Decoupled Intensity-based Nonmetric Visual Servo Control

Geraldo Silveira, Luiz Mirisola, Pascal Morin

HAL Id: hal-02901281
https://hal.sorbonne-universite.fr/hal-02901281v1

Submitted on 17 Jul 2020

Decoupled Intensity-based Nonmetric Visual Servo Control

Geraldo Silveira, Luiz Mirisola, and Pascal Morin

Abstract—**This brief addresses the problem of vision-based robot control where the equilibrium state is defined via a goal image. Specifically, we consider the class of intensity-based nonmetric solutions, which provide for high accuracy, versatility, and robustness. Existing techniques within that class present a fully coupled translational and rotational control error dynamics, what increases analysis complexity and may degrade system performance. This paper proposes a new intensity-based nonmetric visual servoing technique that decouples the translational control error dynamics, regardless of the observed object characteristics, camera displacements, and their relative poses. The obtained system is thus lower-triangular in the general case. For some practical cases, the proposed general technique leads to the Grail of a fully decoupled (i.e., strictly diagonal) linear dynamics. Theoretical analysis of local diffeomorphism, local exponential stability, and of those decoupling properties are provided. Improved performances are also experimentally confirmed using a six-degree-of-freedom robotic manipulator in various positioning and tracking tasks.**

Index Terms—**Visual servoing, vision-based control, direct visual servoing, robot control, nonmetric estimation.**

I. INTRODUCTION

Visual servo control, or simply visual servoing, refers to the use of image feedback to control a robot with respect to the scene. Its typical task consists in stabilizing the robot around a pose that is defined by means of a reference image, also called desired or goal image. Although there exists a myriad of well-established solutions to this problem [1], its overwhelming majority: (i) relies on image features, such as points, lines, ellipses, etc. Thus, they strongly depend on some particular information, on error-prone feature matching procedures, and on various special tuning algorithms; and/or (ii) requires (or estimates, e.g., [2], [3]) some metric information of the object/scene to provide a provably stabilizing control law. This holds even for the so called image-based visual servoing techniques, since depth is part of their necessary interaction matrix. These topics are further discussed next.

Techniques of vision-based estimation and control can generally be classified into feature- or intensity-based. Despite the aforementioned drawbacks, the vast majority of existing visual servoing schemes are indeed based on image features. Possible reasons for this choice are due to their relatively large domain of convergence and/or to the abundant literature on feature-based estimation. Differently, within *intensity-based* estimation (also called pixel-based or direct methods) there are no intermediate steps, such as feature extraction and matching. These techniques directly exploit the intensity value of the pixels so as to recover the sought parameters. Thus, they make use of raw and dense image data, what allow them to attain high levels of versatility and accuracy [4]. Another advantage is their capacity to ensure robustness to arbitrary illumination changes, even in color images [5] and omnidirectional ones [6]. On the other hand, real-time algorithms based on this class rely on local optimization procedures because global ones are usually too time-consuming to be applied in that setting. They have been applied to visual servoing given a goal image in, e.g., [7]–[9]. Let us note that these intensity-based control techniques require or estimate some metric information.

Techniques of visual servoing can also be classified into metric and nonmetric. Their vast majority fall into the former since at least a depth estimate is used in their control laws. That choice is possibly due to that natural description of a scene, readily availability of depth sensors, and/or to the abundant literature on metric reconstruction. As for *nonmetric* visual servoing, there exist only few works on such class of robot control in spite of its increased level of versatility and robustness to errors in the camera parameters [10]–[12]. Recent studies in biological vision, indeed, suggest that the brain processes visual information nonmetrically [11]. A possible reason for such scarcity is the difficulty of finding an interesting nonmetric control error that is diffeomorphic to the camera pose, and is regulated by a control law that does not depend on any metric knowledge (not even a global scalar) of the object. An early work on nonmetric vision-based navigation is given in [12], where a ground robot is used. In [13], four Degrees-of-Freedom (DoF) of a holonomic robot is under control. The nonmetric methods presented in [14] and [15] take control of all six DoF, but consider only planar objects or pure rotations between the reference and initial frames. Recently, a *general* intensity-based nonmetric technique has been presented in [16], called Direct Visual Servoing (DVS), with three variants in [17]. These techniques are general in the sense that all six DoF are stabilized regardless of the object characteristics (e.g., shape, size), camera displacements, and their relative poses. Nevertheless, all aforementioned nonmetric techniques present a fully coupled control error dynamics.

Geraldo Silveira is with the Center for Information Technology Renato Archer (CTI), Division of Robotics and Computer Vision (DRVC), Brazil. E-mail: *Geraldo.Silveira@cti.gov.br*. (Corresponding author)

Luiz Mirisola is with the Technological Institute of Aeronautics (ITA), Division of Computer Science (IEC), Brazil. E-mail: *lgm@ita.br*.

Pascal Morin is with the Sorbonne Université, CNRS, Institut des Systèmes Intelligents et de Robotique (ISIR), 75005 Paris, France. E-mail: *pascal.morin@sorbonne-universite.fr*.

Couplings in the control dynamics increase analysis complexity and may degrade system performance, e.g. rate and domain of convergence, transient behavior. Not surprisingly, a large body of work has been produced on *decoupled* visual servo control. This is specially true for image-based and hybrid strategies since pose-based visual servoing is intrinsically decoupled. In that spirit, the hybrid schemes proposed in [18] for planar objects and the two techniques presented in [19] (one for planar and the other for nonplanar targets) decouple the error dynamics. A hybrid formulation is also developed in [20] using spherical projection of a set of points, whereas [21] deals with the particular case of a spherical object. Within image-based schemes, the control of translation and rotation are also decoupled in [22] using a spherical representation. In this case, object centroid information is used along with a visual measure of rotation. The strategy described in [23] partitions the interaction matrix to isolate motion related to the optical axis. In [24], a vanishing point and the horizon line have been used as image features to decouple the translational and rotational DoF. Nonetheless, existing decoupled visual servoing techniques require some metric information.

This brief proposes a new general intensity-based nonmetric visual servoing technique that decouples the error dynamics. Indeed, the decoupling idea presented in [25] for planar surfaces is here developed and extended to general objects, camera displacements, and their relative poses as in the DVS. In this proposed Decoupled DVS, the translational control error dynamics is decoupled from the rotational one. The obtained system is thus lower-triangular in the general case. For some practical cases, this paper also shows that the proposed technique leads to the Grail of a fully decoupled (i.e., strictly diagonal) linear dynamics. In these cases, each control error element is, hence, linearly related to only one DoF. The new nonmetric control error is theoretically proved to be locally diffeomorphic to the camera pose, and the latter is proved to be locally exponentially stable. All of those decoupling properties and improvements are also experimentally confirmed using a camera-mounted six-DoF robotic arm in various tasks. This paper is a revised and extended version of [26]. Indeed, it clarifies several aspects, presents and discusses many new positioning experiments in distinct configurations and relative to different objects, as well as includes a new tracking task of a fast freely-moving target. Additionally, another practical case of full decoupling is described and also experimentally demonstrated. These revisions and additions better highlight the advantages of this new vision-based control technique.

