

Decoupled Direct Visual Servoing

Geraldo Silveira¹, Luiz Mirisola², and Pascal Morin³

Abstract— This article addresses the problem of direct vision-based robot control where the equilibrium state is defined via a reference image. Direct methods refer to intensity-based nonmetric techniques to perform that stabilization. Intensity-based strategies provide for higher accuracy, whereas not requiring any metric information improves their versatility. However, existing direct techniques either have a coupled error dynamics, or are designed for planar objects only. This paper proposes a new direct technique that decouples the translational motion from the rotational one for the general case of both planar and nonplanar targets under general translational and rotational displacements. Furthermore, for the important case of a fronto-parallel planar object, the proposed technique leads to a fully diagonal interaction matrix. The equilibrium state is made locally exponentially stable for all those cases. These improvements are theoretically proven and experimentally demonstrated using a 6-DoF robotic arm.

I. INTRODUCTION

Visual servoing refers to the use of image feedback to control a robot with respect to the scene. Its typical application consists in stabilizing the robot at a pose defined by means of a reference image, also called goal image. Although there exists a variety of well-established solutions to this problem [2], its vast majority: 1) is based on image features, such as points, lines, etc. Thus, they strongly depend on some particular features, on an error-prone feature matching, and on special tuning procedures; and 2) requires (at least coarse) metric information to provide a provably stabilizing control law. This holds even for image-based visual servoing techniques, where depth estimates are necessary in the interaction matrix. These two topics are discussed next.

Techniques of vision-based estimation 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. This is probably due to its relatively large domain of convergence. Differently, there are no steps of feature extraction and matching within intensity-based techniques of estimation. These techniques directly exploit the intensity value of the pixels so as to recover the needed parameters. Therefore, they make use of raw and dense image data, what allows for attaining high levels of versatility and accuracy. Another advantage refers to their possibility of ensuring robustness to arbitrary illumination changes, even in color images [7].

As for nonmetric visual servoing, in spite of its increased level of versatility and robustness [9], there exist only few works on such class of vision-based control. A possible reason is the difficulty to find an interesting control error that is diffeomorphic to the camera pose (at least around the

equilibrium), and is regulated by a nonmetric control law. An early work on nonmetric visual navigation is given in [1], where a ground robot is used. Recently, a general intensity-based nonmetric technique has been presented in [8], called Direct Visual Servoing (DVS). It is general in the sense that all 6 DoF of a robot are stabilized with respect to both planar and nonplanar objects, under both translational and rotational displacements between the reference and current frames. Nevertheless, the error dynamics in the DVS is coupled.

This article proposes a new decoupled DVS technique. The decoupling idea briefly presented in [3] for planar objects is here developed and extended to general surfaces and displacements as in [8]. The translational motion is thus decoupled from the rotational one in the general scenario. Furthermore, for the important case of a fronto-parallel planar object, it is shown that the proposed technique leads to a fully diagonal interaction matrix. The new control error is theoretically proved to be diffeomorphic to the camera pose around the equilibrium, and the latter is proved to be locally exponentially stable. These improvements are also confirmed with experiments using both synthetic and real data, for both planar and nonplanar objects, simulating and applying a camera-mounted 6-DoF holonomic robot.

II. THEORETICAL BACKGROUND

This section defines the notation used throughout this article and recalls essential models and methods. Let $\|\mathbf{v}\|$, $\hat{\mathbf{v}}$ and \mathbf{v}' denote the Euclidean norm, an estimate, 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 notations $[\mathbf{w}]_{\times}$ and $\text{vex}([\mathbf{w}]_{\times})$ represent, respectively, the antisymmetric matrix associated to the vector $\mathbf{w} = [w_1, w_2, w_3]^T$ and its inverse mapping:

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

A. Two-view Geometry

The relation between corresponding points $\mathbf{p} \leftrightarrow \mathbf{p}^*$ in two perspective images can be described in different geometries. Using projective geometry (which is an extension of Euclidean geometry), the general relation is given by [4]

$$\mathbf{p} \propto \mathbf{G} \mathbf{p}^* + \rho^* \mathbf{e} \in \mathbb{P}^2, \quad (2)$$

where the symbol “ \propto ” denotes proportionality up to a nonzero scale factor, $\mathbf{G} \in \mathbb{SL}(3)$ is a projective homography relative to a plane (projective basis), $\mathbf{e} \in \mathbb{R}^3$ denotes the epipole, and $\rho^* \in \mathbb{R}$ is the projective parallax of the 3D point whose projection in the reference image \mathcal{I}^* is \mathbf{p}^* , relatively to that plane (see Fig. 1). This parallax is proportional to the distance of that 3D point to that plane and is inversely

¹G. Silveira is with CTI, DRVC, Brazil, Geraldo.Silveira@cti.gov.br. This author has also been supported by the brazilian council CNPq.

²L. Mirisola is with UFABC, CMCC, Brazil, luz.mirisola@ufabc.edu.br

³P. Morin is with UPMC, ISIR, France, morin@isir.upmc.fr. This author has been supported by the “Chaire d’excellence en Robotique RTE-UPMC”.

