

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Direct visual servoing with respect to rigid objects

Geraldo Silveira — Ezio Malis

N° 6265

August 2007

Thème NUM



ISSN 0249-6399 ISRN INRIA/RR--6265--FR+ENG



Direct visual servoing with respect to rigid objects

Geraldo Silveira*†, Ezio Malis†

Thème NUM — Systèmes numériques
Projet ARobAS

Rapport de recherche n° 6265 — August 2007 — 24 pages

Abstract: Existing visual servoing techniques which do not need metric information require, on the other hand, prior knowledge about the object's shape and/or the camera's motion. In this paper, we propose a new visual servoing technique which does not require any of them. The method is direct in the sense that: the intensity value of all pixels is used (i.e. we avoid the feature extraction step which introduces errors); and that the proposed control error as well as the control law are fully based on image data (i.e. metric measures are neither required nor estimated). Besides not relying on prior information, the scheme is robust to large errors in the camera's internal parameters. We provide the theoretical proofs that the proposed task function is locally isomorphic to the camera pose, that the approach is motion- and shape-independent, and also that the derived control law ensures local asymptotic stability. Furthermore, the proposed control error allows for simple, smooth, physically valid, singularity-free path planning, which leads to a large domain of convergence for the servoing. The approach is validated through various results using objects of different shapes, large initial displacements as well as large errors in the camera's internal parameters.

Key-words: Visual servoing, visual tracking, direct methods.

This work is also partially supported by the CAPES Foundation under grant no. 1886/03-7, and by the international agreement FAPESP-INRIA under grant no. 04/13467-5.

* CenPRA Research Center – DRVC Division, Rod. Dom Pedro I, km 143,6, Amarais, CEP 13069-901, Campinas/SP, Brazil, Geraldo.Silveira@cenpra.gov.br

† INRIA Sophia-Antipolis – Project ARobAS, 2004 Route des Lucioles, BP 93, 06902 Sophia-Antipolis Cedex, France, FirstName.LastName@sophia.inria.fr

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : +33 4 92 38 77 77 — Télécopie : +33 4 92 38 77 65

Asservissement visuel direct par rapport aux objets rigides

Résumé : Les techniques existantes d'asservissement visuel qui n'ont besoin d'aucune information métrique exigent toutefois, des connaissances a priori sur la forme de l'objet et/ou sur le mouvement de la caméra. Dans cet article, nous proposons une nouvelle technique qui ne nécessite aucune de ces connaissances. La méthode est directe au sens où: l'intensité lumineuse de tous les pixels de l'image est exploitée (on évite donc l'extraction de primitives géométriques qui introduit des erreurs); et où la fonction de tâche proposée ainsi que la loi de commande sont entièrement basées sur des données d'image (i.e. aucune mesure métrique n'est exigée ou estimée). En plus de n'avoir besoin d'aucune de ces informations au préalable, la méthode est robuste aux erreurs de calibration sur les paramètres intrinsèques de la caméra. Nous établissons les preuves théoriques que la fonction de tâche proposée est isomorphe à la pose de la caméra, que l'approche est indépendante de la forme de l'objet et du mouvement de la caméra, et aussi que la loi de commande proposée assure une stabilité locale. En outre, la fonction de tâche proposée permet une planification de trajectoire simple, régulière, physiquement valide et libre de singularités, ce qui mène à un domaine de convergence très grand de l'asservissement. L'approche est validée par de nombreux résultats en utilisant des objets de différentes formes, avec des grands déplacements initiaux ainsi que de grandes erreurs dans les paramètres intrinsèques de la caméra.

Mots-clés : Asservissement visuel, commande référencée vision, méthodes directes.

Contents

1 Introduction **4**

2 Theoretical Background **5**
2.1 Notations 5
2.2 Two-view projective geometry 5
2.3 Direct visual servoing w.r.t. planar objects 6

3 The General Direct Visual Servoing **7**
3.1 The unified direct visual tracking 7
3.2 The proposed task function 8
3.3 The control law and stability analysis 9

4 Path Planning **10**

5 Results **11**

6 Conclusions **12**

Appendix **12**

A Relation between the task function and camera pose **13**

B Demonstrations **13**
B.1 Demonstration of the Lemma 1 13
B.2 Demonstration of the Theorem 1 15
B.3 Demonstration of the Corollary 1 16
B.4 Demonstration of the Theorem 2 17

1 Introduction

Visual servoing consists in controlling the motion of a robot through the feedback of images [1]. Visually servoed systems can then be viewed as regulators of an appropriate task function [2]. This article considers the task functions which can be constructed from the current and the reference images. Furthermore, we focus on techniques that do not use metric information about the observed target. Unfortunately, existing methods which fall within this class require prior knowledge of the object's shape and/or of the camera's motion. Indeed, the technique proposed in e.g. [3] (as any other method which relies solely on the Essential matrix), although not requiring an explicit metric model of the object, requires a non-planar target as well as a sufficient amount of translation to be carried out in order to avoid the degeneracies. With regard to the technique proposed in [4], although also not requiring metric information, it is designed for planar targets only. We remark that even for image-based visual servoing approaches e.g. [5], minimal metric knowledge (the depth distribution) is necessary to provide a stable control law [6]. The 2.5D visual servoing strategy [7] was then proposed to enlarge that domain of stability. However, it requires a coarse metric estimate of the normal vector of the planar target, in order to decide between the two possible solutions of the reconstruction. Another alternative to augment the domain of convergence is to perform a suitable path planning [8]. However, this latter method, when applied to unknown targets (our objective), also requires this coarse metric estimate to accomplish its first phase, i.e. to plan the camera trajectory.

In this work, we propose a new visual servoing technique which does not require or estimate any metric information about the observed target. The proposed control error as well as the control law are fully based on image measurements. We provide here the theoretical proof that the control law, which is extremely simple to compute, ensures local asymptotic stability for the servoing. The control error proposed in this article generalizes [4], which is designed for planar surfaces only. In fact, the proposed method is independent of the object's shape and of the camera's motion. Moreover, another generalization concerns well-established techniques such as [9], which performs a partial Euclidean reconstruction. Our projective formulation also naturally encompasses this latter solution. The theoretical proof of all of these statements is provided in this report. In addition to these attractive generalizations, other improvements are achieved as well. The proposed control error is locally isomorphic to the camera pose, and is also injective around the equilibrium for the entire domain of rotations. The theoretical proof of this isomorphism is also provided here. Furthermore, another important strength of our control error is that it allows for simple, smooth, physically valid, singularity-free path planning. This procedure can considerably enlarge the domain of convergence of the visual servoing.

Another remarkable difference between the proposed approach and those previously mentioned (except for [4]) is how image information is exploited: these latter techniques are feature-based (e.g. points, lines, circle, etc.). This means that a sparse set of carefully chosen, distinct features is firstly extracted in both current and desired images. Correspondences are established afterward based on descriptors together with a robust matching procedure. Here, we directly use the intensity value of all pixels [10]. Therefore, higher

accuracy is achieved since noise is not introduced (there is no feature extraction process) and much more information is exploited. Our direct visual tracking method [11] is coherent with the proposed control law in the sense that it does not make any assumption about both the object's shape and the motion carried out by the camera. That is, the tracking procedure likewise does not require any of these prior knowledges. In fact, the same set of parameters used by the tracking procedure is also used by the control law. Indeed, we strongly believe that both vision and control aspects are intrinsically coupled processes, and are treated here as such. This represents a rupture of paradigm with respect to the vast majority of existing visual servoing techniques to date, where feature extraction process and control computation are formulated separately. Although conceptually appealing, this latter uncoupled, feature-based framework presents some relevant drawbacks. For example, global constraints are not easy to embed into feature correspondence algorithms [12], such as the fact that large portions of the scene move with a coherent rigid motion, or that the appearance changes due to motion of the scene relative to the lights. Attempts to impose these constraints are usually performed a posteriori within this framework. On the other hand, the rigidity of the scene and the robustness to lighting changes can be effectively incorporated within direct methods, see e.g. [13, 14].

The proposed approach is validated through various results using objects of different shapes, large initial displacements, large errors in the camera's internal parameters, as well as path planning examples.

2 Theoretical Background

2.1 Notations

Consider a 3D point projected in the reference image $\mathcal{I}^*$ as a pixel with homogeneous coordinates $\mathbf{p}^* \in \mathbb{P}^2$. Then, its intensity value is denoted by $\mathcal{I}^*(\mathbf{p}^*)$. After displacing the camera by a translation $\mathbf{t} \in \mathbb{R}^3$ and a rotation $\mathbf{R} \in \mathbb{SO}(3)$, another image $\mathcal{I}$ is acquired. That displacement can be represented in homogeneous coordinates as

$$\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \quad \in \mathbb{SO}(3) \times \mathbb{R}^3. \tag{1}$$

The rotation is parameterized here by the axis of rotation $\mathbf{u} \in \mathbb{R}^3 : \|\mathbf{u}\| = 1$ and angle of rotation θ, i.e. $\mathbf{R} = \exp([\mathbf{u}\theta]_\times)$, where $[\mathbf{r}]_\times$ denotes the anti-symmetric matrix associated to the vector $\mathbf{r}$. We also follow the usual notations $\widehat{\mathbf{v}}$, $\mathbf{v}'$, $\mathbf{v}^\top$, $\|\mathbf{v}\|$ to represent respectively an estimate, a modified version, the transpose, and the 2-norm of a variable $\mathbf{v}$. Moreover, $\mathbf{0}$ denotes a matrix of zeros of appropriate dimensions.

