Image-Guided Control of an Endoscopic Robot for OCT Path Scanning

Zhongkai Zhang, Benoît Rosa, Oscar Caravaca-Mora, Philippe Zanne, Michalina J Gora, Florent Nageotte

▶ **To cite this version:**

Zhongkai Zhang, Benoît Rosa, Oscar Caravaca-Mora, Philippe Zanne, Michalina J Gora, et al.. Image-Guided Control of an Endoscopic Robot for OCT Path Scanning. IEEE Robotics and Automation Letters, 2021, 10.1109/LRA.2021.3087085 . hal-03281611

HAL Id: hal-03281611
https://hal.science/hal-03281611

Submitted on 8 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Image-Guided Control of an Endoscopic Robot for OCT Path Scanning

Zhongkai Zhang, Benoît Rosa, Oscar Caravaca-Mora, Philippe Zanne, Michalina J. Gora, Florent Nageotte

***Abstract*—Optical coherence tomography (OCT) endoscopic catheters provide efficient solutions for the non-invasive scan of malignant tissues in internal human organs. In this paper, we investigate the image-guided control of a robotic flexible endoscopic system equipped with an OCT probe for autonomous tissue scanning. The visual control strategy is achieved by combining the image feedback from both a monocular endoscopic camera (eye-to-hand) and an OCT probe (eye-in-hand). Our control strategy allows the OCT probe to automatically track a trajectory defined by the user on the 2D endoscopic image and keep contact with the scanned surface. The orientation, depth, and endoscopic 2D image position are controlled by solving an optimization problem, which is converted to a quadratic programming problem. For the implementation of the visual control strategy, we also consider visibility constraints and actuators limitations. In addition, a marker-based method to estimate the 3D pose of the continuum robot using only 2D images is proposed based on a Kalman filter and a registration technique. The proposed control strategy is validated using both simulation and laboratory experiments.**

***Index Terms*—flexible endoscopy, optical coherence tomography, optimization based control, visual servoing, tissue scanning**

I. Introduction

Optical coherence tomography (OCT) is an imaging modality which allows obtaining cross-sectional images of tissues with micrometer resolution and millimeter penetration depth. Flexible, fiber-based OCT catheters are promising devices which enable minimally invasive scanning of internal organs through natural orifices. Employing OCT probes reduces the need for tissue removal, and ex-situ biopsy can be replaced with in-situ real-time optical measurement. OCT systems have been mostly used for diagnostic in the upper digestive tract [1]. In these cases, large selected segments of the esophagus are scanned at high velocity for subsequent off-line analysis.

Continuum robots embedded with OCT provide promising solutions to obtain medical information from tissues or organs. In [2], a custom-made OCT probe is integrated with a continuum robot to achieve micro and macro scale motion. New systems are also currently under development for assistance to surgical procedures performed with robotic flexible endoscopes in the digestive tube [3]. In such systems, the OCT probe is rather used to analyze specific areas online. For instance, it could be used to check that no tumoral tissues remain at the boundaries after a resection/dissection. An endoscopic camera is available for the user to define the tissues that should be scanned with the OCT probe (such as Fig. 1 (a)).

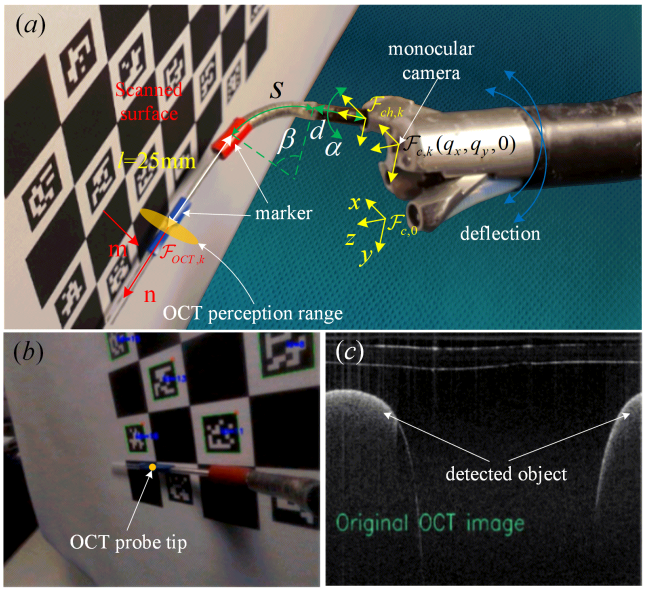

Figure 1. Experimental setup. (a) Overall setup consisting of a scanned surface with QR codes, an OCT probe, and a flexible endoscope robot with a monocular camera; (b) a typical endoscopic view; (c) the corresponding OCT image showing the detected surface.

OCT images can not only provide medical information but also be used to guide surgical devices. An image-guided closed-loop controller has been proposed in [4] to perform an inplane positioning task of a concentric tube robot with respect to a biological sample. This reference employs an eye-to-hand configuration for visual servoing with OCT images being inputs on the control loop. A needle tracking technique is developed in [5] using OCT images to estimate the needle penetration depth in the cornea. A wavelet-based visual servoing using only OCT cross-sectional images is proposed in [6] to control a 6 DoFs robot with an eye-to-hand configuration. It allows to automatically move a biological sample so that its OCT image reach a desired one. Only a limited number of visual servoing strategies have been proposed to combine camera and OCT images. A partitioned camera-OCT based 6 DoFs visual servoing strategy [7] has been proposed for automatic repetitive optical biopsies in an eye-in-hand configuration (3 DoFs controlled by the OCT images and the remaining 3 DoFs controlled by the camera images). However, this was only

The project is supported by the INSERM (French Institute for Health) through the *Plan Physics for Cancer* (project ROBOT), the ATIP-Avenir grant, the ARC Foundation for Cancer research and the University of Strasbourg IdEx. This work is also partially supported by French State Funds managed by the Agence Nationale de la Recherche (ANR) through the Investissements d'Avenir Program, Grants ANR-11-LABX-0004 (Labex CAMI) and ANR-10-IAHU-02 (IHU-Strasbourg).

Authors are with ICube Laboratory, University of Strasbourg, CNRS, UMR7357, France. Contact: zhongkai.zhangzkz@gmail.com

implemented onto a large laboratory robotic platform which is not compatible with minimally invasive procedures.

To achieve precise scanning, the 3D position of the OCT probe and its orientation should be taken into consideration. Besides, due to the limited depth perception (a few mm) of the used OCT probe, contact with the tissue should be maintained during the scanning. Performing such tasks requires controlling several DoFs while relying on both the endoscopic camera and OCT images. They are very difficult to realize by a user, even in telemanipulation. For this reason, we propose to make the scanning automatically. This task is challenging because the conventional monocular camera embedded on the endoscopic robot can only provide 2D images of both the tissue surface and the OCT probe. It is also impossible to estimate the distance between the tissue and the probe using the OCT images when said distance is beyond the perception range. To the best of our knowledge, this is the first paper to investigate the automatic control of a flexible endoscopic robot for OCT path scanning.

