2008 IEEE International Conference on
Robotics and Automation
Pasadena, CA, USA, May 19-23, 2008

3D Structure Identification from Image Moments

Paolo Robuffo Giordano Alessandro De Luca Giuseppe Oriolo
Dipartimento di Informatica e Sistemistica
Università di Roma "La Sapienza"
Via Ariosto 25, 00185 Roma, Italy
{robuffo,deluca,oriolo}@dis.uniroma1.it

Abstract— **In the image-based visual servoing framework, image moments provide an appealing choice as visual features since they can be easily evaluated on any shape on the image plane, and do not require tracking and matching of individual geometric structures between distinct image frames (i.e., the so-called correspondence problem). However, computation of the moment interaction matrix still requires the knowledge of specific unmeasurable 3D quantities relative to the target object, quantities that are usually approximated in practical implementations. Therefore, in this paper we analyze the possibility to estimate on-line the value of such 3D quantities during the camera motion with the only assumption of a target shape with planar limb surface. The proposed estimation scheme builds upon the theory of nonlinear observers, and in particular exploits the basic formulation of the *persistency of excitation* Lemma. Simulation results are then presented in order to support the effectiveness of the proposed approach.**

I. INTRODUCTION

The introduction of visual information in the control loop of robot systems has increased the flexibility and the accuracy of the tasks commonly performed by these systems [1], [2], by providing higher position accuracy, robustness to sensor noise and calibration uncertainties, and reactiveness to environmental changes. This is especially true for the class of mobile robots, where the elaboration of visual cues is often crucial for self-localization and navigation. Another interesting use of visual feedback is the possibility to specify a robotic task in terms of some image features extracted from a target object while the camera/robot is moving through the scene. Two main approaches have been proposed in the past years to deal with this kind of tasks, namely position-based visual servoing (PBVS) and image-based visual servoing (IBVS) schemes, but recently a number of hybrid methods has also been explored [3]–[5]. A thorough presentation and discussion of the different approaches can be found in [1], [6], [7].

In contrast to PBVS methods, which exploit the image features in order to estimate the relative 3D pose between the camera and the target, IBVS schemes compute the error signal directly in terms of quantities extracted from the image plane. Motion of these features is mapped to the velocity twist of the camera via an *interaction matrix* which is then used to control the robot pose by zeroing the image plane error signal. The IBVS approach is usually robust w.r.t. perturbations of the robot/camera models, in particular to calibration errors [8], and more suited to devise feature-based motion strategies aimed at keeping the target always in the field of view of the camera [9]. There are, however, also some drawbacks to be considered. Apart from situations where the interaction matrix loses rank during the motion, local minima of the task error function [10] may be encountered when trying to impose an (infeasible) independent motion to a large number of image features [11]. Moreover, knowledge of some unmeasurable 3D quantities is still needed to correctly compute the interaction matrix. For instance, when considering individual points as visual features, the unknown depth Z of each point is required and must be estimated during the servoing (a common choice is to simply use the constant value at the desired pose). Thus, only local stability can be guaranteed for most IBVS schemes [12].

In the last years, several works have addressed the on-line identification of 3D information for IBVS schemes: Chaumette et al. [13] propose a general methodology to recover the 3D parameters of several geometric primitives (points, lines, cylinders, spheres, etc.) by measuring the current values of the features, of the image motion (the feature time derivatives) and of the camera velocity twist. In [14], two Kalman filter-based algorithms are derived and compared, the first estimating a continuous depth map of the scene, and the second extracting the depth of a discrete set of features. A similar approach is found in [15] where only lateral camera motions are allowed. Adaptive IBVS schemes are devised in [16], [17] for a camera mounted on a nonholonomic mobile robot via an on-line estimation of a constant unknown parameter (the height of the object points and the depth of the target plane at the desired pose, respectively). General solutions to the problem of depth identification for point features have also been proposed in [18]–[20].

Estimating the 3D parameters of geometric primitives (e.g., depths of point features) improves the overall stability of IBVS schemes, and can also be relevant for recovering more complex 3D structures as plane orientations, and similar quantities. It should be noted, however, that tracking and matching individual structures during the camera motion, i.e., solving the so-called correspondence problem, may not be always easy or convenient (think to dense objects as spheres, ellipsoids, etc.). When this is the case, IBVS schemes usually rely on more global (integral) features, like image moments, instead of local descriptors like feature points.

978-1-4244-1647-9/08/$25.00 ©2008 IEEE.

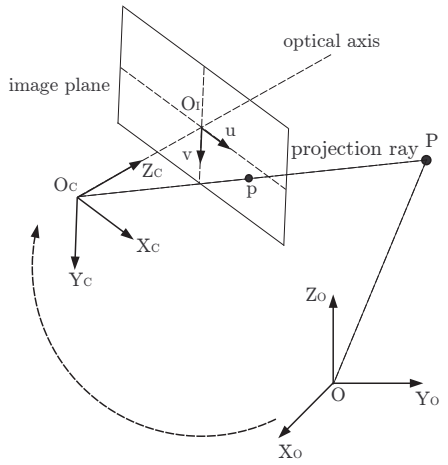

Fig. 1: World and camera frame definition.

Indeed, moments can be directly evaluated on any arbitrary shape on the image plane and are free of the correspondence problem that typically affects the identification of common geometric structures. Furthermore, a suitable combination of moments can be used to control the full pose of a robot, making them an appealing choice for IBVS pose control [21], [22]. Of course, all these nice properties come at a price: when considering moments, the 3D information present in the relative interaction matrix does not reduce to a simple punctual depth, but more general 3D structures are involved. In this respect, the goal of this paper is to explore the possibility to estimate online such 3D quantities having as input the known camera motion and the moments measured on the image plane, under the sole assumption of planar limb surface for the shape of the target object. To this end, we generalize the observer scheme for point features developed in [18] so as to cover the case of moments obtained from generic shapes. Furthermore, we discuss how the use of moments in place of single point features can improve the convergence properties of the mentioned observer, e.g., by reducing the situations where the *persistency of excitation* condition, upon which the observer is built, do not hold.

The paper is organized as follows: in Sect. II we recall the basic kinematic relationships of the camera/target system, while in Sect. III we design a nonlinear observer to estimate the unknown 3D quantities of the moment interaction matrix. Finally, in Sect. IV some simulations are presented in order to show the performance of the proposed observation schemes.

