# Skill-based Human-robot Cooperation in Tele-operated Path Tracking

Nima Enayati, Giancarlo Ferrigno, Elena De Momi

**Abstract** This work proposes a shared-control tele-operation framework that adapts its cooperative properties based on estimated skill level of the operator in performing path tracking. It is hypothesized that different aspects of an operator's performance in executing a tele-operated path tracking task can be assessed through conventional machine learning methods using motion-based and task-related features. To identify performance measures that capture motor skills linked to the studied task, an experiment is conducted where users new to tele-operation, practice towards motor skill proficiency in 7 training sessions. A set of classifiers are then learned from the acquired data and selected features, which can generate a skill profile that comprises estimations of user's various competences. Skill profiles are exploited to modify the behavior of the assistive robotic system accordingly with the objective of enhancing user experience by preventing unnecessary restriction for skilled users. A second experiment is implemented in which novice and expert users execute the path tracking on different pathways while being assisted by the robot according to their estimated skill profiles. Results validate the skill estimation method and hint at feasibility of shared-control customization in tele-operated path tracking.

Keywords **Shared-control, active constraints, virtual fixtures, tele-operation, machine learning, surgery**

## 1 Introduction

Human-robot cooperation is becoming progressively prevalent in applications such as space (Penin, Matsumoto, and Wakabayashi n.d.), search and rescue (Enayati and Najafi 2011), and robotic surgery (Gomes 2011), where automation has not been widely realized due to technological or ethical challenges. With the gradual growth of interest in granting more control to robots (Kranzfelder et al. 2013), more efforts are being expanded on the design of collaborative agents that provide active assistance to operators. Assistance through the haptic channel has been broadly investigated for tele-operation despite remaining challenges (Enayati, De Momi, and Ferrigno 2016). The key concept is to provide haptic cues (e.g. force, torque or vibrations) to the operator to guide (or restrict) the motion of the tool. These methods have shown enhancement in terms of safety and performance in experimental tele-operation setups (S. Bowyer, Davies, and Baena 2014). However, unlike visual or auditory assistance, the guidance/preventive wrenches applied to user's hand explicitly affect motor control and can degrade the transparency and user experience of the tele-operative system (Lawrence 1993) (Enayati et al. 2016). Furthermore, the dissipated energy by the haptic assistive methods can contribute to fatigue in operations lasting several hours. Modulating the intensity of haptic assistance according to the operator's performance and preventing restrictive assistance when not needed can potentially enhance the user-agent interaction.

Haptic assistance can also be exploited in motor skill training and improve the learning curve of motor tasks through the reduction of cognitive load (Marchal-Crespo and Reinkensmeyer 2008). For instance, in training of surgical tasks with complex kinematics, where it has been shown that residents can face difficulties due to poor hand-eye coordination (Finan, Clark, and Rocconi 2010), some form of haptic guidance can speed up the initial motor learning phase. However, there are concerns about creating assistance dependency in trainees which can decrease performance during the actual task (Sigrist et al. 2013). An adaptive assistance can address such concerns by monitoring the skill level of the trainee and gradually lowering the intensity and frequency of the assistance to reduce the risk of creating dependency (Schmidt 1991). Skill assessment can also be employed in non-assisted robotics training programs (Schreuder et al. 2012), where the trainees can be classified through a placement test and receive a training program that is dynamically tailored to their skill profile.

This work discusses a cooperative tele-operation architecture, which adapts its assistive characteristics according to user's skills with the purpose of preventing excessive restriction and improving subjective performance of the system. To study the feasibility of such an approach, a tele-operated path tracking task is chosen for experimentation. The task demands precise motor control, depth perception, hand-eye coordination, and sensory substitution. A set of features are proposed that encompass various aspects of user performance for the studied task, and do not require additional measurements with respect to those available in an active-constraint enabled tele-operation setup. An assembly of classifiers are learned through categorization of the features allowing for a targeted adaptation of the Assistive Method (AM). A first set of experiments is performed on 7 novice users undergoing one training session per day for 7 days to study the learning curve of the task. The results of these experiments are used to select features that better demonstrate the learning

Nima Enayati, Giancarlo Ferrigno, and Elena De Momi are with the department of electronics, information and bioengineering, Politecnico di Milano, 20133, Milan, Italy. {nima.enayati, giancarlo.ferrigno, elena.demomi }@polimi.it.

curve and also to train the classifiers. A second set of experiments with 18 novice and skilled users is performed to validate the skill evaluation method and to demonstrate the effects of the adaptive assistance.

The remainder of this paper is organized as follows. Related work is presented in Sec. II. The methods and the system setup are introduced in Sec. III. In Sec. IV, the results of the experiments are presented and the outcomes are discussed. Finally, conclusions and future work are described in Sec. V.

## 2 Related work

Human-robot cooperation has been employed in several applications such as semi-autonomous cars (Anderson, Karumanchi, and Iagnemma 2012), mobile robots in search and rescue (Gao et al. 2014), and unmanned aerial vehicles (Franchi et al. 2012). In the field of surgical robotics alone, various human-machine collaborative approaches have been proposed. Padoy *et al.* proposed a combination of manual and automatic execution, where portions of a surgical task that do not include interaction with the tissue are performed autonomously, and other portions manually by the surgeon (Padoy and Hager 2011). In (Rafii-Tari et al. 2013), novice operators are assisted in a catheterization procedure by a robotic catheter driver that replicates motion models generated from skilled operators manipulation. A surgical robot was used in (Bauzano, Estebanez, and Muñoz 2014) to assist suturing procedure by holding the needle or stretching the thread in hand-assisted laparoscopic surgery. Some works have focused on designing general frameworks to incorporate different collaborative modalities in surgical robotic workflow (Berthet-rayne et al. 2016; Nichols et al. 2015). Assistance via kinesthetic feedback is among the most widely studied methods and have shown enhancement of performance (Abbott, Marayong, and Okamura 2007; Enayati et al. 2016) and safety (S. A. Bowyer and Baena 2015) in experimental tele-operation setups.

Human-robot cooperation often presents a trade-off between increasing task performance and guaranteeing optimal user experience. It has been observed that in some cases even when robotic assistance improves the performance, some users still prefer feeling being in control of the robot (Dragan and Srinivasa 2013). A user study conducted in (Takayama et al. 2011) found that assisted teleoperation of a mobile remote presence system resulted in superior obstacle avoidance, but concurrently increased the completion time of the task, hinting at an augmented subjective execution cost. To alleviate this, some efforts have been expanded on arbitration between robot's assistance and user's control (Milliken and Hollinger 2016; Takayama et al. 2011). In (Takayama et al. 2011) a 2D simulation was implemented where the user controlled the positions of a mobile robot modeled as a circle to reach a destination while avoiding rectangular obstacles. A keyboard was used as the input device and robotic assistance was in form of altering user inputs to achieve better performance based on user skill simulated as a range from novice to expert. We aim at implementing a physical robotic assistance (e.g. through haptics) for tele-operation that is adapted to user skills considering multiple performance aspects, which to our best knowledge has not been investigated. It is assumed that as the complexity of the motor task increases, different users could demonstrate a variability in different competences that can be better captured by a skill profile rather than a single expertise measure.