II. THEORETICAL BACKGROUND

This section defines the notation used throughout this brief, and recalls essential models and methods. Let $\|\mathbf{v}\|$ and $\mathbf{v}'$ denote the Euclidean norm and a transformed version of the variable $\mathbf{v}$, respectively. An asterisk, e.g., $\mathbf{v}^*$, is used to indicate that $\mathbf{v}$ is defined with respect to the reference frame $\mathcal{F}^*$. The identity matrix and the matrix composed of zeros, all of appropriate dimensions, are written $\mathbf{I}$ and $\mathbf{0}$, respectively. The notations $[\boldsymbol{w}]_\times$ and $\mathrm{vex}([\boldsymbol{w}]_\times)$ represent, respectively, the antisymmetric matrix associated to the vector
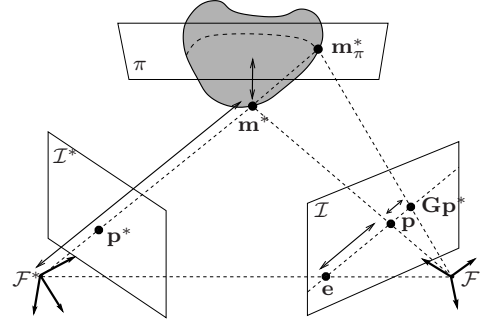

Figure 1. Two-view geometry. Given the 3-D point $\mathbf{m}^*$, its projection $\mathbf{p}^*$ on the image $\mathcal{I}^*$ is related to its projection $\mathbf{p}$ on $\mathcal{I}$ by the point $\mathbf{G}\mathbf{p}^*$, and the point $\mathbf{e}$ multiplied by the projective parallax (2).

$\boldsymbol{w} = \{w_i\}_{i=1}^3 = [w_1, w_2, w_3]^\top$ and its inverse mapping, i.e.,

$$[\boldsymbol{w}]_\times = \begin{bmatrix} 0 & -w_3 & w_2 \\ w_3 & 0 & -w_1 \\ -w_2 & w_1 & 0 \end{bmatrix}, \ \mathrm{vex}([\boldsymbol{w}]_\times) = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}. \quad (1)$$

A. Two-view Nonmetric Geometry

The general relation between corresponding points $\mathbf{p}_i \leftrightarrow \mathbf{p}_i^* \in \mathbb{P}^2$, $i = 1, 2, \ldots, n$, in two perspective images can be described in different geometries. In the projective geometry (which is an extension of the metric stratum), that general relation is written in homogeneous coordinates as [27]

$$\mathbf{p}_i \propto \mathbf{G}\,\mathbf{p}_i^* + \rho_i^*\,\mathbf{e}, \quad (2)$$

where the symbol "$\propto$" denotes proportionality, $\mathbf{G} \in \mathbb{SL}(3)$ is a projective homography relative to a (dominant) plane, $\mathbf{e} \in \mathbb{R}^3$ denotes the epipole on the current image $\mathcal{I}$, and $\rho_i^* \in \mathbb{R}$ is the projective parallax of the 3D point whose projection on the reference image $\mathcal{I}^*$ is $\mathbf{p}_i^*$, relatively to that plane (see Fig. 1). The homography $\mathbf{G}$ can be characterized as

$$\mathbf{G} \propto \mathbf{G}_\infty + \mathbf{e}\,\mathbf{q}^{*\top}, \quad (3)$$

where $\mathbf{q}^* \in \mathbb{R}^3$ is a representation of the image of the line at infinity of such plane, and $\mathbf{G}_\infty \in \mathbb{SL}(3)$ is the homography at infinity. A procedure to estimate all parameters $\{\mathbf{G}, \mathbf{e}, \rho_i^*\}_{i=1}^n$ that relate corresponding pixels (2) is recalled next.

B. Intensity-based Parametric Estimation

Intensity-based estimation exploits the pixel intensities with no intermediate steps, e.g., no feature extraction or matching. Its basic framework is the direct image registration, which consists in seeking the parameters that best transform the current image such that each pixel intensity $\mathcal{I}(\mathbf{p})$ in the transformed image $\mathcal{I}'$ matches as close as possible to the corresponding one in the reference image $\mathcal{I}^*(\mathbf{p}^*)$. Hence, it also amounts to tracking the scene/object between images.

Therefore, a first step consists in devising a suitable photogeometric transformation model. This can be performed by choosing an appropriate model of illumination changes, e.g., [5], along with a warping model $\mathbf{w}(\cdot)$ from (2). The action of such transformation model on pixels is given by [5]

$$\mathcal{I}'(\mathbf{g}, \mathbf{h}, \mathbf{p}^*) = \mathcal{S}(\mathbf{p}^*) \cdot \mathcal{I}\big(\mathbf{w}(\mathbf{g}, \mathbf{p}^*)\big) + \beta \quad \geq 0, \quad (4)$$

where the operator "·" represents elementwise matrix multiplication, and the geometric and photometric parameters are respectively gathered in $\mathbf{g} = \{\mathbf{G}, \mathbf{e}, \rho_i^*\}_{i=1}^n$ and $\mathbf{h} = \{\mathcal{S}, \beta\}$, where $\mathcal{S}$ can be viewed as a surface that compensates for both global and local illumination variations, and $\beta \in \mathbb{R}$ accounts for global changes only.

A typical direct image registration system can then be formulated as the following nonlinear optimization problem:

$$\min_{\mathbf{g}, \mathbf{h}} \frac{1}{2} \sum_{i=1}^n \left[\mathcal{I}'(\mathbf{g}, \mathbf{h}, \mathbf{p}_i^*) - \mathcal{I}^*(\mathbf{p}_i^*) \right]^2, \qquad (5)$$

which seeks to minimize the norm of the vector of image differences, where n is the number of pixels considered for exploitation. Other cost functions may be considered, e.g., an M-estimator [28] if there exist unknown occlusions. Finally, that optimization problem (5) can be solved by standard iterative methods. e.g., Gauss–Newton. For an improved solution in terms of convergence properties, the reader is referred to [5].