2.2 Two-view projective geometry

Projective geometry is an extension of the Euclidean geometry, which describes a larger class of transformations than just rotations and translations, including in particular the

perspective projection performed by a camera [15]. In this general framework, corresponding image points $\mathbf{p} \leftrightarrow \mathbf{p}^*$ are related by

$$\mathbf{p} \propto \underbrace{\left(\mathbf{G}_\infty + \mathbf{e}_p \mathbf{q}^{*\top}\right)}_{\mathbf{G}} \mathbf{p}^* + \rho^* \mathbf{e}_p, \tag{2}$$

where the morphism $\mathbf{G} \in \mathbb{SL}(3)$ includes the homography at infinity $\mathbf{G}_\infty$ and the epipole $\mathbf{e}_p \in \mathbb{R}^3$ in $\mathcal{I}$, while $\rho^* \in \mathbb{R}$ is the projective parallax with respect to a (in general) virtual plane represented in the image $\mathcal{I}^*$ by the vector $\mathbf{q}^* \in \mathbb{R}^3$. Indeed, the Lie group $\mathbb{SO}(3) \times \mathbb{R}^3$ in (1) is a subspace of the Lie group $\mathbb{SL}(3) \times \mathbb{R}^3$, where an homogeneous element of the latter is represented here by

$$\mathbf{Q} = \begin{bmatrix} \mathbf{G} & \mathbf{e}_p \\ \mathbf{0} & 1 \end{bmatrix} \quad \in \mathbb{SL}(3) \times \mathbb{R}^3. \tag{3}$$

From Eq. (2), a warping operator $\mathbf{w}(\,\cdot\,; \mathbf{Q}, \rho^*) : \mathbb{P}^2 \mapsto \mathbb{P}^2$ can thus be defined:

$$\mathbf{p}^* \;\mapsto\; \mathbf{p} = \mathbf{w}(\mathbf{p}^*; \mathbf{Q}, \rho^*). \tag{4}$$

2.3 Direct visual servoing w.r.t. planar objects

The homography-based 2D visual servoing technique proposed in [4] has as objective to control the motion of a camera with respect to a planar object. For this, consider that this object lies on the plane Π, which is not reconstructed either off-line or during the servoing. The method is in fact based on the recovery of the projective homography $\mathbf{G}_\Pi$ induced by this plane between two views. Indeed, given an estimate of the camera's internal parameters $\mathbf{K}$, then $\mathbf{G}_\Pi$ can be transformed into an Euclidean homography through

$$\mathbf{H}_\Pi = \mathbf{K}^{-1}\, \mathbf{G}_\Pi\, \mathbf{K}, \tag{5}$$

as well as a chosen image point (also called control point) can be normalized by

$$\mathbf{m}^* = \mathbf{K}^{-1} \mathbf{p}^*. \tag{6}$$

The task function $\mathbf{e}_\Pi = \left[\mathbf{e}_{\nu\Pi}^\top, \mathbf{e}_{\omega\Pi}^\top\right]^\top \in \mathbb{R}^6$ with

$$\begin{cases} \mathbf{e}_{\nu\Pi} = \mathbf{H}_\Pi \mathbf{m}^* - \mathbf{m}^* \\ [\mathbf{e}_{\omega\Pi}]_\times = \mathbf{H}_\Pi - \mathbf{H}_\Pi^\top \end{cases} \tag{7}$$

is then proven [4] to be locally isomorphic to the camera pose. However, it is non-injective around the equilibrium if the entire domain of rotations is considered, since both $\theta = 0$ and $\theta = \pi$ are mapped to by $\mathbf{e}_{\omega\Pi} = \mathbf{0}$.

Remark 1. *It is easy to verify that $\mathbf{G}_\Pi$ is a particular case of the general morphism $\mathbf{G}$ in (2) for planar targets. In this case, the 3-vector $\mathbf{q}^*$ is the representation of the plane $\Pi = [\mathbf{n}^{*\top}, -d^*]^\top$ in the pixel coordinate system. i.e. $\mathbf{q}^* = \mathbf{K}^{-\top}\mathbf{n}^*$. Of course, given that $\mathbf{p}^*$ is the projection of a 3D point belonging to Π, the parallax from $\mathbf{p}^*$ w.r.t. Π is $\rho^* = 0$.*

3 The General Direct Visual Servoing

This section presents an unified framework where the object can be either planar or non-planar regardless the motion carried out by the camera. Moreover, the proposed task function generalizes previous well-established formulations.

3.1 The unified direct visual tracking

In order to recover the parameters which relate the projection of an object between two views, the Fundamental matrix could be estimated [15]. However, it is not defined for planar scenes, and if the camera undergoes a pure rotation motion the estimation of the translation is degenerate. Hence, a robust method e.g. [16] to detect those degeneracies and to switch between models (affine Fundamental matrix, homography, etc.) has to be used. Alternatively to using these feature-based, multiple hypotheses testing methods, we use a more general visual tracking technique which exploits all image information within a region of the reference image $\mathcal{R}^* \subseteq \mathcal{I}^*$ [11]. The assumption of this tracking procedure is that the camera observes a continuous textured surface. In this case, an efficient second-order optimization technique is applied to minimize directly the intensity discrepancies between the current and the reference images. That is, we seek the parameters $\mathbf{Q}$ and $\boldsymbol{\rho}^*$ (see Subsection 2.2) that minimize the following cost function:

$$\frac{1}{2}\sum_{\mathbf{p}_i^* \in \mathcal{R}^*} \Big[\mathcal{I}\big(\mathbf{w}(\mathbf{p}_i^*; \mathbf{Q}, \rho_i^*)\big) - \mathcal{I}^*(\mathbf{p}_i^*)\Big]^2. \tag{8}$$

The obtained set of parameters can be used as an estimate for the same procedure when a new image is acquired.

With regard to those degeneracies, a conservative solution to recover these parameters is applied in [11]. That is, for every new image an initial attempt to explain the image differences is performed by using $\mathbf{Q}$ only. Afterward, the remainder is corrected with the parallaxes $\boldsymbol{\rho}^*$. Here, we adapt the technique proposed in [13] to our projective formulation. Thus, $\boldsymbol{\rho}^*$ is only used (and whenever used this is made simultaneously with $\mathbf{Q}$) to explain the image discrepancies if and only if the difference between resulting cost values from the (image) optimizations exceeds the image noise. This minimal parameterization framework presents many strengths. First, in the case that $\mathbf{Q}$ has already explained most of those discrepancies, by including the parallaxes afterward would perturb the estimate of $\mathbf{Q}$ in the next iteration. Furthermore, once the optimal parallaxes are obtained, there is no reason to maintain them in the minimization procedure since we deal here with rigid objects only.

Remark 2. *The tracking method does not use any prior information either about the object's shape or about the camera's motion. In fact, adapting the strategy proposed in [13] to our context leads to leaving unaltered the initial parallax values $\boldsymbol{\rho}_0^* = \mathbf{0}$ either if the object is planar or if the camera undergoes a pure rotation motion. This happens because, in these cases, $\mathbf{Q}$ solely explains all image discrepancies.*

3.2 The proposed task function

From the parameters $\mathbf{Q}$ and $\boldsymbol{\rho}^*$ estimated by the visual tracking method (see Subsection 3.1), we propose in this subsection a suitable task function for positioning the robot from an initial pose to the reference (desired) one. Indeed, using Eq. (3) and an estimate of the camera's internal parameters $\mathbf{K}$ permit to obtain

$$\mathbf{H} = \mathbf{K}^{-1}\,\mathbf{G}\,\mathbf{K} \tag{9}$$

and

$$\mathbf{e}'_p = \mathbf{K}^{-1}\mathbf{e}_p. \tag{10}$$

These entities together with the control point $\mathbf{m}^*$ (6) will be used for defining the translational part of the task function.

Next, define both a "projective" axis of rotation $\boldsymbol{\mu} \in \mathbb{R}^3$ (which does not have necessarily unit norm) from $\mathbf{H}$ (9) as

$$[\boldsymbol{\mu}]_\times = \frac{1}{2}\big(\mathbf{H} - \mathbf{H}^\top\big), \tag{11}$$

and a "projective" angle of rotation as

$$\phi = \begin{cases} \arcsin\big(\,\mathrm{sat}(\|\boldsymbol{\mu}\|)\,\big), & \text{if } \big(\mathrm{tr}(\mathbf{H}) - 1\big)/2 \geq 0, \\ \pi - \arcsin\big(\,\mathrm{sat}(\|\boldsymbol{\mu}\|)\,\big), & \text{otherwise,} \end{cases} \tag{12}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix and the function $\mathrm{sat} : \mathbb{R}^* \mapsto [0, 1]$ is defined as

$$\mathrm{sat}(x) = \begin{cases} x, & \text{if } 0 \leq x \leq 1, \\ 1, & \text{if } x > 1, \end{cases} \tag{13}$$

so that $\arcsin(\cdot)$ is a real-valued function. Of course, if $\|\boldsymbol{\mu}\| = 0$ then $\boldsymbol{\mu}$ can be chosen arbitrarily.

