2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Adaptive Motion Scaling for Robot-Assisted Microsurgery Based on Hybrid Offline Reinforcement Learning and Damping Control

Peiyang Jiang*, Wei Li, Yifan Li, Dandan Zhang*

Abstract— **Motion scaling is essential to empower users to conduct precise manipulation during teleoperation for robot-assisted microsurgery (RAMS). A constant, small motion scaling ratio can enhance the precision of teleoperation but hinder the operator from quickly reaching distant targets. The concept of self-adaptive motion scaling has been proposed in previous work. However, previous frameworks required extensive manual tuning of core parameters, which significantly depends on prior knowledge and may potentially lead to non-optimal solutions. This paper presents a hybrid offline reinforcement learning and damping control approach to regulate the motion scaling ratio for different operations during offline training. This method can take user-specific characteristics into consideration and help them achieve better teleoperation performance. Comparisons are made with and without using the adaptive motion-scaling algorithm. 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 time cost for task completion is significantly improved, while the variance of average speed and the total distance for robot operation is reduced.**

I. Introduction

Microsurgery [1], a discipline in surgery that involves the manipulation of microscale anatomical structures, presents significant challenges [2]. These include the need for steady-handed precision, fatigue management, and the necessity for specific, intuitive visual aids [3]. Robotic systems have been proposed to provide surgeons with enhanced capabilities, which can assist surgeons in performing challenging micro-surgical tasks accurately and safely [4]–[6]. Due to the difficulties of microsurgery, robot-assisted microsurgery (RAMS) has been developed to mitigate some of the challenges of microsurgery [7]. Leader-follower control is a prevalent form of control for RAMS [8]. In this approach, to achieve precise operations, the motions of the operator are captured by the leader controller, while these motions are replicated by the follower robot with a scaling ratio applied [9], [10].

Prasad et al. [11] highlighted the significant influence of motion scaling on teleoperation efficiency. Experimental evidence suggested that different motion mapping ratios significantly impact surgical precision, temporal efficiency, and coherence for teleoperation [11]. A small motion mapping ratio might impede long-range operations, necessitating repeated position recalibrations by the leading controller [12]–[14]. Conversely, a large motion coefficient might compromise accuracy during critical operations.

P. Jiang and Y. Li are with the Department of Engineering Mathematics, University of Bristol. W. Li is with the Department of Computing, Imperial College London. D. Zhang is with the Department of Bioengineering, Imperial College London. Corresponding: Dandan Zhang, d.zhang17@imperial.ac.uk

Different from the aforementioned static motion mapping ratios, an adaptive motion scaling concept has been proposed [15]–[18]. In [19], the motion scaling ratio is automatically adjusted to reduce the task completion time for a teleoperated robotic neurosurgical system. Comparisons are made among different modes through a user study involving five subjects. However, 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, where eye-tracking is used in conjunction with other sensing modalities to infer the intended position that the operator is trying to reach, allowing for the motion scaling ratio to be adjusted accordingly [20]. However, this technology is difficult to fully integrate into general surgical systems, and the effectiveness of gaze-assisted intention recognition might not be effective when operators wear glasses.

Zhang et al. [21] introduced a self-adaptive motion scaling framework for remote surgical robot control, where the motion scaling ratio was determined by three components: 1) situation awareness, 2) skill level awareness, and 3) task awareness. This framework has been proven to improve the performance of all subjects, as highlighted by significantly reduced clutching, task completion time, and total path length, as well as improvements in average control speed and control efficiency. However, this approach requires the tuning of many parameters, which consider only users' skill level without taking user-specific characteristics and preferences into account when updating the adaptive motion scaling policy.

Reinforcement Learning (RL) has been proven effective in decision-making for robotic surgery [22]–[25]. With an RL-based approach, the robotic system can pick up the operational features from different users automatically and adjust the control pattern to accelerate the trainees' progress. Therefore, we explore an RL-based approach to infer user-specific characteristics for developing the motion scaling ratio tuning policy, which eliminates the need for manual adjustment of the core parameters. However, using online human feedback directly as a reward function is prohibitively expensive for RL systems that require hundreds or thousands of hours of experience [26]. To practically train RL-based motion scaling ratio tuning policy without providing continuous human feedback, we employ offline RL in this paper.

Offline RL represents a learning paradigm that distills knowledge from pre-collected interaction datasets, which can eliminate the need for real-time environmental interactions with humans in the control loop to provide real-time feedback [27]. Offline RL becomes particularly useful in sce-

narios where real-time interactive training poses challenges [28]. For safety considerations, we integrate damping control into the framework to ensure safety [29], which forms the hybrid offline RL and damping control approach for adaptive motion mapping in RAMS.