C. Direct Visual Servoing

The geometric parameters $\mathbf{g} = \{\mathbf{G}, \mathbf{e}, \rho_i^*\}_{i=1}^n$ estimated using pixel intensities directly (see Section II-B) can be used to visual servoing in robotics. The translational and rotational nonmetric control errors proposed in the Direct Visual Servoing [16], i.e., $\boldsymbol{\varepsilon}_v \in \mathbb{R}^3$ and $\boldsymbol{\varepsilon}_\omega \in \mathbb{R}^3$ respectively, are given as

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_v \\ \boldsymbol{\varepsilon}_\omega \end{bmatrix} = \begin{bmatrix} (\mathbf{H} - \mathbf{I})\mathbf{m}^{*\prime} + \rho^* \mathbf{e}' \\ \vartheta \boldsymbol{\mu} \end{bmatrix}, \qquad (6)$$

where

$$\mathbf{H} = \mathbf{K}^{-1}\mathbf{G}\mathbf{K}, \quad \mathbf{e}' = \mathbf{K}^{-1}\mathbf{e}, \quad \mathbf{m}^{*\prime} = \mathbf{K}^{-1}\mathbf{p}^*, \qquad (7)$$

and $\rho^* \in \mathbb{R}$ is the parallax of the chosen control point $\mathbf{p}^* \in \mathbb{P}^2$. The positive definite matrix $\mathbf{K} \in \mathbb{R}^{3\times3}$ contains the camera intrinsic parameters, i.e., focal lengths, skew and principal point. Even for nonmetric techniques, at least an estimate of such matrix is needed to control all six DoF of a robot. The rotational error $\boldsymbol{\varepsilon}_\omega$ in (6) is computed from the homography $\mathbf{H} \in \mathbb{R}^{3\times3}$ in (7) via

$$\boldsymbol{r} = \frac{1}{2}\text{vex}(\mathbf{H} - \mathbf{H}^\top), \quad \vartheta = \begin{cases} \psi, & \text{if } \text{tr}(\mathbf{H}) \geq 1 \\ \pi - \psi, & \text{otherwise} \end{cases}, \quad \boldsymbol{\mu} = \frac{\boldsymbol{r}}{\|\boldsymbol{r}\|}, \qquad (8)$$

where $\psi = \text{real}\big(\arcsin(\|\boldsymbol{r}\|)\big)$, and the function $\text{tr}(\cdot)$ denotes the trace of a matrix. If $\|\boldsymbol{r}\| = 0$, then $\boldsymbol{\mu}$ is not determined and therefore can be chosen arbitrarily (e.g., $\boldsymbol{\mu} = [0, 0, 1]^\top$). As shown in [29], the linearization of the interaction matrix of $\boldsymbol{\varepsilon}$ (6) at the equilibrium $\mathbf{0}$ is given as

$$\mathbf{L}\big|_{\boldsymbol{\varepsilon}=\mathbf{0}} = -\begin{bmatrix} \dfrac{1}{z^*}\mathbf{I} & -[\mathbf{m}^{*\prime}]_\times \\ \dfrac{1}{2}[\mathbf{q}^{*\prime}]_\times & \mathbf{I} \end{bmatrix}, \qquad (9)$$

where $z^* > 0$ is the metric depth of the chosen control point, and $\mathbf{q}^{*\prime} = \mathbf{K}^\top \mathbf{q}^*$. This matrix reveals that the dynamics of $\boldsymbol{\varepsilon}$ is fully coupled. Let us note that (9) is used only for analysis purposes, i.e., it is not needed during the servoing.

Indeed, let the control inputs be the translational and rotational velocities of the camera, gathered in $\mathbf{v} = [\boldsymbol{v}^\top, \boldsymbol{\omega}^\top]^\top \in \mathbb{R}^6$ respectively. The nonmetric control law

$$\mathbf{v} = \boldsymbol{\Gamma}\boldsymbol{\varepsilon}, \qquad (10)$$

with $\boldsymbol{\Gamma} = \text{diag}(\gamma_v \mathbf{I}, \gamma_\omega \mathbf{I}), \gamma_v, \gamma_\omega > 0$, is proven in [16] to locally stabilize the equilibrium $\boldsymbol{\varepsilon} = \mathbf{0}$ if the control point (7) is chosen such that its parallax ρ^* is sufficiently small.

III. PROPOSED DECOUPLED DIRECT VISUAL SERVOING

This section presents the Decoupled Direct Visual Servoing technique, which extends some results in [25] and [16]. This new intensity-based nonmetric technique decouples the translational control error dynamics from the rotational one, regardless of the object characteristics, camera displacements, and their relative poses. Theoretical analysis for the general case is provided in this section, whereas the next one shows that fully decoupling is realized for some practical cases.

A. Control Error and Some Properties

As in the Direct Visual Servoing [16], the new control error is constructed from the estimated parameters $\mathbf{g} = \{\mathbf{G}, \mathbf{e}, \rho_i^*\}_{i=1}^n$ (see Section II-B), and is defined next.

Definition 3.1: The new *nonmetric control error* $\bar{\boldsymbol{\varepsilon}} \in \mathbb{R}^6$ is

$$\bar{\boldsymbol{\varepsilon}} = \mathbf{M}\boldsymbol{\varepsilon}', \qquad (11)$$

where

$$\mathbf{M} = \begin{bmatrix} 2\mathbf{I} & [\mathbf{m}^{*\prime}]_\times \\ -[\mathbf{c}^{*\prime}]_\times & \mathbf{I} \end{bmatrix} \qquad (12)$$

is a constant (6×6)-matrix, $\mathbf{c}^{*\prime}$ is a control 3-vector, and

$$\boldsymbol{\varepsilon}' = \begin{bmatrix} \boldsymbol{\varepsilon}_v' \\ \boldsymbol{\varepsilon}_\omega' \end{bmatrix} = \begin{bmatrix} (\mathbf{H} - \mathbf{I})\mathbf{m}^{*\prime} + \rho^* \mathbf{e}' \\ \text{vex}(\mathbf{H} - \mathbf{H}^\top) \end{bmatrix} \qquad (13)$$

is a reduced version of the general control error $\boldsymbol{\varepsilon}$ in (6) as the rotational control error is equivalent to the original one via $\boldsymbol{\varepsilon}_\omega' = 2\boldsymbol{r} = 2\vartheta^{-1}\|\boldsymbol{r}\|\boldsymbol{\varepsilon}_\omega$. Furthermore, $\boldsymbol{\varepsilon}_\omega' \approx 2\boldsymbol{\varepsilon}_\omega$ as $\vartheta^{-1}\|\boldsymbol{r}\| \approx 1$ around the equilibrium. Details on choosing the control vector $\mathbf{c}^{*\prime}$ as well as the normalized control point $\mathbf{m}^{*\prime}$ (7) will be given further on (especially in Corollary 3.1).

Remark 3.1: The new control error (11) is constructed without requiring any metric information of the observed object, regardless of its characteristics (e.g., shape, size), of the camera displacements, and of their relative poses. Moreover, all error elements are either control parameters or are estimated using an intensity-based technique.

Theorem 3.1: The control error $\bar{\boldsymbol{\varepsilon}}$ defined in (11) is locally diffeomorphic to the camera pose around the reference image provided that the following conditions are satisfied:

$$\mathbf{m}^{*\prime\top}\mathbf{c}^{*\prime} \neq 2, \qquad (14)$$

$$\mathbf{m}^{*\prime\top}\mathbf{q}^{*\prime} \neq \frac{2}{z^*}. \qquad (15)$$

Furthermore, the linearization of the interaction matrix of $\bar{\boldsymbol{\varepsilon}}$ (11) at this configuration is given by

$$\bar{\mathbf{L}}\big|_{\bar{\boldsymbol{\varepsilon}}=\mathbf{0}} = -\begin{bmatrix} \dfrac{2}{z^*}\mathbf{I} + [\mathbf{m}^{*\prime}]_\times[\mathbf{q}^{*\prime}]_\times & \mathbf{0} \\ -\dfrac{1}{z^*}[\mathbf{c}^{*\prime}]_\times + [\mathbf{q}^{*\prime}]_\times & 2\mathbf{I} + [\mathbf{c}^{*\prime}]_\times[\mathbf{m}^{*\prime}]_\times \end{bmatrix}. \qquad (16)$$

Proof: The proof is presented in [26]. ∎

A nice property of such interaction matrix (16) is its block-triangular structure. This property is exploited next to derive simple stabilizing feedback laws with nice decoupling properties. Let us note that this interaction matrix will be used only for analysis purposes, i.e., it will not be needed for the control computation. Thus, no estimate of the metric depth z^* will be needed.

