

# Automated in-plane OCT-probe Positioning Towards Repetitive Optical Biopsies

M. Ourak<sup>1</sup>, A. De Simone<sup>1</sup>, B. Tamadazte<sup>1</sup>, G. J. Laurent<sup>1</sup>, A. Menciassi<sup>2</sup>, and N. Andreff<sup>1</sup>

**Abstract**—This paper proposes the design of a vision-guided control law for microrobotic-assisted biomedical applications. More precisely, the developed control law is to servo an optical coherence tomography (OCT) system in real-time to carry out repetitive optical biopsy tasks. The OCT images are simultaneously used to perform the optical biopsy and to control the sample holder (microrobotic work-flow). Instead of extracting visual features from the OCT images, the vision-based controller uses the concept of frequency domain information to compute relative motion between two successive OCT B-scan (cross-section) images. Therefore, the visual controller is designed to minimize the error between current and desired images by controlling the microrobotic platform. The proposed approach was experimentally validated, demonstrating more than satisfactory results especially in terms of accuracy, convergence, and robustness.

## I. INTRODUCTION

Currently, endoscopic diagnosis is the primary method for screening and diagnosing cancer and other diseases in internal organs accessible by a natural orifice and often it requires biopsies for *ex-vivo* histopathological examination of the tissue. Early diagnosis is an important issue for efficient treatments of most pathologies and reduction of care costs. Traditional biopsy is an invasive, time-consuming, and destructive procedure that causes delay in diagnosis, possibility of sampling errors and risk of tissue contamination. In traditional endoscopy, it could sometimes be difficult to assess the best area to biopsy as minute lesions are not easily detected, especially in the esophagus, stomach and colon, increasing the false positive rate [1]. Conventional white-light endoscopy includes light-based endoscopes and videoscopes (fiber-optics and cameras) with the aim of improving the sensitivity and specificity of current endoscopic systems. Nowadays, new techniques have been developed achieving better resolution and contrast between normal and damaged tissues [2]. Recent advances in fiber-optics, light sources and detectors allow to perform *optical biopsy*, that is using the light and the optical properties of tissues to perform real-time, and *in-situ* examination [3]. Optical biopsy images can be useful in many clinical scenarios for:

(a) guiding biopsy procedures, (b) reducing sampling errors and costs; (c) reducing the need for surgical removal of tissue samples, and (d) providing real-time feedback on surgical and microsurgical procedures [4].

The most promising optical biopsy techniques are certainly confocal microendoscopy and optical coherence tomography. In this paper, we focus especially on the OCT-based 2D optical biopsy. OCT is based on the principle of low coherence interferometry providing very good lateral and axial resolutions: 4  $\mu\text{m}$  and 3  $\mu\text{m}$ , respectively. It is also able to reach higher penetration depths 1-5 mm compared to only 250  $\mu\text{m}$  for confocal systems. OCT was firstly applied in ophthalmology due to the transparent nature of eyes, their minimal scattering and high light penetration. Recently, imaging of non-transparent tissues has been achieved using longer wavelengths, near infrared, where optical scattering is reduced, allowing OCT applications in other medical fields, such as dermatology. OCT systems are fiber-optic based and can be easily integrated in endoscopic systems for imaging of many internal organs, with applications in cardiology, gastroenterology, and urology, etc. Although many diagnostic endoscopic OCT probes have been reported in the literature, as far as we know, none of these systems use OCT images to guide the endoscope. Thus, vision-based control known as *visual servoing* is a promising approach for robotic control, especially when flexibility, accuracy and robustness are required. In addition to numerous of visual servoing purposes in conventional robotic achievements (using white-light cameras), some studies investigate ultrasound (US) based control for 2D or 3D medical robotic control [7], [8]. As example, echography examination can be realized on a teleoperation mode using an US system which is fixed on a robotic structure so-called US holder [9]. Generally, this teleechography procedures are performed using a US images-based visual servoing essentially to maintain the visibility, in the echography images, of a selected organ [8].

The aim of this paper is to perform B-scan OCT image-based visual servoing in order to control the three degrees-of-freedom (DOF) ( $\mathcal{T} = SE(2) = \mathfrak{R}(2) \times SO(1)$ ) in-plane positioning of a robotic platform with respect to a biological sample. The remaining out-of-plane positioning is not dealt with in this paper since it can be obtained via photometric visual servoing on conventional images. The B-scan images used for the computation of the control law are the cross-sectional OCT images corresponding to a successive single *xy* or *zy* slices of the sample. Instead to use geometrical visual features (i.e., points, lines, etc.) in the design of the interaction matrix that would be hard to compute from poorly

This work is conducted with a financial support from the project NEMRO (ANR-14-CE17-0013-01) funded by the ANR and the financial support of the Franche-Comté region (FRANCHIR), France. It is also performed in the framework of the Labex ACTION (ANR-11-LABEX-01-001)

<sup>1</sup>M. Ourak, A. De Simone, B. Tamadazte, G.J. Laurent, and N. Andreff are with FEMTO-ST Institute, AS2M department, Univ. Bourgogne Franche-Comté/CNRS/ENSMM, 24 rue Savary, 25000 Besançon, France. [brahim.tamadazte@femto-st.fr](mailto:brahim.tamadazte@femto-st.fr)

<sup>2</sup>A. Menciassi is with the BioRobotics Institute, Scuola Superiore Sant'Anna Viale Rinaldo Piaggio 34, Pontedera, Pisa, 56025 Italy. [arianna.menciassi@sssup.it](mailto:arianna.menciassi@sssup.it)

structured OCT images (i.e., very difficult to extract visual features from OCT images), our technique uses the wavelet information in the frequency domain to build the OCT-based visual controller.