The control problem in this paper involves multi-objectives and feedback integration from multiple-sensors. For multi-objective tasks, the suitable strategies are to either separate the task stage by stage [8] or to employ a global task function as the weighted sum of individual task functions [9]. Our main scientific contribution is the formulation of an optimization-based visual servoing control strategy for the automatic scanning control using image feedback from both a camera and an OCT probe. The OCT probe serves as a sensor to detect contact between the probe and the tissue. The control inputs are computed in real-time by solving one constrained quadratic programming problem. Our method has two main advantages: 3D path scanning without building a metric 3D reconstruction, and control inputs computation without switching between multiple control objectives. Unlike hybrid position/force control approaches, where a force sensor [10] is used to control the robot in the direction of contact, the processing of OCT images can only provide binary information on contact existence.

II. Problem Statement

In this section, we first give an overview of the continuum robot equipped with an OCT probe. Then, we briefly introduce the control objectives for an automatic scanning.

A. System introduction

As shown in Fig. 1, the endoscopic robot called STRAS [11] is composed by a main endoscope and a flexible instrument. Actuated by two pairs of cables (joint variables q_{c1} and q_{c2}), the steerable distal part (length: 185 mm, diameter: 16 mm) of the main endoscope can be deflected in two orthogonal directions. A monocular camera is embedded at its distal tip. A flexible instrument can be inserted inside one of the two lateral channels of the endoscope and has three DoFs: rotation α, insertion d in the endoscope channel, and bending β for the short bending distal part (length: $s = 18.5$ mm, diameter: 3.5 mm). All DoFs are actuated from external motors at the proximal side.

A motorized OCT probe [3] is inserted in the bending instrument shaft and extends 25 mm out its distal tip. The probe consists of an external motor, an optical fiber, ball lens, a casing, and the associated electronic and optical elements. Light transmitted through the fiber is deflected by the ball lens, creating an imaging beam orthogonal to the probe. The back-reflected light is then processed to obtain A-line scans. Thanks to the high-speed rotation of the fiber by the external motor, radial images can be reconstructed (see Fig. 1 (c)) at 50Hz. The perception range is 3.5 mm outside of the probe sheath (see Fig. 1 (a)). In this work, we directly use planar radial B-scans images. This OCT acquisition probe has been validated in [3].

B. Overview of the control framework

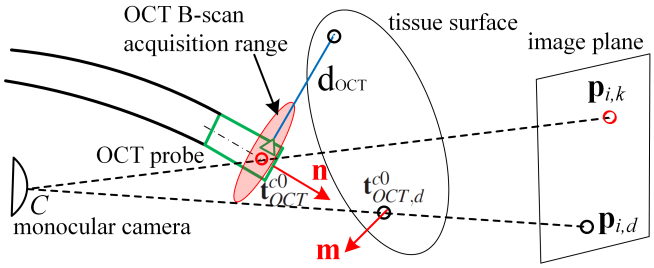

Figure 2. Illustration of the control objectives for tissue scanning. $\mathbf{t}^{c0}_{OCT}$ and $\mathbf{t}^{c0}_{OCT,d}$ are respectively the 3D position of the OCT probe and its desired position. Their corresponding image positions are $\mathbf{p}_{i,k}$ and $\mathbf{p}_{i,d}$, respectively. $\mathbf{m}$ is the normal direction of the surface at point $\mathbf{t}^{c0}_{OCT,d}$. $\mathbf{n}$ is the direction of the OCT probe. d_{OCT} is the shortest distance on the OCT imaging plane between the OCT probe and the tissue.

As shown in Fig. 2, the objective is to control the tip of the OCT probe $\mathbf{t}^{c0}_{OCT}$ to follow a desired path $\mathbf{t}^{c0}_{OCT,d}$ on the tissue under OCT and white light guidance. Because of the limited field of view of the OCT image (B-scan shown as a red disk on Fig. 2), the OCT probe should preferably be maintained in close contact to the target surface, i.e. $\mathrm{d}_{OCT} = 0$. In order to maximize the depth of visualization under the tissue surface, the direction of the OCT probe $\mathbf{n}$ should be orthogonal to the normal direction of the tissue surface $\mathbf{m}$, i.e. $\mathbf{m} \perp \mathbf{n}$. Since only 2D image is available, the objective $\mathbf{t}^{c0}_{OCT} = \mathbf{t}^{c0}_{OCT,d}$ is converted as $\mathrm{d}_{OCT} = 0$ and $\mathbf{p}_{i,k} = \mathbf{p}_{i,d}$ where $\mathbf{p}_{i,k}$ and $\mathbf{p}_{i,d}$ (defined by the user) are respectively the image position of $\mathbf{t}^{c0}_{OCT}$ and $\mathbf{t}^{c0}_{OCT,d}$. Therefore, the control objectives are $\mathbf{m} \perp \mathbf{n}$, $\mathbf{p}_{i,k} = \mathbf{p}_{i,d}$ and $\mathrm{d}_{OCT} = 0$. The only information about the surface that will be used is its local normal $\mathbf{m}$. The required user input is the desired path on the endoscopic image plane, i.e. a sequence of desired image positions $\mathbf{p}_{i,d}$.

The proposed control framework is shown in Fig. 3. The *pose estimation* module is employed to estimate and update geometric features of the environment and the instrument from the endoscopic image. Firstly, the normal direction $\mathbf{m}$ of the scanned surface is estimated using computer vision techniques. Secondly, the module updates $\mathbf{p}_{i,d}$ on the current endoscopic image. Since motions of the camera can be performed to realize the task, it has to be updated at each sampling time. Finally, the *pose estimation* module also implements the algorithm proposed in Section VI-A to estimate $\mathbf{n}$ and $\mathbf{p}_{i,k}$ for the OCT probe.

The *contact detection* module processes OCT images in order to estimate the feedback input η, which depends on

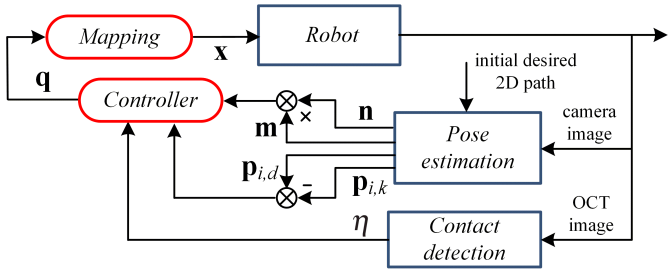

Figure 3. Framework of the proposed visual servoing strategy with four modules (Pose estimation, Contact detection, Controller, and Mapping). The pose estimation module obtains the normal direction of the scanned surface $\mathbf{m}$, the updated desired path with the interpolated point $\mathbf{p}_{i,d}$, the direction $\mathbf{n}$, and the image position $\mathbf{p}_{i,k}$ of OCT probe. The contact detection module defines the parameter η from OCT image to control the approaching speed to the tissue. The control module computes the control input $\mathbf{q}$ in the configuration space, and then $\mathbf{q}$ is mapped to $\mathbf{x}$ in the joint space.

the contact states (see Section VI-B). The control input $\mathbf{q} = \begin{bmatrix} \alpha & \beta & d & q_x & q_y \end{bmatrix}^T$ (q_x and q_y shown in Fig. 1 (a) are camera displacements.) expressed in the configuration space is computed by the *controller* module. Finally, $\mathbf{q}$ in the configuration space is mapped to $\mathbf{x}$ in the joint space by the *mapping* module. The joints of the robot are controlled in position at the low level.

