A Self-Adaptive Motion Scaling Framework for Surgical Robot Remote Control

Dandan Zhang, *Student Member, IEEE*, Bo Xiao, Baoru Huang, Lin Zhang, Jindong Liu, and Guang-Zhong Yang, *Fellow, IEEE*

***Abstract*—Master–slave control is a common form of human–robot interaction for robotic surgery. To ensure seamless and intuitive control, a mechanism of self-adaptive motion scaling during teleoperaton is proposed in this letter. The operator can retain precise control when conducting delicate or complex manipulation, while the movement to a remote target is accelerated via adaptive motion scaling. The proposed framework consists of three components: 1) situation awareness, 2) skill level awareness, and 3) task awareness. The self-adaptive motion scaling ratio allows the operators to perform surgical tasks with high efficiency, forgoing the need of frequent clutching and instrument repositioning. The proposed framework has been verified on a da Vinci Research Kit to assess its usability and robustness. An in-house database is constructed for offline model training and parameter estimation, including both the kinematic data obtained from the robot and visual cues captured through the endoscope. Detailed user studies indicate that a suitable motion-scaling ratio can be obtained and adjusted online. The overall performance of the operators in terms of control efficiency and task completion is significantly improved with the proposed framework.**

***Index Terms*—Learning and adaptive systems, telerobotics and teleoperation, medical robots and systems.**

I. Introduction

ROBOT-ASSISTED Minimally Invasive Surgery (RAMIS) has helped extend the clinical impact of MIS in terms of consistency, accuracy, and safety of procedures [1]. Currently, master-slave control is the primary form of the surgical remote control to ensure that the slave robot accurately executes an operator's command. In a master-slave control framework, the improvement of human-robot interaction can simplify the performance of surgical procedures, bringing improved clinical outcomes [2].

Motion scaling is an essential part of the master-slave paradigm. For performing delicate tasks, measured hand motion of the master manipulators is scaled down and replicated by the slave robots. The use of motion scaling enables the users to conduct precise manipulation during teleoperation [3]. However, the small motion scaling ratio prevents the operator from reaching distant targets without clutching. This is time-consuming as the operator needs to reposition the master by disconnecting the master from the slave. There is therefore the practical need for the robot to cater for different surgical maneuverers and change adaptively the scaling ratio depending on the task being performed [4]. A small ratio is required for delicate motion involved in micro-surgical tasks and a relatively large ratio can be used for gross positioning and reaching to remote targets. In practice, an online adjustable and context-aware motion scaling ratio is beneficial for optimal master-slave mapping and task completion.

Another factor influencing the selection of motion scaling is related to the skill level of the operator. Although the motion scaling ratio can be modified manually by the surgeon, it is difficult to determine an optimal value for a given task. A method was explored in [5] to improve both the speed and precision during the surgical operation, with the main focus on optimizing a generic value without an adaptive mechanism. However, a scaling ratio that is optimal for one user may be unsuitable for another. Furthermore, given the same person, a different scaling ratio may be used depending on the specific task being performed and the cognitive load due to stress, team cooperation and unexpected complications. Therefore, a self-adaptive motion scaling mechanism is required, through which the system can dynamically adjust the motion scaling ratio based on the task requirements, operating condition and the user preference.

In [4], the motion scaling ratio is automatically adjusted to reduce the task completion time for a master-slave robotic neurosurgical system. Comparisons are made among different modes through a user study involving five subjects. However, each mode has its inherent limitations and the linear assumption used for motion-scaling ratio modulation is not appropriate for all surgical tasks.

A gaze-assisted intention recognition scheme has been proposed [6] to achieve adaptive motion scaling. Eye-tracking in conjunction with other sensing modalities is employed to infer the intended position that the operator is trying to reach so that the motion scaling ratio can be changed accordingly. The information is then used to modulate the motion scaling ratio, which enables the operator to reach distant targets more quickly [6]. It is worth noting that adaptive motion scaling with eye tracking

Manuscript received September 10, 2018; accepted December 15, 2018. Date of publication December 28, 2018; date of current version January 11, 2019. This letter was recommended for publication by Associate Editor E. De Momi and Editor P. Valdastri upon evaluation of the reviewers' comments. The work (Ph.D. study) of D. Zhang was supported by the Hamlyn Centre and the Chinese Scholarship Council (CSC). *(Corresponding author: Dandan Zhang.)*

The authors are with the Hamlyn Centre for Robotic Surgery, Imperial College London, London SW7 2AZ, U.K. (e-mail: d.zhang17@imperial.ac.uk; b.xiao@imperial.ac.uk; huangsally5686@gmail.com; lin.zhang11@imperial.ac.uk; jindong.liu@gmail.com; g.z.yang@imperial.ac.uk).