The **contributions** of this paper include:

1) We propose a novel framework that integrates offline RL with damping control to enable adaptive motion scaling;

2) We conduct comprehensive user studies focused on two neurosurgical sub-tasks, which assess the efficacy of our proposed framework by comparing its performance in both fixed and adaptive motion scaling modes.

II. Methodology

A. Hardware System

The hardware system comprises a Sensapex four-axis micro-stage, which can be known as the micromanipulator for the microsurgical robotic system [8]. Two USB digital microscopes (Opti-Tekscope Digital USB Microscope) are used as the vision system. Different types of microsurgical tools can be mounted on the micromanipulator for the execution of different microsurgical tasks. For example, the Micro-25ga Tano polisher and the Micro-25ga maculorhexis forceps are used as microsurgical tools in this paper. While the former is built for needle insertion protocols, the latter is motorized with a Faulhaber motor to facilitate the grasping function of the micro forceps. The configuration of the hardware components of the microsurgical robotic system is illustrated in Fig. 1 (a).

We employ the HoloLens 2 (Microsoft) to provide an immersive experience for the operators when interacting with the robot. It integrates state-of-the-art hand-tracking and speech recognition technologies to enhance user engagement with the virtual content displayed in the optical see-through head-mounted Display. The advantage of using this system is to eliminate the need for a mechanical linkage-based motion tracking device, which can enhance the flexibility of the operator and facilitate a more instinctive control interface over the microsurgical robotic system. The embedded hand-tracking function can capture the leader's hand motion and map to the follower via a motion scaling ratio. The overview of the HoloLens-based human-robot interaction interface is shown in Fig. 1 (b).

B. Overview of Adaptive Motion Scaling Algorithm

The adaptive scaling ratio (factor) $\text{Scale}_{\text{ada}}$ is defined as: $\text{Scale}_{\text{ada}} = \text{Scale}_1 + \text{Scale}_2 + \text{Scale}_3$ (see Algorithm 1), which represents the sum of the damping-based scaling factor Scale_1, the base scaling factor Scale_2, and the offline RL-based scaling factor Scale_3, respectively (see Fig. 1 (c)).

In the initial step of our algorithm, we acquire a prior trajectory, which can also be known as expert demonstration data (denoted as T). This trajectory establishes a ground truth, encapsulating the optimal and desired operations. The acquisition of this trajectory is using the Dynamic Movement Primitives (DMP) method, In practical experiments and subsequent clinical trials, it is necessary to involve experienced surgeons with rich experience to participate in multiple experiments before the experimental operation. These surgeons' operational trajectories are then obtained, and these trajectories are used to generate ideal reference trajectories through the DMPs model. Based on the prior trajectory T, a damping gradient is set up. It implies that the closer the real-time movement is to the trajectory T, the slower the robot's motion. This is achieved by computing the deviation $(error_x, error_y, error_z)$ of the real-time position from the nearest point on T in x, y, z directions, and mapping these deviations to a damping-based scaling factor $scale1$. Specifically, the mapping function is given by:

$$scale1_x = \frac{10000}{1 + \exp(-|error_x|/500)} \tag{1}$$

$$scale1_y = \frac{10000}{1 + \exp(-|error_y|/500)} \tag{2}$$

$$scale1_z = \frac{10000}{1 + \exp(-|error_z|/500)} \tag{3}$$

Fig. 1 (d) illustrates the relationship between the motion mapping coefficient and the corresponding error. Notably, when the error increases, the rate of change in the scale diminishes, converging to a stable value. Such a design choice enhances the system's safety, mitigating potential risks associated with an overly magnified mapping coefficient, especially under conditions of significant error. A base scaling factor $scale2$ is set to a constant value. $scale2 = 6000$ is used in our paper. This value can be adjusted based on the characteristics of different RAMS.

To take user-specific characteristics and varying experience levels into consideration, $scale3$ is used as the component to further regulate the motion scaling ratio for different users, and ensure that their motion curves during operation are smooth, compliant, and safe. $scale3$ is derived from offline RL, which is detailed in Section II-C.

Algorithm 1 Adaptive Motion Scaling Algorithm

procedure AdaptiveMapping
  Initialize: Acquire expert demonstration data
  Generate desired trajectory T using DMP
  For each trial:
  Compute deviation $(error_x, error_y, error_z)$ from nearest point on T
  Compute $scale1$ values using Equation 1-3.
  Define base scaling factor: $scale2 = 6000$
  Learn adjustment factor $scale3$ through offline RL
  Compute adaptive scaling as:
    $\text{Scale}_{\text{ada}} = \text{Scale}_1 + \text{Scale}_2 + \text{Scale}_3$
end procedure

C. Offline Reinforcement Learning

RL is generally formalized as a Markov Decision Process (MDP): $< S, A, R, P, \gamma >$, where S represents the state space, R is the reward function and A is the action space, which is known by the agent. Consider an RL problem where at each time step t an agent performs an action

