

# Direct Visual Servoing using Wavelet Coefficients

Mouloud Ourak, Brahim Tamadazte, *IEEE Member*, Olivier Lehmaan and Nicolas Andreff, *IEEE Member*

**Abstract**—This paper focuses on the design of a 6 degrees-of-freedom (DoF) visual servoing control law. Instead of use of geometric visual features in standard vision-based approaches, the proposed controller makes use of wavelet coefficients as control signal inputs. It uses the multiple resolution coefficients representing the wavelet transform of an image in the spatial domain. The main contributions are the definition of the multiple resolution wavelet interaction model that links the time-variation of wavelet coefficients to the robot spatial velocity and the associated task function controller. The proposed control law was tested and validated experimentally using a commercial micromanipulator in an *eye-to-hand* configuration. To be able to judge the efficiency of the control law, several validation tests were carried out under different conditions of use i.e., large illumination variations, noisy images, partial occlusions and using unknown 3D scenes. It is also demonstrated experimentally that the proposed approach outperforms, the well-known photometry visual servoing as well as a feature-based visual servoing, namely in unfavorable conditions of use.

## I. INTRODUCTION

### A. Overview and Motivations

THE control of a dynamic system through real-time and continuous visual feedback is known as *visual servoing* [1], whereas the continuous observation of objects and the extraction of the corresponding visual features (usually geometric information) is referred to as *visual tracking*. Obviously, there is duality between both themes. In feature-based visual servoing, the success of the designed control law depends imperatively on the success of the spatio-temporal tracking method (speed, accuracy, robustness, redundancy of the visual features, etc.). Designing an efficient and robust visual tracking algorithm is a non-trivial task and it is considered as one of the bottlenecks of the expansion of visual servoing in different applications [2].

To tackle this limitation, new visual servoing paradigms have appeared in the literature which demonstrate that the design of a vision-based control law can completely overcome the visual tracking task in the control loop. These approaches are referred to as *featureless visual servoing* [3] also known as *direct visual servoing methods*. Thus, during the last decade, several methods were proposed using the global image information as the signal input into the control loop: image-intensity [4]–[8], spatio-temporal gradients [9], histogram-based method [10], photometric Gaussian mixtures [11], photometric moments [12], etc. Using global image

information allows also improving both the accuracy and the robustness of visual servoing tasks thanks to the redundancy of the information used to design the interaction model [5]. In [13], the authors have investigated the use of neural network to estimate the interaction matrix of a direct visual servoing under several perturbations (occlusions, lighting variations, etc.). It has been demonstrated that using the neural network, it allowed to overcome the photometric visual servoing drawbacks.

Although these techniques are different from the conventional geometric feature-based control schemes, their aim remains the same. This means that the robot, from its initial pose  $\mathbf{r}(t)$  ( $\mathbf{r} \in SE(3)$ ), must reach accurately a desired pose  $\mathbf{r}^*$  using an optimization algorithm of the associated cost-function. The camera velocity twist is then computed using the time-derivation of the cost function with respect to the camera pose  $\mathbf{r}$ . As mentioned above, the direct methods have demonstrated very interesting behavior in terms of accuracy and robustness. However, it has been shown that these approaches suffer in term of convergence domain as well as important environment variations compared to the traditional visual servoing [5].

To overcome these limits, additional kinds of information have been studied such as mutual information [15]. This method does not consider directly the pure image information (e.g., pixels intensities) but the mutual dependence signal (e.g., the joint-entropy) between a pair of images. Thereby, if two images are perfectly aligned, the mutual information value must be higher than the shifted image. In this work, the authors demonstrated that mutual information-based visual servoing is more robust with respect to the image-intensity-based method, namely in unfavorable conditions.

Another trivial solution to increase the robustness of the visual servoing control law with respect to unfavorable signal-to-noise ratio and important environment variations, is to use *noise-free global image information*. Indeed, instead of using a spatio-temporal image information (intensity, gradient, entropy, etc.), it is possible to consider original image information such as time-frequency image representations. Thereby, this paper addresses a new direct visual servoing scheme which uses wavelet coefficients as the signal input into the control loop. In fact, wavelets theory present a stable mathematical tool for multiscale image representation. Incidentally, wavelet transform is a representation of a signal by an orthonormal series of functions called wavelets [16], [17]. It was demonstrated that wavelet transform yields an efficient and simultaneous representation of the signal in both time and frequency domains as well as a lossless inverse transform. Consequently, wavelet transform has been widely considered in numerous image processing and vision domains: image and data compression (*JPEG2000*) [18], image smoothing and denoising [19], visual tracking [20], etc.

Traditionally, continuous wavelets can be viewed as an

The authors are with FEMTO-ST Institute, AS2M Department, Univ. Bourgogne Franche-Comté, Univ. de Franche-comté/CNRS/ENSMM, 24 rue Savary, F-25000 Besançon, France.

brahim.tamadazte@femto-st.fr

Part of this paper was published in the Proceedings of the International Conference on Robotics and Automation (ICRA) 2016 [14]

This work was supported by French ANR NEMRO Project no ANR-14-CE17-0013-001 and the ROBOT Project: ITMO Cancer no 17CP068-00.

extension of Fourier transform. The main difference consists in that wavelets allow a simultaneous representation of the signal (respectively, the image) in both time and frequency domains, whereas the standard Fourier transform is only localized in the frequency domain. Fourier transform was investigated in the past to design control laws and it has demonstrated interesting properties of robustness and accuracy [21]–[23]. More recently, shearlet image decomposition was studied in [24]–[26] showing also interesting performances to achieve 6 DoF positioning tasks.

## B. Contributions

In this paper, we develop a 6 DoF visual servoing control law based on the multi-resolution wavelet transform. Unlike the image intensities-based approaches, our approach uses the wavelet coefficients to build the control law. To do this, we established the interaction model (i.e., interaction matrix) between the time-variation of the wavelet coefficients and the spatial robot velocity. In this work, we combined *Mallat's* multiple resolution wavelet algorithm [27] with the optical flow constraint equation (OFCE) for the derivation of the interaction matrix. Due the fact that the wavelet transform is multiscale, then the interaction model is also trivially multiscale. This means that, it becomes possible to switch from different interaction matrices (without additional calculation) in function of the positioning task to be performed.

Furthermore, wavelet transform can be seen as a *spatio-temporal filter* which is able to separate the signal of interest from the noise even when there spectral overlap between both (e.g., image contours) [28]. Thereby, to the best of our knowledge, we propose for the first time (put aside our preliminary conference paper [14]) a robust and accurate visual servoing control law tacking into account the wavelet coefficients as the control signal inputs. Complementary to [14], this paper provides a complete and optimized reformulation of the wavelet-based visual servoing control law by taking into account the multiresolution aspect of the wavelet coefficients during the design of the related interaction matrix. For instance, the extended method offers the possibility to adapt the interaction matrix in function during the achievement of the positioning task without any additional computation. As example, the operator can choose a high resolution at the beginning of the positioning task and decrease automatically the resolution in function of the positioning error. This allows enhancing the convergence domain while guaranteeing high accuracy at the end of the positioning task. In fact, we were interested in tuding the impact of the wavelet coefficients resolution on the controller behavior and performances. The proposed method and materials were validated in both simulation and experimental manners using various favorable and unfavorable scenarios. Finally, a comparison study with the photometry as well as the conventional points-based visual servoing methods was provided.

In the remainder of this paper, Section II reminds the basics of a visual servoing scheme. Section III presents a general formulation of the multiresolution wavelet decomposition which allows following the different steps of the

design of the new visual servoing control law. The time-variation of the wavelet coefficients with respect to the camera displacement as well as the derived visual servoing control law are detailed in Section IV while its simulation as well as experimental validations are shown in Section V. Also, the proposed method is compared to both the photometric and the points-based visual servoing methods in Section VI. This paper ends with a discussion about the potential of such controller in medical context, e.g., high accurate automatic positioning of an OCT (Optical Coherence Tomography) probe in case of repetitive optical biopsies investigation (preliminary results are discussed).

## II. BASICS OF VISUAL SERVOING

### A. Feature-based Visual Servoing

According to [1], the aim of a visual servoing is to control the motion of a dynamic system in order to allow a set of geometric visual features  $\mathbf{s}$  ( $\mathbf{s} \in \mathbb{R}^k$ ) defining a robot pose  $\mathbf{r}(t) \in SE(3)$  (i.e.,  $\mathbf{s} = \mathbf{s}(\mathbf{r}(t))$ ) to reach a set of desired ones  $\mathbf{s}^*$  ( $\mathbf{s}^* \in \mathbb{R}^k$ ) by minimizing a visual error given by

$$\mathbf{e} = \mathbf{s}(\mathbf{r}(t)) - \mathbf{s}^* \quad (1)$$

The time-variation of  $\mathbf{s}$  is linked to the velocity twist  $\mathbf{v} = (v_x \ v_y \ v_z \ \omega_x \ \omega_y \ \omega_z)^\top$  of the camera frame by  $\dot{\mathbf{s}} = \mathbf{L}_s \mathbf{v}$ , where  $\mathbf{L}_s \in \mathbb{R}^{k \times 6}$  is commonly called the *interaction matrix* (or *Jacobian image* in certain papers).