B. Control Law and Stability Analysis

In the sequel, consider a camera-mounted six-degree-of-freedom holonomic robot observing a motionless rigid object of unknown shape, size, texture, depth and orientation. Let the control inputs be the translational and rotational velocities of the camera, gathered in the vector $\mathbf{v} = [\boldsymbol{v}^\top, \boldsymbol{\omega}^\top]^\top \in \mathbb{R}^6$.

Definition 3.2: The new *nonmetric control law* is

$$\mathbf{v} = \boldsymbol{\Lambda}\,\bar{\boldsymbol{\varepsilon}}, \tag{17}$$

with a diagonal gain matrix $\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{\Lambda}_v, \boldsymbol{\Lambda}_\omega)$, and $\boldsymbol{\Lambda}_v = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$, $\boldsymbol{\Lambda}_\omega = \mathrm{diag}(\lambda_4, \lambda_5, \lambda_6)$.

Remark 3.2: Following the properties of the control error (see Remark 3.1), the proposed control law (17) is constructed without requiring any metric information of the observed object, and any linearization.

Theorem 3.2: The nonmetric control law (17) ensures local exponential stability of the equilibrium $\bar{\boldsymbol{\varepsilon}} = \mathbf{0}$ provided that the following conditions are satisfied:

$$\mathrm{sign}(\lambda_1) = \mathrm{sign}(\lambda_2) = \mathrm{sign}\left(\frac{2}{z^*} - \mathbf{m}^{*\prime\top}\mathbf{q}^{*\prime}\right), \tag{18}$$

$$\mathrm{sign}(\lambda_4) = \mathrm{sign}(\lambda_5) = \mathrm{sign}(2 - \mathbf{m}^{*\prime\top}\mathbf{c}^{*\prime}), \tag{19}$$

$$\lambda_3, \lambda_6 > 0. \tag{20}$$

Proof: The proof is developed in [26]. ∎

Corollary 3.1: An immediate consequence of (18)–(20) is that local exponential stability of $\bar{\boldsymbol{\varepsilon}} = \mathbf{0}$ is guaranteed $\forall \lambda_i > 0$, $i = 1, 2, \ldots, 6$, if the normalized control point $\mathbf{m}^{*\prime}$ and control vector $\mathbf{c}^{*\prime}$ are chosen such that

$$\mathbf{m}^{*\top}\mathbf{q}^{*\prime} < 2, \tag{21}$$

$$\mathbf{m}^{*\prime\top}\mathbf{c}^{*\prime} < 2. \tag{22}$$

Let us remind that $\mathbf{m}^{*\prime}$ is computed from the chosen image pixel $\mathbf{p}^*$ using (7). In fact, both conditions can always be verified as follows. The condition (21) expresses the perpendicular distance between the dominant plane and the chosen control point since $\mathbf{m}^* = z^*\mathbf{m}^{*\prime}$. This condition can then be easily satisfied if the control point $\mathbf{p}^*$ (and hence $\mathbf{m}^{*\prime}$) is chosen such that its parallax ρ^* is sufficiently small. In fact, one could simply choose a point that has $\rho^* = 0$ as the dominant plane crosses the object. This condition also holds in the DVS [16], where more details on choosing it are provided (after its Corollary 3.2). As for condition (22), it can be easily satisfied by setting, e.g., $\mathbf{c}^{*\prime} = \beta\mathbf{m}^{*\prime}/\|\mathbf{m}^{*\prime}\|^2$, $\forall \beta < 2$. Therefore, from (11) and (17) the obtained closed-loop system

$$\dot{\bar{\boldsymbol{\varepsilon}}} = \bar{\mathbf{L}}\big|_{\bar{\boldsymbol{\varepsilon}}=\mathbf{0}}\,\mathbf{v} = \bar{\mathbf{L}}\big|_{\bar{\boldsymbol{\varepsilon}}=\mathbf{0}}\,\boldsymbol{\Lambda}\,\bar{\boldsymbol{\varepsilon}} \tag{23}$$

can always be made locally exponentially stable around the equilibrium $\forall \lambda_i > 0$, $i = 1, 2, \ldots, 6$, independently of z^*.

Remark 3.3: Equation (23) shows that the proposed technique decouples the control error dynamics. Moreover, let us note that this is obtained regardless of the object characteristics, of the camera displacements, and of their relative poses.

IV. SOME PARTICULAR CASES OF FULL DECOUPLING

As stated in Remark 3.3, the obtained closed-loop system (23) is lower-triangular in the general case. This section presents some particular cases of special interest. Indeed, they occur often in practice and such dynamics becomes strictly diagonal, i.e., a fully decoupled system is obtained. In these cases, not only the translational control error dynamics is decoupled from the rotational one and vice-versa, but also each control error element is independently regulated. Once again, let us remark that the obtained interaction matrices are used only for analysis purposes, i.e., they are not needed to compute the proposed control law (17).

A. Fronto-parallel Planar Object

Consider a planar object such that its scaled normal vector is $\mathbf{n}^* = [0, 0, 1/d^*]^\top$, i.e., a plane fronto-parallel to the reference frame whose perpendicular distance is of $d^* > 0$. Let us choose the control point such that $\mathbf{m}^{*\prime} = [0, 0, 1]^\top$ and set $\mathbf{c}^{*\prime} = [0, 0, 1]^\top$. As consequences, $1/z^* = \mathbf{n}^{*\top}\mathbf{m}^{*\prime} = 1/d^*$, $\mathbf{q}^{*\prime} = \mathbf{n}^*$, the stability conditions (21) and (22) are both satisfied, and the obtained interaction matrix (16) writes

$$\bar{\mathbf{L}}\big|_{\bar{\boldsymbol{\varepsilon}}=\mathbf{0}} = -\left[\begin{array}{cc} \mathrm{diag}\left(\dfrac{1}{z^*}, \dfrac{1}{z^*}, \dfrac{2}{z^*}\right) & \mathbf{0} \\ \mathbf{0} & \mathrm{diag}(1, 1, 2) \end{array}\right]. \tag{24}$$

Therefore, a fully decoupled linear system (23) with strictly negative diagonal elements is obtained $\forall \lambda_i > 0$, $i = 1, 2, \ldots, 6$, i.e.,

$$\dot{\bar{\boldsymbol{\varepsilon}}} = -\mathrm{diag}\left(\frac{\lambda_1}{z^*}, \frac{\lambda_2}{z^*}, \frac{2\lambda_3}{z^*}, \lambda_4, \lambda_5, 2\lambda_6\right)\bar{\boldsymbol{\varepsilon}}. \tag{25}$$

