International Journal of Computer Assisted Radiology and Surgery (2019) 14:117–127
https://doi.org/10.1007/s11548-018-1866-8

ORIGINAL ARTICLE

Conditions for reliable grip force and jaw angle estimation of da Vinci surgical tools

Trevor K. Stephens[1] · John J. O'Neill[1] · Nathan J. Kong[1] · Mark V. Mazzeo[2] · Jack E. Norfleet[2] · Robert M. Sweet[3] · Timothy M. Kowalewski[1]

Received: 30 March 2018 / Accepted: 24 September 2018 / Published online: 4 October 2018

Abstract

Purpose This work presents an estimation technique as well as corresponding conditions which are necessary to produce an accurate estimate of grip force and jaw angle on a da Vinci surgical tool using back-end sensors alone.

Methods This work utilizes an artificial neural network as the regression estimator on a dataset acquired from custom hardware on the proximal and distal ends. Through a series of experiments, we test the effect of estimation accuracy due to change in operating frequency, using the opposite jaw, and using different tools. A case study is then presented comparing our estimation technique with direct measurements of material response curves on two synthetic tissue surrogates.

Results We establish the following criteria as necessary to produce an accurate estimate: operate within training frequency bounds, use the same side jaw, and use the same tool. Under these criteria, an average root mean square error of 1.04 mN m in grip force and 0.17 degrees in jaw angle is achieved. Additionally, applying these criteria in the case study resulted in direct measurements which fell within the 95% confidence bands of our estimation technique.

Conclusion Our estimation technique, along with important training criteria, is presented herein to further improve the literature pertaining to grip force estimation. We propose the training criteria to begin establishing bounds on the applicability of estimation techniques used for grip force estimation for eventual translation into clinical practice.

Keywords Grip force estimation · Surgical robotics · Artificial neural network

Introduction

The introduction of surgical robots such as the da Vinci surgical system has afforded surgeons with improved dexterity, scalable motions, and favorable ergonomic use [17]. Additional upgrades continue to be added the system such as the ability to track robotic end-effector location, though this software is proprietary and currently not widely available. Despite these continued technological improvements, a key aspect missing in these surgical robots is a reliable estimate of grip force at the end effector. Surgical robots are tele-operated, and therefore, there is no physical manipulation of the patient when using these robots. This lack of force feedback to the surgeon presents various problems such as a lack of diagnostic information through palpation [28], increased potential for tissue crush injury [4,10], and a lack of sufficient force application for tasks such as suturing [20]. Additionally, a reliable grip force estimate may prove valuable for ongoing research to automate surgical robot tasks [11,25], identify tissue in real time [5,26,27], as well as create more tissue-realistic models for surgical simulation and training [2,3]. This need for sufficiently realistic simulation is notably acute. Mechanical differences in the behavior of human tissues occur between healthy and diseased tissues, different patients, and even different regions on the same patient. There are, however, enough similarities in the behavior of tissues that appropriate decision can be made based on tissue behaviors. Just as a physician can identify a stiff muscle by touching a live patient (in vivo), muscle stiffness is quantifiable with the proper measurements. Unfortunately, reliable data are insufficient for human tissue elasticity for

✉ Trevor K. Stephens
steph594@umn.edu

[1] Department of Mechanical Engineering, University of Minnesota, Minneapolis, MN, USA

[2] Simulation and Training Technology Center, Army Research Laboratory, Orlando, FL, USA

[3] Department of Urology, University of Washington, Seattle, WA, USA

the purposes of distinguishing healthy tissues from unhealthy ones [6]. A limiting factor in collecting this data is the lack of a reliable method to measure those properties from human tissue samples in vivo [6], especially at a sufficiently large scale to construct a representative database of tissue mechanical responses. The da Vinci robot, however, is in widespread use and in daily contact with these tissues. Accurate grip force and position measurements ascertained from da Vinci procedures may provide unique opportunities to capture and exploit such valuable data.

Prior art has addressed the lack of grip force measurements by first discussing the viability of sensor placement on the surgical robot. There are three candidate locations to place sensors on da Vinci tools: the distal grasping end, the shaft, and the proximal actuating end [24].

Placing sensors directly at the distal grasping end would provide the best-case measurement, as the direct measurement circumvents a need for estimation. Major hindrances previously reported preventing this application include size constraints, cost, and sterilization issues [24]. The concerns with size have recently been addressed with various novel designs for force sensors integrated into the surgical grasper [12,13,23]. However, these designs require either customizations or additions to the existing surgical hardware and likely have not been adopted in clinical use due to remaining concerns of additional cost and sterilization. As explained in [24], sterilization of surgical components which come in contact with the human body is required to help prevent the spread of diseases. The most common of these sterilization techniques is a steam sterilization process known as autoclaving. The sustained high temperature, pressure, and humidity used during autoclaving can destroy electrical components of the sensors, rendering them useless. As [24] explains, other methods for sterilization such as chemical techniques are possible, but have also not been readily adopted due to lengthened time and adverse effects of some chemical agents on tools. The challenges of adopting customized tools as well as new sterilization techniques, even without the mention of added costs, suggest that sensing at the distal end is a challenging prospect.

Placing sensors on the shaft is possibly confounded by the high levels of friction at the insertion port, as well as the aforementioned concerns of sterilization and cost. The proximal end of the da Vinci tool (at the spindles) defaults as a viable option for sensor placement, but suffers from the potential for less accuracy in grasping force given that it is an indirect measurement. Therefore, estimation techniques, although not preferential, become necessary to convert the proximal-end sensor measurements to a distal-end grip force estimate.

Attempts at estimating grip force at the distal end of the da Vinci tool using only proximally located sensors are often called sensorless grip force estimation [19]. Despite this not being a truly sensor-free approach, sensorless refers to the fact that the sensors used in the process are already inherent in the surgical robot. For surgical robots, the back-end measurements available are motor position via encoder readings and motor torque via an estimate of motor current. Motor velocity can be ascertained via differentiation of position readings. Additionally, true motor torque readings could feasibly be added via in-line torque senors on the back end, but are currently not installed.

Most prior art concerning sensorless grip force estimation falls into the category of physics-based modeling of surgical tools to convert proximal measurements to distal estimates of grip force utilizing a model rooted in physical parameters (e.g., friction and hysteresis). In [1], a quasi-static model of a da Vinci tool is developed using a dual tendon–sheath model. In [20], an unscented Kalman filter was utilized to estimate grip force of a 10-mm gripper driven by a Raven-II surgical robot. This filter requires a dynamic model of the surgical instrument, which has several tunable parameters which vary with environment and system configuration [19,20].