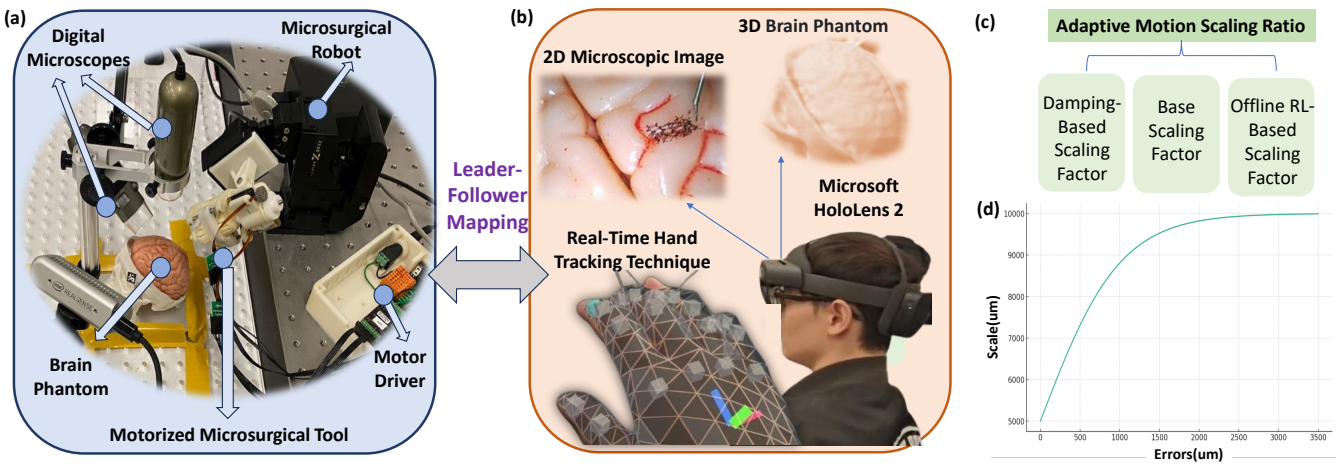

Fig. 1. Overview of the hardware system. (a) The configuration of the microsurgical robotic system. (b) Illustration of the HoloLens-based human-robot interaction interface. (c) Overview of the adaptive scaling algorithm. (d) Variation curve of the motion mapping ratio with respect to error.

$a_t \in A$, observes a new state $s_t \in S$, and receives a reward r_t. $P : S \times A \to S$ is the Markovian transition stochastic function, which defines the unknown dynamics of the environment. $\gamma \in [0, 1)$ is the discount factor, a parameter provided to the RL optimization algorithm favoring short-term rewards.

Due to the limitations of online RL, which requires time-consuming active engagement of agents, this paper employs offline RL to facilitate adaptive motion scaling. In this approach, an agent learns exclusively from previously collected data without the need for additional interactions. Pre-collected data are utilized as task-specific prior information. Initially, random actions are executed within the environment to accumulate data. Subsequently, batch-based learning is adopted during the training phase. Upon completion of the training, the refined model is deployed in the environment. For the development of the offline RL environment, we utilized the 'd3rlpy' library, a renowned open-source Python repository [30]. The goal of using offline RL is to maximize the reward based on user performance, by generating the desired action to regulate the motion scaling ratio.

1) State S: The observation in the RL process in this paper is a one-dimensional array consisting of eleven elements. These elements include the joint angles of the 4-DoFs microrobot (denoted by $j1$, $j2$, $j3$, $j4$), the 3D position of the hand in the predefined world coordinate (x, y, z, and rotation angle θ along the operator's arm), and the closest distance from the robot's end-effector to the ground truth trajectory (denoted by $error_x$, $error_y$, and $error_z$).

2) Action A: The action is a one-dimensional array that represents the motion scaling factor α. This array includes four elements: $s1$, $s2$, $s3$, and $s4$, which represent the motion scaling factors of the four joints of the microrobot respectively .

3) Reward R: The reward for each action can be derived from the combination of the typical evaluation metrics, used for obtaining the indicative performance of the real-time corrected trajectory.

a) Gracefulness (G): The measure of Gracefulness (G) is computed based on the curvature (κ) of a given trajectory. The median curvature along the trajectory generated during the time segment T_i is utilized to ascertain the measure of gracefulness. This is formally defined as $\boldsymbol{G} = \text{Median}(\log_{10}^{\kappa}(T_i))$, where κ is calculated using (4).

$$\kappa = \frac{||\dot{\boldsymbol{\lambda}}(t) \times \ddot{\boldsymbol{\lambda}}(t)||}{||\dot{\boldsymbol{\lambda}}(t)||^3} \tag{4}$$

Here, $\boldsymbol{\lambda}(t)$ denotes points in a 3D space as a vector; $\dot{\boldsymbol{\lambda}}(t)$ and $\ddot{\boldsymbol{\lambda}}(t)$ represent the instantaneous velocity and acceleration of the surgical tool tips, respectively.

b) Smoothness (S): The measure of Smoothness (S) is evaluated based on the rate of changes in acceleration, which is formally expressed as $\boldsymbol{S} = \text{Median}(\log 10^{\phi}(T_i))$, with ϕ calculated using (5).