Theorem 1 (Task function and isomorphism). *The task function*

$$\mathbf{e} = \begin{bmatrix} \mathbf{e}_\nu \\ \mathbf{e}_\omega \end{bmatrix} = \begin{bmatrix} \mathbf{H}\,\mathbf{m}^* + \rho^*\mathbf{e}'_p - \mathbf{m}^* \\ \dfrac{\boldsymbol{\mu}}{\|\boldsymbol{\mu}\|}\phi \end{bmatrix} \in \mathbb{R}^6, \tag{14}$$

is locally isomorphic to the camera pose. Moreover, it is injective around the equilibrium for the entire domain of rotations, since only $\theta = 0$ is mapped to by $\mathbf{e}_\omega = \mathbf{0}$.

Proof. The proof of this isomorphism is presented in Appendix B.2, which uses the results from Lemma 1 in Appendix A. ∎

Remark 3. *A very important note about the task function defined in (14) is that it is constructed from projective entities only, i.e. without measuring or requiring any metric information about the object. This also means that robustness to camera's internal parameters is achieved.*

Remark 4. *Since the epipole is computed in the tracking process, we could use it solely to construct a decoupled translation error, e.g. by defining*

$$\mathbf{e}_\nu = \mathbf{e}'_p \tag{15}$$

instead of that defined in (14). The translation error (15) is decoupled from the rotation motion since $\mathbf{e}'_p = \mathbf{K}^{-1}\mathbf{e}_p = \mathbf{K}^{-1}\mathbf{K}\,\mathbf{t} = \mathbf{t}$ *(neglecting errors in* $\mathbf{K}$*). However, if the object is planar then one is not sure if the recovered* $\mathbf{e}_p$ *corresponds to the true solution (because more than one admissible solution does exist). Nevertheless, the coupling in (14) is not a major concern to the stability because a path planning is performed (see Section 4). In addition to the possible modification (15), we could also have defined the rotational error differently, such as:*

$$[\mathbf{e}_\omega]_\times = 2[\boldsymbol{\mu}]_\times = \mathbf{H} - \mathbf{H}^\top, \tag{16}$$

to replace $\mathbf{H}_\Pi - \mathbf{H}_\Pi^\top$ *as defined in (7), in our general, unified framework. However, remarkable improvements are achieved through* $\mathbf{e}_\omega$ *as defined in (14), which are stated in Corollary 1.*

Corollary 1 (Generality and improvements). *The proposed task function* $\mathbf{e} = \left[\mathbf{e}_\nu^\top, \mathbf{e}_\omega^\top\right]^\top \in \mathbb{R}^6$ *in (14) is a generalization of the one* $\mathbf{e}_\Pi = \left[\mathbf{e}_{\nu\Pi}^\top, \mathbf{e}_{\omega\Pi}^\top\right]^\top \in \mathbb{R}^6$ *defined in (7) for coping with objects of arbitrary shape. Moreover, the proposed control error allows for a straightforward path planning (shown in Section 4). Furthermore, our projective formulation naturally encompasses the hybrid control error* $\mathbf{e}'_\Pi = \left[\mathbf{e}'^\top_{\nu\Pi}, \mathbf{e}'^\top_{\omega\Pi}\right]^\top = \left[(\alpha\mathbf{m} - \mathbf{m}^*)^\top, \theta\mathbf{u}^\top\right]^\top \in \mathbb{R}^6$, $\alpha > 0$, *proposed in [9], which requires a coarse metric estimate of the normal vector of the planar target for recovering* $\mathbf{e}'_{\omega\Pi}$.

Proof. The proof of these statements is presented in Appendix B.3. ∎

3.3 The control law and stability analysis

Theorem 2 (Local stability). *The control law*

$$\mathbf{v} = \boldsymbol{\Lambda}\,\mathbf{e}, \qquad \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_\nu \mathbf{I}_3, \lambda_\omega \mathbf{I}_3), \lambda_\nu, \lambda_\omega > 0, \tag{17}$$

where $\mathbf{v} = \left[\boldsymbol{\nu}^\top, \boldsymbol{\omega}^\top\right]^\top \in \mathbb{R}^6$ *comprises the translational and rotational velocities and the control error* $\mathbf{e} = \left[\mathbf{e}_\nu^\top, \mathbf{e}_\omega^\top\right]^\top \in \mathbb{R}^6$ *as defined in (14), ensures local asymptotic stability provided that the point* $\mathbf{m}^*$ *is chosen such that its parallax relative to the dominant plane of the object is sufficiently small.*

Proof. The proof is presented in Appendix B.4, which uses the results from Lemma 1 in Appendix A. ∎

With respect to the parallax condition in Theorem 2, there always exist a point which has zero parallax (and thus can be chosen as the control point) since, in the formulation, the dominant plane[1] always crosses the object. Therefore, the closed-loop system is always locally asymptotically stable. However, for robustness reasons, it is convenient to choose a point close to the center of the object.

4 Path Planning

Although the method is robust to large camera calibration errors, it is desirable that the trajectory of the control point in the image be as closely as possible to a straight line. With this, a large domain of convergence for the visual servoing is achieved since we enforce that at least such a point always remains in the image. For this, instead of driving $\mathbf{e}(t) \longrightarrow \mathbf{0}$ an appropriate path tracking $\mathbf{e}(t) \longrightarrow \mathbf{e}^*(t)$ is performed. This is accomplished by regulating a time-varying control error

$$\mathbf{e}'(t) = \mathbf{e}(t) - \mathbf{e}^*(t). \tag{18}$$

By abuse of notation, we represent in this section $\mathbf{e}^*$ as the desired control error value to be achieved, instead of a value defined with respect to the reference frame as throughout the article.

The strategy presented in this section is different from [8], where this latter is composed by three phases and requires a coarse metric estimate of the normal vector of the planar target. Indeed, a simple strategy is shown to be sufficient to attain our purposes, i.e. without requiring any metric information and being independent of the object's shape and of the camera's motion. This is due to the properties of the proposed control error (14):

- we need to plan the trajectory of only one point, which means that physically valid camera situations are always specified;
- the projective axis-angle parameterization already provides for a smooth trajectory;
- given the isomorphism, there is no singularity or local minima in the large.

Thus, a linear desired path $\mathbf{e}^*(t) = \left[\mathbf{e}_\nu^{*\top}(t), \mathbf{e}_\omega^{*\top}(t)\right]^\top, \forall t \in [0, T]$, such that $\mathbf{e}^*(0) = \mathbf{e}(0)$ and $\mathbf{e}^*(T) = \mathbf{0}$, can be easily constructed:

$$\mathbf{e}^*(t) = \mathbf{e}^*(0) + \big(\mathbf{e}^*(T) - \mathbf{e}^*(0)\big)\frac{t}{T} = \mathbf{e}(0)\left(1 - \frac{t}{T}\right). \tag{19}$$

Nevertheless, motivated by the fact (see the results from Lemma 1 in Appendix A) that

$$\mathbf{e}_\omega = \frac{\boldsymbol{\mu}}{\|\boldsymbol{\mu}\|}\phi \longrightarrow \mathbf{u}\theta \qquad \text{as} \qquad \mathbf{t} \longrightarrow \mathbf{0}, \tag{20}$$

[1] if the object is not planar, this plane is virtual.

which means that if $\mathbf{t} = \mathbf{0}$ then geodesic rotations will be induced, the rotational part of (19) is slightly changed into

$$\mathbf{e}_\omega^*(t) = \mathbf{e}_\omega(t-1)\left(1 - \frac{t}{T}\right), \tag{21}$$

where the notation $\mathbf{e}_\omega(t-1)$ refers to the last value of $\mathbf{e}_\omega$.

Therefore, considering a motionless target and willing to regulate (18), the control law (17) is transformed into

$$\mathbf{v} = \mathbf{\Lambda}(t)\,\mathbf{e}'(t) + \frac{\partial \mathbf{e}^*(t)}{\partial t}, \tag{22}$$

where the feed-forward term $\partial \mathbf{e}^*(t)/\partial t$ allows compensation of the tracking error and

$$\mathbf{\Lambda}(t) = \mathrm{diag}\big(\lambda_\nu \mathbf{I}_3, \lambda_\omega(t)\mathbf{I}_3\big) \tag{23}$$

$$= \begin{bmatrix} \lambda \mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \lambda \exp(-\gamma \|\mathbf{e}_\nu(t)\|)\mathbf{I}_3 \end{bmatrix}, \quad \lambda, \gamma > 0, \tag{24}$$

is an adaptive gain matrix also motivated by (20): $\lambda_\omega(t)$ is small for large $\|\mathbf{e}_\nu(t)\|$ and $\lambda_\omega(t) \longrightarrow \lambda$ as $\|\mathbf{e}_\nu(t)\| \longrightarrow 0$.