B. Pure Rotation

This case corresponds to a particular camera displacement between the reference and current poses, regardless of the object characteristics and camera-object poses. Let us choose $\mathbf{c}^{*\prime} = \mathbf{0}$. As consequences, $\mathbf{q}^{*\prime} = \mathbf{0}$ (because of the displacement), the stability conditions (21) and (22) are both satisfied $\forall \mathbf{m}^{*\prime} \in \mathbb{R}^3$, and the obtained interaction matrix (16) writes

$$\bar{\mathbf{L}}\big|_{\bar{\boldsymbol{\varepsilon}}=\mathbf{0}} = -2\left[\begin{array}{cc} \dfrac{1}{z^*}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{array}\right]. \tag{26}$$

Once again, a fully decoupled linear system (23) with strictly negative diagonal elements is obtained $\forall \lambda_i > 0$, $i = 1, 2, \ldots, 6$, i.e.,

$$\dot{\bar{\boldsymbol{\varepsilon}}} = -2\,\mathrm{diag}\left(\frac{\lambda_1}{z^*}, \frac{\lambda_2}{z^*}, \frac{\lambda_3}{z^*}, \lambda_4, \lambda_5, \lambda_6\right)\bar{\boldsymbol{\varepsilon}}. \tag{27}$$

V. Experimental Positioning Results

This section reports experimental results using a 6-DoF robotic arm in different situations. In all cases, the control objective consists in positioning the robot such that the current object image coincides with its image captured at the reference pose, called reference image. The control signals do not use metric information, nor extract or match any image feature. Indeed, pixel intensities are directly exploited to estimate the needed parameters. This is performed via direct image registration, which ends up in tracking the object between images (see Section II-B). Although the entire image could be exploited, only an image region (i.e., a template) is processed because of the real time constraints. For the vast majority of the experiments, the template is defined as a rectangle in the middle of the reference image, and its center is chosen as $\mathbf{m}^{*\prime}$. The used CPU runs at 2.8GHz with 4GB RAM, and the applied direct registration algorithms are given in [5], [30].

Comparison results are presented using Direct Visual Servoing (DVS) and the proposed Decoupled DVS. The considered setups do not aim to optimize or compare their rate of convergence, but rather to verify the decoupling properties of the latter. Thus, their control gains are set such that the norms of their control signals are approximately equal at the beginning of the tasks. The norms of the initial displacements vary from 3 to 48 cm in translation, where the latter is the double of the perpendicular depth, and from 27 to 57 degrees in rotation. If those gains are equal, the rate of convergence for the Decoupled DVS is much faster, as its translational dynamics is decoupled from the rotational one.

Unless otherwise stated, the equipments and configurations are identical for all experiments. A coarsely calibrated camera is used, i.e., the hand-eye displacement is measured by hand, the focal length comes from the camera specifications, no skew, and the principal point is assumed to be at the center of the image, which has 640×480 pixels and is captured at 30Hz. In spite of all perturbations (coarse calibrations and large initial displacements) and unknowns about the objects (depth, orientation, size, texture), the positioning is successfully performed for all experiments. The stop condition is reached when the control error norm is less than 10^{-3}. Final Cartesian errors are in submillimeter amounts in translation and in subarcminute in rotation. These characteristics show that the proposed approach possesses a large region of attraction, is highly accurate and versatile. Table I summarizes the three considered positioning setups, which are described next.

A. Generically-placed Planar Object

In this first set of positioning experiments, the observed object is planar and is oriented generically with respect to the image plane. Figure 2 shows this experimental setup, named Setup A. Further details are given in Table I. In particular, let us note the relevant displacement between the reference and initial poses. Its norm of translation and rotation are of 48 cm (about double the perpendicular depth) and 57°, respectively.

1) DVS results: Figure 3 shows the results obtained using the DVS technique for Setup A. The positioning task is successfully performed with no prior knowledge of any

Table I
Specific configurations for each positioning setup.

	Setup A	Setup B	Setup C
Image registration			
object shape	planar	planar	nonplanar
object orientation	general	fronto-parallel	general
template size (pixels)	420×260	420×260	250×250
Initial displacement			
norm of translation (mm)	477.4	477.4	33.3
norm of rotation (deg)	57.1	57.1	26.9
DVS control			
translational gains	0.4	0.3	0.1
rotational gains	0.8	0.8	0.4
Decoupled DVS control			
translational gains	0.1	0.1	0.1
rotational gains	0.45	0.4	0.4
control vector $\mathbf{c}^{*\prime}$	$[0,0,1]^\top$	$[0,0,1]^\top$	$[0,0,0]^\top$

(a) reference pose   (b) initial pose
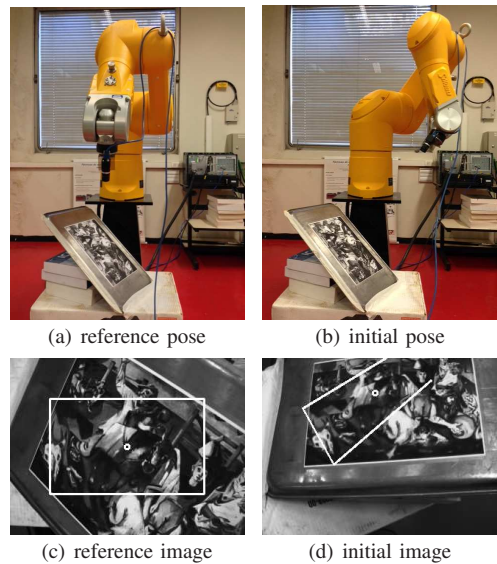
(c) reference image   (d) initial image

Figure 2. Setup A: Generically-placed planar object. (Top) Poses of the camera-mounted robotic arm with respect to the object. (Bottom) Respective images. All pixels within the outlined template are exploited.

metric information of the object. However, a coupling between the translational velocity and the rotational one (they are proportional to the respective control error) can be observed in Fig. 3(a), in particular for the v_y component.

2) Decoupled DVS results: The results obtained using the Decoupled DVS technique for Setup A are given in Fig. 4, and on video (supplementary material) as well. The positioning task is also successfully performed, and the translational velocities are now decoupled from the rotational ones as shown in Fig. 4(a). Let us observe that the rate of convergence is improved. Indeed, although the control gains are much smaller than in the DVS case, the settling time is about the same.