From (1), the variation of the visual error  $\mathbf{e}$  due to the visual sensor velocity is

$$\dot{\mathbf{e}} = \mathbf{L}_s \mathbf{v} - \dot{\mathbf{s}}^* \quad (2)$$

If  $\dot{\mathbf{s}}^* = 0$  and if one want to ensure an exponential decoupled decrease of the error  $\mathbf{e}$  ( $\dot{\mathbf{e}} = -\lambda \mathbf{e}$ ), it becomes possible to express the velocity tensor, using (2), as follows

$$\mathbf{v} = -\lambda \widehat{\mathbf{L}}_s^+ (\mathbf{s}(t) - \mathbf{s}^*) \quad (3)$$

where  $\lambda$  is a positive gain and  $\widehat{\mathbf{L}}_s^+ \left( \widehat{\mathbf{L}}_s^+ = (\widehat{\mathbf{L}}_s^\top \widehat{\mathbf{L}}_s)^{-1} \widehat{\mathbf{L}}_s^\top \right)$  is the *Moore-Penrose* pseudo-inverse of the approximated interaction matrix denoted  $\widehat{\mathbf{L}}_s$ .

Not that the stability issues regarding this kind of controller are discussed in [?].

### B. Direct Visual Servoing

Generally, in a direct visual servoing approach (e.g., photometry), the control task is considered as an alignment between the current image  $\mathbf{I}(\mathbf{r})$  and the desired one  $\mathbf{I}^*$  (case of static desired position). For instance, in [5], the visual servoing task is treated as an optimization problem defined as follows

$$\arg \min_{\mathbf{r}} \|\mathbf{I}(\mathbf{r}) - \mathbf{I}^*\| \quad (4)$$

To solve this optimization problem, [5] consider the optical flow constraint equation of a world point between two successive images. This allows writing

$$\mathbf{I}(x, t) = \mathbf{I}(x + \delta x, t + \delta t) \quad (5)$$

where  $\delta x$  is small displacement of a point  $x$  during a short time interval  $\delta t$ .

The time-derivation of (5) allows writing

$$\dot{\mathbf{I}} + \nabla \mathbf{I}^\top \dot{x} = 0 \quad (6)$$

where  $\nabla$  is the image gradient.

By introducing the 2D-point interaction matrices  $\mathbf{L}_x$  and  $\mathbf{L}_y$  (as defined in [1]) and some mathematical manipulations, the authors of [5] obtain the global interaction matrix for each pixel

$$\mathbf{L}_1 = -\left(\nabla \mathbf{I}_x \mathbf{L}_x + \nabla \mathbf{I}_y \mathbf{L}_y\right) \quad (7)$$

As can be shown, the interaction matrix  $\mathbf{L}_1$  depends on the image intensities as well as the image gradient. This can raise some problems concerning robustness of this approach in case of unstable illumination.

### III. MULTIPLE RESOLUTION WAVELET COEFFICIENTS AS VISUAL SIGNAL

As claimed before, in this paper, the use of wavelet coefficients as visual features to build our control law constitutes the core of this work. Consequently, for a better understanding of the different followed steps to design the visual servoing control law, it is appropriate to remember some basics of the wavelet decomposition.

First, a wavelet transform is defined as a  $L^2$ -inner product between a 2D signal  $\mathcal{F}$  and a wavelet function  $\Psi$  given by

$$\langle \mathcal{F}, \mathcal{G}_{j,\mathbf{m}} \rangle = \iint_{\mathbb{R}^2} \mathcal{F}(x,y) \mathbb{D}_2^j \mathbb{T}_{\mathbf{m}} \Psi(x,y) dx dy \quad (8)$$

where  $j \in \mathbb{Z}$  defines the scale and  $\mathbf{m} = (m_1, m_2) \in \mathbb{Z}^2$  the location of the wavelet  $\Psi_{j,\mathbf{m}} = \mathbb{D}_2^j \mathbb{T}_{\mathbf{m}} \Psi$ . The squeezing and stretching the generator  $\Psi$  are obtained by applying the dyadic scaling operator  $\mathbb{D}_2^j$  which varies the wavelet spatial frequency and by the translation operator  $\mathbb{T}_{\mathbf{m}}$  which defines the wavelet localization.

This transform can be categorized into different groups as continuous [16], discrete [17] and multiple resolution [27]. In [31], we developed a *direct pose-based visual servoing* control law in which the control signal inputs consisted of the frequency spectral information computed by the continuous wavelet transform. However, it has been shown that the calculation of the out-plane rotations remains very challenging. Therefore, in this paper, we will focus of the second family of wavelets, i.e., the multiple resolution decomposition.

#### A. Multiple Resolution Wavelet Transform

The multiple resolution wavelet transform (MRWT) has essentially the ability to represent an image at different resolutions without any loss of information, in both decomposition or re-composition of the image without loss of the signal details. Actually, the difference of information subtracted in the decimation is trivially and simultaneously accessible through the wavelet coefficients. Indeed, at resolution  $2^{-j}$ , the MRWT decomposes the image into sub-images, which are: 1) the new filtered image at a resolution  $2^{-(j+1)}$ , and

2) the *difference* of information between the two consecutive resolutions.

The MRWT method computes the orthogonal projections (coordinates) of the signal  $\mathcal{F}$  on the sub-spaces  $\mathcal{V}_{2^j}$  and  $(\mathcal{W}_{2^j}^k)_{(k=0,\dots,n)}$  related to the new image and the *difference* of information, respectively. More precisely, these coordinates are computed thanks to two functions:

- *scaling function*  $\Phi$  defined by

$$\Phi_{j,\mathbf{m}} = \sqrt{2^{-j}} \Phi_{2^j}(\mathbf{m} - 2^{-j}\mathbf{k}) \quad (9)$$

where  $\mathbf{m}$  is the signal coordinates,  $\mathbf{k} = (k_1, k_2)$  are integers.

- and the *mother wavelet*  $\Psi$  is given by

$$\Psi_{j,\mathbf{m}} = \sqrt{2^{-j}} \Psi_{2^j}(\mathbf{m} - 2^{-j}\mathbf{k}) \quad (10)$$

The multiresolution wavelet decomposition offers numerous interesting properties. For instance, it leads to compute the coefficients coordinates using a linear filtering and decimation ( $\Delta$ ) (i.e., keep one pixel of two). An algorithm was developed by Mallat [27] in which it is proposed to discretize (9) and (10), using low and high pass discrete filters. In our approach, we choose the fourth order *Daubechies* wavelet filters (*db4*) [17] accessible from the wavelets MATLAB toolbox. These filters are defined by the following convolution masks:

- *Scaling function*  $\Phi$  is a low pass discrete filter  $l_{db4}(f_x, \xi_l) = \{-0.01, 0.03, 0.03, -0.18, -0.02, 0.63, 0.71, 0.23\}$ .
- *Mother wavelet*  $\Psi$  is a high pass discrete filter  $h_{db4}(f_x, \xi_h) = \{-0.2, 0.7, -0.63, -0.02, 0.18, 0.03, -0.03, -0.01\}$  and constructed from  $\Phi$ , fully respecting the conditions cited in [27].

The pair of filters is applied to the image through a set of 4 combined operations [27] as follows

$$\Gamma^0 = \left(\Delta \circ l_{db4}\right) \circ \left(\Delta \circ l_{db4}\right) \quad (11)$$

$$\Gamma^H = \left(\Delta \circ h_{db4}\right) \circ \left(\Delta \circ l_{db4}\right) \quad (12)$$

$$\Gamma^V = \left(\Delta \circ l_{db4}\right) \circ \left(\Delta \circ h_{db4}\right) \quad (13)$$

$$\Gamma^D = \left(\Delta \circ h_{db4}\right) \circ \left(\Delta \circ h_{db4}\right) \quad (14)$$

where the operator  $\circ$  represents a convolution operation.

The set of inner products (8) using the  $N \times M$  image  $\mathbf{I}_{(2^{-j})}$  and the wavelet function (11), allows writing

$$\mathbf{I}_{2^{-(j+1)}}(u,v) = \left( \left\langle \mathbf{I}_{2^{-j}}(x,y), \Gamma^0(x,y) \right\rangle \right)_{N \times M} \quad (15)$$

where  $\mathbf{I}_{2^{-(j+1)}}$  is the new image at resolution  $j+1$  (i.e.,  $\frac{N}{2^{j+1}} \times \frac{M}{2^{j+1}}$ ) called the *approximation signal*.

Then, the computation of (8) with the image  $\mathbf{I}_{2^{-j}}$  and the wavelet functions (12) to (14) gives the so-called *detail signal* or difference of information  $(\mathbf{d}_{2^{-(j+1)}}^k)_{(k=H,V,D)}$  as

$$\mathbf{d}_{2^{-(j+1)}}^k(u,v) = \left( \left\langle \mathbf{I}_{2^{-j}}(x,y), \Gamma^k(x,y) \right\rangle \right)_{N \times M}, \forall k = H, V, D \quad (16)$$

with  $H$  the horizontal,  $V$  the vertical, and  $D$  the diagonal orientations.

