



HAL
open science

Image moments: a general and useful set of features for visual servoing

François Chaumette

► **To cite this version:**

François Chaumette. Image moments: a general and useful set of features for visual servoing. IEEE Transactions on Robotics, 2004, 20 (4), pp.713-723. inria-00352019

HAL Id: inria-00352019

<https://inria.hal.science/inria-00352019v1>

Submitted on 12 Jan 2009

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

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

Image Moments: A General and Useful Set of Features for Visual Servoing

François Chaumette, *Member, IEEE*

Abstract—In this paper, we determine the analytical form of the interaction matrix related to any moment that can be computed from segmented images. The derivation method we present is based on Green's theorem. We apply this general result to classical geometrical primitives. We then consider using moments in image-based visual servoing. For that, we select six combinations of moments to control the six degrees of freedom of the system. These features are particularly adequate, if we consider a planar object and the configurations such that the object and camera planes are parallel at the desired position. The experimental results we present show that a correct behavior of the system is obtained if we consider either a simple symmetrical object or a planar object with complex and unknown shape.

Index Terms—Image moments, visual servoing.

I. INTRODUCTION

IMAGE MOMENTS have been widely used in computer vision for a very long time, especially for pattern-recognition applications [11], [15], [17]. It would be interesting to use them in visual servoing since they provide a generic representation of any object, with simple or complex shapes, that can be segmented in an image. They also provide a more geometric and intuitive meaning than the features that have been proposed in [5] from the contour of an object. Attempts to use moments in image-based visual servoing have already been presented in the past. The problem was that the analytical form of the interaction matrix related to image moments was not available. However, this matrix is essential to design a visual-servoing control scheme [8], [12]. That is why, in [1], if the idea of using image moments was expressed rigorously, very coarse approximations were performed in practice to control only four degrees of freedom (DOFs) of a robot using the area, the centroid, and the main orientation of an object in the image. Similarly, in [24], a neural network has been developed to numerically estimate this interaction matrix. Finally, in [2], the interaction matrix related to the moments of order less than three has been determined, but the result was only valid for an ellipse in the image. The first contribution of this paper is that it presents a method to determine the analytical form of the interaction matrix related to any image moment.

A second objective of using moments in image-based visual servoing is to try to determine features that avoid the potential

problems that may appear when redundant image points coordinates are used; local minimum and coupled features that lead to inadequate robot trajectories, for example [3]. In fact, we would like to find again the nice properties of 2 1/2D visual servoing [14], but using visual features extracted only from the current image (to avoid any partial pose estimation at each iteration of the control scheme, which introduces some sensitivity to image noise). A nice step in that direction has been recently presented in [6]. More precisely, we search for six independent visual features, such that the corresponding interaction matrix has a maximal decoupled structure, without any singularity, and such that its condition number is as low as possible (to improve the robustness and the numerical stability of the system [9], [16], [19]). The second contribution of this paper is concerned with this objective.

In the next section, we present a method based on Green's theorem to determine the analytical form of the interaction matrix related to any moment. The most simple case of a planar object is described in text, while the more general case is considered in the Appendix. In Section II-B, we consider more particularly the area and the coordinates of the center of gravity of the object, which are the moments of lowest order zero and one. We then determine the interaction matrix of the centered moments, from which we deduce the one related to the main orientation of the object. In Section III, all of these general results that are valid for objects of any shape are applied to the most simple and usual geometrical primitives: segments, straight lines, circles, spheres, and cylinders. Indeed, the method we propose allows revisiting, in an easy way, the previous works dealing with the modeling of the interaction matrix. After these modeling issues, we propose in Section IV six visual features to control the six DOFs of the system. We will see that the selected features allows the system to have a nice decoupling behavior, if we consider a planar object and the most usual case where the object and camera planes are parallel at the desired position. The experimental results described in Section V validate the proposed choice. Finally, concluding remarks are given in Section VI.

II. INTERACTION MATRIX OF IMAGE MOMENTS

A. Modeling

Let \mathcal{O} be the observed object and $\pi(t)$ the image acquired by the camera at time t . We denote $\mathcal{R}(t)$ the part of $\pi(t)$ where the object projects, and $\mathcal{C}(t)$ the contour of $\mathcal{R}(t)$. As explained at the end of this section, we do not consider the intensity level of each pixel, which means that either binary images are acquired or a spatial segmentation algorithm, providing binary images, is first

Manuscript received December 10, 2002; revised July 25, 2003. This paper was recommended by Associate Editor Y.-H. Liu and Editor S. Hutchinson upon evaluation of the reviewers' comments.

The author is with IRISA/INRIA Rennes, 35 042 Rennes Cedex, France (e-mail: Francois.Chaumette@irisa.fr).

Digital Object Identifier 10.1109/TRO.2004.829463

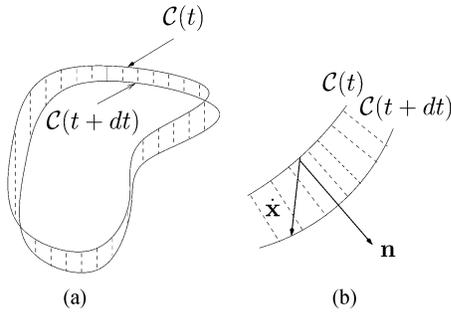


Fig. 1. (a) Time variation of contour $\mathcal{C}(t)$. (b) A more detailed view.

performed on the acquired images. In that case, the moments m_{ij} of \mathcal{O} in the image are defined by

$$m_{ij}(t) = \iint_{\mathcal{R}(t)} f(x, y) dx dy \quad (1)$$

where $f(x, y) = x^i y^j$. We are interested in determining the analytical form describing the time variation \dot{m}_{ij} of moment m_{ij} in the function of the relative kinematic screw $\mathbf{v} = (\mathbf{v}, \boldsymbol{\omega})$ between the camera and the object, where \mathbf{v} and $\boldsymbol{\omega}$ represent the translational and rotational velocity components, respectively. As for classical geometrical features, we will obtain a linear link that can be expressed in the form [8]

$$\dot{m}_{ij} = \mathbf{L}_{m_{ij}} \mathbf{v} \quad (2)$$

where $\mathbf{L}_{m_{ij}}$ is the interaction matrix related to m_{ij} .

In (1), the only part that is a function of time t is $\mathcal{R}(t)$. The time variation of m_{ij} can thus be obtained from the variation of $\mathcal{C}(t)$. More precisely, we have [10]

$$\dot{m}_{ij} = \oint_{\mathcal{C}(t)} f(x, y) \dot{\mathbf{x}}^T \mathbf{n} dl \quad (3)$$

where $\dot{\mathbf{x}}$ is the velocity of contour point $\mathbf{x} = (x, y)$, \mathbf{n} is the unitary vector normal to $\mathcal{C}(t)$ at point \mathbf{x} , and dl is an infinitesimal element of contour $\mathcal{C}(t)$. From a geometrical point of view (see Fig. 1), the variation of m_{ij} is given by computing m_{ij} on the infinitesimal area between $\mathcal{C}(t)$ and $\mathcal{C}(t+dt)$, which is nothing but integrating along $\mathcal{C}(t)$ the product of f (to compute the moment) by the scalar product between $\dot{\mathbf{x}}$ and \mathbf{n} (so that $\mathcal{C}(t)$ reaches $\mathcal{C}(t+dt)$).

If the following conditions are satisfied (which is the case in practice):

- 1) $\mathcal{C}(t)$ is continuous by parts;
- 2) vector $f(x, y) \dot{\mathbf{x}}$ is tangent to $\mathcal{R}(t)$ and continuously differentiable, $\forall \mathbf{x} \in \mathcal{R}(t)$;

we can use the famous Green's theorem [21] and write (3) in the form

$$\dot{m}_{ij} = \iint_{\mathcal{R}(t)} \text{div}[f(x, y) \dot{\mathbf{x}}] dx dy. \quad (4)$$

By developing (4), we finally obtain

$$\dot{m}_{ij} = \iint_{\mathcal{R}} \left[\frac{\partial f}{\partial x} \dot{x} + \frac{\partial f}{\partial y} \dot{y} + f(x, y) \left(\frac{\partial \dot{x}}{\partial x} + \frac{\partial \dot{y}}{\partial y} \right) \right] dx dy. \quad (5)$$

This derivation, usual in physics and mechanics, has also already been used in computer vision. In [4], low-order moments have been considered to estimate the time-to-collision and the surface orientation of an object observed by a moving camera.

However, an affine-motion model was used in the derivation, which was not able to handle the perspective projection model of a camera exactly. The same application has been considered in [7], but using a spherical projection model. Finally, in [18], change in image moments under Lie group transformations has been studied, with a special attention to affine motion. In the following derivations, we consider a full and complete perspective projection model, as well as the six DOFs of the system. Furthermore, moments of any order are addressed in a general way.

