A Semi-Autonomous Control Mode for Flexible Steerable Intraluminal Platforms

Jose Fernando Gonzalez Herrera, Florent Nageotte, Philippe Zanne, Gianni Borghesan, Michel de Mathelin, Emmanuel Vander Poorten, Benoit Rosa

A semi-autonomous control mode for flexible steerable intraluminal platforms

Fernando Gonzalez-Herrera[1,2,*], Florent Nageotte[2], Philippe Zanne[2], Gianni Borghesan[3,4], Michel de Mathelin[2], Emmanuel Vander Poorten[3] and Benoit Rosa[2]

***Abstract*—Flexible steerable intraluminal robot platforms allow treatment and screening of colorectal cancer at an early stage, potentially reducing the associated incidence and mortality rates. Such robotic platforms often rely on a tree-like flexible architecture, with a flexible robotized body carrying both the endoscope camera and two robotized flexible surgical arms at its distal end. Telemanipulating these robotic platforms to correctly perform surgical tasks is technically difficult due to their kinematic complexity and the demanding nature of the task, which leads to potential interruptions in the surgical workflow.**

In this paper, a technique to efficiently control the arms and body and correctly perform complex surgical steps during the endoscopic submucosal dissection procedure is proposed. The technique, referred to as semi-autonomous arm-body control, is based on a quadratic programming controller. Custom-defined tasks synergistically control the arms and body, while avoiding unsafe positions for the arms.

Experiments in a mixed physical-simulated setup with eight users show an increased performance on the task and smoother movements compared to manual telemanipulation, at the expense of a slightly longer operating time. Further study will look at validating the approach in more realistic scenarios.

***Index Terms*—telemanipulation, robot-assisted surgery, endoluminal robotic surgery, optimization based control**

I. Introduction

By 2040 [1] colorectal cancer is predicted to increase to 3.2 million new cases and 1.6 million deaths from the 2020 estimates of 1.9 million cases and 930000 deaths. These numbers put colorectal cancer in the fourth place regarding incidence and mortality [2]. Most cases of colorectal cancers start as tumours on the inner lining of the colon or rectum. Prompt detection and removal of tumours can prevent deaths by anticipating further cancer metastasis. Once the tumour is detected, removal can be done by non-invasive procedures *e.g.* endoscopic mucosal resection, and endoscopic submucosal dissection (ESD). Due to the flexibility of removing malignant tissue of variable size, depth, and spread on the gastrointestinal walls, ESD is of special interest [3].

*This work was supported by the ATLAS project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813782. This work was also partially supported by French State Funds managed by the Agence Nationale de la Recherche (ANR), Grant ANR-18-CE19-0012 (MACROS) and Grant ANR-10-IAHU-02 (IHU-Strasbourg).

[1] ICube, UMR 7357 CNRS-University of Strasbourg, France, and also with Robot-Assisted Surgery Group, Mechanical Engineering, KU Leuven, Belgium.

[2]ICube, UMR 7357 CNRS-University of Strasbourg, France.

[3]Department of Mechanical Engineering, KU Leuven, Belgium.

[4]Core Lab ROB, Flanders Make, 3000 Leuven, Belgium.

* corresponding author : gonzalezherrera@unistra.fr

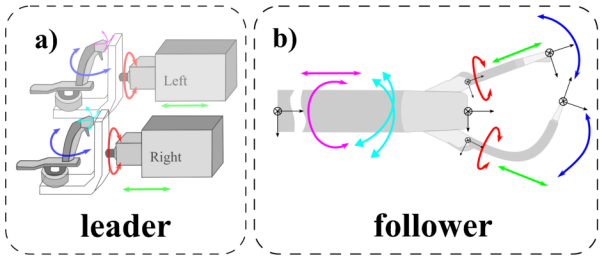

Figure 1. Overview of the STRAS robotic system. The corresponding DOFs of the leader and follower are matched by color. *a)*: leader controller interface. *b)*: follower and the corresponding coordinate frames.

Researchers have developed Flexible Steerable Intraluminal Robots (FSIR) [4]–[8] to help detect and remove malignant tissue. FSIR platforms are introduced via natural orifices. Access through natural orifices reduces the recovery time, body scars, chance of infection, and patient discomfort in comparison to interventions through small incisions or open surgery. These robotic platforms generally have a tree-like architecture, with multiple Degrees of Freedom (DOF). They consist of a main endoscope (body) equipped with a distal camera and with channels where flexible surgical instruments (arms) can be inserted (Fig. 1-b). Once the system has reached the surgical site, most of the tasks are done by the arms which are equipped with knives, snares, and grippers. FSIR platforms are typically telemanipulated, such that the users control the leader side (handles and joysticks, Fig. 1-a) in order to govern movements of the follower side (body and arms which are inside the patient, Fig. 1-b). Users manipulate the handles [9]–[12] to control the position of the tip of the arms, which carry the surgical tools. The arm controlled by the dominant hand of the user (referred to as the dominant arm), is usually used to perform complex tasks, such as marking, cutting, or dissecting tissues. The non-dominant arm most often performs secondary tasks such as grabbing, pushing, or pulling tissue. The leader-follower architecture allows the user to perform complex single-person surgical procedures, such as ESD [13].

During ESD, the user controls the FSIR platform [13]–[15] for realizing several surgical steps: marking the tumor, cutting the boundaries, dissecting the tissue, and removing the dissected specimen. The most complex and important task of ESD is the dissecting stage, where tissue is cut, and special care must be taken by the operator to avoid perforations [13]. The user performs these surgical steps multiple times, to entirely remove the tissue. As shown in Fig. 2, the non-dominant arm is used to grab the tissue and lift it, to provide

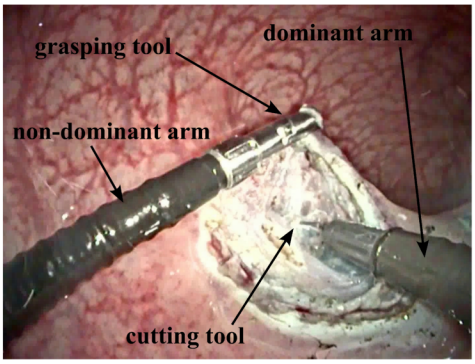

Figure 2. Surgical scenario. Dissection while performing endoscopic submucosal dissection with the STRAS platform.

sufficient space and visibility for cutting. The dominant arm is then used to cut the tissue while the non-dominant arm maintains its position to keep the tissue in place (Fig. 2). To complete the dissection, the user is required to perform these elementary surgical gestures multiple times, in order to completely cut the area of interest. Optimally, the user is required to have *traction* over the tissue [16], [17] while moving the body and cutting at a different tissue location. For establishing traction the non-dominant arm should be positioned such that the tissue is under tension. This helps precisely determine which portion of the tissue will be cut with the cutting tool. After the cut, the tissue may relax urging the non-dominant arm to shift position to re-establish traction again. Correctly performing this task requires the user to control the arms and the body, i.e. to perform coordinated control.

Coordinated control of arm and body movements is complex using manual telemanipulation of FSIR platforms. This is because, for a tree-like architecture, the arms can be moved directly, but also when moving the body, the arms will move. Hence, there is a coupling between the motions of the body and arms. When the user is grabbing tissue with the non-dominant arm and needs to perform a body movement, the user may have to release the tissue, move the body, and re-grab it before resuming the surgical task. In this scenario, the different subsystems are controlled sequentially. Alternatively, the user can compensate for the effect of the body movement on the arms position in real-time, which requires controlling the arms and body and accounting for their effects on each other. The former is detrimental for the patient as the surgical task is not performed optimally while the latter may lead to undesired movements or positions that might be detrimental for performing the surgical task.

The difficulty of the user to perform coordinated control interrupts the surgical workflow and hinders continuous motion. In some cases, the interruption may lead to intervention by an assistant or lead to technical failures [13] that could abort the current surgical procedure. This causes delays which can jeopardize patient safety or significantly prolong the operation time. Coordinated arm-body control by a robotic controller could provide a solution to this problem. Previous research on the control of the body of endoscopic platforms [18], [19] describes the usage of various human interfaces such as pedals [20], voice [21], gaze [22], and joysticks or buttons [23]. In systems with a tree-like architecture, similar approaches have been made using the user's feet [24] or the user's body movements [25] to command the body of FSIR platforms, with no consideration of arms position with respect to the surgical task. As far as the authors are aware, simultaneous coordinated control of the arm and body in FSIR has not yet been considered and tested.

In this paper, we propose an approach, referred to as semi-autonomous arm-body control mode, which performs coordinated control of the body and arms. When the semi-autonomous arm-body control mode is not active, the user is free to telemanipulate the arms. When a body movement is required, the user presses a pedal. The body movements are then controlled by the movements of the dominant arm while the pedal is pressed. During such body movements, the tip positions of the two arms can be automatically kept at a constant relative position to the tissue by the controller. This avoids unwanted tissue re-grabbing and unintentional build-up of stresses on manipulated tissue (and subsequent tissue damage). At the same time, the controller avoids unsafe arm positions (e.g. out of the field of view or too close to the camera). This solution is formulated as a high-level quadratic programming (QP) controller, implemented on the Single-access Transluminal Robotic Assistant for Surgeons (STRAS) robot. The proposed approach is tested with users on a simulated environment controlled by the physical leader side of the STRAS.

The paper is organized as follows: Section II describes the STRAS platform and the formalization of the positions of the arms end-effectors to be controlled. Section III states the problem and outlines the proposed approach. Then, Section IV describes the proposed optimization-based control method. Section V presents the experiments performed, while Section VI outlines the results which are discussed in Section VII. Finally, Section VIII summarizes the main conclusions and presents directions for future work.

II. STRAS PLATFORM

The STRAS robotic platform follows the FSIR architecture depicted in Fig. 1 and their main components are described in Fig. 3. The arms can bend, rotate, and translate, allowing the user to perform complex surgical tasks. The body is not equipped with any surgical tool but contains the camera at its tip. The body can rotate, insert or retract, and bend in one or two orthogonal directions. The complete platform features 10DOF, where each DOF of the follower is mapped to a DOF of the leader. The 3DOF of each arm are mapped to movements on the handles at the leader side, while the 4DOF of the body are mapped to joysticks at the top of the handles. Two pedals C_1 and C_2 are available on the system. C_2 can be pushed by the left foot of the user to activate the non-dominant arm end-effector and grasp a tissue, while the right pedal C_1 activates the electric cutting tool.