B. Fronto-parallel Planar Object

This second set of positioning experiments comprises the same object and robot displacement between the reference and initial poses as in Setup A. However, the planar object is now fronto-parallel to the image plane at the reference pose. This new configuration obviously changes the system behavior.
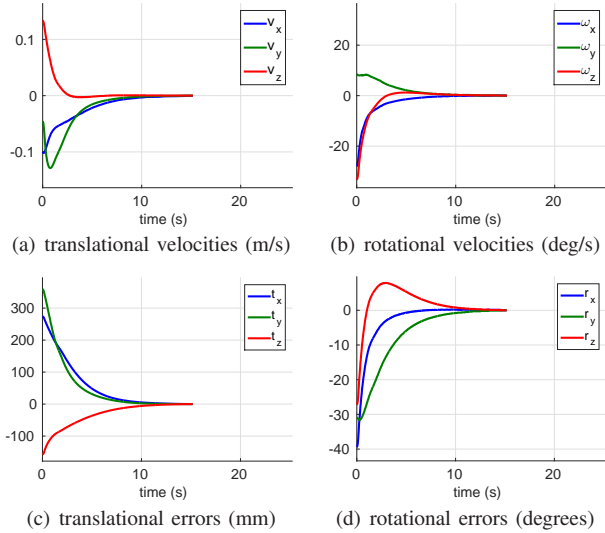
Figure 3. DVS results for the Setup A: Generically-placed planar object. (Top) Control signals. (Bottom) End-effector motion in the Cartesian space.
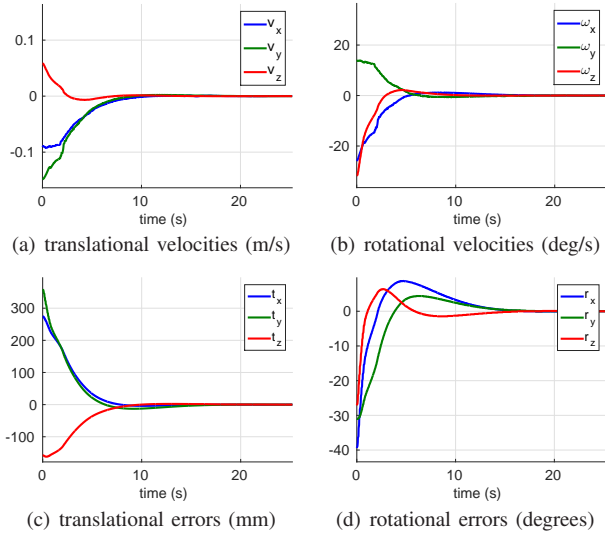

Figure 4. Decoupled DVS results for the Setup A: Generically-placed planar object. (Top) Control signals. A decoupled behavior is observed in 4(a). (Bottom) End-effector motion in the Cartesian space.

(a) reference pose   (b) initial pose
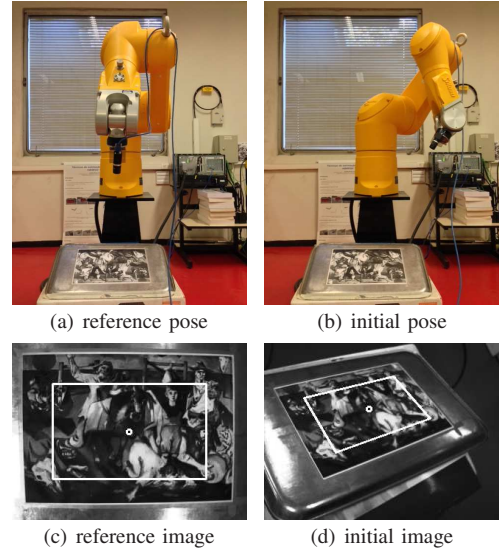
(c) reference image   (d) initial image

Figure 5. Setup B: Fronto-parallel planar object. (Top) Poses of the camera-mounted robotic arm with respect to the object. (Bottom) Respective images. All pixels within the outlined template are exploited.
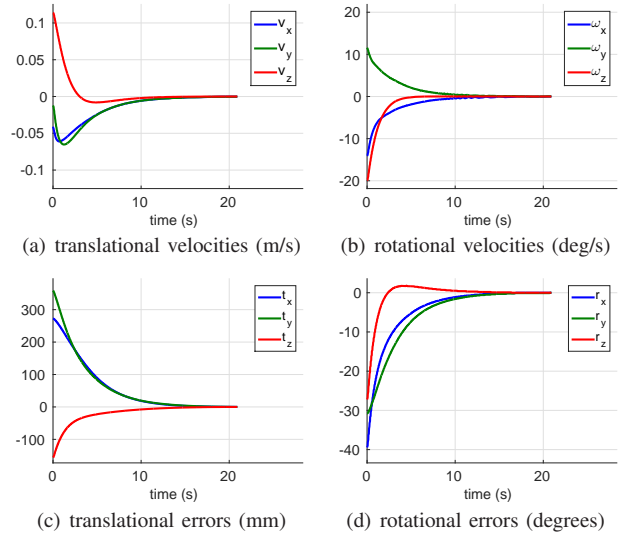

Figure 6. DVS results for the Setup B: Fronto-parallel planar object. (Top) Control signals. (Bottom) End-effector motion in the Cartesian space.

Figure 5 shows this experimental setup, named Setup B. Further details are given in Table I. In particular, the control gains are modified so as to equalize the norms of the control signals for both techniques at the beginning of the task.

1) DVS results: Figure 6 shows the results obtained using the DVS technique for Setup B. Again, the positioning task is successfully performed with no prior knowledge of any metric information of the object. However, a coupling between the translational velocity and the rotational one can be observed in Fig. 6(a), in both v_x and v_y components.

2) Decoupled DVS results: The results obtained using the Decoupled DVS technique for Setup B are given in Fig. 7, and on video (supplementary material) as well. The positioning task is also successfully performed, and the translational velocities are now decoupled from the rotational ones as shown in Fig. 7(a). Let us note that the rate of convergence is again improved. Indeed, although the control gains are much smaller

than in the DVS case, the settling time is about the same.

C. Mostly Rotation using a Nonplanar Object

This third set of positioning experiments comprises a different object and a distinct robot displacement between the reference and initial poses. Indeed, the observed object is here nonplanar (with different texture, size, etc.), and that displacement has relatively large rotational components, whereas the translation is small. Figure 8 shows this experimental setup, named Setup C. Further details are given in Table I. In particular, let us note that equal control gains are applied for both DVS and Decoupled DVS techniques.