Tuning every model parameter was suggested to be too complex of a solution for grip force estimation given the limited number of sensors available [19]. An alternative approach is to utilize regression techniques which do not explicitly rely on the dynamic model of the system, which we will refer to as data-driven modeling. These data-driven models contain parameters which may not directly relate to the physical characteristics of the system, and therefore, understanding an exact physical model is not required. These approaches include kernel-based approaches such as support vector machines (SVM) and Gaussian process regression (GPR) as well as decision trees, bagged trees, random forests, and neural networks. In [19], GPR was used to predict grip force of the Raven surgical robot on a single surgical tool and was capable of outperforming the dynamic physics-based modeling approach. This work illuminated potential advantages from utilizing a data-driven estimation technique.

For each of these data-driven techniques, model parameters are trained using large amounts of input–output data of the system. As mentioned, these model parameters may not necessarily relate to physical properties of the system (e.g., friction), but rather an arbitrary parameter within the model (e.g., weights of interconnected nodes of a neural network). Once the training is complete, the derived model is then used on newly acquired input data to predict the output of the system. Although these data-driven techniques have shown promising accuracy, the lack of physical intuition of the model parameters makes it difficult to understand how well the models generalize to unseen operating conditions. Prior work such as [19] have been limited to single tools, fixed frequencies, and limited grip force ranges.

In this work, we estimate grip force and jaw angle of a da Vinci EndoWrist Surgical tool by utilizing a neural

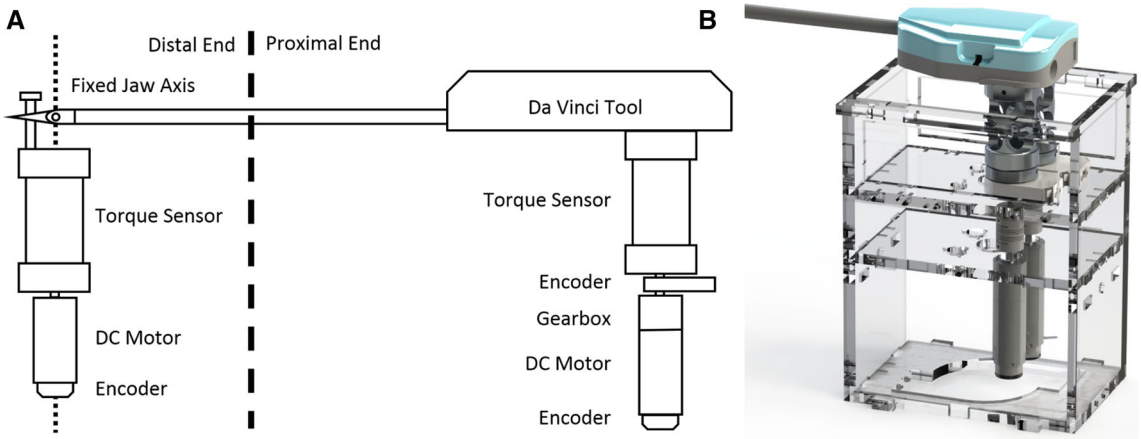

Fig. 1 **a** Conceptual diagram depicting distal and proximal sensors used for data collection, **b** rendered image depicting hardware on proximal end

network. The objective of this research is to test a data-driven regression estimation technique under a spectrum of training conditions to understand generalizability of grip force and jaw angle predictions. The grasps represent typical robotic surgical grasping, and estimation is achieved based on proximal-end measurements alone. We define a typical grasp to be the closing portion of the grasp outputting between 5 and 55 mN m and spanning a range of 50 degrees from the jaw's neutral position. We collect and provide an open-source dataset with time-synced proximal- and distal-end measurements with 0.1 mN m precision and 0.09 degree resolution for distal-end torque and distal-end position, respectively. Through a series of four experimental neural network training conditions, we propose reliability criteria on training conditions to produce acceptable estimates. The four experimental conditions used for training the neural net are the following: a baseline condition (Experiment 1: Baseline), a frequency comparison (Experiment 2: Leave-frequency-out), a jaw comparison (Experiment 3: Leave-jaw-out), and a tool comparison (Experiment 4: Leave-tool-out). Testing the estimation accuracy across several conditions is done to illuminate the generalizability of data-driven estimation techniques, since previous work has reported estimation accuracy only under limited conditions. Utilizing this newly established reliability criteria, we present a case study estimating mechanical material response of two synthetic tissue phantoms and compare these with a ground truth measurement taken at the site of grasping.

Methods

Hardware setup

Data collection was accomplished via custom hardware on the proximal and distal ends of da Vinci EndoWrist surgical tools. The proximal-end hardware drives the spindles while simultaneously measuring position and torque of the spindles. The distal-end hardware provides programmable reaction torques to the jaw while simultaneously measuring the position and torque at the jaw. This allows for testing multiple reaction torque levels at a variety of frequencies and allows measuring reaction torque while the jaw is in motion. A conceptual diagram of both distal- and proximal-end hardware is shown in Fig. 1a. Details of the distal-end hardware can be found in [14]. Details of the proximal-end hardware are included herein.

A rendering of the proximal-end hardware is shown in Fig. 1b for clarity. This hardware consists of a pair of motors, encoders, and torque sensors attached in-line with two of the spindles on the da Vinci endoWrist. For this study, only one spindle was driven at a time. The motors are DCX 19 S Maxon motors (Maxon Motor, Sachseln, Switzerland), which are used to drive the spindles. Each motor has a 35:1 reduced backlash planetary gearhead. The motors are driven by an ESCON Module 24/2 Servo Controller. The torque sensors are FUTEK TFF325 reaction torque sensors (FUTEK Advanced Sensor Technology, Inc., Irvine, California) with capabilities of measuring up to 1.4 Nm of torque; each torque sensor is situated in-line with the spindle to accurately measure torque applied as the motor drives the spindle. The last sensor on the proximal side is a CUI AMT 102 encoder (CUI Inc., Tualatin, OR) with 8192 counts per revolution. The encoder measures spindle position and is also utilized to obtain an estimate of spindle velocity. These sensors placed on the proximal end mimic the best-case scenario of sensor streams potentially available on da Vinci robots and are used for initial verification in finding an accurate estimation of grip force and jaw angle when mapping measurements at the spindle to the end effector of da Vinci tools.