In (5), the terms \dot{x} , \dot{y} , $(\partial \dot{x} / \partial x)$, and $(\partial \dot{y} / \partial y)$ can be linearly expressed to the kinematic screw \mathbf{v} . Indeed, for any point with coordinates $\mathbf{x} = (x, y)$ in the image whose corresponding three-dimensional (3-D) point has depth Z , we have the well-known relation

$$\dot{\mathbf{x}} = \mathbf{L}_{\mathbf{x}} \mathbf{v} \quad (6)$$

where

$$\mathbf{L}_{\mathbf{x}} = \begin{bmatrix} -1/Z & 0 & x/Z & xy & -1-x^2 & y \\ 0 & -1/Z & y/Z & 1+y^2 & -xy & -x \end{bmatrix}. \quad (7)$$

At this step, we consider that the 3-D observed object belongs to a continuous surface, which means that we assume that it does not present any depth discontinuity. In that case, the depth Z of any 3-D object point can be expressed as a continuous function of its image coordinates x and y , as follows:

$$\frac{1}{Z} = \sum_{p \geq 0, q \geq 0} A_{pq} x^p y^q. \quad (8)$$

For instance, if the object is planar or has a planar limb surface (see [8] for more details), whose equation expressed in the camera frame is given by

$$Z = \gamma_1 X + \gamma_2 Y + Z_0$$

then we have

$$\frac{1}{Z} = Ax + By + C \quad (9)$$

where $A = -\gamma_1/Z_0$, $B = -\gamma_2/Z_0$, and $C = 1/Z_0$.

We give in the Appendix the analytical form of the interaction matrix for the general case (8). However, in the following, we will only consider planar objects. Equation (9) will thus be used, instead of (8), to simplify the mathematical developments. Furthermore, even if the general form is available, the exact analytical form we obtain for planar objects can be seen as an approximation for objects with more complex 3-D surfaces. Such approximation is generally sufficiently accurate in practice, thanks to the robustness of the visual-servoing scheme with respect to modeling errors.

Using (9) in (7), (6) can finally be written as

$$\begin{cases} \dot{x} = -(Ax + By + C) \cdot v_x + x(Ax + By + C) \cdot v_z \\ \quad + xy \cdot \omega_x - (1 + x^2) \cdot \omega_y + y \cdot \omega_z \\ \dot{y} = -(Ax + By + C) \cdot v_y + y(Ax + By + C) \cdot v_z \\ \quad + (1 + y^2) \cdot \omega_x - xy \cdot \omega_y - x \cdot \omega_z \end{cases} \quad (10)$$

from which we deduce

$$\begin{cases} \frac{\partial \dot{x}}{\partial x} = -Av_x + (2Ax + By + C)v_z + y\omega_x - 2x\omega_y \\ \frac{\partial \dot{y}}{\partial y} = -Bv_y + (Ax + 2By + C)v_z + 2y\omega_x - x\omega_y \end{cases}. \quad (11)$$

Substituting (10) and (11) into (5), and remembering that $f(x, y) = x^i y^j$, $(\partial f / \partial x) = i x^{i-1} y^j$, and $(\partial f / \partial y) = j x^i y^{j-1}$, we can express (5) under the expected form (2). We obtain, after simple developments

$$\mathbf{L}_{m_{ij}} = [m_{vx} \ m_{vy} \ m_{vz} \ m_{wx} \ m_{wy} \ m_{wz}] \quad (12)$$

where

$$\begin{cases} m_{vx} = -i(Am_{ij} + Bm_{i-1,j+1} + Cm_{i-1,j}) - Am_{ij} \\ m_{vy} = -j(Am_{i+1,j-1} + Bm_{ij} + Cm_{i,j-1}) - Bm_{ij} \\ m_{vz} = (i+j+3)(Am_{i+1,j} + Bm_{i,j+1} + Cm_{ij}) - Cm_{ij} \\ m_{wx} = (i+j+3)m_{i,j+1} + jm_{i,j-1} \\ m_{wy} = -(i+j+3)m_{i+1,j} - im_{i-1,j} \\ m_{wz} = im_{i-1,j+1} - jm_{i+1,j-1}. \end{cases}$$

The time variation of a moment of order $i+j$ can thus be expressed from the moments of order up to $i+j+1$ and, as far as the translational terms are concerned, from the 3-D parameters A , B , and C . This nice property will allow us to easily use image moments in image-based visual servoing.

In fact, all the derivations presented above are valid for the more general definition of image-based moments, that is

$$m_{ij} = \int \int_{\pi} I(x, y) x^i y^j dx dy$$

where $I(x, y)$ is the intensity level of image point with coordinates x and y , and where the integration is performed on all the image plane π . The assumption that the intensity level of a point is not modified during any rigid motion has to be done (i.e., $(\partial I(x, y)) / (\partial t) = 0$). This assumption is usual in computer vision in almost all works dealing with image-motion analysis. It seems possible to use the previous theoretical results on any image (and not only on binary or segmented ones). The problem that occurs in practice lies in the parts of the image that appear or disappear from one image to the next one, which makes the moments non-differentiable on all of the image plane. A region \mathcal{R} or its contour \mathcal{C} (or a set of regions or contours) has to be extracted and tracked to provide a global matching, ensuring that the moments are well defined and differentiable. We can note that a local matching, linking any physical point of the contour to its corresponding point in the next image, is not necessary at all. This is a classical interest of using moments in computer vision. Finally, it may be interesting, in practice, to take into account a normalized intensity level of points in the tracked region for the computation of moments, in order to increase the robustness and stability of the computed values with respect to segmentation noise.

B. Zeroth- and First-Order Moments

From the general form given in (12), we can deduce the form for any particular moment by just selecting the value of i and j . We first consider the most simple case $i = j = 0$ that will give us the interaction matrix related to the area $a = m_{00}$ of the object. We obtain directly from (12)

$$\mathbf{L}_a = [-aA \ -aB \ a(3/Z_g - C) \ 3ay_g \ -3ax_g \ 0] \quad (13)$$

where $1/Z_g = Ax_g + By_g + C$. We can note that \dot{a} depends linearly on a (we have, of course, $\dot{a} = 0$, if $a = 0$). Furthermore, when the object is centered and parallel to the image plane ($A = B = x_g = y_g = 0$), we have $\dot{a} = 0$ for all camera motions but the expected translation v_z along the optical axis.

Similarly, by considering the first-order moments m_{10} and m_{01} , it is easy to determine the interaction matrix related to the coordinates $x_g (= m_{10}/m_{00})$ and $y_g (= m_{01}/m_{00})$ of the center of gravity of an object in the image. We obtain

$$\begin{aligned} \mathbf{L}_{x_g} &= [-1/Z_g \ 0 \ x_{g_{vz}} \ x_{g_{wx}} \ x_{g_{wy}} \ y_g] \\ \mathbf{L}_{y_g} &= [0 \ -1/Z_g \ y_{g_{vz}} \ y_{g_{wx}} \ y_{g_{wy}} \ -x_g] \end{aligned} \quad (14)$$

where

$$\begin{cases} x_{g_{vz}} = x_g/Z_g + 4(An_{20} + Bn_{11}) \\ y_{g_{vz}} = y_g/Z_g + 4(An_{11} + Bn_{02}) \\ x_{g_{wx}} = -y_{g_{wy}} = x_g y_g + 4n_{11} \\ x_{g_{wy}} = -(1 + x_g^2 + 4n_{20}) \\ y_{g_{wx}} = 1 + y_g^2 + 4n_{02} \end{cases}$$

where n_{20} , n_{02} , and n_{11} are the normalized centered moments of order 2, defined as follows:

$$n_{ij} = \mu_{ij}/a \text{ with } \begin{cases} \mu_{20} = m_{20} - ax_g^2 \\ \mu_{02} = m_{02} - ay_g^2 \\ \mu_{11} = m_{11} - ax_g y_g. \end{cases} \quad (15)$$

From this result, we can see that the interaction matrix related to the coordinates of the center of gravity of the object is a generalization of the interaction matrix (7) related to the coordinates of a point. Indeed, for a physical point, we have $n_{20} = n_{11} = n_{02} = 0$, and we can set $A = B = 0$ in (14) to exactly obtain the classical equation (7).

From another point of view, all previous works in visual servoing that used the coordinates of the center of gravity of an object as visual features (for instance, [6]), in fact used with (7) an approximation of the real interaction matrix. However, it was not a crucial problem, since closed-loop visual-servoing schemes are well known to be robust with respect to such modeling errors.

Finally, it is well known and clear from (14) that x_g and y_g are mainly related to translational motion v_x , along the x camera axis and rotational motion ω_y around the y axis (for x_g), and to translational motion v_y and rotational motion ω_x (for y_g).