To describe the kinematics of the follower, a number of frames are defined as shown in the follower side of Fig. 1.

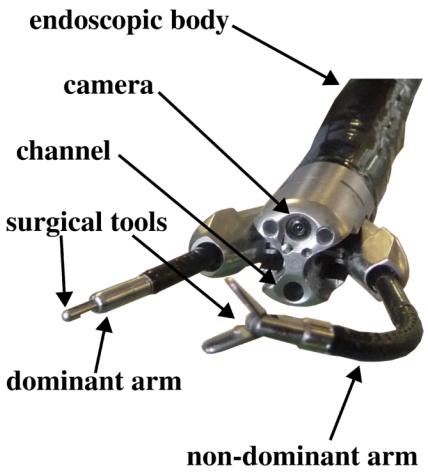

Figure 3. STRAS system components. The dominant and non-dominant arm is carried by the endoscopic body. A camera and an additional working channel are present at the distal end of the body.

The world frame $\{w\}$ is fixed w.r.t the patient. Coordinate frames are also attached to the tips of the endoscope $\{e\}$, the non-dominant $\{l\}$, and dominant arm $\{r\}$. The follower side is described with the following set of configuration variables:

$$\boldsymbol{q} = [\boldsymbol{q_r}, \boldsymbol{q_l}, \boldsymbol{q_e}]^T, \tag{1}$$

where $\boldsymbol{q_r}$ is the configuration of the dominant arm, $\boldsymbol{q_l}$ the one of the non-dominant arm, and $\boldsymbol{q_e}$ the one of the of the body. The model of the follower is then computed thanks to the approach described in [26], where each subsystem is represented as a rigid section followed by a flexible steerable segment described by a constant curvature model [27].

The forward kinematic model can then be described as:

$$\begin{aligned} \mathscr{M}: \mathbb{R}^{10} &\rightarrow SE(3), SE(3), SE(3) \\ \boldsymbol{q} &\mapsto {}_w\boldsymbol{T}^e(\boldsymbol{q_e}), {}_e\boldsymbol{T}^r(\boldsymbol{q_r}), {}_e\boldsymbol{T}^l(\boldsymbol{q_l}) \end{aligned} \tag{2}$$

in which ${}_w\boldsymbol{T}^e(\boldsymbol{q_e})$ is the pose of the endoscope tip $\{e\}$ in $\{w\}$, a function of the configuration variables $\boldsymbol{q_e}$. In the same manner, ${}_e\boldsymbol{T}^r(\boldsymbol{q_r})$ (respectively ${}_e\boldsymbol{T}^l(\boldsymbol{q_l})$) is the pose of the dominant (resp. non-dominant) arm tip $\{r\}$ (resp. $\{l\}$) in the body frame $\{e\}$.

The world frame pose of the arms can then be obtained by chaining the transformations:

$${}_w\boldsymbol{T}^r(\boldsymbol{q_e}, \boldsymbol{q_r}) = {}_w\boldsymbol{T}^e(\boldsymbol{q_e}) {}_e\boldsymbol{T}^r(\boldsymbol{q_r}) \tag{3}$$

$${}_w\boldsymbol{T}^l(\boldsymbol{q_e}, \boldsymbol{q_l}) = {}_w\boldsymbol{T}^e(\boldsymbol{q_e}) {}_e\boldsymbol{T}^l(\boldsymbol{q_l}) \tag{4}$$

which will be denoted ${}_w\boldsymbol{T}^r$, ${}_w\boldsymbol{T}^l$ for the sake of readability, omitting the configuration variables dependency from the transformation matrix. The tip position of the dominant arm in the global frame is described by the translational component of the transformation matrix ${}_w\boldsymbol{T}^r$ denoted ${}_w\boldsymbol{p}^r$. The position of the tip in the local body frame is extracted from ${}_e\boldsymbol{T}^r$ and denoted ${}_e\boldsymbol{p}^r$. The same is done for the non-dominant arm in each frame, ${}_w\boldsymbol{p}^l$ for ${}_w\boldsymbol{T}^l$ and ${}_e\boldsymbol{p}^l$ for ${}_e\boldsymbol{T}^l$. Following prior implementations [26], the jacobians that relate joint velocities to speeds of the tip of the arms and body are defined as $\boldsymbol{J}_l$, $\boldsymbol{J}_r$ and $\boldsymbol{J}_e$.

Once the kinematic model of the follower is established, it is necessary to link the leader configuration variables and the follower. The leader configuration variables are defined as:

$$\boldsymbol{\mathscr{q}} = [\boldsymbol{\mathscr{q}_r}, \boldsymbol{\mathscr{q}_l}, \boldsymbol{\mathscr{q}_e}]^T. \tag{5}$$

In manual teleoperation of the STRAS platform a mapping function $\boldsymbol{g}(\cdot)$ establishes the correspondence between $\boldsymbol{\mathscr{q}}$ and $\boldsymbol{q}$. Arms are controlled in position at joint level by a simple proportional mapping:

$$[\boldsymbol{q_r}, \boldsymbol{q_l}]^T = K[\boldsymbol{\mathscr{q}_r}, \boldsymbol{\mathscr{q}_l}]^T, \tag{6}$$

where K is a constant diagonal matrix that ensures a correspondence between the range of motion of the corresponding leader and follower DOF, following the formalization defined in [26]. The body is rate-controlled via the joysticks on the handle, such that:

$$\dot{\boldsymbol{q}}_{\boldsymbol{e}} = \alpha \boldsymbol{\mathscr{q}_e} \tag{7}$$

where α is a proportionality constant. Combining Eqs. (2) to (4) and Eqs. (6) and (7) provides the full model of the STRAS robot in manual teleoperation mode. Because of hardware limitations and safety concerns, speeds for the arms (Eq. (6)) and the body (Eq. (7)) are bound during teleoperation such that: $\dot{\boldsymbol{q}}^{\,h}_{min} < \dot{\boldsymbol{q}} < \dot{\boldsymbol{q}}^{\,h}_{max}$.

III. Problem Statement

In this paper, we focus on dissection, which is the most challenging part of ESD. To perform it, the clinician has to grab the tissue with its non-dominant arm and lift it to a desired position so as to expose the part to be dissected, having traction over the tissue. Then, the user performs the dissection by cutting the tissue with an electric cutting tool. We model this task using positions of the left and right arm, assuming without loss of generality that the dominant arm is the right arm. Tip position of the dominant and non-dominant arms are described by ${}_w\boldsymbol{p}^r$, ${}_w\boldsymbol{p}^l$ as defined by Eqs. (3) and (4). The envisioned scenario is as follows: The end-effector of the non-dominant arm grabs the tissue and lifts it to a desired position ${}_w\boldsymbol{p}^l_d$, setting ${}_w\boldsymbol{p}^l = {}_w\boldsymbol{p}^l_d$ (Fig. 4-a). The user then cuts the tissue, which we model as the tip position of the dominant arm reaching two successive positions ${}_w\boldsymbol{p}^r_{d1}$ and ${}_w\boldsymbol{p}^r_{d2}$. Reaching ${}_w\boldsymbol{p}^r_{d1}$ or ${}_w\boldsymbol{p}^r_{d2}$ may require moving the body, by controlling its tip position ${}_w\boldsymbol{p}^e$. During such movements, and more generally during the whole cutting phase, the non-dominant arm tip position ${}_w\boldsymbol{p}^l = {}_w\boldsymbol{p}^l_d$ should be kept steady, maintaining traction (Fig. 4-b,c). In this paper, we first tackle the correct performance of dissection in ESD. A correct dissection requires controlling the 10DOF of the arms and body simultaneously.

A second problem is regarding patient safety during the operation. Blindly manipulating the arms or interfering with the surgical site view should be avoided. In our case, this means that, during repositioning, arms should be kept in sight and avoid being brought too close to the camera frame (which is also the body frame $\{e\}$), so as to avoid occluding the view of the surgical site. In addition, it is also desirable to avoid

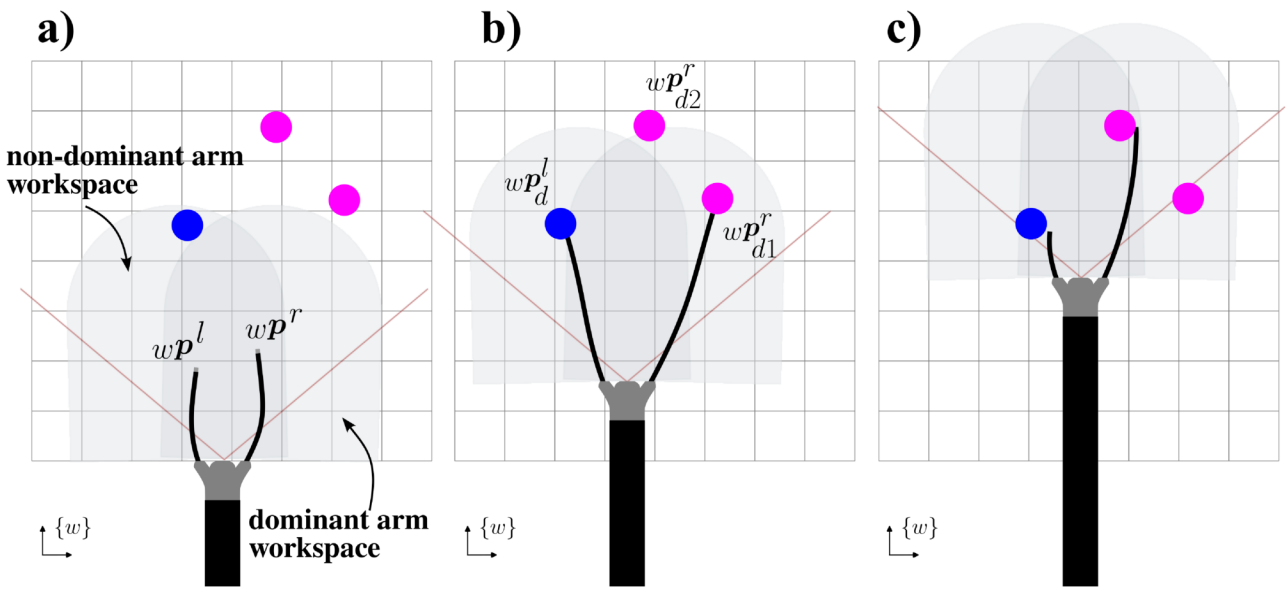

Figure 4. Overview of the problem. Workspace of the arms is depicted in grey and moves as the body moves, dark red line depicts the bounds of the field of view. a) Position of the body allows to reach the target of the non-dominant arm $^w\boldsymbol{p}^l_d$ but not the target of the dominant arm $^w\boldsymbol{p}^r_{d1}$. b) Body is moved to reach $^w\boldsymbol{p}^l_d$ and the new body position allows to reach $^w\boldsymbol{p}^r_{d1}$ but not $^w\boldsymbol{p}^r_{d2}$. c) While $^w\boldsymbol{p}^l_d$ is not violated, the second target position of the dominant arm $^w\boldsymbol{p}^r_{d2}$ is reached while moving the body.

bringing the system close to joint limits in order to avoid the degradation of the user's motion capability.