III. Kinematic Modeling

A. Model of the main endoscope carrying the camera

$\mathscr{F}_{c,0}$, $\mathscr{F}_{c,k}$, $\mathscr{F}_{ch,k}$, and $\mathscr{F}_{OCT,k}$ are respectively the coordinate systems attached to the initial camera position, the current camera position, the exit of endoscope channel, and the OCT imaging plane (see Fig. 1 (a)). When bending the main endoscope, the camera locally moves on a sphere whose radius is close to the length of the endoscope bending section (185 mm). This remains true even when the original endoscope configuration is bent. Moreover, in order to reduce contact forces between the endoscope and the surrounding tissues, the lateral displacement of $\mathscr{F}_{c,k}$ with respect to $\mathscr{F}_{c,0}$ should be limited to typically $\pm$ 3cm. For such small movements, the change of orientation and depth of the camera can be neglected. The camera motion can thus be approximately modeled as a translation in the $x-y$ plane of $\mathscr{F}_{c,0}$.

Based on this assumption, the Jacobian matrix $\mathbf{J}_c \in \mathbb{R}^{3\times 2}$ which maps the tendon displacements $\triangle\mathbf{q}_m = \triangle\begin{bmatrix} q_{c1} & q_{c2} \end{bmatrix}^T$ to the camera displacement $\triangle\mathbf{t}_{ck}^{c0} = \triangle\begin{bmatrix} q_x & q_y & 0 \end{bmatrix}^T$ is constant.

$$\triangle\mathbf{t}_{ck}^{c0} \approx \mathbf{J}_c\triangle\mathbf{q}_m \tag{1}$$

where $\triangle *$ is the increment of $*$ between two subsequent sampling times.

For simplicity, we estimate $\mathbf{J}_c$ offline by measuring the displacement of the camera in a controlled environment [12].

B. Kinematic model of the OCT probe

Based on the assumption of constant curvature [13], the discrete-time translational kinematic equation of the OCT probe is derived with respect to $\mathscr{F}_{c,k}$:

$$\mathbf{t}_{OCT,k+1}^{ck} = \mathbf{t}_{OCT,k}^{ck} + \mathbf{J}_{ins}\triangle\mathbf{q}_{ins} \tag{2}$$

where $\mathbf{q}_{ins} = \begin{bmatrix} \alpha & \beta & d \end{bmatrix}^T$ and $\mathbf{J}_{ins} = \partial\mathbf{t}_{OCT}^{ck}/\partial\mathbf{q}_{ins} \in \mathbb{R}^{3\times 3}$ is the position Jacobian matrix for the flexible instrument. It only depends on the current joint position of the instrument.

With respect to $\mathscr{F}_{c,0}$, the position $\mathbf{t}_{OCT}^{c0}$ and the rotation $\mathbf{R}_{OCT}^{c0}$ of the OCT imaging plane are computed respectively as $\mathbf{t}_{OCT}^{c0} = \mathbf{t}_{OCT}^{ck} + \mathbf{t}_{ck}^{c0}$ and $\mathbf{R}_{OCT}^{c0} \approx \mathbf{R}_{OCT}^{ck}$[1]. Then, the discrete-time translational kinematic equation with respect to $\mathscr{F}_{c,0}$ can be obtained as

$$\mathbf{t}_{OCT,k+1}^{c0} = \mathbf{t}_{OCT,k}^{c0} + \mathbf{J}_r\triangle\mathbf{q} \tag{3}$$

with the position Jacobian matrix $\mathbf{J}_r = \begin{bmatrix} \mathbf{J}_{ins} & \mathbf{E}_{3\times 2} \end{bmatrix} \in \mathbb{R}^{3\times 5}$, $\mathbf{E}_{3\times 2} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}^T$.

In order to map both the movement of the camera and the flexible instrument on the endoscopic image plane, we employ a virtual camera with a fixed frame $\mathscr{F}_{c,0}$. On the image plane of the virtual camera, the discrete-time kinematic equation can be rewritten as

$$\mathbf{p}_{i,k+1} = \mathbf{p}_{i,k} + \mathbf{J}\triangle\mathbf{q} \tag{4}$$

where the image Jacobian matrix $\mathbf{J}$ is computed by $\mathbf{J} = \mathbf{L}_s\mathbf{J}_r$. The interaction matrix $\mathbf{L}_s$ of a 3D point feature (X,Y,Z) can be found in [14].

Given the actuation vector $\mathbf{q}_d = \begin{bmatrix} \alpha & \beta \end{bmatrix}^T$, we have the discrete-time z-axis kinematic equation of the OCT probe with respect to both $\mathscr{F}_{c,0}$ and $\mathscr{F}_{c,k}$[2]:

$$\mathbf{n}_{k+1} = \mathbf{n}_k + \mathbf{J}_d\triangle\mathbf{q}_d \tag{5}$$

where the orientation Jacobian matrix $\mathbf{J}_d$ is computed as $\mathbf{J}_d = \partial\mathbf{n}/\partial\mathbf{q}_d$ with $\mathbf{n}$ corresponding to the last column in the orientation matrix of OCT with respect to $\mathscr{F}_{c,k}$.

IV. Image-Guided Control Strategy

In this section, we propose an optimization-based visual control strategy to simultaneously achieve the above-mentioned control objectives while simultaneously enforcing visibility and actuator constraints. First, we explain how the control objectives can be converted to a constrained quadratic programming (QP) problem. Then, we show the design of all its components.

A. Optimization-based high level control

We express the objectives as three optimization problems which are then integrated with the endoscope displacement limitation into a single QP problem. The integrated QP problem is written as:

$$\min_{\triangle\mathbf{q}} \quad \frac{1}{2}\triangle\mathbf{q}^T\mathbf{M}\triangle\mathbf{q} + c^T\triangle\mathbf{q} \tag{6}$$

$$\text{s.t.} \quad \mathbf{S}_l \le \mathbf{A}_c\triangle\mathbf{q} \le \mathbf{S}_u \tag{7}$$

$$\triangle\mathbf{q}_l \le \triangle\mathbf{q} \le \triangle\mathbf{q}_u \tag{8}$$

[1] As explained in the previous section, it is considered that the camera is not subject to changes of orientation

[2] The insertion d does not act on the direction $\mathbf{n}$.

where $\mathbf{M} = \chi_o \mathbf{M}_o + \chi_i \mathbf{M}_i + \chi_d \mathbf{M}_d + \chi_c \mathbf{M}_c + \chi_a \mathbf{M}_a$ and $c = \chi_o c_o + \chi_i c_i + \chi_d c_d + \chi_c c_c$. In the next subsections, we explain the semi-positive definite matrices $\mathbf{M}_o$, $\mathbf{M}_i$, $\mathbf{M}_d$, $\mathbf{M}_c$, the vectors c_o, c_i, c_d, c_c, the *visibility constraint* (7), and the *actuator velocity constraint* (8). By adding an identity matrix $\mathbf{M}_a$ in the objective function, $\mathbf{M}$ becomes positive definite so that solving (6) provides a unique solution.