C. Centered Moments

We now consider the centered moments μ_{ij} , defined by

$$\mu_{ij} = \int \int_{\mathcal{R}} (x - x_g)^i (y - y_g)^j dx dy. \quad (16)$$

Knowing the relation between μ_{ij} and the moments m_{kl} of order up to $i+j$, given as follows:

$$\mu_{ij} = \sum_{k=0}^i \sum_{l=0}^j \binom{i}{k} \binom{j}{l} (-x_g)^{i-k} (-y_g)^{j-l} m_{kl} \quad (17)$$

we can determine the interaction matrix related to μ_{ij} , using (12). We can even obtain the analytical form of this matrix expressed with the centered moments by using the reciprocal relation to (17) that links any m_{kl} to the centered moments up to order $k+l$

$$m_{kl} = \sum_{m=0}^k \sum_{n=0}^l \binom{k}{m} \binom{l}{n} x_g^{k-m} y_g^{l-n} \mu_{mn}. \quad (18)$$

After some (tedious) developments, we obtain the following simple form:

$$\mathbf{L}_{\mu_{ij}} = [\mu_{vx} \ \mu_{vy} \ \mu_{vz} \ \mu_{wx} \ \mu_{wy} \ \mu_{wz}] \quad (19)$$

where

$$\begin{cases} \mu_{vx} = -(i+1)A\mu_{ij} - iB\mu_{i-1,j+1} \\ \mu_{vy} = -jA\mu_{i+1,j-1} - (j+1)B\mu_{ij} \\ \mu_{vz} = -A\mu_{wy} + B\mu_{wx} + (i+j+2)C\mu_{ij} \\ \mu_{wx} = (i+j+3)\mu_{i,j+1} + ix_g\mu_{i-1,j+1} \\ \quad + (i+2j+3)y_g\mu_{ij} - 4in_{11}\mu_{i-1,j} - 4jn_{02}\mu_{i,j-1} \\ \mu_{wy} = -(i+j+3)\mu_{i+1,j} - (2i+j+3)x_g\mu_{ij} \\ \quad - jy_g\mu_{i+1,j-1} + 4in_{20}\mu_{i-1,j} + 4jn_{11}\mu_{i,j-1} \\ \mu_{wz} = i\mu_{i-1,j+1} - j\mu_{i+1,j-1}. \end{cases}$$

This result can also be obtained by applying Green's theorem directly to the definition of the centered moments (16). Indeed, we thus obtain

$$\begin{aligned} \mu_{ij} = \int \int_{\mathcal{R}} & \left[i(x-x_g)^{i-1}(y-y_g)^j(\dot{x}-x_g) \right. \\ & + j(x-x_g)^i(y-y_g)^{j-1}(\dot{y}-y_g) \\ & \left. + (x-x_g)^i(y-y_g)^j \left(\frac{\partial \dot{x}}{\partial x} + \frac{\partial \dot{y}}{\partial y} \right) \right] dx dy \quad (20) \end{aligned}$$

from which, using (10) and (14), we find again (19).

From the analytical form of μ_{vx} and μ_{vy} , we can note that all centered moments are invariant with respect to translational motions parallel to the image plane only if the image plane is parallel to the object ($\mu_{vx} = \mu_{vy} = 0$ when $A = B = 0$).

If we consider the centered moments of order 2, that is, μ_{20} , μ_{02} , and μ_{11} , we obtain from (19)

$$\begin{aligned} \mathbf{L}_{\mu_{20}} &= [\mu_{20vx} \quad -B\mu_{20} \quad u_{20} \quad t_{20} \quad s_{20} \quad 2\mu_{11}] \\ \mathbf{L}_{\mu_{02}} &= [-A\mu_{02} \quad \mu_{02vy} \quad u_{02} \quad t_{02} \quad s_{02} \quad -2\mu_{11}] \\ \mathbf{L}_{\mu_{11}} &= [\mu_{11vx} \quad \mu_{11vy} \quad u_{11} \quad t_{11} \quad s_{11} \quad \mu_{02} - \mu_{20}] \quad (21) \end{aligned}$$

with

$$\begin{cases} \mu_{20vx} = -3A\mu_{20} - 2B\mu_{11} \\ \mu_{02vy} = -2A\mu_{11} - 3B\mu_{02} \\ \mu_{11vx} = -2A\mu_{11} - B\mu_{02} \\ \mu_{11vy} = -2B\mu_{11} - A\mu_{20} \\ u_{ij} = -As_{ij} + Bt_{ij} + 4C\mu_{ij} \\ s_{20} = -7x_g\mu_{20} - 5\mu_{30} \\ t_{20} = 5(y_g\mu_{20} + \mu_{21}) + 2x_g\mu_{11} \\ s_{02} = -5(x_g\mu_{02} + \mu_{12}) - 2y_g\mu_{11} \\ t_{02} = 7y_g\mu_{02} + 5\mu_{03} \\ s_{11} = -6x_g\mu_{11} - 5\mu_{21} - y_g\mu_{20} \\ t_{11} = 6y_g\mu_{11} + 5\mu_{12} + x_g\mu_{02} \end{cases}$$

where centered moments of order 3 are given by [see (17)]

$$\begin{cases} \mu_{30} = m_{30} - 3x_gm_{20} + 2ax_g^3 \\ \mu_{21} = m_{21} - 2x_gm_{11} - y_gm_{20} + 2ax_g^2y_g \\ \mu_{12} = m_{12} - 2y_gm_{11} - x_gm_{02} + 2ax_gy_g^2 \\ \mu_{03} = m_{03} - 3y_gm_{02} + 2ay_g^3. \end{cases}$$

In visual servoing, we could hope that, using as visual features \mathbf{s} the six moments of order less than 3 (that is, $\mathbf{s} = (a, x_g, y_g, \mu_{20}, \mu_{11}, \mu_{02})$), the six DOF of the system could be controlled. However, the interaction matrix \mathbf{L}_s related to \mathbf{s} is always of maximal rank 5 when an object appears with a symmetrical shape in the image, such as ellipses, squares, or rectangles. Furthermore, for nonsymmetrical objects, even if \mathbf{L}_s is of full rank 6, its condition number is really too high to provide satisfactory results (the control scheme is unstable). As will be shown in Section IV, few moments of order at least

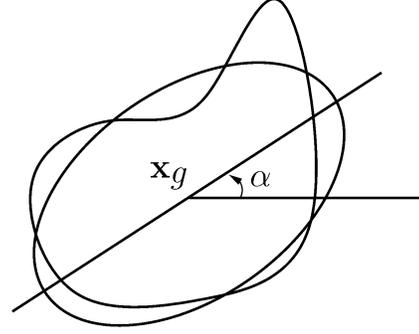


Fig. 2. Orientation α of an object is defined as the orientation of the ellipse obtained using the value of the object moments of order less than 3.

equal to 3 thus have to be involved in the selection of the visual features.

Despite this, an interesting feature can be selected from the moments of order 2. It is the well-known object orientation α defined by (see [15] and Fig. 2)

$$\alpha = \frac{1}{2} \arctan \left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \right). \quad (22)$$

Since $(d \arctan(x))/(dt) = (\dot{x})/(1+x^2)$, we have

$$\mathbf{L}_\alpha = ((\mu_{20} - \mu_{02})\mathbf{L}_{\mu_{11}} - \mu_{11}(\mathbf{L}_{\mu_{20}} - \mathbf{L}_{\mu_{02}}))/\Delta$$

where $\Delta = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2$. Using (21), we obtain

$$\mathbf{L}_\alpha = [\alpha_{vx} \quad \alpha_{vy} \quad \alpha_{vz} \quad \alpha_{wx} \quad \alpha_{wy} \quad -1] \quad (23)$$

with

$$\begin{cases} \alpha_{vx} = aA + bB \\ \alpha_{vy} = -cA - aB \\ \alpha_{vz} = -A\alpha_{wy} + B\alpha_{wx} \\ \alpha_{wx} = -bx_g + ay_g + d \\ \alpha_{wy} = ax_g - cy_g + e \\ a = \mu_{11}(\mu_{20} + \mu_{02})/\Delta \\ b = [2\mu_{11}^2 + \mu_{02}(\mu_{02} - \mu_{20})]/\Delta \\ c = [2\mu_{11}^2 + \mu_{20}(\mu_{20} - \mu_{02})]/\Delta \\ d = 5[\mu_{12}(\mu_{20} - \mu_{02}) + \mu_{11}(\mu_{03} - \mu_{21})]/\Delta \\ e = 5[\mu_{21}(\mu_{02} - \mu_{20}) + \mu_{11}(\mu_{30} - \mu_{12})]/\Delta. \end{cases}$$

We can note that α is invariant with respect to any translational motion when the object plane is parallel to the image plane ($\alpha_{vx} = \alpha_{vy} = \alpha_{vz} = 0$ when $A = B = 0$). Of course, we can also note the direct link between the variation of α and the rotational motions ω_z around the optical axis.

From the previous results, we have exhibited four visual features that have a direct and intuitive link with particular 3-D motions. They are the area a , the coordinates x_g and y_g of the center of gravity of the object, and the orientation α . We will see in Section IV how to select two supplementary visual features when we are interested in controlling the six DOFs of the system. In the next section, we consider very particular and simple cases.

III. FROM MOMENTS TO SIMPLE GEOMETRICAL PRIMITIVES

We have already seen in Section II-B that a point can be retrieved from the center of gravity of an object whose area is null. In this section, we will see that the interaction matrix related to simple geometrical primitives can be obtained easily from

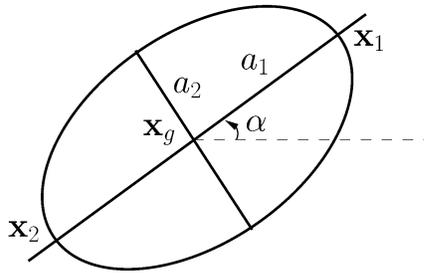


Fig. 3. Usual representation of an ellipse.

the general previous results. We first consider ellipses, and then segments and straight lines. These primitives have already been studied in the past [2], [8]. We here revisit the obtained results with a completely different and systematic approach.