Digital Object Identifier 10.1109/LRA.2018.2890200

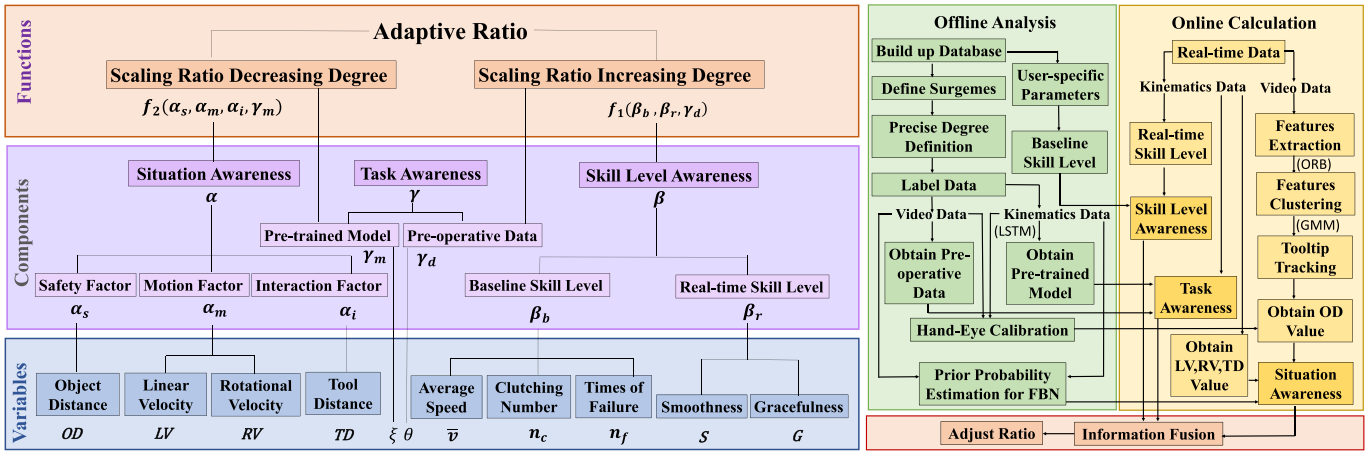

Fig. 1. Overview of the self-adaptive motion scaling framework and the flow charts of the offline analysis and the online calculation.

is still difficult to be fully integrated to general surgical systems and the effectiveness of gaze-assisted intention recognition can be problematic when operators wear glasses. Therefore, a more general solution is required to combine the kinematic data with the vision data for the construction of an adaptive mechanism.

In existing work, it was reported that the parameters could be automatically tuned by using Bayesian inference to optimize a human-robot interface in a user-specific manner [7], the subjects were required to perform the same tasks for multiple trials to grade the robot's behavior or select preferred parameters, which is not realistic in clinical settings.

In order to address the aforementioned limitations, a novel self-adaptive motion scaling framework is proposed in this letter. The framework utilizes both vision and kinematic data to provide complementary information. Real-time vision contains semantic information (e.g., instrument motion characteristics and tool-tissue interaction), which is equally or more informative than kinematics data for determining surgical situation awareness [8]. The kinematic data, on the other hand, contains the detailed trajectories and poses of the end-effectors, as well as the status of the end-effectors of the instruments.

Probabilistic modeling can be employed to recognize the advent of specific surgical procedures that require the precise operation, which represents situation awareness of the system. Skill level assessment is incorporated to realize user-specific parameter estimation. Surgical episode identification is necessary to take task-driven factors into consideration and ensure flexible adaption to different surgical scenarios.

The overall structure of the rest of this paper is organized as follows. Firstly, the methodology is introduced in Section II. Secondly, detailed user studies are described in Section III. The results are presented and analyzed in Section IV and finally, conclusions are drawn in Section V.

II. Methodology

In this section, the architecture and technical details of the construction of the self-adaptive motion scaling framework are illustrated, which dictates the final adaptive ratio.

A. System Overview

The overall framework is illustrated in Fig. 1. The adaptive ratio is determined by three components, including 1) situation awareness, 2) skill level awareness, and 3) task awareness. Situation awareness can be regarded as an online pattern recognition of different operational situations (e.g., tool-tissue interaction, dexterous and bimanual operation). User-specific factors can be addressed by skill level assessment, while task-specific factors are incorporated into the framework to ensure generalizability of the system.

1) Situation Awareness (α):

- **Safety Factor** (α_s): When the end-effectors of the slave robot are operating within a short distance to the tissue, the scaling ratio should be small to ensure safety. The value of α_s is assigned as the probability of tool-tissue interaction.
- **Motion Factor** (α_m): When the motion of the robot varies extensively, the operator is assumed to conduct a dexterous operation, like positioning and targeting. The value of α_m is assigned as the probability of dexterous operation.
- **Interaction Factor** (α_i): When the distance between a pair of instruments of the slave robot is small, it is assumed that the slave robots are involved in the bimanual operation. The value of α_i is assigned as the probability of bimanual operation.

2) Skill Level Awareness (β):

- **Baseline Skill Level** (β_b): Baseline skill level β_b is adopted as a constraint to limit the adaptive ratio within a reasonable range. The value of β_b depends on the skill level of the operators given by their previous operation performance.
- **Real-time Skill Level** (β_r): Real-time skill level β_r is introduced to adjust the skill level of the operator online during the teleoperation process. The skill level of the operator is not fixed, because the operators may encounter various situations when performing different procedures.

3) Task Awareness (γ):

- **Pre-trained Model** (γ_m): Different surgical tasks may have different inherent characteristics. A pre-trained model for surgical gesture recognition through offline training is useful to incorporate the task-specific features. γ_m is deter-

mined by the online surgical gesture classification results as the real-time precise degree evaluation.

- **Pre-operative Data** (γ_d): Pre-operative data can be used to predict the complexity of procedures at different phases of the operation. Therefore, the framework can combine the information from the pre-operative data as the conservative degree of surgical skill assessment when conducting the operation in critical areas.

B. Situation Awareness

Bayesian networks can capture the relationship of causality by incorporating prior knowledge in a parameter estimation process. In a Bayesian network, suppose that there are n child nodes in total and each child node has only one corresponding parent node. $P\left(C_i|F(C_i)\right)(i=1,2,\ldots,n)$ refers to the conditional occurrence probability of C_i (child node) given that $F(C_i)$ (parent node) occurs.

Fuzzy Bayesian Network (FBN) utilizes fuzzy logic [9] to simplify continuous data and generate soft evidence with uncertainty in the Bayesian network. Suppose that $\widetilde{C}_i = \left\{\widetilde{C}_{i1}, \widetilde{C}_{i2}\ldots, \widetilde{C}_{ir}\right\}$ represents the fuzzy set for variable C_i generated by the corresponding membership function $\mu_{ij}(j=1,2\ldots r)$, where $\widetilde{C}_{ij} = \mu_{ij}(x)$ is the j-th fuzzy state of $\forall x \in C_i$. r denotes the total number of fuzzy states, which can be determined the specific application scenarios. The grade of the membership function is used as the probability value. Considering the fuzzy state mapping, $\hat{P} = P\left(F_l(C_i)|\widetilde{C}_i\right)$ can be obtained by (1):

$$\hat{P} = \sum_{j=1}^{r} \frac{P\left(\widetilde{C}_{ij}|F_l(C_i)\right) \times P(F_l(C_i))}{\sum_{k=1}^{m} P\left(\widetilde{C}_{ij}|F_k(C_i)\right) \times P(F_k(C_i))} \times \widetilde{C}_{ij} \quad (1)$$

where $F_l(C_i)(l=1,2,\ldots,m)$ represents the parent node at state l.