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

Here, σ corresponds to the duration of the time segment which can be regulated by the real-time skill level updating rate, while v_p denotes the peak velocity.

c) Spectral Arc Length (SAL): This metric represents the arc length of the Fourier magnitude spectrum within an adaptive frequency range [31].

d) Average Velocity (A_v): This metric represents the average speed of the robot during a period of control, which is calculated by the total path length of the end-effector divided by the duration of time.

The equation for the reward is defined by (6).

$$\text{reward} = (-G - S) \times 15 + \left(\frac{A_v - 120}{50}\right) + \left(\frac{200 - \text{SAL}}{50}\right) \tag{6}$$

D. Comparisons between BCQ and CQL

We compare two popular offline RL algorithms: Batch-Constrained deep Q-learning (BCQ) [32] and Conservative Q-Learning (CQL) [33]. Batch-Constrained Deep Q-Learning (BCQ) addresses some of the inherent challenges in learning from fixed datasets. It ensures that the learned policy does not deviate drastically from the data-collection

policy, thereby avoiding regions of the state-action space where the Q-function might be inaccurately estimated due to lack of data. By constraining the policy in this manner, BCQ minimizes the extrapolation errors that are common in traditional deep Q-learning methods [32]. CQL mitigates this challenge by introducing a conservative mechanism into the Q-learning paradigm. It ensures that the estimated Q-values for unseen or less-frequented actions in the dataset are not over-optimistic [33]. It integrates a regularization term that penalizes excessive Q-value estimates beyond what the data supports.

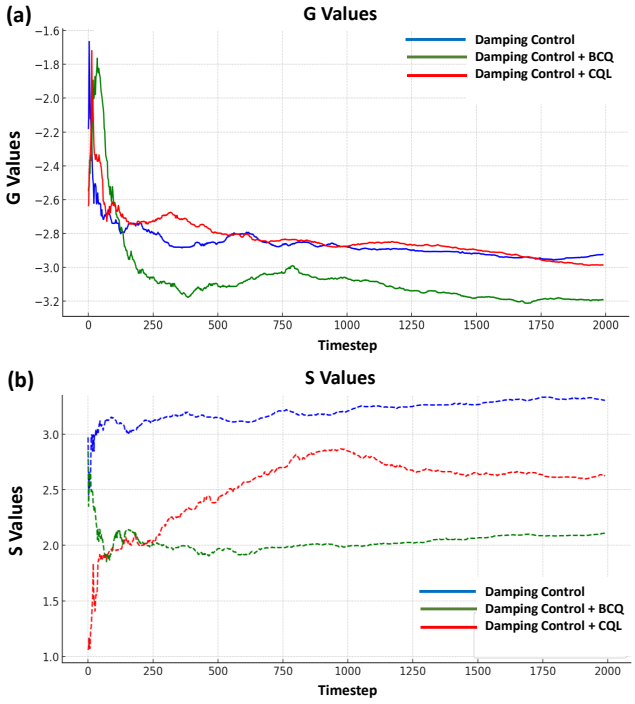

Fig. 2. Performance comparisons among i) damping control, ii) damping control + BCQ, iii) damping control + CQL for optimizing metrics (a) G and (b) S respectively.

We used BCQ and CQL offline RL to train the corresponding models, and then combine the same damping control with trained model to observe the experimental results. Fig. 2 visualizes the metric values of G and S for the same operator undertaking the same microsurgical task for comparisons. It can be seen that BCQ demonstrates superior performance to CQL. Additionally, the convergence time during BCQ's training is notably shorter than that of CQL. Consequently, the BCQ algorithm was chosen for subsequent user studies (see Algorithm 2).

Algorithm 2 Batch-Constrained Q-learning (BCQ)

Input: Offline dataset $\mathcal{D}$, Behavior policy π_b
Initialize:
Train a Variational Auto-Encoder (VAE) using $\mathcal{D}$ to learn distribution of actions for various states.
Initialize Q-function Q and target Q-function Q'.
Sample a batch of transitions (s, a, r, s') from $\mathcal{D}$.
Generate action a_{gen} from VAE using current state s.
Compute target value:
  $y = r + \gamma Q'(s', a_{gen})$
Update Q-function using MSE loss:
  $\mathcal{L}(Q) = \mathbb{E}_{(s,a,r,s')\sim\mathcal{D}}\left[(Q(s,a) - y)^2\right]$
Update Q-function parameters using gradient descent.
Periodically copy weights of Q to Q'.
Output policy:
Given a state s, the policy π generates an action a by sampling from VAE.