A. Ellipses

The equation of an ellipse is given by (see Fig. 3)

$$\frac{(x - x_g + t(y - y_g))^2}{a_1^2(1 + t^2)} + \frac{(y - y_g - t(x - x_g))^2}{a_2^2(1 + t^2)} - 1 = 0$$

where x_g and y_g are the coordinates of the center of gravity of the ellipse, a_1 and a_2 are its major and minor axes, respectively, and t is directly linked to its orientation (more precisely, we have $t = \tan \alpha$). This representation is degenerated when the ellipse becomes a circle (in that case, t is no longer defined). That is why it is better to represent an ellipse with the coordinates of its center of gravity and the normalized centered moments of order 2 defined in (15). These moments can easily be expressed from the parameters a_1 , a_2 , and t . Indeed, we have

$$\begin{cases} n_{20} = (a_1^2 + a_2^2 t^2) / 4(1 + t^2) \\ n_{02} = (a_1^2 t^2 + a_2^2) / 4(1 + t^2) \\ n_{11} = t (a_1^2 - a_2^2) / 4(1 + t^2) \end{cases} \quad (24)$$

from which we can determine the reciprocal relations

$$\begin{cases} a_1^2 = 2 \left(n_{02} + n_{20} + \sqrt{(n_{20} - n_{02})^2 + 4n_{11}^2} \right) \\ a_2^2 = 2 \left(n_{02} + n_{20} - \sqrt{(n_{20} - n_{02})^2 + 4n_{11}^2} \right) \\ t = \left(n_{02} - n_{20} + \sqrt{(n_{20} - n_{02})^2 + 4n_{11}^2} \right) / 2n_{11}. \end{cases}$$

From this equation, the link between the area a of the ellipse and the normalized centered moments is easily obtained as follows:

$$a = \pi a_1 a_2 = 4\pi \sqrt{n_{20} n_{02} - n_{11}^2}.$$

Furthermore, the equation of an ellipse can be expressed using the minimal, complete, and nonambiguous representation $\mathbf{p} = (x_g, y_g, n_{20}, n_{02}, n_{11})$ in the following form:

$$\begin{aligned} g(\mathbf{x}, \mathbf{p}) &= n_{02}x^2 + n_{20}y^2 - 2n_{11}xy + 2(n_{11}y_g - n_{02}x_g)x \\ &\quad + 2(n_{11}x_g - n_{20}y_g)y + n_{02}x_g^2 + n_{20}y_g^2 \\ &\quad - 2n_{11}x_g y_g + 4n_{11}^2 - 4n_{20}n_{02} = 0. \end{aligned} \quad (25)$$

The interaction matrix related to the coordinates of the center of gravity of the ellipse are given in (14). As for the normalized centered moments, we obtain from their definition (15)

$$\mathbf{L}_{n_{ij}} = \frac{1}{a} \mathbf{L}_{\mu_{ij}} - \frac{n_{ij}}{a} \mathbf{L}_a.$$

Using (21) and (13) and knowing that the centered moments of order 3 are all equal to zero in the case of an ellipse (i.e., $\mu_{30} = \mu_{21} = \mu_{12} = \mu_{03} = 0$), we finally obtain

$$\begin{aligned} \mathbf{L}_{n_{20}} &= [n_{20_{vx}} \quad 0 \quad n_{20_{vz}} \quad n_{20_{wx}} \quad n_{20_{wy}} \quad 2n_{11}] \\ \mathbf{L}_{n_{02}} &= [0 \quad n_{02_{vy}} \quad n_{02_{vz}} \quad n_{02_{wx}} \quad n_{02_{wy}} \quad -2n_{11}] \\ \mathbf{L}_{n_{11}} &= [n_{11_{vx}} \quad n_{11_{vy}} \quad n_{11_{vz}} \quad n_{11_{wx}} \quad n_{11_{wy}} \quad n_{02} - n_{20}] \end{aligned} \quad (26)$$

where

$$\begin{cases} n_{20_{vx}} = -2(An_{20} + Bn_{11}) \\ n_{02_{vy}} = -2(An_{11} + Bn_{02}) \\ n_{11_{vx}} = -An_{11} - Bn_{02} \\ n_{11_{vy}} = -An_{20} - Bn_{11} \\ n_{ij_{vz}} = -An_{ij_{wy}} + Bn_{ij_{wx}} + 2Cn_{ij} \\ n_{20_{wx}} = 2(y_g n_{20} + x_g n_{11}) \\ n_{20_{wy}} = -4x_g n_{20} \\ n_{02_{wx}} = 4y_g n_{02} \\ n_{02_{wy}} = -2(y_g n_{11} + x_g n_{02}) \\ n_{11_{wx}} = 3y_g n_{11} + x_g n_{02} \\ n_{11_{wy}} = -y_g n_{20} - 3x_g n_{11}. \end{cases}$$

We find again the analytical form presented in [2]. We recall that, if the ellipse in the image results from the projection of a 3-D circle, the 3-D parameters A , B , and C are directly obtained from the parameters of the plane which the circle belongs to [see (9)]. In that case, the interaction matrix $\mathbf{L}_{\mathbf{p}}$ related to the five parameters \mathbf{p} is always of full rank 5, but in the case where the 3-D circle is parallel to the image plane ($A = B = 0$) and centered. In that particular case, the ellipse becomes a centered circle (such that $n_{20} = n_{02}$ and $x_g = y_g = n_{11} = 0$) and the rank of $\mathbf{L}_{\mathbf{p}}$ is only 3 (since $\mathbf{L}_{n_{20}} = \mathbf{L}_{n_{02}}$ and $\mathbf{L}_{n_{11}} = \mathbf{0}_{1 \times 6}$).

In the case where the ellipse in the image results from the projection of a sphere, we have [2]

$$\begin{cases} A = X_0 / (X_0^2 + Y_0^2 + Z_0^2 - R^2) \\ B = Y_0 / (X_0^2 + Y_0^2 + Z_0^2 - R^2) \\ C = Z_0 / (X_0^2 + Y_0^2 + Z_0^2 - R^2) \end{cases}$$

where X_0 , Y_0 , and Z_0 are the coordinates of the center of the sphere and R is its radius. In that case, the interaction matrix related to \mathbf{p} is always of rank 3, which means that only three parameters are necessary and sufficient to represent the image of a sphere. This result is intuitive, since a sphere is represented by only four independent parameters. In order to obtain a minimal and robust representation, we recommend simply using the area and the coordinates of the center of gravity of the ellipse when a sphere is considered. Indeed, in that case, the related interaction matrix is always of full rank 3.

B. Segments

A segment of length l can be seen as an ellipse with major axis a_1 equal to $l/2$ and minor axis a_2 equal to 0 (see Fig. 3). Of course, the middle of the segment corresponds exactly to the

center of gravity of that particular ellipse, and similarly, the orientation of the segment is nothing but the orientation of the ellipse. We deduce from (24) the value of the normalized centered moments in the case of a segment, as follows:

$$\begin{cases} n_{20} = a_1^2/4(1+t^2) = l^2 \cos^2 \alpha/16 \\ n_{02} = a_1^2 t^2/4(1+t^2) = l^2 \sin^2 \alpha/16 \\ n_{11} = t a_1^2/4(1+t^2) = l^2 \sin \alpha \cos \alpha/16 \end{cases} \quad (27)$$

from which we deduce $l = 4\sqrt{n_{20} + n_{02}}$. We can thus determine the interaction matrix related to l from $\mathbf{L}_{n_{20}}$ and $\mathbf{L}_{n_{02}}$. More precisely, we have

$$\mathbf{L}_l = \frac{8}{l} (\mathbf{L}_{n_{20}} + \mathbf{L}_{n_{02}}). \quad (28)$$

Using (27) in (14), (23), and (28), we finally obtain a simple form for the interaction matrix related to the four parameters of a segment, as follows:

$$\begin{aligned} \mathbf{L}_{x_g} &= [-1/Z_g \quad 0 \quad x_{g_{vz}} \quad x_{g_{wx}} \quad x_{g_{wy}} \quad y_g] \\ \mathbf{L}_{y_g} &= [0 \quad -1/Z_g \quad y_{g_{vz}} \quad y_{g_{wx}} \quad y_{g_{wy}} \quad -x_g] \\ \mathbf{L}_l &= [-Dc \quad -Ds \quad l_{vz} \quad l_{wx} \quad l_{wy} \quad 0] \\ \mathbf{L}_\alpha &= [Ds/l \quad -Dc/l \quad \alpha_{vz} \quad \alpha_{wx} \quad \alpha_{wy} \quad -1] \end{aligned} \quad (29)$$

where $c = \cos \alpha$, $s = \sin \alpha$, $D = 1/Z_1 - 1/Z_2 = Alc + Bls$, and $Z_g = 2Z_1 Z_2 / (Z_1 + Z_2) = 1/(Ax_g + By_g + C)$, Z_1 and Z_2 are the depth of the end points \mathbf{x}_1 and \mathbf{x}_2 of the segment, and Z_g is the depth of the middle point \mathbf{x}_g of the segment. Furthermore, we have

$$\begin{cases} x_{g_{vz}} = x_g/Z_g + Dlc/4 \\ y_{g_{vz}} = y_g/Z_g + Dls/4 \\ x_{g_{wx}} = -y_{g_{wy}} = x_g y_g + l^2 sc/4 \\ x_{g_{wy}} = -(1 + x_g^2 + l^2 c^2/4) \\ y_{g_{wx}} = 1 + y_g^2 + l^2 s^2/4 \\ l_{vz} = l/Z_g + D(x_g c + y_g s) \\ l_{wx} = l(y_g(1 + s^2) + x_g cs) \\ l_{wy} = -l(x_g(1 + c^2) + y_g cs) \\ \alpha_{vz} = D(cy_g - sx_g)/l \\ \alpha_{wx} = s(cy_g - sx_g) \\ \alpha_{wy} = -c(cy_g - sx_g). \end{cases}$$

As expected, we find again the analytical form given in [8] for a segment, and we can note that the interaction matrix given by (29) is always of full rank 4, and has a quite nice decoupling form when the segment is parallel to the image plane (in which case, $D = 0$).