The paper is organized as follow: Section II presents the basics of optical biopsy and OCT operating system. Section III describes the designed OCT-based visual servoing. The experimental set-up and materials are discussed in Section IV while Section V reports the obtained experimental results.

## II. OCT PRINCIPLE

OCT is based on the measure of the echo time delay and the intensity of the back-scattered light from the sample that provides information about the different scattering layers of the tissue. The working principle is similar to ultrasound imaging but, since light travels faster than sound, a direct time of flight measurement is impossible. This is indirectly estimated comparing the back-scattered light from the sample with light that has travelled a known reference path, using a *Michelson* interferometer (Fig. 1(a)).

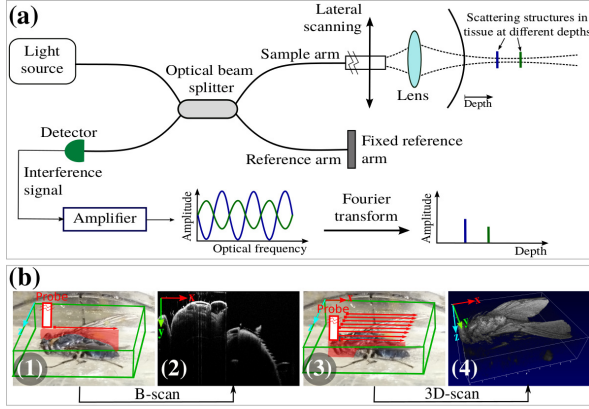


Fig. 1: (a) principle of the spectral-domain OCT system, and (b) OCT system acquisition: (b-1) an image of a fly with B-scan axis, (b-2) B-scan OCT image, (b-3) 3D-scan axes, and (b-4) 3D-scan OCT image.

Light from a low coherence source is split into reference and sample arm. In the reference arm, light is reflected by a mirror, whereas in the sample arm it is focused on the sample by objective lenses. Then, the back-scattered light from the sample interferes with the light returning from the reference arm; the signal is revealed by a photo-detector and sent to the signal processing unit to be analyzed. This interferometric signal is detected only when there is a constructive interference between the two signals i.e., the lengths of sample and reference path are matched. In Time-Domain OCT (TD-OCT) this is achieved by scanning the reference mirror in order to record the signals from all layers, whereas in Fourier-Domain OCT (FD-OCT) the system is fixed recording the spectral interference pattern. The depth reflectivity profile of the sample at the focal spot (A-scan i.e. 1D signal) is then obtained by performing the Fourier transform of the recorded spectral signal. Moving the focused

beam across the sample in a straight line generate 2D cross-sectional image (B-scan) which is the concatenation of A-scans [10]. In our work, the second family of OCT i.e., FD-OCT is considered (Fig. 1(a)).

## III. OCT-BASED VISUAL SERVOING

### A. Planar Pose Estimation from OCT images

1) *Wavelet Transform*: Wavelet transform has emerged as an useful mathematical tool for time-frequency decomposition of signals, as a complement to standard Fourier analysis that decomposes the signal in frequency components only and cannot transform non-stationary signals. Further, the wavelet decomposes the signal into a set of elementary waveforms called *wavelets* formed by dilatation, translation and rotation of a *Mother function* denoted  $\psi$  [11].

Let  $\mathbf{I}(x, y) \in \mathbb{R}^2$  be an image, its continuous wavelet transform is defined by

$$\mathcal{W}(a, b, \Phi) = \left\langle \mathbf{I}(x) | \overline{\Psi_{(a, b, \Phi)}(x)} \right\rangle \quad (1)$$

$$= \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} \mathbf{I}(x) \overline{\Psi_{(a, b, \Phi)}(x)} d^2x \quad (2)$$

where  $\mathcal{W}$  is the wavelet signal,  $a \in \mathbb{R}_*$ ,  $b \in \mathbb{R}^2$  and  $\Phi \in \mathbb{R}$  are the scale, translations and rotation parameters of  $\mathcal{W}$ , respectively, and  $\Psi_{(a, b, \Phi)}(x)$  is the complex conjugate of the mother function  $\psi$ .