II. Pin-hole camera model

With reference to Fig. 1, consider an inertial world reference frame $\mathcal{F}_O : \{O;\, \vec{X}_O,\, \vec{Y}_O,\, \vec{Z}_O\}$ and a pin-hole camera associated to the moving frame $\mathcal{F}_C : \{O_C;\, \vec{X}_C,\, \vec{Y}_C,\, \vec{Z}_C\}$, with $\vec{Z}_C$ coincident with the camera optical axis. The image plane, perpendicular to the optical axis, lies at a distance λ (the focal length) from O_C, and is endowed with a 2D reference frame $\mathcal{F}_I : \{O_I;\, \vec{u},\, \vec{v}\}$ with axes parallel to $\vec{X}_C$ and $\vec{Y}_C$, respectively. Furthermore, let vector $[v_C^T\ \omega_C^T]^T \in \mathbb{R}^6$ represent the linear/angular velocity of $\mathcal{F}_C$ w.r.t. $\mathcal{F}_O$ expressed in $\mathcal{F}_C$. From standard kinematics, the apparent velocity of a point $P = [X\ Y\ Z] \in \mathbb{R}^3$ in $\mathcal{F}_C$, induced by the camera motion, is

$$\dot{P} = -v_C - [\omega_C]_\times P \tag{1}$$

where $[u]_\times \in so(3)$ is the 3×3 skew-symmetric matrix associated to a vector $u \in \mathbb{R}^3$. Equation (1) can be rearranged in matrix form as

$$\begin{bmatrix} \dot{X} \\ \dot{Y} \\ \dot{Z} \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & 0 & -Z & Y \\ 0 & -1 & 0 & Z & 0 & -X \\ 0 & 0 & -1 & -Y & X & 0 \end{bmatrix} \begin{bmatrix} v_C \\ \omega_C \end{bmatrix}. \tag{2}$$

The pin-hole camera projects a 3D point P in $\mathcal{F}_C$ with homogeneous coordinates $\bar{P} = [X\ Y\ Z\ 1]^T$ into a 2D point p with homogeneous normalized coordinates $\bar{p} = [p_u\ p_v\ 1]^T = [X/Z\ Y/Z\ 1]^T$. The image plane measurement (in pixels) of point p is given by $\tilde{p} = [\tilde{p}_u\ \tilde{p}_v\ 1]^T = A\bar{p}$, where A is a nonsingular matrix containing the camera intrinsic parameters, i.e.,

$$A = \begin{bmatrix} \lambda k_u & -\lambda k_u / \tan\delta & u_0 \\ 0 & \lambda k_v / \sin\delta & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{3}$$

with $[u_0\ v_0]^T$ being the coordinates of the principal point (in pixels), λ the focal length (in meters), k_u and k_v the magnifications in the $\vec{u}$ and $\vec{v}$ directions (in pixel/meters), and δ the angle between these axes. In order to simplify the notation, in the following we will assume that any quantity is expressed in the normalized space. This is equivalent to assume a calibrated camera, i.e., full knowledge of the calibration matrix A.

Given a vector of visual features $f = [f_1 \dots f_s]^T \in \mathbb{R}^s$, the velocity twist $(v_C,\, \omega_C)$ of the camera is mapped into $\dot{f}$ by a $s \times 6$ matrix $J_v(f,\, \chi)$ called the *interaction matrix*

$$\dot{f} = J_v(f, \chi) \begin{bmatrix} v_C \\ \omega_C \end{bmatrix}, \tag{4}$$

where χ is a vector representing 3D information associated to f. It is possible to determine the interaction matrix for many features of interest, see [2] for the case of points, lines, planes, circles, and [21], [22] for the set of image moments. The 3D information represented by χ largely depends on the particular feature extracted from the selected geometric structure. In the case of a point P (the most simple shape), χ reduces to the unknown depth Z while, for more complex shapes, additional quantities are required, like radii of spheres, plane orientations of planar shapes, etc. In all cases, however, depth is always present in χ even if not in an explicit way as in the case of point features.

III. Design of the nonlinear observer

As outlined in the introduction, a common problem to pure IBVS settings is the knowledge of 3D quantities χ in the interaction matrix (4). Among the various approximations, one interesting possibility is to obtain an estimation $\widehat{\chi}(t)$ to be used in place of $\chi(t)$ during the servoing. In [18], we proposed an on-line identification scheme for the depth

Z of point features based on the *persistency of excitation* lemma [23]. The idea was to interpret Z as an unmeasurable time-varying state with known (nonlinear) dynamics upon which a suitable observer could be designed by taking advantage of the aforementioned lemma. The goal of this section is to show how the same basic formulation can be exploited in order to estimate the unknown vector χ relative to the interaction matrix of moments.

As a preliminary step, we recall the *persistency of excitation* Lemma upon which our observer will be built.

Lemma 1: Consider the linear time-varying system

$$\left\{ \begin{array}{rcll} \dot{\xi} & = & W\xi + \Omega^T(t)z, & \xi \in \mathbb{R}^n \\ \dot{z} & = & -\Lambda\Omega(t)S\xi, & z \in \mathbb{R}^p \end{array} \right. \tag{5}$$

where W is an $n\times n$ Hurwitz matrix, S is an $n\times n$ symmetric positive definite matrix such that $W^TS+SW=-Q$, with Q symmetric positive definite, and Λ is a $p\times p$ symmetric positive definite matrix. If $\|\Omega(t)\|$, $\|\dot{\Omega}(t)\|$ are uniformly bounded and the *persistency of excitation* condition is satisfied, i.e., there exist two positive real numbers T and γ such that

$$\int_t^{t+T} \Omega(\tau)\Omega^T(\tau)\mathrm{d}\tau \geq \gamma I > 0, \qquad \forall\, t \geq t_0, \tag{6}$$

then $(\xi,\, z) = 0$ is a globally exponentially stable equilibrium point. ■

The key idea in using this Lemma is the following: given a state vector $x = [x_1^T \;\; x_2^T]^T \in \mathbb{R}^{n+p}$ where only the state subset x_1 is directly measurable, design an update law for the estimated state $\widehat{x} = [\widehat{x}_1^T \;\; \widehat{x}_2^T] \in \mathbb{R}^{n+p}$ such that, by letting $\xi = x_1 - \widehat{x}_1$ and $z = x_2 - \widehat{x}_2$ be the error sub-vectors, the associated error dynamics matches formulation (5). When this manipulation is possible, Lemma 1 guarantees exponential convergence of the error system, or, in other words, that values of the unmeasurable variables x_2 can be inferred from knowledge of x_1. In this context, condition (6) plays the role of an *observability* test, i.e., estimation of x_2 is possible iff there does not exist a $\bar{t}$ such that $\forall t > \bar{t},\ \det(\Omega(t)\Omega^T(t)) \equiv 0$. Such a requirement is violated whenever matrix $\Omega(t)$ ultimately loses rank during the camera motion, or if $n < p$, so that $\Omega(t)\Omega^T(t)$ is structurally singular. As a consequence, in order to estimate p independent quantities, one must necessarily exploit $n \geq p$ independent measurements. Moreover, note that, if $p = 1$, $\Omega(t)$ becomes a row vector: in this case condition (6) is satisfied iff the norm of $\Omega(t)$ (i.e., at least one component) does not ultimately vanish over time.