IV. Semi-autonomous control

The proposed approach involves the triggering of autonomous motions of the arms end-effectors and the body to allow coordinated control. Following the modelling presented in the previous section, an overview of the proposed approach is presented in Section IV-A. The problem set as control objectives is described in Section IV-B. An optimization-based formulation of the problem is presented in Section IV-C covering the objectives and constraints.

A. Overview of the approach

To best describe our proposed approach, let us first consider a situation where no body movement is required. In this case, the user telemanipulates the dominant and non-dominant arms as in standard telemanipulation, the system being then governed by Eqs. (2) to (4) and Eqs. (6) and (7).

To move the body, the user presses the pedal C_1 (see Fig. 5). The relation between leader side $\boldsymbol{q_r}$, $\boldsymbol{q_l}$ and follower side $\boldsymbol{q_r}$, $\boldsymbol{q_l}$ is then disconnected, and the semi-autonomous arm-body control mode takes control of the 10DOF of the follower (note that in this mode, the joystick is not used). The movements of the body are then controlled by the movements of the user telemanipulating the dominant arm handle on the leader interface, as detailed hereafter.

First, using Eq. (2) with $K\boldsymbol{q}$ as an input, one can compute the tip position $_e\boldsymbol{p}^r_v$ of a *virtual* dominant arm that would have resulted from the handle motion. $_e\boldsymbol{p}^r_v$ is, in turn, used to compute a desired body speed $\dot{\boldsymbol{q}}_v$ [1]:

$$\dot{\boldsymbol{q}}_v = k_e \boldsymbol{J}_e^{\dagger}({}_e\boldsymbol{p}^r_v - {}_e\boldsymbol{p}^r) \tag{8}$$

where k_e is a scalar gain to match manual telemanipulation speeds and $\boldsymbol{J}_e^{\dagger}$ is the Moore-Penrose pseudo inverse of the estimated jacobian $\boldsymbol{J}_e$. In other terms, the user controls the body movements by controlling the tip of a virtual arm, which is the dominant arm of the user (shown in semi-transparent green in Fig. 5).

As described in Section III, body movements impact the tip position of the arms with respect to the surgical site and such effect should be compensated. We therefore add three objectives, which will be formulated as control objectives and solved using a QP formulation in the next subsections.

First, the dominant arm should not move from its world position $\{w\}$ during body movements, such that $_w\dot{\boldsymbol{p}}^r = 0$. In other words, the tip dominant arm position of the leader $_e\boldsymbol{p}^r_v$ moves the body, while the follower dominant arm $_e\boldsymbol{p}^r$ and body $_w\boldsymbol{p}^e$ compensate, so as to keep the world position $_w\boldsymbol{p}^r$ still during body movement avoiding interfering with the surgical task as little as possible.

The non-dominant arm behaviour during body movements, however, depends on the surgical step. In this work, we assume that the user is performing a dissection task. In this task, the user grasps the tissue with the non-dominant arm and lifts it to a desired position $_w\boldsymbol{p}^l_d$. Such that grasping is activated by pressing pedal C_2 in the STRAS system. Therefore, if C_2 is not activated, the non-dominant arm can move together

[1]The endoscope body having 4 DOFs, there is a redundancy for the displacement of the body. The rotation of the endoscope body being an unusual and possibly disturbing motion of the user because of the resulting rotation of the camera, it is here not actuated and the last component of $\dot{\boldsymbol{q}}_v$ is zero.

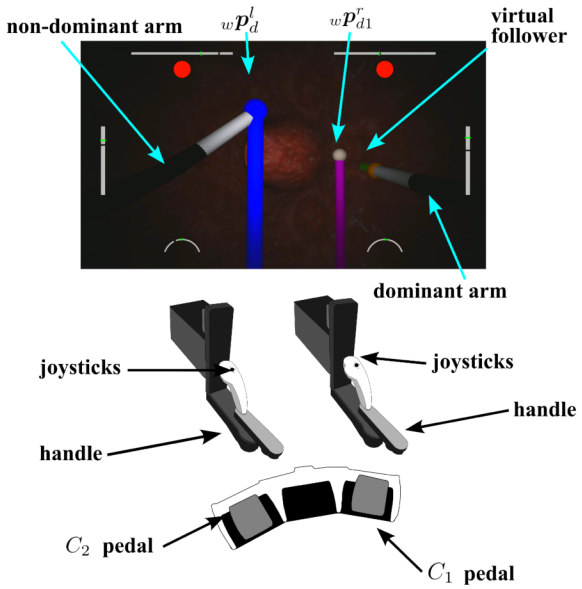

Figure 5. Overview of the main components of the simulator setup. Reaching targets are depicted in blue and pink for the non-dominant and dominant arms. Follower in black with a grey tip, virtual follower in semi-transparent green.

with the body since it is not interacting with tissue, reducing constraints on the control system. If pedal C_2 is active a tissue is being grabbed. It is then paramount to maintain the position of the non-dominant arm during body movements, i.e. keeping ${}_w\dot{\boldsymbol{p}}^l = 0$. This task is particularly challenging to perform manually and is carried out automatically by the control system during body movements using the semi-autonomous arm-body control approach.

Finally, we also consider safety. The semi-autonomous arm-body controller should not result in the arms being placed in an unsafe position, i.e. outside of the field of view, too close to the body (which holds the camera), or close to joint limits.

Given the described behaviours, the semi-autonomous approach allows to: *i)* have free control of the arms when no body movements are required, *ii)* control the body with the dominant arm, so as to re-position the endoscope with respect to the surgical site –by anchoring the dominant arm position in the world frame, and *iii)* correctly perform the most demanding surgical task –dissection– in which the position of the arms end-effectors should be maintained while the body is moved, in order to efficiently reach subsequent cutting locations. This three-way behaviour allows to assist the user not only in the most demanding stage, which is dissection, but also during the other stages of the ESD. The proposed approach could also be used to move the body while marking the tissue and simplifying the task by maintaining the world position of the dominant arm while accounting for safety constraints.

B. Formulation of the control problems as objectives

In this section, we formulate the control problem previously described (Section III) as high-level objectives. Keeping the absolute position of the dominant arm still, or ${}_w\dot{\boldsymbol{p}}^r = 0$, is expressed by:

$$\mathscr{H}_1 = \min \left\| {}_w\boldsymbol{p}^r_0 - {}_w\boldsymbol{p}^r \right\| \tag{9}$$

where the subscript $_0$ refers to the position at the beginning of the body movement (i.e. when the pedal C_1 is pressed). Similarly,

$$\mathscr{H}_2 = \min \left\| {}_w\boldsymbol{p}^l_0 - {}_w\boldsymbol{p}^l \right\| \tag{10}$$

for when the absolute position of the non-dominant arm is required to be maintained —during dissection after grabbing the tissue. The objective $\mathscr{H}_2$ is active only when the C_2 pedal is active, that is when the non-dominant arm is grabbing – keeping traction over– the tissue.

Secondly, the dominant arm is set to be in a safe position w.r.t. the body frame ${}_e\boldsymbol{p}^r_d$. The translational component $\boldsymbol{q_r}|_{trans}$ of the configuration variables of the dominant arm in Eq. (9), defines a safe position as:

$$\lambda^-_{trans} \leq \boldsymbol{q_r}|_{trans} \leq \lambda^+_{trans} \tag{11}$$

which effectively repels joint limits. Where λ^-_{trans} is the lower bound and λ^+_{trans} is the upper bound. The position of the dominant arm in the image is given by its projection, $img(\cdot)$. The arm defines a safe position as those in which the tip of the dominant arm is visible, which in horizontal component of the image $img({}_e\boldsymbol{p}^r)|_x$ is given by:

$$\lambda^-_x \leq img({}_e\boldsymbol{p}^r)|_x \leq \lambda^+_x \tag{12}$$

and in its vertical component $img({}_e\boldsymbol{p}^r)|_y$ as

$$\lambda^-_y \leq img({}_e\boldsymbol{p}^r)|_y \leq \lambda^+_y. \tag{13}$$

When the dominant arm leaves the previously defined safe zone in the image or is close to the translational bounds, the arm is returned inside the safe bounds while the body and arms are being controlled by two objectives. A simplified overview of the safe zone in which the dominant arm should be kept is depicted in Fig. 6. While the body is moving, the dominant arm avoids leaving the safe zone, by Eqs. (11) to (13). Maintaining the arm in a safe position is obtained by setting the following objective:

$$\mathscr{H}_3 = \min \left\| {}_e\boldsymbol{p}^r_d - {}_e\boldsymbol{p}^r \right\| \tag{14}$$

in the Cartesian space, where the goal position ${}_e\boldsymbol{p}^r_d$ is the centroid of the workspace. The arm should also remain inside the field of view of the endoscopic camera, which can be obtained by setting the following objective:

$$\mathscr{H}_4 = \min \left\| img({}_e\boldsymbol{p}^r_d) - img({}_e\boldsymbol{p}^r) \right\| \tag{15}$$

with $img({}_e\boldsymbol{p}^r_d)$ being the image center. When the arm is in a safe position the objectives $\mathscr{H}_3$, $\mathscr{H}_4$ are deactivated.