III. Empirical Validation and Results Analysis

A. Experiment Design and Participants

To assess the efficacy of the adaptive motion scaling mapping ratio in comparison to the fixed mapping ratio for operators, this study designed two microsurgical subtasks frequently employed in neurosurgery. The first task required participants to maneuver the robot's end effector along a simulated vascular pathway on a brain phantom and execute needle punctures at two specific locations. The second task involved using micro forceps to retrieve a micro-object from the brain phantom, simulating a biopsy procedure.

Seven participants (2 females and 5 males) were invited to join the user studies. They had different levels of teleoperation experience: one had experience of more than 1 year, another two had 6 months of experience, and the remaining participants had no experience. Before the experiment, each participant underwent identical training sessions, and their performance data were collected for offline RL model training. In the official experiment, participants were required to complete six trials: three employing a fixed motion mapping ratio and three utilizing an adaptive scaling motion mapping ratio. These trials were conducted in an alternating fashion to mitigate potential biases. In total, participants generated 54 trials of data for analysis in this comparative study. Data recorded for analysis included the robot's kinematic information, microscopic observations, needle puncture locations in Task 1, object retrieval outcomes in Task 2, and the time taken to complete each task. This comprehensive dataset was used to evaluate the performance differences between the two motion scaling strategies.

At the end of the experiment, participants were asked to fill out a questionnaire. The NASA-TLX [34] was used as the questionnaire to evaluate the mental (Q1), physical (Q2), and temporal (Q3) demands, as well as the frustration (Q4), effort (Q5), and performance (Q6) of the participants using the proposed framework, with scores ranging from 0 to 10.

Here is the list of evaluation metrics used to quantify participants' performance during user studies.

- **Variance of Speed (VarS):** VarS represents the variance of the robot's speed during the execution of a task (unit:$(\mu m/s)^2$)
- **Total Distance (TD):** TD represents the total distance covered by the microsurgical robot's end-effectors in a single trial of operation (unit: μm).
- **Time Cost (TC):** TC refers to the time required for a

TABLE I. Results based on quantitative and subjective evaluation of Task 1 (Average across trials)

	VarS$(\mu m/s)^2$	TD(mm)	TC(s)	Mean Error (μm)	Variance of Error $(\mu m/s)^2$	Q1	Q2	Q3	Q4	Q5	Q6
Fixed	34752.8	1316.6	139.8	2285.05	80.94	8.25	8.00	6.75	4.75	7.25	7.00
Adaptive	6213.2	411.5	99.7	50.87	18.60	5.50	4.75	3.79	2.50	2.75	2.5

participant to finish one trial of task (unit: s).

- **Mean Needle Insertion Error:** This is used to determine the distance deviation between the actual needle insertion point and the predefined position, as well as the variance among these deviations. It reflects the precision and repeatability of the needle insertion task (unit: μm, $(\mu m/s)^2$).

B. Results

1) Task 1: In the context of surgical robots, a lower VarS is indicative of smoother operations by the robot. This could potentially enhance the safety and reliability of surgical procedures. Furthermore, a reduction in TD and TC signifies that surgeons can perform surgeries more efficiently. As observed in Table I, compared to the fixed motion mapping coefficient, the results obtained using the adaptive coefficient indicate lower values for VarS and TC. The average TD covered by the microsurgical robot when utilizing a fixed motion mapping ratio is $1316.6mm$. In contrast, the adoption of an adaptive motion mapping ratio results in a significantly reduced average TD of $411.5mm$. Additionally, TC decreases from $139.8s$ to $99.7s$. These findings suggest that adaptive motion mapping ratios can enhance the efficiency and smoothness of microrobotic movements.

The Shapiro-Wilk test was employed to assess if the differences conformed to the assumption of normal distribution. If the p-value was less than 0.05, the data might not have adhered to the assumption of a normal distribution. Based on the Shapiro-Wilk test, all variables satisfied the normal distribution. For the fixed motion scaling mode, the p-values of Shapiro-Wilk test for VarS, TD, and TC are 0.572, 0.065, and 0.370, respectively. For the adaptive motion scaling mode, the p-values are 0.330, 0.114, and 0.312, respectively. The paired-sample t-test was used to analyze whether the results had significant differences. Fig.3 (a) shows the values for VarS, TD, and TC. The calculated p-values were 0.023, 0.027, and 0.031, respectively. Given a p-value threshold set at 0.05, it can be concluded that all three metrics exhibit statistically significant differences for Task 1.

Furthermore, the positions of the needle insertions during the experiment were recorded. Table I presents the mean and variance of the needle insertion error. It is evident that the trials using the adaptive motion mapping ratio showed smaller average error values and variance compared to those utilizing a fixed ratio. This indicates enhanced precision and repeatability brought by the adaptive motion scaling mode. In addition, Table I presents the scores from the NASA-TLX questionnaire for Task 1. Across all six questions, scores derived from using the adaptive motion mapping were lower than those using the fixed motion mapping coefficient, indicating the operators felt less burden when using the adaptive motion scaling mode.

