

PCA-based Visual Servoing using Optical Coherence Tomography

B. Dahroug¹, B. Tamadazte^{1,2}, and N. Andreff¹

Abstract—This article deals with the development of a vision-based control law to achieve high-accuracy automatic six degrees of freedom (DoF) positioning tasks. The objective of this work is to be able to replace a biological sample under an optical device for a non-invasive depth examination at any given time (i.e., performing repetitive and accurate optical characterizations of the sample). The optical examination, also called optical biopsy, is performed thanks to an optical coherence tomography (OCT) system. The OCT device is used to perform a 3-dimensional optical biopsy, and as a sensor to control the robot motion during the repositioning process. The proposed visual servoing controller uses the 3D pose of the studied biological sample estimated directly from the C-scan OCT images using a Principal Component Analysis (PCA) framework.

The proposed materials and methods were experimentally validated using a spectral-domain OCT and a 6-DoF robotic platform. The obtained results have demonstrated the pertinence of such methods which offer a positioning accuracy around 0.052 ± 0.03 mm (mean error \pm standard deviation) for linear errors and $0.41 \pm 0.16^\circ$ for angular ones over a $8 \times 9 \times 3.5$ mm³ workspace.

I. INTRODUCTION

A. Motivations

A large number of diseases are diagnosed only by a medical procedure commonly referred to as "biopsy" [1] which is eventually the case for several types of cancers. This medical procedure is also known as the *physical biopsy* which requires the extraction of sample cells or tissues for *ex-vivo* optical and/or chemical examination to prove the presence or extent of disease (Fig. 1(a)).

Furthermore, when the suspected tissue is collected, a second step is necessary. It concerns the *ex-vivo* analysis, usually performed under an optical microscope, in some cases also analyzed chemically, by a pathologist. However, this procedure, even if it remains a landmark in diagnosis and medicine, raises several questions. As reported in [2], the added clinical value decreases in some cases, because of the procedure invasiveness and the different cases of contraindications (e.g., presence of infectious risk factors, risk of haemorrhaging, risk of malignant degeneration, risk of dispersion of cancer cells during the sampling process, etc.). More recently, other complementary or alternative methods have emerged, such as the *optical biopsy* technique

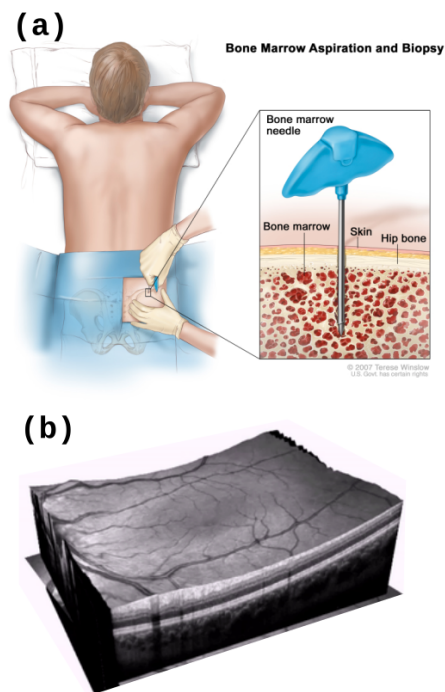


Fig. 1: Physical biopsy versus optical biopsy: (a) an example of a physical biopsy procedure using a hollow needle and (b) an optical biopsy for a retina acquired by OCT system.

(Fig. 1(b)). This procedure refers to the examination of the suspected tissue under an optical microscope at a cellular resolution [3]. Numerous optical biopsy techniques have been developed over the last two decades, such as fluorescence confocal microscopy [4], [5], multi-photon microscopy [6], optical coherence tomography [7], [8], and others at the research stage.

This article focuses specifically on the OCT modality, notably through the automatic control of the acquisition of 3D optical biopsies. In particular, the ability to accurately reposition an OCT probe (respectively, the observed sample) at given pose (i.e., position and orientation), where the physician can assess the sample evolution over time, and then carry out comparative studies of the tissue. To tackle this problem, we investigated the development of a visual feedback control law, also called visual servoing [9]. As can be seen in (Fig. 1(b)), OCT images are characterized by an unfavorable signal-to-noise ratio (similar to the ultrasound images), low texture, no geometric visual features that can be easily detected, extracted and tracked over time. As a result, using geometric visual features [10] as usually reported in

¹FEMTO-ST Institute, AS2M department, Univ. Bourgogne Franche-Comté/CNRS/ENSMM, 24 Rue Alain Savary, 25000 Besançon, France. firstname.lastname@femto-st.fr.

² Institute for Intelligent Systems and Robotics, University of Sorbonne, CNRS, UMR 7222, 4 place Jussieu, 75005 Paris, France. This work was supported by French ANR NEMRO Project no ANR-14-CE17-0013-001 and the INSERM ROBOT Project: ITMO Cancer no 17CP068-00.

the literature is highly challenging.

B. Background

OCT technology [11] has been considered in numerous medical and surgical applications. In [12]–[14], the authors discussed the integration of an OCT device into the surgical workflow for providing to the physician a real-time and high-resolution visual feedback of the operative site (observed tissues). A miniaturized OCT in the form of millimeter size probe [15] was also considered for such surgical task. OCT usage has also shown significant added clinical value, namely in laser surgery [16]. The cross-section OCT images (also called *B-scan*) were used for both detecting the position of the laser in real-time and evaluating the quality of ablation/resection of pathological tissues [17]. In the same manner, the B-scan OCT images were employed for guiding a robotic platform to carry out a laser ablation of pathological tissue [18], and to track a surgical instrument during an intervention [19]. These preliminary works are original and engaging despite a lack of autonomy of the proposed approaches.

More recently, it was suggested to perform highly accurate and automatic 3-DoF positioning task which consists of the replacing a biological sample under the OCT probe for further examination. The proposed approach was based on using a featureless visual servoing control law [20]. The controller was derived using wavelets representation of the B-scan OCT images. The contribution of this work is that using wavelets representation allowed overcoming the unfavorable signal-to-noise ratio, which generally characterizes the OCT images [21]. Therefore, the authors in [22], [23] used the wavelets coefficients as visual inputs in the vision-based control law for regulating the sample pose error within the in-plane robotic stages. Similarly, 6-DoF positioning task was performed by merging both OCT data (for the in-plane DoF) and conventional camera (for the out-plane DoF) in the same visual servoing control scheme [24]. Note that the clinical objectives of these investigations concern the achievement of repetitive and accurate optical biopsies in the aim to monitor the variation of the suspected tissue over time.