In order to maintain the continuity of the signal and choose the orientation of the representation, we select the anisotropic 2D *Morlet* mother function  $\psi$  defined in [12] by

$$\psi(a, b, \Phi) = \psi\left(\frac{\mathbf{r}_{-\Phi}(x - b)}{a}\right) \quad (3)$$

with  $\mathbf{r}_{-\Phi}$  is the  $3 \times 3$  rotation matrix carried by the angle  $\Phi$ ,

$$\text{given by } \mathbf{r}_{-\Phi} = \begin{pmatrix} \cos(\Phi) & -\sin(\Phi) & 0 \\ \sin(\Phi) & \cos(\Phi) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

2) *Wavelets in the Spectral Domain*: The method presented in this paper uses the continuous wavelet transform in spectral domain to position B-scan images that are translated and rotated (i.e.,  $\mathcal{T} = SE(2)$ ) with respect to one another as defined in [13].

Let us consider  $\mathbf{I}(x, y)$  (expressed in the camera frame  $\mathcal{R}_c$ ) and  $\mathbf{I}^*(x, y)$  (expressed in  $\mathcal{R}_{c^*}$ ) the current and desired images, respectively.

a) *Case of a pure translation between the desired and the current image*: Let us consider a pure translation  ${}^{c^*}\mathbf{t}_c = (\delta x, \delta y, 0)^\top$  in the current camera frame  $\mathcal{R}_c$  regarding to the camera frame in the desired position  $\mathcal{R}_{c^*}$ . The wavelet transform  $\mathcal{W}_\mathbf{I}$  of the current image can be expressed according to that of the desired position  $\mathcal{W}_{\mathbf{I}^*}$  after translation by

$$\mathcal{W}_\mathbf{I}(x, y) = \mathcal{W}_{\mathbf{I}^*}(x - \delta x, y - \delta y) \quad (4)$$

By applying a Fourier transform to (4), it is possible to split the contributes of the translations as

$$\mathcal{F}\mathcal{W}_\mathbf{I}(\xi, \eta) = \left( \mathcal{F}\mathcal{W}_{\mathbf{I}^*}(\xi, \eta) \right) e^{-j2\pi(\xi\delta x + \eta\delta y)} \quad (5)$$

where  $\mathcal{FW}_I$  and  $\mathcal{FW}_{I^*}$  are the spectral wavelet signal of  $\mathcal{W}_I$  and  $\mathcal{W}_{I^*}$ , respectively, expressed in the 2D frequency  $(\xi, \eta)$  domain. Consequently, the cross-power spectrum of the spectral wavelet signals yields

$$\frac{\mathcal{FW}_I(\xi, \eta) \overline{\mathcal{FW}_{I^*}(\xi, \eta)}}{|\mathcal{FW}_I(\xi, \eta) \mathcal{FW}_{I^*}(\xi, \eta)|} = e^{-j2\pi(\xi\delta x + \eta\delta y)} \quad (6)$$

where  $\overline{\mathcal{FW}_{I^*}(\xi, \eta)}$  is the complex conjugate of  $\mathcal{FW}_{I^*}(\xi, \eta)$ .

From (6), it is possible now to get the phase shift and then the translation  ${}^{c*}\mathbf{t}_c$  between  $\mathbf{I}$  and  $\mathbf{I}^*$ .

b) *Case of a planar motion between the desired and the current image:* Let us now consider the case of translation  ${}^{c*}\mathbf{t}_c$  and rotation  $\boldsymbol{\theta}\mathbf{u} = (0, 0, \delta\theta)^\top$  that is the axes/angle representation of the rotation between  $\mathbf{I}(x, y)$  and  $\mathbf{I}^*(x, y)$ , then the wavelet representation is written as

$$\begin{aligned} \mathcal{W}_I(x, y) = \mathcal{W}_{I^*}(x \cos \delta\theta + y \sin \delta\theta - \delta x, \\ -x \sin \delta\theta + y \cos \delta\theta - \delta y) \end{aligned} \quad (7)$$

Furthermore, the Fourier transform of (7) splits the contributions of translational motions as

$$\begin{aligned} \mathcal{FW}_I(\xi, \eta) = \left( \mathcal{FW}_{I^*}(\xi \cos \delta\theta + \eta \sin \delta\theta, \right. \\ \left. -\xi \sin \delta\theta + \eta \cos \delta\theta) \right) e^{-j2\pi(\xi\delta x + \eta\delta y)} \end{aligned} \quad (8)$$

However, the contribution of the rotation only appears in the magnitude of (8). To extract it, we need to compute the magnitude  $\mathcal{M}\mathcal{FW}_I$  and  $\mathcal{M}\mathcal{FW}_{I^*}$  of the spectral signal  $\mathcal{FW}_I$  and  $\mathcal{FW}_{I^*}$ , respectively. This leads to write

$$\begin{aligned} \mathcal{M}\mathcal{FW}_I(\xi, \eta) = \mathcal{M}\mathcal{FW}_{I^*}(\xi \cos \delta\theta + \eta \sin \delta\theta, \\ -\xi \sin \delta\theta + \eta \cos \delta\theta) \end{aligned} \quad (9)$$

Thus, it is possible to extract the rotation from (9) by applying the polar representation of the magnitude as follow

$$\mathcal{M}\mathcal{FW}_I(\rho, \theta) = \mathcal{M}\mathcal{FW}_{I^*}(\rho, \theta - \delta\theta) \quad (10)$$

where  $\rho$  is the polar coordinates representation.

