_2 \in \mathbb{R}$ is a threshold. The outcome determines the child node to which x is routed. For instance, 0 may represent the left child node while 1 may represent the right child node. The leaf nodes of the tree either store class probability distributions or class labels based on the training samples they receive. During testing, for a test sample x, each tree returns a probability distribution $p_t(y|x)$ stored on the leaf node it falls into, and the class label is obtained via averaging.

2) Random Forest Classifier validation: due to the high imbalances in the number of gestures reported in Figure 5 we validate our classifier using Stratified k Fold(SKF) methodology. In SKF cross-validation, the folds are created in a way that they contain approximately the same proportion of predictor labels as the original dataset. Since the lowest number of samples for a gesture is n=5, we create 5 test and train sets. This was done to maintain near 20% of samples for testing and guarantee the presence of each gesture in the splits.

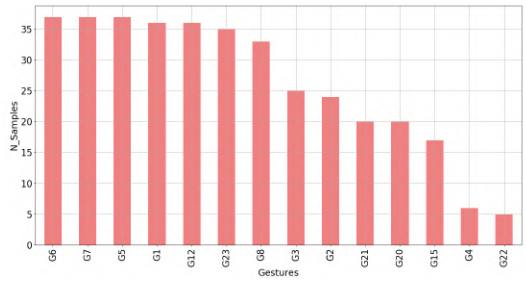

Fig. 5: Histogram for gestures distribution, considering only gesture labels annotated in the proposed dataset.

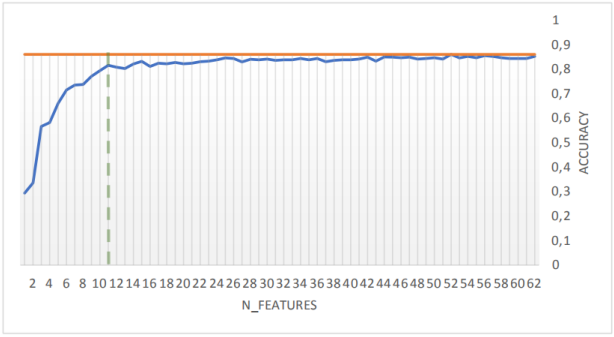

Fig. 6: Accuracy curve considering an increasing number of features.

TABLE II: Average classification accuracy, considering the different subset of metrics considered for the experimental evaluation.

Features	Average Accuracy
Joint space	83.01%
Cartesian space	75.27%
All metrics	86.51%

III. Results and Discussions

A. Temporal and Spatial calibration

The mean temporal desynchronization between video frames and PSMs are of 140*ms*. The mean error for spatial calibration is 1.5 mm for PSM1 and 1.6 mm for PSM2. Errors in spatial calibration are attributable to the deformability of the phantom.

B. Automatic classification of surgical gestures

In this section, we will discuss some of the results obtained from the automatic classification of surgical gestures using as descriptive features, the metrics presented in section II-A. The experiment was accomplished on the dataset described in section II-B. Specifically, we analyze the contribution of pose (distance and acceleration), orientation (Cartesian space and joints space), and effort (joints space) metrics to the gesture classification, observing changes on accuracy performance.

In table II we report the mean accuracy obtained using Stratified k Fold ($k = 5$) methodology on considered three groups of metrics. The highest average accuracy with a score of 86.51% is obtained using both Cartesian and joint space. In Figure 6 we show that few features are highly discriminative for the dataset since the average accuracy grows very quickly. The ten most important metrics used for automatic classification are shown in Table IV. Five (shown in bold) out of the ten most important features are from joint space and six of them are orientation based features.