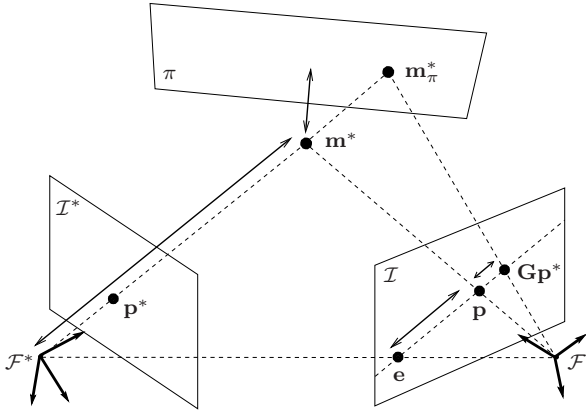


Fig. 1. Two-view geometry. Given the 3-D point \mathbf{m}^* , its projection \mathbf{p}^* in the image \mathcal{I}^* is related to its projection \mathbf{p} in \mathcal{I} by the point $\mathbf{G}\mathbf{p}^*$, and the point \mathbf{e} multiplied by the projective parallax. This parallax is indeed the ratio of the distances of the aligned points $\mathbf{G}\mathbf{p}^*$, \mathbf{e} and \mathbf{p} .

proportional to its depth. The epipole is proportional to the translation between the current and reference frames; and the homography \mathbf{G} can be characterized as

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

where the 3-vector \mathbf{q}^* is a representation of the line at infinity of such plane in the reference image, and $\mathbf{G}_\infty \in \text{SL}(3)$ is the homography at infinity. Such homography is proportional to a matrix that is similar to a rotation matrix. A procedure to estimate the parameters $\{\mathbf{G}, \mathbf{e}, \rho^*\}$ in (2) is recalled next.

B. Direct Visual Servoing: Estimation Aspects

In Direct Visual Servoing (DVS) [8], the estimation procedure to obtain all needed parameters directly exploits the pixel intensities without any feature extraction or matching. The basic framework for such intensity-based estimation is the direct image registration. Direct image registration consists in searching for the parameters that best transform the current image such that each pixel intensity in the current image $\mathcal{I}(\mathbf{p})$ is matched as closely as possible to the corresponding one in the reference image $\mathcal{I}^*(\mathbf{p}^*)$.

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., [7], along with a warping model $\mathbf{w}(\cdot)$ from (2). Formally, the action of this transformation model on pixels is given by

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

where the operator “ \cdot ” represents the componentwise multiplication of matrices, and the geometric and photometric parameters are respectively gathered in $\mathbf{g} = \{\mathbf{G}, \mathbf{e}, \rho^*\}$ 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 $\beta \in \mathbb{R}$.

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

$$\min_{\substack{\mathbf{g}=\{\mathbf{G}, \mathbf{e}, \rho^*\} \\ \mathbf{h}=\{\mathcal{S}, \beta\}}} \frac{1}{2} \sum_{i=1}^n \underbrace{[\mathcal{I}'_{gh}(\mathbf{g}, \mathbf{h}, \mathbf{p}_i^*) - \mathcal{I}^*(\mathbf{p}_i^*)]^2}_{d_i(\mathbf{g}, \mathbf{h})}, \quad (5)$$

which seeks to minimize the norm of the vector of image differences $\mathbf{d}(\mathbf{g}, \mathbf{h}) = \{d_i(\mathbf{g}, \mathbf{h})\}_{i=1}^n$, i.e.,

$$\mathbf{d}(\mathbf{g}, \mathbf{h}) = \begin{bmatrix} \mathcal{I}'_{gh}(\mathbf{g}, \mathbf{h}, \mathbf{p}_1^*) - \mathcal{I}^*(\mathbf{p}_1^*) \\ \mathcal{I}'_{gh}(\mathbf{g}, \mathbf{h}, \mathbf{p}_2^*) - \mathcal{I}^*(\mathbf{p}_2^*) \\ \vdots \\ \mathcal{I}'_{gh}(\mathbf{g}, \mathbf{h}, \mathbf{p}_n^*) - \mathcal{I}^*(\mathbf{p}_n^*) \end{bmatrix} \in \mathbb{R}^n, \quad (6)$$

where n is the number of pixels considered for exploitation. Other cost functions may be considered instead of that widely used sum-of-squared-differences. For example, a robust function [5], e.g., an M-estimator, may be used if there exist unknown occlusions. Finally, the nonlinear optimization problem in (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 [7].

C. Direct Visual Servoing: Control Aspects

The geometric parameters $\mathbf{g} = \{\mathbf{G}, \mathbf{e}, \rho^*\}$ estimated using pixel intensities (see Section II-B) can be used to visual servoing in robotics. The translational and rotational nonmetric control errors proposed in the DVS, i.e., $\varepsilon_v \in \mathbb{R}^3$ and $\varepsilon_\omega \in \mathbb{R}^3$ respectively, are given as

$$\varepsilon = \begin{bmatrix} \varepsilon_v \\ \varepsilon_\omega \end{bmatrix} = \begin{bmatrix} (\mathbf{H} - \mathbf{I})\mathbf{m}' + \rho^*\mathbf{e}' \\ \vartheta\boldsymbol{\mu} \end{bmatrix}, \quad (7)$$

where

$$\mathbf{H} = \mathbf{K}^{-1}\mathbf{G}\mathbf{K}; \quad \mathbf{e}' = \mathbf{K}^{-1}\mathbf{e}; \quad \mathbf{m}' = \mathbf{K}^{-1}\mathbf{p}^*, \quad (8)$$

and $\rho^* \in \mathbb{R}$ is the parallax of the chosen control point $\mathbf{p}^* \in \mathbb{P}^2$. The positive definite matrix $\mathbf{K} \in \mathbb{R}^{3 \times 3}$ 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 *always* needed to control all six degrees of freedom of a robot, as it moves in the Euclidean space. The rotational error ε_ω in (7) is computed from the homography $\mathbf{H} \in \mathbb{R}^{3 \times 3}$ via