C. Contributions

In this article, it is proposed a new framework for repetitive optical biopsy acquisition that go beyond the state-of-the-art related to use the OCT volume image (also called *C-scan*) in a visual servoing control loop for managing the out-plane motion. Our work also investigated an efficient method to control the entire 6-DoF based only using OCT data without any additional external sensor. The most challenging task of this work is to compute the biological sample pose in the OCT frame (respectively, the robot frame) and *vice versa* without advanced image processing or 3D reconstruction algorithms. To tackle this, we used a Principal Component Analysis (PCA) which is an orthogonal transformation framework [25]–[27] to recover, as accurate as possible, the 3D pose of an object. PCA has demonstrated its relevance and effectiveness in a wide range of engineering sciences

such as image processing, data registration and fusion, supervised learning, and clustering. Earlier this year, authors [28] proposed a new formulation of 2D direct visual servoing based on PCA decomposition of the image. In this work, conventional images were projected on a new orthogonal basis in aim to compactly express the visual information, i.e., using the image coefficients obtained by the PCA approach.

Our development goes further by applying the PCA method on C-scan OCT data. The proposed PCA-based 3D direct visual servoing is validated experimentally using a spectral-domain OCT system which is mounted in an eye-to-hand configuration with respect to the robotic platform. The validation scenario consists of replacing the sample at its desired pose at which an initial 3D optical biopsy was acquired. The obtained results in terms of accuracy and controller behaviour are very satisfactory since the linear error was evaluated as 0.052 ± 0.03 mm which is close to the human cell which its size vary from 0.1 mm to 0.03 mm.

In the remainder of this article, Section II presents the proposed methodology to express analytically the 3D direct visual servoing controller. Section III discusses the experimental robotic setup as well as the proposed validation scenarios to assess the relevance of the developed methods. Also, the obtained results were discussed and analyzed.

II. METHODOLOGY

As mentioned above, one of the challenging task in this work is the 3D pose estimation of the sample viewed by the OCT device (i.e., expressed in the OCT frame \mathfrak{O}). To tackle this, we investigated the use of OCT C-scan data to express the sample pose in the OCT frame (respectively, robot frame \mathfrak{R}) without using advanced and complex image processing algorithms or any other 3D reconstruction of the sample. Instead of formulating the object 3D pose directly using the C-scan image which is high dimensional data (e.g., $1024 \times 860 \times 640$ voxels), we opted for a PCA-based representation for the OCT data to reduce both the complexity and the estimation process.

A. Notations and Symbols

In this article, the used notations and symbols are summarized in TABLE I.

Symbol	Description
\mathbf{D}	the data set describes the OCT C-scan as a point cloud
d_{ij}	the i^{th} point in the data set \mathbf{D} and j represents the point information such as its position or its intensity
$\bar{\mathbf{d}}$	the centroid of the point cloud
\mathbf{B}	the covariance matrix of data set \mathbf{D}
b_i^2	the variance of the i^{th} variable
b_{ij}	the covariance of the i^{th} and j^{th} variable
\mathbf{A}	a diagonal matrix with the eigenvalues of \mathbf{B}
\mathbf{V}	the eigenvector matrix of \mathbf{B}
\mathbf{g}_i	the i^{th} eigenvector of \mathbf{V}
\mathbf{v}_i	the normalized vector of \mathbf{g}_i
$\mathbf{I}_{n \times n}$	a $n \times n$ identity matrix
${}^c\mathbf{M}_o$	the transformation matrix which describes the 3D pose of the object frame \mathfrak{O} with respect the OCT frame \mathfrak{O}

${}^c\mathbf{R}_o$	the rotation matrix which describes the object frame \mathcal{O} with respect the OCT frame \mathcal{C}
${}^c\bar{\mathbf{t}}_o$	the translation vector which describes the object frame \mathcal{O} with respect the OCT frame \mathcal{C}
$\theta\mathbf{u}$	the rotation represented in the angle/axis parameterization form
\mathbf{s}, \mathbf{s}^*	the current and the reference features vectors, respectively
$\mathbf{e}, \dot{\mathbf{e}}$	visual servoing vector error and its time derivative
\mathbf{v}_c	saptial velocity vector which gathers the instantaneous linear \mathbf{v}_c and angular ω_c velocities of the camera/robot interaction matrix which relates the $\dot{\mathbf{e}}$ to \mathbf{v}_c
\mathbf{L}_{3D}	pseudo-inverse of the interaction matrix
\mathbf{L}_{3D}^+	a 1×3 row vector of zeros
$\mathbf{0}_{1 \times 3}$	a 3×3 zeros matrix
$\mathbf{0}_{3 \times 3}$	the skew-symmetric matrix associated to the vector \mathbf{u}
$[\mathbf{u}]_{\times}$	the cardinal sinus
λ	a positive gain coefficient of the visual servoing controller
λ_t, λ_r	positive gain coefficients for regulating the translation and rotation errors, respectively

TABLE I: Notations summary.

B. PCA-based 3D Pose Estimation

The PCA is a powerful linear algebra method, namely used for simplifying the complexity of large-dimensional data thanks to a more compact representation. Indeed, this method can be considered mathematically as an orthogonal linear transformation that maps a data set to a new coordinate system, called *eigenspace*. This new space has its bases, called *eigenvectors* or *principal components* which are obtained by projecting the data on them.

Let us consider a data set $\mathbf{D} \in \mathbb{R}^{n \times m}$ which describes the OCT C-scan as a point cloud. The matrix \mathbf{D} collects the n points, where the m -dimension of each point could gather the 3D position (x, y, z) expressed in the C-scan frame \mathcal{C} with a voxel intensity (e.g., grayscale).