## B. Building the Visual Features Vector

The mathematical model for optical flow constraint is based on the time-independent of the brightness assumption [30]. Indeed, it is possible to generalize the optical flow constraint equation (OFCE) given in (5) to a whole image  $\mathbf{I}$ . Thereby, according to [32], the OFCE can be extended to the MRWT. Therefore, for each image point having the coordinates  $p = (u, v)^\top$ , it is possible to write

$$\left\langle \mathbf{I}_{2^{-(j+1)}}(u, v), \frac{\partial \Gamma^H(u, v)}{\partial u} \right\rangle \dot{u} + \left\langle \mathbf{I}_{2^{-(j+1)}}(u, v), \frac{\partial \Gamma^V(u, v)}{\partial v} \right\rangle \dot{v} + \frac{\partial}{\partial t} \left\langle \mathbf{I}_{2^{-(j+1)}}(u, v), \Gamma^0(u, v) \right\rangle = 0 \quad (17)$$

To simplify the notation, let us introduce

$$\mathbf{g}_{2^{-(j+1)}}^H(u, v) \triangleq \left\langle \mathbf{I}_{2^{-j}}(u, v), \frac{\partial \Gamma^H(u, v)}{\partial u} \right\rangle \quad (18)$$

$$\mathbf{g}_{2^{-(j+1)}}^V(u, v) \triangleq \left\langle \mathbf{I}_{2^{-j}}(u, v), \frac{\partial \Gamma^V(u, v)}{\partial v} \right\rangle \quad (19)$$

and reformulate (17) as

$$\mathbf{g}_{2^{-(j+1)}}^H(u, v) \dot{u} + \mathbf{g}_{2^{-(j+1)}}^V(u, v) \dot{v} + \frac{\partial}{\partial t} \mathbf{I}_{2^{-(j+1)}}(u, v) = 0 \quad (20)$$

or, in a matrix form

$$\frac{d}{dt} \mathbf{I}_{2^{-(j+1)}}(u, v) = \begin{bmatrix} \mathbf{g}_{2^{-(j+1)}}^H(u, v) & \mathbf{g}_{2^{-(j+1)}}^V(u, v) \end{bmatrix} \begin{bmatrix} \dot{u} \\ \dot{v} \end{bmatrix} \quad (21)$$

where,  $\dot{u}$  and  $\dot{v}$  are obtained from  $\dot{x}$  and  $\dot{y}$  by the camera calibration matrix. Thus,

$$\begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix} = \begin{pmatrix} \alpha_u & 0 \\ 0 & \alpha_v \end{pmatrix} \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} \quad (22)$$

where  $\alpha_u$  and  $\alpha_v$  represent the focal length in terms of pixels and  $\dot{x}$  and  $\dot{y}$  represent the velocity of an image point of the coordinates  $(x, y)^\top$ .

So, by merging (22) in (23), we obtain

$$\frac{d}{dt} \mathbf{I}_{2^{-(j+1)}}(u, v) = \begin{bmatrix} \mathbf{g}_{2^{-(j+1)}}^H(u, v) & \mathbf{g}_{2^{-(j+1)}}^V(u, v) \end{bmatrix} \begin{pmatrix} \alpha_u & 0 \\ 0 & \alpha_v \end{pmatrix} \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} \quad (23)$$

Furthermore, it is possible to express  $\dot{x}$  and  $\dot{y}$  as function of the camera velocity tensor  ${}^c\mathbf{v}_c$  thanks to the 2D-point interaction matrix  $\mathbf{L}_{2D}$  as defined in [1]. Thus,

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \mathbf{L}_{2D}^c \mathbf{v}_c \quad (24)$$

where,

$$\mathbf{L}_{2D} = \begin{pmatrix} -1/Z & 0 & x/Z & xy & -(1+x^2) & y \\ 0 & -1/Z & y/Z & -(1+y^2) & -xy & -x \end{pmatrix} \quad (25)$$

By introducing (24) and (25) in (23), and by omitting the pixel coordinates in various frames for the sake of readability, we finally get

$$\frac{\partial}{\partial t} \mathbf{I}_{2^{-(j+1)}} = - \begin{bmatrix} \mathbf{g}_{2^{-(j+1)}}^H & \mathbf{g}_{2^{-(j+1)}}^V \end{bmatrix} \begin{pmatrix} \alpha_u & 0 \\ 0 & \alpha_v \end{pmatrix} \mathbf{L}_{2D} {}^c\mathbf{v}_c \quad (26)$$

The interaction matrix associated to the wavelet coefficients is then expressed in each pixel as follows

$$\mathbf{L}_{w_{2^{-(j+1)}}} = - \begin{bmatrix} \mathbf{g}_{2^{-(j+1)}}^H & \mathbf{g}_{2^{-(j+1)}}^V \end{bmatrix} \begin{pmatrix} \alpha_u & 0 \\ 0 & \alpha_v \end{pmatrix} \mathbf{L}_{2D} \quad (27)$$

From (26), it can be underlined that the interaction matrix is built using the time-derivative wavelet coefficients. The depth  $Z$  is estimated once at the desired position.

Now, let us consider the down sampled image  $\mathbf{I}_{2^{-(j+1)}}$  of size of  $\frac{N}{2^{(j+1)}} \times \frac{M}{2^{(j+1)}}$ , as a visual feature vector  $\mathbf{w}$ . Thus,

$$\mathbf{w}_{2^{-(j+1)}} = \text{vec}(\mathbf{I}_{2^{-(j+1)}}) = \begin{bmatrix} I_{2^{-(j+1)}(1,1)} \\ I_{2^{-(j+1)}(1,2)} \\ \vdots \\ I_{2^{-(j+1)}(M,N)} \end{bmatrix} \quad (28)$$

Thus, we can write the whole wavelet variation in function of the robot velocity thanks to the whole wavelet interaction matrix  $\mathfrak{L}_{w_{2^{-(j+1)}}}$  as

$$\dot{\mathbf{w}} = \mathfrak{L}_{w_{2^{-(j+1)}}} {}^c\mathbf{v}_c \quad (29)$$

which is obtained by the vertical concatenation of all  $\mathbf{L}_{w_{2^{-(j+1)}}}$ 's in (26) as follows

$$\mathfrak{L}_{w_{2^{-(j+1)}}} = \begin{bmatrix} \mathbf{L}_{w_{2^{-(j+1)}}(1,1)} \\ \mathbf{L}_{w_{2^{-(j+1)}}(1,2)} \\ \vdots \\ \mathbf{L}_{w_{2^{-(j+1)}}(M,N)} \end{bmatrix} \quad (30)$$

## IV. CONTROL

### A. Control Law

Having the analytical wavelet interaction matrix, it becomes trivial to design the related control law. Actually, following (3) allows writing the following controller

$${}^c\mathbf{v}_c = -\lambda \widehat{\mathfrak{L}}_w^+ (\mathbf{w} - \mathbf{w}^*) \quad (31)$$

Actually, to ensure a stable and smooth control, we use the *Levenberg-Marquardt* method [33] rather than the usual *Gauss-Newton* or *Gradient* descent optimization as addressed in [5]. The choice of *Levenberg-Marquardt* leads us to write the robot velocity in the camera's frame  $\mathcal{R}_c$  as

$${}^c\mathbf{v}_c = -\lambda \left( \mu \mathbb{I}_{6 \times 6} + \widehat{\mathfrak{L}}_w^\top \widehat{\mathfrak{L}}_w \right)^{-1} \widehat{\mathfrak{L}}_w^+ (\mathbf{w} - \mathbf{w}^*) \quad (32)$$

where  $\mu$  is a constant positive scalar, and  $\mathbb{I}_{6 \times 6}$  is the identity matrix.