Furthermore, it emerges that the joint space analysis (made possible by the dataset we introduced) improves the quality of the classification. Also, the joints sub-division in two classes (base unit and instrument unit) allows obtaining important features for the classification as shown by Table IV. Table III shows Precision, Recall, and F-Score for each gesture and three groups of metrics: Cartesian space metrics, joints space metrics, and combined analysis of Cartesian and joint metrics. Precision evaluates the fraction of correctly classified instances among the ones classified as positive, Recall is a metric that quantifies the number of correct positive predictions made out of all positive predictions that could have been made; unlike precision that only comments on the correct positive predictions out of all positive predictions, recall indicates missed positive predictions. In this way, recall provides some notion of the coverage of the positive class. F-Score provides a way to combine precision and recall into a single measure that captures both properties. Table III shows that some gestures have better metric scores in proposed joint space and most of the gestures are better recognized in the combined Cartesian and joint space. Our hypothesis that joint space helps capture more information and improves the performance of the model is validated: our proposed orientation based metrics in joint space are a good set of features that can be used to develop methods for automated classification of surgical gestures.

The classification method, however, fails for some gestures, like for example for G4 (*Transferring needle from left to right*) and G22 (*Transferring needle from right to left*). The reason is that the number of occurrences of such gestures is very low: 4 occurrences for G4 and 6 occurrences for G22. Another reason could be related to the intrinsic similarity of G4 and G22: from kinematic variable it is difficult to estimate which instrument is holding the needle and therefore distinguish between these two gestures.

IV. Conclusions and Future Works

With this work, we address the lack in the literature about the use of orientation-based metrics in both Cartesian space and joint space to objectively describe the characteristics of a surgical sub-operation of which the main activity is composed. The effectiveness of metrics has been validated using them as features for an automatic classification algorithm. Some surgical sub-operations can be well described using metrics based on Cartesian position, other sub-operations can instead be well described using orientation-based metrics. We have also introduced a new dataset that contains information both in the Cartesian space and in the joint space. It was used to make experimental validation of the proposed metrics.

This work will later be extended to allow an automatic segmentation of the main surgical procedure and to permit

TABLE III: Support, precision, recall and F-Score of considered surgical gestures.

Gestures	Support	Cartesian space			Joints space			Cartesian + Joint space		
		Precision	Recall	F-Score	Precision	Recall	F-Score	Precision	Recall	F-Score
G1	7.2000	0.7148	0.6429	0.6639	0.7864	**0.8357**	0.7993	**0.8556**	**0.8357**	**0.8363**
G2	4.8000	**0.9029**	**0.9200**	**0.9044**	0.8095	0.9000	0.8485	0.8762	0.9000	0.8810
G3	5.0000	0.8529	0.8800	0.8600	0.8333	0.9200	0.8727	**0.8600**	**0.9200**	**0.8873**
G4	1.2000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
G5	7.4000	0.6495	0.6464	0.6337	0.6752	0.8143	0.7259	**0.7844**	**0.8357**	**0.7907**
G6	7.4000	0.8950	0.9429	0.9115	**0.9492**	**0.9714**	**0.9597**	0.9092	**0.9714**	0.9374
G7	7.4000	0.8051	0.9464	0.8655	**0.8833**	0.9750	0.9249	**0.8833**	**1.0000**	**0.9366**
G8	6.6000	0.6267	0.6524	0.6217	0.7992	0.7238	0.7453	**0.8444**	**0.8429**	**0.8359**
G12	7.2000	0.7853	0.7571	0.7577	0.8929	**0.8929**	**0.8894**	**0.9095**	0.8679	0.8857
G15	3.4000	0.7333	**0.5500**	**0.6248**	0.5000	0.2500	0.3267	**0.8333**	0.4167	0.5343
G20	4.0000	0.9143	0.7500	0.7588	**0.9500**	0.8000	0.8548	0.9200	**0.9500**	**0.9270**
G21	4.0000	**1.0000**	0.8500	0.8800	**1.0000**	**0.9000**	**0.9333**	**1.0000**	**0.9000**	**0.9333**
G22	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
G23	7.0000	0.5813	0.6857	0.6109	0.7214	0.7714	0.7222	**0.7922**	**0.8571**	**0.7936**

TABLE IV: Ten most significant features for gesture classification. Bold font is used to indicate proposed joint space metrics.