For the sake of illustration, when considering a point feature $p = [p_u \; p_v]^T$, it is possible to recover formulation (5) by setting $x_1 = [p_u \; p_v]^T$ and $x_2 = 1/Z$, as shown in [18]. In this case, the persistency of excitation conditions states that estimation of x_2 is possible iff

1) the camera is moving with a non-zero linear velocity v_C;
2) the camera is not translating along the projection ray of point p,

i.e., the well-known fact that recovering depth requires a nonzero (and known) camera translational motion [24].

Now consider a generic $(i,\, j)$-th order moment m_{ij} evaluated on the image plane projection of a 3D object $\mathcal{O}$. Assume that $\mathcal{O}$ is planar or has a planar limb surface [2] with plane equation

$$\vec{n} \cdot P + d = 0 \tag{7}$$

in the camera frame, where $\vec{n} = [n_x \; n_y \; n_z]^T \in \mathbb{S}^2$ is the plane unit normal and d the plane distance to the origin of $\mathcal{F}_C$. The depth Z of any 3D point P lying on this plane can be expressed in terms of its image coordinates p as

$$\frac{1}{Z} = Ap_u + Bp_v + C, \tag{8}$$

where

$$\left[\begin{array}{c} A \\ B \\ C \end{array} \right] = -\vec{n}/d. \tag{9}$$

The interaction matrix $J_{m_{ij}}$ of m_{ij} has the expression

$$\dot{m}_{ij} = J_{m_{ij}}(m_{kl},\, \chi) \left[\begin{array}{c} v_C \\ \omega_C \end{array} \right], \tag{10}$$

where m_{kl} stands for generic $(k,\, l)$-th moments of order up to $i + j + 1$, and $\chi = [A\, B\, C]^T$. Hence, when considering moments of any shape, the 3D information represented by χ always reduces to the plane normal $\vec{n}$ scaled by the plane distance d, i.e., the 'depth' of the plane. In the following, we discuss three possible solutions for the estimation of $[A\, B\, C]^T$ depending on the initial assumptions made on the quantities directly measurable.

A. General case

Equation (10) can be rearranged linearly in $(A,\, B,\, C)$ as

$$\begin{aligned} \dot{m}_{ij} =& A\lambda_A(m_{kl},\, v_C) + B\lambda_B(m_{kl},\, v_C) + C\lambda_C(m_{kl},\, v_C) + \\ & \lambda_D(m_{kl},\, \omega_C), \end{aligned} \tag{11}$$

where $\lambda_i(\cdot)$ are known scalar functions of measurable quantities (moments and camera velocity) [21]. Let $x_1 = [m_{i_1j_1} \ldots m_{i_nj_n}]^T \in \mathbb{R}^n$ be a collection of n generic moments, and $x_2 = [A \;\; B \;\; C]^T \in \mathbb{R}^p$, $p = 3$. From (11), we can rewrite the x_1 dynamics in the compact form

$$\begin{aligned} \dot{x}_1 &= \left[\begin{array}{ccc} \lambda_{A_1} & \lambda_{B_1} & \lambda_{C_1} \\ \vdots & \vdots & \vdots \\ \lambda_{A_n} & \lambda_{B_n} & \lambda_{C_n} \end{array} \right] \left[\begin{array}{c} A \\ B \\ C \end{array} \right] + \left[\begin{array}{c} \lambda_{D_1} \\ \vdots \\ \lambda_{D_n} \end{array} \right] = \\ &= \Gamma x_2 + \Pi. \end{aligned} \tag{12}$$

Hence, by defining the update law for $\widehat{x}_1$ as

$$\dot{\widehat{x}}_1 = \Gamma\widehat{x}_2 + \Pi + K_1(x_1 - \widehat{x}_1), \quad K_1 > 0, \tag{13}$$

we get a $\dot{\xi} = \dot{x}_1 - \dot{\widehat{x}}_1$ error dynamics

$$\dot{\xi} = -K_1\xi + \Gamma z \tag{14}$$

that matches exactly the first row of (5) with $z = x_2 - \widehat{x}_2$, $W = -K_1$ and $\Omega^T(t) = \Gamma(t)$. Note that Lemma 1 requires the boundedness of $||\Gamma(t)||$ and $||\dot{\Gamma}(t)||$. In our case, this is guaranteed as long as the camera velocity $(v_C,\, \omega_C)$ keeps

bounded with bounded derivatives and a finite image plane size is assumed, so that the measured moments are bounded.

Now it remains to design an update law for $\widehat{x}_2$ which can yield a $\dot{z}$ dynamics as close as possible to the second row of (5). To this end, we first need an explicit expression of $\dot{x}_2$. From (9), it is

$$\dot{x}_2 = -\frac{\dot{\vec{n}}d - \vec{n}\dot{d}}{d^2} \tag{15}$$

and, since $\vec{n}$ is a free vector expressed in $\mathcal{F}_C$, from standard kinematics we have

$$\dot{\vec{n}} = -[\omega_C]_\times \vec{n}. \tag{16}$$

Expression of $\dot{d}$ can be obtained as follows: among the points belonging to (7) consider point $P_n = -d\vec{n}$, i.e. the point on the plane which lies at a distance d along the direction of $\vec{n}$. From (7), it is $\dot{d} = -\dot{\vec{n}} \cdot P_n - \vec{n} \cdot \dot{P}_n$ and, since $\vec{n}$ and P_n are parallel, $\vec{n} \cdot P_n = 0$. By exploiting the kinematics of P_n given by (1), we have

$$\vec{n} \cdot \dot{P}_n = \vec{n} \cdot (-v_C - [\omega_C]_\times P_n) = -\vec{n} \cdot v_C,$$

where, again, the fact that $\vec{n}$ and P_n are parallel is used. In conclusion, we obtain

$$\dot{d} = \vec{n} \cdot v_C. \tag{17}$$

By plugging (16) and (17) into (15), we get the searched relation

$$\dot{x}_2 = [\omega_C]_\times \frac{\vec{n}}{d} + \left(\frac{\vec{n}}{d} \cdot v_C\right) \frac{\vec{n}}{d}$$

which, using (9), can be explicitly rewritten in terms of $(A,\, B,\, C)$ as

$$\begin{aligned} \dot{x}_2 &= \begin{bmatrix} A^2 & AB & AC \\ AB & B^2 & BC \\ AC & BC & C^2 \end{bmatrix} v_C - [\omega_C]_\times x_2 = \\ &= \Theta(x_2) v_C - [\omega_C]_\times x_2. \end{aligned} \tag{18}$$