Our system (Fig. 6) is designed in an *eye-to-hand* configuration. Thus, the relation between the robot velocity  $\dot{\mathbf{q}}$  and the camera one  ${}^c\mathbf{v}_c$  is obtained as follows

$$\dot{\mathbf{q}} = -{}^e\mathbf{K}_b {}^b\mathbf{V}_c {}^c\mathbf{v}_c \quad (33)$$

where  ${}^e\mathbf{K}_b$  is the inverse kinematic Jacobian matrix of the robot in the base frame  $\mathcal{R}_b$ , and  ${}^b\mathbf{V}_c$  is the transformation

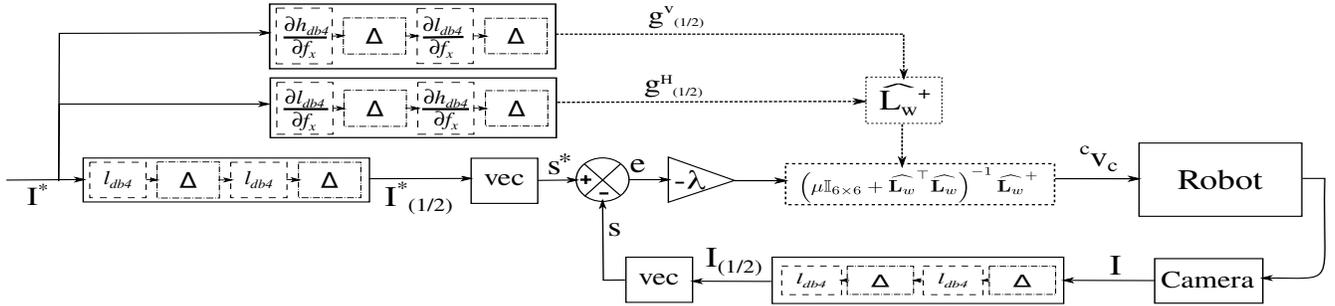


Fig. 1: Closed loop visual servoing using the half resolution wavelet coefficients as visual feature.

matrix associated to the velocity change frame from  $\mathcal{R}_b$  to  $\mathcal{R}_c$  (i.e., the camera frame). The latter is constructed as

$${}^b\mathbf{V}_c = \begin{bmatrix} {}^b\mathbf{R}_c & [{}^b\mathbf{t}_c]_{\wedge} & {}^b\mathbf{R}_c \\ 0 & & \end{bmatrix} \quad (34)$$

where  ${}^b\mathbf{R}_c$  is the  $3 \times 3$  rotation matrix from  $\mathcal{R}_c$  to  $\mathcal{R}_b$ ,  ${}^b\mathbf{t}_c$  is the  $3 \times 1$  associated translation vector, and " $\wedge$ " is the skew symmetric matrix associated to the vector cross-product. The proposed control law follows the flowchart illustrated in Fig. 1.

### B. Impact of the Wavelet Resolution on the Control Performances

As mentioned above, the multiresolution interaction matrix is designed. Thereby, it is possible to intuitively switch from one resolution to another without additional calculation. Here, the idea is to judge the capability of the controller to work with high resolution *approximation* signal as well as with low ones. Thus, we computed and plotted the cost-functions of different wavelet coefficients resolutions  $j = [0, 1, 2, 3]$  using *Lena* photography as the scene viewed by the camera Fig. 2. This will allow us to objectively choose the *approximation* signal resolution  $j$  in order to obtain the best cost-function possible (i.e., convex shape, wide at the top, tight at the bottom and having a single global minimum and smooth as possible). As claimed in the introduction, direct visual servoing techniques are distinguished by their high performances in terms of accuracy and robustness to the expense of lower convergence domain with respect to the traditional feature-based visual servoing. However, this limitation of the convergence capabilities of the wavelet-based visual servoing can be tackled using the multiresolution aspect of the developed interaction matrix. For instance, it is possible to consider high scale of the wavelet coefficients (e.g.,  $j = 0$ ) to increase the convergence domain and switch continuously (during the positioning task) towards a lower resolution (e.g.,  $j = 1, 2, \dots$  etc.) until reaching the desired position in order to guarantee high accuracy. Switching from interaction matrix resolution to another can be performed in function of the positioning error  $\mathbf{e}$ : higher value of  $\mathbf{e}$  involves higher value of the scale  $j$  and conversely.

## V. VALIDATION

### A. Simulation Validation

1) *Accuracy*: First the proposed controller was evaluated in simulation using a developed simulator using the Visual

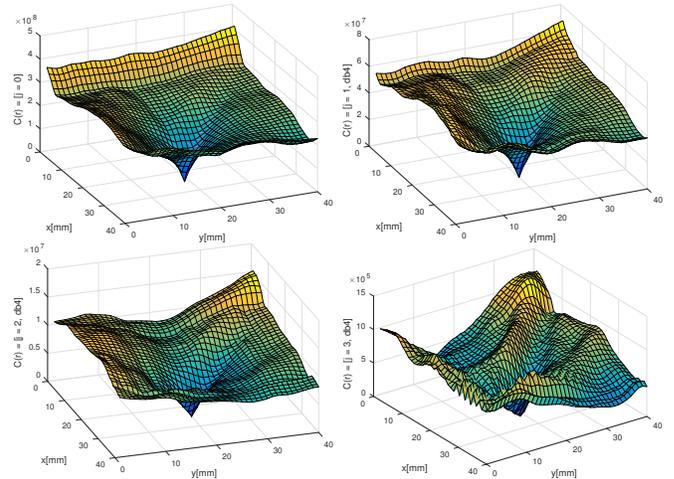


Fig. 2: From left to right and from top to bottom: computed cost-functions for  $j = [0, 1, 2, 3]$  resolutions, respectively.

Servoing Platform (ViSP) framework (visp.inria.fr). The simulation scenario consists of the achievement of an automatic 6 DoF positioning task (in  $SE(3)$ ) under nominal conditions. We also quantify the convergence domain depending on the used resolution  $j$  in the interaction matrix.

Firstly, the test of nominal conditions is performed using the photography of *Lena*. Fig. 3 depicts some snapshots captured during the positioning task achievement. As can be seen in Fig. 3(d), the normalized difference image between  $\mathbf{I}$  and  $\mathbf{I}^*$  in completely gray (perfect overlapping between the final and desired images) which means that the robot reaches accurately the desired position. Both initial and final positioning errors are reported in TABLE I. As can be highlighted, the final linear and rotational errors (obtained using the simulated anthropomorphic robot encoders) are extremely low.

TABLE I: [Simulation - Nominal Conditions] - Numerical values in case of nominal conditions ( $\Delta T_i$  (mm) and  $\Delta R_i$  ( $^\circ$ ))

pose	$\Delta T_x$	$\Delta T_y$	$\Delta T_z$	$\Delta R_x$	$\Delta R_y$	$\Delta R_z$
initial	-15	-15	10	8	-7	-5
desired	0.0	0.0	0	0.0	0.0	0.0
<b>final error</b>	<b><math>10^{-8}</math></b>	<b><math>10^{-7}</math></b>	<b><math>10^{-5}</math></b>	<b><math>10^{-6}</math></b>	<b><math>10^{-5}</math></b>	<b><math>10^{-6}</math></b>

The error decay in each DoF as well as the norm of the global error are recorded and plotted in Fig. 4(a) and (b), respectively. One can underline that all the errors converge towards zero. By the same manner, the camera linear and

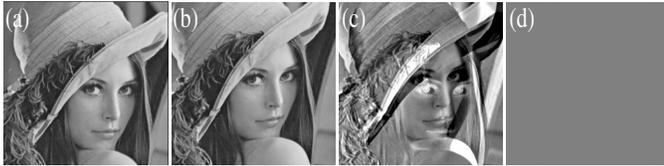
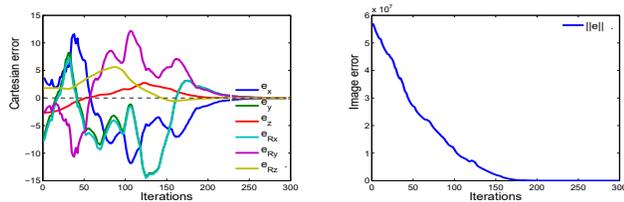
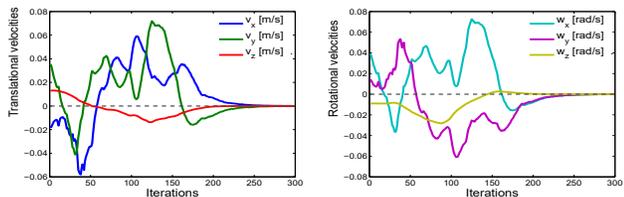


Fig. 3: [Simulation - Nominal Conditions] - Image sequence captured during the 3D positioning task. (a):  $\mathbf{I}(\mathbf{r})$ , (b):  $\mathbf{I}(\mathbf{r})^*$ , (c):  $\mathbf{I}(\mathbf{r}) - \mathbf{I}(\mathbf{r})^*$ , and (d) final error showing that the controller reaches accurately  $\mathbf{r}^*$ .

angular velocities are plotted in Fig. 4(c) and (d), respectively.



(a) Left: Cartesian error progression, and right: norm of the error in the image



(b) Camera velocities. Left: linear velocities  $(v_x, v_y, v_z)^T$  and right: angular velocities  $(\omega_x, \omega_y, \omega_z)^T$ .

Fig. 4: [Simulation - Nominal Conditions] - Errors and velocities evolution vs. number of iterations.

2) *Convergence Domain*: Here the idea is to perform the same positioning task, i.e., using the same initial and desired positions as well as the same control gains, however we modify the resolution  $j$  from one test to another. The aim is to show the variation of the convergence domain as a function of the resolution  $j$ . Thereby, spiderweb-like representations were used to compare each convergence domain depending on the chosen resolution (Fig. 5). Each line from the center to the border of the spiderweb represents an initial error between the initial image  $\mathbf{I}$  and the desired image  $\mathbf{I}^*$  in translation (left spiderweb) or rotation (right spiderweb) in the positive or negative directions of displacement.