Again, in surgical robotics skill assessment has been broadly investigated. The premise is that of replacing subjective human-based evaluation of trainees by a more quantitative analysis (Van Hove et al. 2010). In numerous works, gesture segmentation algorithms have been developed that could be exploited for skill estimation as a distance-based method, where a user's performance is evaluated according to its model's statistical distance from that of an expert. Lin *et al.* developed a surgical gesture segmentation method from kinematics measurement of a da Vinci robot (Intuitive Surgical, USA) using linear discriminant analysis (Lin et al. 2006). In (Fard et al. 2017) the temporal sequence of surgical gestures is segmented in an unsupervised manner and gradual transitions between them is modelled using fuzzy membership scores. Inspired by the success of Markov Model (MM) and its constituent Hidden Markov Model (HMM) in fields such as speech recognition, considerable number of works have employed these methods to classify user skills in laparoscopic procedures (Rosen et al. 2006), virtual simulators (Megali et al. 2006) and tele-operated robotic systems (Reiley and Hager 2009). Skill level is generally incorporated into the MM by developing different MMs based on data acquired from users with different levels of expertise (Rosen et al. 2006). Recently a public benchmark surgical activity dataset captured from the da Vinci robot has been released that contains synchronized video and kinematic data from three standard surgical tasks performed by subjects with different skill levels (Ahmidi et al. 2017). The authors also present the results of six gesture recognition and segmentation techniques applied on the dataset. These approaches have provided the community with a broader understanding of surgical procedures by exposing the internal hierarchy of states encompassed in various surgical tasks. However, for targeted customization of a collaborative control scheme, a skill estimation method with interpretable outcomes is essential. In other words, while a statistical distance from expert performance can be helpful in modulating the overall magnitude of assistance, to adapt multiple properties of assistance, the skill estimation result must involve descriptive performance metrics. Such metrics play a critical role in determining the success of skill evaluation methods (Kim et al. 2010). Inspired by the way expert users instruct and comment upon the performance of novices, some quantitative performance metrics have been defined in the literature that are useful across a large variety of tasks (Cotin et al. 2002). Common objective descriptive performance measures in teleoperation include completion time (Rosen et al. 2006), (Chmarra et al. 2010), interaction force/torque (Trejos et al. 2013), (Richards et al. 2000), path length (Chmarra et al. 2010), (Rosen et al. 2006), and motion smoothness (Cotin et al. 2002)(Chmarra et al. 2010). Such metrics have been employed to classify the skills of the users as novice, intermediate and expert for a laparoscopic surgery simulator using a fuzzy classifier in (Hajshirmohammadi and Payandeh 2007). Chmarra *et al.* trained a linear discriminant analysis classifier to classify the performance of users in a box trainer for laparoscopic surgery (Chmarra et al. 2010).

The study of performance in motor tasks can highly benefit from an understanding of the neuroscience of motor learning. Motor adaptation, where the effects of perturbation on motor performance is studied, has received a particular attention in neuroscience literature (Ghez, Krakauer, and Ghilardi 1999; Sainburg and Wang 2002; Shadmehr, Smith, and Krakauer 2010). Although arguably less extensively studied, the improvement of motor task itself—where the principal performance objective is to reduce variable error in the absence of perturbation—has been investigated in works such as (Liu, Mayer-Kress, and Newell 2006) and (Shmuelof, Krakauer, and Mazzoni 2012). No universally accepted definition of motor skill learning metrics can be found in the literature, but the features proposed to capture the concepts of motor skill learning share an emphasis on speed, accuracy, and overall efficiency. In (Shmuelof, Krakauer, and Mazzoni 2012) for example, a 2D path tracking motor skill task was designed and skill learning at the task level was defined as a change in the speed-accuracy trade-off function. Results suggest that path tracking skill acquisition was associated with a large reduction in trajectory variability. This study does not intend to provide an in-depth study of motor learning and its neurophysiological aspects. We however investigate the features that can capture the motor skill difference of novice and expert users in a multiple-day tele-operated path tracking experiment. These performance metrics are selected to uncover the underlying characteristics of skills needed to execute a tele-operated path tracking task.

The contributions of this work can be summarized as the study of a set of performance measures that demonstrate the learning curve of tele-operated path tracking and the implementation of a cooperative tele-operation architecture with assistive properties that are adapted to the user's estimated skill profile.

## 3 Methods

Following the development order, the experimental setup and the task, defined metrics, designed AMs and the implemented machine learning algorithms are described in this section.

### 3.1 Tele-operation setup

The hardware of the tele-operation setup (Fig. 1) comprised an LWR4+ robot as the slave, a Sigma 7 haptic device (Force Dimension, Switzerland) as the master, a foot-switch, a 720p webcam, a desktop computer with a 3.6 GHz Core i7-4790 CPU (Intel, Corp.) running a real-time patched Linux kernel (Kernel 3.5.7 patched with Xenomai 2.6.2.) for the control loop and a laptop with a 2.5 GHz Core i7-6500 CPU running a generic Linux kernel (3.19.0) for machine learning and vision applications.

The software architecture of the setup, depicted in Fig. 2, was developed as a modular platform. The control modules were written in C++ as real-time Open RObot COntrol Software (OROCOS) (Bruyninckx 2001) components and the vision and classifier modules were developed as Robot Operating System (ROS) nodes written in C++ and Python respectively. The OROCOS components ran periodically at 500Hz in a synchronized mode triggered by new measurements from the slave. The vision node ran at approximately 25 Hz that was the frame-per-second rate of images captured by the webcam. Communication between the ROS nodes and OROCOS components was on a local DHCP-managed network. Master and slave were coupled as long as the foot-switch was pressed. Master's motion was downscaled by a factor of 3.

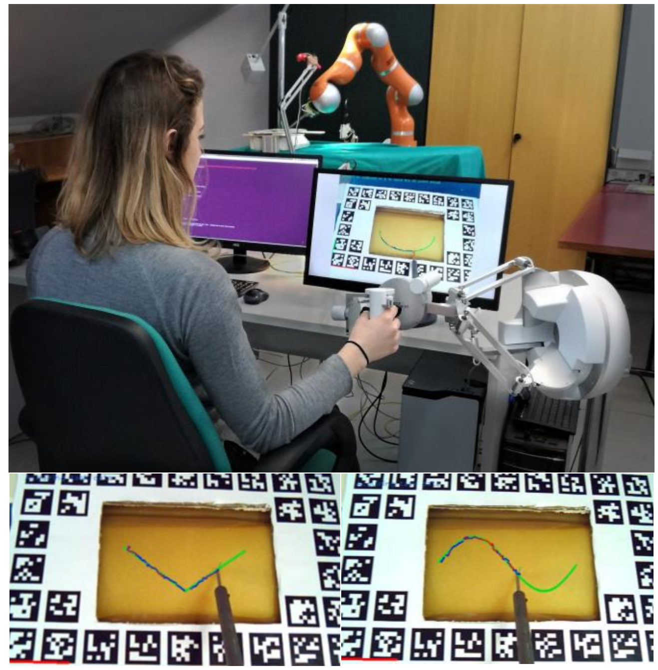

**Fig. 1** The tele-operation setup comprised an LWR4+ robot and a Sigma 7 master device. To perform the task the user placed the tool on a silicone phantom and followed the path overlaid on the camera view shown on the display. The bottom images show the paths used in the test experiments (left: task 1, right: task 2).