The objectives defined previously are activated depending on the states of the pedals. When no pedal is active, the body is static and the user has full range control of the arms by moving the leader console handles. Pressing C_1 activates $\mathscr{H}_1$, and if the dominant arm is in an unsafe position $\mathscr{H}_3$ and $\mathscr{H}_4$ become active such that the dominant arm keeps its absolute position, away from the joint limits and in a visible location, while the body moves. When C_2 is pressed, $\mathscr{H}_2$ is active.

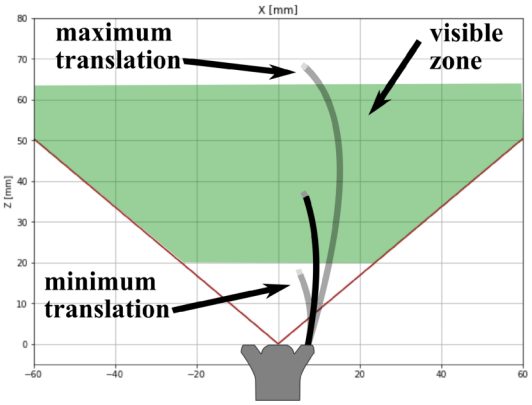

Figure 6. Overview of the defined safe zones for the dominant arm. The green area is a simplification of the safe zone. This safe zone is visible and is within the desired translation range.

C. Optimization-Based High Level Control

In order to deal with the multiple objectives we adopt a quadratic form of the objectives defined in Section IV-B. The optimization objectives are in the form $\min \|p^* - p\|$, which requires to minimize the difference between p^* and p. This cannot be done in a single step when the error becomes large, instead, at each time step an intermediate objective is required to be minimized as:

$$\min \|k(p^* - p)\|_2 \,. \tag{16}$$

Each of the optimization objectives takes the form of Eq. (16) to then be formulated in quadratic form. The quadratic form writes:

$$\min_{\dot{\boldsymbol{q}}} \quad \frac{1}{2}\, \dot{\boldsymbol{q}}^{\,T} \boldsymbol{H}_t \dot{\boldsymbol{q}} + \boldsymbol{c}_t^T \dot{\boldsymbol{q}} \tag{17}$$

$$s.t. \quad \dot{\boldsymbol{q}}^{\,+} \leq \dot{\boldsymbol{q}} \leq \dot{\boldsymbol{q}}^{\,-} \tag{18}$$

where $\boldsymbol{H}_t = \sum \gamma_i \boldsymbol{H}_i$ and $\boldsymbol{c}_t = \sum \gamma_i \boldsymbol{c}_i$. $\boldsymbol{H}_i$ is the Hessian matrix associated with a given objective $\mathscr{H}_i$ while $\boldsymbol{c}_i$ is its constant vector for each of the i previously defined objectives (Eqs. (9), (10), (14) and (15)), and γ_i is a given objective weight.

The intermediate step defined in Eq. (16) is reformulated using the quadratic form with the constant $\boldsymbol{c}_i$ being defined as $\boldsymbol{c}_i = -k\boldsymbol{J}_i^{\,T}(\boldsymbol{p}^* - \boldsymbol{p})$ where $\boldsymbol{J}_i$ is the task Jacobian associated with objective i. The approximated Hessian matrix is computed by $\boldsymbol{H}_i = \boldsymbol{J}_i^{\,T}\boldsymbol{J}_i$. This representation is used by the solver [28] to solve for a vector of optimization variables $\dot{\boldsymbol{q}}_{qp}$ that quadratically reduce the errors in position, and is subject to linear inequality constraints, Eq. (18).

The positions of the arms end-effectors are set as output variables. The output variables to be regulated are: the position of the dominant arm $\{r\}$ w.r.t. the endoscope base frame $\{e\}$, also $\{r\}$ w.r.t. the world base $\{w\}$, and the position of the non-dominant arm, $\{l\}$ in $\{w\}$, grouped in:

$$\boldsymbol{p} = \left[{}_e\boldsymbol{p}^r, {}_w\boldsymbol{p}^r, {}_w\boldsymbol{p}^l\right]^T, \tag{19}$$

while the input variables $\dot{\boldsymbol{q}}$ are the commanded velocities to the arms and body, as in Eq. (1), such that

$$\dot{\boldsymbol{q}}_{qp} = \left[\dot{\boldsymbol{q}}_{\boldsymbol{r}}\,, \dot{\boldsymbol{q}}_{\boldsymbol{l}}\,, \dot{\boldsymbol{q}}_{\boldsymbol{e}}\right]^T \tag{20}$$

is the result of solving Eq. (17) and it is added to the body movements defined in Eq. (8) as:

$$\dot{\boldsymbol{q}} = \dot{\boldsymbol{q}}_{qp} + \left[\boldsymbol{0}_{1\times 3}, \quad \boldsymbol{0}_{1\times 3}, \quad \dot{\boldsymbol{q}}_v\right]^T. \tag{21}$$

Given the desired output, we can use a single generalised Jacobian matrix for all objectives, which relates actuation and output space variables, and is a 10×9 Jacobian matrix defined as:

$$\boldsymbol{J} = \begin{bmatrix} {}_e\boldsymbol{J}_r & 0 & 0 \\ {}_e\boldsymbol{J}_r & {}_w\boldsymbol{J}_e & 0 \\ 0 & {}_w\boldsymbol{J}_e & {}_e\boldsymbol{J}_l \end{bmatrix}, \quad \dot{\boldsymbol{p}} = \boldsymbol{J}\,\dot{\boldsymbol{q}}\,. \tag{22}$$

Such that $\boldsymbol{J}$ is computed at each time step following the formulation presented in [26] for each of the subsystem Jacobians. Note that J is a generalized Jacobian linking $\dot{\boldsymbol{q}}_{qp}$ to $\dot{\boldsymbol{p}}$. It is therefore populated differently depending on the active objectives at a given time, yielding the corresponding $\boldsymbol{J}_i$: objective $\mathscr{H}_1$ acts on the dominant arm ($\dot{\boldsymbol{q}}_{\boldsymbol{r}}$) and the body ($\dot{\boldsymbol{q}}_{\boldsymbol{e}}$), $\mathscr{H}_3$ and $\mathscr{H}_4$ act only on the dominant arm ($\dot{\boldsymbol{q}}_{\boldsymbol{r}}$) such that, Eq. (22) is populated until on the last three rows. When C_2 is pressed and $\mathscr{H}_2$ is active, the Jacobian $\boldsymbol{J}$ is populated completely since $\mathscr{H}_2$ acts on $\dot{\boldsymbol{q}}_{\boldsymbol{l}}$.

The optimization variables $\boldsymbol{q}$ should stay within ranges $[{}_i\boldsymbol{q}_{min}, {}_i\boldsymbol{q}_{max}]$. This is achieved by introducing lower and upper bounds:

$$\dot{\boldsymbol{q}}^{\,\diamond}_{-} = \delta({}_i\boldsymbol{q}_{min} - \boldsymbol{q}) \tag{23}$$

$$\dot{\boldsymbol{q}}^{\,\diamond}_{+} = \delta({}_i\boldsymbol{q}_{max} - \boldsymbol{q}) \tag{24}$$

depending on the distance of $\boldsymbol{q}$ to ${}_i\boldsymbol{q}_{min}$ and ${}_i\boldsymbol{q}_{max}$, with δ being the convergence factor (set to 0.1) as introduced in [29]. Eq. (18) uses the minimum value for the lower bound between the hardware limits $\dot{\boldsymbol{q}}^{\,h}_{min}$ and the joint limits $\dot{\boldsymbol{q}}^{\,\diamond}_{-}$ presented in Eq. (23), such that the minimum speed is given by $\dot{\boldsymbol{q}}^{\,+} = \min(\dot{\boldsymbol{q}}^{\,\diamond}_{-}, \dot{\boldsymbol{q}}^{\,h}_{min})$. The upper bound is obtained similarly as $\dot{\boldsymbol{q}}^{\,-} = \min(\dot{\boldsymbol{q}}^{\,\diamond}_{+}, \dot{\boldsymbol{q}}^{\,h}_{max})$.

V. Experiments

A. Experimental setup

In order to avoid the perturbations induced by non-linearities of the real STRAS system [30], the proposed semi-autonomous arm-body control mode is compared to manual telemanipulation in a virtual simulator (see Fig. 5). On the leader side, users manipulate the physical leader interface of the STRAS system, which gives leader joint values $\boldsymbol{q}$ as input. Such values are then used to control a simulated follower. The kinematics of the follower implement Eqs. (2) to (4), which are then used to render a virtual scene using VTK [31]. Target positions ${}_w\boldsymbol{p}^l_d$, ${}_w\boldsymbol{p}^r_{d1}$ and ${}_w\boldsymbol{p}^r_{d2}$ are displayed as white spheres with a radius of 3mm in order to simulate the dissection task, as described in Section III, and the colon environment is simulated as a textured cylinder around a curved axis. Note that no collisions are implemented in the simulation.

During the user experiments, a Python script reads the leader side positions of the STRAS system using sockets through a wired Ethernet connection. The Python script creates the VTK environment that is presented to the user on a screen on top of the leader side, as when the user telemanipulates the system with the real follower (Fig. 7). The weights for the objectives

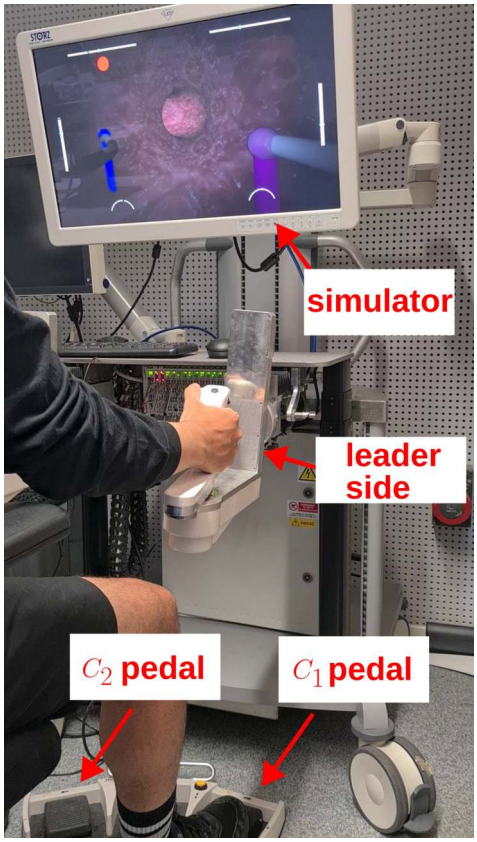

Figure 7. Experimental setup used. Simulated follower visualization through the VTK environment using the physical leader side of the STRAS robotic system.