In the proposed framework, tool-tissue interaction indicates that the surgical tools approach a critical area of the anatomy. Dexterous operation represents the situation that the surgical tools are involved in relatively big movements, while bimanual operation represents the tool-tool interaction. Tool-tissue interaction, dexterous operation and bimanual operation are the child nodes of precise operation, while object distance, linear velocity, rotational velocity and tool distance are evidence at the fuzzy mapping level. True and false represent the two states for nodes (tool-tissue interaction, dexterous operation and bimanual operation), the probability of which can be regarded as fuzzy values. Small and big are two fuzzy states for the evaluation of the nodes at the fuzzy mapping level. The structure of the FBN used in this letter for situation awareness is shown in Fig. 2.

As for online data acquisition, object distance (OD) is the characteristic distance between the end-effectors and the nearest critical area for surgical operation, which is estimated by visual cues from the endoscope. Linear velocity (LV) and rotational velocity (RV) can be acquired by differentiating the real-time pose of the end-effectors of the slave robot. Tool distance (TD) is the Euclidean distance between the trajectory generated by

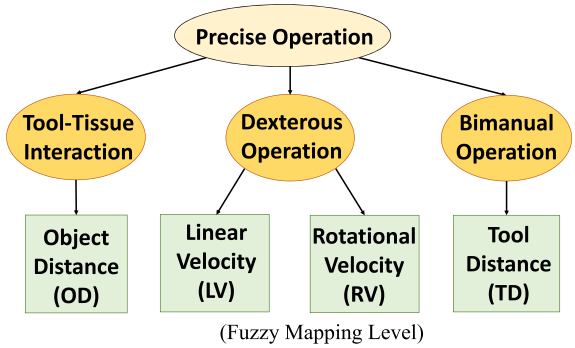

Fig. 2. The Fuzzy Bayesian Network (FBN) for the situation awareness.

the two end-effectors of the slave robot. The value of OD, LV, RV and TD are calculated by (2).

$$\begin{cases} OD = \max(\min(P_{li}),\ \min(P_{ri}))(i=1,2,\ldots,s) \\ LV = ||\boldsymbol{\lambda}_*(t) - \boldsymbol{\lambda}_*(t-\Delta t)||/\Delta t(*=l,r) \\ RV = ||\boldsymbol{\Omega}_*(t) - \boldsymbol{\Omega}_*(t-\Delta t)||/\Delta t(*=l,r) \\ TD = ||\boldsymbol{\lambda}_l(t) - \boldsymbol{\lambda}_r(t)|| \end{cases} \quad (2)$$

where P_{li} and P_{ri} $(i=1,2,\ldots,s)$ are the distance between the end-effectors of the surgical tools to the features on the left image and that on the right image respectively. In the above equation, s indicates the number of clustering centers of the feature points acquired by raw images. $\boldsymbol{\lambda_r}(t)$ and $\boldsymbol{\lambda_l}(t)$ are the 3D trajectory profile, while $\boldsymbol{\Omega_r}(t)$ and $\boldsymbol{\Omega_l}(t)$ are the 3D orientation profile of the tooltips of the right and left surgical tools at time step t respectively. Δt is the time interval which depends on the the kinematic data acquisition rate.

After acquiring all the required continuous data, the fuzzy values are estimated, and are used as evidences for the FBN. α_s, α_m, and α_i represent the probability of tool-tissue interaction, dexterous operation and bimanual operation respectively. With these obtained values and other prior information, the posterior probability of precise operation $P(\alpha_s, \alpha_m, \alpha_i)$ can be estimated based on (1), which is assigned to α and can be adjusted online.

C. Skill Level Awareness

The skill level can be determined by both baseline and real-time skill levels.

1) Baseline Skill Level: Baseline skill level β_b is a fixed parameter that can be estimated by the previous performance of the operators. To this end, the operators historical data for task execution is utilized for analysis.

For MIS, Objective Structured Assessment of Technical Skills (OSATS) is normally used [10]. However, the grading process is time-consuming. To simplify this process, three metrics are selected to form the baseline skill assessment in this letter. These include the time-normalized instrument path length $\bar{v}$ of the slave robot end-effectors, the number of clutching n_c on the master side and the number of failure n_f (e.g., the number of dropping the peg or rubber ring during transfer process), which have been reported to demonstrate some aspects of the surgeon's expertise in [11].

Suppose that R_u and R_d are the upper and lower limit of skill assessment range respectively. Baseline skill level β_b can be determined by (3).

$$\beta_b = R_d + (R_u - R_d) \times \min\left(\frac{\bar{v}}{\max(\bar{v})}, \frac{n_c + n_f}{\max(n_c + n_f)}\right) \tag{3}$$

where $\max(\bar{v})$ is the maximum average control speed among all the trials in the database, $\max(n_c + n_f)$ is the maximum number of the sum of n_c and n_f during one trial based on the previously recorded data.

2) Real-Time Skill Level: The real-time skill level can be adjusted at the end of a time segment $T_i (i = 1, 2, \ldots, n)$, where T_n is the current time segment, t represents the current time point. By monitoring the motion quality, real-time skill level adjustment can be realized.