The LWR4+ kinematics controller component performed main control operations including inverse and forward kinematics, redundancy resolution, trajectory generation, and reference frame transformations. The AC (Active Constraint) geometry and AC force generator components, when activated, generated the assistive force for the haptic device based on the user profile and data received from the other components. The task performance evaluation component computed the performance metrics during acquisitions and wrote the results to a file when acquisitions ended. These result files were then used either off-line for the training of the classifiers, or online by the user classifier component to create the user profile. Finally the vision node displayed the camera images, messages, reference path, and drawn paths on the display, and estimated the pose of the camera with respect to the task frame by detecting a set of ArUco markers (Garrido-Jurado et al. 2014) attached to the task phantom. The transformation from the task reference frame to the slave's base frame was estimated in

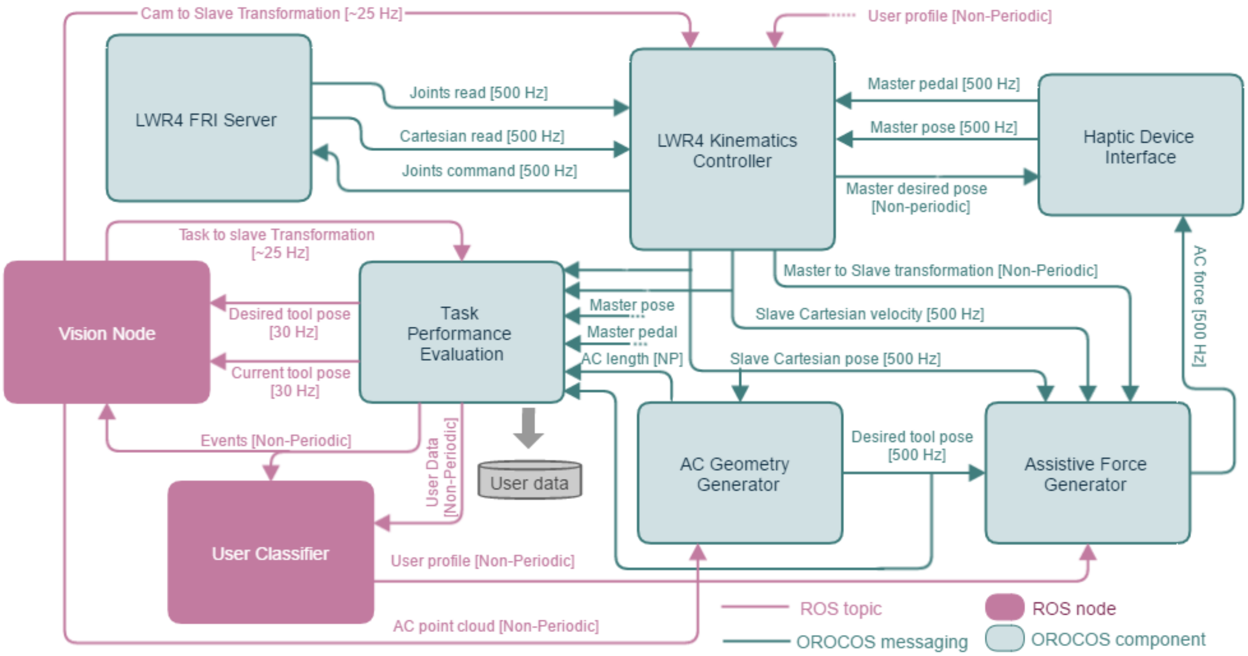

**Fig.2** Software components of the tele-operation setup

advance through a calibration process.

### 3.2 The Task and Experiments

The task consisted of Cartesian path tracking. The reference path and user's trajectory were overlaid on the camera image. To provide a cue for the user to better percept the depth through the 2D display, the color of the current point of the trajectory transitioned from blue to red relative to the depth of penetration into the phantom.

Since the developed tele-operation setup is an experimental system, to investigate the performance of expert users, such users had to be trained. An initial 5-day, 7-subject training program was scheduled in which each user exercised with the system by executing a single task 12 times per session. To ensure that all subjects reached their highest performance, the training was later extended to 7 days as the subjects still showed signs of progress. The training sessions took 25.3 minutes in average. The subjects of this "training" experiment were 6 males, 1 female, all engineering students, right-handed, with the average age of 24.2 (SD = 4.4) and with little/no experience with the setup. The users were considered skilled at the 7$^{th}$ day and the acquisitions of that last day and the first day were labeled as skilled and novice respectively and used in the learning of the classifiers.

A test experiment was performed that comprised two tasks with path shapes different from that of the training experiment (Fig. 1) to evaluate the skill assessment method and study the outcomes of the assistance. Participants were 6 skilled users from the training experiment and 12 new novice users (5 females and 7 males, right-handed, average age 26.5, SD age = 1.8). The novices were divided in two groups of 6. The novice-1 group performed all the repetitions with no assistance to study intra-session learning effects. The novice-2 group and the skilled group, for each path, first performed the task 3 times with no assistance. At the end of each repetition, the skill level of the user was estimated. The average of the skill probabilities estimated for the three repetitions was then used to customize the AMs and the users performed 3 repetitions with assistance. To summarize, all the three groups (novice-1, novice-2 and skilled) performed a task three times with no assistance, followed by three more repetitions that were assisted for two groups (novice-2 and skilled) and non-assisted for the other group (novice-1). The experiments were carried out in accordance with the recommendations of our institution with written informed consent from the subjects in accordance with the declaration of Helsinki.

### 3.3 Performance measures

A set of performance measures encompassing task outcome and competence in using the tele-operation setup was defined as follows. These metrics were analyzed in the training experiment to investigate their possible correlation with skill level as the users gained more experience.

- Root-Mean-Square Error (RMSE) ε: a frequently used measure of tracking error calculated for $n$ samples as:

$$\varepsilon = \sqrt{\sum_{k=1}^{n} e_k^2 \,/\, n} \tag{1}$$

The tracking error $e_k$ at time sample k is defined as the Euclidean distance between the position of the tool-tip $\boldsymbol{p}_k^{tt}$ and the closest point on reference path $\boldsymbol{p}_k^{cpr}$:

$$e_k = \left\| \boldsymbol{p}_k^{tt} - \boldsymbol{p}_k^{cpr} \right\| \tag{2}$$

- Maximum Error E: While representing accuracy, maximum error can give some insight on user's interaction with the device and the continuity of his/her motion.

- Average tracking velocity $\nu$: Higher experience and dexterity may lead to a more determined and faster tracking in average.
- Maximum tracking velocity $V$: While higher mean tracking velocity may imply higher experience, peaks in the tracking velocity profile can be due to discontinuous motion often associated with novice users.
- Number of tracking segments $\lambda$: The number of segments that takes a user to finish a tracking task can show the dexterity of the user in performing the task. This was calculated as the number of times the tool lost and regained contact with the phantom.
- Number of clutch engagements $\kappa$: Due to the limited workspace of the master device, it is common to relate the position of the master device to that of the slave in an incremental manner, through a foot switch. The user may need to use the foot switch often if the motion of the master is downscaled and it is suspected that novice users tend to disengage the foot-switch less frequently.
- Master workspace (WS) limit metric $\omega$: Despite the simplicity of the idea of clutching in tele-operation, in our previous works we observed that novice users tend to reach the limits of the master device's workspace frequently, which is often followed by confusion and reduction of performance. This is of particular importance when the WS of the master device is relatively small (often the case for devices with parallel mechanisms) and when the slave tracks downscaled motion of the master. To investigate this behavior, a measure was defined as:

$$\omega = \frac{t_{out}}{T} \tag{3}$$

where $t_{out}$ is the sum of the time durations spent close to the workspace limits and $T$ is the total task duration. The workspace of the master device was approximated as a semi-sphere shown in Fig. 3 and $t_{out}$ was incremented when the end-effector of the master device was placed out of this semi-sphere.
- Total Master Displacement $\delta$: Overall path length of tool tip is a commonly used performance metric in tele-operation. Here we consider the overall displacement of the master device as a possible measure of user's motion efficiency, calculated as the sums of distances between consecutive positions of the master's end effector $\boldsymbol{p}_k^m$:

$$\delta = \sum_{k=2}^{n} \|\boldsymbol{p}_k^m - \boldsymbol{p}_{k-1}^m\| \tag{4}$$

- Task Duration $\tau$: The time a user spends in executing a task is a popular measure often used as an indication of performance.