Hence, by choosing the update law

$$\dot{\widehat{x}}_2 = \Theta(\widehat{x}_2) v_C - [\omega_C]_\times \widehat{x}_2 + K_2 \Gamma^T \xi, \quad K_2 > 0, \tag{19}$$

we get a $\dot{z}$ error dynamics

$$\dot{z} = (\Theta(x_2) - \Theta(\widehat{x}_2)) v_C - [\omega_C]_\times z - K_2 \Gamma^T \xi \tag{20}$$

which, by setting $\Lambda = K_2$ and $S = I$, results very close to the formulation in (5), the only differences being the first two terms in (20). The last step is to prove stability of the closed-loop error system (14)–(20) despite the presence of the unwanted terms in (20).

Proposition 1: Using the observer (13)–(19), the origin of the error system (14)–(20) is exponentially stable as long as the conditions of Lemma 1 are verified, in particular condition (6).

Proof: Let $e = [\xi^T \;\; z^T]^T$ be the error vector and rewrite (14)–(20) as

$$\begin{aligned} \dot{e} &= \begin{bmatrix} -K_1 & \Gamma \\ -K_2 \Gamma^T & 0 \end{bmatrix} e + \begin{bmatrix} 0 \\ (\Theta(x_2) - \Theta(\widehat{x}_2)) v_C - [\omega_C]_\times z \end{bmatrix} = \\ &= A(t) e + g(e, t) \end{aligned} \tag{21}$$

where we interpreted the term $\Theta(x_2) - \Theta(\widehat{x}_2)$ as a function of e. The quantity $g(e, t)$ can be seen as a perturbation term of the nominal system $\dot{e} = A(t)e$ which is guaranteed to be globally exponentially stable by Lemma 1. Note that $g(e, t)$ is a vanishing perturbation, i.e., $g(0, t) = 0,\ \forall t$. Therefore, if $||g(e, t)||$ is sufficiently small, the exponential stability of (21) is (locally) preserved. Due to the boundedness of $||\Gamma(t)||$ and $||\dot{\Gamma}(t)||$, the nominal system is an exponentially stable slowly varying linear system, and therefore there exists a suitable Lyapunov function $V(e, t)$ such that

$$\begin{gathered} c_1 e^T e \le V \le c_2 e^T e \\ \dot{V}(e,\, t) = \frac{\partial V}{\partial t} + \frac{\partial V}{\partial e} A(t) e \le -c_3 \|e\|^2 \\ \left\| \frac{\partial V}{\partial e} \right\| \le c_4 \|e\|, \end{gathered}$$

with $c_1 \dots c_4$ positive constants. Let $S_c = \{e \mid V(e, t) \le c\}$ be a level set of function V. Since V is radially unbounded, S_c is a compact set. Due to the assumed boundedness of $(v_C,\, \omega_C)$, $g(e, t)$ is (locally) Lipschitz and there exists a positive constant M such that $||g(e, t)|| \le M||e||$ in S_c. Using the Lyapunov candidate V for the perturbed system we get

$$\dot{V}(e, t) \le -c_3 ||e||^2 + \left\| \frac{\partial V}{\partial e} \right\| ||g(e, t)|| \le -c_3 ||e||^2 + c_4 M ||e||^2.$$

If M is small enough to satisfy the bound $M < c_3/c_4$, $\dot{V}$ is negative definite on S_c. Therefore, if the initial error $e(t_0)$ is such that

$$||e(t_0)||^2 \le \frac{V(e(t_0), t_0)}{c_1} \le \frac{c}{c_1}, \tag{22}$$

system (14)–(20) converges exponentially to the origin. Note that, besides being a vanishing perturbation w.r.t. e, the term $g(e,\, t)$ also vanishes for $(v_C,\, \omega_C) = (0,\, 0)$ — see (21). Hence, regardless of the initial error $e(t_0)$, the value of M can always be made arbitrarily small by suitably slowing down the camera motion. Moreover, a less conservative estimation on the initial error norm can be obtained by considering that observer (13)–(19) can be initialized with the measured states x_1. In this case, (22) reduces to

$$||e(t_0)||^2 = ||z(t_0)||^2 \le \frac{c}{c_1}.$$

■

This result demonstrates the possibility to use the measured moments and the known camera velocity to estimate vector $[A\ B\ C]^T$ with the only assumption that the object considered has a planar limb surface. Note, however, that the persistency of excitation condition (6) is supposed to hold. As explained before, this is equivalent to assume that $\Gamma^T(t)$ does not ultimately lose rank over time, and that $n \ge 3$ moments are included in vector x_1.

In practice, the choice (both in number and kind) of the image moments to be used for estimation is crucial to meet condition (6). In fact, as (12) shows, the structure of matrix $\Gamma(t)$ depends on such moments. Our current efforts are aimed

at the identification of a meaningful set of moments suitable for robust convergence. As a first evaluation, simulation results are presented in Sect. IV in order to illustrate the observer behavior when area and barycenter are chosen as moments. This choice, for instance, can improve the overall convergence properties of the estimation w.r.t. the case of a target point feature, since the information relative to the area proves to be relevant to reduce the situations in which condition (6) is not met. In any case, since functions $\lambda_i(m_{kl}, v_C)$ in (12) are such that $\lambda_i(m_{kl}, 0) \equiv 0$ [21], the persistency of excitation requires, again, that the camera must necessarily move with a nonzero linear velocity v_C in order to have a converging estimation process.

B. Plane orientation $\vec{n}$ known

In some cases, it is possible to obtain plane orientation $\vec{n}$ by a direct evaluation. For instance, if the homography matrix H between the current and the desired views is available, $\vec{n}$ can be recovered by suitably decomposing H [24]. Computation of the homography typically requires the tracking and matching of several distinct points on the current/desired images, but there also exist techniques to obtain H from a dense unstructured object [25]–[27].