#	Feature
1	distance on z-axis of PSM1
2	acceleration on z-axis of PSM1
3	distance on z-axis of PSM2
4	acceleration on z-axis of PSM2
5	**mean effort on sixth joint of PSM2**
6	**mean effort on sixth joint of PSM1**
7	**mean effort on PSM2's Instrument Unit**
8	**angular Displacement of PSM2's fifth joint**
9	rate of change of PSM1
10	**angular Displacement of PSM2's forth joint**

an automatic assessment of surgical skills as shown by [6] based on the metrics introduced. We are also working for extending the dataset with more experts demonstrations.

REFERENCES

[1] I. S. Inc., "https://isrg.gcsweb.com/static-files/880bf027-e866-4c32-b910-5332467cd8dc."

[2] J. Dwivedi and I. Mahgoub, "Robotic surgery -a review on recent advances in surgical robotic systems," 01 2012.

[3] R. DiPietro, N. Ahmidi, A. Malpani, M. Waldram, G. Lee, M. Lee, S. Vedula, and G. Hager, "Segmenting and classifying activities in robot-assisted surgery with recurrent neural networks," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, pp. 1–16, 04 2019.

[4] J. Chen, N. Cheng, G. Cacciamani, P. Oh, M. Lin-Brande, D. Remulla, I. Gill, and A. Hung, "Objective assessment of robotic surgical technical skill: A systematic review," *The Journal of Urology*, vol. 201, 07 2018.

[5] S. Gearhart, M. Wang, M. Gilson, B. Chen, and D. Kern, "Teaching and assessing technical proficiency in surgical subspecialty fellowships," *Journal of surgical education*, vol. 69, pp. 521–8, 07 2012.

[6] Y. Sharon, T. Lendvay, and I. Nisky, "Instrument orientation-based metrics for surgical skill evaluation in robot-assisted and open needle driving," 09 2017.

[7] C. Reiley, H. Lin, D. Yuh, and G. Hager, "Review of methods for objective surgical skill evaluation," *Surgical endoscopy*, vol. 25, pp. 356–66, 02 2011.

[8] P. Besl and H. McKay, "A method for registration of 3-d shapes. ieee trans pattern anal mach intell," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 14, pp. 239–256, 03 1992.

[9] N. Ahmidi, L. Tao, S. Sefati, Y. Gao, C. Lea, B. Bjar, L. Zappella, S. Khudanpur, R. Vidal, and G. Hager, "A dataset and benchmarks for segmentation and recognition of gestures in robotic surgery," *IEEE Transactions on Biomedical Engineering*, vol. PP, pp. 1–1, 01 2017.

[10] I. S. Inc., ""http://www.intuitive-foundation.org/dvrk/","

[11] Y. Li, B. Hannaford, and J. Rosen, "Raven: Open surgical robotic platforms," 06 2019.

[12] G. Menegozzo, D. DallrAlba, C. Zandon, and P. Fiorini, "Surgical gesture recognition with time delay neural network based on kinematic data," pp. 1–7, 04 2019.

[13] M. DeGroot and M. Schervish, *Probability and Statistics*, vol. 151. 01 2002.

[14] D. Qian, X. Pei, and X. Li, *The Application of Equivalent Mean Square Error Method in Scalable Video Perceptual Quality*, pp. 3–7. 01 2018.

[15] E. Kerre and M. Nachtegael, "Fuzzy techniques in image processing," *Physica-Verlag*, vol. 52, 01 2000.

[16] B. Singh, D. Singh, G. Singh, N. Sharma, and V. Kumar, "Motion detection for video surveillance," pp. 578–584, 07 2014.

[17] K. Arun, T. Huang, and S. Blostein, "Least-squares fitting of two 3-d point sets. ieee t pattern anal," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-9, pp. 698 – 700, 10 1987.

[18] S. Li, C. Xu, and M. Xie, "A robust o(n) solution to the perspective-n-point problem," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, 01 2012.

[19] G. Louppe, *Understanding Random Forests: From Theory to Practice*. PhD thesis, 10 2014.

[20] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, I. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-iearn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, 01 2011.

[21] M. Fernandez-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?," *Journal of Machine Learning Research*, vol. 15, pp. 3133–3181, 10 2014.

[22] G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts, "Understanding variable importances in forests of randomized trees," vol. 26, 12 2013.

[23] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 10 2001.