$$\mathbf{r} = \frac{1}{2} \text{vex}(\mathbf{H} - \mathbf{H}^\top), \quad (9)$$

$$\vartheta = \begin{cases} \text{real}(\arcsin(\|\mathbf{r}\|)), & \text{if } \text{tr}(\mathbf{H}) \geq 1, \\ \pi - \text{real}(\arcsin(\|\mathbf{r}\|)), & \text{otherwise,} \end{cases} \quad (10)$$

$$\boldsymbol{\mu} = \frac{\mathbf{r}}{\|\mathbf{r}\|}, \quad (11)$$

where the function $\text{tr}(\cdot)$ denotes the trace of a matrix. If $\|\mathbf{r}\| = 0$, then $\boldsymbol{\mu}$ is not determined and therefore can be chosen arbitrarily (e.g., $\boldsymbol{\mu} = [0, 0, 1]^\top$).

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

$$\mathbf{v} = \lambda \varepsilon, \quad (12)$$

with $\lambda > 0$, is proven in [8] to locally stabilize the equilibrium $\varepsilon = \mathbf{0}$ if the control point (8) is chosen such that its parallax is sufficiently small.

III. PROPOSED DECOUPLED TECHNIQUE

This section presents the proposed decoupled direct visual servoing technique, which extends some results from [3]. This new technique is intensity-based, nonmetric, and the behavior of the translational motion is decoupled from the rotational one. This is demonstrated for the general case of both planar and nonplanar objects under general translational and rotational displacements. Then, we show that it leads to a fully decoupled system for the important case of a fronto-parallel planar target.

A. Control Error and Some Properties

As in the DVS [8], the new control error is constructed from the estimated parameters $\mathbf{g} = \{\mathbf{G}, \mathbf{e}, \rho^*\}$ (see Section II-B). For the sake of simplicity, let us work with a reduced version of the general control error in (7) as

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

whose rotational control error is equivalent to the original one via $\varepsilon'_\omega = 2\mathbf{r} = 2\vartheta^{-1}\|\mathbf{r}\|\boldsymbol{\varepsilon}_\omega$. Furthermore, around the equilibrium it can be shown that $\varepsilon'_\omega \approx 2\varepsilon_\omega$ as $\vartheta^{-1}\|\mathbf{r}\| \approx 1$. This nonmetric control error is general in the sense that it deals with both planar and nonplanar objects, under both translational and rotational displacements between the reference and initial frames. Indeed, as in the general relation (2), it does not assume $\rho^*\mathbf{e}' = \mathbf{0}$.

The new nonmetric control error is defined as

$$\bar{\boldsymbol{\varepsilon}} = \mathbf{M} \boldsymbol{\varepsilon}', \quad (14)$$

where

$$\mathbf{M} = \begin{bmatrix} 2\mathbf{I} & [\mathbf{m}^{*'}]_\times \\ -[\mathbf{c}^{*'}]_\times & \mathbf{I} \end{bmatrix} \quad (15)$$

is a constant (6×6) -matrix, and $\mathbf{c}^{*'}$ is a 3-vector. Details on choosing this vector will be given further on. More explicitly, the new control error (14) writes

$$\bar{\boldsymbol{\varepsilon}} = \begin{bmatrix} 2[(\mathbf{H} - \mathbf{I})\mathbf{m}^{*'} + \rho^*\mathbf{e}'] + [\mathbf{m}^{*'}]_\times \text{vex}(\mathbf{H} - \mathbf{H}^\top) \\ \text{vex}(\mathbf{H} - \mathbf{H}^\top) - [\mathbf{c}^{*'}]_\times [(\mathbf{H} - \mathbf{I})\mathbf{m}^{*'} + \rho^*\mathbf{e}'] \end{bmatrix}. \quad (16)$$

It is important to note that this control error is constructed without requiring any metric information of the object, regardless of its shape and of the camera displacement.

Theorem 3.1: The control error $\bar{\boldsymbol{\varepsilon}}$ defined in (14) is locally diffeomorphic to the camera pose around the reference image provided that

$$\mathbf{m}^{*'\top} \mathbf{c}^{*' } \neq 2, \quad (17)$$

$$\mathbf{m}^{*'\top} \mathbf{q}^{*' } \neq \frac{2}{z^*}, \quad (18)$$

where $z^* > 0$ is the depth of the control point \mathbf{p}^* , and $\mathbf{q}^{*' } = \mathbf{K}^\top \mathbf{q}^*$. Furthermore, the linearization of the interaction matrix of $\bar{\boldsymbol{\varepsilon}}$ at this configuration is given by

$$\bar{\mathbf{L}}|_{\bar{\boldsymbol{\varepsilon}}=\mathbf{0}} = - \begin{bmatrix} \frac{2}{z^*} \mathbf{I} + [\mathbf{m}^{*'}]_\times [\mathbf{q}^{*' }]_\times & \mathbf{0} \\ -\frac{1}{z^*} [\mathbf{c}^{*'}]_\times + [\mathbf{q}^{*' }]_\times & 2\mathbf{I} + [\mathbf{c}^{*'}]_\times [\mathbf{m}^{*'}]_\times \end{bmatrix}. \quad (19)$$