If $\vec{n}$ is known, estimation of $[A\ B\ C]^T$ is considerably simplified since the unmeasurable quantities only reduce to the plane distance d (see (9)). Indeed, in this case we can set $x_1 = [m_{i_1 j_1} \dots m_{i_n j_n}]^T \in \mathbb{R}^n$, as before, and $x_2 = 1/d$. As a consequence, (11) can be rearranged as

$$\dot{m}_{ij} = -\frac{n_x \lambda_A(m_{kl},\, v_C) + n_y \lambda_B(m_{kl},\, v_C) + n_z \lambda_C(m_{kl},\, v_C)}{d} + \lambda_D(m_{kl},\, \omega_C) = \frac{\lambda(\vec{n}, m_{kl}, v_C)}{d} + \lambda_D(m_{kl},\, \omega_C), \tag{23}$$

and dynamics of x_1 becomes

$$\dot{x}_1 = \frac{1}{d}\begin{bmatrix} \lambda_1(\vec{n}, m_{kl}, v_C) \\ \vdots \\ \lambda_n(\vec{n}, m_{kl}, v_C) \end{bmatrix} + \begin{bmatrix} \lambda_{D_1} \\ \vdots \\ \lambda_{D_n} \end{bmatrix} = \Gamma_2 x_2 + \Pi. \tag{24}$$

By choosing the update law

$$\dot{\widehat{x}}_1 = \Gamma_2 \widehat{x}_2 + \Pi + K_1(x_1 - \widehat{x}_1), \quad K_1 > 0, \tag{25}$$

we obtain the same $\dot{\xi}$ error dynamics as in (14) with $\Omega^T(t) = \Gamma_2(t)$. Expression of $\dot{x}_2$ can be derived from (17) as

$$\dot{x}_2 = -\frac{\vec{n} \cdot v_C}{d^2} = -x_2^2 \vec{n} \cdot v_C,$$

from which, by designing the update law

$$\dot{\widehat{x}}_2 = -\widehat{x}_2^2 \vec{n} \cdot v_C + K_2 \Gamma_2^T \xi, \quad K_2 > 0, \tag{26}$$

we get the $\dot{z}$ error dynamics

$$\dot{z} = -(x_2^2 - \widehat{x}_2^2)\vec{n} \cdot v_C - K_2 \Gamma_2^T \xi. \tag{27}$$

The first term in (27) may be again considered as a vanishing perturbation term $g(e,t)$, so that exponential convergence of observer (25)–(26) can be proven by following the same arguments given for the general case (Sect. III-A).

Concerning condition (6), the same former considerations about number and kind of moments to be included in x_1 hold also in this case. There is, however, a slight difference which may be important in practical implementations: while $\Gamma(t)$ in (12) is a $n \times 3$ matrix, $\Gamma_2(t)$ is always a column vector of dimension n. Hence, as discussed at the beginning of Sect. III, it is sufficient that one component of $\Gamma_2(t)$ does not vanish over time for the persistency of excitation condition to hold. In many practical situations, this can result in a milder constraint than requiring, as in the general case, full-rankness of matrix $\Gamma(t)$ over time. Such difference is, of course, due to the assumed knowledge of $\vec{n}$ which reduces the number of unknown parameters to be estimated.

C. Case of a sphere

In the previous developments, we addressed the estimation of $\chi(t)$ under the sole assumption that object $\mathcal{O}$ possesses a planar limb surface, but without posing other special requirements on its geometric structure. Of course, if some additional information about $\mathcal{O}$ is available, one can exploit this knowledge in order to obtain an improved estimation scheme tailored for the specific case. As an illustrative example, in this section we consider the design of the estimation algorithm when object $\mathcal{O}$ is a sphere. This case has also a practical relevance in the mobile robotics field when, e.g., robots are committed with visual tasks involving tracking/positionig w.r.t. a ball, and similar scenarios.

Consider a 3D sphere, with center $P_0 = [X_0\ Y_0\ Z_0]^T$ and radius R, represented by the equation

$$(X - X_0)^2 + (Y - Y_0)^2 + (Z - Z_0)^2 - R^2 = 0.$$

The sphere is an example of a 3D object with a planar limb surface, and, in this case, (8) becomes

$$\frac{1}{Z} = \frac{X_0}{K} p_u + \frac{Y_0}{K} p_v + \frac{Z_0}{K}, \tag{28}$$

where $K = X_0^2 + Y_0^2 + Z_0^2 - R^2$ [2]. From (28) and (8), it follows

$$\begin{bmatrix} A \\ B \\ C \end{bmatrix} = \frac{1}{K}\begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} = -\frac{\vec{n}}{d}, \tag{29}$$

implying that $\vec{n}$ lies on the ray passing through the sphere center P_0. The projection of a sphere on the image plane is the ellipse

$$(X_0 p_u + Y_0 p_v + Z_0)^2 - K(p_u^2 + p_v^2 + 1) = 0, \tag{30}$$

with an equivalent expression in terms of image moments

$$\begin{aligned} &n_{02}p_u^2 + n_{20}p_v^2 - 2n_{11}p_u p_v + 2(n_{11}y_g - n_{02}x_g)p_u + \\ &+ 2(n_{11}x_g - n_{20}y_g)p_v + n_{02}x_g^2 + n_{20}y_g^2 \\ &- 2n_{11}x_g y_g + 4n_{11}^2 - 4n_{20}n_{02} = 0, \end{aligned} \tag{31}$$

where $\bar{p}_g = [x_g\ y_g\ 1]^T$ is the ellipse barycenter in homogeneous coordinates, and n_{ij} are normalized centered moments of order $i + j$ [21]. By equating (30) and (31), it follows

$$\begin{aligned} x_g &= \frac{X_0 Z_0}{Z_0^2 - R^2} \\ y_g &= \frac{Y_0 Z_0}{Z_0^2 - R^2}, \end{aligned}$$

which, plugged into (28), yields

$$Z_g = \frac{Z_0^2 - R^2}{Z_0} \tag{32}$$

as the barycenter depth, i.e., the depth of the point on the limb plane whose projection is $\bar{p}_g$. In order to evaluate $\vec{n}$ in terms of image quantities, one could hope that the 3D barycenter backprojection $P_g = Z_g \bar{p}_g$ and the sphere center P_0 were aligned. Indeed, in this case, it would be possible to obtain $\vec{n}$ as the direction of the measured barycenter $\bar{p}_g$. Unfortunately, while $x_g Z_g = X_0$ and $y_g Z_g = Y_0$, from (32) it is $Z_g \neq Z_0$, so that P_g and P_0 do not share the same 3D direction (actually, $Z_g < Z_0$, i.e., P_g is always in front of P_0). Note that, however, if $R \ll Z_0$, i.e., if the sphere radius is small compared to the distance of the sphere center from the camera, (32) can be approximated as $Z_g \simeq Z_0$, and (29) becomes

$$\begin{bmatrix} A \\ B \\ C \end{bmatrix} = \frac{1}{Z_g} \begin{bmatrix} \frac{x_g}{x_g^2 + y_g^2 + 1} \\ \frac{y_g}{x_g^2 + y_g^2 + 1} \\ \frac{1}{x_g^2 + y_g^2 + 1} \end{bmatrix} = \frac{\vec{n}_g}{Z_g}. \tag{33}$$