It can be highlighted that the convergence domain of the proposed approach is naturally related to the wavelet coefficients resolution. This means that higher resolution provides larger convergence domain and conversely. Also, the proposed method showed a better functioning in  $Z+$  direction comparing to  $Z-$  one. This can be explained, at least partially, by the fact the  $Z$  used in 2D interaction matrix (25) is approximated (not computed accurately) only one time at the desired position).

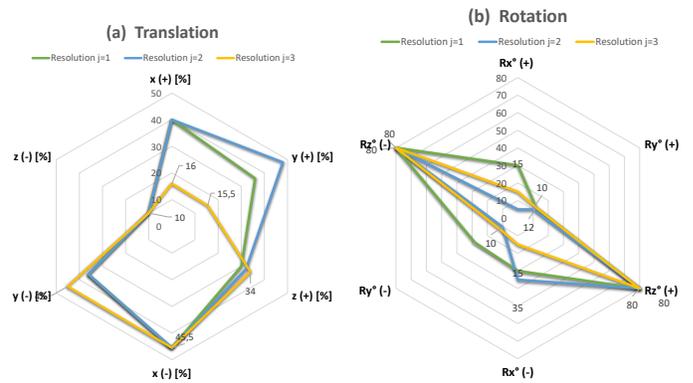


Fig. 5: Convergence domain study of wavelet coefficients resolution  $j = [1, 2, 3]$ .

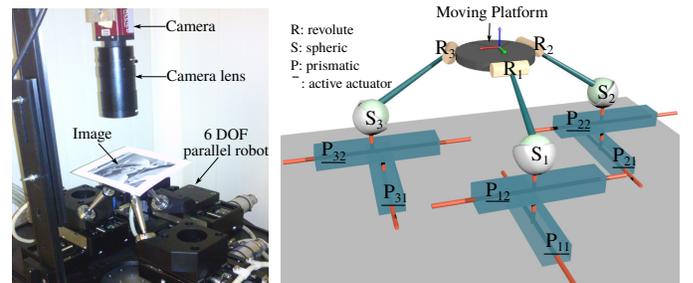


Fig. 6: Global view on the 6 DoF experimental platform.

## B. Experimental Validation

To demonstrate the effectiveness of the developed controller, we have designed an *eye-to-hand* experimental set-up (Fig. 6). It is based on a 6 DoF parallel robotic structure: a **3PPSR** robot SpaceFAB SF-3000 BS from Micos<sup>1</sup>. The latter is characterized with the following features: translation ranges  $(T_x, T_y, T_z)_{max}^T = (50, 100, 12.7)^T [mm]$  and rotation ranges  $(R_x, R_y, R_z)_{max}^T = (10, 10, 10)^T [^\circ]$ , a linear resolution of  $0.2\mu m$  (repeatability of  $\pm 0.5\mu m$ ) and an angular one of  $0.0005^\circ$  (repeatability of  $\pm 0.0011^\circ$ ). Also, a Firewire monochrome CCD camera is attached on top of the robot platform. The camera runs at 25 frames per second (fps) for a resolution of  $640 \times 480$  pixels. Two computers equip the experimental platform: the first one (a 3.20-GHz i5 core Intel CPU with a Linux distribution) is dedicated to computer vision and the proposed control (image acquisition, and control calculation, etc.) when the second one (a 2.33-GHz Xeon Intel CPU with a Windows distribution) is used for the robot inner control (inner PID loop, static and differential kinematic models). The computers communicate asynchronously using a TCP/IP protocol.

As can be underlined, the robotic workcell is dedicated to micromanipulation and micropositioning tasks. This means that, the linear and angular stages are limited in terms of translation and rotation ranges, comparing to a conventional robotic platform. Also, the used camera is equipped with a high magnification objective-lens in order to fit the micrometric resolution provided by the robotic setup. Consequently, the

<sup>1</sup><http://www.pimicos.com>

positioning task studied below are limited to small displacements. However, in visual servoing, these displacements are a function of the field of view of the used camera i.e., the wider the field of vision, the larger the displacements and *vice versa*.

### C. Experimental Validation Methodology

In order to study the behavior of the developed controller, several scenarios are carried out, in different conditions of use: 1) favorable: nominal conditions, 2) unfavorable: illumination variations, partial occlusions, different scenes (2D and 3D), etc. To do that, the studied positioning task consists of reaching a desired position  $\mathbf{r}^* \in SE(3)$  (desired image) from an arbitrary initial position  $\mathbf{r} \in SE(3)$  (initial image). The final positioning error is computed using the high resolution robot encoders supplied by the robot software. Likewise, the proposed control law is also compared to both the photometric visual servoing and traditional points-based visual servoing.

### D. Nominal Conditions

The first test consists of the validation of the wavelet-based control law in nominal conditions, i.e., using a stable illumination source and precise calibration parameters (camera parameters matrix  $\mathbf{K}_{3 \times 3}$  and *eye-to-hand* transformation  $({}^c\mathbf{V}_b)_{6 \times 6}$ ). The initial error is  $\Delta\mathbf{r}[\text{mm}, \text{deg}] = (11, 11, 5, 9, 8, 5)^T$

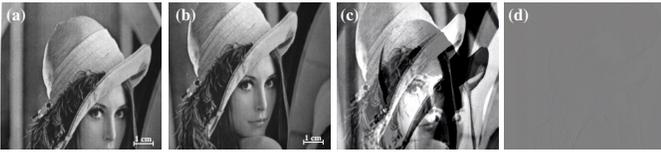


Fig. 7: [Experiment - Nominal Conditions] - Achievement of a 3D positioning task. (a)  $\mathbf{I}(\mathbf{r}^*)$ , (b)  $\mathbf{I}(\mathbf{r})$ , (c) the initial image difference, and (d) the final one.

In Fig. 7(a) and (b) are depicted the desired and the initial images, respectively, when Fig. 7(e) shows the initial error ( $\mathbf{I}(\mathbf{r}) - \mathbf{I}(\mathbf{r}^*)$ ) between the desired and the initial images. The image difference at convergence is represented in Fig. 7(f) showing that the controller converges smoothly and accurately towards  $\mathbf{r}^*$ . Note that a completely gray image difference is synonymous of a perfect overlapping between the final and desired images.

TABLE II: [Experiment - Nominal Conditions] - Numerical values in case of nominal conditions ( $\Delta T_i$  (mm) and  $\Delta R_i$  ( $^\circ$ ))

pose	$\Delta T_x$	$\Delta T_y$	$\Delta T_z$	$\Delta R_x$	$\Delta R_y$	$\Delta R_z$
initial	-6.001	6.001	78.001	4.001	4.001	3.001
desired	5.000	-5.000	74.000	-5.000	-4.000	-2.000
reached	5.009	-4.992	74.036	-4.981	-4.061	-2.014
<b>final error</b>	<b>0.009</b>	<b>0.008</b>	<b>0.036</b>	<b>0.019</b>	<b>0.061</b>	<b>0.014</b>

The positioning errors ( $\mathbf{e} = \mathbf{r} - \mathbf{r}^*$ ) are computed using the robot encoders and forward kinematic model (assumed perfectly calibrated). TABLE II gives the final obtained error in the Cartesian space demonstrating the accuracy of the proposed approach as shown by the following numerical values: average linear error =  $17.66\mu\text{m}$  and average angular error =  $0.031^\circ$ .

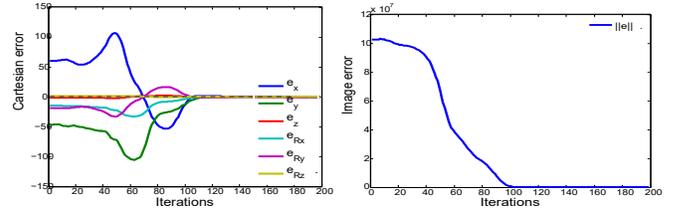


Fig. 8: [Experiment - Nominal Conditions] - Cartesian errors evolution vs. number of iterations  $i$ . right: for each DoF, left: norm of the error.

In Fig. 8(a)(left) is illustrated the pose error decay with respect to the number of iterations  $i$  (each iteration takes 0.4 seconds for non-optimized implementation). It can be seen that the different components converge to their desired values. In the same way, Fig. 8(a)(right) depicts the norm of error decay  $\|\mathbf{e}\| = \|\mathbf{I}(\mathbf{r}) - \mathbf{I}(\mathbf{r}^*)\|$ . It can be noticed that the experimental results look like the simulation ones and confirm our expectations in terms of accuracy and behavior.

### E. Robustness to Illumination Variations and Specularity

In these experiments, the appearance of the scene was modified during the positioning task by changing the illumination conditions and or by creating specularity effects in the scene. The image-intensities methods (e.g., photometry) are usually defeated because of unstable lighting conditions during the positioning task. Also, non-optimal light conditions can be a problem for feature-based methods when the extraction of geometric features (visual tracking algorithms) is not trivial.



Fig. 9: [Experiment - Illumination Variations] - Snapshots captured during the positioning task performed under lighting variation, (a) is  $\mathbf{I}^*$  acquired under a stable lighting, (b) is  $\mathbf{I}$  captured using another lighting source, (c) initial error and (d) final error.