5 Results

In this section, we report a diverse set of results to validate our direct visual servoing technique. We use the word "direct" to express both that there is no feature extraction process, and that the control error and control law are computed using only image measurements. For their computation, all pixels within an area of interest are exploited. For all results, this area is delimited by a red grid. The visual servoing task consists in positioning the camera with respect to a rigid object independently of its shape. For this, a reference image is stored at the reference (desired) pose. After displacing the camera (whilst tracking the object in the image) to another pose, the objective is then to drive the camera back to this desired pose. It should be noted that the technique is also independent of the displacement between the initial and desired poses, i.e. it may comprise pure translations, pure rotations or a combination of both. In order to have a real ground truth, we constructed synthetic objects of different shapes and, to simulate realistic situations as closely as possible, real textured images are mapped onto them. For all images shown here, blue marks are used to depict the motion of the control point ($\mathbf{m}^*$ in Eq. (14)) in the image plane, whilst its planned path is projected in green. This latter is typically composed of 1000 points with an adaptive gain (24) with $\lambda = \gamma = 10$.

In Fig. 1, it is shown that the proposed method can cope with planar objects, including thus the previously proposed task function (7) which is designed for this particular surface. The control law is stable: both translational and rotational velocities converge to zero. At convergence, the visual information coincides with the reference image, and the camera is positioned at the reference pose very accurately. Errors less than 1mm for the translation

and less than 0.1° for the rotation are achieved. Figure 1 also shows the evolution of the Cartesian displacement (in meters and in radians). The blue marks in the reference image depict the straight line performed by the control point, as desired. In this same set of results, a comparison of the behavior of the control errors by using the two methods (defined by (7) and ours) can be seen in the bottom plots. We can observe that a rapidly increasing rotational error is obtained by the existing technique. This behavior may lead to failure of the system. Yet another improvement concerns a positioning task for a rotation of $\theta = 180^\circ$, which cannot be performed by the existing method. The results achieved by using the proposed task function for this case are presented in Fig. 2, and without any path planning, i.e. using Eq. (17) with $\lambda_\nu = \lambda_\omega = 1$. Since a pure rotation is given, the result will be the same regardless the shape of the object and so we use a sphere. Nevertheless, it is shown in Fig. 3 that the technique can also deal with this non-planar object under rotational and translational displacements ($\theta_0 = 84^\circ$ and $\|\mathbf{t}_0\| = 0.72$m) without performing any additional modification. Once again, observe that the control point follows a straight line in the image.

In last set of results, we set up a challenging scenario: the object was an hyperbolic paraboloid (the horse's saddle); the used focal lengths were almost the double of the true ones, i.e. instead of $f_u = f_v = 500$ we used $\widehat{f}_u = 900$ and $\widehat{f}_v = 800$; and a large initial displacement ($\theta_0 = 162^\circ$ and $\|\mathbf{t}_0\| = 0.3$m) was carried out. Even in this case, the visual servoing was performed successfully. See Fig. 4 for the corresponding results. This demonstrates that the proposed technique also copes with non-planar objects, that the strategy is robust to large errors in the camera's internal parameters, and that the servoing has a very large domain of convergence.

6 Conclusions

In this paper, we have proposed a new approach to vision-based control that does not require or estimate any metric information. Our general technique is independent of the object's shape and of the camera's motion. Thus, we do not rely on prior knowledge (leading to system flexibility), and we achieve robustness to errors in the calibration parameters. The sole requirement is that the object has to be rigid. In addition, the visual tracking exploits all image information without any feature extraction process, allowing us to attain high levels of accuracy for the positioning whilst being computationally efficient. This latter strength is also important since real-time performance is always a major concern in robotic systems. Finally, the proposed control law ensures a very large domain of convergence for the servoing due to a straightforward path planning scheme. Hence, visual servoing tasks can be performed despite large initial displacements.

Acknowledgments

This work is also partially supported by the CAPES Foundation under grant no. 1886/03-7, and by the international agreement FAPESP-INRIA under grant no. 04/13467-5.

Appendix

Consider a 3D point, which has coordinates $\mathcal{X}^* = [X^*, Y^*, Z^*]^\top$ relative to the reference frame, and in front of the camera $Z^* > 0$. Its location can be expressed with respect to the current camera frame through

$$\mathcal{X} = \mathbf{R}\mathcal{X}^* + \mathbf{t}. \tag{25}$$

Consider the pinhole model. In this case, the projection of $\mathcal{X}^*$ and $\mathcal{X}$ in the normalized image plane is respectively

$$\mathbf{m}^* = [x^*,\, y^*, 1]^\top = \frac{1}{Z^*}\mathcal{X}^* \qquad \text{and} \qquad \mathbf{m} = [x,\, y,\, 1]^\top = \frac{1}{Z}\mathcal{X}. \tag{26}$$

A Relation between the task function and camera pose

Let us state an important result which is largely used through this report. For this, neglect for the moment the errors in the camera's internal parameters.

Lemma 1 (Task function and camera pose). *The task function (14) is expressed as a function of the camera pose by using:*

$$\mathbf{e}_\nu = \big((\mathbf{R} - \mathbf{I}_3)\mathcal{X}^* + \mathbf{t}\big)\frac{\beta}{Z^*}, \quad \beta > 0, \tag{27}$$

and $\mathbf{e}_\omega = \frac{\boldsymbol{\mu}}{\|\boldsymbol{\mu}\|}\phi$ *with*

$$\boldsymbol{\mu} = \sin(\theta)\mathbf{u} + \frac{1}{2}[\mathbf{q}^{*\prime}]_\times \mathbf{t} \tag{28}$$

and Eq. (12).

B Demonstrations

B.1 Demonstration of the Lemma 1

Proof (Task function and camera pose). We provide a constructive proof. Let us start by expressing $\mathbf{e}_\nu$ as a function of the camera pose. From Eqs. (9) and (10), we have

$$\mathbf{G} = \mathbf{K}\,\mathbf{H}\,\mathbf{K}^{-1} \quad \text{and} \quad \mathbf{e}_p = \mathbf{K}\,\mathbf{e}'_p. \tag{29}$$

Further, if errors in $\mathbf{K}$ are not considered then $\mathbf{e}'_p = \mathbf{K}^{-1}\mathbf{e}_p = \mathbf{K}^{-1}\mathbf{K}\,\mathbf{t} = \mathbf{t}$.

By injecting (29) in (2), multiplying both terms of the result by $\mathbf{K}^{-1}$, and then using (6) yield

$$\mathbf{m} \propto \mathbf{H}\,\mathbf{m}^* + \rho^*\mathbf{e}'_p = \alpha\,\mathbf{m}, \quad \alpha > 0. \tag{30}$$

This permits to write

$$\mathbf{e}_\nu = \mathbf{H}\,\mathbf{m}^* + \rho^*\mathbf{e}'_p - \mathbf{m}^* \tag{31}$$

$$= \alpha\,\mathbf{m} - \mathbf{m}^*. \tag{32}$$

From the scale factor

$$\alpha = \beta\frac{Z}{Z^*}, \quad \beta > 0, \tag{33}$$

and Eq. (26), we can rewrite (32) as

$$\mathbf{e}_\nu = (\mathcal{X} - \mathcal{X}^*)\frac{\beta}{Z^*}, \tag{34}$$

where β is only a normalization factor which depends on the reconstruction algorithm. Finally, Eq. (25) permits to obtain the desired relation between $\mathbf{e}_\nu$ and the camera pose:

$$\mathbf{e}_\nu = \big((\mathbf{R} - \mathbf{I}_3)\mathcal{X}^* + \mathbf{t}\big)\frac{\beta}{Z^*}. \tag{35}$$

With respect to $\mathbf{e}_\omega$, knowing that

$$\mathbf{G}_\infty = \mathbf{K}\,\mathbf{R}\,\mathbf{K}^{-1} \quad \text{and} \quad \mathbf{e}_p = \mathbf{K}\,\mathbf{t} \tag{36}$$

and using (9) yield

$$\mathbf{H} = \mathbf{R} + \mathbf{t}\mathbf{q}^{*\prime\top} \tag{37}$$

with

$$\mathbf{q}^{*\prime} = \mathbf{K}^\top\mathbf{q}^*. \tag{38}$$

Then, the projective axis of rotation (11) can be written

$$[\boldsymbol{\mu}]_\times = \frac{1}{2}\big(\mathbf{H} - \mathbf{H}^\top\big) \tag{39}$$

$$= \frac{1}{2}\big(\mathbf{R} + \mathbf{t}\mathbf{q}^{*\prime\top} - \mathbf{R}^\top - \mathbf{q}^{*\prime}\mathbf{t}^\top\big). \tag{40}$$

Using Rodrigues' formula

$$\mathbf{R} = \mathbf{I}_3 + \sin(\theta)[\mathbf{u}]_\times + \big(1 - \cos(\theta)\big)[\mathbf{u}]_\times^2 \tag{41}$$

we have

$$\mathbf{R} - \mathbf{R}^\top = 2\sin(\theta)[\mathbf{u}]_\times, \tag{42}$$

and together with the property

$$\mathbf{t}\mathbf{q}^{*\prime\top} - \mathbf{q}^{*\prime}\mathbf{t}^\top = \big[\,[\mathbf{q}^{*\prime}]_\times\mathbf{t}\,\big]_\times, \tag{43}$$

we obtain the important relation from Eq. (40):

$$\boldsymbol{\mu} = \sin(\theta)\mathbf{u} + \frac{1}{2}[\mathbf{q}^{*\prime}]_{\times}\mathbf{t}. \tag{44}$$

The projective angle of rotation follows directly by injecting the norm of (44) in (12):

$$\phi = \begin{cases} \arcsin\big(\,\mathrm{sat}\big(\|\sin(\theta)\mathbf{u} + \frac{1}{2}[\mathbf{q}^{*\prime}]_{\times}\mathbf{t}\|\big)\,\big), & \text{if } \big(\mathrm{tr}(\mathbf{H}) - 1\big)/2 \geq 0, \\ \pi - \arcsin\big(\,\mathrm{sat}\big(\|\sin(\theta)\mathbf{u} + \frac{1}{2}[\mathbf{q}^{*\prime}]_{\times}\mathbf{t}\|\big)\,\big), & \text{otherwise.} \end{cases} \tag{45}$$

Finally, by using Eqs. (44), (45) and (37) the desired relation between $\mathbf{e}_{\omega}$ and the camera pose is achieved. ∎

B.2 Demonstration of the Theorem 1

This demonstration uses the results from Lemma 1 in Appendix A.