TABLE II. Questionnaire result of Task 2 (Average Score)

	Q1	Q2	Q3	Q4	Q5	Q6
Fixed	6.20	8.20	7.60	8.75	7.60	7.60
Adaptive	3.40	4.60	5.94	6.39	4.20	4.80

TABLE III. Results based on quantitative of Task 2 (Average across trials)

	VarS$(\mu m/s)^2$	TD (μm)	TC (s)
Fixed	5392.18	233058.50	72.03
Adaptive	759.20	210348.8	29.33

2) Task 2: Table III and Fig. 3 (b) summarize the evaluation results for Task 2. Based on the results, the metrics obtained using the adaptive mode consistently exhibit lower values compared to those derived from the fixed mode. This indicates that for Task 2, employing the adaptive mode, as opposed to using a fixed one, allows operators to achieve smoother robot operations. It also enables faster task completion, thereby reducing operator fatigue. Similar to the statistical analysis conducted for Task 1, an assessment was made to determine if the experimental data under the conditions of VarS, TD, and TC conform to a normal distribution. For the fixed motion scaling mode, the p-values of Shapiro-Wilk test for VarS, TD, and TC are 0.0003, 0.153, 0.746, respectively. For the adaptive motion scaling mode, the p-values are 0.105, 0.439, 0.171, respectively. The results from the Shapiro-Wilk test indicate that the data for the variable VarS does not follow a normal distribution, whereas the remaining variables do. The Wilcoxon signed-rank test was applied to data that did not conform to a normal distribution, while the paired-sample t-test was employed for the rest. The resulting p-values are 0.594, 0.767, and 0.006, respectively. With a significance threshold set at 0.05, it can be inferred that there is a statistically significant difference between the fixed motion scaling mode and adaptive motion scaling scale for the TC metric. Table II displays the NASA-TLX scores for Task 2. For metrics Q1-Q6, the scores using the adaptive motion scaling are consistently lower than those using the fixed motion scaling ratio.

3) Discussions: The adaptive motion mapping framework, as evidenced by the results from the experiment, has showcased its potential to improve the teleoperation experience. The metrics VarS, TD, and TC indicate that the adaptive motion scaling algorithm offers a smoother and more efficient operation compared to the fixed motion scaling mode. However, it's worth noting that the impact of the motion mapping ratio is more evident in complex tasks. Compared to Task 1, the values of variables VarS and TD for Task 2 did not exhibit significant differences. A potential explanation is that Task 2 is simpler than Task 1. The operators worked over shorter durations and distances, where the effect caused by the motion scaling ratio is

Fig. 3. Visualization of the box plot results of Task 1 (a) and Task 2 (b), including the comparisons between Fixed Mode and Adaptive mode based on metrics VarS, TD, TC.

not significant in this case. It can be concluded that the adaptive motion mapping algorithm is particularly beneficial in intricate microsurgical procedures. Conversely, for less complex microsurgical tasks, the fixed motion scaling mode might be adequately effective.

Furthermore, regarding the VarS values for Task 2, increased variance can be observed in the initial performance of operators with no prior experience using MR devices. This heightened variance can be largely attributed to their lack of familiarity with the MR devices and initial nervousness during the operation, compared to more experienced operators. This observation emphasizes the importance of considering operator experience when evaluating the efficacy of motion mapping strategies in surgical robot operations.

IV. Conclusions and Future Work

This paper proposes an adaptive motion scaling algorithm that combines conventional damping control with an offline RL-based pretrained policy. User studies were conducted based on two microsurgical sub-tasks to evaluate the effectiveness of the proposed algorithm. Results show that with adaptive motion scaling, the performance of the participants improved significantly when the microsurgical tasks became more complex. A smoother trajectory was observed, and less total distance and time were required for task completion. The use of adaptive motion scaling can reduce surgeon fatigue and associated safety risks. Needle insertion experiments showed that adaptive motion mapping led to more accurate insertions, with reduced error and variance compared to the fixed ratio. NASA-TLX questionnaire scores for both microsurgical subtasks were consistently lower when using the adaptive motion mapping, suggesting improved task performance and reduced cognitive load on surgeons. Overall, the adaptive motion scaling control mode was found to be superior in terms of operational smoothness, precision, repeatability, and reduced cognitive load for surgeons compared to the fixed motion scaling control mode.