As can be seen in Fig. 9(c) and (d), despite the change of the scene texture during the positioning task process, the controller converges to the desired position (the error decreases smoothly to become almost null). The final positioning error remains very low: the linear average error is  $45.66\mu\text{m}$  when the angular average is  $0.10^\circ$  (see TABLE III).

TABLE III: [Experiment - Illumination Variations] - Numerical values in case of lighting variations ( $\Delta T_i$  (mm) and  $\Delta R_i$  ( $^\circ$ ))

pose	$\Delta T_x$	$\Delta T_y$	$\Delta T_z$	$\Delta R_x$	$\Delta R_y$	$\Delta R_z$
initial	-5.001	5.001	78.001	3.001	3.001	3.001
desired pose	3.000	-3.000	74.000	-7.000	-4.000	-2.000
reached pose	3.091	-3.021	73.975	-7.066	-5.110	-2.139
<b>final error</b>	<b>0.091</b>	<b>0.021</b>	<b>0.025</b>	<b>0.066</b>	<b>0.110</b>	<b>0.139</b>

As for the nominal conditions, we plotted the regulation to zero of the error in Cartesian space (Fig. 10(a)). In fact, the controller demonstrates a high accuracy even under unstable lighting source.

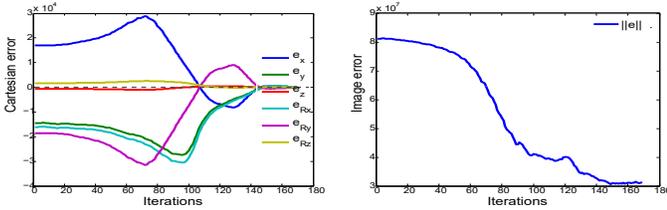


Fig. 10: [Experiment - Illumination Variations] Left: error progression, and right: norm of the error vs. number of iterations.

### F. Robustness to Partial Occlusions

These experiments deal with the study of the behavior of the control law under partial occlusions of the viewed scene during the visual servoing process. This means that, the desired image is acquired in normal conditions (Fig. 11(a)) and during the positioning task achievement, a part (almost  $\frac{1}{3}$ ) of the current images is removed (replaced by a white region-of-interest (ROI)) (Fig. 11(b)). As can be seen in Fig. 11(c) and (d), despite these external perturbations our control law converges to the desired pose without difficulty (the control law behavior is almost similar to that of the nominal conditions test).



Fig. 11: [Experiment - Partial Occlusions] - visual servoing task achievement under partial occlusions. (a)  $I^*$ , (b)  $I$ , (c) initial image difference, and (d) final one.

TABLE IV: [Experiment - Partial Occlusions] - Numerical values in case of partial occlusion ( $\Delta T_i$  (mm) and  $\Delta R_i$  ( $^\circ$ ))

pose	$\Delta T_x$	$\Delta T_y$	$\Delta T_z$	$\Delta R_x$	$\Delta R_y$	$\Delta R_z$
initial	-3.001	3.001	76.000	2.000	2.000	23.000
desired	0.000	0.000	74.000	-1.000	-1.000	-1.000
reached	0.110	0.235	74.116	-1.230	-0.982	-1.536
<b>final error</b>	<b>0.110</b>	<b>0.235</b>	<b>0.116</b>	<b>0.230</b>	<b>0.098</b>	<b>0.536</b>

Once again, the proposed approach remains robust despite the presence of partial occlusions in the image. TABLE IV gives the numerical values of the error in the Cartesian space where we can see that the average errors are  $150.36\mu\text{m}$  and  $0.288^\circ$  in linear and angular stages, respectively.

Also, another conclusion can be made on the wavelet method. The fact that the proposed approach assumes a linearization around the desired position (one of the direct visual servoing methods characteristics) which does not make it possible to ensure a global convergence.

The evolution of the Cartesian errors as well as the norm of the error are depicted in Fig. 12(a). The wavelet control law

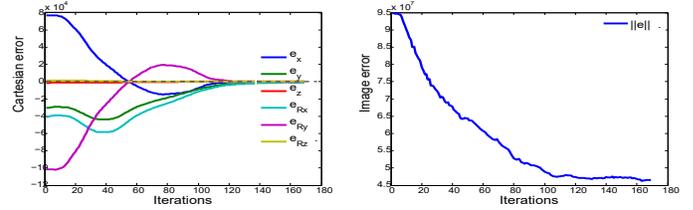


Fig. 12: [Experiment - Partial Occlusions] Left: error evolution Cartesian space, right: norm of the error vs. number of iterations.

shows the same convergence rate with respect to the nominal conditions test despite the partial occlusions (proof of robustness). This is possible because the current and the desired images share sufficiently of information (wavelet coefficients) even with the additional white ROI on the current images.

### G. Robustness to Depth Approximation

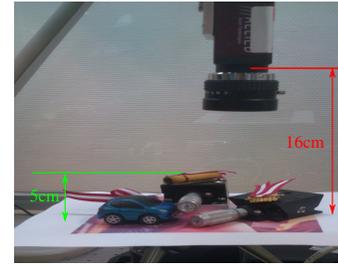


Fig. 13: Front view of the experimental set-up using a 3D scene.

The different positioning tasks presented above are performed using a fronto-parallel scene. Therefore, the effects of the depth of the scene is limited. Here, the planar scene is replaced by a 3D one (Fig. 14) where the depth  $Z$  is estimated as for 2D scene and the interaction matrix is computed again at the desired pose as for the others experiments. Note that the height (depth variations) of the 3D object is 5cm and the camera position is situated at 16cm of the scene (using another objective-lens). Also, the camera is inclined with regards to the 3D scene. As can be seen in Fig. 14, the wavelet-based control law converges smoothly to the desired pose. Indeed, the  $Z$  variations only causes a bias in the wavelet coefficients computation which are negligible as shown by the numerical values of the final error (average error of  $248\mu\text{m}$  and  $0,103\text{deg}$  in linear and angular stages, respectively) (TABLE V) which are in the same order (despite a slight deterioration) as for the nominal conditions results.

TABLE V: [Experiment - 3D Scene] - Numerical values in case of a 3D scene ( $\Delta T_i$  (mm) and  $\Delta R_i$  ( $^\circ$ ))

pose	$\Delta T_x$	$\Delta T_y$	$\Delta T_z$	$\Delta R_x$	$\Delta R_y$	$\Delta R_z$
initial	-4.001	4.001	74.001	0.001	-2.001	-2.001
desired	5.001	-5.001	79.001	2.001	-6.001	2.001
reached	4.771	-5.027	79.373	1.388	-6.381	2.101
<b>final error</b>	<b>0.279</b>	<b>0.026</b>	<b>0.372</b>	<b>0.388</b>	<b>0.380</b>	<b>0.100</b>

Concerning the behaviour of the control law, as shown in Fig. 15(a), the regulation of the error in the Cartesian space



Fig. 14: [Experiment - 3D Scene] - Achievement of the positioning task using a 3D object. (a):  $I^*$ , (b):  $I$ , (c): the initial error and (d): the final one.

presents the same aspect as for the 2D scene. Furthermore, the efficiency of the wavelet-based control law remains reliable even in case of a 3D scene (with a coarse estimation of  $Z$ ).

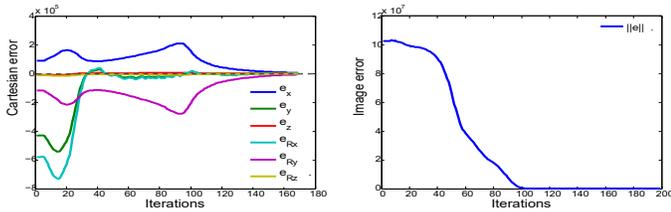


Fig. 15: [Experiment - 3D Scene] Left: error evolution in Cartesian space, right: norm of the error vs. number of iterations.

## VI. COMPARAISON AND DISCUSSION

The photometry-based visual servoing as well as the 4-points-based one (both source codes are provided by their authors within the ViSP framework<sup>2</sup>) are experimented and compared to our control law. These comparison tests were performed in both favourable and unfavourable conditions of use.

### A. Wavelet vs. Photometry

1) *Nominal Conditions*: First, let us start with a comparison in nominal conditions. TABLE VI gives the final errors in the Cartesian space. It can be seen that, the wavelet method present slightly the same accuracy (especially in translation) as the photometry method [5] in the Cartesian space. In fact, the wavelet-based method presents an average translation error of **17.66 $\mu$ m** (respectively, **17.33 $\mu$ m** for the photometry one) and an average rotation error of **0.019 $^\circ$**  (respectively, **0.051 $^\circ$**  for the photometry).

TABLE VI: [Experiment - Nominal Conditions] Numerical values of the positioning error (wavelet vs. photometry) ( $\Delta T_i$  (mm) and  $\Delta R_i$  ( $^\circ$ ))

pose	$\Delta T_x$	$\Delta T_y$	$\Delta T_z$	$\Delta R_x$	$\Delta R_y$	$\Delta R_z$
initial error	4.000	4.000	2.000	3.000	3.000	3.000
wavelet	<b>0.013</b>	<b>0.012</b>	<b>0.028</b>	<b>0.019</b>	<b>0.019</b>	<b>0.018</b>
photometry	0.008	0.015	0.029	0.054	0.044	0.055

The norms of the error in both methods are shown in Fig. 16(left) while the robot trajectory is plotted in Fig. 16(right). It can be underlined that despite the same experimental conditions (lighting source, camera and camera/robot

calibration parameters, and initial and desired positions), the wavelet-based method presents a better robot's behavior in the 3D space.