Proof: Since $\bar{\boldsymbol{\varepsilon}} = \mathbf{M} \boldsymbol{\varepsilon}'$ and \mathbf{M} is a constant matrix, $\bar{\mathbf{L}} = \mathbf{M} \mathbf{L}'$ with $\bar{\mathbf{L}}$ (resp. \mathbf{L}') the interaction matrix of $\bar{\boldsymbol{\varepsilon}}$ (resp. $\boldsymbol{\varepsilon}'$). From [6], the linearization of \mathbf{L}' at $\boldsymbol{\varepsilon}' = \mathbf{0}$ is given as

$$\mathbf{L}'|_{\boldsymbol{\varepsilon}'=\mathbf{0}} = - \begin{bmatrix} \frac{1}{z^*} \mathbf{I} & -[\mathbf{m}^{*'}]_\times \\ [\mathbf{q}^{*' }]_\times & 2\mathbf{I} \end{bmatrix}, \quad (20)$$

and (19) follows from (15) and (20). To conclude the proof, there remains to show that the matrix (19) is invertible when both (17) and (18) are satisfied. Since $\bar{\mathbf{L}}|_{\bar{\boldsymbol{\varepsilon}}=\mathbf{0}} = \mathbf{M} \mathbf{L}'|_{\boldsymbol{\varepsilon}'=\mathbf{0}}$, $\bar{\mathbf{L}}|_{\bar{\boldsymbol{\varepsilon}}=\mathbf{0}}$ is invertible provided that both \mathbf{M} and $\mathbf{L}'|_{\boldsymbol{\varepsilon}'=\mathbf{0}}$ are invertible. Let us first consider \mathbf{M} . From Schur's formula, and (15), \mathbf{M} is invertible provided that the matrix

$$\mathbf{M}_0 = 2\mathbf{I} + [\mathbf{c}^{*'}]_\times [\mathbf{m}^{*'}]_\times \quad (21)$$

is invertible. Consider any matrix \mathbf{Q} such that $\mathbf{Q} \mathbf{m}^{*' } = \|\mathbf{m}^{*' }\| \mathbf{b}_3$ with $\mathbf{b}_3 = [0, 0, 1]^\top$ the third canonical vector. Using the fact that for any 3-vector \mathbf{x} , $\mathbf{Q}[\mathbf{x}]_\times \mathbf{Q}^\top = [\mathbf{Q}\mathbf{x}]_\times$, one verifies by multiplying (21) on the left by \mathbf{Q} and on the right by \mathbf{Q}^\top that

$$\mathbf{Q} \mathbf{M}_0 \mathbf{Q}^\top = 2\mathbf{I} + [\mathbf{Q}\mathbf{c}^{*' }]_\times [\|\mathbf{m}^{*' }\| \mathbf{b}_3]_\times. \quad (22)$$

Since $\det(\mathbf{Q} \mathbf{M}_0 \mathbf{Q}^\top) = \det(\mathbf{M}_0)$, a straightforward calculation yields

$$\det(\mathbf{M}_0) = 2(2 - \langle \mathbf{Q}\mathbf{c}^{*' }, \|\mathbf{m}^{*' }\| \mathbf{b}_3 \rangle)^2 \quad (23)$$

$$= 2(2 - \langle \mathbf{c}^{*' }, \mathbf{m}^{*' } \rangle)^2. \quad (24)$$

This shows that \mathbf{M}_0 (and consequently \mathbf{M}) is invertible provided the condition (17) is satisfied. A similar reasoning shows that $\mathbf{L}'|_{\boldsymbol{\varepsilon}'=\mathbf{0}}$ is invertible provided (18) is satisfied. ■

A nice property of the linearized interaction matrix in (19) is its block-triangular structure. This property is exploited next to derive simple stabilizing feedback laws with nice decoupling properties.

B. Control Law and Stability Analysis

In the sequel, consider a camera-mounted 6-DoF holo-robot observing a motionless rigid object of unknown shape. Let the control inputs be the translational and rotational velocities of the camera, gathered in the vector $\mathbf{v} \in \mathbb{R}^6$.

The nonmetric control law is simply defined as

$$\mathbf{v} = \boldsymbol{\Lambda} \bar{\boldsymbol{\varepsilon}}, \quad (25)$$

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

Theorem 3.2: The nonmetric control law (25) ensures local exponential stability of the equilibrium $\bar{\boldsymbol{\varepsilon}} = \mathbf{0}$ provided that the following conditions are satisfied:

$$\text{sign}(\lambda_1) = \text{sign}(\lambda_2) = \text{sign}\left(\frac{2}{z^*} - \mathbf{m}^{*'\top} \mathbf{q}^{*' }\right), \quad (26)$$

$$\text{sign}(\lambda_4) = \text{sign}(\lambda_5) = \text{sign}(2 - \mathbf{m}^{*'\top} \mathbf{c}^{*' }), \quad (27)$$

$$\lambda_3, \lambda_6 > 0. \quad (28)$$

Proof: From the expression (19) of the linearization of the interaction matrix and applying the control law (25), the linearization of the closed-loop system at $\bar{\boldsymbol{\varepsilon}} = \mathbf{0}$ writes

$$\dot{\bar{\boldsymbol{\varepsilon}}} = \mathbf{A} \bar{\boldsymbol{\varepsilon}}, \quad (29)$$

with

$$\mathbf{A} = \bar{\mathbf{L}}|_{\bar{\varepsilon}=0} \mathbf{A} = \begin{bmatrix} \mathbf{A}_{vv} & \mathbf{0} \\ \mathbf{A}_{vw} & \mathbf{A}_{ww} \end{bmatrix} \quad (30)$$

and hence

$$\mathbf{A}_{vv} = - \left[\frac{2}{z^*} \mathbf{I} + [\mathbf{m}^{*'}]_{\times} [\mathbf{q}^{*'}]_{\times} \right] \mathbf{A}_v, \quad (31)$$

$$\mathbf{A}_{ww} = - \left[2\mathbf{I} + [\mathbf{c}^{*'}]_{\times} [\mathbf{m}^{*'}]_{\times} \right] \mathbf{A}_w. \quad (32)$$