Gracefulness (G) is quantified based on the curvature (κ) of a trajectory, which indicates the straightness of the path [12], [13]. Median curvature along the trajectory generated during time segment T_i is used to determine gracefulness. It is defined as follows:

$$G = \text{Median}(\log_{10}^{\kappa}(T_i)) \tag{4}$$

where κ is calculated based on (5).

$$\kappa = \frac{||\dot{\boldsymbol{\lambda}}(t) \times \ddot{\boldsymbol{\lambda}}(t)||}{||\dot{\boldsymbol{\lambda}}(t)||^3} \tag{5}$$

where $\boldsymbol{\lambda}(t)$ represents the points in 3D space in the form of vector. $\dot{\boldsymbol{\lambda}}(t)$ and $\ddot{\boldsymbol{\lambda}}(t)$ are instantaneous velocity and acceleration of the surgical tooltips respectively, which can be calculated directly by computing the first and second derivatives of the positions of the tooltips.

Smoothness (S) is quantified based on the rate of changes in acceleration [14], which can reflect the continuity or non-intermittency of a movement.

A valid, sensitive, practical and reliable measure should be considered for smoothness estimation [15]. Compared to root mean square jerk, normalized mean absolute jerk and spectral arc length, the number of peaks and dimensionless jerk have the advantage of being independent of its amplitude and duration. Log dimensionless jerk is able to eliminate the inherent lack of sensitivity in the physiological range brought by the dimensionless jerk and the number of peaks during measurement [15]. Therefore, Smoothness is defined as follows.

$$S = \text{Median}(\log_{10}^{\phi}(T_i)) \tag{6}$$

where ϕ is calculated based on (7).

$$\phi = \frac{\sigma^5}{v_p^2} \int_{t-\sigma}^{t} \left| \frac{d^3 \boldsymbol{\lambda}(t)}{dt^3} \right|^2 dt \tag{7}$$

where σ is the duration of the time segment which can be determined by the real-time skill level updating rate, v_p is the peak value of velocity within the time segment.

As the log dimensionless jerk is sensitive to measurement noise, the kinematic data are filtered at first before being used for the real-time skill level assessment to ensure reliability. This can reduce the influence from the noisy data. The real-time skill level β_r can therefore be determined as follows:

$$\beta_r = \max\left\{\frac{1}{1 + e^{S - \varepsilon_1}}, \frac{1}{1 + e^{G - \varepsilon_2}}\right\} \tag{8}$$

where parameter ε_1 and ε_2 can be tuned based on the application scenarios.

Frequency-based features and entropy-based features were used in some literature for skill assessment. While end-to-end learning methods without feature extraction were employed in [11], [16]. However, this letter is mainly focused on the self-adaptive motion scaling framework construction. Details of the evaluation metric and parameter estimation can be readily extended to other situations.

D. Task Awareness

1) Pre-Operative Data: Current imaging modalities such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT), X-ray, fluoroscopy and ultrasound are widely used for pre-operative imaging and reconstruction [17]. Therefore, pre-operative data can be collected beforehand to define the targeted area for operation. Critical areas are pre-defined, and the minimal distance between the surgical tooltips and the centers of the critical areas can be monitored online.

γ_d can be calculated as follows:

$$\gamma_d = (1 - \gamma_{do}) + \gamma_{do} \times \frac{2}{1 + e^{\delta_d - \theta}} \tag{9}$$

where $\gamma_d \in (1 - \gamma_{do}, 1 + \gamma_{do})$. γ_{do} and δ_d are parameters that control the adaptive range and distribution of γ_d respectively. $\theta = \min D_j (j = 1, \ldots q)$ is a variable. D_j is the distance (unit:mm) between the surgical tooltip to the j_{th} critical areas. q is the total number of pre-defined critical areas, which depends on the complexity of the surgical tasks and the targeted anatomy.

2) Pre-Trained Model: For adaptive motion scaling, the surgical gestures are classified based on the level of precision. Hidden Markov Model (HMM) [18], Long Short Term Memory (LSTM) based Recurrent Neural Networks (RNNs) [19] or other time sequence processing technologies [20] can be employed to assist surgical episode identification and to incorporate characteristics of different tasks in the proposed framework. The relationship between the precise degree and the surgemes are pre-defined and can be modified by different surgical tasks. A classification model can be trained by utilizing the existing database. Considering its favorable performance of end-to-end learning in many situations, LSTM is utilized for classification in this letter.

$\xi (\xi = 0, 1, \ldots (N - 1))$ represents the precise degree that determines the value of γ_m, which depends on the LSTM classification results. $\gamma_m(\xi)$ can be calculated as follows:

$$\gamma_m(\xi) = (1 - \gamma_{mo}) + \gamma_{mo} \times \frac{2\xi}{N - 1} \tag{10}$$

where $\gamma_m(\xi) \in [1 - \gamma_{mo}, 1 + \gamma_{mo}]$. γ_{mo} is a parameter that control the adaptive range of γ_m, N is the total number of levels

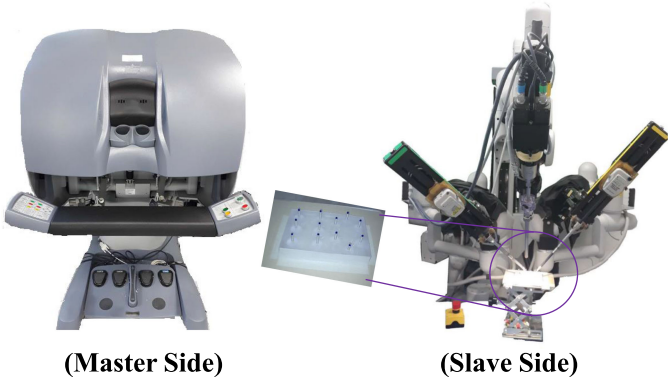

Fig. 3. The experimental platform for user studies.

for precise degree. The higher the ξ value, the more precise requirement the surgeme needs.