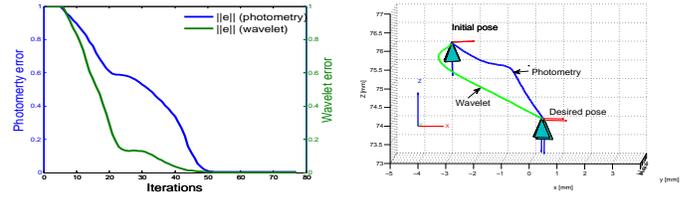


Fig. 16: (left) norm of the error (wavelet vs. photometry), and (right) robot trajectory (wavelet vs. photometry).

2) *Salt and Pepper Noise*: The last experimental test concerns a comparison of both methods under a *salt and pepper* noise of density of 70%. In Fig. 17 is depicts some snapshots grabbed during the positioning task performing. As can be highlighted, the proposed controller reaches the desired position despite the image noise. This can be also seen in Fig. 18 which represents the error regulation towards zero for both approaches. In fact, the wavelet-based control law converges without specific difficulties when the photometry method failed.

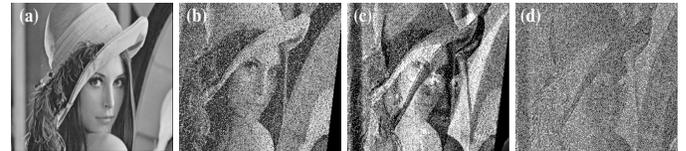


Fig. 17: [Experiment - Noisy Images] visual servoing task performing under *salt and pepper* noise. (a)  $I^*$  without noise, (b)  $I$  with noise, (c) initial error and (d) final error at convergence.

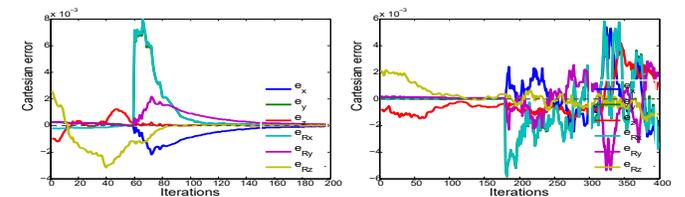


Fig. 18: [Experiment - Noisy Images] Error evolution in the Cartesian space for each DoF, (left): wavelet, and (right): photometry.

TABLE VII: [Experiment - Noisy Images] Numerical values of the positioning error (wavelet vs. photometry) ( $\Delta T_i$  (mm) and  $\Delta R_i$  ( $^\circ$ ))

pose	$\Delta T_x$	$\Delta T_y$	$\Delta T_z$	$\Delta R_x$	$\Delta R_y$	$\Delta R_z$
initial	-29.35	-29.35	9.80	-8	-5	10
wavelet	<b>0.131</b>	<b>1.573</b>	<b>0.12</b>	<b>0.987</b>	<b>2.188</b>	<b>0.028</b>
photometry	3.78	28.46	16.99	19.93	-7.53	6.96

TABLE VII summarizes the obtained numerical values, thus the average translation error for the wavelet approach is **60 $\mu$ m** and **1.06 $^\circ$**  for translation and rotation stages, respectively, when the photometry method failed. This proves that our

<sup>2</sup>www.visp.inria.fr

controller is more robust to noise disturbances comparing to the photometry technique.

### B. Wavelet vs IBVS

1) *Nominal Conditions*: In this section, we compare the popular 4-point-based visual servoing approach to our method. The experimental scenario remains the same (i.e., performing of a positioning task in  $SE(3)$ ). In order to ensure the same experimental conditions for both methods, we add four white markers into the photography of *Lena* to be tracked easily in case of IBVS method. In parallel, the wavelet method use the rest of the image to perform the same positioning task. The estimated accuracy is reported in TABLE VIII.

TABLE VIII: [Nominal Conditions] Numerical values of the positioning error (wavelet vs. IBVS) ( $\Delta T_i$  (mm) and  $\Delta R_i$  ( $^\circ$ ))

pose	$\Delta T_x$	$\Delta T_y$	$\Delta T_z$	$\Delta R_x$	$\Delta R_y$	$\Delta R_z$
initial	4.000	4.000	2.000	3.000	3.000	3.000
wavelet	<b>0.013</b>	<b>0.012</b>	<b>0.028</b>	<b>0.019</b>	<b>0.019</b>	<b>0.018</b>
IBVS	0.022	0.015	0.014	0.036	0.025	-0.001

Furthermore, Fig. 19(a) gives the error decay in each visual servoing task. It can be highlighted that our method presents a rapid decay of the norm error comparing to the IBVS one. Similarly, Fig. 19(b) shows the robot trajectory in each case. Also, the wavelet-based control law presents a similar trajectory (but takes opposite path) compared to the IBVS method.

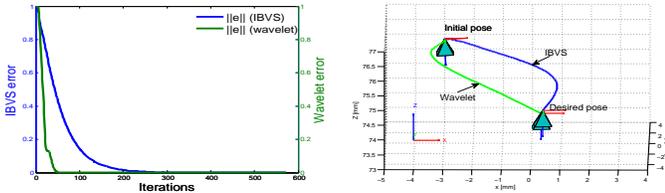


Fig. 19: Wavelet vs. IBVS: (a) norm of the error (b) robot trajectory.

### C. Towards OCT-based Visual Servoing

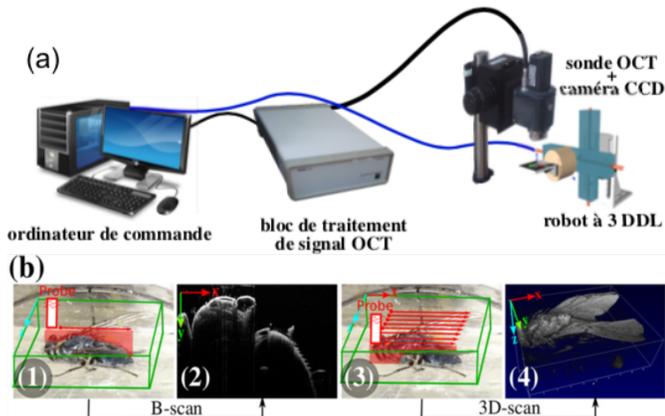


Fig. 20: Overview of the robotic setup equipped with an OCT device. (a): experimental setup, (b): operating of a OCT device.

The final objectives of the proposed work targets medical applications. More precisely, it will concern the automation of optical biopsies acquisition using an Optical Coherence Tomography (OCT) system. The latter is widely used in clinical purposes, e.g., for *in-situ* tissue characterization, especially in ophthalmology, dermatology, gastrointestinal, etc. Indeed, OCT provides very good lateral and axial resolution:  $4\mu\text{m}$  and  $3\mu\text{m}$ , respectively, and higher penetration depths 1-5 mm (Fig. 20). In addition, OCT imaging system offers real-time acquisition (several dozen images per frame) entirely compatible to robot-vision applications.

Therefore, in order to demonstrate the ability of the proposed wavelet control law to work even using untextured medical images (e.g., OCT images), we implemented a positioning task in  $SE(2) = \mathfrak{R}(2) \times SO(1)$ . The control signal inputs are the wavelet coefficients computed from the B-Scan OCT images (i.e.,  $xz$  or  $yz$  slices of the viewed biological sample). The task achievement works as follows: 1) the operator defines (grabs) an optical image of the sample at time  $t$ , and 2) the robotic platform which holds the sample must retrieve automatically the initially acquired OCT image in case of temporal monitoring of the tissue. Thereby, the wavelet control law moves the positioning platform towards the desired position using the wavelet coefficients computed from the B-Scan images. As can be seen in (Fig. 21), the control law allows converging to the desired optical biopsy without much difficulties (as can be seen in the image difference) despite the low signal-to-noise rate which characterizes OCT images.

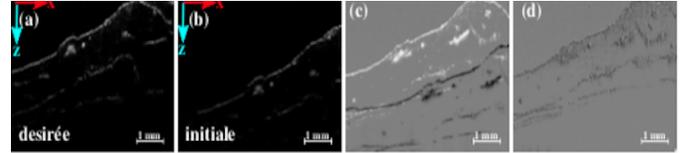


Fig. 21: OCT image snapshots acquired during the positioning task performing: (a) desired OCT image, (b) initial OCT image, (c) initial OCT image difference, and (d) final OCT image difference.