γ are set as in Table I. When the arm is in an unsafe position the values of γ_3 and γ_4 are increased to 0.15; and γ_1,γ_2 are decreased to 0.4 and 0.3 in order to return the arm into the safe zone. Safe image bounds in the image $\lambda_x^{+,-}$ and $\lambda_y^{+,-}$ are set by the normalized coordinates of the screen from 0.1 to 0.9 in both coordinated axes, with a 1920x1080 px resolution. The bounds $\lambda_{trans}^{-,+}$ are set within the 0-72mm range of motion for the arms.

Table I
PARAMETERS USED IN THE EXPERIMENTAL SETUP. VALUES OF THE WEIGHTS γ AND THE PROPORTIONAL GAINS k FOR EACH OBJECTIVE ACTIVE DURING SEMI-AUTONOMOUS CONTROL MODE.

Variable	γ_1	γ_2	γ_3	γ_4	$k_{1,2,3,4}$
Value	0.5	0.4	0.05	0.05	0.25

B. Experimental task

A trial consists of the user sequentially reaching the three targets ${}_w\boldsymbol{p}_d^l$, ${}_w\boldsymbol{p}_{d1}^r$ and ${}_w\boldsymbol{p}_{d2}^r$, which emulates the dissection stage of ESD, as described in Section III.

Starting from a given reference position, the user first reaches target ${}_w\boldsymbol{p}_d^l$ with the non-dominant arm. This first stage, which will be referred to as Segment I in the following, simulates tissue grabbing. It is considered successful when the user keeps the non-dominant arm tip within a ball of radius 3 mm centred on ${}_w\boldsymbol{p}_d^l$ for two seconds. Deviating too far from ${}_w\boldsymbol{p}_d^l$ resets the timer to 0.

Segment II of the experiment simulates the dissection. The user is required to reach ${}_w\boldsymbol{p}_{d1}^r$ and ${}_w\boldsymbol{p}_{d2}^r$ with the dominant arm while staying on ${}_w\boldsymbol{p}_d^l$ with the non-dominant arm. Similarly to Segment I, a 2s threshold is used to validate a target being reached. In practice, ${}_w\boldsymbol{p}_{d1}^r$ is displayed when Segment I is finished, and ${}_w\boldsymbol{p}_{d2}^r$ is displayed only when ${}_w\boldsymbol{p}_{d1}^r$ is validated. If the user deviates significantly (more than 5mm) from ${}_w\boldsymbol{p}_d^l$, the user will need to reach ${}_w\boldsymbol{p}_d^l$ again and stay for two seconds before being able to validate the current target (${}_w\boldsymbol{p}_{d1}^r$ or ${}_w\boldsymbol{p}_{d2}^r$).

Once the sequence of three targets is fulfilled, the trial is considered finished. New positions of the three targets are then randomly drawn to start a new trial. Positions are uniformly randomly generated, such that reaching target ${}_w\boldsymbol{p}_d^l$ by the dominant arm and reaching ${}_w\boldsymbol{p}_{d1}^r$ by the non-dominant arm require moving the body. Reaching target ${}_w\boldsymbol{p}_{d1}^r$ could be within reach of the dominant arm or require moving the body, depending on the current body position. The positions mimic the difficulty of performing real dissection during an ESD.

C. Protocol

The experimental protocol is constructed to last between 30 and 40 minutes. The user is first introduced to the platform, task, and control modes for 5-10 minutes, and then performs a 15 minutes training. During this training time, the user receives instructions regarding the manual teleoperation and semi-autonomous control mode, as well as specific instructions to perform the task correctly. The user is also instructed to perform smooth and effective movements, avoiding abrupt changes of direction and/or speed that could be detrimental to the patient in the real surgical scenario. Furthermore, the user is instructed to avoid placing the tools out of sight or too close to the camera. The user then performs as many trials as possible during the 15 minutes training time, with the possibility to ask questions to the investigator.

During the experiments, after the training stage has been finished, the user alternates 5 trials with the semi-autonomous arm-body control mode and 5 trials with the manual telemanipulation mode (using the same target positions in both) until the time (10 minutes) runs out, without any interference from the investigator.

After the experiment is finished, the user fills out a NASA TLX [32] questionnaire. The NASA TLX questionnaire evaluates the subjective opinion of users for both control modes. Afterwards, the user is asked for a general opinion of both control modes and is allowed to give feedback for further improvement of the setup or the proposed control mode.

D. Quantitative evaluation metrics

Three categories of performance metrics are considered: time, kinematics and surgical task. Metrics are adapted from [33], [34] considering the architecture of the FSIR platforms, especially the STRAS system.

The time performance metrics measure the raw duration of Segment I T_{sI}, of Segment II T_{sII}, and of the complete

trial T_s. Kinematic metrics measure how the STRAS system is controlled. For the kinematics category, the first set of metrics relates to body movements. A body movement starts when the speed $\|\dot{\boldsymbol{q}}_{\boldsymbol{e}}\|$ goes from 0 to a positive value and stops when $\|\dot{\boldsymbol{q}}_{\boldsymbol{e}}\| < \epsilon$, with ϵ denoting a small body motion threshold. N_b^{mvt} measures the number of independent movements of the body per trial. $T_{b,sI}^{mvt}$ measures the total duration of body movements in Segment I (respectively $T_{b,sII}^{mvt}$ for Segment II) and T_b^{mvt} during the complete trial. The total travelled distance by the body for the complete trial is measured by d_b. Finally, $f_b = N_b^{mvt}/T_b^{mvt}$ is the frequency of body movements (i.e. number of body movements per second) during a given trial. T_b^{idle} is the amount of time the body is idle (i.e. $\|\dot{\boldsymbol{q}}_{\boldsymbol{e}}\| < \epsilon$). Similarly, d_l and f_l measure the total distance travelled and frequency of movements of the non-dominant arm. Such metrics are not computed for the **dominant** arm, because the triggering of body movements during semi-autonomous arm-body control mode would interfere and not provide meaningful output.

A second set of metrics relates to the smoothness of the movements. We use the spectral arc length w to quantify the smoothness [35], which was shown to be correlated with surgical skill and smooth operation [36] and was used in several studies with similar design [37]. We compute the smoothness $w_{b,sII}$ of the body (resp. $w_{l,sII}$ of the non-dominant arm) during Segment II. Note here that the numbers we compute are the opposite of the number computed classically and defined in [35]. This operation is done to obtain a positive number which increases as the smoothness diminishes (i.e. the lower the better, as for most of the other metrics, aside from the $T_b^{idle}$and N_b^{mvt}). Again, metrics relating to the dominant arm are not considered for the reasons exposed above.

Surgical performance metrics measure how well the surgical task was performed. The distance between the non-dominant arm and its objective ${}_w\boldsymbol{p}_d^l$ at a given time t is given by:

$$\boldsymbol{d}_l\left(t\right) = \left\|{}_w\boldsymbol{p}_d^l - {}_w\,\boldsymbol{p}^l(t)\right\|_2 \tag{25}$$

and should be maintained to the minimum during Segment II. Since deviating too much from ${}_w\boldsymbol{p}_d^l$ will require to perform extra surgical gestures, the maximum recorded distance for each trial is of interest. Maximum recorded distance is set as:

$$d_{max} = \max_{t=t_1..t_2} \boldsymbol{d}_l(t), \tag{26}$$

where t_1 and t_2 are the starting and ending times of Segment II for a given trial, respectively. T_v, measures the time extent during which the non-dominant arm is deviating from ${}_w\boldsymbol{p}_d^l$ during Segment II:

$$T_v = \int_{t_1}^{t_2} \left(\boldsymbol{d}_l\left(t\right) > d_{thr}\right) dt, \tag{27}$$

where d_{thr} is a distance threshold set to 5 mm. Finally, we measure the time during which the arms are outside of the field of view during a trial by $T_{r,o}$ and $T_{l,o}$ for the dominant and non-dominant arms respectively.

VI. Results

After obtaining the ethical approval from our competent local notified body (Comité d'Ethique de la Recherche from University of Strasbourg, agreement CER 2022-56) the experimental protocol was tested with a group of eight users. All users were novice users in the sense that they had no prior experience with an FSIR system or with ESD surgery. Two users had clinical background.