A Teensy 3.5 microcontroller was used for all experimentation. The proximal- and distal-end data streams were

synchronized for Experiments 1 through 4. The sensor data streams were collected at 1 kHz and saved to a microSD card. The entire dataset of time-synced proximal- and distal-end measurements is provided as an open-source dataset to further test and compare estimation techniques.[1] The distal-end torque measurements have a precision of 0.1 mN m, and the distal-end encoder has a resolution of 0.09 degrees. The torque precision was determined by taking 10,000 consecutive samples of the torque sensor in a motionless state and calculating the corresponding standard deviation of these measurements. The angle resolution was determined from the encoder's datasheet.

This hardware setup allows for full jaw range of motion but does not currently allow testing of different wrist pitch or roll angles. The spindles actuating pitch and roll degrees of freedom were fixed in their nominal states with physical restraints attached directly to the spindles at the proximal end. This was done to avoid the near threefold variation in grip force observed when changing between extreme postures as reported by [18].

Artificial neural network

The neural network utilized in this work consisted of an input layer, a hidden layer, and an output layer. This structure was chosen due to Kolmogorov's theorem, which states that any continuous function can be represented arbitrarily well with a single hidden layer [8]. The three nodes in the input layer consisted of the following features measured by sensors on the proximal end: motor position, motor velocity, and motor torque. These nodes were interconnected with the single hidden layer, which was comprised of 30 nodes. The number of nodes in this hidden layer was determined by iterating across several network sizes (5–60 nodes) and found that there were diminishing returns after 30 nodes. The hidden layer nodes were interconnected with a single-output node. Two separate types of neural networks were trained for simplicity and reduction in training time during development: one with end-effector torque as the output feature and one with end-effector position as the output feature. An alternative approach could be to train a single multi-output neural network as opposed to multiple single-output networks. A potential benefit to the multi-output neural network could be parameter sharing within the input layer; this could lead to improved accuracy if the angle and force estimates share a common underlying representation trying to be learned. This work did not analyze the multi-output network and remains a possibility for future work.

All input and output features were normalized using min–max normalization as a preprocessing step and randomized to remove patterns within the dataset. The activation function at the hidden layer was a log-sigmoid transfer function, and the activation function at the output layer was a purely linear transfer function. The training method was Bayesian regularization backpropagation [7,21] as implemented in the MATLAB Neural Network Toolbox (The MathWorks, Inc., Natick, Massachusetts). The training data were collected using the aforementioned hardware with the sensors on the proximal and distal sides as described. For each trained neural network, datasets were randomly partitioned into five equally sized folds to preform five-fold cross-validation.

Dataset

For each network, the input features were position, velocity, and torque as acquired on the proximal end. Position was run through a low-pass Butterworth filter (5th order, 50 Hz cutoff frequency) to smooth out discretized encoder data. Velocity was found by implementing a noise-robust numerical differentiation method as described by Holoborodko (11th order, non-causal numerical differentiator) [9]. The training data consisted of 20-s sinusoidal grasp trajectories with pairwise combinations of five frequencies ranging from 0.1 to 0.5 Hz and four resistive torques ranging from 5 to 55 mN m.

The selected frequency range was determined by encapsulating the range of jaw velocities found within an existing database of laparoscopic tasks collected using the electronic data generation and evaluation (EDGE) device [16]. This database consisted of 124 subjects across eight academic urology training programs in the USA with a total of 454 recordings on tasks such as peg transfer, pattern cutting, suturing, and clip applying tasks. For the surrogate tissue manipulation tasks (peg transfer and pattern cutting), the max jaw velocity averaged across all users was 0.7 ± 0.2 rad/s. Our training dataset included velocities which reached a peak of nearly 1.4 rad/s, which is well sufficient to encapsulate realistic jaw velocities as measured in surgical training tasks.

The resistive torques were applied by sending a constant current command to the distal-end motor to guarantee the torque sensor did not lose contact with the jaw, and therefore provide continuous measurements. The four current levels were commanded as constant values, but the torque values were allowed to fluctuate corresponding to changing tool dynamics throughout the trajectory. This was seen as a benefit as we were able to span a wider range of torque values to fill out the training set. The maximum torque value reaches a peak of approximately 5.5 Newtons as computed from the fixed 10-mm lever arm. These grip force/torques fall within the range of measured da Vinci tool grip forces as reported in the literature [22].

The error between the testing data (y) and neural network estimate ($\hat{y}$) is computed in two different approaches: mean absolute error (MAE) as defined in Eq. 1 and root mean

[1] The full dataset is hosted at https://github.com/MRDLab/mis-tool-characterization.

square error (RMSE) as defined in Eq. 2. These two error metrics are reported in each table

$$\mathrm{MAE} = \frac{\sum_{t=1}^{T_f} |\hat{y}_t - y_t|}{T_f}, \tag{1}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=1}^{T_f} \left(\hat{y}_t - y_t\right)^2}{T_f}}. \tag{2}$$

Additionally, to generate boxplots the error is computed at each time step by computing the difference between the estimate and the measured value (i.e., $\hat{y}_t - y_t$). The absolute value is intentionally omitted to preserve the sign of the difference to visually determine whether the error is zero mean. Boxplots are created with the middle line representing the median value in the dataset. The bottom line of the box represents the 25th percentile (q_1) and the top line of the box represents the 75th percentile (q_2) of the data. The whiskers extend to the smallest and largest data points which are not considered outliers. Data points are considered outliers if they fall below the threshold (γ^-) in Eq. 3 or above the threshold (γ^+) in Eq. 4. Outliers are plotted with a dot symbol

$$\gamma^- = q_1 - 1.5(q_3 - q_1) \tag{3}$$

$$\gamma^+ = q_3 + 1.5(q_3 - q_1). \tag{4}$$

Experiment 1: Baseline

Experiment 1 establishes a baseline of using a neural network to predict grip force and jaw angle for a single frequency on a single tool. This represents a best-case scenario for estimation. The tool used was the left jaw of a Maryland Bipolar Forceps tool. For convenience, other standard regression techniques which have not previously been used within the literature for grip force estimation are also presented for comparison. The regression techniques tested include decision trees, bagged trees, random forests, and support vector machines. Gaussian process regression, which was used in [19], was not used in this comparison, but the results of that work will be compared to our results later as part of the discussion section. For each of these regression techniques, several parameter values were tested to provide a more general comparison.