An example of a geometric interpretation of PCA applied to a sample of 3D points within a spheroid shape is shown in Fig. 2. The centroid $\bar{\mathbf{d}}$ of such spheroid cloud is obtained as follows:

$$\bar{d}_j = \frac{1}{n} \sum_{i=1}^n d_{ij}, \quad j = 1, 2, 3 \quad (1)$$

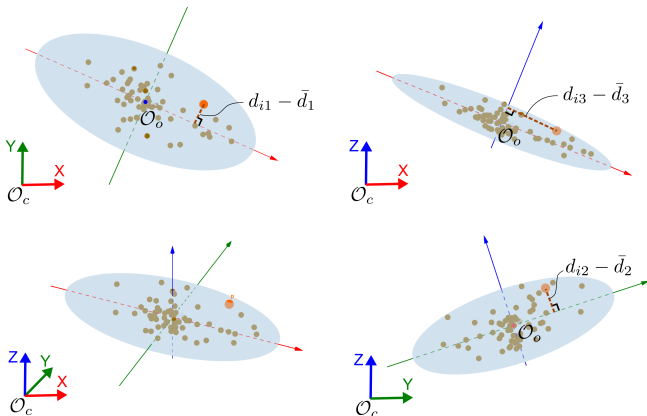


Fig. 2: Geometric representation of PCA.

where $\bar{\mathbf{d}} = [\bar{d}_1, \bar{d}_2, \bar{d}_3]^\top$ is the point cloud centroid in xyz -coordinates, $\mathbf{p}_i = [d_{i1}, d_{i2}, d_{i3}]^\top$ is the 3D position of the i^{th} point in xyz -axes respectively, and n is the length of data set.

The covariance matrix, also called the correlation matrix, provides relevant information about the point cloud rotation. It is consequently obtained by

$$\mathbf{B}_{m \times m} = \begin{bmatrix} b_{11}^2 & b_{12} & \cdots & b_{1m} \\ b_{12} & b_{22}^2 & \cdots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{1m} & b_{2m} & \cdots & b_{mm}^2 \end{bmatrix} \quad (2)$$

whereby b_i^2 is the variance of the i^{th} variable, and b_{ij} is the covariance between the i^{th} and j^{th} variables.

Furthermore, the variance is a measure of how the data is spread out, and it is formulated as

$$b_i^2 = \frac{\sum_{k=1}^n (d_{ik} - \bar{d}_k)^2}{n-1}, \quad \text{where } 1 \leq k \leq m \quad (3)$$

while the covariance is a measure of how much the dimensions vary from the mean with respect to each other and it is computed by

$$b_{ij} = \frac{\sum_{k=1}^n (d_{ik} - \bar{d}_k)(d_{jk} - \bar{d}_k)}{n-1}, \quad \text{where } j \neq k \text{ and } j \neq i. \quad (4)$$

Since the data cloud has three dimensions, the variable m varies as x , y and z . Consequently, the covariance matrix is deduced as

$$\mathbf{B}_{3 \times 3} = \begin{bmatrix} cov(\mathbf{d}_x, \mathbf{d}_x) & cov(\mathbf{d}_x, \mathbf{d}_y) & cov(\mathbf{d}_x, \mathbf{d}_z) \\ cov(\mathbf{d}_y, \mathbf{d}_x) & cov(\mathbf{d}_y, \mathbf{d}_y) & cov(\mathbf{d}_y, \mathbf{d}_z) \\ cov(\mathbf{d}_z, \mathbf{d}_x) & cov(\mathbf{d}_z, \mathbf{d}_y) & cov(\mathbf{d}_z, \mathbf{d}_z) \end{bmatrix}. \quad (5)$$

The vector $\mathbf{d}_x = [d_{11}, d_{21}, \dots, d_{n1}]$ is the column vector in the data set \mathbf{D} which represents the x -position of all points. Similarly, the column vectors \mathbf{d}_y and \mathbf{d}_z are the yz -position, respectively.

For instance, the covariance of the element (1,2) from the previous matrix equals to

$$b_{12} = cov(\mathbf{d}_x, \mathbf{d}_y) = \frac{1}{n-1} \sum_{i=1}^n (d_{i1} - \bar{d}_1)(d_{i2} - \bar{d}_2). \quad (6)$$

The PCA method decomposes the covariance matrix to

$$\mathbf{V}^\top \mathbf{B} \mathbf{V} = \mathbf{A}. \quad (7)$$

The matrix \mathbf{A} is a diagonal matrix which gathers the eigenvalues of \mathbf{B} . The elements of \mathbf{A} (i.e., a_1, a_2, \dots, a_m) are deduced by the characteristic equation $\|\mathbf{A} - a_i \mathbf{I}\| = 0$, where \mathbf{I} is a $m \times m$ identity matrix.

The columns of matrix \mathbf{V} are the eigenvectors of \mathbf{B} . These characteristic vectors can be calculated by solving the equation $\|\mathbf{B} - a_i \mathbf{I}\| \mathbf{g}_i = 0$, where $1 \leq i \leq m$. By normalizing the vector \mathbf{g}_i as $\mathbf{v}_i = \frac{\mathbf{g}_i}{\sqrt{\mathbf{g}_i^\top \mathbf{g}_i}}$, the matrix \mathbf{V} is

formed by these unit vectors. The fact that these unit vectors are orthogonal to each other, the matrix \mathbf{V} is considered as the rotation matrix of the point cloud. Therefore, the instantaneous transformation that maps the object pose into the OCT frame \mathfrak{Q} is defined by

$${}^c\mathbf{M}_o = \begin{bmatrix} \mathbf{V} & \bar{\mathbf{d}} \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} = \begin{bmatrix} {}^c\mathbf{R}_o & {}^c\bar{\mathbf{t}}_o \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix}. \quad (8)$$

where ${}^c\mathbf{R}_o$ is a 3×3 matrix that describes the object rotation with respect to the OCT frame \mathfrak{Q} , while ${}^c\bar{\mathbf{t}}_o$ is a 3×1 column vector that describes the object translation with respect to the OCT frame \mathfrak{Q} , and $\mathbf{0}_{1 \times 3}$ is a row vector of zeros.

C. 3D Direct Visual Servoing Control Law

After deducing the instantaneous pose of the observed object from the OCT C-scan, a classical visual servoing controller [9] is applied. The feature vector $\mathbf{s} = (\bar{\mathbf{t}}, \theta \mathbf{u})$ is defined as the pose vector which gathers the translation $\bar{\mathbf{t}}$ of the point cloud and its rotation $\theta \mathbf{u}$ in form of angle/axis parameterization. Therefore, the error to be regulated to zero is deduced as the difference between the current features vector and the reference one, i.e., $\mathbf{e} = \mathbf{s} - \mathbf{s}^*$.

The time variation of the error is related to the spatial velocity of the camera/robot $\underline{\mathbf{v}}_c$ by the interaction matrix $\mathbf{L}_{3D} \in \mathbb{R}^{6 \times 6}$ as

$$\dot{\mathbf{e}} = \mathbf{L}_{3D} \underline{\mathbf{v}}_c \quad (9)$$

in which $\underline{\mathbf{v}}_c = (\mathbf{v}_c, \omega_c)$ gathers the instantaneous linear and angular velocities of the camera/robot.