Proof (Isomorphism). The proof consists in demonstrating that $\mathbf{e} = \big[\mathbf{e}_{\nu}^{\top}, \mathbf{e}_{\omega}^{\top}\big]^{\top} = \mathbf{0}$ if and only if $\theta = 0$ and $\mathbf{t} = \mathbf{0}$. We remark that the domain of the angle of rotation also includes $\theta = \pi$.

First of all, it is evident that if $\theta = 0$ and $\mathbf{t} = \mathbf{0}$ then $\mathbf{e} = \mathbf{0}$ ($\Longleftarrow$). However, we need to prove the implication in the other direction ($\Longrightarrow$): if $\mathbf{e} = \mathbf{0}$ then $\theta = 0$ and $\mathbf{t} = \mathbf{0}$. That is, we have to show that the homogeneous non-linear system of equations $\mathbf{e} = \mathbf{0}$ has an unique solution which is $\theta = 0$ and $\mathbf{t} = \mathbf{0}$, $\forall \mathbf{q}^*$ and $\forall \mathcal{X}^*$ such that $Z^* > 0$. We start by constructing such a system of equations, which is given as

$$\begin{cases} (\mathbf{R} - \mathbf{I}_3)\mathcal{X}^* + \mathbf{t} = \mathbf{0} \\ \sin(\theta)\mathbf{u} + \dfrac{1}{2}[\mathbf{q}^{*\prime}]_{\times}\mathbf{t} = \mathbf{0} \\ \big(\mathrm{tr}(\mathbf{H}) - 1\big)/2 \geq 0. \end{cases} \tag{46}$$

The first equation of (46) comes directly from Eq. (27) since $\beta > 0$ and $Z^* > 0$. Thus, one obtains directly

$$\mathbf{t} = (\mathbf{I}_3 - \mathbf{R})\mathcal{X}^* = \big(\mathbf{I}_3 - \exp([\mathbf{u}\theta]_{\times})\big)\mathcal{X}^*. \tag{47}$$

The second equation and the inequality in (46) were constructed by injecting Eqs. (44) and (45) in (14) together with the following facts. The statement $\mathbf{e}_{\omega} = \boldsymbol{\mu}\phi/\|\boldsymbol{\mu}\| = \mathbf{0}$ implies $\phi = 0$ since $\boldsymbol{\mu}/\|\boldsymbol{\mu}\|$ is an unit (projective) axis of rotation. This implies in the inequality in (46) given that $\mathrm{sat}(\|\boldsymbol{\mu}\|) \in [0, 1]$ and then $\arcsin\big(\mathrm{sat}(\|\boldsymbol{\mu}\|)\big) \in [0, \pi/2]$. Thus, $\phi = 0$ if and only if $\|\boldsymbol{\mu}\| = 0$.

Pre-multiplying the second equation of (46) by $\mathbf{t}^{\top}$ and then injecting (47), one obtains

$$\big[\big(\mathbf{I}_3 - \exp([\mathbf{u}\theta]_{\times})\big)\mathcal{X}^*\big]^{\top}\sin(\theta)\mathbf{u} = 0, \tag{48}$$

using the property

$$\mathbf{b}^\top [\mathbf{a}]_\times \mathbf{b} = 0. \tag{49}$$

Given that by definition $Z^* > 0$, $\big(\mathbf{I}_3 - \exp([\mathbf{u}\theta]_\times)\big)\mathcal{X}^* = \mathbf{0}$ if and only if $\theta = 0$. On the other hand, the solutions of $\sin(\theta)\mathbf{u} = \mathbf{0}$ are $\theta = 0$ or $\theta = \pi$, since $\mathbf{u}$ is an unit axis of rotation. Both cases imply $\mathbf{q}^{*\prime} = \mathbf{0}$ and/or $\mathbf{t} = \mathbf{0}$ from the second equation of (46). Then, using this implication together with (37) and the identity

$$\cos(\theta) = \big(\mathrm{tr}(\mathbf{R}) - 1\big)/2 \tag{50}$$

permits to make conclusions from the inequality in (46):

$$\big(\mathrm{tr}(\mathbf{H}) - 1\big)/2 \geq 0 \implies \big(\mathrm{tr}(\mathbf{R}) - 1\big)/2 \geq 0 \tag{51}$$

$$\implies |\theta| \leq \pi/2. \tag{52}$$

Therefore, the only solution to (48) is $\theta = 0$, $\forall \mathbf{q}^*$ and $\forall \mathcal{X}^*$ such that $Z^* > 0$. In addition, by applying this solution to (47), we must have $\mathbf{t} = \mathbf{0}$. ∎

B.3 Demonstration of the Corollary 1

Proof (Generality and improvements). The generality regards to coping with objects of arbitrary shape. That is, besides coping with non-planar objects, the proposed task function (14) comprises the control error (7) as well, which deals with planar objects only. This is easy to show given the following facts. From Remark 2, our visual tracking method provides the parallax $\rho_i^* = 0$, $\forall \mathbf{m}_i^*$, if the object is planar. Moreover, for this target Remark 1 states that the dominant plane is in fact the plane Π on which the object lies, i.e. $\mathbf{q}^* = \mathbf{K}^{-\top}\mathbf{n}^*$. Applying this result to (38) gives

$$\mathbf{q}^{*\prime} = \mathbf{K}^\top \mathbf{q}^* = \mathbf{K}^\top \mathbf{K}^{-\top} \mathbf{n}^* = \mathbf{n}^*. \tag{53}$$

Then, the proposed task function (14) is rewritten as

$$\begin{cases} \mathbf{e}_\nu = \mathbf{H}\,\mathbf{m}^* + \rho^* \mathbf{e}'_p - \mathbf{m}^* = \mathbf{H}\,\mathbf{m}^* - \mathbf{m}^* = \mathbf{H}_\Pi \mathbf{m}^* - \mathbf{m}^* \\ \mathbf{e}_\omega = \dfrac{\boldsymbol{\mu}}{\|\boldsymbol{\mu}\|}\phi = \boldsymbol{\eta}\big(\mathbf{H}_\Pi - \mathbf{H}_\Pi^\top\big), \end{cases} \tag{54}$$

where the function $\boldsymbol{\eta} : \mathbb{R}^{3\times 3} \mapsto \mathbb{R}^3$ is given by

$$\boldsymbol{\eta}\big(\mathbf{H}_\Pi - \mathbf{H}_\Pi^\top\big) = \frac{\boldsymbol{\mu}_\Pi}{\|\boldsymbol{\mu}_\Pi\|}\phi_\Pi \tag{55}$$

with

$$\big[\boldsymbol{\mu}_\Pi\big]_\times = \Big[\sin(\theta)\mathbf{u} + \frac{1}{2}[\mathbf{n}^*]_\times \mathbf{t}\Big]_\times = \frac{1}{2}\big(\mathbf{H}_\Pi - \mathbf{H}_\Pi^\top\big), \tag{56}$$

by injecting (53) in (44), and in (45) to obtain ϕ_Π.