The proposed motion scaling algorithm can extend its applicability beyond RAMS to various types of biomedical robot control [35]. The leader-follower control system can be improved to facilitate more seamless control. For instance, integrating eye-tracking capabilities [36], such as those found in the HoloLens, could allow the system to predict and adapt to the user's intentions in real-time, thereby improving the intuitiveness of the control mechanism in a user-centric manner. Additionally, exploring the concept of human-robot shared control offers a promising avenue to augment the efficiency of remote microsurgical robot operations [37]. By merging the distinct advantages of human intelligence with robotic precision, a more effective operational synergy could be achieved. Furthermore, the incorporation of wearable haptic devices into the MR system presents an opportunity to further enhance the precision, safety, and efficiency of teleoperation [38], since providing operators with a tactile sense of their interactions is pivotal for delicate microsurgical tasks.

Acknowledgements

The University of Bristol Ethics Committee granted ethical approval (ref No.10389) before initiating the study. The authors would like to acknowledge all participants who contributed to the user studies.

REFERENCES

[1] S. Tamai, "History of microsurgery," *Plastic and Reconstructive Surgery*, vol. 124, no. 6S, pp. e282–e294, 2009.

[2] C. J. Payne, K. Vyas, D. Bautista-Salinas, D. Zhang, H. J. Marcus, and G.-Z. Yang, "Shared-control robots," *Neurosurgical Robotics*, pp. 63–79, 2021.

[3] P. D. Le Roux, H. Das, S. Esquenazi, and P. J. Kelly, "Robot-assisted microsurgery: a feasibility study in the rat," *Neurosurgery*, vol. 48, no. 3, pp. 584–589, 2001.

[4] D. Zhang, J. Chen, W. Li, D. B. Salinas, and G.-Z. Yang, "A microsurgical robot research platform for robot-assisted microsurgery research and training," *International journal of computer assisted radiology and surgery*, vol. 15, no. 1, pp. 15–25, 2020.

[5] O. M. Omisore, S. Han, J. Xiong, H. Li, Z. Li, and L. Wang, "A review on flexible robotic systems for minimally invasive surgery," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 1, pp. 631–644, 2020.

[6] K.-W. Kwok, H. Wurdemann, A. Arezzo, A. Menciassi, and K. Althoefer, "Soft robot-assisted minimally invasive surgery and interventions: Advances and outlook," *Proceedings of the IEEE*, vol. 110, no. 7, pp. 871–892, 2022.

[7] D. Zhang, W. Si, W. Fan, Y. Guan, and C. Yang, "From teleoperation to autonomous robot-assisted microsurgery: A survey," *Machine Intelligence Research*, vol. 19, no. 4, pp. 288–306, 2022.

[8] J. Lin, X. Guo, W. Fan, W. Li, Y. Wang, J. Liang, W. Liu, L. Wei, and D. Zhang, "Tims: A tactile internet-based micromanipulation system with haptic guidance for surgical training," *arXiv preprint arXiv:2303.03566*, 2023.

[9] S. C. Low and L. Phee, "A review of master–slave robotic systems for surgery," *International Journal of Humanoid Robotics*, vol. 3, no. 04, pp. 547–567, 2006.

[10] S. Parsa, H. A. Maior, A. R. E. Thumwood, M. L. Wilson, M. Hanheide, and A. G. Esfahani, "The impact of motion scaling and haptic guidance on operators' workload and performance in teleoperation," in *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 2022, pp. 1–7.

[11] 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," *Journal of the American College of Surgeons*, vol. 199, no. 6, pp. 863–868, 2004.

[12] D. Zhang, J. Liu, L. Zhang, and G.-Z. Yang, "Hamlyn crm: A compact master manipulator for surgical robot remote control," *International journal of computer assisted radiology and surgery*, vol. 15, pp. 503–514, 2020.

[13] W. Zhang, H. Cheng, L. Zhao, L. Hao, M. Tao, and C. Xiang, "A gesture-based teleoperation system for compliant robot motion," *Applied Sciences*, vol. 9, no. 24, p. 5290, 2019.

[14] D. Zhang, Y. Guo, J. Chen, J. Liu, and G.-Z. Yang, "A handheld master controller for robot-assisted microsurgery," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 394–400.

[15] 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 *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 6425–6430.

[16] S. A. Heredia-Pérez, K. Harada, M. A. Padilla-Castañeda, M. Marques-Marinho, J. A. Márquez-Flores, and M. Mitsuishi, "Virtual reality simulation of robotic transsphenoidal brain tumor resection: Evaluating dynamic motion scaling in a master-slave system," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 15, no. 1, p. e1953, 2019.

[17] F. Richter, R. K. Orosco, and M. C. Yip, "Motion scaling solutions for improved performance in high delay surgical teleoperation," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 1590–1595.

[18] S.-K. Hsia, Y.-H. Chuang, and C.-W. Chen, "Auto-tuned motion scaling in teleoperation based on human reaction model identification," *IEEE Robotics and Automation Letters*, vol. 7, no. 1, pp. 318–325, 2021.