In case that the reference feature vector equals to $\mathbf{s}^* = (*\bar{\mathbf{t}}, \mathbf{0})$, then the interaction matrix related to \mathbf{e} is determined by

$$\mathbf{L}_{3D} = \begin{bmatrix} -\mathbf{I}_{3 \times 3} & [\bar{\mathbf{t}}]_{\times} \\ \mathbf{0}_{3 \times 3} & \mathbf{L}_{\theta \mathbf{u}} \end{bmatrix} \quad (10)$$

where $\mathbf{I}_{3 \times 3}$ is a 3×3 identity matrix, $[\bullet]_{\times}$ is the skew-symmetric matrix associated to a vector, and $\mathbf{L}_{\theta \mathbf{u}}$ is given by [29]

$$\mathbf{L}_{\theta \mathbf{u}} = \mathbf{I}_{3 \times 3} - \frac{\theta}{2} [\mathbf{u}]_{\times} + \left(1 - \frac{\text{sinc} \theta}{\text{sinc}^2 \frac{\theta}{2}} \right) [\mathbf{u}]_{\times}^2 \quad (11)$$

where $\text{sinc} x$ is the sinus cardinal.

At this stage, the spatial velocity $\underline{\mathbf{v}}_c$ can be deduced for ensuring an exponential decoupled reduction of the error (i.e., $\dot{\mathbf{e}} = -\lambda \mathbf{e}$) as

$$\underline{\mathbf{v}}_c = -\lambda \mathbf{L}_{3D}^{-1} \mathbf{e} \quad (12)$$

where λ is a gain coefficient, since \mathbf{L}_{3D} is square and has a closed-form inverse [29].

III. EXPERIMENTAL VALIDATION

The proposed methods and materials were validated using an OCT device mounted in an eye-to-hand configuration to a parallel robotic platform on which are placed the samples to be analysed (Fig. 3).

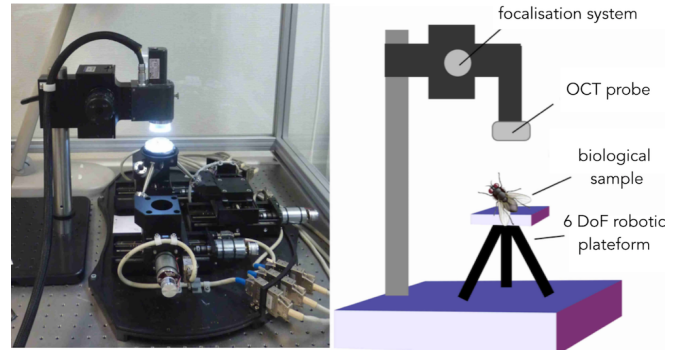


Fig. 3: Illustration of the experimental framework used to validate the proposed methods.

A. Optical Coherence Tomography

A Telesto-II 1325nm spectral domain from ThorLabs (www.thorlabs.com) has been adopted in this study. It provides 1D depth (A-scan), 2D cross-sectional (B-scan) or 3D volumetric (C-scan) images with micrometer resolution ($5.5 \mu\text{m}$ and $7 \mu\text{m}$ for axial and lateral resolutions, respectively) and millimeter depth (3.5 mm of penetration). The Telesto-II allows a maximum field-of-view of $10 \times 10 \times 3.54 \text{ mm}^3$ with the highest A-scan line rate of 76 kHz. The OCT probe is also equipped with a coaxial CCD (Charge Coupled Device) camera of 640×480 pixels of resolution. This camera is useful to adjust the focus plane by a fine-tuning along the optical axis. Consequently, the operator can regulate the OCT focus since both the OCT and the camera focus plans are the same. Besides that, the operator can define directly in the CCD image plane which part of the sample should be scanned (Fig. 4).

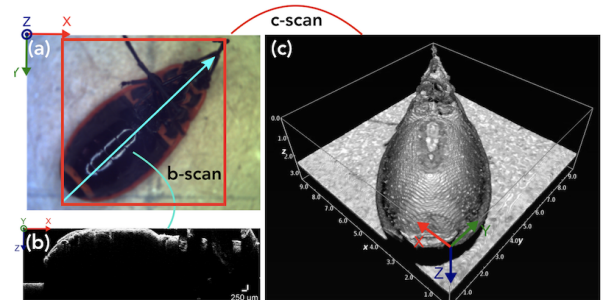


Fig. 4: Example of B-scan and C-scan OCT data. (a) A white light image of a fly, (b) cross-section (B-scan) OCT image acquired at the arrow position, and (c) the C-scan OCT (volume) acquired inside the rectangular box.

Moreover, the Thorlabs API (Application Programming Interface) provides a threshold tool which supplies the sample external surface as shown in Fig. 4(c). Afterwards, the acquired C-scan is segmented by using the PCL (Point Cloud Library) [30]. The segmentation parameters are constant when the samples studied are of the same type, for example, taken from the same organ.

B. Robotic Sample Holder

The sample (organic or inorganic) to be characterized is placed or fixed on the robotic platform. The latter consists of a 6-DoF parallel robot from Physical Instruments (www.physikinstrumente.com) (SpaceFAB SF-3000 BS). Each DoF is individually actuated by a high-resolution continuous DC motor which is equipped with a shaft encoder.

The robotic system is characterized by the following features: i) translation ranges $(t_x, t_y, t_z)_{\max} = (50 \text{ mm}, 100 \text{ mm}, 12.7 \text{ mm})$ with a linear resolution of $0.2 \mu\text{m}$ and a repeatability of $\pm 0.5 \mu\text{m}$, and ii) rotation ranges $(r_x, r_y, r_z)_{\max} = (10^\circ, 10^\circ, 10^\circ)$ with an angular resolution of 0.0005° and a repeatability of $\pm 0.0011^\circ$.

Two computers equip the experimental platform: the first (a 2.33 GHz Core Intel CPU with Windows 7 OS) is used for implementing the inner control laws of the robot, and the second one (a 3.50 GHz Xeon Intel CPU with Windows 7 OS) is dedicated to the image acquisition (OCT and camera) as well as the image processing and visual servoing algorithms. Both the computers communicate with each other thanks to an asynchronous TCP/IP protocol.