Furthermore, we remark that the knowledge of $\mathbf{q}^{*\prime}$ is not required in our projective framework, regardless of the object's shape. In an Euclidean framework, $\mathbf{q}^{*\prime} = \mathbf{0}$ must be enforced as well as this latter requires perfect camera parameters. These requirements can be shown directly by injecting Eq. (26) in (25), together with Eqs. (6), (36) and afterward dividing each term of the result by Z^*:

$$\mathbf{p} \propto \mathbf{K}\,\mathbf{R}\,\mathbf{K}^{-1}\mathbf{p}^* + \frac{1}{Z^*}\mathbf{K}\mathbf{t}. \tag{57}$$

In turn, using (36) yields

$$\mathbf{p} \propto \mathbf{G}_\infty \mathbf{p}^* + \frac{1}{Z^*}\mathbf{e}_p, \tag{58}$$

which is only a stratum of the general relation (2). This demonstrates that our task function (14) is also a generalization of the hybrid control error $\mathbf{e}'_\Pi = \left[\mathbf{e}'^{\top}_{\nu\Pi}, \mathbf{e}'^{\top}_{\omega\Pi}\right]^\top \in \mathbb{R}^6$ with

$$\begin{cases} \mathbf{e}'_{\nu\Pi} = \left[\frac{Z}{Z^*}x - x^*, \frac{Z}{Z^*}y - y^*, \frac{Z}{Z^*} - 1\right]^\top = \alpha\mathbf{m} - \mathbf{m}^* \\ \mathbf{e}'_{\omega\Pi} = \mathbf{u}\theta \end{cases} \tag{59}$$

proposed in [9], with the advantage of not needing the coarse metric estimate of the normal vector to perform its required partial Euclidean reconstruction (to recover $\mathbf{e}'_{\omega\Pi} = \mathbf{u}\theta$). That is, if $\mathbf{q}^{*\prime} = \mathbf{0}$ (or $\mathbf{t} = \mathbf{0}$) then (44) and (45) yield the equivalences $\boldsymbol{\mu}/\|\boldsymbol{\mu}\| = \mathbf{u}$, $\phi = \theta$. Also, the translational error $\mathbf{e}'_{\nu\Pi}$ in (59) is shown to be equivalent to ours by using the result (32), with $\beta = 1$ from their feature-based projective reconstruction.

Besides these generalities, improvements in the behavior of the servoing is attained by using the proposed rotational control error. In fact, it determines explicitly in which quadrant the projective angle of rotation operates, instead of using simply (16) (or Eq. (7) for a planar object). This is particularly important for the initial conditions $\theta_0 > \pi/2$ and $\mathbf{t}_0 \approx \mathbf{0}$. In this situation, by using for example (7) and then (56), $\mathbf{e}_{\omega\Pi} \approx 2\sin(\theta)\mathbf{u}$ and the norm of this error is initially increased during the servoing since $\sin(\theta_0) < 1$ and $\sin(\theta) \longrightarrow 1$ as $\theta \longrightarrow \pi/2$, because the second quadrant is never specified. This may lead to system failure. In addition, this non-injection does not allow for a straightforward path planning. The task function (7) and the corresponding (16) are also non-injective around the equilibrium point if $\theta = \pi$ belongs to the codomain of these task functions. ∎

B.4 Demonstration of the Theorem 2

For this, we need to derivate the proposed control error with respect to time and then to obtain the closed-loop equation:

$$\dot{\mathbf{e}} = \begin{bmatrix} \dot{\mathbf{e}}_\nu \\ \dot{\mathbf{e}}_\omega \end{bmatrix} = \mathbf{L}\begin{bmatrix} \boldsymbol{\nu} \\ \boldsymbol{\omega} \end{bmatrix} = \mathbf{L}\,\mathbf{v} = \mathbf{L}\,\boldsymbol{\Lambda}\,\mathbf{e}, \tag{60}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_\nu \mathbf{I}_3, \lambda_\omega \mathbf{I}_3), \lambda_\nu, \lambda_\omega > 0$ and $\mathbf{L}$ denotes the interaction matrix, which is never used to perform the visual servoing. The matrix $\mathbf{L}$ is necessary only for analysis purposes.

Proof (Local stability). The proof consists in analyzing the behavior of the closed-loop system (60) around the equilibrium. Hence, only local stability will be demonstrated here. This can be achieved by using the results from Lemma 1 in Appendix A. Let us start with $\mathbf{e}_\nu$:

$$\dot{\mathbf{e}}_\nu = (\dot{\mathbf{R}}\mathcal{X}^* + \dot{\mathbf{t}})\frac{\beta}{Z^*}. \tag{61}$$

By injecting the well-known relations

$$\dot{\mathbf{t}} = -\boldsymbol{\nu} - [\boldsymbol{\omega}]_\times \mathbf{t} \tag{62}$$

$$\dot{\mathbf{R}} = -[\boldsymbol{\omega}]_\times \mathbf{R} \tag{63}$$

in (61), using (27) together with the property

$$[\mathbf{a}]_\times \mathbf{b} = -[\mathbf{b}]_\times \mathbf{a}, \tag{64}$$

we obtain

$$\dot{\mathbf{e}}_\nu = -\frac{\beta}{Z^*}\boldsymbol{\nu} + \left[\mathbf{e}_\nu + \beta\mathbf{m}^*\right]_\times \boldsymbol{\omega}. \tag{65}$$

With respect to $\mathbf{e}_\omega$, we have:

$$\dot{\mathbf{e}}_\omega = \frac{\partial \sin(\theta)\mathbf{u}}{\partial t} + \frac{1}{2}[\mathbf{q}^{*\prime}]_\times \dot{\mathbf{t}}. \tag{66}$$

By using the relation

$$\frac{\partial \sin(\theta)\mathbf{u}}{\partial t} = -\mathbf{L}_\omega \boldsymbol{\omega} \tag{67}$$

with

$$\mathbf{L}_\omega = \mathbf{I}_3 - \frac{\sin(\theta)}{2}[\mathbf{u}]_\times - \sin^2\left(\frac{\theta}{2}\right)(2\mathbf{I}_3 + [\mathbf{u}]_\times^2) \tag{68}$$

and Eqs. (62), (64), we obtain:

$$\dot{\mathbf{e}}_\omega = -\frac{1}{2}[\mathbf{q}^{*\prime}]_\times \boldsymbol{\nu} - \left(\mathbf{L}_\omega - \frac{1}{2}[\mathbf{q}^{*\prime}]_\times [\mathbf{t}]_\times\right)\boldsymbol{\omega}. \tag{69}$$

By using Eqs. (65), (69) and setting $\beta = 1$ without loss of generality, the interaction matrix is finally given as

$$\mathbf{L} = \begin{bmatrix} -\frac{1}{Z^*}\mathbf{I}_3 & [\mathbf{e}_\nu + \mathbf{m}^*]_\times \\ -\frac{1}{2}[\mathbf{q}^{*\prime}]_\times & -\mathbf{L}_\omega + \frac{1}{2}[\mathbf{q}^{*\prime}]_\times [\mathbf{t}]_\times \end{bmatrix}. \tag{70}$$

Then, we may proceed to the evaluation of (60) around $\mathbf{e} = \left[\mathbf{e}_\nu^\top, \mathbf{e}_\omega^\top\right]^\top = \mathbf{0}$:

$$\dot{\mathbf{e}} = \mathbf{L}\big|_{\mathbf{e}=\mathbf{0}} \boldsymbol{\Lambda}\, \mathbf{e} \tag{71}$$

$$= -\begin{bmatrix} \lambda_\nu \frac{1}{Z^*}\mathbf{I}_3 & -\lambda_\omega [\mathbf{m}^*]_\times \\ \lambda_\nu \frac{1}{2}[\mathbf{q}^{*\prime}]_\times & \lambda_\omega \mathbf{I}_3 \end{bmatrix} \mathbf{e}, \tag{72}$$

whose eigenvalues of $\mathbf{L}\big|_{\mathbf{e}=\mathbf{0}}\,\mathbf{\Lambda}$ are given by

$$\begin{bmatrix} -\lambda_\omega \\ -\frac{\lambda_\nu}{Z^*} \\ -\frac{\lambda_\omega Z^*+\lambda_\nu-\sqrt{\lambda_\omega^2 Z^{*2}+\lambda_\nu^2-2\lambda_\nu\lambda_\omega Z^*(1-Z^*\mathbf{q}^{*\prime\top}\mathbf{m}^*)}}{2Z^*} \\ -\frac{\lambda_\omega Z^*+\lambda_\nu+\sqrt{\lambda_\omega^2 Z^{*2}+\lambda_\nu^2-2\lambda_\nu\lambda_\omega Z^*(1-Z^*\mathbf{q}^{*\prime\top}\mathbf{m}^*)}}{2Z^*} \\ -\frac{\lambda_\omega Z^*+\lambda_\nu-\sqrt{\lambda_\omega^2 Z^{*2}+\lambda_\nu^2-2\lambda_\nu\lambda_\omega Z^*(1-Z^*\mathbf{q}^{*\prime\top}\mathbf{m}^*)}}{2Z^*} \\ -\frac{\lambda_\omega Z^*+\lambda_\nu+\sqrt{\lambda_\omega^2 Z^{*2}+\lambda_\nu^2-2\lambda_\nu\lambda_\omega Z^*(1-Z^*\mathbf{q}^{*\prime\top}\mathbf{m}^*)}}{2Z^*} \end{bmatrix}. \tag{73}$$

Therefore, if $\lambda_\nu > 0$, $\lambda_\omega > 0$, and

$$\lambda_\omega^2 Z^{*2} + \lambda_\nu^2 - 2\lambda_\nu\lambda_\omega Z^*\big(1 - Z^*\mathbf{q}^{*\prime\top}\mathbf{m}^*\big) \;<\; (\lambda_\omega Z^* + \lambda_\nu)^2, \tag{74}$$

whose simplification of this latter inequality gives

$$\mathbf{q}^{*\prime\top}\mathcal{X}^* < 2 \tag{75}$$

using (26) and $Z^* > 0$, then all eigenvalues of $\mathbf{L}\big|_{\mathbf{e}=\mathbf{0}}\,\mathbf{\Lambda}$ shown in (73) have strictly negative real part. As a matter of fact, the condition (75) expresses the distance between the chosen point and the reference plane. Given that this reference plane represents the dominant plane of the object in our projective formulation, this condition can be easily satisfied if the point is chosen such that its parallax ρ^* is sufficiently small. In fact, we could use simply a point which has $\rho^* = 0$ (since in the formulation the dominant plane always crosses the object) but, for robustness reasons, it is convenient to choose a point close to the center of the object. Consequently, the closed-loop system (60) is always locally asymptotically stable. ■

References

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

[2] C. Samson, B. Espiau, and M. le Borges, *Robot Control: the Task Function Approach.* Oxford University Press, 1990.