E. Real-Time Motion Scaling Ratio

Thus far, the methods for constructing the major components of the self-adaptive motion scaling framework have been illustrated. The final adaptive ratio $R(t)$ can be determined as the function of all the adjustable factors as shown in (11).

$$R(t) = f_1(\beta_b, \beta_r, \gamma_d) - f_2(\alpha_s, \alpha_m, \alpha_i, \gamma_m) \tag{11}$$

where $f_1(.)$ is the function for the scaling ratio increasing degree estimation, $f_2(.)$ is the function for the scaling ratio decreasing degree estimation (see (12)).

$$\begin{cases} f_1(\beta_b, \beta_r, \gamma_d) = \beta_b + \beta_r \times \gamma_d \\ f_2(\alpha_s, \alpha_m, \alpha_i, \gamma_m) = \gamma_m \times P(\alpha_s, \alpha_m, \alpha_i) \end{cases} \tag{12}$$

$f_1(.)$ is calculated based on real-time motion analysis, performance evaluation of the operator, as well as the pre-operative data. It is affected by the skill level of the operator, and the distance of the surgical tools to the operational targets.

$f_2(.)$ is determined by probabilistically based situation awareness and surgical gesture recognition. It is influenced by the inherent characteristics of the surgical procedures and online surgemes precise degree estimation during operation.

III. User Studies

In this section, detailed illustrations of user studies for validating the proposed framework are presented.

A. Experimental Setup

In order to verify the effectiveness of the proposed motion scaling framework, detailed user studies are conducted on a dVRK [21], which includes the Patient Side Manipulators (PSMs), the Endoscopic Camera Manipulator (ECM) and the Master Tool Manipulators (MTMs). This research platform has similar hardware components as the da Vinci Surgical System (Intuitive Surgical Inc., CA, United States) has. Fig. 3 illustrates the master-slave system used in the user studies.

Thanks to the dVRK-ROS Bridge, both kinematics and vision data can be captured for offline analysis and online calculation. The end-effectors' pose of the PSMs and MTMs can be subscribed from a ROS based topic (collected at 100 frames per second). The stereo stream from the endoscopic system is captured by a Kona 4 PCIe frame grabber (AJA Video System, CA, United States). The stereo vision data is recorded at 25 frames per second with a resolution of 720 × 576. Chessboard calibration [22] is utilized to estimate the rigid transform relationship between the frames of the PSMs' end-effectors and the coordinate frames of the stereo images.

TABLE I
Surgical Gesture Vocabulary and Precise Degree

Index	Surgemes Description	Precise Degree ξ		
		N=2	N=3	N=4
S1	Transfer Object Operation	1	1	2
S2	Grasp Object with Right Hand	1	2	3
S3	Move to Next Target with Right Hand	0	0	0
S4	Locate Object with Right Hand	1	1	1
S5	Grasp Object with Left Hand	1	2	3
S6	Move to Next Target with Left Hand	0	0	0
S7	Locate Object with Left Hand	1	1	1

A peg board for standard FLS (Fundamentals of Laparoscopic Surgery) training, along with a rubber ring, was employed for the ring transfer task in the user studies. The rigs on the peg board were defined as the potential tool-tissue interaction areas, the features of which were extracted by the ORB (Oriented Fast and Rotated Brief) algorithm as raw data. GMM (Gaussian Mixture Model) was employed to cluster the raw data to obtain 12 central points, which were regarded as the centers of tool-tissue interaction areas.

B. Database and Procedures

A database for ring transfer task was collected in-house beforehand for parameter estimation and model training. In the database, a trial is a sequence of data performed by one subject during a specific task. Data for each trial includes both kinematics and stereo vision data.

Surgical tasks can be decomposed into basic rudimentary gestures that represent well-defined surgical motion unit, named surgemes [23]. Seven types of surgemes are defined for the ring transfer task (see Table I) in this letter. The database was manually annotated with the surgemes defined at the frame level. The kinematics data was labeled after the registration with the video data.

Suppose that there are K types of surgemes for a surgical task (e.g., $K = 7$ for ring transfer task). A N-class classification model can be trained with the labeled data based on LSTM method ($2 \leq N \leq K$). As shown in Table I, the precise degree ξ is defined under the condition of $N = 2, 3, 4$ respectively. Considering that the task requirements for the ring transfer task is relatively simple, the total number of precise degree $N = 2$ was used in this letter, with surgemes $S3$ and $S6$ defined as coarse surgemes ($\xi = 0$) and the remaining ones defined as precise surgemes ($\xi = 1$). A 2-class classification model for precise surgemes and coarse surgemes can therefore be trained with the labeled data, which can be loaded for online classification. Six critical areas ($q = 6$) are pre-defined (see Fig. 4). Other task-specific parameters $\gamma_{mo} = 0.1$, $\gamma_{do} = 0.1$ and $\delta_d = 10$ are

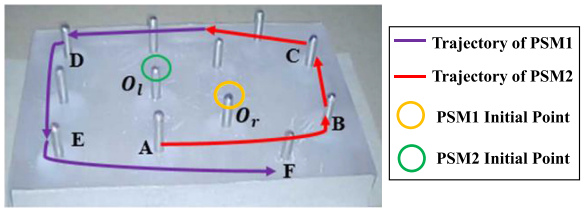

Fig. 4. Illustration of the ring transfer task.

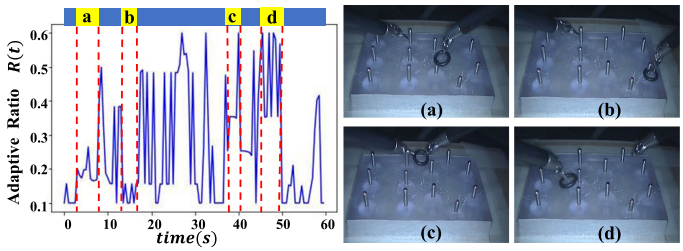

Fig. 5. An example of the evolution of the adaptive scaling motion ratio with typical surgemes during a trial. (a) S3: Move to Next Target with Right Hand (approach critical areas); (b) S2: Grasp Object with Right Hand (tool-tissue interaction); (c) S1: Transfer Object Operation (bimanual operation); (d) S6: Move to Next Target with Left Hand (far from critical areas).

chosen for the user studies. In this way, pre-operative data and the pre-trained model can be obtained beforehand. It is worth noting that the total level of precise degree, the number of critical areas and other parameters can be customized when transferred to other application scenarios.