C. Validation Scenarios

The proposed controller has been experimentally validated using different biological samples (e.g., an ant and a fly) and solid 3D objects (Fig. 5). The operator places the

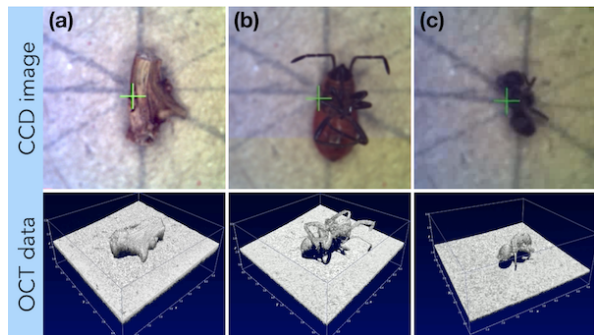


Fig. 5: Example of used samples: (first row) the CCD camera images and (second row) the corresponding OCT C-scan images.

sample on the robotic platform and performs a first optical characterization (i.e., 3D optical biopsy acquisition and examination) of the sample. This first acquired 3D optical biopsy is considered as the desired/reference pose for further examinations. Consequently, in case that the operator needs to perform a new characterization of the sample to observe its change, the sample is then placed at any location on the platform that is visible by the OCT (within the OCT field-of-view). To carry out this comparison, the sample must be placed in the same spatial pose as the first examination as accurate as possible i.e., at almost at the cellular resolution. This means that the controller should converges with an accuracy of about $50 \mu\text{m}$. To reposition manually the sample at the reference position is very challenging if not impossible. However, the proposed 3D direct visual servoing controller

can accurately achieve this due to the high-resolution OCT data used for controlling the robot motion to the convergence toward the reference pose. In the following, we will discuss the results obtained experimentally on different samples and positioning tasks achievement.

Finally, for better behavior of the proposed controller, the gain coefficient λ in (12) is opted to have different values between the linear and angular robotic stages. Thereby, λ is defined as follows

$$\lambda = \begin{bmatrix} \lambda_t & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_t & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_t & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_r & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_r & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_r \end{bmatrix}. \quad (13)$$

D. Experimental Results

In these trials, the OCT field-of-view was fixed to $8 \times 9 \times 3.5 \text{ mm}^3$ for ensuring a "reasonable" OCT frame-rate (the larger the field-of-view, the shorter the frame-rate). For instance, in this work, we tuned the spatial resolution of our OCT device to have a voxel spacing of $20 \times 20 \times 3.46 \mu\text{m}^3$ which generates a volume of size $400 \times 450 \times 1024 \text{ voxels}$ every two seconds.

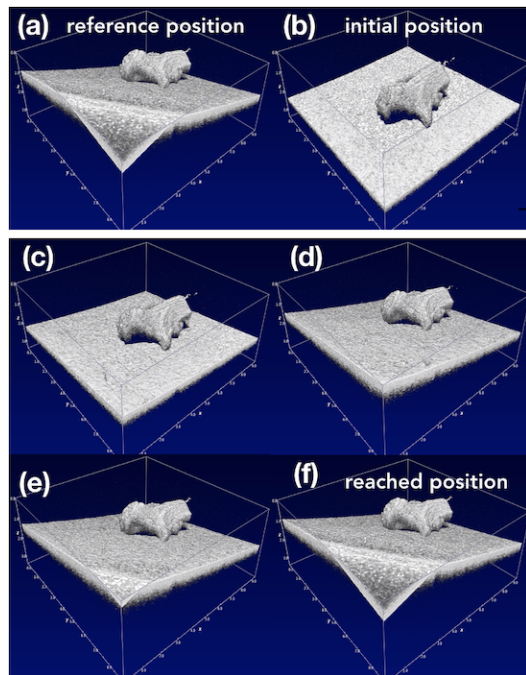


Fig. 6: Image sequence showing the 3D positioning task achievement. Snapshot (a) shows the desired pose, (b) that of the initial pose, (c)-(e) represent intermediate pose acquired during the positioning task and (f), the reached pose.

1) *Validation trial 1:* During this trial, a solid organic sample (a tiny part of a grapes branch, Fig. 5(a)) used to evaluate the performance of the proposed visual servoing control law. As mentioned, the desired pose s^* is initially defined during a first examination. The sample is

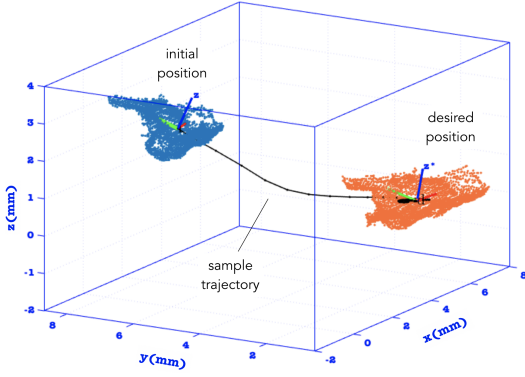


Fig. 7: Plot of the 3D trajectory performed by the sample (respectively, the robot) during the pose regulation task.

afterwards placed at an arbitrary pose, considered as the initial one (s) to begin the control loop. For this trial (test 3 in TABLE II), the initial error is estimated as $e_i = (2.6 \text{ mm}, -5.6 \text{ mm}, -1.8 \text{ mm}, -10.6^\circ, -9.6^\circ, -19.7^\circ)$.

Figure 6 depicts an image sequence grabbed during the positioning task achievement. One can highlight that the control law allows converging accurately towards the desired pose. Indeed, the desired position (Fig. 6(a)) is almost similar to the reached one (Fig. 6(f)). For this trial, the average linear error is estimated to $mean(t_x, t_y, t_z) = 0.075 \text{ mm}$ while the average angular one to $mean(r_x, r_y, r_z) = 0.344^\circ$.

Furthermore, the sample trajectory was recorded and plotted in Fig. 7. It is known that a 3D visual servoing scheme allows obtaining a straight-line trajectory of the sample's gravity center. This is almost the case for the trajectory shown in Fig. 7.

The regulation to zero of the error e versus the number of iterations for each robotic stage is recorded and plotted separately in Fig. 8. As can be noticed, the convergence is not carried out exponentially because the 3D OCT data (C-scan images) are generally noised (i.e., speckle-like noise) which characterizes this type of imaging modality, likewise ultrasounds images. Indeed, the acquired B-scan images contain a noise level 0.0102, based on the method proposed in [32], while the classical lena image has a noise level around 0.0032. Besides that, the C-scan images contain a noise level about 26.736. Even under unfavourable working conditions, the proposed controller still accurate and robust.