The first five measures are outcome measures related to the studied task of this work, the last four can be used in studies involving different tasks as three of them estimate the efficiency with which the operator uses the tele-operation setup and one (task duration) that can be used as a general performance metric for most tasks.

### 3.4 Assistive methods

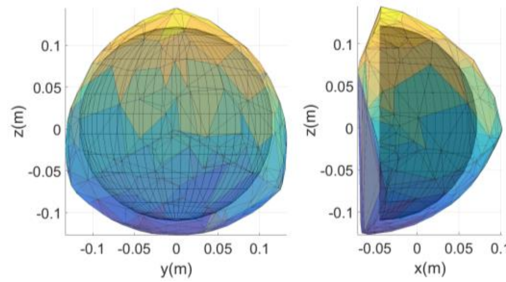


**Fig. 3** The workspace of the Sigma7 device was estimated by finding a 3D convex hull of a set of measured points. To calculate the master WS limit metric ω, a semi sphere was used to approximate the WS.

Four AMs were deployed, three of which attempted to explicitly improve the task performance through kinesthetic guidance, called haptic AMs, and one that helped the operator in interacting with the master device and avoid the workspace limits. These methods are explained in the following.

***AM1:*** *Guidance active constraint*

Guidance ACs aim at guiding user's motion through haptic enforcement. The enforcement method implemented in this work for the guidance AC was the non-energy-storing method described in (Enayati et al. 2016). Such an AC enforcement attempts to redirect the motion of the user toward the AC path (the desired path) through generated force applied to user's hand and unlike energy-storing methods (such as elastic enforcement), guarantees that no force is applied when the user is not moving the tool.

***AM2:*** *Isotropic viscosity*

Increasing the stiffness/viscosity of the controller has demonstrated potential to improve performance in surgical tasks by diminishing high-frequency component of the user's input (Beretta et al. 2015). The second AM consisted of an isotropic viscous force aiming at improving the continuity of naïve operator's motion by limiting velocity-peaks of the tool's motion often generated unintentionally and leading to large errors.

***AM3:*** *Elastic penetration constraint*

Limited depth perception and lack of haptic feedback are two common factors that can affect the quality of a path tracking task by causing deep penetrations or frequent loss of contact with the tissue that results in a scattered trajectory with too many segments. Having the position of the AC path and assuming that the axis of the tool is quasi-parallel to the viewing axes of the camera (that is often the case in laparoscopic surgery) an assistive force can be generated that gently pushes the tool towards the camera (and therefore away from the tissue) with an intensity proportional to the depth of the penetration passed the AC path along the camera axis. This can be thought of as a simulated and weak haptic feedback. The user then has to maintain a constant force that can simplify the tracking procedure. Note that this AM may not be easily applicable to complicated pathways in a highly deformable environment.

***AM4****: Master Workspace Helper (MWH)*

As it was discussed in subsection III-C when the input from the master device is downscaled, low-skilled users frequently get confused after hitting the workspace's limits and continue moving along the boundary of the WS instead of disengaging the foot-switch and moving the master's end-effector to an appropriate point inside the WS. To address this issue, an AM was implemented that displayed a message on the screen when the user reached the limits. If the user released the foot-switch (i.e. disconnecting the slave from the master) while the message was shown, the master would automatically move its end-effector to the center of the workspace, creating space for the user's maneuver.

The engagement level of these 4 AMs were set as the maximum applied force for haptic methods and on/off for the MWH, according to the estimated skill profile of the user. The maximum force values and other related parameters of haptic methods are reported in table 1, where $B$ and $K$ respectively stand for viscosity and elasticity coefficients.

**Table 1** Parameters Used in the Haptic Assistive Methods

| Guidance AC | | Isotropic viscosity | | Penetration constraint | |
|---|---|---|---|---|---|
| Param. | Value | Param. | Value | Param. | Value |
| $F_{max}$ | 3 N | $F_{max}$ | 1 N | $F_{max}$ | 2 N |
| $B$ | 80 Ns/m | $B$ | 20 Ns/m | $K$ | 400 N/m |
| $D_m^*$ | 0.002 m | | | | |

* Boundary smoothing parameter explained in (Enayati et al. 2016)

### 3.5 Classification and parameter setting

The skill profile comprised five parameters, one of which was an overall binary classification of user's overall skill and the other four described specific competences used for adapting the AMs. The overall skill estimation was used simply as an informative evaluation that could be used for training purposes, and the customization of the AMs was performed based on the specific competence estimations. A Support Vector Machine (SVM) classifier was implemented as the overall skill classifier that produced a binary user skill level as "skilled" or "novice" based on a 7-element feature vector X:

$$\mathrm{X} = \{\varepsilon, \mathrm{E}, V, \omega, \delta, \lambda, \tau\}$$

SVMs provide a versatile approach for classification due to the possibilities of using various kernels. A Radial Basis Function (RBF) kernel was used as it showed slightly a higher accuracy with respect to other kernels. Features $\nu$ and $\kappa$ were not used since they did not show significant correlation with the experience level, as it will be seen in the results section. The goal was to set the parameter $\Phi_j$ of assistive method $j$ based on specific aspects of user's performance $\psi_j$ that can be enhanced by that AM. These specific performance aspects and their corresponding feature set $x_j$ (subsets of the feature vector X) were defined as follows:

| | | |
|---|---|---|
| AM1: | Accuracy/time: | $x_1 = \{\varepsilon, \tau\}$ |
| AM2: | Motion consistency: | $x_2 = \{\mathrm{E}, V\}$ |
| AM3: | Trajectory quality: | $x_3 = \{\lambda\}$ |
| AM4: | Master WS Handling: | $x_4 = \{\omega, \delta\}$ |

The skill profile $\Psi$ for each user was constructed from specific performance aspect $\psi_j$:

$$\Psi = \{\psi_1,\ \psi_2,\ \psi_3,\ \psi_4\}$$

User skill profiles were mapped to AM's parameter space $\Phi$ through a set of mapping functions $M$:

$$M : \Psi \rightarrow \Phi$$

$$\Phi_j = \left[0,\ F_j^{max}\right] \quad j = 1,2,3 \qquad (5)$$
$$\Phi_j = \{0, 1\} \quad j = 4$$

The continuous range of the haptic AMs' ($j = 1,2,3$) parameter set necessitates a continuous performance evaluation that can be formulated as a regression problem, while the MWH's on/off binary parameter set ($j = 4$) leads to a classification problem. However, both problems can be addressed by using the probabilistic output of a classifier that can be easily mapped into the continuous parameters, and passed through a threshold to produce a binary output for setting the parameter of MWH. Thus, assuming label y as 1 for high and 0 for low skill level, the performance evaluation can be seen as the probability of being skilled in a specific performance aspect given the classifiers coefficients $\theta_j$ and feature vector $x_j$:

$$\psi_j = P_j\left(y = 1 \middle| x_j, \theta_j\right) \qquad (6)$$

A linear mapping function was used for the haptic AMs that simply scaled the maximum force intensity of the AM, $F_j^{max}$ by the probability value $P_j$. The binary output for activating/deactivating the MWH was generated by applying a threshold of 0.5:

$$\varphi_j = \left(1 - \psi_j\right) F_j^{max} \quad j = 1,2,3$$
$$\varphi_j = \begin{cases} 1 & \psi_j \leq 0.5 \\ 0 & \psi_j > 0.5 \end{cases} \quad j = 4 \qquad (7)$$

Logistic regression was selected as the classification method which by definition generates a probability function output as:

$$P_j\left(y = 1 \middle| x_j, \theta_j\right) = \frac{1}{1 + e^{-\theta_j^T h(x_j)}} \qquad (8)$$

where h is the hypothesis function. L2 penalty was applied to prevent overfitting. The classifiers were implemented using the Scikit-learn Python module (Pedregosa et al. 2011) and learned based on the training experiment data. In learning of all the 5 classifiers, the data were first standardized to ensure maximum learning efficiency. This is commonly done by removing the mean and scaling to unit variance. However, to prevent the negative influence of outliers, the median and the interquartile range were used to scale the data.