For offline analysis, all the user-specific and task-specific parameters, as well as the prior probability of the FBN can be estimated through the database and fitted into the scaling ratio adjusting functions. For online calculation, the duration of the time segment $\sigma = 1$ s was used to update the operator' skill level, while task awareness can be achieved by estimating the real-time precise degree and monitoring the distance to the targeted area. All the variables required for FBN inference can be obtained to realize situation awareness when combining real-time visual cues with kinematic data of the robot. Flow charts of the detailed procedures of the offline analysis and the online calculation are shown in Fig. 1. Thus far, the self-adaptive motion scaling framework can be set up when all the information is fused together.

C. Participants and Tasks

Seven non-clinical participants (1 female and 6 males; aged 26.6 ± 2.7) were recruited in the user studies. Six of the participants are right-handed, while one of them is left-handed. Three of the subjects have teleoperation experience, and two of them are familiar with the dVRK. One of the participants has been using the dVRK system for four years, while the other one has six-month experience.

The ring transfer task was simplified so that all the subjects could finish the whole procedures with ease. PSM1 and PSM2 represent the right arm and the left arm of the slave robot, the initial positions of which are located at around at around O_r and O_l (see Fig. 4). As shown in Fig. 4, the whole procedures for the ring transfer task are described as follows:

1) Right Arm Movements: Control PSM1 to grasp the ring from A, then transfer and place it on B. Similarly, grasp the ring from B, then transfer and place it on C.
2) Bimanual Operation: Control PSM1 to grasp the ring from C, then pass the ring to PSM2, then control PSM2 to place the ring on D.
3) Left Arm Movements: Control PSM2 to grasp the ring from D, then transfer and place it on E. Finally, grasp the ring from E, transfer and place it on F.

All subjects had a practice session for basic skills acquisition. Those who had no experience on teleoperation were required to get familiar with all the surgemes of the ring transfer task by executing each gesture at least three times until they met the baseline proficiency to be included in the user studies. Then, all the qualified subjects were asked to go through the whole procedure twice to get accustomed to the experimental protocols.

During the formal user studies, the subjects were asked to perform the same task for three to five times by transferring the ring as quickly as possible with each of the two control modes, namely, via i) fixed mode, ii) adaptive mode. Thus, overall each participant performed the task for six to ten times in total. In order to remove outliers, the best trial and worst trial in terms of task completion time for each subject were abandoned. A total of 24 trials were collected. In order to mitigate the learning effects, the sequence in which the subjects performed the tasks via i) fixed mode, ii) adaptive mode was randomized. The participants were informed of the sequence after they finished all the trials.

IV. Results and Analysis

In this section, behavioral analysis is presented at first. Result analysis is conducted based on the operational performance and cognitive workload of the participants in the user studies.

A. Behavioral Analysis

A survey was conducted after the user studies to collect subjective feedback from the subjects. All the subjects agreed that they feel significantly more natural and comfortable when using the adaptive mode.

A typical example of the adjustment of the motion scaling ratio during an operation is shown in Fig. 5, the data of which lasted for one minute approximately, with 0.5 hz online acquisition rate. The range of the overall scaling ratio is 0.1–0.6. Compared to the period when a tool was far from critical areas, the range of the adaptive ratio was significant reduced when the tool approached critical area (Fig. 5(a) vs. Fig. 5(d)). It is evident that the range of the adaptive range of the scaling ratio of precise surgeme is much smaller than that of the coarse surgeme (Fig. 5(b) vs. Fig. 5(d)). The range of the scaling ratio of the bimanual operation motion is less than that of the single hand motion without tool-tool interaction (Fig. 5(c) vs. Fig. 5(d)).

Fig. 6 shows the trajectories of the master manipulators during operation with and without using the self-adaptive motion

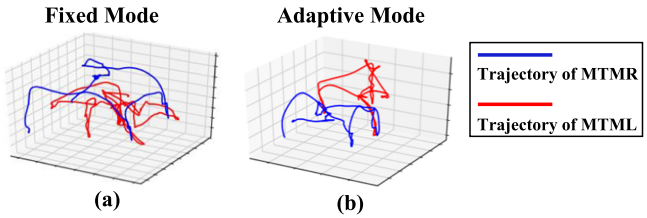

Fig. 6. Trajectories of the end-effectors of master manipulators (MTMR and MTML respectively) with (a) fixed mode, (b) adaptive mode.

scaling framework. MTMR and MTML represent the right and left arm of the master manipulators respectively. It can be seen that with the adaptive mode, the trajectories are more consistent, which demonstrates clear motion primitives during the operation. The trajectories are much more chaotic with the fixed mode, which indicates poor ergonomics for the teleoperation.

B. Quantitative Evaluation Metrics

The usability of the proposed framework is analyzed quantitatively through the following metrics:

1) Completion Time: Completion time measured in seconds (s) represents the length of the time for a subject to finish a single trial of the ring transfer task.

2) Master Robot Trajectory: The total path length of master robot end-effectors is measured over a single trial in meters (m).

3) Slave Robot Trajectory: The total path length of slave robot end-effectors is measured over a single trial in metres (m).

4) Control Efficiency: The ratio between the total path length of the slave robot end-effectors and that of the master robot. This metric can reveal how efficiently the operation's movements were mapped to the slave robot to fulfill the surgical procedures. The higher the control efficiency, the less the physical demand is required by an operator to manipulate the slave robot.

5) Clutching Number: The times of the user using the clutching mechanism to reposition the master manipulators.