To compute the rotation motion, it amounts to reproduce the operations (5) and (6).

## B. OCT-based 2D Pose Visual Servoing

1) *Background:* Generally, the aim of visual servoing scheme is to minimize the error  $\mathbf{e}(t)$  between the current and the desired visual features  $\mathbf{s}(t)$  and  $\mathbf{s}^*$ , respectively [6].

$$\mathbf{e}(t) = \mathbf{C}(\mathbf{s}(\mathbf{m}(t)) - \mathbf{s}^*) \quad (11)$$

where  $\mathbf{C} \in \mathbb{R}^{6 \times k}$  is a combination matrix of full rank, and  $\mathbf{s}(t)$  depends on a set of image measurements  $\mathbf{m}(t)$ .

Thereby, visual servoing controller relies on the establishment of a relationship between the time-variation of  $\mathbf{s}$  and the camera velocity tensor  $\mathbf{v} = (v_x, v_y, v_z, \omega_x, \omega_y, \omega_z)^\top$ . This relationship is given by

$$\dot{\mathbf{s}} = \mathbf{L}_e \mathbf{v} \quad (12)$$

where  $\mathbf{L}_e \in \mathbb{R}^{k \times 6}$  is the *interaction matrix*.

Using (11) and (12), it is possible to link the camera velocity to the time-variation of the error  $\dot{\mathbf{e}}$  of the task-function by writing

$$\dot{\mathbf{e}} = \mathbf{C} \mathbf{L}_e \mathbf{v} \quad (13)$$

Imposing the exponential decrease of the error and using (13), we can write

$$\mathbf{v} = -\lambda \mathbf{L}_e^+ \mathbf{e} \quad (14)$$

where  $\mathbf{L}_e^+ \in \mathbb{R}^{6 \times k}$  is the *Moore-Penrose* pseudo-inverse of  $\mathbf{L}_e$ . In a real system, it is impossible to know perfectly this matrix, thus we need to use its approximation  $\widehat{\mathbf{L}_e^+}$ . Consequently, (14) becomes

$$\mathbf{v} = -\lambda \widehat{\mathbf{L}_e^+} \mathbf{e} \quad (15)$$

with  $\widehat{\mathbf{L}_e^+}$  which can be computed at each sampling time or at the desired configuration, and  $\lambda$  which is a positive gain. This control is trivially Lyapunov stable if  $\mathbf{L}_e \mathbf{L}_e^+$  is definite positive.

2) *Proposed Control:* As it can be highlighted, using the global image information in the frequency domain provides a direct estimation in-plane of translations and rotation (i.e., the 2D pose) of the OCT probe (receptively a camera). Consequently, it is trivial to choose a pose-based visual servoing technique to servo the OCT probe.

Let us consider  $\mathbf{s} = [{}^{c*}\mathbf{t}_c, \boldsymbol{\theta}\mathbf{u}]^\top$ ,  $\mathbf{s}^* = 0$ ,  $\mathbf{e} = 0$  the current and desired poses, respectively. From [6], it is possible to define the following decoupled interaction matrix

$$\mathbf{L}_e = \begin{bmatrix} {}^{c*}\mathbf{R}_c & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{L}_{\boldsymbol{\theta}\mathbf{u}} \end{bmatrix} \quad (16)$$

where  ${}^{c*}\mathbf{R}_c$  is the rotation matrix between the current and the desired camera frame, and  $\mathbf{L}_{\boldsymbol{\theta}\mathbf{u}}$  is given by

$$\mathbf{L}_{\boldsymbol{\theta}\mathbf{u}} = \mathbb{I}_{3 \times 3} - \frac{\theta}{2} [\mathbf{u}]_{\wedge} + \left( 1 - \frac{\text{sinc}(\theta)}{\text{sinc}^2(\frac{\theta}{2})} \right) [\mathbf{u}]_{\wedge}^2 \quad (17)$$

where  $\mathbb{I}$  is the identity matrix, and *sinc* is the sinus cardinal.

If we take into account the fact that translation and rotation are perfectly decoupled from each other. Then, it is possible to write the following simple controller scheme by projecting the 3D Pose visual servoing controller [6] onto the planar constraints

$$\begin{cases} v_c &= -\lambda \text{diag}(1, 1, 0) {}^{c*}\mathbf{R}_c^\top {}^{c*}\mathbf{t}_c \\ \omega_c &= -\lambda \text{diag}(0, 0, 1) \boldsymbol{\theta}\mathbf{u} \end{cases} \quad (18)$$

where  $v_c$  is the translation velocity, and  $\omega_c$  is the rotation velocity in the camera frame  $\mathcal{R}_c$ .