## VII. CONCLUSION

In this paper, it was presented a novel 6 DoF visual servoing controller based on the use of the multiple resolution wavelet coefficients as the visual signal in the control loop instead of geometric visual features. The developed interaction matrix is multiple scale which means that it is possible to switch intuitively from a resolution to another without additional calculation. The method was successfully validated following several simulation and experimental scenarios (optimal conditions, lighting variations, partial occlusions, 3D scene). The obtained results have demonstrated the efficiency of the developed controller in terms of accuracy (some tens of micrometers and few tenths of millidegree in the translation and rotation Cartesian space, respectively), convergence, and robustness (the controller has kept on working under different external disturbances). Also, the developed wavelet-based visual servoing at half resolution was compared to the photometry-based

approach as well as the popular IBVS method. The ground truth tests have proven that the wavelet method is at least as efficient with regards to both methods in nominal conditions and better in case of unfavorable ones (e.g., and under noise).

More broadly, this is the first work on 6 DoF wavelet-based visual servoing, and there remains much work to do. Moreover, enhancement of the computation time by using GPU hardware, or the combination of *compressed sensing* with the proposed approach must be foreseen. Also, it is interesting to investigate a method to compute the interaction matrix relatively to the time not to the desired pose which will increase sensitively the convergence domain of the wavelet-based visual servoing.

## REFERENCES

- [1] F. Chaumette and S. Hutchinson, "Visual servo control. I. basic approaches," *IEEE Rob. & Aut. Mag.*, vol. 13, no. 4, pp. 82–90, 2006.
- [2] M. Fanni, A. Khalifa, "A New 6-DOF Quadrotor Manipulation System: Design Kinematics Dynamics and Control" , *IEEE/ASME Trans. on Mech.*, vol. 22, pp. 1315-1326, 2017.
- [3] D. Koichiro, "A direct interpretation of dynamic images with camera and object motions for vision guided robot control," *Int. J. of Comp. Vis.*, vol. 37, no. 1, pp. 7–20, 2000.
- [4] V. Kallem et al., "Kernel-based visual servoing," *IEEE/RSJ IROS*, vol. 12, no. 5, pp. 1975–1980, 2007.
- [5] C. Collewet and E. Marchand, "Photometric visual servoing," *IEEE Trans. on Rob.*, vol. 27, no. 4, pp. 828–834, 2011.
- [6] B. Tamadazte, N. Piat, and E. Marchand, "A direct visual servoing scheme for automatic nanopositioning," *IEEE/ASME Trans. on Mech.*, vol. 17, no. 4, pp. 728–736, 2012.
- [7] G. Caron, E. Marchand, E.-M. Mouaddib, "Photometric visual servoing for omnidirectional cameras", *Autonomous Rob.*, Springer Verlag, 2013, 35 (2-3), pp.177-193.
- [8] C. Teuliere, E. Marchand, "A dense and direct approach to visual servoing using depth maps", *IEEE Trans. on Robotics*, 30(5):1242-1249, 2014.
- [9] E. Marchand and C. Collewet, "Using image gradient as a visual feature for visual servoing," in *IEEE/RSJ IROS*, 2010, pp. 5687–5692.
- [10] Q. Bateux, E. Marchand, "Histograms-based Visual Servoing. *IEEE Rob. and Auto. Let.*, 2(1):80-87, 2017.
- [11] N. Crombez, G. Caron and E. M. Mouaddib, "Photometric Gaussian mixtures based visual servoing," *IEEE/RSJ IROS*, Hamburg, 2015, pp. 5486-5491.
- [12] M. Bakthavatchalam, O. Tahri, F. Chaumette, "A Direct Dense Visual Servoing Approach using Photometric Moments", *IEEE Trans. on Rob.*, 2018, 34 (5), pp.1226-1239.
- [13] Q. Bateux, E. Marchand, J. Leitner, F. Chaumette, P. Corke, "Training Deep Neural Networks for Visual Servoing", In *IEEE ICRA*, pp. 3307-3314, 2018.
- [14] M. Ourak, B. Tamadazte, O. Lehmann, and N. Andreff, "Wavelet-based 6 dof visual servoing," *IEEE ICRA*, pp. 3414–3419, 2016.
- [15] A. Dame and E. Marchand, "Mutual Information-Based Visual Servoing," in *IEEE Trans. on Rob.*, vol. 27, no. 5, pp. 958-969, 2011.
- [16] Y. Meyer, "Wavelet -algorithms and applications," *Wavelet -Algorithms and Applications Soc. for Ind. and App. Math. Translation.*, 142 p., vol. 1, 1993.
- [17] I. Daubechies *et al.*, *Ten lectures on wavelet* , SIAM, 1992, vol. 61 ISBN. 978-0-89871-274-2.
- [18] M. Marcellin, *JPEG2000 Image Compression Fundamentals, Standards and Practice: Image Compression Fundamentals, Standards, and Practice*, Springer Science & Business Media, 2002, ISBN. 0-7923-7519-X.
- [19] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Trans. on Ima. Proc.*, vol. 12, no. 11, pp. 1338–1351, 2003.
- [20] C. He, Y. F. Zheng, and S. C. Ahalt, "Object tracking using the gabor wavelet transform and the golden section algorithm," *IEEE Trans. on Multimedia*, vol. 4, no. 4, pp. 528–538, 2002.
- [21] N. Marturi, B. Tamadazte, S. Dembélé and N. Piat, "Visual servoing schemes for automatic nanopositioning under scanning electron microscope." in *IEEE ICRA*, pp. 981-986, 2014.
- [22] V. Guelpa, G. J. Laurent, B. Tamadazte, et al., Single Frequency-based Visual Servoing for Microrobotics Applications, *IEEE/RSJ IROS*, pp. 4099-4104, 2016.
- [23] N. Marturi, B. Tamadazte, S. Dembélé and N. Piat, "Image-Guided Nanopositioning Scheme for SEM," in *IEEE Trans. on Auto. on Sci. and Eng.*, vol. 15, no. 1, pp. 45-56, 2018.
- [24] L. A. Dufлот, B. Tamadazte, A. Krupa, and N. Andreff, "Shearlet-based vs. Photometric-based Visual Servoing for Robot-assisted Medical Applications", *IEEE/RSJ IROS*, pp. 4099-4104, 2016.
- [25] L. A. Dufлот, A. Krupa, B. Tamadazte, and N. Andreff, Towards Ultrasound-based Visual Servoing using Shearlet Coefficients, *IEEE ICRA*, pp. 3420-3425, 2016.
- [26] L.-A. Dufлот, R. Reisenhofer, B. Tamadazte, N. Andreff, A. Krupa, "Wavelet and Shearlet-based Image Representations for Visual Servoing", *The Int. J. of Rob. Res.*, 2018, 0278364918769739.
- [27] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. on Pat. Anal. and Mach. Intel.*, vol. 11, no. 7, pp. 674–693, 1989.
- [28] Y. Xu, J. B. Weaver, D. M. Healy, and J. Lu, "Wavelet transform domain filters: a spatially selective noise filtration technique," *IEEE Trans. on Image Proc.*, vol. 3, no. 6, pp. 747-758, 1994.
- [29] L. E. Ramos-Velasco, "Visual servoing using discrete wavelet transform," in *World Auto. Cong. (WAC)*, 2012, pp. 1–6.
- [30] B. K. Horn and B. G. Schunck, "Determining optical flow," in *1981 Tech. Symp. East. Int. Soc. for Optics and Photonics*, 1981, pp. 319–331.
- [31] 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," 2016 *IEEE ICRA*, 2016, pp. 4186-4191.
- [32] C. Bernard, "Wavelet and ill-posed problems: optic flow estimation and scattered data interpolation," *Centre de Mathématiques Appliquées, École Polytechnique, France*, 1999.
- [33] M. Lourakis, "A brief description of the levenberg-marquardt algorithm implemented by levmar," *Foundation of Res. and Tech.*, pp. 1–6, 2005.



**Mouloud Ourak** holds a Ph.D. in Automation from the "Université de Franche-Comté" in 2016 and a MS degree in Robotics and Vision from the "Université de Montpellier" in 2013, in France. He is currently a temporary assistant-professor at the high school engineering "ENSMM" and temporary researcher at the AS2M department at the FEMTO-ST Institute (Besançon, France) where his research activities concern mainly image-guided surgery.



**Brahim Tamadazte** holds a Ph.D. in Automation and Computer Science from the "Université de Franche-Comté" in 2009 and a MS degree in Robotics and Intelligent Systems from the "Université Pierre et Marie Curie (Paris VI)" in 2005, in France. Since 2012, he is with the FEMTO-ST Institute in the AS2M Dep. as Senior Scientist CNRS researcher, FR). Currently, he works in surgical microrobotics surgery (conception and control) and visual servoing using OCT imaging systems.



**Nicolas Andreff** received the Engineer degree (M.S. degree) in computer sciences and applied mathematics from ENSEEIHT, Toulouse, France, in 1994, and the Ph.D. degree in computer graphics, computer vision, and robotics from INPG, Grenoble, FR, in 1999. He is currently Professor at the FEMTO-ST Institute, Univ. de Franche-Comté, FR. He created the Biomedical Micro-/Nano-Robotics group of the AS2M Dep. in 2012. His research interests range from vision-based control and parallel kinematics to intracorporeal microrobotics.