[19] S. Ko, A. Nakazawa, Y. Kurose, K. Harada, M. Mitsuishi, S. Sora, N. Shono, H. Nakatomi, N. Saito, and A. Morita, "Intelligent control of neurosurgical robot mm-3 using dynamic motion scaling," *Neurosurgical focus*, vol. 42, no. 5, p. E5, 2017.

[20] G. Gras, K. Leibrandt, P. Wisanuvej, P. Giataganas, C. A. Seneci, M. Ye, J. Shang, and G.-Z. Yang, "Implicit gaze-assisted adaptive motion scaling for highly articulated instrument manipulation," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 4233–4239.

[21] D. Zhang, B. Xiao, B. Huang, L. Zhang, J. Liu, and G.-Z. Yang, "A self-adaptive motion scaling framework for surgical robot remote control," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 359–366, 2018.

[22] H. Su, Y. Hu, Z. Li, A. Knoll, G. Ferrigno, and E. De Momi, "Reinforcement learning based manipulation skill transferring for robot-assisted minimally invasive surgery," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 2203–2208.

[23] R. Zhu, D. Zhang, and B. Lo, "Deep reinforcement learning based semi-autonomous control for robotic surgery," *arXiv preprint arXiv:2204.05433*, 2022.

[24] S. Datta, Y. Li, M. M. Ruppert, Y. Ren, B. Shickel, T. Ozrazgat-Baslanti, P. Rashidi, and A. Bihorac, "Reinforcement learning in surgery," *Surgery*, vol. 170, no. 1, pp. 329–332, 2021.

[25] K. Fan, Z. Chen, G. Ferrigno, and E. De Momi, "Learn from safe experience: Safe reinforcement learning for task automation of surgical robot," *IEEE Transactions on Artificial Intelligence*, 2024.

[26] E. Ferreira and F. Lefevre, "Reinforcement-learning based dialogue system for human–robot interactions with socially-inspired rewards," *Computer Speech & Language*, vol. 34, no. 1, pp. 256–274, 2015.

[27] M. Yoo, S. Cho, and H. Woo, "Skills regularized task decomposition for multi-task offline reinforcement learning," *Advances in Neural Information Processing Systems*, vol. 35, pp. 37 432–37 444, 2022.

[28] R. F. Prudencio, M. R. Maximo, and E. L. Colombini, "A survey on offline reinforcement learning: Taxonomy, review, and open problems," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[29] F. Petit and A. Albu-Schäffer, "State feedback damping control for a multi dof variable stiffness robot arm," in *2011 IEEE international conference on robotics and automation*. IEEE, 2011, pp. 5561–5567.

[30] T. Seno and M. Imai, "D3rlpy: An offline deep reinforcement learning library," *J. Mach. Learn. Res.*, vol. 23, no. 1, jan 2022.

[31] S. Balasubramanian, A. Melendez-Calderon, and E. Burdet, "A robust and sensitive metric for quantifying movement smoothness," *IEEE transactions on biomedical engineering*, vol. 59, no. 8, pp. 2126–2136, 2011.

[32] S. Fujimoto, D. Meger, and D. Precup, "Off-policy deep reinforcement learning without exploration," in *International conference on machine learning*. PMLR, 2019, pp. 2052–2062.

[33] A. Kumar, A. Zhou, G. Tucker, and S. Levine, "Conservative q-learning for offline reinforcement learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1179–1191, 2020.

[34] S. G. Hart, "Nasa-task load index (nasa-tlx); 20 years later," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, no. 9. Sage publications Sage CA: Los Angeles, CA, 2006, pp. 904–908.

[35] D. Zhang, T. E. Gorochowski, L. Marucci, H.-T. Lee, B. Gil, B. Li, S. Hauert, and E. Yeatman, "Advanced medical micro-robotics for early diagnosis and therapeutic interventions," *Frontiers in Robotics and AI*, vol. 9, p. 1086043, 2023.

[36] S. Ban, Y. J. Lee, K. J. Yu, J. W. Chang, J.-H. Kim, and W.-H. Yeo, "Persistent human–machine interfaces for robotic arm control via gaze and eye direction tracking," *Advanced Intelligent Systems*, p. 2200408, 2023.

[37] D. Zhang, Z. Wu, J. Chen, R. Zhu, A. Munawar, B. Xiao, Y. Guan, H. Su, W. Hong, Y. Guo *et al.*, "Human-robot shared control for surgical robot based on context-aware sim-to-real adaptation," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 7694–7700.

[38] W. Fan, X. Guo, E. Feng, J. Lin, Y. Wang, J. Liang, M. Garrad, J. Rossiter, Z. Zhang, N. Lepora *et al.*, "Digital twin-driven mixed reality framework for immersive teleoperation with haptic rendering," *IEEE Robotics and Automation Letters*, 2023.