As shown in Fig. 2, the experimental setup is an *eye-to-hand* configuration. Thus, the relation between the robot velocity  $\dot{\mathbf{q}}$  and the camera  $\mathbf{v}$  is

$$\dot{\mathbf{q}} = -{}^e\mathbf{K}_e^{-1}(\mathbf{q}) {}^e\mathbf{M}_o {}^o\mathbf{M}_c \mathbf{v} \quad (19)$$

where  ${}^e\mathbf{K}_e^{-1}(\mathbf{q})$  is the inverse kinematic Jacobian expressed in the end effector frame  $\mathcal{R}_e$ ,  ${}^e\mathbf{M}_o$  is the twist transformation matrix from the robot base frame  $\mathcal{R}_o$  to the end effector frame  $\mathcal{R}_e$ , and  ${}^o\mathbf{M}_c$  is the twist transformation matrix from

the camera frame  $\mathcal{R}_c$  to the robot base frame  $\mathcal{R}_b$ . The latter two are of the form

$${}^b\mathbf{M}_a = \begin{bmatrix} {}^b\mathbf{R}_a & [{}^b\mathbf{t}_a]_{\wedge} [{}^b\mathbf{R}_a] \\ \mathbf{0}_{3 \times 3} & {}^b\mathbf{R}_a \end{bmatrix} \quad (20)$$

where  ${}^b\mathbf{R}_a$  is the  $3 \times 3$  rotation matrix from the  $\mathcal{R}_a$  to  $\mathcal{R}_b$  frames,  ${}^b\mathbf{t}_a$  is the  $3 \times 1$  associated translation, and " $[\cdot]_{\wedge}$ " is the skew symmetric matrix associated to the cross-product.

#### IV. MATERIALS

##### A. Experimental Setup

To evaluate the performances of the proposed controller, an experimental setup was implemented (Fig. 2). It includes an OCT imaging system (a *Telesio-II 1300 nm*) from Thor-Labs<sup>1</sup>, and a microrobotic platform with 3 DOF (2 translation stages and 1 rotation stage) (TABLE I summarizes the motors features) in which the sample holder is fixed. The OCT imaging system can provide 1D depth (A-scan), 2D cross-sectional (B-scan) or 3D volumetric (3D scan) images (Fig. 3) with micrometric resolution ( $5.5 \mu\text{m}$  and  $7 \mu\text{m}$  for axial and lateral resolution, respectively) and millimetres of imaging depth (3.5 mm of depth). The Telesio-II allows a maximum field of view of  $10 \times 10 \times 3.5 \text{ mm}^3$  with a maximum A-Scan line rate of 76 kHz. The sample can also be viewed by a conventional CCD camera ( $640 \times 480$  pixels of resolution) placed on the top of the sample holder and it is rigidly connected to the OCT probe. The A-scan lines and the CCD camera are registered with each other. The OCT images acquisition and the controller are running on a 3.5GHz Xeon Intel CPU with a Windows distribution.

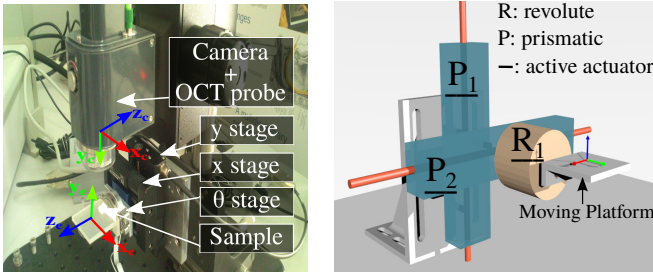


Fig. 2: Global view on the 3DOF experimental platform.

TABLE I: Features of the different microrobotic stages.

stage	product reference	features
xy	M-111-DG PI Mercury	stroke: 15 mm backlash: $2 \mu\text{m}$ min. inc. motion : $0.05 \mu\text{m}$
$\theta$	SR3610S	stroke: $2\pi$ resolution: $< 0.17 \mu\text{rad}$

Figure 3 shows some examples of OCT images acquired using a biological sample (grapes). Fig. 3(b) represents the 3D scan image acquired by scanning, line by line inside a red square defined in the white light camera image Fig. 3(a). While Fig. 3(c) shows a B-scan image along the  $x$  axis and Fig. 3(d) shows a B-scan image along the  $z$  axis.

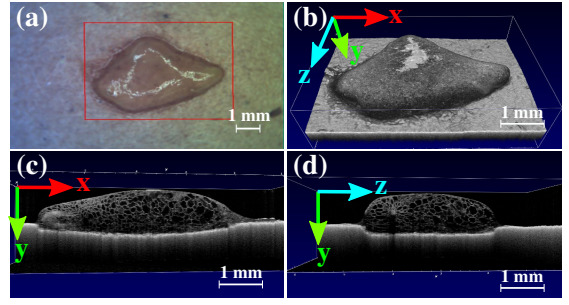


Fig. 3: (a) white light image with the defined boundary scan, (b) a 3D optical biopsy (OCT volume), (c)  $xy$  B-scan OCT image and (d)  $zy$  B-scan OCT image.