[3] R. Basri, E. Rivlin, and I. Shimshoni, "Visual homing: surfing on the epipoles," *Int. J. of Comp. Vision*, vol. 33, no. 2, pp. 22–39, 1999.

[4] S. Benhimane and E. Malis, "Homography-based 2D visual servoing," in *IEEE International Conf. on Robotics and Automation*, USA, 2006.

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

[6] E. Malis and P. Rives, "Robustness of image-based visual servoing with respect to depth distribution errors," in *IEEE International Conference on Robotics and Automation*, 2003.

[7] 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.

[8] Y. Mezouar and F. Chaumette, "Path planning for robust image-based control," *IEEE Trans. on Rob. and Autom.*, vol. 18, pp. 534–549, 2002.

[9] 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.

[10] M. Irani and P. Anandan, "All about direct methods," in *Workshop on Vision Algorithms: Theory and practice*, 1999.

[11] E. Malis, "An efficient unified approach to direct visual tracking of rigid and deformable surfaces," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, USA, 2007.

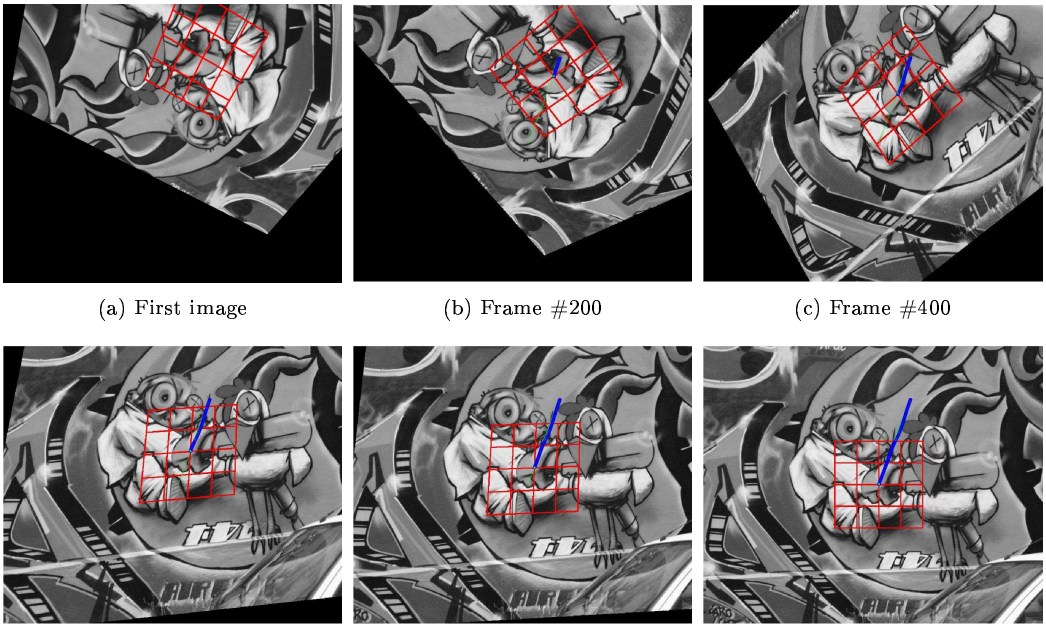
[12] H. Jin, P. Favaro, and S. Soatto, "A semi-direct approach to structure from motion," *The Visual Computer*, vol. 6, pp. 377–394, 2003.

[13] G. Silveira, E. Malis, and P. Rives, "An efficient direct method for improving visual SLAM," in *IEEE International Conference on Robotics and Automation*, Italy, 2007.

[14] G. Silveira and E. Malis, "Real-time visual tracking under arbitrary illumination changes," in *IEEE Computer Vision and Pattern Recognition*, USA, 2007.

[15] O. Faugeras, Q.-T. Luong, and T. Papadopoulo, *The geometry of multiple images.* The MIT Press, 2001.

[16] P. H. S. Torr, A. Zisserman, and S. J. Maybank, "Robust detection of degenerate configurations whilst estimating the fundamental matrix," *Comp. Vision and Image Underst.*, vol. 71, no. 3, pp. 312–333, 1998.



(a) First image (b) Frame #200 (c) Frame #400

(d) Frame #600 (e) Frame #800 (f) Reference image

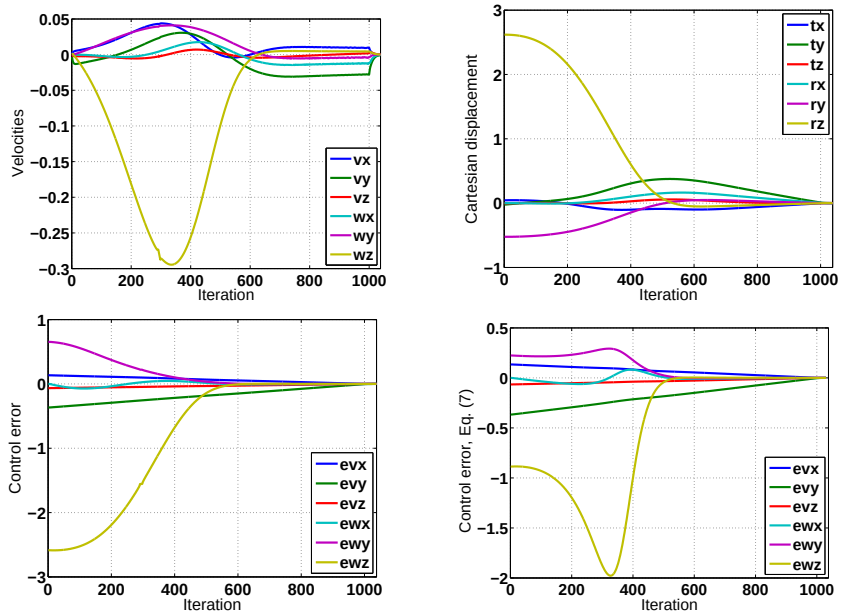
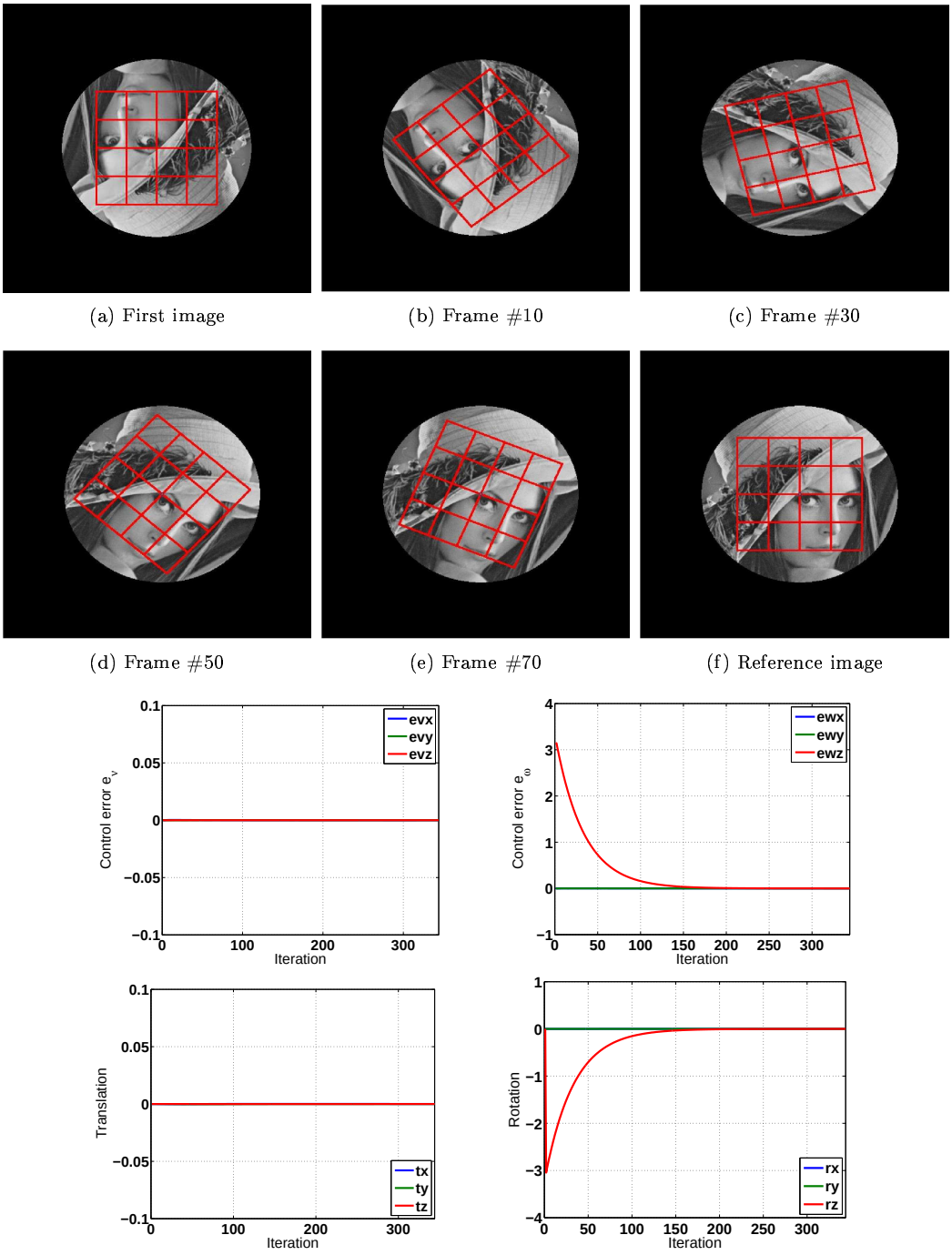