Therefore, under this approximation, the only unmeasurable quantity reduces to Z_g and it is possible to proceed similarly as in Sect. III-B, by setting $x_1 = [m_{i_1 j_1} \dots m_{i_n j_n}]^T \in \mathbb{R}^n$ and $x_2 = 1/Z_g$. Dynamics of x_1 and $\dot{\hat{x}}_1$ are given by (24) and (25), where d is replaced by Z_g, and $\vec{n}$ by $\vec{n}_g$. Furthermore, by using the last row of (2), we have

$$\begin{aligned} \dot{x}_2 = -\frac{\dot{Z}_g}{Z_g^2} \simeq -x_2^2 \dot{Z}_0 = x_2^2 (v_{C_z} + Y_0 \omega_{C_x} - X_0 \omega_{C_y}) = \\ = x_2^2 v_{C_z} + x_2 (y_g \omega_{C_x} - x_g \omega_{C_y}). \end{aligned} \tag{34}$$

The update law for $\widehat{x}_2$ is then chosen as

$$\dot{\widehat{x}}_2 = \widehat{x}_2^2 v_{C_z} + \widehat{x}_2 (y_g \omega_{C_x} - x_g \omega_{C_y}) + K_2 \Gamma_2^T \xi \tag{35}$$

which yields the $\dot{z}$ error dynamics

$$\dot{z} = (x_2^2 - \widehat{x}_2^2) v_{C_z} + z (y_g \omega_{C_x} - x_g \omega_{C_y}) - K_2 \Gamma_2^T \xi. \tag{36}$$

Since the first two perturbation terms in (36) are, again, vanishing for $z = 0$, convergence of observer (25)–(35) can be proved an in the previous sections.

It is interesting to note that, for a sphere, the design of the observer structure is conceptually equivalent to the situation discussed in Sect. III-B. Indeed, in both cases, the plane normal direction $\vec{n}$ is directly evaluated in terms of image data, and the only unknown quantity becomes the 'depth' of the target object. The only relevant difference is that the special geometric structure of the sphere allows a direct evaluation of $\vec{n}$, while in the previous (and more general) case a homography decomposition between current and desired view may be needed in order to obtain the same information.

Fig. 2: Webots simulation environment with a mobile manipulator carrying a camera mounted on the end-effector. As target objects, we considered the case of a planar "F" shape (left) and of a sphere (right).

IV. Simulations

In this section, we present two simulations which show the performances of the estimation schemes for a generic planar shape with known normal (Sect. III-B), and for a sphere (Sect. III-C). Ongoing research efforts are currently devoted to select a suitable set of moments for the general case of Sect. III-A. The algorithms were implemented in the Webots environment [28] by considering a camera mounted on the end-effector of a mobile manipulator made of a unicycle-like platform carrying a polar 2R arm (see Fig. 2). A video clip of these simulations is also attached to the paper. The idea was to test the performance of the observer against the measurement noise automatically introduced by the Webots engine (roughly equivalent to a white noise with standard deviation $\sigma = 0.1$ pixels added to the extracted image data). Such a noise is also representative of errors on the input camera velocity $(v_C,\, \omega_C)$, since both disturbances have comparable effects on the observer behavior.

In the first simulation, we considered a planar "F" shape (Fig. 2, left), and tested the observer (25)–(26) by relying on the area a and the barycenter $(x_g,\, y_g)$ for the estimation of the plane distance d. Hence, in this case it is $x_1 = [a\ x_g\ y_g]^T \in \mathbb{R}^n$, $n = 3$, $x_2 = 1/d$, and $\Gamma_2(t) \in \mathbb{R}^3$. The robot was commanded with a periodic predefined motion according to the velocity profiles $v(t) = 0.7 \sin 0.4\,\pi\, t$, $\dot{q}_1(t) = 0.2 \sin 1.6\,\pi\, t$, and $\dot{q}_2(t) = 0.1 \sin 0.8\,\pi\, t$, where v is the platform linear velocity and $(\dot{q}_1,\, \dot{q}_2)$ the first/second link velocities. The observer was initialized with $\widehat{x}_2(t_0) = 1/\widehat{d}(t_0) = 0.667$ m, and gains $K_1 = 25$ and $K_2 = 8000$.

Results of the simulation are presented in Figs. 3–5. In particular, Fig. 3 shows how the estimate $\widehat{d}(t)$ approaches the true value $d(t)$ after about 12 sec of motion, while, in Fig. 4, we report the behavior of the plane normal $\vec{n} = [n_x\ n_y\ n_z]^T$ which was assumed to be measured independently through an homography decomposition. Furthermore, Fig. 5 depicts the behavior of $||\Gamma_2(t)||$, showing that the choice of moments in x_1 meets the persistency of excitation condition ($||\Gamma_2(t)||$ does not ultimately vanish over time).

In the second simulation (Fig. 2, right), we considered a

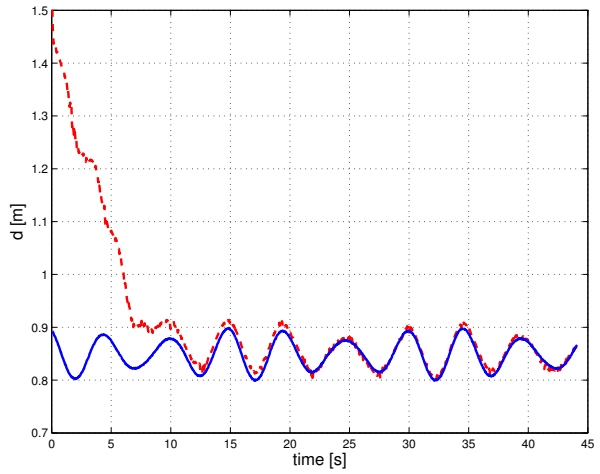

Fig. 3: First simulation. Behavior of d (solid blue line) and $\widehat{d}$ (dashed red line) over time.