6) Average Velocity: Average velocity measured in millimeter per seconds (mm/s) can be regarded as time-normalized instrument path length. It is obtained by dividing the slave robot end-effectors total path length measured over a single trial by the respective task completion time.

7) Cognitive Workload: National Aeronautical Space Agency-Task Load Index (NASA-TLX) questionnaire can be used to measure the subjects' cognitive workload for different mapping strategies by scoring six subjective subscales [24]. This subjective questionnaire is validated through the weighted score method. The average scores for the fixed mode and the adaptive mode for each subject are provided as the final results of cognitive workload comparison.

C. Performance Analysis

Normality tests (Shapiro-Wilk test) at 0.05 significance level were performed before subsequent statistical analysis. The obtained experimental data of evaluation metrics (completion time and master robot trajectory) reveal non-parametric nature,

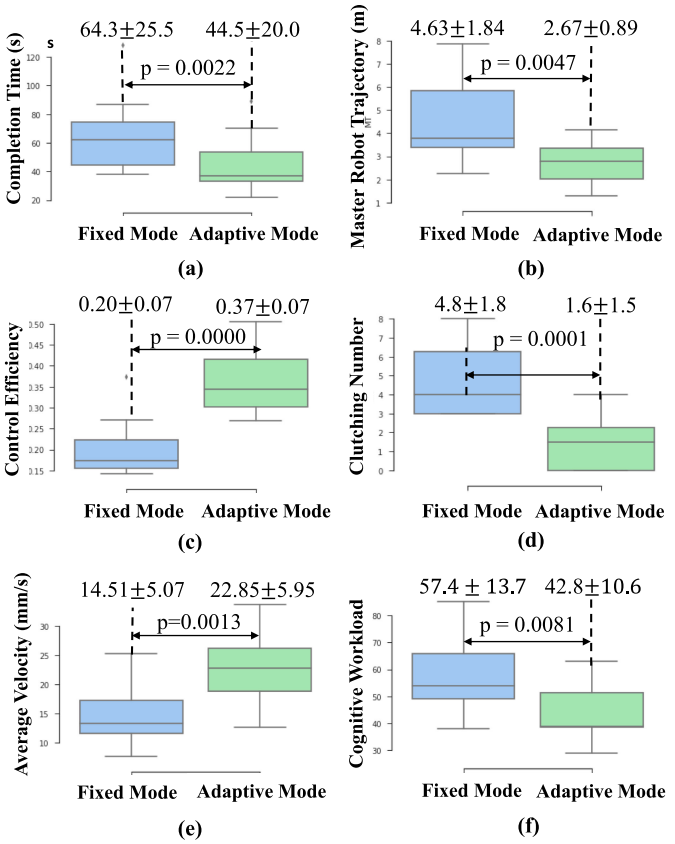

Fig. 7. Box plot results, including the comparisons in terms of (a) Completion Time; (b) Master Robot Trajectory; (c) Control Efficiency; (d) Clutching Number; (e) Average Velocity; (f) Cognitive Workload.

while the data of other metrics satisfy the normal distribution assumption.

The user study was a within-subject design, with all the subjects completing repetitions of two control modes. Therefore, Wilcoxon signed-rank tests were conducted for non-parametric statistical comparison between variables (completion time and master robot trajectory), while T-tests were conducted for the other metrics to justify the statistical differences between the two control modes. A p-value <0.05 is considered significant.

The total path length of the slave robot trajectory of all the subjects was calculated for the fixed mode and the adaptive mode, the average value of which was 0.851 m and 0.923 m respectively. One potential reason for the increased path length is that the users can operate with higher speed when performing coarse motions with the adaptive mode, which results in moving towards the next target without carefully selecting an optimal trajectory. However, the results are not statistically significant ($p = 0.425$).

Fig. 7 shows the results of comparison presented in the form of the box plot with evaluation metrics that have statistic differences ($p < 0.05$). Respective average values and standard deviation values are also included.

The completion time over the trials was significantly shorter for the adaptive mode compared to the fixed mode (44.5 s vs. 64.3 s, $p = 0.0022$). In terms of the average speed, the

subjects could complete the same procedures much faster with the adaptive mode than with the fixed mode (22.85 mm/s vs. 14.51 mm/s, $p = 0.0013$). Using the adaptive mode, reduced clutching was observed (1.6 vs. 4.8, $p = 0.0001$), which contributed to the smooth operational workflow. This makes a significance, as the subjects can improve the control efficiency (0.37 vs. 0.20, $p = 0.0022$), by reducing the total path length of the master robot (2.67 m vs. 4.63 m, $p = 0.0047$), due to the reduced requirement of repositioning. Moreover, the subjects' cognitive workload could be significantly reduced when using the adaptive framework ($p = 0.0081$), since the average weighted score values for the adaptive mode and the fixed mode are 42.8 and 57.4 respectively.

From the comparison of results between two different modes, it can be concluded that with the proposed self-adaptive motion scaling framework, the teleoperation efficiency and the overall operators' performance can be significantly improved.

V. Conclusion and Future Works

This letter introduces a novel approach to dynamically adjust the motion scaling ratio. By combining the visual cues and the motion data obtained from the slave robot, along with prior knowledge, operator skill assessment and task information, the system is capable of modifying the motion scaling accordingly.

Based on the results of user studies, the self-adaptive motion scaling framework is proved to have the ability to seamlessly adjust the scaling ratio, which results in improved efficiency of teleoperation. The effects of the adaptive scaling are noticeable by the subjects, because of the significantly reduced clutching, task completion time and total path length of the master robot. The improvements of the average control speed and the control efficiency are also evident.

The self-adaptive motion scaling framework can be readily adapted to other surgical tasks, provided that the surgemes and the corresponding precise degree are well-defined. Future work will include testing the self-adaptive motion scaling framework on different master-slave systems via different surgical tasks (e.g., knot-tying, needle passing and suturing). More objective evaluation metrics can be explored to improve the reliability of baseline skill assessment in more complicated surgical scenarios. Moreover, the effect of situation awareness can be significantly improved when the database is enlarged.