Both the linear (v_x, v_y, v_z) (in mm/s) and the angular ($\omega_x, \omega_y, \omega_z$) (in degree/s) velocities expressed in the OCT frame are shown in Fig. 9(a) and (b), respectively. The behavior of the OCT velocities is similar to that of the error regulation. The reasons why they are not perfectly smooth are the same as for those of the error, i.e., the inherent noise aspect of OCT data. However, one can remark that the computed velocities are relatively smooth, a proof that the 3D pose estimation method is relatively accurate despite the use of noised OCT measures.

2) *Others validation trials:* Other tests were carried out using the biological samples (e.g., a fly and an ant, Fig. 5(b) and (c), respectively). The validation scenario remains sim-

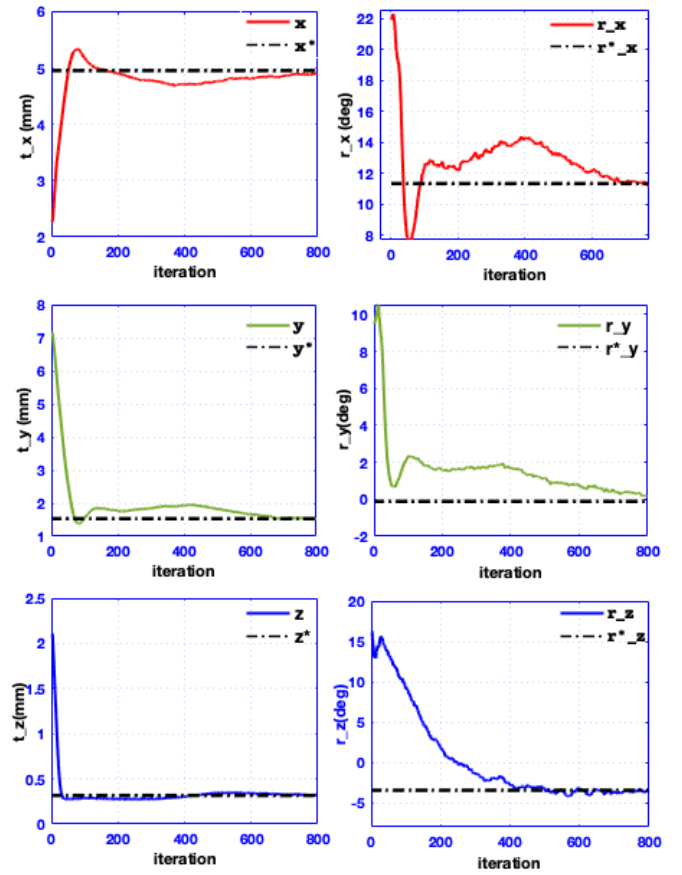


Fig. 8: Plotting of the error evolution for each DoF during the regulation task 1 (left column represents the linear stages while the right one those of the angular stages).

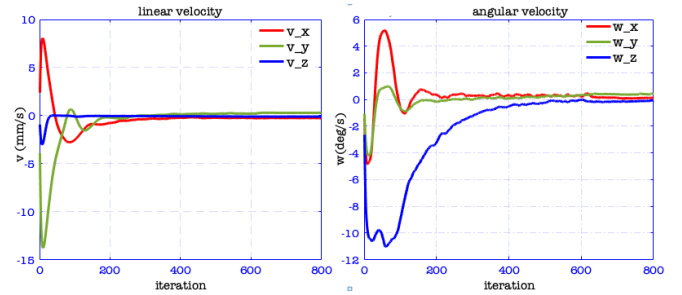


Fig. 9: Representation of the camera velocities recorded during the pose regulation task 1 (left are the linear velocities and right the angular ones).

ilar to those of trial 1. However, these samples have more details that can complicate the pose computation of the object in the OCT frame.

As can be seen in Fig. 10, the proposed control law allows converging the sample from an arbitrary initial position to the reference one. In this trial (test 9 in TABLE II), the initial error is measured as $e_i = (3.9 \text{ mm}, -4.4 \text{ mm}, 0.3 \text{ mm}, -4.8^\circ, -13.6^\circ, -19.4^\circ)$. The reached position (Fig. 10(e)) is almost similar to the desired one (Fig. 10(a)). The average linear error is $mean(t_x, t_y, t_z) = 0.027 \text{ mm}$, while the average angular

sample	test number	type of measure	translations (mm)				rotations (degree)				gain control
			t_x	t_y	t_z	$\ t\ $	θ_{u_x}	θ_{u_y}	θ_{u_z}	$\theta\ \mathbf{u}\ $	
branch	1	initial pose	2.04	7.16	1.84		18.42	3.79	1.05		$\lambda_t = 3$
		reached pose	5.61	1.76	0.44		16.51	4.84	1.41		$\lambda_r = 0.52$
		initial error	3.57	-5.37	-1.40		-1.60	1.05	0.02		
		final error	0.001	0.037	-0.001	0.013±0.02	0.305	-0.005	-0.343	0.218±0.185	
branch	2	initial pose	4.05	5.2	1.36		17.77	4.87	1.01		$\lambda_t = 3$
		reached pose	5.28	4.26	1.19		8.16	-7.64	10.14		$\lambda_r = 0.52$
		initial error	1.236	-0.948	-0.18		-9.767	-12.294	9.173		
		final error	0.003	-0.005	-0.006	0.004±0.001	0.007	0.182	-0.01	0.066±0.1	
branch	3	initial pose	2.02	6.83	1.56		23.1	9.94	10.28		$\lambda_t = 3$
		reached pose	5.28	1.51	0.34		11.57	-0.88	-8.48		$\lambda_r = 0.52$
		initial error	3.17	-5.23	-1.256		-11.318	-10.107	-18.866		
		final error	-0.09	0.09	-0.04	0.075±0.03	0.212	0.717	-0.103	0.344±0.328	
fly	4	initial pose	2.74	5.8	0.78		-11.06	2.43	17.45		$\lambda_t = 3$
		reached pose	5.17	3.58	0.68		-6.82	0.49	-2.77		$\lambda_r = 0.52$
		initial error	2.316	-2.0	-0.218		5.814	-1.57	-21.838		
		final error	-0.11	0.218	-0.116	0.149±0.06	1.577	0.372	-1.613	1.18±0.706	
fly	5	initial pose	2.8	6	0.7		-2.8	1.9	15.5		$\lambda_t = 3$
		reached pose	5.2	3.8	0.4		-3.3	-0.4	-5.1		$\lambda_r = 0.52$
		initial error	2.465	-2.07	-0.218		-0.12	-2.604	-21.434		
		final error	0.034	0.077	0.003	0.038±0.037	0.416	-0.246	-0.787	0.483±0.276	
fly	6	initial pose	2.78	5.93	0.69		-2.69	2.42	15.92		$\lambda_t = 2$
		reached pose	5.24	3.61	0.52		-3.65	-1.07	-6.66		$\lambda_r = 0.698$
		initial error	2.436	-2.071	-0.231		0.045	-3.187	-23.596		
		final error	-0.022	0.256	-0.056	0.112±0.126	1.01	0.306	-1.018	0.778±0.409	
fly	7	initial pose	2.72	5.81	0.81		-11.03	2.40	17.00		$\lambda_t = 3$
		reached pose	5.12	3.72	0.64		-6.18	-0.15	-5.97		$\lambda_r = 0.52$
		initial error	2.334	-1.98	-0.243		5.484	-2.195	-23.229		
		final error	-0.066	0.139	-0.0716	0.092±0.04	0.628	0.359	-0.26	0.416±0.19	
ant	8	initial pose	5.13	3.54	0.91		6.95	11.8	-12.55		$\lambda_t = 3$
		reached pose	5.64	3.28	1.72		5.48	-1.96	-16.79		$\lambda_r = 0.52$
		initial error	0.498	-0.25	0.816		-1.236	-13.731	-4.045		
		final error	-0.001	0	0.001	0.001±0.001	0.045	0.026	0.231	0.101±0.114	
ant	9	initial pose	2.16	7.36	1.39		16.42	13.35	-9.25		$\lambda_t = 3$
		reached pose	6.06	2.95	1.75		11.12	1.92	-28.38		$\lambda_r = 0.52$
		initial error	3.948	-4.384	0.349		-4.849	-13.626	-19.37		
		final error	0.05	0.023	-0.008	0.027±0.021	0.452	-0.199	-0.238	0.296±0.136	
degraded fly	10	initial pose	1.65	6.71	0.77		-4.25	0.67	14.15		$\lambda_t = 0.8$
		reached pose	4.27	2.46	1.09		-12.89	-8.3	-8.23		$\lambda_r = 0.2$
		initial error	2.63	-4.23	0.327		-8.88	-8.9	-22.26		
		final error	0.03	0.003	0.005	0.004±0.001	0.373	0.03	0.112	0.172±0.179	
mean error ± STD			0.0516±0.03				0.406±0.162				