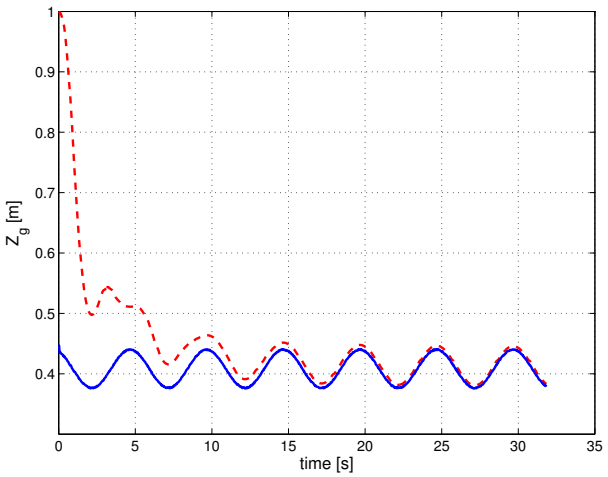

Fig. 6: Second simulation. Behavior of Z_g (solid blue line) and $\widehat{Z}_g$ (dashed red line) over time.

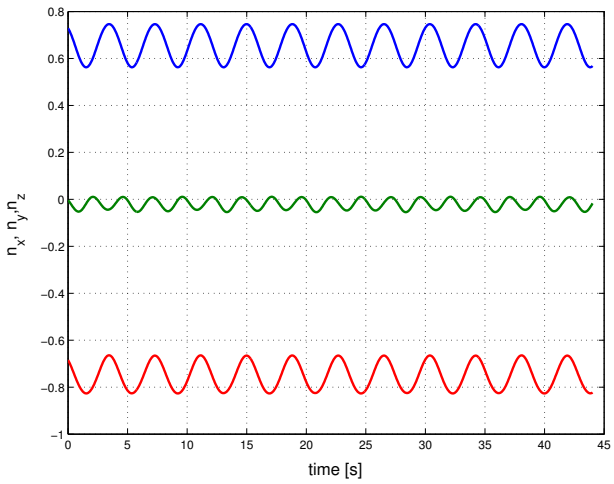

Fig. 4: First simulation. Behavior of $\vec{n} = [n_x\ n_y\ n_z]^T$ over time.

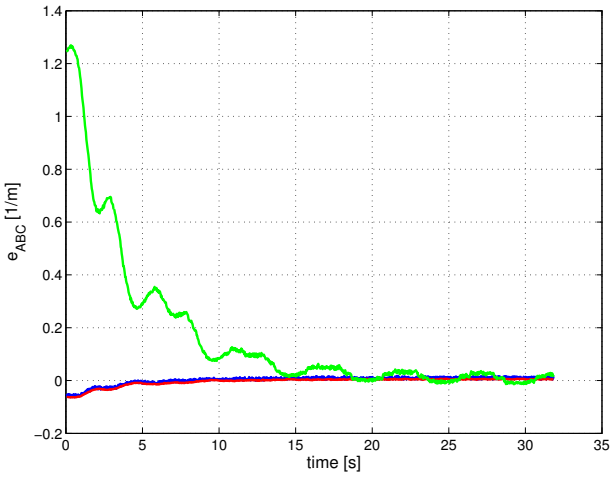

Fig. 7: Second simulation. Behavior of $e_{ABC} = [A\ B\ C]^T - [\widehat{A}\ \widehat{B}\ \widehat{C}]^T$ over time.

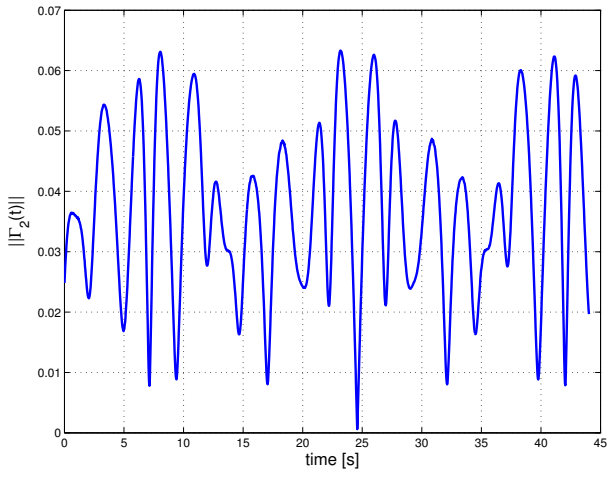

Fig. 5: First simulation. Behavior of $||\Gamma_2||$ over time.

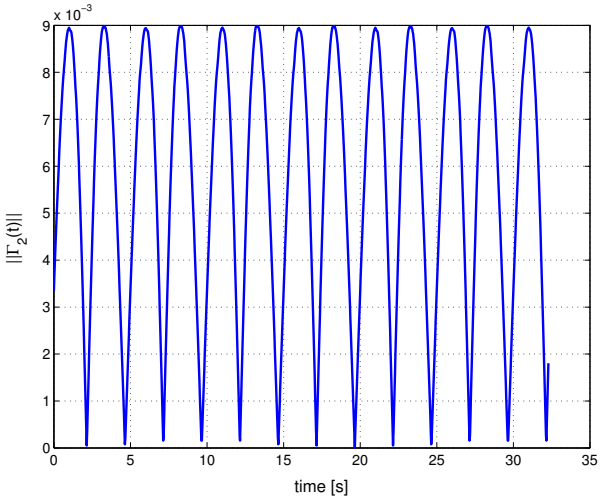

Fig. 8: Second simulation. Behavior of $||\Gamma_2||$ over time.

sphere with radius $R = 0.07$ m, lying at a distance of about 0.4 m from the camera, and the same moments exploited before (area and barycenter). In order to test the performance of the observer (25)–(35), we discarded the link velocity commands $(\dot{q}_1, \dot{q}_2)$, while keeping the previous platform linear velocity command v. As a result, the camera moves is such a way that the sphere barycenter stays almost fixed at the center of the image, i.e., the center of the sphere lies on the camera optical axis during the backward/forward camera motion. This choice was meant to demonstrate the potential benefits of using moments for 3D structure estimation. Indeed, in this situation, an estimation scheme designed for a point feature (like the one proposed in [18]) could not correctly recover the feature depth since the camera linear velocity v_C and the projection ray of the point feature would be almost coincident, thus yielding an ill-conditioned problem (see Sect. III). On the other hand, exploiting the area a besides barycenter (x_g, y_g) makes the estimation possible. This can be verified from Figs. 6–8 which show, as before, the good convergence properties of the observation scheme. Figure 6 illustrates how the estimate $\widehat{Z}_g$ approaches the true value Z_g obtained from (32) after about 10 sec of motion, while Fig. 7 reports the behavior of the error vector $e_{ABC} = [A\ B\ C]^T - [\widehat{A}\ \widehat{B}\ \widehat{C}]^T$, where the first term is evaluated according to (29), and the second is given by (33) with Z_g replaced by its estimate $\widehat{Z}_g$. Hence, one can check that, although neglecting the sphere radius R, observer (25)–(35) is able to recover the actual value of $[A\ B\ C]^T$ in an accurate way. Finally, despite the unfavorable arrangement of the sphere/camera relative motion, the persistency of excitation condition is still satisfied, as can be checked from Fig. 8.