For this experiment, all five frequencies were used independently to train each of the regression models. For each frequency, five-fold cross-validation was performed. For each fold, the designated test data, which were not used in the training process, were run through the regression model and averaged to provide estimates of end-effector position and torque.

Experiment 2: Leave-frequency-out

For all remaining experiments, neural networks were used as the only regression model to examine training conditions.

To test the generalizability of the neural network approach for frequencies not present in the training data, a leave-frequency-out cross-validation study was performed across each frequency. Five separate neural nets were trained leaving out an entire frequency each time. The designated test data consisted of all five folds of data from Experiment 1, corresponding to the left-out frequency.

Since all five frequencies were used, this provided two examples of extrapolation (the end points of 0.1 Hz and 0.5 Hz) and three examples of interpolation.

Experiment 3: Leave-jaw-out

The next experiment was conducted to test whether a single neural network can be applied successfully to both right and left jaws without retraining. This constituted a leave-jaw-out cross-validation study. First, the right jaw's data streams had to be transformed into the left jaw's coordinate system by reflecting about the origin.

The left jaw's data were run through a neural net trained on the left jaw's training data (labeled as same), as well as a neural net trained on the right jaw's training data (labeled as opposite) for each of the five randomly partitioned folds. The right jaw's data were also tested against both neural nets for each of the folds. The experimental data labeled as same also provide a baseline estimate of how the neural net performs when trained on all training data from a specific jaw.

Experiment 4: Leave-tool-out

This experiment utilized the datasets from two different tools. Both tools were Maryland Bipolar Forceps tools of a similar age and use level. For simplicity, the inter-tool comparison utilized only the left jaw. The first tool's test data were run through a neural net trained on the first tool's training data (labeled as same), as well as a neural net trained on the other tool's data (labeled as opposite) for each of the five randomly partitioned folds. The other tool's test data were also tested against both neural nets for each of the folds.

Reliability criteria determination

Reliability criteria are the determined training conditions which we propose that were found to produce an estimation with considerably lower root mean square error (RMSE). The reliability criteria were determined post hoc after conducting the four experiments outlined. Throughout the paper, an asterisk was used to mark conditions which fell within the reliability criteria. This asterisk was applied to the plots and

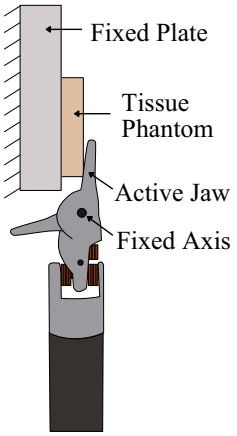

Fig. 2 Diagram of case study setup

tables in the results section in post once determining these criteria. The post hoc analysis consisted of finding clear separation between the RMSE between experimental conditions; the conditions with lower RMSE were considered to be part of the reliability criteria.

Material response case study

To evaluate the efficacy of our estimation technique and reliability criteria, a case study was set up to compare two different synthetic tissue's torque–displacement curves between two approaches: proximal-end sensors with estimation from the neural net and distal-end sensors at the jaw.

The proximal-end measurements were collected across the same frequency range and for the same duration as Experiment 1, which resulted in 30 total grasps. The neural network used to estimate force and jaw angle from these proximal measurements adhered to the reliability criteria. The distal-end measurements were taken by affixing a da Vinci tool jaw directly to the distal-end torque sensor and actuating the grasp via the distal-side motor. The cables of the da Vinci tool were cut to allow the jaw to rotate freely without back-driving the entire cable–pulley system. The motor command was slowly ramped up to 100% over 20-s while measuring position and torque.

This setup provided equal comparison of jaw grasping area between both data collections for fair comparison of torque–displacement curves. Additionally, our technique to estimate torque makes this experimental setup tool-agnostic, as any da Vinci tool could be used as long as a proper neural network is trained. For this work, we trained on a Maryland Bipolar forceps and used it for the case study because of the widespread use of this tool. For each data collection, only one jaw was used, compressing the phantom between the jaw and a fixed plate. The test setup is depicted in Fig. 2.

All 30 grasps from the proximal end were averaged by fitting a second-order polynomial to the data and reported with the corresponding 95% confidence interval.

Table 1 Jaw angle and torque estimation error across several standard regression techniques and parameters

Regression method	RMSE (deg)	RMSE (mN m)
Neural network (1 hidden layer, 30 nodes)	0.08	0.71
Decision tree (leaf size = 4)	0.22	0.48
Decision tree (leaf size = 12)	0.23	0.59
Decision tree (leaf size = 36)	0.25	0.78
Bagged trees (no feature bagging, leaf size = 4)	0.12	0.37
Bagged trees (no feature bagging, leaf size = 12)	0.13	0.46
Bagged trees (no feature bagging, leaf size = 36)	0.14	0.63
Random forest (feature bagging = 2, leaf size = 4)	0.11	0.33
Random forest (feature bagging = 2, leaf size = 12)	0.12	0.40
Random forest (feature bagging = 2, leaf size = 36)	0.15	0.56
Support vector machine (linear kernel)	0.02	4.3
Support vector machine (quadratic kernel)	0.02	2.0
Support vector machine (cubic kernel)	0.02	3.3
Support vector machine (fine Gaussian kernel)	0.03	0.99
Support vector machine (medium Gaussian kernel)	0.02	1.1
Support vector machine (coarse Gaussian kernel)	0.02	1.8

Results

Experiment 1 results: Baseline

The results comparing standard regression techniques are given in Table 1. Random forests resulted in lowest RMSE for torque estimation (0.20 mN m). Neural networks resulted in slightly higher RMSE, and SVM methods produced the highest RMSE across all kernel selections. Conversely, for jaw angle estimation SVM methods resulted in the lowest RMSE (0.02 degrees), with neural networks than decision tree-based methods (decision tree, bagged tree, and random forest) producing higher RMSE. The decision tree-based methods exhibited an undesirable frequency-dependent error. These approaches performed very well at low frequencies (0.1–0.2 Hz), but exhibited significant drop-off at higher frequencies (0.4–0.5 Hz). Neural networks did not exhibit any frequency-dependent degradation of performance.