Table II
SUMMARY OF THE RESULTS OF THE EXPERIMENTS. THE EXPERIMENTS COMPARE THE MANUAL (**MAN**) AND SEMI-AUTONOMOUS (**SEM**) MODES IN A MULTIPLE-REACHING TASK. METRICS ARE SEPARATED BY DASHED LINES, FIRST GROUP OF METRICS ARE TIME-RELATED, FOLLOWED BY KINEMATIC AND SURGICAL TASK METRICS. STATISTICALLY SIGNIFICANT DIFFERENCES WHERE SEMI-AUTONOMOUS ARM-BODY CONTROL MODE OUTPERFORMS MANUAL MODE ARE MARKED WITH '*' ($p \leq 0.05$). P-VALUES MARKED WITH $^{\triangleleft}$ ($p \geq 0.95$) INDICATE THAT MANUAL MODE OUTPERFORMS SEMI-AUTONOMOUS ARM-BODY CONTROL MODE.

metric	A_w	p-**value**	**median MAN**	**IQR MAN**	**median SEM**	**IQR SEM**
T_s	small	1.0000$^{\triangleleft}$	25.23	17.22	31.00	21.69
T_{sI}	small	0.9924$^{\triangleleft}$	16.78	8.40	18.18	12.34
T_{sII}	small	1.0000$^{\triangleleft}$	7.58	5.45	11.43	13.56
N_b^{mvt}	small	0.0000*	2.00	6.50	2.00	2.00
d_b	small	1.0000$^{\triangleleft}$	46.36	29.60	57.71	39.89
f_b	small	0.0000*	0.19	0.33	0.00	0.16
T_b^{mvt}	small	0.9260	4.01	5.81	4.12	6.14
$T_{b,sI}^{mvt}$	small	0.4977	1.47	3.26	0.00	3.36
$T_{b,sII}^{mvt}$	small	0.9334	1.92	2.42	2.12	2.64
T_b^{idle}	small	1.0000$^{\triangleleft}$	22.04	14.60	25.02	21.85
d_l	small	1.0000$^{\triangleleft}$	12.37	8.57	14.88	10.87
f_l	small	0.9016	0.83	0.85	1.03	0.99
T_l^{mvt}	large	0.0000*	40.00	34.50	4.00	6.25
$w_{l,sII}$	mod.	0.0078*	8.46	15.69	4.59	7.95
$w_{b,sII}$	mod.	0.0091*	5.54	12.39	4.03	3.18
T_v	large	0.0000*	3.30	5.25	0.00	0.00
d_{max}	large	0.0000*	7.17	9.45	1.65	1.34
$T_{r,o}$	small	0.1435	0.00	0.10	0.00	0.00
$T_{l,o}$	small	0.8435	0.00	0.00	0.00	0.00

Preliminary experiments showed a non-gaussian distribution for most cases. Therefore, we report the median and interquartile range (IQR) of each metric per control mode. For the same reason, statistical significance was analyzed using a one-sided Wilcoxon signed-rank test, with the null hypothesis being that the medians of both modes are equal and the alternate hypothesis being that the median of the manual mode is greater than that of the semi-autonomous arm-body control mode. In this context, a p-value of $p < 0.05$ indicates a statistically significant advantage of the semi-autonomous arm-body control mode for a given metric (i.e. smaller value in the proposed approach). Alternatively, if $p > 0.95$, the manual mode is statistically significantly better. The p-value is an indicator of statistically significant differences between two distributions, but does not say anything about the size of the difference. In order to quantify this, we report the effect size using the non-parametric estimator for common language effect size A_w [38], which is an effect size metric adapted to non-parametric distributions. Values of 0.56, 0.64, 0.71 depict small, moderate and large effect sizes, respectively [39].

Table II presents an overview of the metrics.

Results show that users took less time to perform the overall task as well as the segments I and II by using the manual mode, as evidenced by the median values of T_s, T_{sI} and T_{sII}. It is worth noting that the IQR of those metrics is also lower during manual teleoperation, suggesting a more homogeneous spread of user performance in terms of timing than in semi-autonomous arm-body control mode. The effect size is however small.

Results on the kinematic performance metrics are mixed. Users realized fewer numbers of body movements N_b^{mvt} and with a smaller frequency f_b in semi-autonomous arm-body control mode, but travelled a larger distance with the body, as shown by the values of d_b. On the other hand, the idle body time T_b^{idle} was shorter and less distance was travelled by the non-dominant arm d_l in manual teleoperation. However, the effect size associated with those metrics is small. The duration of non-dominant arm movements T_l^{mvt}, the smoothness of both body $w_{l,sII}$ and the non-dominant arm movements $w_{b,sII}$ were statistically significantly better in semi-autonomous arm-body control mode with large and moderate effect sizes. The smoothness metric in particular shows an important decrease in the IQR, showing a more homogeneous performance of the users in that mode.

Finally, the semi-autonomous arm-body control mode clearly outperforms the manual teleoperation mode regarding the two main surgical performance metrics. The duration of violating the reaching task of the non-dominant arm T_v has a median of 0 for the semi-autonomous arm-body control mode and of 3.30 in the manual teleoperation mode, with statistically significant p-values, and a large effect size. Similarly, the max recorded distance d_{max} between the non-dominant arm and ${}_w\boldsymbol{p}_d^l$ is lower during the semi-autonomous arm-body control mode, with a median of 1.65 compared to 7.17 for manual mode, again with statistically significant p-value and a large effect size. The last two metrics reporting the amount of time the dominant (respectively non-dominant) arm spends outside of the field of view $T_{r,o}$(respectively $T_{l,o}$) does not give significant insights with values close to zero in all cases.

Qualitative metrics evaluating the user's opinion using the NASA TLX questionnaire are reported in Fig. 8. The purple bars represent the users' opinions on the manual teleoperation mode, while the blue bars represent the semi-autonomous arm-body control mode. Overall the users rank both control modes similarly, with the notable exception of the temporal demand. In fact, users felt more pressured during the semi-autonomous arm-body control mode, which also affected their frustration.

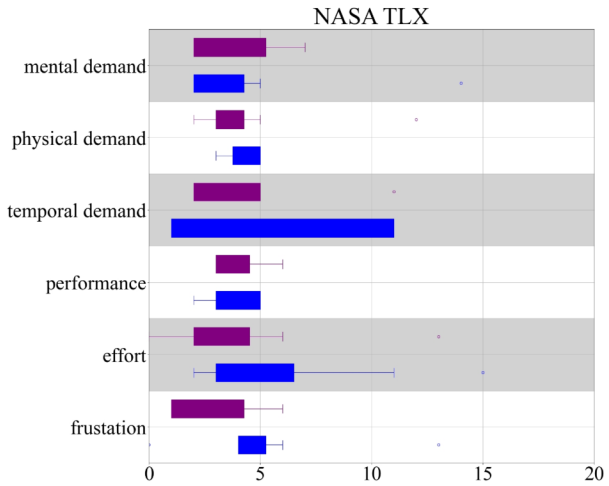

Figure 8. NASA TLX results from the manual and semi-autonomous arm-body control mode. Purple depicts manual control. Semi-autonomous control mode is depicted in blue.

VII. Discussion

This section presents and discusses the major outcomes of the study. ESD requires users to consider the intricate relationship between the arms and body of FSIR platforms to correctly perform the surgical task. For this reason, we proposed a semi-autonomous control method that assists users in handling the arm-body relationship allowing them to focus on the task at hand. During the dissection stage of ESD, the user is required to grab a tissue and maintain it in a position that eases cutting while controlling the body, to then continue cutting with the other arm –traction– which is hard to correctly perform with FSIR platforms [40] due to the coupled architecture – which is also present in a more intricate way in manual flexible endoscopes [41], [42]. During dissection through manual telemanipulation, commanding both arms and the body is a task that involves controlling 10DOF –while controlling one arm and the body requires 7DOF. During the semi-autonomous control mode, the arms are directly telemanipulated through the handles of the leader console while body control is set by the user commanding a pedal and a leader handle. The proposed approach reduces the above-mentioned 10DOF to be controlled to correctly perform dissection to 3DOF –those of the dominant arm handle and actuation of the pedal. The user only focuses on body control through the handle while the arms can be locked into place.

The user study showed that the semi-autonomous arm-body control mode allowed performing the simulated dissection step in a better and safer way, as evidenced by the values of the task-related metrics, at a penalty cost with an increment on the temporal demand. Despite the lack of clinical training by six of the users, no clear evidence of better or worse performance is noticed. The group without clinical background did not report any special training or abilities, which further validates the ease of use of the approach. As for many other robotic assistance systems, this is at the expense of the operating time [43], which is slightly longer in the semi-autonomous arm-body control mode, both for the whole duration and for the two stages of the simulated dissection. One should note here that the IQR of the reported timing metrics T_s, T_{sI} and T_{sII} is larger in the semi-autonomous arm-body control mode, suggesting some users took more time than others in this mode. This conclusion is supported by the NASA TLX questionnaire outputs on Fig. 8, for which one can clearly see a larger spread of the *temporal demand* responses by the users in semi-autonomous arm-body control mode.

Examples of experimental data showing distances between

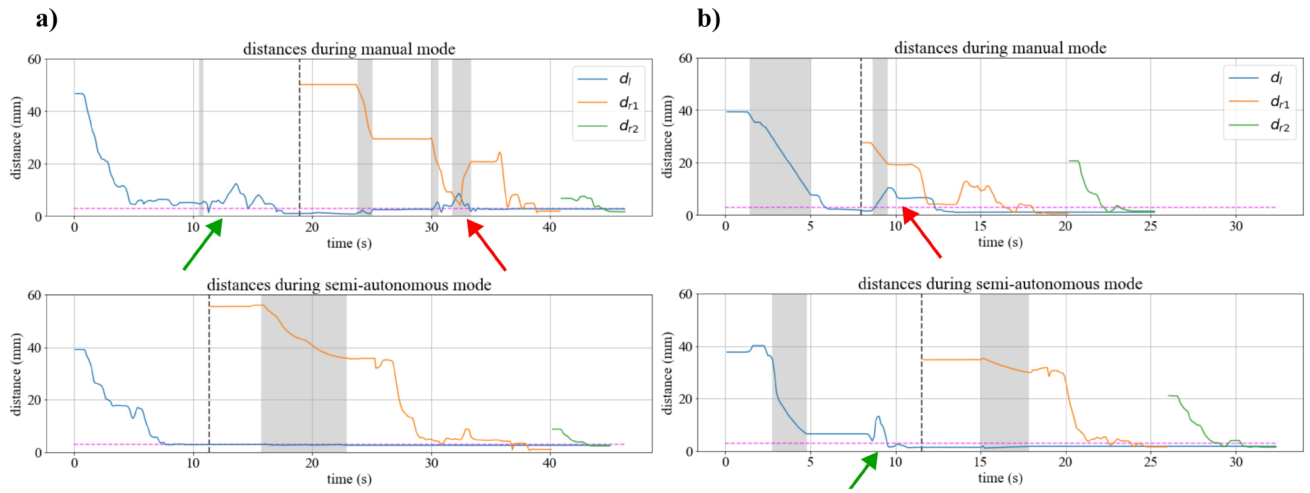

Figure 9. Example of distances of the arms to the objectives in manual (upper) and semi-autonomous arm-body control mode (lower). a) and b) are trials done by two different users. Body movements are depicted with grey highlights. $\boldsymbol{d}_l$ is the distance between the non-dominant arm and its objective ${}_w\boldsymbol{p}_d^l$, $\boldsymbol{d}_{r1}$ (resp. $\boldsymbol{d}_{r2}$) is the distance between the dominant arm and its first objective ${}_w\boldsymbol{p}_{d1}^r$ (resp. second objective ${}_w\boldsymbol{p}_{d2}^r$). Distances are plotted only for the relevant stages. The thresholds for fulfilling the reaching tasks are represented by the horizontal dashed pink line, of 3 mm. The vertical dashed line represents the transition between stage I and stage II of the dissection. Red arrowheads represent the violation of the surgical task after the ${}_w\boldsymbol{p}_d^l$ objective has been reached (after the black dashed vertical line). Green arrowhead shows an increase in the distance of ${}_w\boldsymbol{p}_d^l$ but is before the reaching objective by the non-dominant arm has been attained.