C. Straight Lines

A straight line in the image can be seen as a segment of infinite length and whose center of gravity belongs to the straight line (see Fig. 4). Using the (ρ, θ) representation that defines the configuration of the straight line by

$$x \cos \theta + y \sin \theta - \rho = 0$$

we have

$$\theta = \alpha + \frac{\pi}{2} \quad (30)$$

$$\rho = x_g \cos \theta + y_g \sin \theta. \quad (31)$$

The coordinates x_g and y_g have the following form:

$$\begin{cases} x_g = \rho \cos \theta + \lambda \sin \theta \\ y_g = \rho \sin \theta - \lambda \cos \theta \end{cases} \quad (32)$$

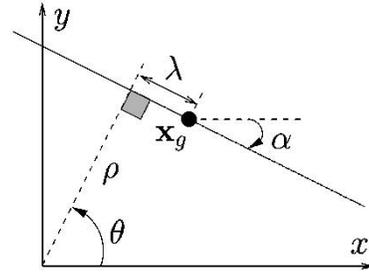


Fig. 4. Representation of a straight line.

where λ is unknown (see Fig. 4). Let us note that the derivations described briefly below are simpler if we consider $\lambda = 0$, but that is not necessary at all (and not perfectly justified from a theoretical point of view), and we will see that all of the terms in which λ is induced disappear, as expected.

First, from (30), we deduce directly $\mathbf{L}_\theta = \mathbf{L}_\alpha$. Using the interaction matrix \mathbf{L}_α related to the orientation of a segment, given in (29), and using $\sin \alpha = -\cos \theta$, $\cos \alpha = \sin \theta$ and (32), we directly obtain

$$\mathbf{L}_\theta = [\lambda_\theta c\theta \quad \lambda_\theta s\theta \quad -\lambda_\theta \rho \quad -\rho c\theta \quad -\rho s\theta \quad -1] \quad (33)$$

where $c\theta = \cos \theta$, $s\theta = \sin \theta$, and $\lambda_\theta = Bc\theta - As\theta$.

Furthermore, we obtain, by differentiating (31) to yield

$$\begin{aligned} \mathbf{L}_\rho &= c\theta \mathbf{L}_{x_g} + s\theta \mathbf{L}_{y_g} - s\theta x_g \mathbf{L}_\theta + c\theta y_g \mathbf{L}_\theta \\ &= c\theta \mathbf{L}_{x_g} + s\theta \mathbf{L}_{y_g} - \lambda \mathbf{L}_\theta. \end{aligned}$$

Using (14), (33), and noting from (27) that

$$\begin{cases} n_{20}/(n_{20} + n_{02}) = \cos^2 \alpha = \sin^2 \theta \\ n_{02}/(n_{20} + n_{02}) = \sin^2 \alpha = \cos^2 \theta \\ n_{11}/(n_{20} + n_{02}) = \cos \alpha \sin \alpha = -c\theta s\theta \end{cases}$$

we obtain, after simple developments

$$\mathbf{L}_\rho = [\lambda_\rho c\theta \quad \lambda_\rho s\theta \quad -\lambda_\rho \rho \quad (1 + \rho^2)s\theta \quad -(1 + \rho^2)c\theta \quad 0] \quad (34)$$

where $\lambda_\rho = -(A\rho c\theta + B\rho s\theta + C)$. As expected, we thus find again the analytical form of \mathbf{L}_θ and \mathbf{L}_ρ given in [8].

If the considered 3-D object is a straight line, parameters A , B , and C are directly obtained from the equation of a plane to which the 3-D straight line belongs. The only condition is that the chosen plane does not contain the camera optical center. If the considered 3-D object is a cylinder, its projection in the image can be represented by two straight lines (in all nondegenerated cases) and parameters A , B , and C can have the same value for both lines. More precisely, we have [2]

$$\begin{cases} A = X_O/(X_O^2 + Y_O^2 + Z_O^2 - R^2) \\ B = Y_O/(X_O^2 + Y_O^2 + Z_O^2 - R^2) \\ C = Z_O/(X_O^2 + Y_O^2 + Z_O^2 - R^2) \end{cases}$$

where R is the radius of the cylinder, and where X_O , Y_O , and Z_O are the coordinates of the point of the cylinder axis nearest to the camera optical center.

We have shown in this section how the interaction matrix related to simple and classical geometrical primitives can be determined by considering particular forms of moments. We now come back to the general case by searching for adequate combinations of moments able to control the six DOFs of a robot by image-based visual servoing.

IV. SELECTION OF FEATURES FOR VISUAL SERVOING

From the results of Section II, four adequate visual features are available from moments of order less than 3: coordinates x_g and y_g of the center of gravity; area a ; and orientation α . They are particularly interesting, since they are closely related to v_x and ω_y, v_y and ω_x, v_z , and ω_z , respectively. In fact, these four features have already been used in the partitioned system presented in [6] (with a more intuitive and approximative modeling of the interaction matrix). The question is now to determine supplementary visual features in order to decouple ω_y from v_x and ω_x from v_y . We have already seen in Section II-C that using only moments of order less than 3 was not possible. This is why we propose to select two combinations of moments of order up to 3 and not combinations of moments of higher order. Indeed, it seems preferable to use moments of order as low as possible to reduce the sensitivity of the control scheme to image noise. For that, the well-known skewness terms defined by $S_x = \mu_{30}/\mu_{20}^{3/2}$ and $S_y = \mu_{03}/\mu_{02}^{3/2}$ could be used [17], [24]. However, a better decoupled behavior will be obtained if we are able to determine invariants to translational motions and rotational motion around the optical axis (since we already have visual features to control these motions). Determining such invariants is a very difficult problem [22]. Our choice is based on the famous Hu's invariants to image-based motion [11]. We recall three of them, given as follows:

$$\begin{cases} I_1 = c_1^2 + s_1^2 (= \Delta) \\ I_2 = c_2^2 + s_2^2 \\ I_3 = \mu_{20} + \mu_{02} \end{cases} \quad (35)$$

where

$$\begin{cases} c_1 = \mu_{20} - \mu_{02}, & s_1 = 2\mu_{11} \\ c_2 = \mu_{03} - 3\mu_{21}, & s_2 = \mu_{30} - 3\mu_{12}. \end{cases}$$

These invariants can be combined using products and ratios to also obtain invariants to scale. In such a way, we propose using

$$\begin{cases} P_x = I_1/I_3^2 \\ P_y = aI_2/I_3^3. \end{cases} \quad (36)$$

Indeed, when the object is parallel to the image plane, the interaction matrix related to any of these two features, denoted \mathbf{L}^{\parallel} , has the expected form

$$\mathbf{L}^{\parallel} = [0 \quad 0 \quad 0 \quad l_{wx} \quad l_{wy} \quad 0]. \quad (37)$$

Unfortunately, the value of l_{wx} and l_{wy} (whose analytical form is too complex to be given here) is equal to 0 for P_x and P_y when a centered symmetrical object is observed in the image. In fact, this problem appears for all of Hu's invariants. A different set has to be selected for symmetrical objects in order to avoid this isolated singularity. In that case, from the form of Hu's invariants (which look like the basic trigonometric equation), we propose using

$$\begin{cases} s_x = (c_2c_3 + s_2s_3)/K \\ s_y = (s_2c_3 - c_2s_3)/K \end{cases} \quad (38)$$

where $c_3 = c_1^2 - s_1^2, s_3 = 2s_1c_1$, and $K = I_1I_3^{(3/2)}/\sqrt{a}$. These features have been obtained by looking for the analytical and numerical forms of the interaction matrix related to several

possible trigonometric combinations and several possible normalizations. More precisely, we have chosen two invariants to 2D rotations and scale, valid for symmetrical objects (such that $\mu_{30} = \mu_{21} = \mu_{12} = \mu_{03} = 0$), such that the first one is as independent as possible from ω_y , while the second is as independent as possible from ω_x . For the proposed features and for the configurations where the object is parallel to the image plane, the interaction matrix \mathbf{L}^{\parallel} has the same nice form as (37). For non-symmetrical objects, s_x and s_y are not invariant to 2D rotations, since they do not correspond to combinations of Hu's invariants. This implies that the coefficient of their interaction matrix corresponding to ω_z is no more equal to 0. That is why we propose using s_x and s_y only for symmetrical objects, and P_x and P_y in all other cases.

Finally, we have to note that the translational terms of the interaction matrix related to P_x and P_y (and to s_x and s_y for a symmetrical object) are equal to 0 only when the object is parallel to the image plane. The behavior far away from these configurations may be unsatisfactory. However, as we will see in the next section, where we have considered the case where object and image planes are parallel at the desired position, the results are satisfactory, even if the object is not parallel to the image plane at the beginning of the positioning task.