We first determine the analytic expression of the eigenvalues of \mathbf{A} . Since \mathbf{A} is block triangular, its eigenvalues consist of the union of the eigenvalues of \mathbf{A}_{vv} and \mathbf{A}_{ww} :

$$\sigma(\mathbf{A}) = \sigma(\mathbf{A}_{vv}) \cup \sigma(\mathbf{A}_{ww}), \quad (33)$$

with $\sigma(\cdot)$ denoting the spectrum of a matrix. Let us first consider \mathbf{A}_{vv} . If \mathbf{x} denotes an eigenvector of \mathbf{A}_{vv} associated with the eigenvalue μ , then for any rotation matrix \mathbf{Q} , $\mathbf{Q}\mathbf{x}$ is an eigenvector of $\mathbf{Q}\mathbf{A}_{vv}\mathbf{Q}^{\top}$ associated with the same eigenvalue μ . Thus, $\sigma(\mathbf{A}_{vv}) = \sigma(\mathbf{Q}\mathbf{A}_{vv}\mathbf{Q}^{\top})$. As in the proof of Theorem 3.1, let us choose \mathbf{Q} such that $\mathbf{Q}\mathbf{m}^{*'} = \|\mathbf{m}^{*'}\| \mathbf{b}_3$ with $\mathbf{b}_3 = [0, 0, 1]^{\top}$. Then, a straightforward calculation shows that $\mathbf{Q}\mathbf{A}_{vv}\mathbf{Q}^{\top}$ is upper-triangular so that its diagonal terms correspond to its spectrum:

$$\begin{aligned} \sigma(\mathbf{Q}\mathbf{A}_{vv}\mathbf{Q}^{\top}) &= \sigma(\mathbf{A}_{vv}) = \\ & \left\{ -\lambda_1 \left(\frac{2}{z^*} - \mathbf{m}^{*'\top} \mathbf{q}^{*'} \right), -\lambda_2 \left(\frac{2}{z^*} - \mathbf{m}^{*'\top} \mathbf{q}^{*'} \right), -\lambda_3 \frac{2}{z^*} \right\}. \end{aligned}$$

Applying the same procedure to \mathbf{A}_{ww} yields

$$\begin{aligned} \sigma(\mathbf{Q}\mathbf{A}_{ww}\mathbf{Q}^{\top}) &= \sigma(\mathbf{A}_{ww}) = \\ & \left\{ -\lambda_4 (2 - \mathbf{m}^{*'\top} \mathbf{c}^{*'}), -\lambda_5 (2 - \mathbf{m}^{*'\top} \mathbf{c}^{*'}), -2\lambda_6 \right\}. \end{aligned}$$

The proof directly follows from the above expressions of $\sigma(\mathbf{A}_{vv})$ and $\sigma(\mathbf{A}_{ww})$. ■

Theorem 3.2 provides explicit conditions on the control gains so as to ensure local exponential stability of the closed-loop system. Since both $\mathbf{m}^{*'}$ and $\mathbf{c}^{*'}$ are defined by the user, the unique important constraint is on the choice of λ_1 and λ_2 .

However, since $\mathbf{m}^* = z^* \mathbf{m}^{*'}$, a direct consequence of (26)–(28) is that stability is guaranteed $\forall \lambda_i > 0$, $i = 1, 2, \dots, 6$, if the two following conditions are satisfied:

$$\mathbf{m}^{*'\top} \mathbf{q}^{*'} < 2, \quad (34)$$

$$\mathbf{m}^{*'\top} \mathbf{c}^{*'} < 2. \quad (35)$$

Let us note that these conditions can always be verified. The condition (34), which also holds in the original DVS, expresses the perpendicular distance between the chosen control point and the reference plane. Given that this plane corresponds to the dominant plane of the object, this condition can be easily satisfied if the control point is chosen such that its parallax $\rho^* \in \mathbb{R}$ is sufficient small. In fact, one could simply choose a point that has $\rho^* = 0$ as the dominant plane crosses the object. As for condition (35), it represents the length of the projection of $\mathbf{m}^{*'}$ onto $\mathbf{c}^{*'}$. It can then be easily satisfied by setting, e.g., $\mathbf{c}^{*' = \beta \mathbf{m}^{*'}/\|\mathbf{m}^{*'}\|^2}$, $\forall \beta < 2$. The closed-loop system (29) is thus always locally exponentially stable at the equilibrium, $\forall \lambda_i > 0$. Furthermore, the behavior of the translational control error is decoupled from the rotational one, as shown in (30).