the tip of the arms and the target positions for two different users are shown in Fig. 9-a and Fig. 9-b. One can see that the semi-autonomous arm-body control mode allows moving the body (grey highlights in the figures) while not impacting the distance $\boldsymbol{d}_l$ between the non-dominant arm and its target point. On the other hand, in the manual telemanipulation mode, there is a clear influence of body movements on $\boldsymbol{d}_l$ when reaching for the dominant arm targets and moving the body, with large deviations from ${}_w\boldsymbol{p}_d^l$ potentially leading to tissue damage and/or loss of traction which will require re-grabbing (red arrowheads on Fig. 9-a and Fig. 9-b). This is likely due to the fact that handling coordinated multi-arm movements with an FSIR platform such as the STRAS is extremely complex in manual teleoperation. Such spikes in the movement of the non-dominant arm in the manual mode are also the likely reason why smoother movements were performed by the non-dominant arm and body in the semi-autonomous arm-body control mode, as evidenced by the values and statistical analysis of $w_{l,sII}$ and $w_{b,sII}$.

Our approach shows promising results regarding the performance of the task, by easing traction –here modeled as the positional task for the non-dominant arm. Moreover, our approach avoids unnecessary movements as demonstrated by the kinematics metrics performing better with the proposed approach. Our study has however several limitations which could be addressed in future work.

First, our model of the dissection stage of ESD is made of static positions which the user should reach sequentially. While it is a good first-order approximation of the task, it is also a simplification in two major aspects. The first one is that we consider in our study the environment to be perfectly known and static. In practice, the environment would be moving with physiological movement. Therefore, estimation algorithms would be needed to estimate the current stage of the operation (marking, dissection, etc.) as well as the positions to be reached and maintained. Such estimations would add uncertainty to the control system which would need to be tackled in future work. Secondly, we model the tissue-grabbing task of the dissection stage as the non-dominant arm keeping a given position in space. Even with minimal physiological movement (which can be achieved by insufflation of the colon during ESD), the desired task is more akin to keeping a desired tension, which would evolve in time as the underlying tissue is being dissected. A more complex formulation of the control problem accounting for the evolution of the desired position ${}_w\boldsymbol{p}_d^l$ during the tissue dissection, would therefore be needed.

Second, during the semi-autonomous arm-body control mode the user has control of the arms when not moving the body. This means that the world frame position of the arms which should be kept constant during body movements is a user input. Depending on the current state $\boldsymbol{q}$ of the robot, it may not be possible to fulfil all objectives, leading to a violation of the non-dominant arm tissue grabbing task. This happened on a few occasions during the experiments and required the user to stop the semi-autonomous control, move the body to a more favourable position, and perform a re-grabbing. Such behaviour leads to a considerable time penalty and is likely the source of increased temporal demand for the users (see Fig. 8), as well as increased time in semi-autonomous arm-body control mode. Coupling our proposed approach with a robust motion planning algorithm for anticipating such events would be a possible solution to mitigate this problem, potentially leading to a more intuitive and effective semi-autonomous control mode.

Third, a limitation of the study is that it was not tested on the physical follower of the STRAS system. This choice was made in order to evaluate the effect of the proposed control approach, without undue disturbance from the non-linearities

of the physical system [30]. Translating our proposed approach on the physical system suffers from one major hurdle: state estimation of the robot follower, which is impacted by non-linear frictional and backlash behaviour as well as by forces on the arms when interacting with the tissue. This problem could be tackled using endoscopic vision as a sensing input. When coupled with the model of the robotic system, it was shown to be sufficient for state/shape detection using marker-based [44], [45] or markerless [46] solutions. To enhance the robustness of the controller, coupling such vision-based approaches with our proposed semi-autonomous control approaches would likely require reformulating some of the objectives in the image space, using visual servo control. Such developments are beyond the scope of the present study.

Fourth, one should note that our results show an improvement in the kinematic (smoothness of movements) and task-related metrics (performing the reaching tasks with the dominant arm without violating the positional task –traction– on the non-dominant arm) with the proposed control mode, but with longer operating times. This increase in duration is potentially linked to the limited duration of the experiments. Users spend 40 minutes –at most– to get familiar with the platform, receive instructions, and perform the training and the experiments. We decided for such a short duration due to the availability of users and to avoid stress and/or fatigue. During this short period reaching a plateau in the learning curve [47] is challenging [48], and duration is not expected to be heavily improved –similarly to experiments in laparoscopy [49]. In fact, in laparoscopic robotic surgery learning curves are reported to be between 20 and 40 cases [50], [51]. Future work could involve a longitudinal study [52] with more users under the simulated scenario and the usage of the physical STRAS system, as to validate the gained skills and the fidelity of the proposed simulation with novice users. It would also be interesting to evaluate the proposed control approach with both novice and expert users.

Finally, one should also note that our study specifically focused on the STRAS system for practical reasons (the leader and follower systems being available for testing in our lab). The tree-like architecture of the STRAS system is however common to many FSIR platforms [5], [8], [26]. Therefore, results are very likely to transfer easily to such platforms. The same limitation could be formulated for the specific dissection stage considered in this work. Dissection being the most difficult part of one of the main procedures targeted by FSIR platforms, it is natural to focus on it. The proposed control formulation, however, could be extended beyond ESD and dissection, since many surgical scenarios feature grasping [17], tissue tensioning [16], and body or camera movements. Future work could therefore consider extending our proposed approach to multi-arm continuum robotic structures for urology [53], fetal [7], or brain surgery [54].

VIII. Conclusions and Future Work

In this paper, we proposed a semi-autonomous control approach for handling the complex multi-DOF coordination required to perform ESD using FSIR. This task is particularly challenging, as it requires performing movements with the arms while considering the coupled arm-body relationship of FSIR platforms. Our proposed control approach leaves the user free to focus on fulfilling the task, while body repositioning movements are handled automatically by a QP-based controller. The controller is formulated by considering specific features of the task itself. Specifically, we keep one or both arms static with respect to the surgical targets while moving the body, depending on the surgical step being performed, which is extremely complex to achieve using manual telemanipulation.

The approach has been validated in a simulated environment using the STRAS system, showing a strong increase in performance of the surgical task at the expense of a slight increase in time. Validation tests with the physical STRAS system will be performed in future work.

References

[1] E. Morgan, M. Arnold, A. Gini, V. Lorenzoni, C. Cabasag, M. Laversanne, J. Vignat, J. Ferlay, N. Murphy, and F. Bray, "Global burden of colorectal cancer in 2020 and 2040: Incidence and mortality estimates from globocan," *Gut*, vol. 72, no. 2, pp. 338–344, 2023.

[2] The International Agency for Research on Cancer (IARC), "Global cancer observatory," https://gco.iarc.fr/today, accessed: 2022-3-7.

[3] Y. Saito, T. Sakamoto, S. Fukunaga, T. Nakajima, S. Kuriyama, and T. Matsuda, "Endoscopic submucosal dissection (esd) for colorectal tumors," *Digestive Endoscopy*, vol. 21, pp. S7–S12, 2009.

[4] A. De Donno, L. Zorn, P. Zanne, F. Nageotte, and M. de Mathelin, "Introducing stras: A new flexible robotic system for minimally invasive surgery," in *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 1213–1220.

[5] M. Hwang and D.-S. Kwon, "K-flex: a flexible robotic platform for scar-free endoscopic surgery," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 16, no. 2, p. e2078, 2020.

[6] M. Mitsuishi, A. Morita, N. Sugita, S. Sora, R. Mochizuki, K. Tanimoto, Y. M. Baek, H. Takahashi, and K. Harada, "Master–slave robotic platform and its feasibility study for micro-neurosurgery," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 9, no. 2, pp. 180–189, 2013.

[7] T. Vandebroek, M. Ourak, C. Gruijthuijsen, A. Javaux, J. Legrand, T. Vercauteren, S. Ourselin, J. Deprest, and E. Poorten, "Macro-micro multi-arm robot for single-port access surgery," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 425–432.

[8] D. J. Abbott, C. Becke, R. I. Rothstein, and W. J. Peine, "Design of an endoluminal notes robotic system," in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2007, pp. 410–416.

[9] P. Wisanuvej, G. Gras, K. Leibrandt, P. Giataganas, C. A. Seneci, J. Liu, and G.-Z. Yang, "Master manipulator designed for highly articulated robotic instruments in single access surgery," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 209–214.

[10] J. Ahn, J. Kim, H. Lee, M. Hwang, and D.-S. Kwon, "A highly intuitive and ergonomic redundant joint master device for four-degrees of freedom flexible endoscopic surgery robot," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 17, no. 1, pp. 1–14, 2021.

[11] A. De Donno, F. Nageotte, P. Zanne, L. Zorn, and M. de Mathelin, "Master/slave control of flexible instruments for minimally invasive surgery," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 483–489.

[12] P. Berthet-Rayne, G. Gras, K. Leibrandt, P. Wisanuvej, A. Schmitz, C. A. Seneci, and G.-Z. Yang, "The i 2 snake robotic platform for endoscopic surgery," *Annals of biomedical engineering*, vol. 46, pp. 1663–1675, 2018.