Specifically for neural networks, the average RMSE across all frequencies in Experiment 1 was 0.08 degrees and 0.71 mN m. For perspective, the average angular operating range spanned 52 degrees and output an average torque load

Table 2 Jaw angle and torque cross-validation results, where an asterisk represents experimental conditions considered to be a part of the reliability criteria

	Jaw angle estimation (deg)			Torque estimation (mN m)		
	MAE	95th%	RMSE	MAE	95th%	RMSE
Exp 1 Baseline*	0.06	0.15	0.08	0.51	1.37	0.71
Exp 2 Interpolation*	0.43	1.16	0.56	1.24	3.90	1.92
Exp 2 Extrapolation	1.04	3.87	1.64	4.29	19.8	7.21
Exp 3 Same*	0.09	0.23	0.13	0.65	1.85	0.96
Exp 3 Opposite	5.63	6.75	5.75	7.97	11.6	8.73
Exp 4 Same*	0.10	0.25	0.14	0.65	1.88	1.01
Exp 4 Opposite	4.97	6.16	5.05	6.75	11.3	7.67

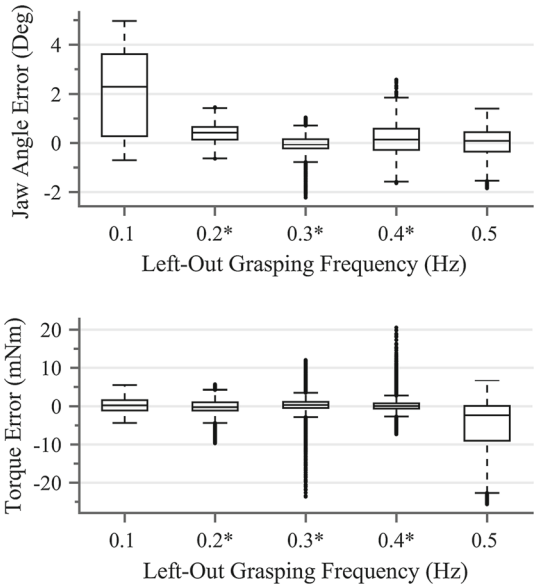

Fig. 3 Experiment 2: Leave-frequency-out cross-validation results. For each frequency $N = 26{,}868$ time samples on average. An asterisk represents experimental conditions considered to be part of the reliability criteria

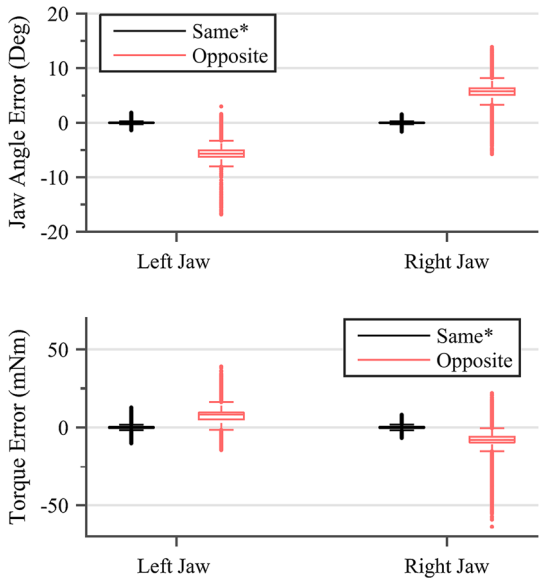

Fig. 4 Experiment 3: Leave-jaw-out cross-validation results. For each jaw $N = 134{,}869$ time samples on average. An asterisk represents experimental conditions considered to be part of the reliability criteria

of 32 mN m across all experimentation. The MAE, 95th percentile of MAE, and RMSE for this experiment and all other experiments are reported in Table 2.

Experiment 2 results: Leave-frequency-out

The grip force and jaw angle signed errors for Experiment 2 are reported in Fig. 3. The top plot shows error in jaw angle estimation, and the bottom plot shows error in grip force estimation. Extrapolation outside of the training frequencies led to a larger magnitude of errors than interpolation within the training frequencies. For jaw angle, the worst result was extrapolation to 0.1 Hz with an RMSE of 2.66 degrees. For torque, the worst result was extrapolation to 0.5 Hz with an RMSE of 8.86 mN m.

Experiment 3 results: Leave-jaw-out

For the comparison of the left and right jaws of the same tool, when training on one jaw and testing on the other, the overall average RMSE in angle was 5.75 degrees and in torque was 8.73 mN m. This is compared with an average RMSE of 0.13 degrees and 0.96 mN m for each jaw trained on itself. The jaw comparison is shown in Fig. 4.

Experiment 4 results: Leave-tool-out

For the comparison of two different tools, when training on one tool and testing on the other, the overall average RMSE in angle was 5.05 degrees and in torque was 7.67 mN m. This is compared with an average RMSE of 0.14 degrees and 1.01 mN m for each tool trained on itself. The tool comparison is shown in Fig. 5.

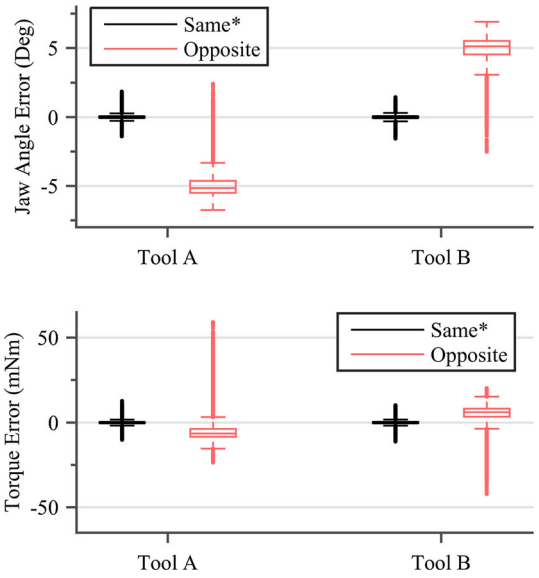

Fig. 5 Experiment 4: Leave-tool-out cross-validation results. For each tool $N = 134{,}749$ time samples on average. An asterisk represents experimental conditions considered to be part of the reliability criteria