B. Orientation control

To make sure that the OCT probe is parallel to the surface at the scanned point, we need $\mathbf{n}^T \mathbf{m} = 0$ which can be achieved by solving:

$$\min_{\triangle \mathbf{q}_d} \quad \left[(\mathbf{n}_k + \mathbf{J}_{da} \triangle \mathbf{q})^T \mathbf{m} \right]^2 \tag{9}$$

where $\mathbf{J}_{da} = \begin{bmatrix} \mathbf{J}_d & \mathbf{0}_{3\times 3} \end{bmatrix}$ is the augmented $\mathbf{J}_d$ which is used to take into account all actuated DoFs. (9) can be converted as a QP problem with the following standard form:

$$\min_{\triangle \mathbf{q}} \quad \frac{1}{2} \triangle \mathbf{q}^T \mathbf{M}_o \triangle \mathbf{q} + (\triangle \mathbf{q})^T c_o \tag{10}$$

where $\mathbf{M}_o = \mathbf{J}_{da}^T \mathbf{m} \mathbf{m}^T \mathbf{J}_{da}$ and $c_o = \mathbf{J}_{da}^T \mathbf{m} \mathbf{n}_k^T \mathbf{m}$.

C. Image position control

An image-based visual servoing strategy is employed to achieve the objective $\mathbf{p}_{i,k} = \mathbf{p}_{i,d}$. Given the current image position $\mathbf{p}_{i,k}$ and its desired one $\mathbf{p}_{i,d}$, the control input can be computed by

$$\min_{\triangle \mathbf{q}} \quad \| \mathbf{J} \triangle \mathbf{q} + \mathbf{p}_{i,k} - \mathbf{p}_{i,d} \|_2 \tag{11}$$

In order to compute all the five control variables, the objective function is obtained based on the kinematic equation (4) which is built on the image plane of the virtual camera introduced in Section III-B.

Similarly, (11) can also be converted as a QP problem:

$$\min_{\triangle q} \quad \frac{1}{2} \triangle \mathbf{q}^T \mathbf{M}_i \triangle \mathbf{q} + \triangle \mathbf{q}^T c_i \tag{12}$$

where $\mathbf{M}_i = \mathbf{J}^T \mathbf{J}$ is also semi-positive definite and $c_i = \mathbf{J}^T \left(\mathbf{p}_{i,k} - \mathbf{p}_{i,d} \right)$.

Remark 1. The 2D error vector $\mathbf{p}_{i,k} - \mathbf{p}_{i,d}$ in (11) is supposed to be measured on the image plane of the virtual camera. However, the virtual image plane does no exist for the robotic system. Because of the minor rotation of $\mathscr{F}_{c,k}$ with respect to $\mathscr{F}_{c,0}$, we have $\mathbf{p}_{i,k} - \mathbf{p}_{i,d} \approx \mathbf{p}_{i,k}^{ck} - \mathbf{p}_{i,d}^{ck}$ where $\mathbf{p}_{i,k}^{ck}$ and $\mathbf{p}_{i,d}^{ck}$ are respectively the current and the desired image positions (in pixel) of the probe on the real image plane with respect to $\mathscr{F}_{c,k}$.

D. Contact control

In order to achieve contact between the OCT probe and the tissue, the movement of the probe is controlled along the normal direction $\mathbf{m}$ of the scanned surface. To design the controller, the kinematic equation (3) is projected along $\mathbf{m}$ as

$$p_{c,k+1} = p_{c,k} + \mathbf{m}^T \mathbf{J}_r \triangle \mathbf{q} \tag{13}$$

where $p_c = \mathbf{m}^T \mathbf{t}_{OCT}^{c0}$ is the variable obtained by projecting $\mathbf{t}_{OCT}^{c0}$ along $\mathbf{m}$.

Then, the control inputs to maintain the contact are computed by solving the following optimization problem:

$$\min_{\triangle \mathbf{q}} \quad \| \mathbf{m}^T \mathbf{J}_r \triangle \mathbf{q} + p_{c,k} - p_{c,d} \|_2 \tag{14}$$

where $p_{c,k} - p_{c,d}$ is the distance between the OCT probe and the tissue along the normal at the desired position.

In (14), $p_{c,d}$ is unknown due to the lack of 3D environment perception. Actually, $p_{c,k} - p_{c,d}$ can be used to define the approaching speed, for the probe, to the tissue. Instead of computing it directly, we employ a variable η to capture the speed. η can be estimated from the OCT image at each step (see Section VI-B). (14) can then be converted to:

$$\min_{\triangle \mathbf{q}} \quad \| \mathbf{m}^T \mathbf{J}_r \triangle \mathbf{q} + \eta \|_2 \tag{15}$$

which can, in turn, be converted to a QP problem:

$$\min_{\triangle \mathbf{q}} \quad \frac{1}{2} \triangle \mathbf{q}^T \mathbf{M}_d \triangle \mathbf{q} + \triangle \mathbf{q}^T c_d \tag{16}$$

where $\mathbf{M}_d = \mathbf{J}_r^T \mathbf{m} \mathbf{m}^T \mathbf{J}_r$ and $c_d = \mathbf{J}_r^T \mathbf{m} \eta$. $\mathbf{M}_d$ is also semi-positive definite.

In order to avoid perturbations caused by noise in the OCT images, η is quantized into three values which depend on three contact states (see details in Section VI-B): I ($\eta = \eta_1$ when no object is detected in the OCT image), II ($\eta = \eta_2$ when an object is detected but there is no contact), and III ($\eta = \eta_3$ when contact is detected).

The value of η dictates the speed and direction of the probe movement with respect to the tissue. When no object is detected, the probe should move towards the tissue. As soon as an object is detected in the image, the speed should be reduced, therefore $\eta_1 > \eta_2 > 0$. In order to avoid pushing the surface too much, the OCT probe is controlled to retract slightly away from the tissue when the contact is detected. It is achieved by changing the direction of η, i.e. $\eta_3 < 0$. However, the contact situation is desirable and should be maintained as long as possible. Therefore, we employ a relative small η_3, i.e. $-\eta_3 < \eta_2$.

E. Endoscope displacement limitation

Deflecting the main endoscope, i.e. moving the camera, may be necessary for achieving the scanning objectives. Nevertheless, a limitation for the endoscope displacement is used to avoid strong interactions with surrounding organs.

Given the current camera position $\mathbf{t}_{ck}^{c0}$ with respect to $\mathscr{F}_{c,0}$, the incremental displacement $\triangle \mathbf{t}_{ck}^{c0}$ is computed by solving the optimization problem: $\min_{\triangle \mathbf{t}_{ck}^{c}} \| \mathbf{t}_{ck}^{c0} + \triangle \mathbf{t}_{ck}^{c0} \|_2$ which minimizes the absolute camera translation at each step. Considering that $\triangle \mathbf{t}_{ck}^{c0} = \begin{bmatrix} \mathbf{0}_{3\times 2} & \mathbf{I}_{2\times 2} \end{bmatrix} \triangle \mathbf{q}$ (see Section III-A), this optimization problem can also be converted as the following QP problem with an extended optimization variable $\triangle \mathbf{q}$.

$$\min_{\triangle \mathbf{q}} \quad \frac{1}{2} \triangle \mathbf{q}^T \mathbf{M}_c \triangle \mathbf{q} + \triangle \mathbf{q}^T c_c \tag{17}$$

where $\mathbf{M}_c = \begin{bmatrix} \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 2} \\ \mathbf{0}_{2\times 3} & \mathbf{I}_{2\times 2} \end{bmatrix}$ is semi-positive definite and $c_c =$

$\left[\begin{array}{cc} \mathbf{0}_{1\times 3} & \left(\mathbf{t}_{ck}^{c0}\right)^T \end{array}\right]^T$.