TABLE II: Summary of different trials achieved with constant gain values.

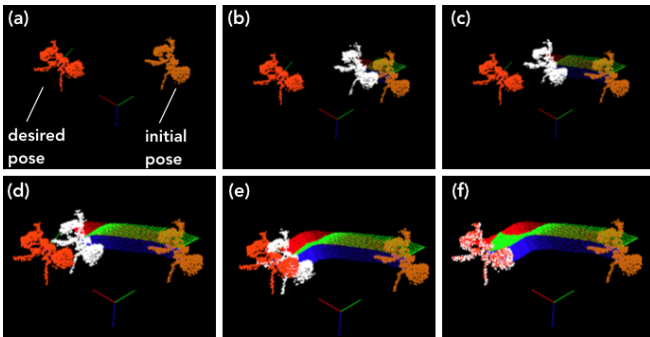


Fig. 10: Image sequence depicting the performing of a 6 DoF positioning task 2. (a) both the initial and desired poses, (b) - (e) some intermediate poses while the sample converges towards the desired one, and (f) the final pose which overlaps the desired one. Note in this example, we segmented the 3D object for a better illustration of the performed positioning task (see Multimedia Extension 1).

one to $mean(r_x, r_y, r_z) = 0.296^\circ$. As can be noticed, even the shape of studied sample is more complex, the controller is still accurate.

A last trial (test 10 in TABLE II) was conducted to assess the controller robustness to the modification of the observed sample (i.e., tissue degradation). Therefore, the scenario of the first trial is modified by comparing an old C-scan image of a fly, which is saved during a previous test, with a new fly sample which misses different part of its body (e.g., its legs and antennas). Afterwards, the controller guides the robot by reducing the pose error between the old C-scan image (the reference pose) and the new sample which is located in an arbitrary initial pose (see the attached multimedia). The controller showed that it is still accurate and robust, where the average linear error is $mean(t_x, t_y, t_z) = 0.004 \text{ mm}$ and the angular one is $mean(r_x, r_y, r_z) = 0.173^\circ$.

TABLE II gathers the different conducted tests. One can be noticed that the controller is highly accurate in translation as well as in rotation, even some shortcomings concerning the rotation around the z -axis (optical axis). Besides that, the convergence region is relatively large compared to state-of-the-art methods dealing with the use of OCT data as control signal in a visual servoing control loop, such as those reported in [23], [31]. In these methods, the convergence area is around $\pm 4 \text{ mm}$ in translation and $\pm 10^\circ$. In our work, to ensure the convergence of the proposed method, the operator

has to make sure to position the sample manually in the field-of-view of the OCT which is smaller than that of the robot used for automatic positioning.

IV. CONCLUSION

In this article, we have demonstrated that the use of a direct 3D visual servoing allows performing high accurate positioning tasks. This work was developed in the context of the optical characterization of tissues using an optical coherence tomography system. The challenge was to be able to place automatically (at almost the cellular resolution) the sample at the same pose for carrying out several time-spaced examinations and studying the evolution of tissues. To tackle this problem, we developed an original visual servoing method which uses the OCT 3D data (C-scan images) as the visual information. To compute the 3D pose of the sample in the OCT frame, we used a Principal Component Analysis framework.

The proposed materials and methods were successfully validated using an experimental robotic setup equipped with a spectral-domain OCT system. The obtained results in terms of accuracy, versatility and behavior are satisfactory to consider future investigations on a more clinical context in collaboration with an anatomical pathology laboratory.