[13] A. Légner, M. Diana, P. Halvax, Y.-Y. Liu, L. Zorn, P. Zanne, F. Nageotte, M. De Mathelin, B. Dallemagne, and J. Marescaux, "Endoluminal surgical triangulation 2.0: A new flexible surgical robot. preliminary pre-clinical results with colonic submucosal dissection," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 13, no. 3, p. e1819, 2017, e1819 RCS-16-0150.R2.

[14] D. T. H. de Moura, H. Aihara, P. Jirapinyo, G. Farias, K. E. Hathorn, A. Bazarbashi, A. Sachdev, and C. C. Thompson, "Robot-assisted endoscopic submucosal dissection versus conventional esd for colorectal lesions: outcomes of a randomized pilot study in endoscopists without prior esd experience (with video)," *Gastrointestinal endoscopy*, vol. 90, no. 2, pp. 290–298, 2019.

[15] P. W. Y. Chiu, K. Y. Ho, and S. J. Phee, "Colonic endoscopic submucosal dissection using a novel robotic system (with video)," *Gastrointestinal Endoscopy*, vol. 93, no. 5, pp. 1172–1177, 2021.

[16] S. Portolés, P. Vanbiervliet, B. Rosa, C. Tomassetti, C. Meuleman, E. B. Vander Poorten, and D. Reynaerts, "Force control for tissue tensioning in precise robotic laser surgery," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 579–585.

[17] A. A. Shahkoo and A. A. Abin, "Autonomous tissue manipulation via surgical robot using deep reinforcement learning and evolutionary algorithm," *IEEE Transactions on Medical Robotics and Bionics*, vol. 5, no. 1, pp. 30–41, 2023.

[18] A. Pandya, L. A. Reisner, B. King, N. Lucas, A. Composto, M. Klein, and R. D. Ellis, "A review of camera viewpoint automation in robotic and laparoscopic surgery," *Robotics*, vol. 3, no. 3, pp. 310–329, 2014.

[19] R. D. Ellis, A. J. Munaco, L. A. Reisner, M. D. Klein, A. M. Composto, A. K. Pandya, and B. W. King, "Task analysis of laparoscopic camera control schemes," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 12, no. 4, pp. 576–584, 2016.

[20] Y.-F. Wang, D. R. Uecker, and Y. Wang, "A new framework for vision-enabled and robotically assisted minimally invasive surgery," *Computerized Medical Imaging and Graphics*, vol. 22, no. 6, pp. 429–437, 1998.

[21] K. Zinchenko, C.-Y. Wu, and K.-T. Song, "A study on motion control of a robotic endoscope holder using speech recognition," in *2016 IEEE International Conference on Industrial Technology (ICIT)*. IEEE, 2016, pp. 1472–1475.

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

[23] A. Bihlmaier, "Endoscope robots and automated camera guidance," in *Learning Dynamic Spatial Relations*. Springer, 2016, pp. 23–102.

[24] Y. Huang, W. Lai, L. Cao, E. Burdet, and S. J. Phee, "Design and evaluation of a foot-controlled robotic system for endoscopic surgery," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2469–2476, 2021.

[25] S. Zuo, T. Chen, X. Chen, and B. Chen, "A wearable hands-free human-robot interface for robotized flexible endoscope," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 3953–3960, 2022.

[26] F. Nageotte, L. Zorn, P. Zanne, and M. De Mathelin, "Stras: A modular and flexible telemanipulated robotic device for intraluminal surgery," in *Handbook of Robotic and Image-Guided Surgery*. Elsevier, 2020, pp. 123–146.

[27] R. J. Webster III and B. A. Jones, "Design and kinematic modeling of constant curvature continuum robots: A review," *The International Journal of Robotics Research*, vol. 29, no. 13, pp. 1661–1683, 2010.

[28] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *Journal of Machine Learning Research*, vol. 17, no. 83, pp. 1–5, 2016.

[29] B. Faverjon and P. Tournassoud, "A local based approach for path planning of manipulators with a high number of degrees of freedom," in *Proceedings. 1987 IEEE international conference on robotics and automation*, vol. 4. IEEE, 1987, pp. 1152–1159.

[30] T. Poignonec, P. Zanne, B. Rosa, and F. Nageotte, "Towards in situ backlash estimation of continuum robots using an endoscopic camera," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 4788–4795, 2020.

[31] W. Schroeder, K. Martin, and B. Lorensen, "The visualization toolkit, 4th edn. kitware," *New York*, 2006.

[32] 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.

[33] T. Da Col, A. Mariani, A. Deguet, A. Menciassi, P. Kazanzides, and E. De Momi, "Scan: System for camera autonomous navigation in robotic-assisted surgery," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 2996–3002.

[34] A. M. Jarc and M. J. Curet, "Viewpoint matters: objective performance metrics for surgeon endoscope control during robot-assisted surgery," *Surgical endoscopy*, vol. 31, pp. 1192–1202, 2017.

[35] S. Balasubramanian, A. Melendez-Calderon, A. Roby-Brami, and E. Burdet, "On the analysis of movement smoothness," *Journal of neuroengineering and rehabilitation*, vol. 12, no. 1, pp. 1–11, 2015.

[36] S. Estrada, C. Duran, D. Schulz, J. Bismuth, M. D. Byrne, and M. K. O'Malley, "Smoothness of surgical tool tip motion correlates to skill in endovascular tasks," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 5, pp. 647–659, 2016.

[37] O. Caravaca-Mora, P. Zanne, G. Liao, N. Zulina, L. Heroin, L. Zorn, M. De Mathelin, B. Rosa, F. Nageotte, and M. J. Gora, "Automatic intraluminal scanning with a steerable endoscopic optical coherence tomography catheter for gastroenterology applications," *Journal of Optical Microsystems*, vol. 3, no. 1, pp. 011 005–011 005, 2023.

[38] J. Ruscio, "A probability-based measure of effect size: robustness to base rates and other factors." *Psychological methods*, vol. 13, no. 1, p. 19, 2008.

[39] J. C.-H. Li, "Effect size measures in a two-independent-samples case with nonnormal and nonhomogeneous data," *Behavior research methods*, vol. 48, pp. 1560–1574, 2016.

[40] Z. Meng, Z. Huang, B. Deng, L. Ling, Y. Ning, and S. M. Rafiq, "Robotic-assisted vs non-robotic traction techniques in endoscopic submucosal dissection for malignant gastrointestinal lesions," *Frontiers in Oncology*, vol. 12, p. 1062357, 2022.

[41] K. Kume, "Endoscopic mucosal resection and endoscopic submucosal dissection for early gastric cancer: current and original devices," *World Journal of Gastrointestinal Endoscopy*, vol. 1, no. 1, p. 21, 2009.

[42] C. C. Thompson, M. Ryou, N. J. Soper, E. S. Hungess, R. I. Rothstein, and L. L. Swanstrom, "Evaluation of a manually driven, multitasking platform for complex endoluminal and natural orifice transluminal endoscopic surgery applications (with video)," *Gastrointestinal endoscopy*, vol. 70, no. 1, pp. 121–125, 2009.

[43] H. G. Kenngott, L. Fischer, F. Nickel, J. Rom, J. Rassweiler, and B. Müller-Stich, "Status of robotic assistance—a less traumatic and more accurate minimally invasive surgery?" *Langenbeck's archives of surgery*, vol. 397, pp. 333–341, 2012.

[44] P. Cabras, F. Nageotte, P. Zanne, and C. Doignon, "An adaptive and fully automatic method for estimating the 3d position of bendable instruments using endoscopic images," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 13, no. 4, p. e1812, 2017.

[45] Z. Zhang, B. Rosa, O. Caravaca-Mora, P. Zanne, M. J. Gora, and F. Nageotte, "Image-guided control of an endoscopic robot for oct path scanning," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5881–5888, 2021.

[46] L. Sestini, B. Rosa, E. De Momi, G. Ferrigno, and N. Padoy, "A kinematic bottleneck approach for pose regression of flexible surgical instruments directly from images," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2938–2945, 2021.

[47] K. Siau, N. D. Hawkes, and P. Dunckley, "Training in endoscopy," *Current Treatment Options in Gastroenterology*, vol. 16, pp. 345–361, 2018.

[48] R. M. Jiménez-Rodríguez, M. Rubio-Dorado-Manzanares, J. M. Díaz-Pavón, M. L. Reyes-Díaz, J. M. Vazquez-Monchul, A. M. Garcia-Cabrera, J. Padillo, and F. De la Portilla, "Learning curve in robotic rectal cancer surgery: current state of affairs," *International journal of colorectal disease*, vol. 31, pp. 1807–1815, 2016.

[49] E. Leijte, I. de Blaauw, F. Van Workum, C. Rosman, and S. Botden, "Robot assisted versus laparoscopic suturing learning curve in a simulated setting," *Surgical endoscopy*, vol. 34, pp. 3679–3689, 2020.

[50] S. W. Wong and P. Crowe, "Factors affecting the learning curve in robotic colorectal surgery," *Journal of Robotic Surgery*, vol. 16, no. 6, pp. 1249–1256, 2022.

[51] M. S. Choi, S. H. Yun, C. K. Oh, J. K. Shin, Y. A. Park, J. W. Huh, Y. B. Cho, H. C. Kim, and W. Y. Lee, "Learning curve for single-port robot-assisted rectal cancer surgery," *Annals of Surgical Treatment and Research*, vol. 102, no. 3, pp. 159–166, 2022.

[52] T. S. Zhu, N. Godse, D. R. Clayburgh, and U. Duvvuri, "Assessing the learning curve associated with a novel flexible robot in the pre-clinical and clinical setting," *Surgical endoscopy*, pp. 1–10, 2022.

[53] R. J. Hendrick, S. D. Herrell, and R. J. Webster, "A multi-arm hand-held robotic system for transurethral laser prostate surgery," in *2014 IEEE international conference on robotics and automation (ICRA)*. Ieee, 2014, pp. 2850–2855.

[54] K. Price, J. Peine, M. Mencattelli, Y. Chitalia, D. Pu, T. Looi, S. Stone, J. Drake, and P. E. Dupont, "Using robotics to move a neurosurgeon's hands to the tip of their endoscope," *Science Robotics*, vol. 8, no. 82, p. eadg6042, 2023.