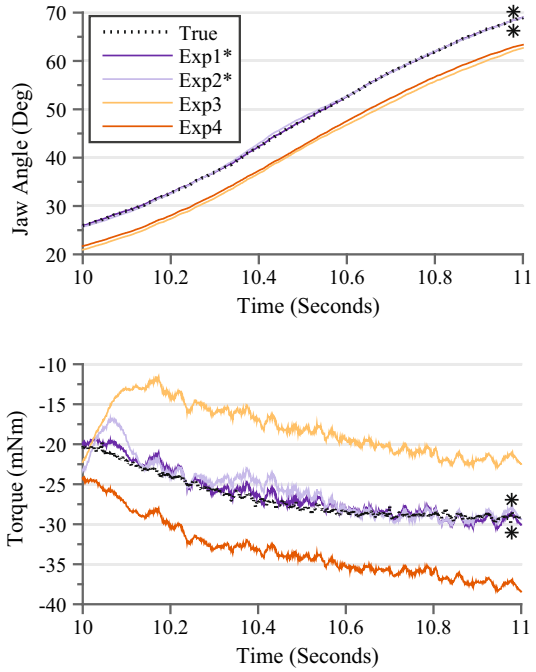

Fig. 6 Sample time series for each experiment taken from the 0.3 Hz dataset over a 1-s duration. The legend applies to both plots and is clarified as follows: Exp1 = Baseline, Exp2 = Leave-frequency-out interpolation, Exp3 = Leave-jaw-out opposite, and Exp4 = Leave-tool-out opposite. An asterisk represents conditions which met the reliability criteria

Time-series results

For visual comparison between the four experiments, a one-second sample from the time-series data was extracted from the 0.3 Hz data and run through each of the experiments. This is shown in Fig. 6, where the front-end sensor values (true value) are compared with the estimates from four different neural nets, corresponding to the four experiments: Experiment 1: Baseline, Experiment 2: Leave-frequency-out interpolation, Experiment 3: Leave-jaw-out opposite, and Experiment 4: Leave-tool-out opposite.

Reliability criteria results

The post hoc analysis of determining the separation line between RMSE values which fall in and out of the reliability criteria is shown in Fig. 7. The separation line for the reliability criteria was conservatively established at 4 mN m and 1 degree. The experimental conditions which met the reliability criteria include Experiment 1: Baseline, Experiment 2: Leave-frequency-out interpolation, Experiment 3: Leave-jaw-out same, and Experiment 4: Leave-tool-out same. The aggregated results comparing conditions which fell within and outside our criteria are shown in Fig. 8.

Case study results

The torque–displacement curves from all three measurement methods are shown in Fig. 9. The direct jaw measurement values stayed within the 95% confidence interval for both

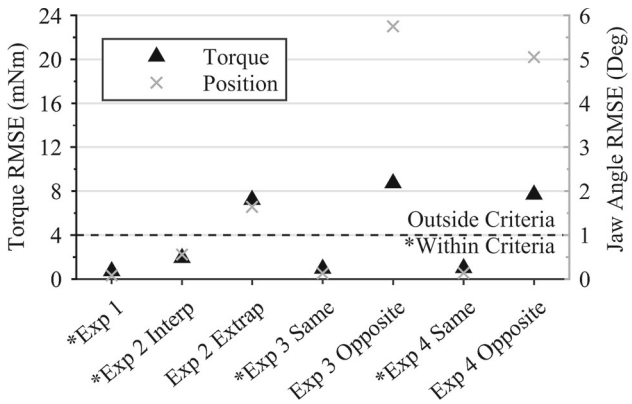

Fig. 7 Post hoc analysis to find separation in RMSE values for reliability criteria. The experimental conditions which fall below the line are considered within the reliability criteria and have been marked with an asterisk throughout the paper

synthetic tissues. The grasp starts at the upper right, at zero torque and zero angle, and proceeds leftward and downward.

Discussion

The baseline experiment (Experiment 1) compared several different regression methods for the estimation of grip force and jaw angle. No single method performed best for both

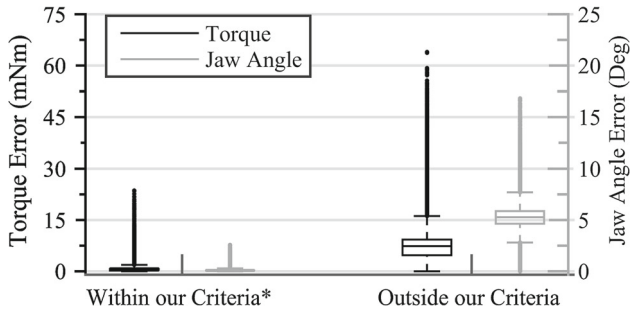

Fig. 8 Comparison of aggregated errors with and without our reliability criteria. The asterisk marks conditions within our reliability criteria which consists of Experiment 1: Baseline, Experiment 2: Leave-frequency-out interpolation, Experiment 3: Leave-jaw-out same, and Experiment 4: Leave-tool-out same for a total of 598 test grasps resulting in $N = 757{,}997$ time samples. For the data outside our reliability criteria, there were 472 test grasps resulting in $N = 589{,}153$ time samples

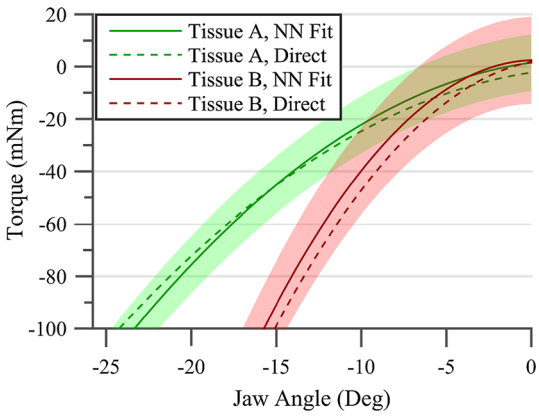

Fig. 9 Case study results for synthetic tissues with 95% confidence interval band around neural net fits with $N = 30$ grasps. Initial contact is at zero degrees with decreasing jaw angle representing a closing grasp

jaw angle and torque estimation. In regard to torque estimation, which is the more difficult prospect, the neural networks performed well, but did not produce the lowest RMSE on average. The SVM methods all performed worse than neural networks irregardless of the kernel selected. The ensemble methods utilizing decision trees (bagged trees and random forests) as well as decision trees alone resulted in lower RMSE, and it may appear that they are the superior method. However, a closer inspection of the results indicates that these methods involving decision trees all performed very well at the low frequencies (0.1–0.2 Hz), but saw significant drop-off at the higher frequencies (0.4–0.5 Hz). Considering that a main contribution of this work was to investigate the effect of variable training frequencies (Experiment 2), we decided to proceed with neural networks as the chosen estimation technique, as they performed more consistently, albeit at a slightly higher RMSE. Additionally, our previous work [15] utilized a similar neural network architecture which is useful for comparison.