### 3.6 Statistical Analysis

The statistical analysis performed in this work comprise solely checking for statistical significance difference between acquired data samples with the null hypothesis that data samples belong to the same population. The analysis was performed in MATLAB. The Friedman test was used which is a nonparametric version of balanced one-way ANOVA and adjusts for possible intra-sample effects (here being samples acquired from the same user). The multiple comparison function *multcompare* was used to perform a pairwise comparison on the test data generated by the Friedman method. For classifiers, accuracy was first evaluated through cross-validation. To account for the dependencies of samples coming from the same user, the partitioning was done based on samples acquired from the same user resulting in a 7-fold cross validation scheme.

## 4 Results and Discussion

### 4.1 Training

The results of the training experiments, depicted in Fig. 4, reveal a learning curve captured by all the measured metrics except for mean tracking velocity. Improvements can be observed in terms of the medians suggesting that in average the users gained more skills in performing the task. Another considerable enhancement can be seen in terms of the reduction of variability among the samples, hinting at the efficacy of the training in diminishing the dependency of performance on each user's pre-training skills. Since the first and last day samples were planned to be used in learning of the classifiers as novice and skilled data classes respectively, it had to be validated that these sets were statistically different. All the metrics showed significant difference ($p<0.001$) except for the mean tracking velocity and number of clutch engagements, which were thus not used as features in the classification.

**Table 2** Classification accuracy from cross-validation

| Classifier | Overall skill class | Accuracy/ time | Motion consistency | Trajectory quality | WS |
|---|---|---|---|---|---|
| Accuracy % (SD) | 97.2 (0.05) | 94.8 (0.05) | 87.5 (0.06) | 84.5 (0.15) | 94.0 (0.04) |

Reported in table 2, the classification accuracies suggest that the classifiers are able to distinguish individuals with different skills. As expected the specific classifiers are less accurate in predicting the skill levels. Nonetheless, the achieved accuracies are sufficient for the purpose of customization of the assistive methods.

### 4.2 Test

The skill profiles generated for each user of the groups that received assistance (novice-2 and skilled) in both tasks are shown in Fig. 5. Each data point represents the average value calculated from 3 placements acquisitions (non-assisted) of a user. Although the vertical axis indicates the probability of being high-skilled, the binary outputs of the overall skill and WS handling classifiers are also included in the plot so that the complete skill profile of each user can be inspected. To evaluate the accuracy of the overall skill classifier the acquisition of all the 12 novice users (novice-1 and novice2) and the 6 skilled users were used. Out of the 36 classification instances 34 were correctly classified yielding a test accuracy of %94.4. 2 novice users (one from each group) were labeled as skilled in task 1 due to their above-average performance. All novice users were however classified as novices in the second task that proved to be more demanding than task 1. The WS handling classifier identified all skilled users correctly and activated the WS helper for all novice users but 3 in the first task. The other 3 specific classifiers, used to set the gain of the corresponding AMs, generated probabilities that are lower for novice users resulting in higher assistance gains in the subsequent acquisitions. The probabilities are distributed among users such that the entire ranges of the gains were exploited. The yielded skill profiles confirm that certain skills can vary considerably among users. While one user may perform well in one aspect and bad in another, a different user can demonstrate a converse performance profile. The information provided by the skill profiles allow for an interpretable assessment of skills that can be exploited in training programs to either communicate the competences that need more attention form the trainee, or to automatically reprogram exercises to enhance the weaker performance aspects. Considering assisted tele-operation, tailoring the haptic assistive methods to the skill profiles can provide assistance when and to the amount needed and prevent unnecessarily restrictive AMs.

The performance metrics acquired from the test experiments with and without assistance are depicted in Fig. 6. Each bar contains 18 samples acquired from 6 users of each subject

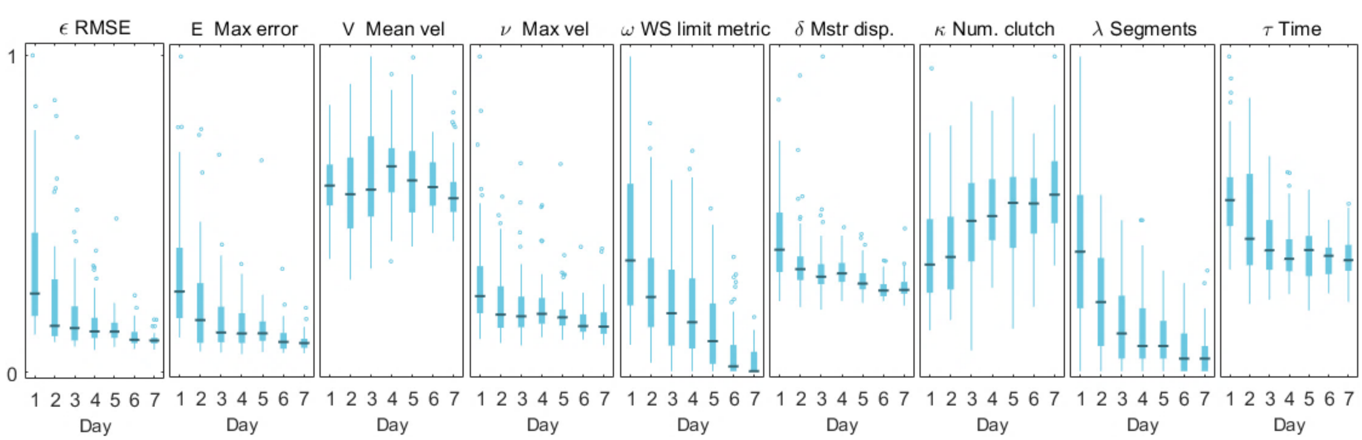


**Fig. 4** The results of the training experiment. Each bar contains 84 samples (7 users and 12 repetitions per user). Central marks are median; bar edges are the 25th and 75th percentiles; whiskers are values within 1.5 times the IQR and dots are outliers. For a better visualization the samples were normalized using the following maximum values: $\epsilon$ = 12.9 [mm], E = 39.7 [mm], $\nu$ = 5.9 [mm/s], V = 92.8 [mm/s], $\omega$ = 0.8, $\delta$ = 3.1 [m], $\kappa$ = 12.5, $\lambda$ = 29, $\tau$ = 169.2 [s]