V. Conclusions

By borrowing techniques from nonlinear observer theory, we developed an estimation framework to recover on-line the unmeasurable 3D quantities related to the interaction matrix of image moments. The problem was first addressed in the general case, under the only assumption of a target object with planar limb surface. Then, additional results were presented for more specific cases. Simulation results support the effectiveness of the proposed technique.

In the future, we will analyze the use of different combinations of moments to evaluate the pros and cons of each possible choice. Furthermore, we are planning to implement these estimation techniques on a manipulator equipped with an eye-in-hand camera, so as to obtain an experimental validation of the overall approach.

References

[1] S. Hutchinson, G. D. Hager, and P. I. Corke, "A tutorial on visual servo control," *IEEE Trans. on Robotics and Automation*, vol. 12, no. 5, pp. 651–670, 1996.

[2] B. Espiau, F. Chaumette, and P. Rives, "A new approach to visual servoing in robotics," *IEEE Trans. on Robotics and Automation*, vol. 8, no. 3, pp. 313–326, 1992.

[3] E. Malis, F. Chaumette, and S. Boudet, "2-1/2-D visual servoing," *IEEE Trans. on Robotics and Automation*, vol. 15, no. 2, pp. 238–250, 1999.

[4] E. Malis and F. Chaumette, "Theoretical improvements in the stability analysis of a new class of model-free visual servoing methods," *IEEE Trans. on Robotics and Automation*, vol. 18, no. 2, pp. 176–186, 2002.

[5] E. Cervera, A. D. Pobil, F. Berry, and P. Martinet, "Improving image-based visual servoing with three-dimensional features," *Int. J. of Robotics Research*, vol. 11, no. 10–11, pp. 821–839, 2003.

[6] F. Chaumette and S. Hutchinson, "Visual servo control. I. Basic approaches," *IEEE Robotics & Automation Mag.*, vol. 13, no. 4, pp. 82–90, 2006.

[7] ——, "Visual servo control. II. Advanced approaches," *IEEE Robotics & Automation Mag.*, vol. 14, no. 1, pp. 109–118, 2007.

[8] B. Espiau, "Effect of camera calibration errors on visual servoing in robotics," *Proc. 3rd Int. Symp. on Experimental Robotics*, pp. 182–192, 1993.

[9] P. I. Corke and S. A. Hutchinson, "A new partitioned approach to image-based visual servo control," *IEEE Trans. on Robotics and Automation*, vol. 17, no. 4, pp. 507–515, 2001.

[10] C. Samson, B. Espiau, and M. L. Borgne, *Robot Control: The Task Function Approach*. Oxford University Press, 1991.

[11] F. Chaumette, "Potential problems of stability and convergence in image-based and position-based visual servoing," *The Confluence of Vision and Control*, vol. 237, pp. 66–78, 1998.

[12] E. Malis and P. Rives, "Robustness of image-based visual servoing with respect to depth distribution errors," *Proc. of the 2003 IEEE Int. Conf. on Robotics and Automation*, vol. 1, pp. 1056–1061, 2003.

[13] F. Chaumette, S. Boukir, P. Bouthemy, and D. Juvin, "Structure from controlled motion," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, no. 5, pp. 492–504, 1996.

[14] L. Matthies, R. Szelinski, and T. Kanade, "Kalman filter-based algorithms for estimating depth from image sequences," *Int. J. of Computer Vision*, vol. 3, pp. 209–236, 1989.

[15] C. E. Smith and N. P. Papanikolopoulos, "Computation of shape through controlled active exploration," *Proc. of the 1994 IEEE Int. Conf. on Robotics and Automation*, vol. 3, pp. 2516–2521, 1994.

[16] F. Conticelli, B. Allotta, and P. K. Khosla, "Image-based visual servoing of nonholonomic mobile robots," *Proc. of the 38th Conf. on Decision and Control*, pp. 3496–3501, 1999.

[17] Y. Fang, W. E. Dixon, D. M. Dawson, and P. Chawda, "Homography-based visual servo regulation of mobile robots," *IEEE Trans. on Systems, Man, and Cybernetics, Part B*, vol. 35, no. 5, pp. 1041–1050, 2005.

[18] A. De Luca, G. Oriolo, and P. Robuffo Giordano, "On-line estimation of feature depth for image-based visual servoing schemes," *Proc. 2007 IEEE Int. Conf. on Robotics and Automation*, pp. 2823–2828, 2007.

[19] N. Metni and T. Hamel, "Visual tracking control of aerial robotic systems with adaptive depth estimation," *Int. J. of Control, Automation, and Systems*, vol. 5, no. 1, pp. 51–60, 2007.

[20] W. E. Dixon, Y. Fang, D. M. Dawson, and T. J. Flynn, "Range identification for perspective vision systems," *IEEE Trans. on Automatic Control*, vol. 48, no. 12, pp. 2232–2238, 2003.

[21] F. Chaumette, "Image moments: A general and useful set of features for visual servoing," *IEEE Trans. on Robotics and Automation*, vol. 20, no. 4, pp. 713–723, 2004.

[22] O. Tahri and F. Chaumette, "Point-based and region-based image moments for visual servoing of planar objects," *IEEE Trans. on Robotics*, vol. 21, no. 6, pp. 1116–1127, 2005.

[23] R. Marino and P. Tomei, *Nonlinear Control Design: Geometric, Adaptive and Robust*. Prentice Hall, London, 1995.

[24] Y. Ma, S. Soatto, J. Košecká, and S. S. Sastry, *An Invitation to 3-D Vision*, S. Antman, J. Marsden, L. Sirovich, and S. Wiggins, Eds. Springer–Verlag New York, 2004, vol. 26.

[25] G. Chesi, E. Malis, and F. Cipolla, "Automatic segmentation and matching of planar contours for visual servoing," *Proc. 2000 IEEE Int. Conf. on Robotics and Automation*, pp. 2753–2758, 2000.

[26] S. Benhimane and E. Malis, "Real-time image-based tracking of planes using efficient second-order minimization," *Proc. 2004 IEEE/RSJ Int. Conf. on Intelligent Robots Systems*, pp. 943–948, 2004.

[27] ——, "Homography-based 2D visual servoing," *Proc. 2006 IEEE Int. Conf. on Robotics and Automation*, pp. 2397–2402, 2006.

[28] http://www.cyberbotics.com.