REFERENCES

- [1] E. Sausville and D. Longo, "Principles of cancer treatment: surgery, chemotherapy, and biologic therapy," *Harrisons Principles of Internal Medicine*, vol. 16, no. 1, p. 464, 2005.
- [2] 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.
- [3] T. Meyer, M. Schmitt, O. Guntinas-Lichius, and J. Popp, "Toward an all-optical biopsy," *Opt. and Phot. News*, vol. 30, pp. 26–33, 2019.
- [4] R. K. Orosco, R. Y. Tsien, and Q. T. Nguyen, "Fluorescence imaging in surgery," *IEEE Rev. Biomed. Eng.*, vol. 6, pp. 178–187, 2013.
- [5] S. Puliatti, L. Bertoni, G. M. Pirola, *et al.*, "Ex-vivo fluorescence confocal microscopy: The first application for real-time pathologic examination of prostatic tissue," *BJU Int.*, 2019.
- [6] S. R. Kantelhardt, J. Leppert, J. W. Kantelhardt, E. Reusche, G. Hüttmann, and A. Giese, "Multi-photon excitation fluorescence microscopy of brain-tumour tissue and analysis of cell density," *Acta Neurochirurgica*, vol. 151, no. 3, p. 253, 2009.
- [7] G. J. Tearney, M. E. Brezinski, B. E. Bouma, S. A. Boppart, C. Pitris, J. F. Southern, and J. G. Fujimoto, "In vivo endoscopic optical biopsy with optical coherence tomography," *Science*, vol. 276, no. 5321, pp. 2037–2039, 1997.
- [8] A. F. Fercher, W. Drexler, C. K. Hitzenberger, and T. Lasser, "Optical coherence tomography-principles and applications," *Reports on Progress in Physics*, vol. 66, no. 2, p. 239, 2003.
- [9] F. Chaumette and S. Hutchinson, "Visual servo control. i. basic approaches," *IEEE Rob. & Auto. Mag.*, vol. 13, pp. 82–90, 2006.
- [10] N. J. Cowan and D. E. Chang, "Geometric visual servoing," *IEEE Trans. on Rob.*, vol. 21, no. 6, pp. 1128–1138, 2005.
- [11] O. M. Carrasco-Zevallos, C. Viehland, B. Keller, *et al.*, "Review of intraoperative optical coherence tomography: technology and applications," *Biomed. Opt. Express*, vol. 8, no. 3, pp. 1607–1637, 2017.
- [12] J. P. Ehlers, J. Goshe, W. J. Dupps, *et al.*, "Determination of feasibility and utility of microscope-integrated optical coherence tomography during ophthalmic surgery: the discover study rescans results," *JAMA Ophthalmology*, vol. 133, no. 10, pp. 1124–1132, 2015.
- [13] C. Viehland, B. Keller, O. M. Carrasco-Zevallos, *et al.*, "Enhanced volumetric visualization for real time 4d intraoperative ophthalmic swept-source oct," *Biomed. Opt. Express*, vol. 7, no. 5, pp. 1815–1829, 2016.
- [14] M. Zhou, M. Hamad, J. Weiss, *et al.*, "Towards robotic eye surgery: Marker-free, online hand-eye calibration using optical coherence tomography images," *IEEE Rob. and Auto. Lett.*, vol. 3, no. 4, pp. 3944–3951, 2018.
- [15] S. S. Gurbani, P. Wilkening, M. Zhao, *et al.*, "Robot-assisted three-dimensional registration for cochlear implant surgery using a common-path swept-source optical coherence tomography probe," *J. of Biomed. Opt.*, vol. 19, no. 5, p. 057004, 2014.
- [16] S. A. Boppart, J. Herrmann, C. Pitris, *et al.*, "High-resolution optical coherence tomography-guided laser ablation of surgical tissue," *J. of Surg. Res.*, vol. 82, no. 2, pp. 275–284, 1999.
- [17] B. Y. Leung, P. J. Webster, J. M. Fraser, and V. X. Yang, "Real-time guidance of thermal and ultrashort pulsed laser ablation in hard tissue using inline coherent imaging," *Lasers in surgery and medicine*, vol. 44, no. 3, pp. 249–256, 2012.
- [18] Y. Zhang, T. Pfeiffer, M. Weller, *et al.*, "Optical coherence tomography guided laser cochleostomy: Towards the accuracy on tens of micrometer scale," *BioMed Res. Int.*, vol. 2014, 2014.
- [19] H. Yu, J.-H. Shen, R. J. Shah, N. Simaan, and K. M. Joos, "Evaluation of microsurgical tasks with oct-guided and/or robot-assisted ophthalmic forceps," *Biomed. Opt. Express*, vol. 6, pp. 457–472, 2015.
- [20] L.-A. Dufлот, B. Tamadazte, N. Andreff, and A. Krupa, "Wavelet-based visual servoing using oct images," in *IEEE Int. Conf. on Biomed. Rob. and Biomechatronics*, 2018, pp. 621–626.
- [21] T. Murakami and K. Ogawa, "Speckle noise reduction of optical coherence tomography images with a wavelet transform," in *IEEE Int. Coll. on Sig. Proces. & Its Appl.*, 2018, pp. 31–34.
- [22] M. Ourak, A. De Simone, B. Tamadazte, G. J. Laurent, A. Menciassi, and N. Andreff, "Automated in-plane oct-probe positioning towards repetitive optical biopsies," in *IEEE Int. Conf. on Rob. and Auto.*, 2016, pp. 4186–4191.
- [23] M. Ourak, B. Tamadazte, O. Lehmann, and N. Andreff, "Direct visual servoing using wavelet coefficients," *IEEE/ASME Trans. on Mech.*, vol. 24, no. 3, pp. 1129–1140, 2019.
- [24] M. Ourak, B. Tamadazte, and N. Andreff, "Partitioned camera-oct based 6 dof visual servoing for automatic repetitive optical biopsies," in *IEEE/RSJ Int. Conf. on Intel. Rob. and Syst.*, 2016, pp. 2337–2342.
- [25] J. E. Jackson, *A user's guide to principal components*. John Wiley & Sons, 2005.
- [26] L. I. Smith, "A tutorial on principal components analysis," Tech. Rep., 2002.
- [27] J. Shlens, "A tutorial on principal component analysis," *arXiv preprint arXiv:1404.1100*, 2014.
- [28] E. Marchand, "Subspace-based direct visual servoing," *IEEE Rob. and Auto. Lett.*, vol. 4, no. 3, pp. 2699–2706, 2019.
- [29] E. Malis, F. Chaumette, and S. Boudet, "2 1/2 d visual servoing," *IEEE Tans. on Rob. and Auto.*, vol. 15, no. 2, pp. 238–250, 1999.
- [30] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," in *IEEE Int. Conf. on Rob. and Auto.*, Shanghai, China, May 9-13 2011.
- [31] L.-A. Dufлот, R. Reichenhofer, B. Tamadazte, N. Andreff, and A. Krupa, "Wavelet and shearlet-based image representations for visual servoing," *The Int. J. of Rob. Res.*, vol. 38, no. 4, pp. 422–450, 2018.
- [32] D. Garcia, "Robust smoothing of gridded data in one and higher dimensions with missing values," *Computational statistics & data analysis*, vol. 54, no. 4, pp. 1167–1178, 2010.