F. Visibility constraint

The visibility constraint is enforced to make sure that the probe stays within the endoscope field of view. The visibility constraint is unrelated to the camera position $\mathscr{F}_{ck}$ because the probe moves with the camera during a camera motion. Therefore, the kinematic model in the image plane for the tip of the probe $\mathbf{s}_m$ with respect to $\mathscr{F}_{c,k}$ can be written as

$$\mathbf{s}_{m,k+1} = \mathbf{s}_{m,k} + \mathbf{J}_{m,ins} \triangle \mathbf{q}_{ins} \tag{18}$$

where $\mathbf{J}_{m,ins} = \mathbf{L}_s \mathbf{J}_{ins}$ is the image Jacobian matrix for the flexible instrument with respect to $\mathscr{F}_{ck}$.

Given the bounds $\left[\begin{array}{cc} \mathbf{s}_{min} & \mathbf{s}_{max} \end{array}\right]$ of the image position ($\mathbf{s}_{m,k}$ in pixel) of the probe tip, the visibility constraint is achieved by

$$\mathbf{s}_{min} - \mathbf{s}_{m,k} \leq \mathbf{J}_{m,ins} \triangle \mathbf{q}_{ins} \leq \mathbf{s}_{max} - \mathbf{s}_{m,k} \tag{19}$$

In this expression, the control horizon is set to one step for the optimization. This could generate abrupt motions or even instabilities because of the limited response time to deal with hard visibility constraints. Instead of using (19) directly, we therefore add a positive parameter $N > 1$ to smooth the motion. Then, (19) can be modified as

$$\mathbf{S}_l / N \leq \mathbf{A}_c \triangle \mathbf{q} \leq \mathbf{S}_u / N \tag{20}$$

where $\mathbf{A}_c = \left[\begin{array}{cc} \mathbf{J}_{m,ins} & \mathbf{0}_{2\times 2} \end{array}\right]$, $\mathbf{S}_l = \mathbf{s}_{min} - \mathbf{s}_{m,k}$, and $\mathbf{S}_u = \mathbf{s}_{max} - \mathbf{s}_{m,k}$. The constraint (20) allows to increase the response time (N steps) in a simple way. It is equivalent to assume that a constant control input $\triangle \mathbf{q}$ is computed and applied for the N future steps.

G. Actuators constraints

Due to constraints on the mechatronic architecture, the actuator values on the configuration space are limited to $\left[\begin{array}{cc} \mathbf{q}_{min} & \mathbf{q}_{max} \end{array}\right]$. Given the current actuator value $\mathbf{q}$, $\triangle \mathbf{q}$ is limited by $\mathbf{q}_{min} \leq \mathbf{q} + \triangle \mathbf{q} \leq \mathbf{q}_{max}$. We also limit its velocity, i.e. $\triangle \mathbf{q} \in \left[\begin{array}{cc} \mathbf{q}_{v,min} & \mathbf{q}_{v,max} \end{array}\right]$. Then, we have $\triangle \mathbf{q} \in [\mathbf{q}_{min} - \mathbf{q}, \mathbf{q}_{max} - \mathbf{q}] \cap [\mathbf{q}_{v,min}, \mathbf{q}_{v,max}]$. The upper bound $\triangle \mathbf{q}_u$ and the lower bound $\triangle \mathbf{q}_l$ are computed respectively as $\triangle \mathbf{q}_u = \min(\mathbf{q}_{v,max}, \mathbf{q}_{max} - \mathbf{q})$ and $\triangle \mathbf{q}_l = \min(\mathbf{q}_{v,min}, \mathbf{q}_{min} - \mathbf{q})$.

H. Parameter selection

The values of the different weighting parameters come from our task priorities. The orientation, image position, and contact control are primary tasks and are more important than the endoscope displacement limitation, which is itself more important than the QP regularization with an identity matrix. Therefore, the parameters χ_c and χ_a should be set to small positive values and we need $0 < \chi_a < \chi_c \ll \min(\chi_o, \chi_i, \chi_d)$. The three main tasks require only four DoFs, whereas the system generally provides five. If the scanned point is reachable, the solution of (6) can make the three objectives hold at the same time for any positive χ_o, χ_i, and χ_d [15].

The software packages qpOASES [16] is used to compute the control inputs in the configuration space. The actuator inputs are then computed by the mapping module.

V. Simulations

We validate the control strategy for positioning and tracking under the situation where the visibility constraint, the actuator constraint, and the endoscope displacement limitation are all activated. The Jacobian matrix in (1) is estimated around the straight configuration based on the theoretical model in [13]. The desired path is defined as $[30 - 30\cos(k/30), -30\sin(k/30), 0]$ (the path is a circle with radius 30 mm) with respect to $\mathscr{F}_{c0}$. This trajectory represents the boundaries of a lesion, which could be assessed before or after Endoscopic Submucosal Dissection. The normal direction of the scanned surface is set to $\mathbf{m} = (-0.5, 0, -0.866025)$. To build the simulation environment, a 3D desired path is defined so that we can compute the 3D tracking error. However, the 3D path is not controlled directly. For the computation of control inputs, only the normal direction of the surface, the 2D desired image position and the contact state are used.

The chosen control parameters are $\eta_1 = 2$ mm, $\eta_2 = 0.2$ mm, $\eta_3 = -0.1$ mm, $\chi_o = \chi_i = \chi_d = 1$, $\chi_c = 10^{-5}$, $\chi_a = 10^{-6}$, and $N = 5$. The range of the actuators for the flexible instrument are limited to $\alpha \in \left[\begin{array}{cc} -4.71 & 6.28 \end{array}\right]$ rad, $\beta \in \left[\begin{array}{cc} -2.1 & 2.1 \end{array}\right]$ rad, and $d \in \left[\begin{array}{cc} 0 & 100 \end{array}\right]$ mm. The velocities of all five actuators are constrained by $\left[\begin{array}{cc} \triangle q_l & \triangle q_u \end{array}\right]$ where $\triangle q_l = -\triangle q_u = [-0.157\text{rad}, -0.157\text{rad}, -2\text{mm}, -2\text{mm}, -2\text{mm}]$. The perception depth of the OCT probe is set to 3 mm. These values are representative of the real STRAS robotic system.

Fig. 4 shows the 3D path of the OCT probe (a), probe mapped on the image plane (b), the camera position (c), and the tracking error (d). The tracking errors are computed by $\mid \mathbf{n}^T \cdot \mathbf{m} \mid$(orientation error), $\parallel \mathbf{p}_i - \mathbf{p}_{i,d} \parallel$(image error in pixels), $\mathbf{m}^T \left(\mathbf{t}_{OCT}^{c0} - \mathbf{t}_{OCT,d}^{c0}\right)$(depth error in mm), and $\parallel \mathbf{t}_{OCT}^{c0} - \mathbf{t}_{OCT,d}^{c0} \parallel$(3D error in mm).

The defined image boundary (red rectangle in Fig. 4 (b)) can effectively enforce the visibility constraint. The endoscope displacement is limited in a small range thanks to $\mathbf{M}_c$ (see Fig. 4 (c)). For the first 30 steps, the desired scanned point remains constant to allow the initial positioning process, during which the OCT probe will reach the first desired point on the trajectory. As shown in Fig. 4 (d), the tracking errors converge to small values. The scanning task is then switched to the tracking process to allow the probe to follow the desired path. The convergence of the image and depth (the distance between the probe and the surface) can ensure the convergence of the probe onto the desired 3D path (see Fig. 4 (a)).