1) DVS results: Figure 9 shows the results obtained using the DVS technique for Setup C. The task is successfully performed, but the coupling interferes in the translational velocity, imposing a nearly zero v_z and degrading performance.
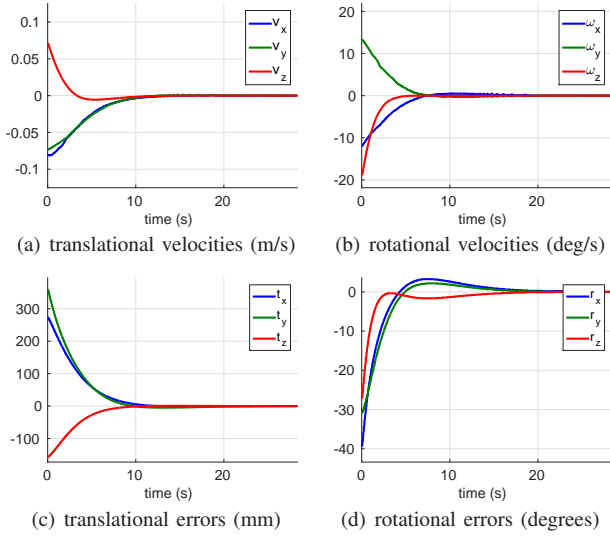
Figure 7. Decoupled DVS results for the Setup B: Fronto-parallel planar object. (Top) Control signals. A decoupled behavior is observed in Fig. 7(a). (Bottom) End-effector motion in the Cartesian space.
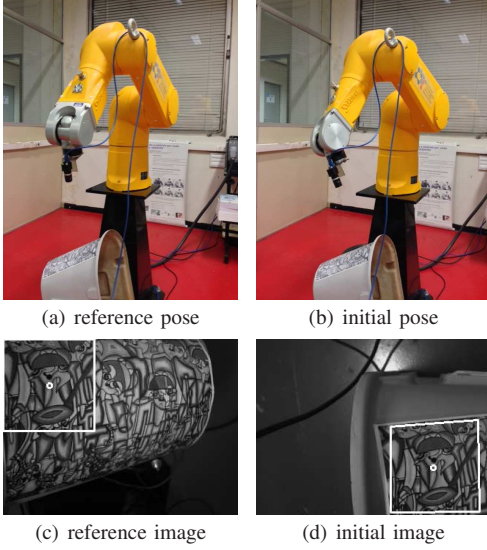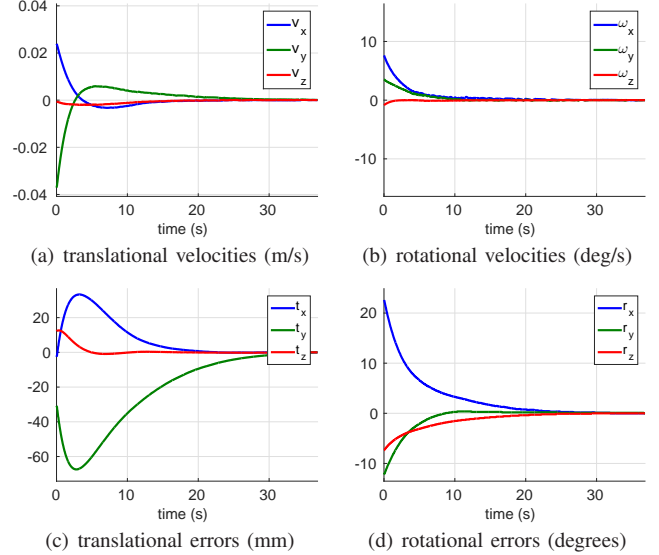

Figure 9. DVS results for the Setup C: Mostly rotation using a nonplanar object. (Top) Control signals. (Bottom) End-effector motion in Cartesian space.

Figure 8. Setup C: Mostly rotation using a nonplanar object. (Top) Poses of the camera-mounted robotic arm with respect to the object. (Bottom) Respective images. All pixels within the outlined template are exploited.
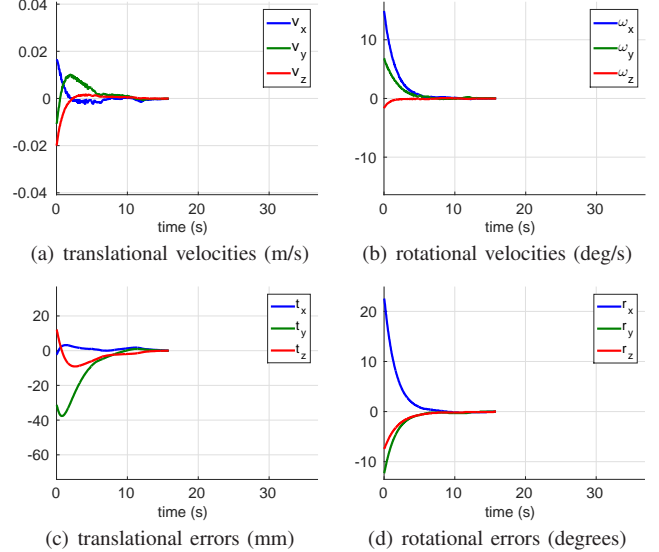

Figure 10. Decoupled DVS results for the Setup C: Mostly rotation using a nonplanar object. (Top) Control signals. (Bottom) End-effector motion in the Cartesian space. Much faster convergence is observed with respect to Fig. 9.

2) Decoupled DVS results: The results obtained using the Decoupled DVS technique for Setup C are given in Fig. 10, and on video (supplementary material) as well. The task is successfully performed with no interferences of the rotational motion in the translational one as shown in Fig. 10(a). Given its decoupling properties, the evolution of the translation error differs significantly between Figs. 9(c) and 10(c). The rate of convergence of the Decoupled DVS is in fact much higher relatively to the DVS, with a much lower settling time.

VI. EXPERIMENTAL TRACKING RESULTS

All aforementioned experiments have consisted of a positioning task with respect to a motionless object. To demonstrate the versatility and performance of the proposed Decoupled DVS, an experiment of tracking a fast freely-moving

object has been conducted as well. In this experiment the camera framerate is 120Hz, the template size has 320×240 pixels, and the control gains are $1.8 \exp(-0.9\,\|\bar{\varepsilon}(t)\|)$. These variable gains induce faster robot motion when the control errors are small. Let us remark that neither error integrators nor state observers are added to the system, what could still improve its performance. The obtained results are available in the accompanying video (supplementary material), where some excerpts are shown in Fig. 11. The control signals and norm of control errors are given in Fig. 12, where fast robot motion in all six DoF can be observed. Indeed, translational and rotational velocities of more than 0.7 m/s and 70 degrees/s, respectively, are successfully attained even for an observed object of unknown depth, orientation, size, texture, and undergoing disparate and large displacements.
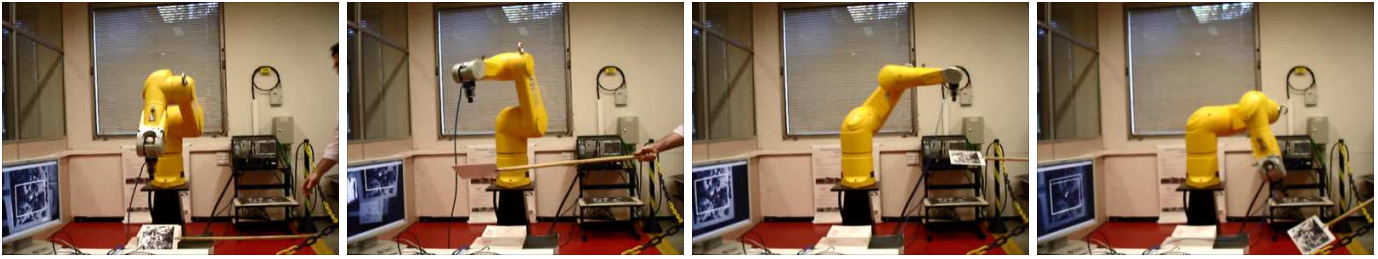
Figure 11. Excerpts of a tracking task using a camera-mounted robot and a fast freely-moving planar object of unknown depth, orientation, size, and texture.
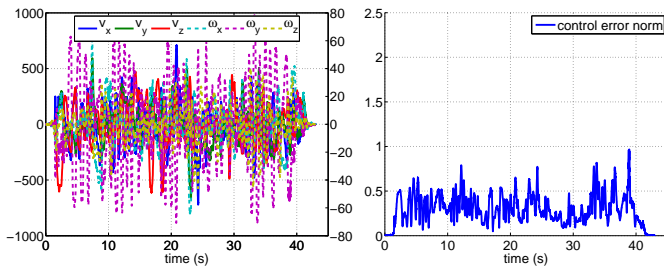

Figure 12. Control signals (left axis in mm/s and right one in degrees/s) and norm of control errors, respectively, for the tracking experiment shown in Fig. 11. Fast robot motion is observed in all six degrees of freedom.