Utilizing the reliability criteria, the neural network accurately estimated grip force (1.04 mN m RMSE) and jaw angle (0.17 degree RMSE) as depicted in Fig. 8. The criteria we proposed for a reliable estimate include testing on the same tool, the same jaw, and at a frequency which falls within the range of training data. The separation of RMSE values as detailed in Fig. 7 helped determine these criteria. A more thorough discussion of the implications from each experimental training condition is included sequentially as part of this discussion.

As shown in Experiment 1, the estimation with neural networks worked very well when trained and tested at specific frequency levels (RMSE: 0.08 degrees and 0.71 mN m). As expected, this situation represents the best-case scenario: train the neural network at a given frequency and operate the grasper at the same frequency. Our grip force error results compare favorably to previously reported data-centric approaches. For comparison, the results from [19], which utilized GPR, yield an average error of 0.07 N for grasps with a peak grasping force of roughly 1 N. Our approach, when converted into force units, resulted in an RMSE of 0.07 N for grasps with a peak force of approximately 5.5 N. Our estimation errors were equivalent to the GPR approach even though the nominal grasping forces used in our experimentation were higher. Although it is not always possible to restrict grasping in practice to a specific frequency, there are applications where this is feasible. One example is in measuring torque–displacement curves from in-vivo or in-situ tissues. Experimentation could be conducted where grasping is performed on these tissues at a programmed frequency which is consistent with the frequency of the training data.

Experiment 2 results provide insight on the efficacy of this estimation in cases where it may not be possible to restrict grasping frequencies to a single frequency. The leave-frequency-out analysis showed that grasping at frequencies which were interpolated within the training data was feasible (RMSE: 0.56 degrees and 1.92 mN m). However, grasping at frequencies which were extrapolated outside the training data resulted in poorer estimation (RMSE: 1.64 degrees and 7.21 mN m). This establishes bounds on the operating ranges which grasping can most accurately be performed; the bounds are the lowest and highest frequencies present in the training dataset. Automated surgical subtasks could practically be implemented over a known frequency range and benefit from this estimation by training on data encapsulating the operating frequencies. Additionally, this estimate could be useful in haptic implementation, by either limiting the frequency of grasping or by only providing estimates when grasping is conducted within the bounds of the training dataset.

Experiment 3 results highlighted the inaccuracies in using the opposite jaw's model for estimation (RMSE: 5.63 degrees and 8.73 mN m). This suggests that it is important to train sep-

arate neural networks for each jaw if both jaw estimates are needed. These data strongly suggest avoiding assumptions of symmetry between jaws, as the variance between jaws for a given tool is even larger than the magnitude of variance between two tools (compare Figs. 4, 5).

Experiment 4 results illuminate some possible limitations worth considering when applying this estimation technique. The leave-tool-out experiment resulted in higher-than-normal estimation errors (RMSE: 5.05 degrees and 7.67 mN m). This becomes pertinent for applications which perform tasks across multiple tools (e.g., performing consecutive surgeries with different tools or switching tools during surgery). There are a few potential solutions to rectify this estimation approach even if multiple tools are required.

The first potential solution is to retrain a tool at least one time before it is ever used. The training process could feasibly be accomplished with the tool attached to the da Vinci robot with the aid of a portable calibration device, such as the distal sensing unit used in this experimentation and explained in [14]. The training process is non-trivial, but achievable in the workflow of surgical setup, as the total training time for this experimentation was less than 7 min for non-optimized implementation. A second alternative is unverified, but we hypothesize that a single, representative tool model may be used in conjunction with a minor calibration routine specific to each tool to account for offsets in the estimation. This minor calibration would be far less extensive than retraining an entire neural network. As seen from the results of Experiment 4, the estimation error in the leave-tool-out study is non-mean, yet the errors are still Gaussian with a comparably small standard deviation. We postulate that there may exist an approximately linear offset that is inherent to each tool which can be subtracted from the estimate to provide an adjustment to real jaw angle and force measurements. For example, this is shown in Fig. 6 where an offset of approximately 5 degrees would provide a significantly better fit for Experiments 3 and 4. This offset may be due to something unavoidable such as manufacturing discrepancies between tools.

The case study was accurate in estimating torque–displacement curves of two synthetic tissue phantoms relying on proximal sensors alone when applying the neural net. The estimate and associated confidence bounds also increasingly separated as the grasper closed. This allows the two synthetic tissues to be differentiated with monotonically increasing confidence for possible applications of tissue identification.

This case study also highlights the potential for obtaining in vivo or in situ tissue measurements. Equipping surgical robotic tools with the capability to accurately estimate forces at the tool–tissue interface could potentially provide a wealth of in-vivo or in-situ measurements, since surgical robots interact with patients daily. The mechanical properties of healthy and diseased tissue could additionally be collected and characterized to improve surgical simulators for more effective and realistic training.

Future work pertaining to this research falls into two main categories: (1) improving and verifying extensibility of the estimation technique and (2) testing applications and other case studies with the estimation in place. For the first category, our future work consists of training the neural network with commanded current instead of measured torque, training on a wider range of torques and frequencies, as well as modifying the hardware to allow training on combinations of roll, pitch, and yaw angles. The extension to incorporating roll, pitch, and yaw is not trivial as the input spindles are not completely decoupled. This will likely require additional sensors at each spindle and contribute to a larger feature space for training. Additionally, the decision tree-based methods, particularly the ensemble methods of random forests and bagged trees tested in experiment 1, merit further investigation. As mentioned, these methods exhibited a drop-off in accuracy at higher frequencies, but it is worth determining if they can be used to hone in on a better estimation technique, especially for grip force estimation. An investigation of GPR is also left for future work, as this work only compared our results to existing literature which utilized GPR.

Future work pertaining to the second category includes testing real-time use of the torque estimates for haptic feedback, as well as obtaining torque–displacement curves of tissues in vivo. Future testing on both human and synthetic tissues would enable a tissue measurements database which is valuable for surgical simulator design. For all future work, the reliability criteria separation line set at 4 mN m and 1 degree will allow for a benchmark for comparison.

Conclusion

In summary, the estimation of grip force and jaw angle using a neural network is reliably accomplished given the following conditions: train and test on each jaw, on the same tool, and at a frequency that falls within the bounds of the training data. These conditions resulted in 1.04 mN m and 0.17 degree RMSE for grip force and jaw angle, respectively. Future studies toward an accurate grip force measurement for the da Vinci robot grasper may provide a new dimension for understanding, modeling, and augmenting tissue–tool interactions during surgery and surgical training.