In Fig. 4 (d), the chattering phenomenon is generated by the backward velocity induced by the value of η_3. A smaller value of η_3 would reduce the chattering effect (which also transmits to the endoscopic image, because of slight coupling, and to the 3D position errors). However, backward motions are useful because they avoid pushing too hard on the surface during scanning. In practice, this parameter would need to be adjusted keeping in mind this tradeoff.

Other sets of parameters (χ_o, χ_i, χ_d) have been tested ((100,100,100) and (1,100,1)). These parameters have very limited impact on the positioning, with set (100,100,100) slightly decreasing the maximum error on the orientation (see accompa-

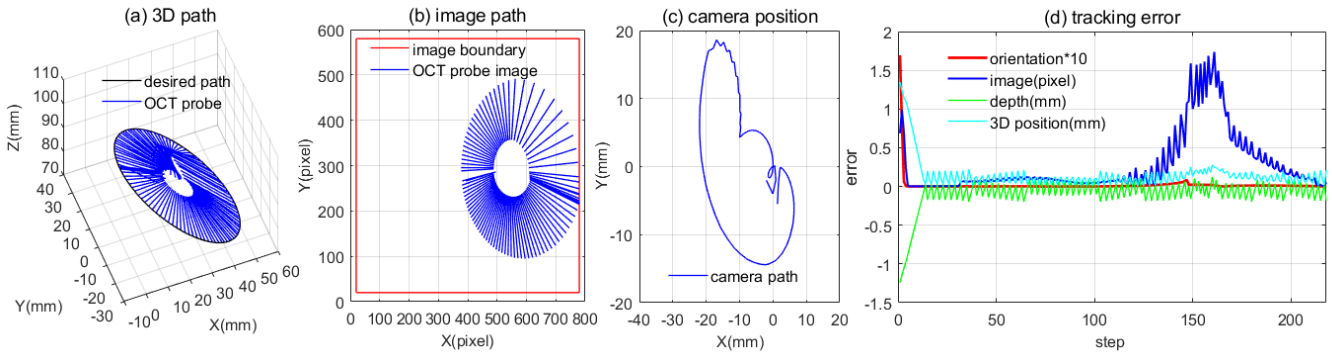

Figure 4. Simulation-based validation for positioning and tracking with visibility constraint. (a), (b), (c), and (d) are respectively the 3D OCT probe, the probe image position, the camera position with respect to $\mathscr{F}_{c0}$, and the tracking errors. In (a), the end of each blue line segment, which remains close to the desired path, is the probe position. The red rectangle in (b) is the image boundary for the visibility constraint. The blue segments in (a) and (b) are drawn every 2 steps.

$\psi(^\circ)$	Errors : average (max)			
	orientation	image (pixel)	depth (mm)	3D (mm)
no error	0.00 (0.01)	0.70 (2.97)	0.09 (0.24)	0.18 (0.47)
33	0.06 (0.22)	0.98 (2.82)	0.10 (0.55)	0.22 (0.58)
62	0.08 (0.29)	1.08 (3.36)	0.18 (1.00)	0.30 (1.22)
84	0.14 (0.47)	1.26 (4.85)	0.48 (2.62)	0.58 (2.66)

Table I
TRACKING ERROR IN PRESENCE OF A RANDOM DISTURBANCE WITHIN A CONE OF APERTURE ψ ON THE TISSUE NORMAL $\mathbf{m}$.

nying video). This global absence of effect is expected as long as the inequalities given in Section IV-H are respected. Our control method works well for different trajectories. Some examples are shown in the accompanying video. Table I shows the influence of adding a random disturbance on the normal direction $\mathbf{m}$ of the tissue at each time step k, simulating estimation errors. If this error stays within a cone of 33 degrees aperture, the orientation control is affected but the overall tracking errors increase only marginally. Larger normal errors, however, affect all indices (orientation, image, depth and 3D errors).

VI. EXPERIMENTAL VALIDATION

This section presents proof-of-concept validation experiments performed on the experimental setup depicted on Fig. 1. Note that the tissue was replaced by a planar board fitted with QR codes. This simplification allows computing easily the normal direction of the tissue surface $\mathbf{m}$, as well as to update the desired path. We first introduce a marker-based method to estimate the pose of the probe. Then, we propose the image processing method to obtain the contact state between the probe and the scanned surface. Finally, we show the trajectory tracking results.

A. OCT probe pose estimation

Due to kinematic modeling errors for the flexible instrument, it is not advisable to compute the 3D pose of the probe based on the analytical model directly. We propose a marker-based pose estimation strategy to estimate the 3D pose using the monocular endoscopic images in real-time. Two markers are attached to the probe with a distance of 25 mm. The strategy consists of two steps: image position estimation and pose estimation using model registration.

A Kalman Filter is employed to estimate the image position of the markers. By implementing prediction, an estimation can be obtained even if the image features are lost, for instance due to marker occlusions or specularities. In the situation where image features are lost, the update step is only based on the analytical model.

The image positions of the two markers are estimated using Kalman Filter as $\mathbf{p}_{reg} = \begin{bmatrix} \mathbf{p}_{1,k} & \mathbf{p}_{2,k} \end{bmatrix}^T$ at the sampling time k. We denote $\mathbf{t}^{ck}_{reg} = \begin{bmatrix} \mathbf{t}^{ck}_{1,k} & \mathbf{t}^{ck}_{2,k} \end{bmatrix}^T$ and $\mathbf{J}_{reg} = \begin{bmatrix} \mathbf{J}_{1,k} & \mathbf{J}_{2,k} \end{bmatrix}^T$ as respectively the markers' 3D position and their corresponding Jacobian matrices with respect to $\mathscr{F}_{c,k}$. The measurement matrix $\mathbf{H}_{reg} = \begin{bmatrix} \mathbf{H}_{1,k} & \mathbf{H}_{2,k} \end{bmatrix}^T$ is employed to map the 3D positions to image positions with $\mathbf{H}_{i,k} = \frac{1}{Z_i} \begin{bmatrix} \mathbf{K} & \mathbf{p}_0 \end{bmatrix}$ where Z_i is the depth of ith marker and $\mathbf{p}_0$ is the location of the principal point in the image plane, which are known from camera calibration. Motivated by the work in [17], an optimization-based method is proposed for the registration:

$$\min_{\triangle \mathbf{q}_{ins}} \quad \| \mathbf{H}_{reg} \left(\mathbf{t}^{ck}_{reg} + \mathbf{J}_{reg} \triangle \mathbf{q}_{ins} \right) - \mathbf{p}_{reg} \|_2 \tag{21}$$

which can be converted to a QP problem:

$$\min_{\triangle \mathbf{q}_{ins}} \quad \frac{1}{2} (\triangle \mathbf{q}_{ins})^T \mathbf{M}_{reg} \triangle \mathbf{q}_{ins} + (\triangle \mathbf{q}_{ins})^T c_{reg} \tag{22}$$

where $\mathbf{M}_{reg} = (\mathbf{H}_{reg}\mathbf{J}_{reg})^T (\mathbf{H}_{reg}\mathbf{J}_{reg})$ and $c_{reg} = \mathbf{H}_{reg}\mathbf{J}_{reg} \left(\mathbf{H}_{reg}\mathbf{t}^{ck}_{reg} - \mathbf{p}_{reg} \right)$. The variables α, β, and d are computed by accumulating the incremental control input $\triangle \mathbf{q}_{ins}$. They are used to compute the probe's 3D pose based on the direct kinematic model. Moreover, all the Jacobian (except J_c) are expressed analytically in function of α, β, and d. The estimates are used to obtain numerical values on-line. In order to have a unique solution for (22), at least two markers are needed to recover the three variables.