V. EXPERIMENTAL RESULTS

In the presented experimental results, obtained with a 6-DOFs eye-in-hand system, we have used the classical control law

$$\mathbf{v}_c = -\lambda \mathbf{L}_{\mathbf{s}|\mathbf{s}=\mathbf{s}^*}^{-1} (\mathbf{s} - \mathbf{s}^*) \quad (39)$$

where \mathbf{v}_c is the camera kinematic screw sent to the low-level robot controller, λ is a positive gain, and $\mathbf{L}_{\mathbf{s}|\mathbf{s}=\mathbf{s}^*}^{-1}$ is the inverse of the interaction matrix related to \mathbf{s} computed for the desired value \mathbf{s}^* . We recall that $\mathbf{L}_{\mathbf{s}|\mathbf{s}=\mathbf{s}^*}^{-1}$ exists, since the six visual features we have selected are such that $\mathbf{L}_{\mathbf{s}|\mathbf{s}=\mathbf{s}^*}$ is of full rank 6. Using this control scheme, only the local stability of the system is demonstrated, since the sufficient positivity condition

$$\mathbf{L}_{\mathbf{s}|\mathbf{s}=\mathbf{s}^*}^{-1} \mathbf{L}_{\mathbf{s}}(\mathbf{t}) > 0$$

is only ensured in a neighborhood of $\mathbf{s} = \mathbf{s}^*$ (see [8] for more details). Using $\mathbf{L}_{\mathbf{s}}^{-1}$ instead of $\mathbf{L}_{\mathbf{s}|\mathbf{s}=\mathbf{s}^*}^{-1}$ in control scheme (39) would possibly give even better results, but we currently have no way to estimate the 3-D parameters A, B , and C involved in the interaction matrix when objects of unknown shape are considered.

A. Case of a Simple Symmetrical Object

We first consider a simple rectangular object whose corners can easily be tracked in the acquired images. The desired camera position is such that the rectangle is parallel to the image plane, at a range of 50 cm from the camera optical center ($A = B = 0, C = 2$), and such that the rectangle is centered and horizontal in the image [see Fig. 5(a)]. The displacement that the camera has to realize is approximately composed of a rotation of $-35^\circ, -20^\circ$, and -65° around camera axis x, y , and z , respectively, and of a translation of 10, -25 , and 15 cm along these axes. The image acquired at the initial camera position is displayed in Fig. 5(b).

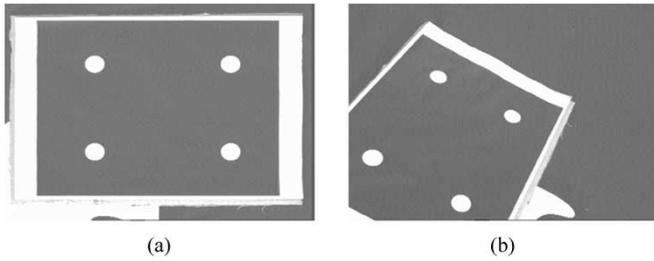


Fig. 5. (a) Desired image. (b) Initial image.

Using the coordinates x_i and y_i of the four corners as visual features, the numerical value of the interaction matrix for the desired position is given by

$$\mathbf{L}_s|_{s=s^*} = \begin{bmatrix} -2 & 0 & -0.3 & 0.01 & -1.02 & -0.10 \\ 0 & -2 & -0.2 & 1.01 & -0.01 & 0.14 \\ -2 & 0 & 0.3 & -0.01 & -1.02 & -0.10 \\ 0 & -2 & -0.2 & 1.01 & 0.01 & -0.14 \\ -2 & 0 & 0.3 & 0.01 & -1.02 & 0.10 \\ 0 & -2 & 0.2 & 1.01 & -0.01 & -0.14 \\ -2 & 0 & -0.3 & -0.01 & -1.02 & 0.10 \\ 0 & -2 & 0.2 & 1.01 & 0.01 & 0.14 \end{bmatrix}.$$

As usual, this matrix does not present any interesting decoupling properties, and we can note that its condition number is approximately equal to 180. The results obtained are depicted in Fig. 7. Hopefully, the camera reaches its desired position, but we can see that the corner trajectories in the image, as well as the camera trajectory (that can be extrapolated from the computed velocity), are not particularly satisfactory.

We now describe the results obtained for the same experiment exactly, but using the six visual features based on image moments selected in Section IV. In practice, we determine the shape of the quadrilateral that corresponds to the image of the rectangle from the position of the corners, and then compute the moments of the quadrilateral [20]. We thus have as visual features

$$\mathbf{s} = (x_g, y_g, a, s_x, s_y, \alpha)$$

whose desired value \mathbf{s}^* is given by

$$\mathbf{s}^* = (0, 0, a^*, 0, 0, 0).$$

We can note that the selected visual features allow one to specify, easily and in an intuitive way, the desired position or the desired image to be reached. For example, modifying the depth between the camera and the object can be specified by modifying only the desired value of the object area a^* in the image. That is a supplementary advantage of determining a minimal set of visual features directly related to the 3-D translational and rotational motions.

At the desired value \mathbf{s}^* , the interaction matrix presents excellent decoupling properties, since it is upper triangular with only two non-null values in the upper part. More precisely, we have

$$\mathbf{L}_s|_{s=s^*} = \begin{bmatrix} -2 & 0 & 0 & 0 & -1.02 & 0 \\ 0 & -2 & 0 & 1.01 & 0 & 0 \\ 0 & 0 & 0.2 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.15 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.19 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

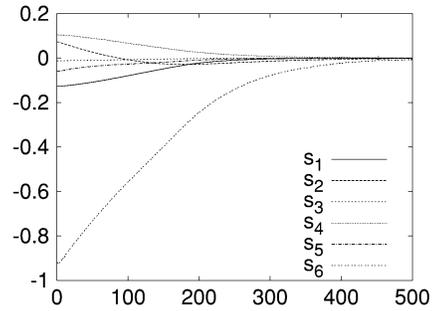


Fig. 6. Results for the rectangle using image moments as visual features: image moments error (m).

whose condition number is now equal to 17.1. The numerical stability of the system is thus greatly improved using moments-based features instead of points features (gain with scale equal to more than 10). We also note that the numerical value of $\mathbf{L}_s|_{s=s^*}$ is directly obtained from the analytical forms determined in this paper. More precisely, it just needs the numerical value of the moments of order up to 4 and C . Since computing moments of a quadrilateral is very easy [20], the rate of the control loop is the same using image moments or image-points coordinates (that is, 50 Hz). The obtained results are depicted in Figs. 6 and 8. By comparing Figs. 7 and 8, we can immediately observe the improvements obtained in the camera and image-corners trajectories, even at the beginning of the camera displacement (where the value and form of the current interaction matrix \mathbf{L}_s is completely different from $\mathbf{L}_s|_{s=s^*}$).

B. Case of a Complex Object: The “Whale”

We finally present the results obtained for a nonsymmetrical planar object whose shape is quite complex [see Fig. 9(a) and (b)]. In practice, the moments are now computed by simple summations on the binarized images. This simple image processing and all of the computations involved in an iteration of the control loop are again performed at the video rate. We now use as visual features

$$\mathbf{s} = (x_g, y_g, a, P_x, P_y, \alpha).$$

The numerical value of the interaction matrix computed from the desired image depicted in Fig. 9(a) is given by

$$\mathbf{L}_s|_{s=s^*} = \begin{bmatrix} -2 & 0 & 0.09 & 0.01 & -1.04 & 0.01 \\ 0 & -2 & 0.01 & 1.02 & -0.01 & -0.05 \\ 0 & 0 & 0.26 & 0.01 & -0.01 & 0 \\ 0 & 0 & 0 & 0.08 & -0.11 & 0 \\ 0 & 0 & 0 & 0.04 & 0.05 & 0 \\ 0 & 0 & 0 & -0.00 & -0.05 & -1 \end{bmatrix}.$$

We can note that $\mathbf{L}_s|_{s=s^*}$ is no longer purely triangular (such a form has not been reached yet) and presents more non-null values in the upper part than previously (mainly because the “whale” is not centered in the desired image). The condition number of $\mathbf{L}_s|_{s=s^*}$ is now equal to 42.1, which is still far better than using points coordinates in \mathbf{s} (it would be impossible to use points coordinates in that case).

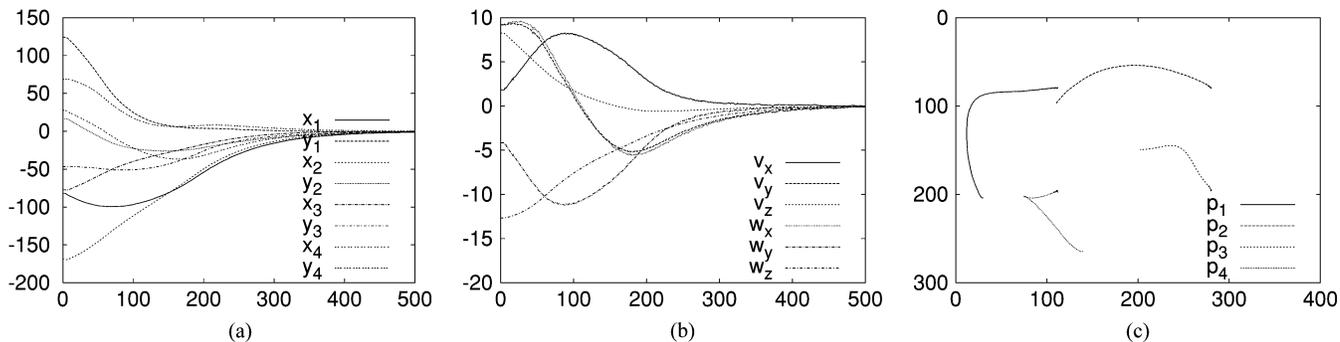


Fig. 7. Results for the rectangle using point coordinates as visual features. (a) Point coordinates error (pixels) versus iteration number. (b) Computed camera velocity (cm/s and dg/s) versus iteration number. (c) Image points trajectory.