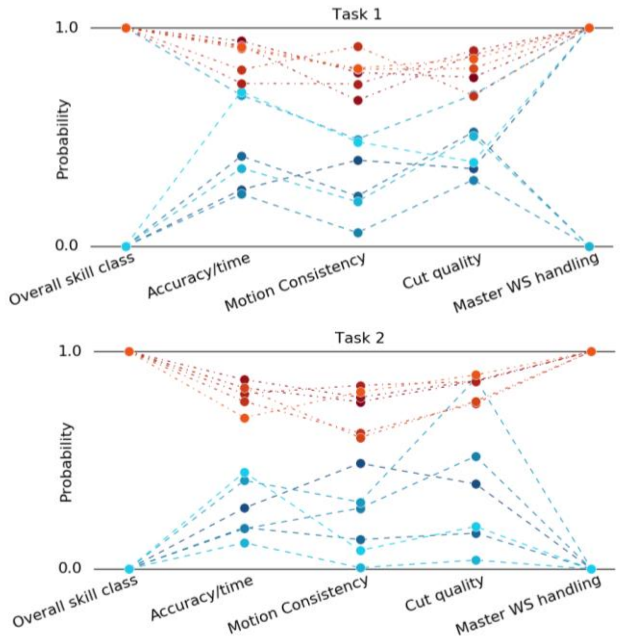


**Fig. 5** User profiles generated in the placement tests for the novice-2 and skilled groups. Vertical axis indicates the probability of being high-skilled. Overall class and Master WS handling have binary outputs and reported here for a concise visualization of a user profile. Cold and warm colors represent novice-1 and skilled subjects respectively. The AMs' assistance levels for each user are proportional to these evaluations as described in section 3.5

group. Although based on the training phase analysis it was deemed unlikely that a considerable training could happen in a single day session, a group of subjects (novice1) performed all the repetitions with no assistance to ensure no learning bias is present in the test data. The statistical significance of the difference among populations of each task, shown by a horizontal line above the bars, was determined using the Friedman test ($p<0.05$). The results hint at a statistically significant enhancement in the performance of novice users (novice-2 group) when the AMs are used, in terms of all metrics except for time. Performance variability among the novice users is considerably reduced with the AMs as can be inspected from the plots and the interquartile ranges (IQR) reported in Tables 3 and 4. The non-assisted novices (novice1 group) did not show significant improvements ruling out the possibility of intra-session learning bias. No significant difference was found between the assisted and non-assisted repetitions of the skilled users. A comparison of the metrics shows that the AMs have improved the performance of the novice users to the point that it is not statistically different from that of the skilled subjects along all metrics. Note that the overall performance enhancement of the novice users (especially in terms of variability) with the AMs has been achieved while the gains of these methods were highly variably among these users resulting in variable energetic activity of assistive methods for each user. Although setting the same fully active assistive method for all users could result in a similar or higher performance improvement, as mentioned in section I, there is a trade-off between the level of activity of an assistive method and the subjective quality of the tele-operation. Expert users such as surgeons are highly sensitive to any medium that may interfere with their intended motion of the surgical tools and losing the

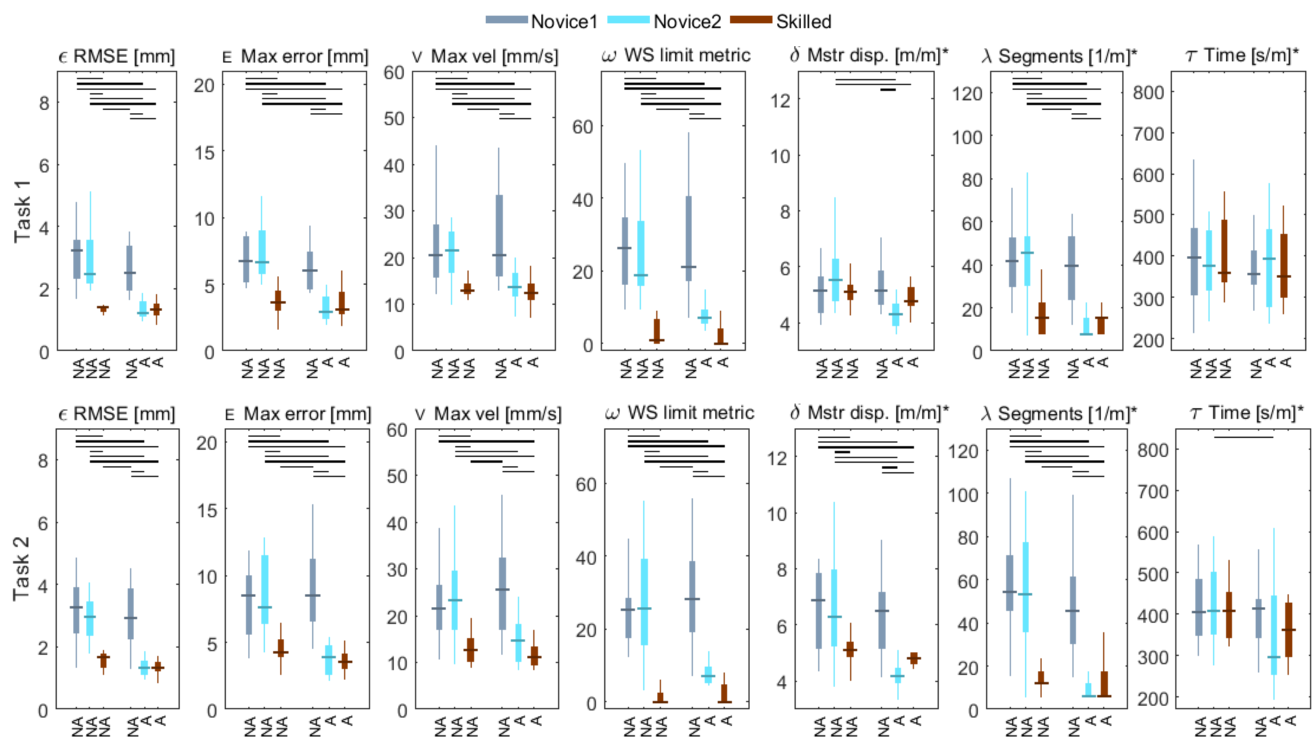


**Fig. 6** Results of the test experiment for both tasks. NA: Not-Assisted, A: Assisted. The performance metrics for both tasks of the test experiment and all user groups. The novice-1 group did not receive any assistance to investigate inter-session learning bias. Statistical significant differences ($p<0.05$) are depicted by horizontal lines on top. Outliers are not displayed in this figure for a better visualization (no outlier has been removed from the results analysis). *Divided by path length (T1=0.168 m, T2=0.132 m).

**Table 3** Calculated Performance metrics of the test experiment task 1

| | Novice1 | | Novice2 | | Skilled | | Novice1 | | Novice2 Assisted | | Skilled Assisted | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Med. | IQR | Med. | IQR | Med. | IQR | Med. | IQR | Med. | IQR | Med. | IQR |
| RMSE [mm] | 3.2 | 1.3 | 2.5 | 1.4 | 1.4 | 0.2 | 2.5 | 1.4 | 1.2 | 0.5 | 1.3 | 0.4 |
| Max err [mm] | 6.8 | 3.5 | 6.6 | 3.3 | 3.8 | 1.5 | 6.0 | 2.9 | 2.9 | 1.7 | 3.1 | 1.7 |
| Max vel [mm/s] | 20.5 | 11.5 | 21.5 | 8.8 | 13.0 | 2.2 | 205 | 17.5 | 13.7 | 5.1 | 12.4 | 3.6 |
| WS limit metric | 26.3 | 18.6 | 18.7 | 17.9 | 0.8 | 6.8 | 21.1 | 23.3 | 7.2 | 4.1 | 0 | 4.3 |
| Mstr disp.* [m/m] | 5.1 | 1.3 | 5.6 | 1.5 | 5.1 | 0.5 | 5.1 | 1.2 | 4.3 | 0.8 | 4.8 | 0.7 |
| Segments* [1/m] | 41 | 23 | 45 | 23 | 15 | 15 | 40 | 29 | 8 | 8 | 15 | 8 |
| Time* [s/m] | 396.1 | 161.9 | 375.4 | 146.1 | 360.1 | 153.2 | 357.0 | 84.7 | 395.1 | 188.1 | 3508 | 155.0 |

*values scaled by the path length.