B. OCT image processing

OCT images can be used to detect the contact state which is needed for the control design. In Fig. 5 (d) showing a B-scan OCT image in polar coordinates, the white pixels correspond to the sheath of the probe and the canned surface in the field of view. To detect the surface, we first calibrate the values of h_s

Algorithm 1 Implementation of configuration control

1. **Initialization**: $\eta_{1,2,3}, N, \chi_{o,i,d,c,a}, e_{o,t}, e_{i,t}, k=0, \mathbf{J}_c, \mathbf{p}_{i,d}$
2. **For** each step **do**
3. $\quad \mathbf{p}_{i,d}(k) \leftarrow$Update desired 2D image position
4. $\quad \mathbf{m} \leftarrow$Update normal direction
5. $\quad \mathbf{p}_{i,k}, n \leftarrow$Pose estimation
6. $\quad \eta \leftarrow$Contact detection
7. $\quad e_0, e_i \leftarrow$Compute tracking errors
8. $\quad \mathbf{S}_l, \mathbf{S}_u \leftarrow$Define incremental image boundary
9. $\quad \triangle\mathbf{q}_l, \triangle\mathbf{q}_u \leftarrow$Define incremental actuator boundary
10. $\quad \mathbf{M}, c, \mathbf{A}_c \leftarrow$Compute components for QP
11. $\quad \triangle\mathbf{q} \leftarrow$Solve (6)
12. $\quad \mathbf{q} \leftarrow$Accumulate $\triangle\mathbf{q}$
13. $\quad$ Robot actuation
14. $\quad$ **If** $e_o < e_{o,t}$, $e_i < e_{i,t}$, and contact is detected
15. $\quad\quad k = k+1$
16. **End**

and h_m, respectively the thickness of the probe sheath and of the marker in the image. Pixels closer than $h_s + h_m$ pixels from the top of the image are discarded. An image pixel is treated as a noise if the number of its neighbor image pixels is smaller than a predefined threshold. We can find the contour of each group of pixels. If the number of image pixels within a contour is smaller than another predefined threshold, the corresponding group of pixels will also be filtered out.

The remaining white pixels correspond to the surface. h_i is then obtained as the shortest vertical distance between the OCT probe and the tissues (see Fig. 5 (d)). Finally, the contact state is obtained based on the relation between h_i and $h_s + h_m$: contact detected ($h_i = h_s + h_m$), object detected but no contact ($h_i > h_s + h_m$) and no object (no white pixels in the filtered image). Fig. 5 shows an example of OCT image with h_i, h_s, and h_m (d), as well as the corresponding processed image (e).

C. Implementation of the control strategy

The overall implementation of the proposed control method is summarized in Algorithm 1.

The normal direction $\mathbf{m}$ of the QR board can be obtained using the CharucoBoard Class in OpenCV. $\mathbf{m}$ is estimated at each sampling time in order to deal with situations where the normal direction changes locally during the scanning. However, for simplifying the experiments, we employ a fixed planar surface with a constant $\mathbf{m}$.

We record the orientation error e_o, the image error e_i, and the contact state at each sampling time. In the absence of an external sensor which could provide ground truth, the orientation error is computed from state estimates. Due to model errors (such as backlash), it can be difficult to bring errors to small values in a single step. A simple and reasonable strategy is to define two threshold errors $e_{o,t}$ and $e_{i,t}$. If $e_o \geq e_{o,t}$, or $e_i \geq e_{i,t}$, or no contact is detected, a positioning task is employed instead of tracking the desired path continuously. When $e_o < e_{o,t}$, $e_i < e_{i,t}$, and the contact is detected, we switch the controller to the tracking mode and OCT images are recorded for medical analysis.

D. Trajectory tracking results with visibility constraints

At the beginning of the scanning, the robot is telemanipulated to a position close to the area to scan. Then, the initial desired path is defined on the image plane. At each step, we estimate the image position and the orientation of the probe (see (c) in Fig. 5). Besides, both the normal direction of the scanned surface (see (a) in Fig. 5) and the desired 2D image position (see (b) in Fig. 5) are updated to take into account the camera movement for the computation of the control inputs. The thresholds of the tracking errors are defined as $e_{o,t} = 0.1$ mm and $e_{i,t} = 10$ pixels. A screenshot of the experiment is shown in Fig.5 where the visibility constraint is activated within a user-defined rectangle (see Fig. 5 (b)).

In Fig. 6, we illustrate the image position of the probe (a), the tracking error without employing the error threshold (b) and tracking error with error threshold (c). The desired image path (see (a)) is gathered after updating the initial desired position at each sampling time. In (b), the first 30 steps are used for the initial positioning task to reduce the initial error and the tracking task is carried out during the remaining steps.

For the tracking process without using the error threshold (see (b), step 30-320), the average and maximum image position errors are respectively 5.61 pixels (0.679 mm) and 17.4 pixels (2.14 mm). However, for the tracking process with error threshold (see (c)), the average and maximum image position errors reduce to 4.64 pixels (0.563 mm) and 10 pixels (1.39 mm), the latter being equal to be the $e_{i,t}$ threshold. The orientation errors for both cases converge to a small range with an average error of 0.016. Note that this error is an estimation computed from instrument state. It could differ from the real error. But practically we observe the expected behaviour of the probe direction. It can be noted that the scheme is robust to robot model errors, and to approximations made on Jacobians. All points on the path can be accurately scanned, ensuring the complete assessment of the desired line on the surface, which is a desired feature for optical biopsies. This experiment therefore shows the feasibility of automatic path following for OCT scanning of a surface. Moreover, realizing such a path tracking by telemanipulating the robot while maintaining contact and adequate orientation would be almost impossible which provides a real practical interest.

VII. Conclusion

We have developed an optimization-based visual servoing control framework to provide automatic scanning using a flexible endoscopic robot embedded with an OCT probe providing B-scans images. The proposed control framework allows to track a desired path on a surface using the information provided by the endoscopic camera and the OCT probe. The framework is able to handle all degrees of freedom consistently, whatever the current state of the probe with respect to the surface. The proposed method has been validated using both simulation and experiment for automatic scanning. Our main aimed application is the pre-operative and post-operative assessment of the boundaries of cancerous lesions in colorectal surgery.