C. A Fully Decoupled Matrix

The obtained interaction matrix (19) is lower triangular in the general case. This section presents a case of special interest. Indeed, it occurs very often in practice and such matrix is purely diagonal. 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 $\mathbf{m}^{*' = [0, 0, 1]^{\top}}$ and set $\mathbf{c}^{*' = \mathbf{m}^{*'}$. As consequences,

$$\frac{1}{z^*} = \mathbf{n}^{*\top} \mathbf{m}^{*' = \frac{1}{d^*}, \quad (36)$$

$$\mathbf{q}^{*' = \mathbf{n}^*, \quad (37)$$

and the stability conditions (34) and (35) are satisfied. Furthermore, the obtained interaction matrix (19) writes

$$\bar{\mathbf{L}}|_{\bar{\varepsilon}=0} = - \begin{bmatrix} \frac{1}{z^*} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{z^*} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{2}{z^*} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}. \quad (38)$$

Therefore, a fully decoupled linear system (29) with strictly negative diagonal elements is obtained $\forall \lambda_i > 0$. Once again, let us remark that such matrix is used only for analysis purposes, i.e., it is not needed for servoing the robot.

IV. EXPERIMENTAL RESULTS

This section reports experimental results using both synthetic and real data, simulating and using a camera-mounted six-DoF robotic arm. In all cases, the control objective consists in stabilizing the robot such that the current image of the object coincides with its image captured at the reference pose. The control error and control law are both calculated at the signal level, i.e., they do not use either image features or metric information of the object. Indeed, pixel intensities are directly exploited to estimate all needed projective parameters. The direct image registration technique described in [7] is used for this purpose. Comparative results are presented using the Direct Visual Servoing (DVS) and the proposed decoupled DVS, all using a control gain of $\lambda = 0.4$ and a stop condition on the norm of the control errors of 10^{-3} .

A. Synthetic Data

This first set of experiments uses a nonplanar object. The used target is indeed a hyperbolic paraboloid, also known as horse's saddle, whose center is placed 1m away from the reference camera pose. The displacement of the initial camera pose relative to the reference one is of $[0.17, -0.11, 0.01]$ meters (norm of 0.2m, i.e., of 20% of the center's depth) in translation, and of $[-0.1, -0.31, 1.04]$ radians (norm of 62.4°) in rotation. All these information is obviously not available for the control computation. The focal lengths are set to 500 pixels, no skew, and the principal point as the middle of the image, which has 550×418 pixels. The applied sampling period is of 30 ms, leading to a framerate of 33.33Hz. The control point is chosen such that $\mathbf{m}^{*' = [0, 0, 1]^{\top}}$. For the sake of simplicity, all pixel intensities within the region of interest of size 200×150 pixels (i.e., within the grid) are exploited, and no illumination variations

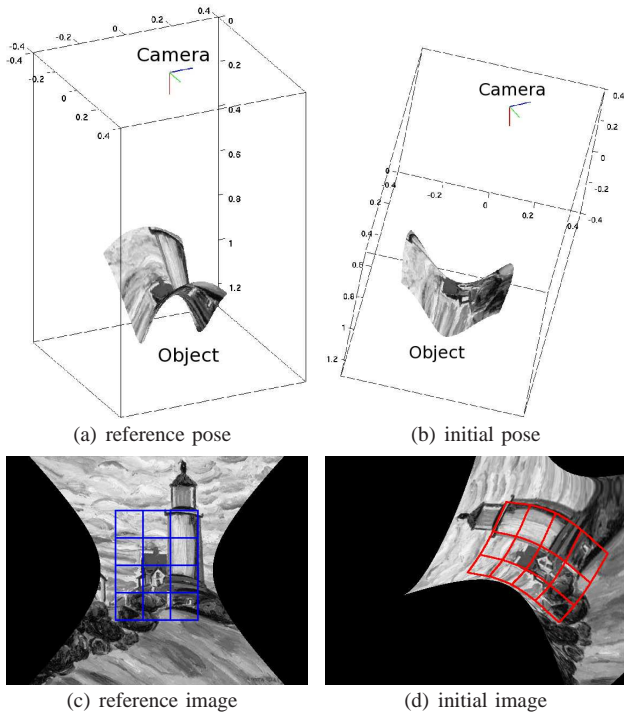


Fig. 2. Setup of the experiment using synthetic data. The target is nonplanar (a horse’s saddle), and all pixels within the outlined image region (the grid) are exploited. (Top) Configurations of the camera frame with respect to the target, seen from different viewpoints. Only the exploited part of the object is shown. (Bottom) Images as viewed by the camera at those relative poses.

have been imposed. The interested reader may refer to [7] for numerous other estimation experiments, including varying illumination conditions, different objects and color cameras. The setup for this experiment is shown in Fig. 2.

1) *DVS*: This section presents the results obtained using the reduced control error (13) under that setup. Let us remark that such setup is particularly unfavourable for most 2D visual servoing techniques due to the relatively large initial rotational displacement around the \bar{z} -axis. This does not hold for the DVS technique. Indeed, it successfully performs the task accurately, whose convergence is established after 905 images without nearly any camera retreat. The corresponding results are shown in Fig. 3. Nevertheless, one can observe that there exists a coupling of the translational velocities with rotational ones, in particular v_x for such experiment.

2) *Decoupled DVS*: This section presents the results obtained using the decoupled control error (16) with $\mathbf{c}^{*l} = \mathbf{m}^{*l}$. As for the DVS technique, it successfully performs the task without nearly any camera retreat, whose convergence is now established after only 762 images. The rate of convergence is thus improved. Indeed, the coupling in the translational velocities observed in the previous DVS is far less severe using this technique. Furthermore, it fully disappears near the equilibrium. See Fig. 4 for the corresponding results.