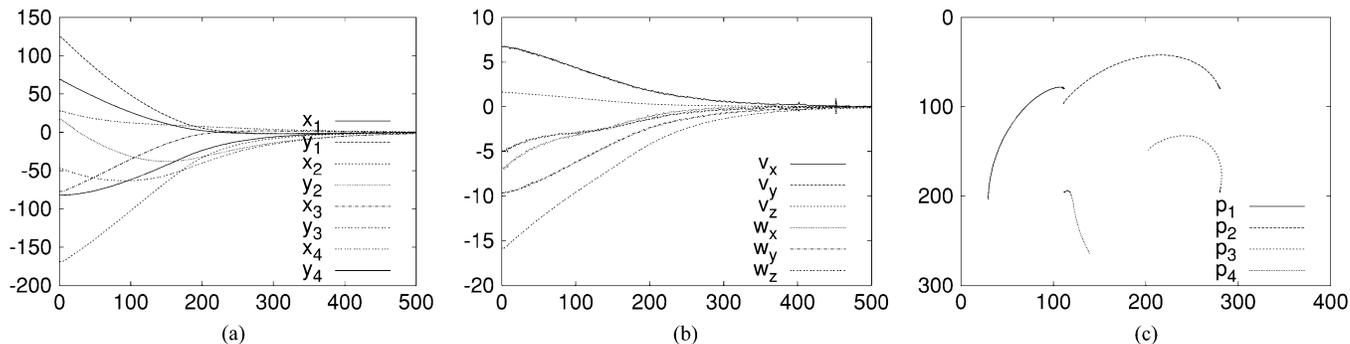


Fig. 8. Results for the rectangle using image moments as visual features. (a) Point coordinates error (pixels). (b) Computed camera velocity (cm/s and dg/s). (c) Image points trajectory.

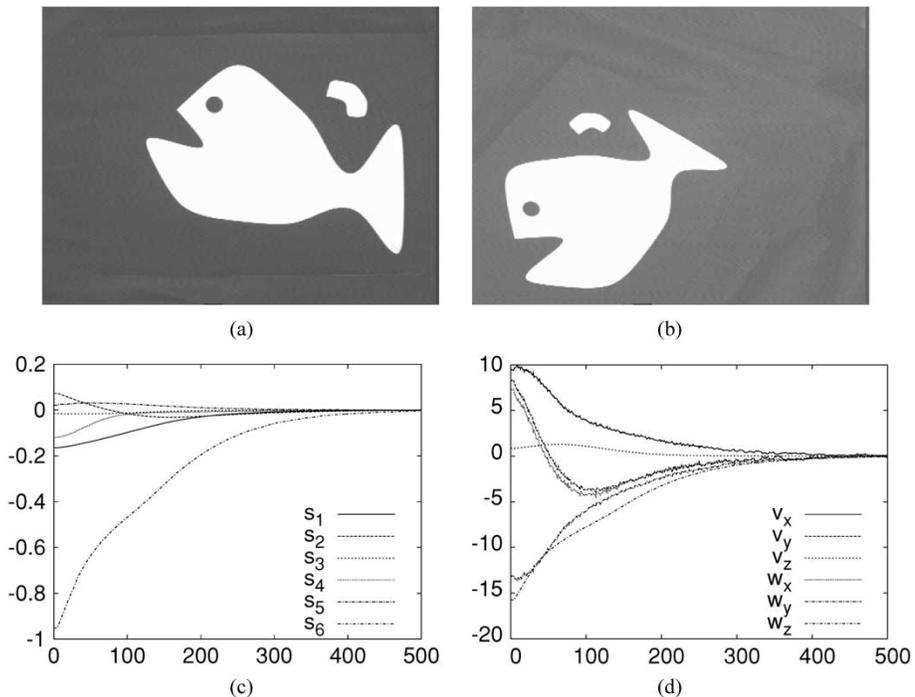


Fig. 9. Results for the “whale.” (a) Desired image. (b) Initial image. (c) Image moments error (m). (d) Computed camera velocity (cm/s and dg/s).

The results obtained for the same displacement to realize as before are depicted in Fig. 9. Even if the decreasing behavior of the visual features is similar for the rectangle and for the “whale” [compare Figs. 6 and 9(c)], the camera trajectory is a little bit less satisfactory [compare Figs. 8(b) and 9(d)]. However, the convergence is obtained, even for the large

considered displacement, and all of the “whale” remains in the camera field of view during the positioning task. We can finally note that the small white part outside the “whale” [see Fig. 9(a) and (b)], which is taken into account in the computation of the moments, does not perturb the behavior of the system at all.

VI. CONCLUSION

This paper has presented a first step to using image moments in visual servoing. The improvements with respect to classical visual servoing seem to be significant, since we obtained a generic representation able to handle simple geometrical primitives, but also complex objects with unknown shapes. The visual features we have selected by combinations of moments present nice decoupling and stability properties. However, a great deal of work remains to be done. First, we would like to adapt our selection of visual features to the case where the object and camera planes are not parallel at the desired position. Indeed, in that case, since the interaction matrix related to the selected visual features has no particular decoupling form, the obtained behavior is satisfactory for small displacements to realize, but may be unsatisfactory when large displacements around x and y camera axes have to be realized (as for all existing image-based visual-servoing schemes). Future work will be also devoted to studying the robustness of our scheme with respect to calibration errors and to partial occlusions. We also would like to try to demonstrate the global stability of the system. This will certainly be quite complex, but may be possible, thanks to the selection of a minimal number of decoupled features. We are also convinced that it is possible to improve the selection of the visual features, for example, following the recent developments presented in [13] (which were concerned with only one DOF) and [23].

APPENDIX

We give here the analytical form of the interaction matrix related to any image moment, in the general case, where the object belongs to a continuous non-planar surface. As already stated in Section II-A, the depth Z of any object point can be expressed as a continuous function of its image coordinates x and y

$$\frac{1}{Z} = \sum_{p \geq 0, q \geq 0} A_{pq} x^p y^q.$$

Following exactly the same reasoning as before, changes in the determination of the interaction matrix will only appear in (10) and (11), from which we obtain

$$\mathbf{L}_{m_{ij}} = [m_{vx} \quad m_{vy} \quad m_{vz} \quad m_{wx} \quad m_{wy} \quad m_{wz}]$$

where only the first three terms that correspond to translational motions are modified (since depth has no influence for rotational motions). More precisely, we have

$$\begin{cases} m_{vx} = -\sum_{p,q} (i+p) A_{pq} m_{i+p-1, j+q} \\ m_{vy} = -\sum_{p,q} (j+q) A_{pq} m_{i+p, j+q-1} \\ m_{vz} = \sum_{p,q} (i+j+p+q+2) A_{pq} m_{i+p, j+q}. \end{cases}$$

We can note that this general form obviously contains the particular form given in (12) for a planar object (where we had set $A_{00} = C, A_{10} = A, A_{01} = B$ and all other terms A_{pq} equal to 0).

From the previous result, we easily deduce the terms that correspond to the translational part of the interaction matrices related to the area a , and the coordinates x_g and y_g of the object center of gravity

$$\begin{cases} a_{vx} = -\sum_{p,q} p A_{pq} m_{p-1,q} \\ a_{vy} = -\sum_{p,q} q A_{pq} m_{p,q-1} \\ a_{vz} = \sum_{p,q} (p+q+2) A_{pq} m_{pq} \\ x_{g_{vx}} = \frac{1}{a} \sum_{p,q} A_{pq} (-(p+1)m_{pq} + px_g m_{p-1,q}) \\ x_{g_{vy}} = \frac{1}{a} \sum_{p,q} A_{pq} (-qm_{p+1,q-1} + qx_g m_{p,q-1}) \\ x_{g_{vz}} = \frac{1}{a} \sum_{p,q} A_{pq} ((p+q+3)m_{p+1,q} - (p+q+2)x_g m_{pq}) \\ y_{g_{vx}} = \frac{1}{a} \sum_{p,q} A_{pq} (-pm_{p-1,q+1} + py_g m_{p-1,q}) \\ y_{g_{vy}} = \frac{1}{a} \sum_{p,q} A_{pq} (-(q+1)m_{pq} + qy_g m_{p,q-1}) \\ y_{g_{vz}} = \frac{1}{a} \sum_{p,q} A_{pq} ((p+q+3)m_{p,q+1} - (p+q+2)y_g m_{pq}). \end{cases}$$

To compute the terms that correspond to the centered moments μ_{ij} , a more simple form can be derived if the polynomial that links the depth Z of a point to the corresponding image point is decomposed around the object center of gravity

$$\frac{1}{Z} = \sum_{p \geq 0, q \geq 0} B_{pq} (x - x_g)^p (y - y_g)^q.$$

We have the following relations:

$$\begin{aligned} B_{pq} &= \sum_{k=p} \sum_{l=q} A_{kl} \binom{k}{p} \binom{l}{q} x_g^{k-p} y_g^{l-q} \\ A_{pq} &= \sum_{k=p} \sum_{l=q} B_{kl} \binom{k}{p} \binom{l}{q} (-x_g)^{k-p} (-y_g)^{l-q} \end{aligned}$$

which, for a planar object, yields

$$\begin{aligned} B_{00} &= 1/Z_g = Ax_g + By_g + C \\ B_{10} &= A_{10} = A \\ B_{01} &= A_{01} = B. \end{aligned}$$