**Table 4** Calculated Performance metrics of the test experiment task 2

| | Novice1 | | Novice2 | | Skilled | | Novice1 | | Novice2 Assisted | | Skilled Assisted | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Med. | IQR | Med. | IQR | Med. | IQR | Med. | IQR | Med. | IQR | Med. | IQR |
| RMSE [mm] | 3.3 | 1.5 | 3.0 | 1.1 | 1.7 | 0.4 | 2.9 | 1.6 | 1.3 | 0.4 | 1.3 | 0.3 |
| Max err [mm] | 8.5 | 4.4 | 7.6 | 5.2 | 4.3 | 1.4 | 8.5 | 4.6 | 3.9 | 2.2 | 3.5 | 1.1 |
| Max vel [mm/s] | 21.5 | 9.6 | 23.3 | 12.7 | 12.7 | 5.0 | 25.6 | 15.4 | 14.8 | 8.3 | 11.1 | 4.0 |
| WS limit metric | 25.4 | 11.0 | 25.6 | 24.0 | 0.0 | 2.5 | 28.1 | 19.5 | 7.1 | 4.6 | 0 | 4.7 |
| Mstr disp.* [m/m] | 6.9 | 2.7 | 6.3 | 2.7 | 5.1 | 0.6 | 6.5 | 2.0 | 4.2 | 0.6 | 4.8 | 0.4 |
| Segments* [1/m] | 54 | 26 | 54 | 42 | 12 | 6 | 46 | 31 | 6 | 6 | 6 | 12 |
| Time* [s/m] | 403.9 | 136.6 | 406.8 | 152.7 | 407.7 | 112.6 | 414.0 | 94.0 | 292.1 | 192.0 | 362.2 | 133.3 |

*values scaled by the path length.

liberty of motion to an over-restraining assistive method can degrade their evaluation of such assisted tele-operation system.

As a measure of the assistive methods' level of activity the integral of the exerted assistive force's magnitude over time was calculated. Known as impulse in classical mechanics, the integral of force applied to an object over a time interval is equivalent to the vector change in its linear momentum. The product of impulse and duration in fact is equal to average force. Note that not having the measurements of the exerted force by the user to the haptic device, the work done by the assistive forces cannot be estimated explicitly. The average of impulse magnitudes calculated for 3 repetitions of each user per task are reported in Fig. 7. As expected, novice users have received a more active assistance compared to the skilled group. The activity level is also different among the novice

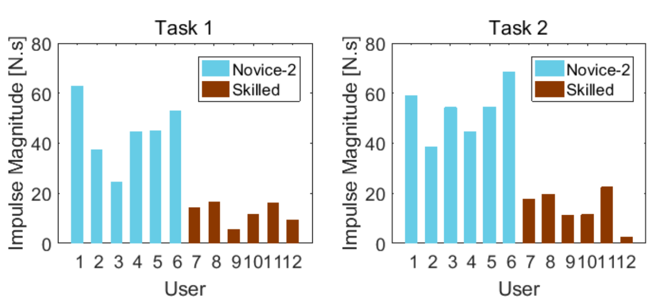


**Fig. 7** The time integral of assistive force's magnitude applied to each user. Each bar represents the average of the impulse magnitudes of 3 repetitions performed by a user.

users as the estimated skill levels (and accordingly their performances) have been different. Considering the performance results discussed in the previous paragraph, this suggests that some novice users have achieved performances comparable to that of skilled users without the full application of AMs. The lower impulse magnitudes of the assistive methods for skilled users confirms the lower engagement of the guiding forces which can reduce the amount of dissipated user energy and optimize the human-robot interaction.

## 5 Conclusions

The presented work introduced a tele-operated human-robot cooperation framework where the assistance is adapted to user skill profile. The skill profiles were generated through a set of classifiers based on motion- and task-related features, that are found to depict the training curve associated with the studied path tracking task. It must however be noted that the obtained classification results are for a relatively simple and continuous task of path tracking. Complex tasks such as suturing in surgical teleoperation may require further deciphering and segmentation. It was shown that novice users could demonstrate a variability in different competences that can be better captured by a skill profile. Such interpretable assessment can also be valuable in robotic training programs to accelerate the training through a more targeted approach. Furthermore, we studied the feasibility of adapting haptic assistive methods to robotic systems' operator skills. An experimental tele-operation framework was introduced that employs the generated skill profiles to customize a set of haptic AMs so that users receives an assistance tailored to their skill levels. It was validated that such an approach can improve the assistance/subjective-quality trade-off by preventing excessive energy dissipation for users with higher skill levels while not deteriorating the performance enhancement expected from such assistive methods. Our future works will focus on diversification of tasks—focusing on surgical applications—and experimentation using the da Vinci research kit, increasing the similarity of the tasks to those performed in the operating room and performing validation experiments employing skilled surgeons and novice residents.

## 6 Acknowledgements

This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. H2020-ICT-26-2016-1 -732515 (SMARTsurg project).

## 7 References

Abbott, JJ, P Marayong, and AM Okamura. 2007. "Haptic Virtual Fixtures for Robot-Assisted Manipulation." *Robotics research*: 49–64. http://link.springer.com/chapter/10.1007/978-3-540-48113-3_5 (October 29, 2014).

Ahmidi, N et al. 2017. "A Dataset and Benchmarks for Segmentation and Recognition of Gestures in Robotic Surgery." *IEEE transactions on bio-medical engineering* 9294(c): 1–16.

Anderson, Sterling J, Sisir B. Karumanchi, and Karl Iagnemma. 2012. "Constraint-Based Planning and Control for Safe, Semi-Autonomous Operation of Vehicles." In *2012 IEEE Intelligent Vehicles Symposium*, IEEE, 383–88.

Bauzano, Enrique, B Estebanez, and Victor F Muñoz. 2014. "Collaborative Human – Robot System for HALS Suture Procedures." *IEEE SYSTEMS JOURNAL*: 1–10.

Beretta, E. et al. 2015. "Enhanced Torque-Based Impedance Control to Assist Brain Targeting during Open-Skull Neurosurgery: A Feasibility Study." *International Journal of Medical Robotics and Computer Assisted Surgery* 12(3): 326–41. http://doi.wiley.com/10.1002/rcs.1690.

Berthet-rayne, Pierre, Maura Power, Hawkeye King, and Guang-zhong Yang Fellow. 2016. "Hubot : A Three State Human-Robot Collaborative Framework for Bimanual Surgical Tasks Based on Learned Models." In *International Conference on Robotics and Automation*, , 715–22.

Bowyer, SA, BL Davies, and F Rodriguez y Baena. 2014. "Active Constraints/virtual Fixtures: A Survey." *IEEE Transactions on Robotics* 30(1): 138–57.

Bowyer, Stuart A., and Ferdinando Rodriguez Y Baena. 2015. "Dissipative Control for Physical Human-Robot Interaction." *IEEE Transactions on Robotics* 31(6): 1281–93.

Bruyninckx, H. 2001. "Open Robot Control Software: The OROCOS Project." In *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164)*, IEEE, 2523–28.

Chmarra, Magdalena K. et al. 2010. "Objective Classification of Residents Based on Their Psychomotor Laparoscopic Skills." *Surgical Endoscopy* 24(5): 1031–39.

Laparoscopic Skills Trainers: The Weakest Link!" *MICCAI - Medical Image Computing and Computer-Assisted Intervention* 2488: 35–43.

Dragan, A. D., and S. S. Srinivasa. 2013. "A Policy-Blending Formalism for Shared Control." *The International Journal of Robotics Research* 32(7): 790–805.

Enayati, Nima, Eva C. Alves Costa, Giancarlo Ferrigno, and Elena De Momi. 2016. "A Dynamic Non-Energy-Storing Guidance Constraint with Motion Redirection for Robot-Assisted Surgery." In *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, Daejeon, Korea.