B. Real Data

This second set of experiments uses a planar object, which is placed about 0.7m away from the reference pose. The displacement of the initial robot pose relative to the reference one is of $[0.13, -0.23, -0.08]$ meters (norm of 0.27m, i.e., of about 38% of the depths) in translation, and

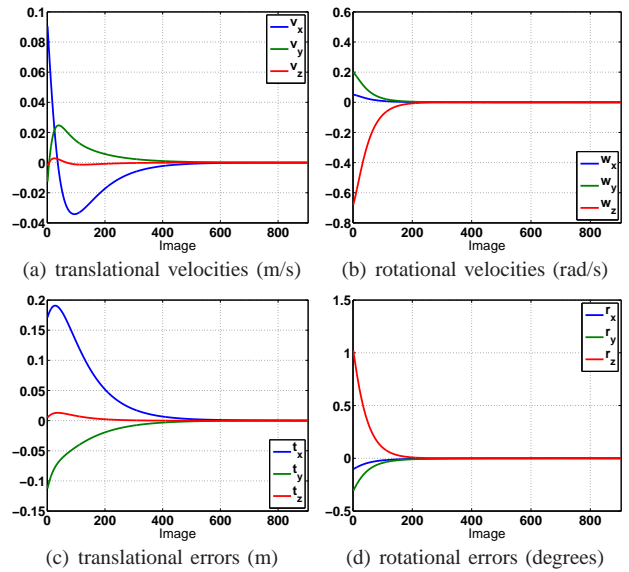


Fig. 3. Direct visual servoing with respect to a nonplanar object. (Top) Control inputs, i.e., the camera velocities. The translational velocities are coupled with the rotational ones, in particular v_x for such experiment. (Bottom) Motion of the camera in the Cartesian space towards convergence.

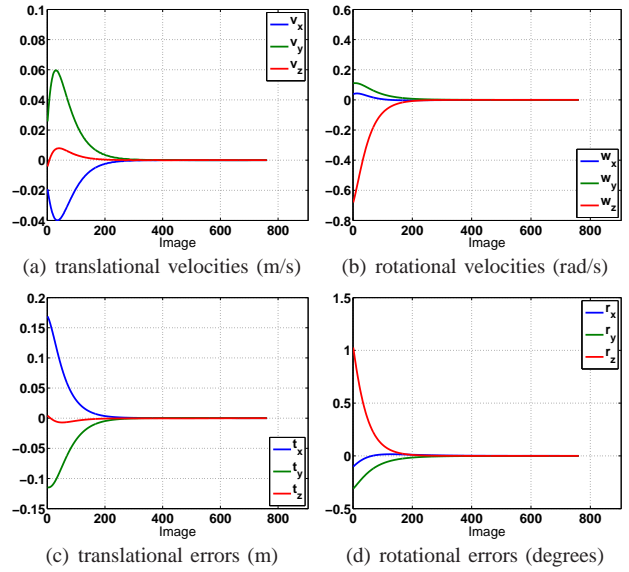


Fig. 4. Decoupled direct visual servoing with respect to a nonplanar object. (Top) Control inputs, i.e., camera velocities. A decoupling behavior of the translational velocities is observed as we compare them with Fig. 3. (Bottom) Motion of the camera in the Cartesian space towards convergence.

of $[-20.3, 2.17, 14.59]$ degrees (norm of 25°) in rotation. These are obviously unknown by the algorithms. To show the robustness of the techniques, a coarsely calibrated webcam is used. Indeed, the focal lengths are simply set to 420 pixels, no skew and the principal point as the middle of the image, which has 320×240 pixels. This camera is placed on the end-effector of a 6-axis robotic arm, and the hand/eye calibration is also coarsely set. The framerate is of about 30Hz, which is the maximum of the webcam. The chosen reference template has 70×70 pixels (to satisfy real-time constraints), and the control point is chosen as its center. See Fig. 5 for this setup.

1) *DVS*: As in the previous set of experiments, the reduced control error (13) is used here. The corresponding

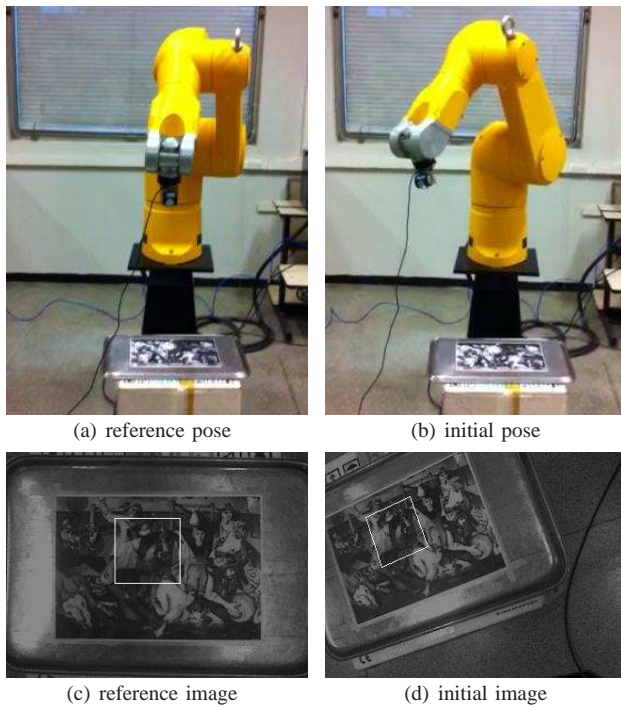


Fig. 5. Setup of the the experiment using real data. It uses a planar object and a coarsely calibrated camera-mounted robotic arm. (Top) Configurations of the robot relative to the target. (Bottom) Images as viewed by the mounted camera at those poses. All pixels within the outlined region are exploited.

results are shown in Fig. 6. Once again, the servoing task is successfully performed, with convergence after 417 images. Nevertheless, there exist a coupling of the translational velocities with rotational ones.