VII. CONCLUSIONS

This brief has proposed a general intensity-based nonmetric visual servoing technique that decouples the control error dynamics. This new decoupled technique is general in the sense that all 6-DoF of a holonomic robot are stabilized regardless of the observed object characteristics, camera displacements, and their relative poses. The proposed strategy exploits the pixel intensities without extracting or matching image features, does not require any metric information of the object, and is proven to locally exponentially stabilize the equilibrium. Various comparison results with a state-of-the-art direct technique using a camera-mounted 6-DoF robotic arm confirm the improvements. Moreover, a tracking task of a fast freely-moving object has also been possible given its high rate of convergence and versatility. In future work we plan to extend the proposed framework to control other mechanical systems, such as nonholonomic and underactuated robots.

REFERENCES

[1] F. Chaumette and S. Hutchinson, "Visual servo control part I: Basic approaches," *IEEE Robotics & Automation Magazine*, pp. 82–90, 2006.
[2] A. De Luca, G. Oriolo, and P. Robuffo Giordano, "Feature depth observation for image-based visual servoing: Theory and experiments," *Int. Journal of Robotics Research*, vol. 27, no. 10, pp. 1093–1116, 2008.
[3] B. Li, Y. Fang, G. Hu, and X. Zhang, "Model-free unified tracking and regulation visual servoing of wheeled mobile robots," *IEEE Transactions on Control Systems Technology*, vol. 24, no. 4, pp. 1328–1339, 2016.
[4] M. Irani and P. Anandan, "All about direct methods," in *Workshop on Vision Algorithms: Theory and practice*, 1999.
[5] G. Silveira and E. Malis, "Unified direct visual tracking of rigid and deformable surfaces under generic illumination changes in grayscale and color images," *Int. Journal of Comp. Vision*, vol. 89, pp. 84–105, 2010.
[6] G. Silveira, "Photogeometric direct visual tracking for central omnidirectional cameras," *Journal of Mathematical Imaging and Vision*, 2012.
[7] ——, "On intensity-based 3D visual servoing," *Robotics and Autonomous Systems*, vol. 62, no. 11, pp. 1636–1645, 2014.
[8] A. Dame and E. Marchant, "Mutual information-based visual servoing," *IEEE Trans. on Robotics*, vol. 27, no. 5, pp. 958–969, 2011.
[9] S. Han, A. Censi, A. D. Straw, and R. M. Murray, "A bio-plausible design for visual pose stabilization," in *IEEE/RSJ IROS*, 2010.
[10] C. Zeller and O. Faugeras, "Applications of non-metric vision to some visual guided tasks," in *IEEE Intelligent Vehicles Symposium*, 1994.
[11] L. Thaler and M. A. Goodale, "Beyond distance and direction: The brain represents target locations non-metrically," *Journal of Vision*, vol. 10, no. 3, pp. 1–27, 2010.
[12] P. A. Beardsley, I. D. Reid, A. Zisserman, and D. W. Murray, "Active visual navigation using non-metric structure," in *IEEE International Conference on Computer Vision*, 1995, pp. 58–64.
[13] V. Kallem, M. Dewan, J. Swensen, G. Hager, and N. Cowan, "Kernel-based visual servoing," in *IEEE/RSJ IROS*, USA, 2007.
[14] M. Vargas and E. Malis, "Visual servoing based on an analytical homography decomposition," in *Joint IEEE CDC-ECC*, Spain, 2005.
[15] S. Benhimane and E. Malis, "Homography-based 2D visual servoing," in *IEEE ICRA*, USA, 2006.
[16] G. Silveira and E. Malis, "Direct Visual Servoing: Vision-based estimation and control using only nonmetric information," *IEEE Transactions on Robotics*, vol. 28, no. 4, pp. 974–980, 2012.
[17] G. Silveira, "On intensity-based nonmetric visual servoing," *IEEE Transactions on Robotics*, vol. 30, no. 4, pp. 1019–1026, 2014.
[18] E. Malis, F. Chaumette, and S. Boudet, "2D 1/2 visual servoing," *IEEE Trans. on Robotics and Automation*, vol. 15, no. 2, pp. 238–250, 1999.
[19] K. Deguchi, "Optimal motion control for image-based visual servoing by decoupling translation and rotation," in *IEEE/RSJ IROS*, Canada, 1998.
[20] H. Hadj-Abdelkader, Y. Mezouar, and P. Martinet, "Decoupled visual servoing based on the spherical projection of a set of points," in *IEEE ICRA*, 2009, pp. 1110–1115.
[21] R. T. Fomena and F. Chaumette, "Improvements on visual servoing from spherical targets using a spherical projection model," *IEEE Trans. on Robotics*, vol. 25, no. 4, pp. 874–886, 2009.
[22] R. Mahony, T. Hamel, and F. Chaumette, "A decoupled image space approach to visual servo control of a robotic manipulator," in *IEEE ICRA*, 2002, pp. 3781–3786.
[23] P. I. Corke and S. Hutchinson, "A new partitioned approach to image-based visual servo control," in *Trans. on Robotics and Automation*, vol. 17, no. 4, pp. 507–515, 2001.
[24] P. Rives and J. R. Azinheira, "Linear structures following by an airship using vanishing point and horizon line in a visual servoing scheme," in *IEEE ICRA*, 2004, pp. 255–260.
[25] H. de Plinval, P. Morin, P. Mouyon, and T. Hamel, "Visual servoing for underactuated VTOL UAVs: a linear, homography-based approach," in *IEEE ICRA*, 2011, pp. 3004–3010.
[26] G. Silveira, L. Mirisola, and P. Morin, "Decoupled direct visual servoing," in *IEEE/RSJ IROS*, Japan, 2013, pp. 71–76.
[27] O. Faugeras, Q.-T. Luong, and T. Papadopoulo, *The geometry of multiple images*. The MIT Press, 2001.
[28] P. J. Huber, *Robust Statistics*. John Wiley & Sons, 1981.
[29] G. Silveira and E. Malis, "Direct visual servoing with respect to rigid objects," INRIA, Research Report 6265, 2007.
[30] E. Malis *et al.*, "ESM basic visual tracking software," 2007, http://esm.gforge.inria.fr/ESM.html.