One of the remaining limitations is the modeling which does not consider nonlinear phenomena, like backlash. This mod-

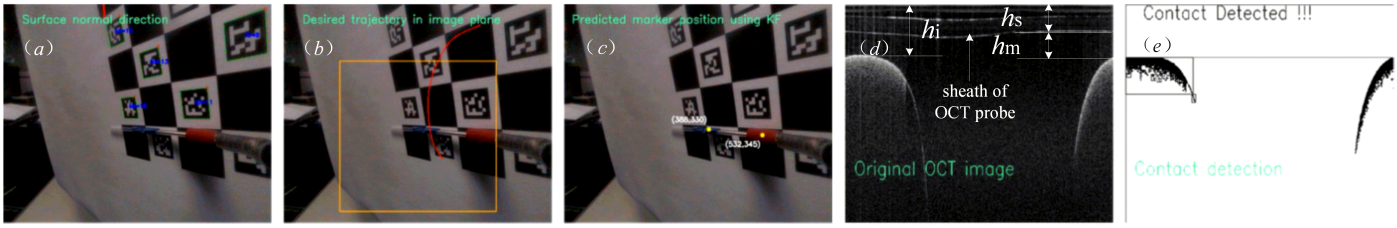

Figure 5. Screenshot of the tracking experiment. The surface normal direction is estimated based on the perception of the QR codes in (a). In (b), the desired image path is updated and the image boundary is defined. The predicted image position of the markers is shown in (c). The original OCT image and the processed image for the contact detection are shown in (d) and (e), respectively.

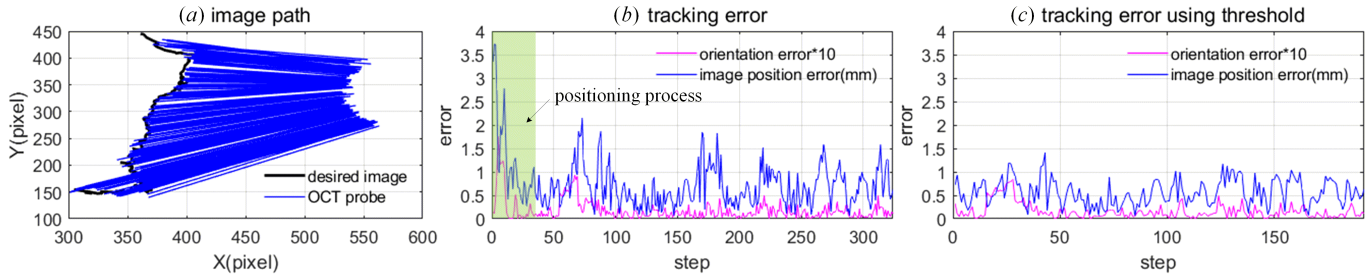

Figure 6. Tracking performances on the image plane. (a) is the image position and orientation of the OCT probe. (b) and (c) are the tracking errors without and with error threshold, respectively. In (a), the position of the OCT beam lies at the end of each blue line segment, which is close to the desired path. For visualization purposes, the orientation errors are increased by 10 times in (b) and (c). The colored area (step 0-30) in (b) is the positioning process.

eling simplification results in larger tracking errors especially when the actuators driving tendons change direction. An error compensation step would be useful to increase the scanning speed and decrease time spent outside of the desired trajectory. While the control framework provides satisfactory results in the proposed setup, some limitations will have to be tackled in order to get closer to a real medical application. As a proof of concept, the scanned tissue is replaced with a board with QR codes and the probe is equipped with markers. Future work will investigate the use of computer vision techniques, such as [18], to estimate the normal of the tissue surface and the pose of the OCT probe.

REFERENCES

[1] M. J. Gora, M. J. Suter, G. J. Tearney, and X. Li, "Endoscopic optical coherence tomography: technologies and clinical applications," *Biomedical optics express*, vol. 8, no. 5, pp. 2405–2444, 2017.

[2] G. Del Giudice, L. Wang, J.-H. Shen, K. Joos, and N. Simaan, "Continuum robots for multi-scale motion: Micro-scale motion through equilibrium modulation," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2537–2542, IEEE, 2017.

[3] O. C. Mora, P. Zanne, L. Zorn, F. Nageotte, N. Zulina, S. Gravelyn, P. Montgomery, M. de Mathelin, B. Dallemagne, and M. J. Gora, "Steerable oct catheter for real-time assistance during teleoperated endoscopic treatment of colorectal cancer," *Biomedical Optics Express*, vol. 11, no. 3, pp. 1231–1243, 2020.

[4] Y. Baran, K. Rabenorosoa, G. J. Laurent, P. Rougeot, N. Andreff, and B. Tamadazte, "Preliminary results on oct-based position control of a concentric tube robot," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3000–3005, IEEE, 2017.

[5] B. Keller, M. Draelos, G. Tang, S. Farsiu, A. N. Kuo, K. Hauser, and J. A. Izatt, "Real-time corneal segmentation and 3d needle tracking in intrasurgical oct," *Biomedical Optics Express*, vol. 9, no. 6, pp. 2716–2732, 2018.

[6] L.-A. Duflot, B. Tamadazte, N. Andreff, and A. Krupa, "Wavelet-based visual servoing using oct images," in *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*, pp. 621–626, IEEE, 2018.

[7] M. Ourak, B. Tamadazte, and N. Andreff, "Partitioned camera-oct based 6 dof visual servoing for automatic repetitive optical biopsies," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2337–2342, IEEE, 2016.

[8] F. Nageotte, P. Zanne, C. Doignon, and M. De Mathelin, "Stitching planning in laparoscopic surgery: Towards robot-assisted suturing," *The International Journal of Robotics Research*, vol. 28, no. 10, pp. 1303–1321, 2009.

[9] E. Malis, G. Morel, and F. Chaumette, "Robot control using disparate multiple sensors," *The International Journal of Robotics Research*, vol. 20, no. 5, pp. 364–377, 2001.

[10] M. C. Yip and D. B. Camarillo, "Model-less hybrid position/force control: a minimalist approach for continuum manipulators in unknown, constrained environments," *IEEE Robotics and Automation Letters*, vol. 1, no. 2, pp. 844–851, 2016.

[11] L. Zorn, F. Nageotte, P. Zanne, A. Legner, B. Dallemagne, J. Marescaux, and M. de Mathelin, "A novel telemanipulated robotic assistant for surgical endoscopy: Preclinical application to esd," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 4, pp. 797–808, 2017.

[12] L. Ott, F. Nageotte, P. Zanne, and M. de Mathelin, "Robotic assistance to flexible endoscopy by physiological-motion tracking," *IEEE Transactions on Robotics*, vol. 27, no. 2, pp. 346–359, 2011.

[13] 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.

[14] F. Chaumette and S. Hutchinson, "Visual servo control. i. basic approaches," *IEEE Robotics & Automation Magazine*, vol. 13, no. 4, pp. 82–90, 2006.

[15] Z. Zhang, J. Dequidt, J. Back, H. Liu, and C. Duriez, "Motion control of cable-driven continuum catheter robot through contacts," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1852–1859, 2019.

[16] H. J. Ferreau, C. Kirches, A. Potschka, H. G. Bock, and M. Diehl, "qpoases: A parametric active-set algorithm for quadratic programming," *Mathematical Programming Computation*, vol. 6, no. 4, pp. 327–363, 2014.

[17] Z. Zhang, A. Petit, J. Dequidt, and C. Duriez, "Calibration and external force sensing for soft robots using an rgb-d camera," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2356–2363, 2019.

[18] L. Li, X. Li, S. Yang, S. Ding, A. Jolfaei, and X. Zheng, "Unsupervised learning-based continuous depth and motion estimation with monocular endoscopy for virtual reality minimally invasive surgery," *IEEE Transactions on Industrial Informatics*, 2020.