The experimental validation scenario consists of performing automatic, repetitive and accurate optical biopsies using the proposed method. More precisely, using an OCT image as a reference (respectively position), the visual controller must retrieve accurately the same image (respectively, the same position) during other examinations few days later. The aim of this is to observe and to supervise quantitatively and qualitatively the pathology evolution on the tissue.

##### B. Software Implementation

The experiments were carried under MATLAB/SIMULINK thanks to the graphical library developed in our lab: **cvLink**<sup>2</sup>. MATLAB usage gives a lot of advantages for fast prototyping and numerical data treatment. In the present case, software efficiency is not an issue because the image acquisition frequency does not exceed a few frames per second.

The developed library allows to control robotic stages and haptic interfaces using dedicated SIMULINK blocks [14]. For this work, we developed two new blocks: to grab B-scans from the Telesio-II system and to encapsulate the visual servoing computation (Fig. 4). The processing of the wavelets and the control signals are done in C++ using the open-source ViSP library within the SIMULINK blocks.

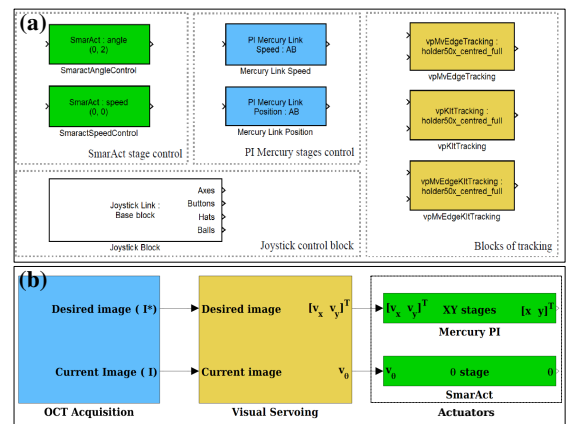


Fig. 4: Overview of the blocks of MATLAB/SIMULINK for (a) OCT acquisition and (b) robotic stages control.

<sup>1</sup><https://www.thorlabs.de/>

<sup>2</sup>[sourcesup.renater.fr/cvlink](https://sourcesup.renater.fr/cvlink)



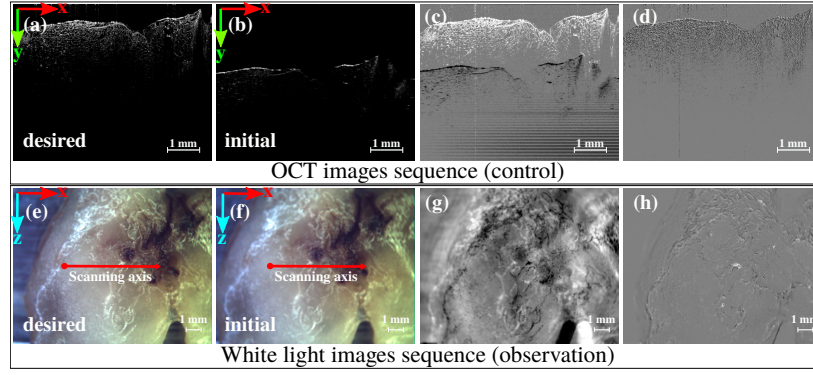


Fig. 5: (first line) OCT image snapshots acquired during the  $\mathfrak{R}(2)$  positioning task: (a) desired OCT image, (b) initial OCT image, (c) initial OCT image difference, and (d) final OCT image difference. (second line) the corresponding white light images captured by the top CCD camera.

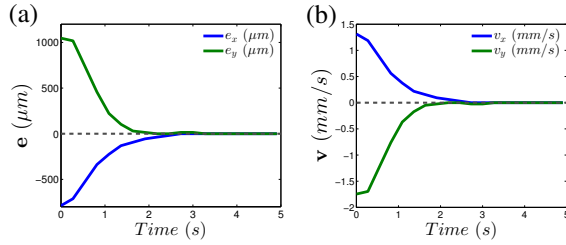


Fig. 6: (a) Cartesian error  $\Delta \mathbf{r}_i$ , and (b) velocities (in mm/s).

## V. EXPERIMENTAL RESULTS

### A. 2 DOF Positioning Task

The first test consists of performing 2 DOF positioning ( $xy$  translations) task. A first biopsy (B-scan OCT image) is performed (ultimately by a clinician) by defining a line to scan in the white light image of a biological sample. Thereby, the visual controller must reach iteratively this reference image in a closed-loop mode.

$$\mathbf{I}_{diff} = \frac{(\mathbf{I} - \mathbf{I}^*) + 255}{2} \quad (21)$$

To visualize the error, we compute and plot the image difference in addition to the control signal  $\mathbf{e}$ .