Acknowledgment

The authors would like to acknowledge G. Gras for discussions and all the subjects for taking part in the user studies.

References

[1] V. Vitiello, S.-L. Lee, T. P. Cundy, and G.-Z. Yang, "Emerging robotic platforms for minimally invasive surgery," *IEEE Rev. Biomed. Eng.*, vol. 6, pp. 111–126, 2013.

[2] K. Fujii, G. Gras, A. Salerno, and G.-Z. Yang, "Gaze gesture based human robot interaction for laparoscopic surgery," *Med. Image Anal.*, vol. 44, pp. 196–214, 2018.

[3] S. M. Prasad, S. M. Prasad, H. S. Maniar, C. Chu, R. B. Schuessler, and R. J. Damiano, Jr., "Surgical robotics: Impact of motion scaling on task performance," *J. Amer. College Surgeons*, vol. 199, no. 6, pp. 863–868, 2004.

[4] S. Ko *et al.*, "Intelligent control of neurosurgical robot mm-3 using dynamic motion scaling," *Neurosurgical Focus*, vol. 42, no. 5, 2017, Art. no. E5.

[5] R. Cassilly, M. D. Diodato, M. Bottros, and R. J. Damiano, Jr., "Optimizing motion scaling and magnification in robotic surgery," *Surgery*, vol. 136, no. 2, pp. 291–294, 2004.

[6] G. Gras *et al.*, "Implicit gaze-assisted adaptive motion scaling for highly articulated instrument manipulation," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2017, pp. 4233–4239.

[7] G. Gras, C. A. Seneci, P. Giataganas, and G.-Z. Yang, "Gaze-assisted adaptive motion scaling optimization using graded and preference based bayesian approaches," in *Proc. IEEE Int. Conf. Robot. Autom.* 2018, pp. 6425–6430.

[8] L. Zappella, B. Béjar, G. Hager, and R. Vidal, "Surgical gesture classification from video and kinematic data," *Med. Image Anal.*, vol. 17, no. 7, pp. 732–745, 2013.

[9] L. A. Zadeh *et al.*, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965.

[10] J. Martin *et al.*, "Objective structured assessment of technical skill (OSATS) for surgical residents," *Brit. J. Surgery*, vol. 84, no. 2, pp. 273–278, 1997.

[11] M. J. Fard, S. Ameri, R. D. Ellis, R. B. Chinnam, A. K. Pandya, and M. D. Klein, "Automated robot-assisted surgical skill evaluation: Predictive analytics approach," *The Int. J. Medical Robot. Comput. Assisted Surgery*, vol. 14, no. 1, p. e1850, Wiley Online Library, 2018.

[12] T. N. Judkins, D. Oleynikov, and N. Stergiou, "Objective evaluation of expert and novice performance during robotic surgical training tasks," *Surgical Endoscopy*, vol. 23, no. 3, p. 590–597, 2009.

[13] M. K. Chmarra, S. Klein, J. C. de Winter, F.-W. Jansen, and J. Dankelman, "Objective classification of residents based on their psychomotor laparoscopic skills," *Surgical Endoscopy*, vol. 24, no. 5, pp. 1031–1039, 2010.

[14] A. Ghasemloonia, Y. Maddahi, K. Zareinia, S. Lama, J. C. Dort, and G. R. Sutherland, "Surgical skill assessment using motion quality and smoothness," *J. Surgical Educ.*, vol. 74, no. 2, pp. 295–305, 2017.

[15] S. Balasubramanian, A. Melendez-Calderon, A. Roby-Brami, and E. Burdet, "On the analysis of movement smoothness," *J. Neuroengineering Rehabil.*, vol. 12, no. 1, pp. 112–122, 2015.

[16] Z. Wang and A. M. Fey, "SATR-DL: Improving surgical skill assessment and task recognition in robot-assisted surgery with deep neural networks," in *Proc. Annual Int. Conf. IEEE Eng. Medicine Biology Soc.*, vol. 2018, Jul. 2018, pp. 1793–1796, doi: 10.1109/EMBC.2018.8512575.

[17] S.-L. Lee *et al.*, "From medical images to minimally invasive intervention: computer assistance for robotic surgery," *Computerized Med. Imag. Graph.*, vol. 34, no. 1, pp. 33–45, 2010.

[18] L. Tao, E. Elhamifar, S. Khudanpur, G. D. Hager, and R. Vidal, "Sparse hidden Markov models for surgical gesture classification and skill evaluation," in *Proc. Int. Conf. Inf. Process. Comput.-Assisted Interventions*, 2012, pp. 167–177.

[19] R. DiPietro *et al.*, "Recognizing surgical activities with recurrent neural networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2016, pp. 551–558.

[20] N. Ahmidi *et al.*, "A dataset and benchmarks for segmentation and recognition of gestures in robotic surgery," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2025–2041, Sep. 2017.

[21] P. Kazanzides, Z. Chen, A. Deguet, G. S. Fischer, R. H. Taylor, and S. P. DiMaio, "An open-source research kit for the da Vinci surgical system," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2014, pp. 6434–6439.

[22] C. Chen, S. Stitt, and Y. F. Zheng, "Robotic eye-in-hand calibration by calibrating optical axis and target pattern," *J. Intell. Robot. Syst.*, vol. 12, no. 2, pp. 155–173, 1995.

[23] H. C. Lin, I. Shafran, D. Yuh, and G. D. Hager, "Towards automatic skill evaluation: Detection and segmentation of robot-assisted surgical motions," *Comput. Aided Surgery*, vol. 11, no. 5, pp. 220–230, 2006.

[24] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task load Index): Results of empirical and theoretical research," in *Advances Psychol.*, vol. 52. Amsterdam, The Netherlands: Elsevier, 1988, pp. 139–183.