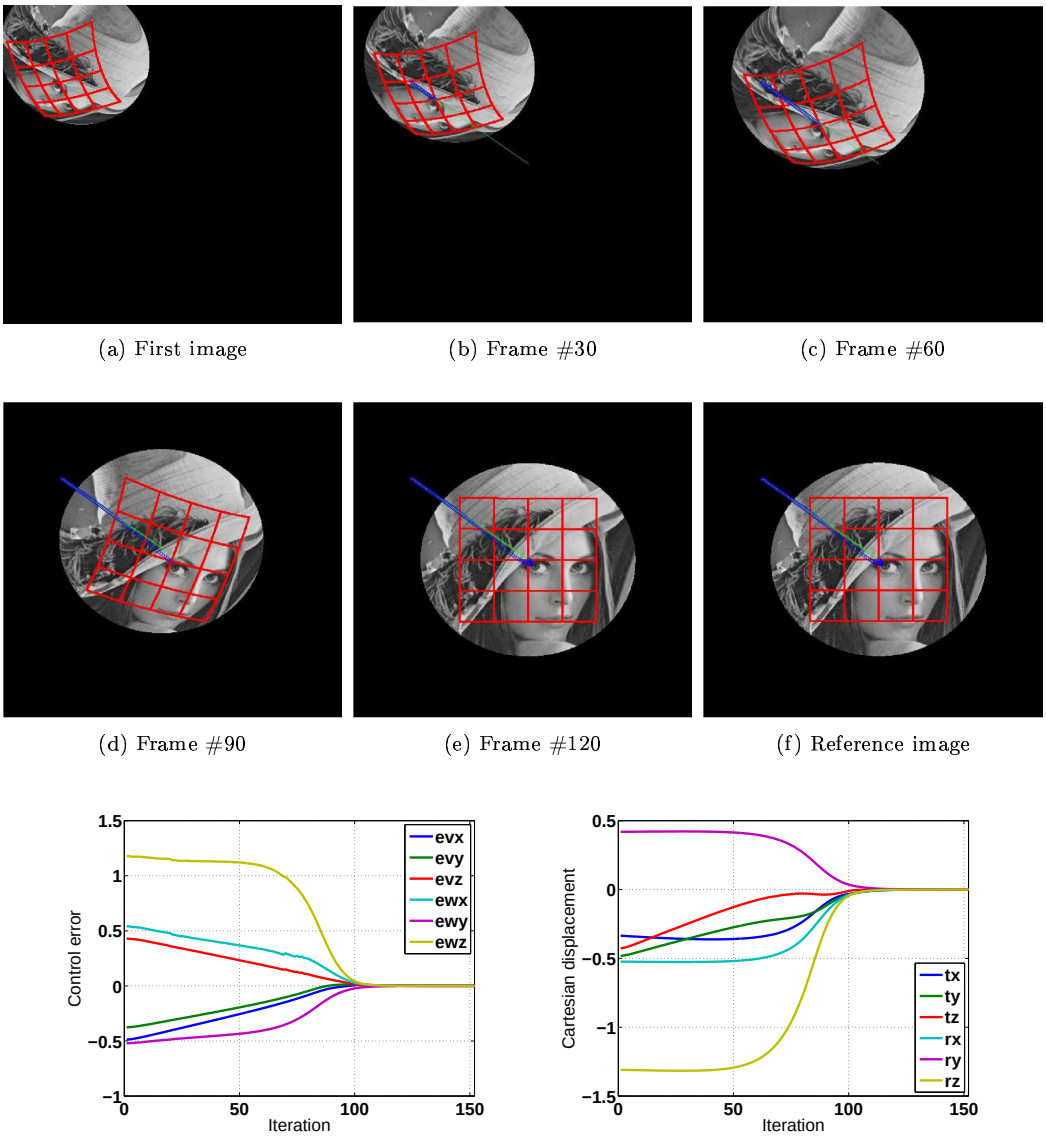
Figure 1: Direct visual servoing with respect to a planar object from an initial pose which is different from the desired one by large rotations and translations. Bottom: evolution of the control errors by using the proposed task function whilst, on the right, by using (7). Observe the rapidly increasing rotational error for the existing method, before converging to zero.



(a) First image (b) Frame #10 (c) Frame #30

(d) Frame #50 (e) Frame #70 (f) Reference image

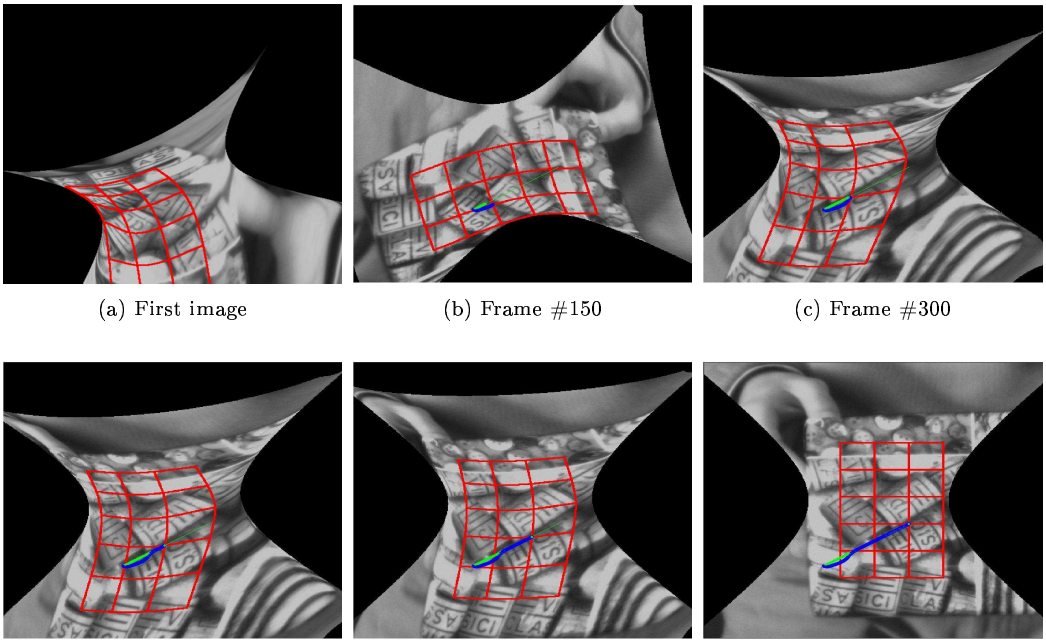
Figure 2: Direct visual servoing with respect to a sphere from an initial pose differing from the desired one by a pure rotation motion ($\theta = 180°$).



(a) First image (b) Frame #30 (c) Frame #60

(d) Frame #90 (e) Frame #120 (f) Reference image

Figure 3: Direct visual servoing with respect to a sphere from an initial pose differing from the desired one by rotations and translations. Path planning is composed of only 100 points.



(a) First image (b) Frame #150 (c) Frame #300

(d) Frame #500 (e) Frame #700 (f) Reference image

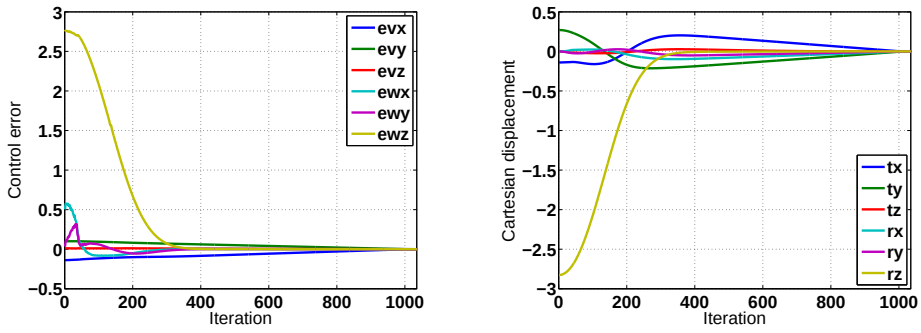


Figure 4: Direct visual servoing with respect to an hyperbolic paraboloid using very poor camera's internal parameters (almost the double of the true focal lengths), and large displacements between the initial and the desired poses.



Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399