Figure 5(a)-(d) depicts an OCT image sequence representing the achievement of a planar positioning task. In this experiment, the initial pose error between the initial OCT image  $\mathbf{I}(\mathbf{r}_0)$  (chosen in arbitrary way) and the desired OCT image  $\mathbf{I}(\mathbf{r}^*)$  was measured to be  $\Delta \mathbf{r}_{init}(\mu\text{m}) = (1000, 1000)$ . Figure 5(a)-(b) show the initial and desired images grabbed in different conditions of use, respectively. The initial image difference  $\mathbf{I}_{diff}$  ((21)) is shown in Fig. 5(c) while the final one is illustrated in Fig. 5(d) demonstrating the smoothly convergence of the proposed controller. Indeed, the final positioning error, computed using the high resolution robot encoders supplied by the robot software, was measured to be  $\Delta \mathbf{r}_{final}(\mu\text{m}) = (5, 5)$  illustrating accuracy of the approach. Also, it can be highlighted that the controller is robust despite the difference between the desired and the current images.

For a better illustration of the positioning task performing, we also recorded the white light images sequence as depicted

in Fig. 5(e)-(h). As can be shown in Fig. 5(f), the biologist/clinician draws a scanning line whereby he/she seeks a possible tissue pathology by analyzing the B-scan OCT image (i.e., a micrometric resolution 2D *depth*-section in  $y$  axis). Therefore, the visual controller moves automatically the sample towards a desired scanning line analyzed some-time ago (minutes, hours, days, etc.) (Fig. 5(e)).

The positioning error decay in each stage is plotted in Fig. 6(a) while Fig. 6(b) shows the evolution of the spatial velocity over the time. As can be underlined, the controller converges exponentially to the desired position after almost 3 seconds for almost all initial positions.

### B. 3 DOF Positioning Task

The previous experiment was reproduced by adding the rotation stage  $\theta$  in the positioning task. So, the validation scenario remains the same. Figure 7(a)-(d) represents some OCT image snapshots showing the  $SE(2)$  positioning task with an initial error estimated to be  $\Delta \mathbf{r}_{init}(\mu\text{m}, \text{deg}) = (1000, 1000, 3.20)$ . Figure 7(a)-(b) depict the initial and desired OCT images, respectively when Fig. 7(c) illustrates the initial image difference between  $\mathbf{I}(\mathbf{r}_0)$  and  $\mathbf{I}(\mathbf{r}^*)$  and Fig. 7(d) the final one. By analyzing the latter, it can be underlined that the controller converges to the desired position after thirty seconds. The measured final error was  $\Delta \mathbf{r}_{final}(\mu\text{m}, \text{deg}) = (81, 10, 0.10)$  that proves the efficiency of the proposed method.

Again for this example, the white light images were recorded and shown in Fig. 7(e)-(f). The latter depicts the positioning task achievement viewed by the upper CCD camera. As can highlighted the controller converges with accuracy demonstrated by the final image difference Fig. 7(h). The Cartesian error as well as the velocities versus time (seconds) were illustrated in Fig. 8. However, small adding time consumption appears in the convergence caused by the rotation extraction process.

### C. Repeatability Study

The positioning task was repeated several times with different initial positions (TABLE II) in order to judge the

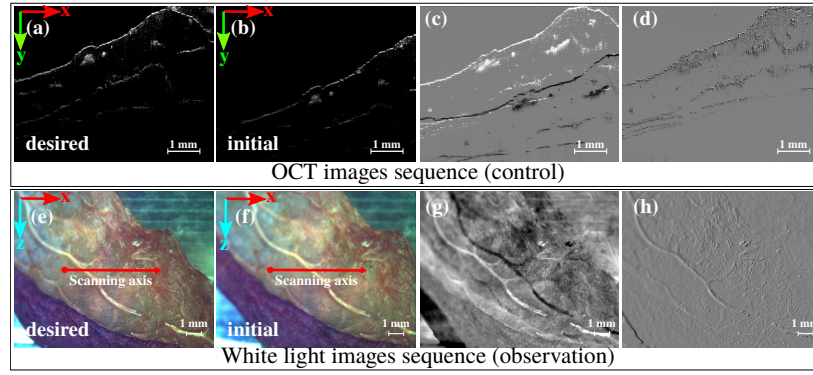


Fig. 7: (first line) OCT image snapshots recorded during the  $SE(2)$  positioning task, and (second line) the corresponding white light images captured by the top CCD camera.

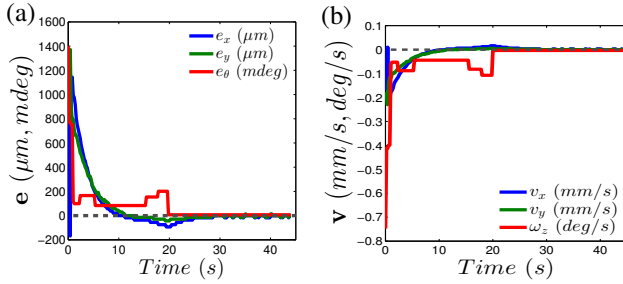


Fig. 8: (a) Cartesian errors  $\Delta \mathbf{r}_i$  (in mm and deg), and (b) velocities (in mm/s and deg/s) versus time.

repeatability of the proposed visual controller. Thereby, it was experimentally assessed that in each test the controller reaches successfully the desired position. The average final positioning error is  $\text{mean}(\mathbf{e})(\mu\text{m}, \text{deg}) = (26.3, 7.50, 0.133)$  and  $\text{STD}(\mu\text{m}, \text{deg}) = (0.036, 0.0029, 0.058)$ .