Enayati, Nima, Elena De Momi, and Giancarlo Ferrigno. 2016. "Haptics in Robot-Assisted Surgery: Challenges and Benefits." *IEEE Reviews in Biomedical Engineering* 9: 49–65.

Enayati, Nima, and Farid Najafi. 2011. "Design and Manufacturing of a Tele-operative Rescue Robot with a Novel Track Arrangement." *Industrial Robot: An International Journal* 38(5): 476–85.

Fard, Mahtab Jahanbani, Sattar Ameri, Ratna Babu Chinnam, and R. Darin Ellis. 2017. "Soft Boundary Approach for Unsupervised Gesture Segmentation in Robotic-Assisted Surgery." *IEEE Robotics and Automation Letters* 2(1): 171–78.

Finan, Michael A., Michael E. Clark, and Rodney P. Rocconi. 2010. "A Novel Method for Training Residents in Robotic Hysterectomy." *Journal of Robotic Surgery* 4(1): 33–39.

Franchi, Antonio et al. 2012. "Shared Control : Balancing Autonomy and Human Assistance with a Group of Quadrotor UAVs." *IEEE Robotics & Automation Magazine* 19(3): 57–68.

Gao, Ming, Jan Oberlander, Thomas Schamm, and J. Marius Zollner. 2014. "Contextual Task-Aware Shared Autonomy for Assistive Mobile Robot Teleoperation." In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, 3311–18.

Garrido-Jurado, S., R. Muñoz-Salinas, F.J. Madrid-Cuevas, and M.J. Marín-Jiménez. 2014. "Automatic Generation and Detection of Highly Reliable Fiducial Markers under Occlusion." *Pattern Recognition* 47(6): 2280–92.

Ghez, Claude, John W. Krakauer, and Maria-Felice Ghilardi. 1999. "Independent Learning of Internal Models for Kinematic and Dynamic Control of Reaching." *Nature Neuroscience* 2(11): 1026–31.

Gomes, Paula. 2011. "Surgical Robotics: Reviewing the Past, Analysing the Present, Imagining the Future." *Robotics and Computer-Integrated Manufacturing* 27(2): 261–66.

Hajshirmohammadi, I, and S Payandeh. 2007. "Fuzzy Set Theory for Performance Evaluation in a Surgical Simulator." *Presence-Teleoperators and Virtual Environments* 16(6): 603–22.

Van Hove, P. D. et al. 2010. "Objective Assessment of Technical Surgical Skills." *British Journal of Surgery* 97(7): 972–87.

Kim, Stephen et al. 2010. "Lack of a Discriminatory Function for Endoscopy Skills on a Computer-Based Simulator." *Surgical Endoscopy* 24(12): 3008–15. http://link.springer.com/10.1007/s00464-010-1077-z (May 10, 2017).

Kranzfelder, Michael et al. 2013. "Toward Increased Autonomy in the Surgical OR: Needs, Requests, and Expectations." *Surgical Endoscopy and Other Interventional Techniques* 27(5): 1681–88.

Lawrence, D.a. 1993. "Stability and Transparency in Bilateral Teleoperation." *IEEE Transactions on Robotics and Automation* 9(5): 624–37.

Lin, Henry C, Izhak Shafran, David Yuh, and Gregory D Hager. 2006. "Towards Automatic Skill Evaluation: Detection and Segmentation of Robot-Assisted Surgical Motions." *Computer aided surgery* 11(5): 220–30.

Liu, Yeou-Teh, Gottfried Mayer-Kress, and Karl M. Newell. 2006. "Qualitative and Quantitative Change in the Dynamics of Motor Learning." *Journal of Experimental Psychology: Human Perception and Performance* 32(2): 380–93.

Marchal-Crespo, Laura, and David J Reinkensmeyer. 2008. "Haptic Guidance Can Enhance Motor Learning of a Steering Task." *Journal of motor behavior* 40(6): 545–56.

Megali, Giuseppe, Stefano Sinigaglia, Oliver Tonet, and Paolo Dario. 2006. "Modelling and Evaluation of Surgical Performance Using Hidden Markov Models." *IEEE Transactions on Biomedical Engineering* 53(10): 1911–19.

Milliken, Lauren, and Geoffrey A Hollinger. 2016. "Modeling User Expertise for Choosing Levels of Shared Autonomy." In *Planning for Human-Robot Interaction Shared Autonomy and Collaborative Robotics Workshop, Robotics: Science and Systems Conference*.

Nichols, Kirk A. et al. 2015. "Models of Human-Centered Automation in a Debridement Task." *IEEE International Conference on Intelligent Robots and Systems* 2015–Decem: 5784–89.

Padoy, Nicolas, and Gregory D. Hager. 2011. "Human-Machine Collaborative Surgery Using Learned Models." *Proceedings - IEEE International Conference on Robotics and Automation*: 5285–92.

Pedregosa, Fabian et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12(Oct): 2825–30.

Penin, L.F., K. Matsumoto, and S. Wakabayashi. "Force Reflection for Time-Delayed Teleoperation of Space Robots." In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065)*, IEEE, 3120–25.

Rafii-Tari, Hedyeh et al. 2013. "Learning-Based Modeling of Endovascular Navigation for Collaborative Robotic Catheterization." In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer Berlin Heidelberg, 369–77.

Reiley, Carol E., and Gregory D. Hager. 2009. "Task versus Subtask Surgical Skill Evaluation of Robotic Minimally Invasive Surgery." In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, London: Springer Berlin Heidelberg, 435–42.

Richards, C. et al. 2000. "Skills Evaluation in Minimally Invasive Surgery Using Force/torque Signatures." *Surgical Endoscopy* 14(9): 791–98.

Rosen, Jacob et al. 2006. "Generalized Approach for Modeling Minimally Invasive Surgery as a Stochastic Process Using a Discrete Markov Model." *IEEE Transactions on Biomedical Engineering* 53(3): 399–413.

Sainburg, Robert L., and Jinsung Wang. 2002. "Interlimb Transfer of Visuomotor Rotations: Independence of Direction and Final Position Information." *Experimental Brain Research* 145(4): 437–47.

Schmidt, Richard A. 1991. "Frequent Augmented Feedback Can Degrade Learning: Evidence and Interpretations." In *Tutorials in Motor Neuroscience*, Dordrecht: Springer Netherlands, 59–75.

Schreuder, H. W R et al. 2012. "Training and Learning Robotic Surgery, Time for a More Structured Approach: A Systematic Review." *BJOG: An International Journal of Obstetrics and Gynaecology* 119(2): 137–49.

Shadmehr, Reza, Maurice A. Smith, and John W. Krakauer. 2010. "Error Correction, Sensory Prediction, and Adaptation in Motor Control." *Annual Review of Neuroscience* 33(1): 89–108.

Shmuelof, L., J. W. Krakauer, and P. Mazzoni. 2012. "How Is a Motor Skill Learned? Change and Invariance at the Levels of Task Success and Trajectory Control." *Journal of Neurophysiology* 108(2): 578–94.

Sigrist, Roland, Georg Rauter, Robert Riener, and Peter Wolf. 2013. "Augmented Visual, Auditory, Haptic, and Multimodal Feedback in Motor Learning: A Review." *Psychonomic Bulletin & Review* 20(1): 21–53.

Takayama, Leila, Eitan Marder-Eppstein, Helen Harris, and Jenay M. Beer. 2011. "Assisted Driving of a Mobile Remote Presence System: System Design and Controlled User Evaluation." In *2011 IEEE International Conference on Robotics and Automation*, IEEE, 1883–89.

Trejos, Ana Luisa et al. 2013. "The Application of Force Sensing to Skills Assessment in Minimally Invasive Surgery." In *2013 IEEE International Conference on Robotics and Automation*, IEEE, 4370–75.