2) *Decoupled DVS*: Again, the decoupled control error (16) with $\mathbf{c}^{*'} = \mathbf{m}^{*'}$ is used here. The task is also successfully performed, with convergence after 309 images. The rate of convergence is thus improved. The coupling behavior previously observed is indeed reduced. See Fig. 7 for the corresponding results.

V. CONCLUSIONS

This article has proposed a general intensity-based non-metric visual servoing technique that decouples the error dynamics. This new decoupled technique is general in the sense that it deals with both planar and nonplanar objects, under both translation and rotation displacements between the reference and initial frames. The proposed technique directly 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 state. Comparative results with a state-of-the-art direct technique using a camera-mounted 6-DoF robotic arm confirm the improvements. In future work we plan to exploit the proposed framework for the control of other mechanical systems, such as nonholonomic and underactuated robots.

REFERENCES

- [1] 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*, pages 58–64, 1995.
- [2] F. Chaumette and S. Hutchinson. Visual servo control part I: Basic approaches. *IEEE Robot. Autom. Mag.*, pages 82–90, 2006.

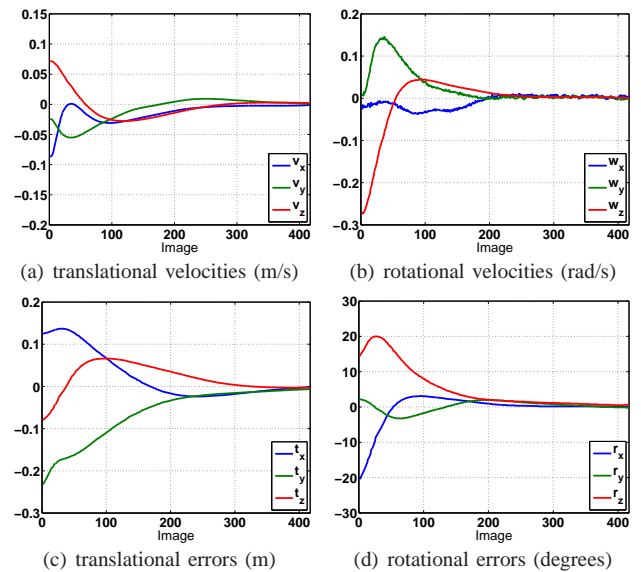


Fig. 6. Direct visual servoing with respect to a planar object using a coarsely calibrated camera and robot. (Top) Camera velocities. Observe that there exists a coupling amongst them, in particular v_x and w_y for such experiment. (Bottom) Motion in the Cartesian space towards convergence.

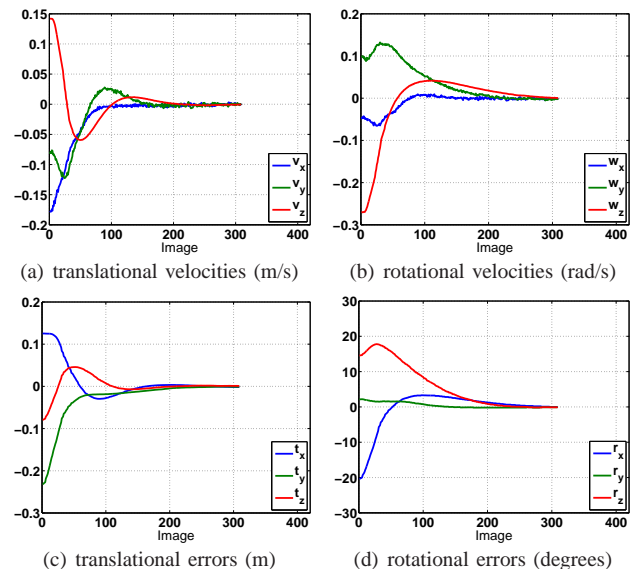


Fig. 7. Decoupled direct visual servoing with respect to a planar object using a coarsely calibrated camera and robot. (Top) Camera velocities. A decoupling behavior is observed as we compare them with Fig. 6. (Bottom) Motion of the end-effector in the Cartesian space towards convergence.

- [3] H. de Plinval, P. Morin, P. Mouyon, and T. Hamel. Visual servoing for underactuated VTOL UAVs: a linear, homography-based approach. In *Proc. IEEE Int. Conf. Robot. Autom.*, pages 3004–3010, 2011.
- [4] O. Faugeras, Q.-T. Luong, and T. Papadopoulos. *The geometry of multiple images*. The MIT Press, 2001.
- [5] P. J. Huber. *Robust Statistics*. John Wiley & Sons, 1981.
- [6] G. Silveira and E. Malis. Direct visual servoing with respect to rigid objects. Research Report 6265, INRIA, 2007.
- [7] 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*, 89(1):84–105, 2010.
- [8] G. Silveira and E. Malis. Direct Visual Servoing: Vision-based estimation and control using only nonmetric information. *IEEE Transactions on Robotics*, 28(4):974–980, 2012.
- [9] L. Thaler and M. A. Goodale. Beyond distance and direction: The brain represents target locations non-metrically. *J. Vis.*, 10(3):1–27, 2010.