TABLE II: Repeatability study.

$N^\circ$	error	$x (\mu\text{m})$	$y (\mu\text{m})$	$\theta (\text{deg})$	$T (\text{seconds})$
1	$\Delta \mathbf{r}_{\text{initial}}$	1000	1000	0	10
	$\Delta \mathbf{r}_{\text{final}}$	10	5	0	
2	$\Delta \mathbf{r}_{\text{initial}}$	500	500	1.5	15
	$\Delta \mathbf{r}_{\text{final}}$	5	5	0.1	
3	$\Delta \mathbf{r}_{\text{initial}}$	1000	1000	2	24
	$\Delta \mathbf{r}_{\text{final}}$	5	5	0.1	
4	$\Delta \mathbf{r}_{\text{initial}}$	1000	1000	3.2	44
	$\Delta \mathbf{r}_{\text{final}}$	80	10	0.2	

## VI. CONCLUSION

In this paper, an OCT-based visual servoing for automatic and repetitive optical biopsies was presented. The developed visual controller operates by estimating the relative motion between two images using their frequency spectral information. This is because the OCT images are affected from the high image noise and weak texture. Therefore, tracking of visual features in OCT images is always a challenging task. Consequently, our approach overcomes completely the tracking step using the global image information i.e., the spectral wavelet information in the design of the controller.

The visual controller was validated using an experimental set-up containing an OCT imaging system and 3 DOF microrobotic system. The obtained results have demonstrated the efficiency of the proposed approach in terms of convergence, sufficient accuracy for the clinicians requirements, and repeatability. Future work will focus on going closer to the clinical objective (repetitive optical biopsies) by assessing the robustness of the proposed method with respect to structural evolution of the tissue.

## REFERENCES

- [1] C. Macaulay, P. Lane, et al., "In vivo pathology: microendoscopy as a new endoscopic imaging modality," *Gastrointestinal Endoscopy Clinics of North America*, vol. 14, no. 3, pp. 595–620, 2004.
- [2] M. Bruno, "Magnification endoscopy, high resolution endoscopy, and chromoscopy; towards a better optical diagnosis," *Gut*, vol. 52, no. suppl 4, pp. iv7–iv11, 2003.
- [3] T. D. Wang and J. Van Dam, "Optical biopsy: a new frontier in endoscopic detection and diagnosis," *Clinical Gastroenterology and Hepatology*, vol. 2, no. 9, pp. 744–753, 2004.
- [4] W. Jung and S. A. Boppart, "Modern trends in imaging v: Optical coherence tomography for rapid tissue screening and directed histological sectioning," *Analytical Cellular Pathology*, vol. 35, no. 3, pp. 129–143, 2012.
- [5] J. G. Fujimoto, C. Pitris, et al., "Optical coherence tomography: an emerging technology for biomedical imaging and optical biopsy," *Neoplasia*, vol. 2, no. 1, pp. 9–25, 2000.
- [6] F. Chaumette, et al., "Visual servo control, part I: Basic approaches," *IEEE Rob. and Auto. Mag.*, vol. 13, no. 4, pp. 82–90, 2006.
- [7] P. M. Novotny, J. A. Stoll, P. E. Dupont, and R. D. Howe, "Real-time visual servoing of a robot using three-dimensional ultrasound," in *IEEE Int. Conf. on Rob. and Auto.*, pp. 2655–2660, 2007.
- [8] R. Mebarki, A. Krupa, and F. Chaumette, "2-D ultrasound probe complete guidance by visual servoing using image moments," *IEEE Trans. on Rob.*, vol. 26, no. 2, pp. 296–306, 2010.
- [9] A. Krupa, D. Folio, C. Novales, P. Vieyres, and T. Li, "Robotized tele-echography: an assisting visibility tool to support expert diagnostic," *IEEE Sys. J.*, 2015.
- [10] J. Izatt and M. Choma, "Theory of optical coherence tomography," in *Optical coherence tomography*. Springer, pp. 47–72, 2008.
- [11] I. Daubechies et al., *Ten lectures on wavelets*. SIAM, vol. 61, 1992.
- [12] S. Mallat, *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.
- [13] B. S. Reddy and B. N. Chatterji, "An fft-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. on image processing*, vol. 5, no. 8, pp. 1266–1271, 1996.
- [14] A. Kudryavtsev, G. J. Laurent, et al., "Analysis of cad model-based visual tracking for microassembly using a new block set for matlab/simulink," *Int. J. of Optomechatronics*, pp. 1–7, 2015.