Using this form, we obtain

$$\begin{cases} a_{vx} = -\sum_{p,q} p B_{pq} \mu_{p-1,q} \\ a_{vy} = -\sum_{p,q} q B_{pq} \mu_{p,q-1} \\ a_{vz} = \sum_{p,q} B_{pq} ((p+q+2)\mu_{pq} + px_g \mu_{p-1,q} + qy_g \mu_{p,q-1}) \\ x_{g_{vx}} = -\sum_{p,q} (p+1) B_{pq} n_{pq} \\ x_{g_{vy}} = -\sum_{p,q} q B_{pq} n_{p+1,q-1} \\ x_{g_{vz}} = \sum_{p,q} B_{pq} (\gamma n_{p+1,q} + (p+1)x_g n_{pq} + qy_g n_{p+1,q-1}) \end{cases}$$

where $\gamma = p + q + 3$ (and we recall that n_{ij} is the centered moment normalized by the area $n_{ij} = \mu_{ij}/a$). Similarly, we obtain

$$\begin{cases} y_{g_{vx}} = -\sum_{p,q} p B_{pq} n_{p-1,q+1} \\ y_{g_{vy}} = -\sum_{p,q} (q+1) B_{pq} n_{pq} \\ y_{g_{vz}} = \sum_{p,q} B_{pq} (\gamma n_{p,q+1} + (q+1)y_g n_{pq} + px_g n_{p-1,q+1}). \end{cases}$$

After some simple (but tedious) developments using (20) and the previous equations, we finally obtain the terms related to

the translational part of the interaction matrix related to any centered moment $\mu_{i,j}$ as follows:

$$\begin{cases} \mu_{vx} = -\sum_{p,q} \gamma_1 B_{pq} \mu_{i+p-1,j+q} \\ \quad -i\mu_{i-1,j} x_{g_{vx}} - j\mu_{i,j-1} y_{g_{vx}} \\ \mu_{vy} = -\sum_{p,q} \gamma_2 B_{pq} \mu_{i+p,j+q-1} \\ \quad -i\mu_{i-1,j} x_{g_{vy}} - j\mu_{i,j-1} y_{g_{vy}} \\ \mu_{vz} = \sum_{p,q} B_{pq} (\gamma_3 \mu_{\gamma_1,\gamma_2} + \gamma_1 x_g \mu_{\gamma_1-1,\gamma_2} + \gamma_2 y_g \mu_{\gamma_1,\gamma_2-1}) \\ \quad -i\mu_{i-1,j} x_{g_{vz}} - j\mu_{i,j-1} y_{g_{vz}} \end{cases}$$

where $\gamma_1 = i+p$, $\gamma_2 = j+q$, and $\gamma_3 = \gamma_1 + \gamma_2 + 2$. From these last equations, we can deduce, if needed, the analytical form of the interaction matrix related to any combination of moments, such as, for example, the object orientation α . In all cases, the results are a generalization of those obtained by considering a planar object (i.e., the planar case can be retrieved from the general case by canceling all the terms A_{pq} and B_{pq} , such that $p+q > 1$).

To conclude, we can see from all the previous equations that, if the object is not planar, or if its limb surface is not a plane, the translational terms of the interaction matrix related to any moments combination requires the value of moments of order higher than $i+j+2$. Furthermore, an approximation of all the terms A_{pq} or B_{pq} is also necessary in the visual-servoing scheme. Even if not validated through experimental results, approximating the object surface by a plane seems to be a good compromise, in practice, between modeling errors and complexity, especially thanks to the well-known robustness properties of closed-loop visual-servoing schemes with respect to modeling errors. Regardless, the results presented in this Appendix may be useful for an accurate theoretical stability analysis of the system.

REFERENCES

- [1] Z. Bien, W. Jang, and J. Park, "Characterization and use of Feature-Jacobian matrix for visual servoing," in *Visual Servoing*, K. Hashimoto, Ed, Singapore: World Scientific, 1993, pp. 317–363.
- [2] F. Chaumette, P. Rives, and B. Espiau, "Classification and realization of the different vision-based tasks," in *Visual Servoing*, K. Hashimoto, Ed, Singapore: World Scientific, 1993, pp. 199–228.
- [3] F. Chaumette, "Potential problems of stability and convergence in image-based and position-based visual servoing," in *The Confluence of Vision and Control*, ser. LNCIS Series. Berlin, Germany: Springer-Verlag, 1998, 237, pp. 66–78.
- [4] R. Cipolla, *Active Visual Inference of Surface Shape*. Berlin, Germany: Springer-Verlag, 1995, vol. 1016, Lecture Notes in Computer Science.
- [5] C. Collewet and F. Chaumette, "A contour approach for image-based control of objects with complex shape," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, vol. 1, Takamatsu, Japan, Nov. 2000, pp. 751–756.
- [6] P. Corke and S. Hutchinson, "A new partitioned approach to image-based visual servo control," *IEEE Trans. Robot. Automat.*, vol. 17, pp. 507–515, Aug. 2001.
- [7] Z. Duric, A. Rosenfeld, and J. Duncan, "The applicability of Green's theorem to computation of rate of approach," *Int. J. Comput. Vis.*, vol. 31, no. 1, pp. 83–98, Feb. 1999.

- [8] B. Espiau, F. Chaumette, and P. Rives, "A new approach to visual servoing in robotics," *IEEE Trans. Robot. Automat.*, vol. 8, pp. 313–326, June 1992.
- [9] J. Feddema, C. Lee, and O. Mitchell, "Automatic selection of image features for visual servoing of a robot manipulator," in *Proc. IEEE Int. Conf. Robotics and Automation*, vol. 2, Scottsdale, AZ, May 1989, pp. 832–837.
- [10] W. H. Hayt, *Engineering Electromagnetics*, 5th ed. New York: McGraw-Hill, 1989.
- [11] M. K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inform. Theory*, vol. 8, pp. 179–187, 1962.
- [12] S. Hutchinson, G. Hager, and P. Corke, "A tutorial on visual servo control," *IEEE Trans. Robot. Automat.*, vol. 12, pp. 651–670, Oct. 1996.
- [13] R. Mahony, P. Corke, and F. Chaumette, "Choice of image features for depth-axis control in image-based visual servo control," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, vol. 1, Lausanne, Switzerland, Oct. 2002, pp. 390–395.
- [14] E. Malis, F. Chaumette, and S. Boudet, "2 1/2D visual servoing," *IEEE Trans. Robot. Automat.*, vol. 15, pp. 238–250, Apr. 1999.
- [15] R. Mukundan and K. R. Ramakrishnan, *Moment Functions in Image Analysis: Theory and Application*, Singapore: World Scientific, 1998.
- [16] B. Nelson and P. Khosla, "The resolvability ellipsoid for visual servoing," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Seattle, WA, June 1994, pp. 829–832.
- [17] R. J. Prokop and A. P. Reeves, "A survey of moment-based techniques for unoccluded object representation and recognition," in *Proc. Computer Vision, Graphics, Image Processing Conf.*, vol. 54, Sept. 1992, pp. 438–460.
- [18] J. Sato and R. Cipolla, "Extracting group transformations from images moments," *Comput. Vis. Image Understanding*, vol. 73, no. 1, pp. 29–42, 1999.
- [19] R. Sharma and S. Hutchinson, "Optimizing hand/eye configuration for visual servo systems," in *Proc. IEEE Int. Conf. Robotics and Automation*, vol. 1, Nagoya, Japan, May 1995, pp. 172–177.
- [20] C. Steger, "On the calculation of arbitrary moments of polygons," Munchen Univ., Munchen, Germany, Tech. Rep. FGBV-96-05, Oct. 1996.
- [21] J. Stewart, *Calculus*, 2nd ed. Pacific Grove, CA: Brooks/Cole, 1991.
- [22] L. Van Gool, T. Moons, E. Pauwels, and A. Oosterlink, "Vision and Lie's approach to invariance," *Image Vis. Comput.*, vol. 13, no. 4, pp. 259–277, May 1995.
- [23] O. Tahri and F. Chaumette, "Application of moment invariants to visual servoing," in *Proc. IEEE Int. Conf. Robotics and Automation*, vol. 3, Taipei, Taiwan, Sept. 2003, pp. 4276–4281.
- [24] G. Wells, C. Venaille, and C. Torras, "Vision-based robot positioning using neural networks," *Image Vis. Comput.*, vol. 14, no. 10, pp. 715–732, Dec. 1996.



François Chaumette (M'98) was born in Nantes, France, in 1963, and graduated from École Nationale Supérieure de Mécanique, Nantes, in 1987. He received the Ph.D degree and "Habilitation à Diriger des Recherches" in computer science from the University of Rennes, Rennes, France, in 1990 and 1998, respectively.

Since 1990, he has been with IRISA/INRIA, Rennes, where he is now "Directeur de Recherches." His research interests include robotics, computer vision, and especially the coupling of these two

research domains (visual servoing and active perception).

Dr. Chaumette received the AFCET/CNRS Prize for the best French thesis in automatic control in 1991. He was also the corecipient, with Ezio Malis, of the 2002 King-Sun Fu Memorial Best IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION Paper Award. He is currently an Associate Editor of the IEEE TRANSACTIONS ON ROBOTICS.