Funding Research was sponsored in part by the Army Research Laboratory and was accomplished under Cooperative Agreement No. W911NF-14-2-0035. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the US Government. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. Additionally, this material is based upon work supported in part by the National Science

Foundation Graduate Research Fellowship under Grant No. 00039202. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical standard This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent This article does not contain patient data.

References

1. Anooshahpour F, Polushin IG, Patel RV (2014) Quasi-static modeling of the da Vinci instrument. In: International conference on intelligent robots and systems (IROS). IEEE, pp 1308–1313
2. Barocas V, Tenorio LM, Devine K, Lee J, Sweet R, Norfleet J, Mazzeo M (2016) A paradigm for materials design for surgical simulators, with specific application to the pleura and needle decompression. J Med Dev 10(3):030934
3. Basdogan C, De S, Kim J, Muniyandi M, Kim H, Srinivasan MA (2004) Haptics in minimally invasive surgical simulation and training. IEEE Comput Graph Appl 24(2):56–64
4. De S, Rosen J, Dagan A, Hannaford B, Swanson P, Sinanan M (2007) Assessment of tissue damage due to mechanical stresses. Int J Robot Res 26(11–12):1159–1171
5. Dockter R, O'Neill J, Stephens T, Kowalewski T (2016) Feasibility of tissue classification via da Vinci endowrist surgical tool. In: Hamlyn symposium on medical robotics, pp 64–65
6. Egorov V, Tsyuryupa S, Kanilo S, Kogit M, Sarvazyan A (2008) Soft tissue elastometer. Med Eng Phys 30(2):206–212
7. Foresee FD, Hagan MT (1997) Gauss–Newton approximation to Bayesian learning. IEEE Int Conf Neural Netw 3:1930–1935
8. Hecht-Nielsen R (1987) Kolmogorov's mapping neural network existence theorem. In: Proceedings of the international conference on neural networks. IEEE Press, pp 11–14
9. Holoborodko P (2008) Smooth noise robust differentiators. http://www.holoborodko.com/pavel/numerical-methods/numerical-derivative/smooth-low-noise-differentiators/
10. Jones D, Jaffer A, Nodeh AA, Biyani CS, Culmer P (2018) Analysis of mechanical forces used during laparoscopic training procedures. J Endourol. https://doi.org/10.1089/end.2017.0894
11. Kehoe B, Kahn G, Mahler J, Kim J, Lee A, Lee A, Nakagawa K, Patil S, Boyd WD, Abbeel P, Goldberg K (2014) Autonomous multilateral debridement with the raven surgical robot. In: International conference on robotics and automation (ICRA). IEEE, pp 1432–1439
12. Kim U, Lee DH, Yoon WJ, Hannaford B, Choi HR (2015) Force sensor integrated surgical forceps for minimally invasive robotic surgery. IEEE Trans Robot 31(5):1214–1224
13. Kim U, Kim YB, Seok DY, So J, Choi HR (2016) A new type of surgical forceps integrated with three-axial force sensor for minimally invasive robotic surgery. In: 2016 13th international conference on ubiquitous robots and ambient intelligence (URAI). IEEE, pp 135–137
14. Kong NJ, Stephens TK, ONeill JJ, Kowalewski TM (2017) Design of a portable dynamic calibration instrument for da Vinci Si tools. In: 2017 design of medical devices conference. American Society of Mechanical Engineers, pp V001T11A023–V001T11A023
15. Kong NJ, Stephens TK, Kowalewski TM (2018) Da Vinci tool torque mapping over 50,000 grasps and its implications on grip force estimation accuracy. In: 2018 international symposium on medical robotics (ISMR). IEEE, pp 1–6
16. Kowalewski TM, Sweet R, Lendvay TS, Menhadji A, Averch T, Box G, Brand T, Ferrandino M, Kaouk J, Knudsen B, Landman J, Lee B, Schwartz B, McDougall E (2016) Validation of the aua blus tasks. J Urol 195(4):998–1005
17. Lanfranco AR, Castellanos AE, Desai JP, Meyers WC (2004) Robotic surgery: a current perspective. Ann Surg 239(1):14–21
18. Lee C, Park YH, Yoon C, Noh S, Lee C, Kim Y, Kim HC, Kim HH, Kim S (2015) A grip force model for the da vinci end-effector to predict a compensation force. Med Biol Eng Comput 53(3):253–261
19. Li Y, Hannaford B (2017) Gaussian process regression for sensorless grip force estimation of cable-driven elongated surgical instruments. IEEE Robot Autom Lett 2(3):1312–1319
20. Li Y, Miyasaka M, Haghighipanah M, Cheng L, Hannaford B (2016) Dynamic modeling of cable driven elongated surgical instruments for sensorless grip force estimation. In: International conference on robotics and automation (ICRA). IEEE, pp 4128–4134
21. MacKay DJ (1992) A practical Bayesian framework for backpropagation networks. Neural Comput 4(3):448–472
22. Mucksavage P, Kerbl DC, Pick DL, Lee JY, McDougall EM, Louie MK (2011) Differences in grip forces among various robotic instruments and da Vinci surgical platforms. J Endourol 25(3):523–528
23. Nakai A, Kuwana K, Saito K, Dohi T, Kumagai A, Shimoyama I (2017) Mems 6-axis force-torque sensor attached to the tip of grasping forceps for identification of tumor in thoracoscopic surgery. In: 2017 IEEE 30th international conference on micro electro mechanical systems (MEMS). IEEE, pp 546–548
24. Puangmali P, Althoefer K, Seneviratne LD, Murphy D, Dasgupta P (2008) State-of-the-art in force and tactile sensing for minimally invasive surgery. IEEE Sens J 8(4):371–381
25. Shademan A, Decker RS, Opfermann JD, Leonard S, Krieger A, Kim PC (2016) Supervised autonomous robotic soft tissue surgery. Sci Transl Med 8(337):337ra64
26. Stephens T, Kong N, Dockter R, O'Neill J, Sweet R, Kowalewski T (2018) Blended shared control utilizing online identification: regulating grasping forces of a surrogate surgical grasper. Int J Comput Assist Radiol Surg 13(6):769–776
27. Stephens TK, Meier ZC, Sweet RM, Kowalewski TM (2015) Tissue identification through back end sensing on da Vinci endowrist surgical tool. J Med Dev Tech Brief 9(3):030939
28. Yamamoto T, Vagvolgyi B, Balaji K, Whitcomb LL, Okamura AM (2009) Tissue property estimation and graphical display for teleoperated robot-assisted surgery. In: International conference on robotics and automation (ICRA). IEEE, pp 4239–4245