Direct visual servoing with respect to rigid objects

Geraldo Silveira, Ezio Malis

Direct Visual Servoing with respect to Rigid Objects

Geraldo Silveira and Ezio Malis

Abstract— **In this work, we propose a new and unified visual servoing technique which is independent both on the object's shape and on the camera motion. The method is direct in the sense that: (i) pixel intensities are promptly used without any feature extraction process. This means that all possible image information is exploited; (ii) the proposed control error as well as the control law are fully based on image data, i.e. no metric 3D measure is either required or estimated. We provide the theoretical proof that the proposed control error is globally isomorphic to the camera pose, as well as that such an error is motion- and shape-independent, and also that the derived control law is locally stable. Given that only image information is used to compute that error, the proposed scheme is robust to large camera calibration errors. Furthermore, the proposed control error allows for a simple, smooth, physically valid, singularity-free path planning and thus, for achieving a very large domain of convergence. The advantages of the approach with respect to existing methods are demonstrated through a diverse set of results, which includes objects of different shapes, large initial displacements as well as large errors in the camera parameters.**

I. INTRODUCTION

Visual servoing consists in controlling the motion of a robot through the feedback of images [1]. Visually servoed systems can be then 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. The reader is referred to the E-3D visual servoing technique [3] for the case where the reference image is not necessary, as well as the visibility constraints do not apply.

Specifically, we propose in this work a new approach to vision-based control whose main characteristics are:

- it does not require or estimate any metric 3D information. Thus, we do not rely on prior information, as well as we achieve robustness to calibration parameters;
- it is independent both on the object's shape and on the camera motion. With this, besides not requiring prior information, system flexibility is attained. The sole requirement about the object is to be rigid;
- it possesses a very large domain of convergence. This is highly desirable so that tasks can be performed despite large initial displacements;
- it exploits all possible image information. We aim to attain high levels of accuracy;
- it is computationally efficient. This is important since real-time performance is always a major concern in robotic systems.

G. Silveira is with INRIA Sophia-Antipolis – Project ICARE, 2004 Route des Lucioles, BP 93, 06902 Sophia-Antipolis Cedex, France, and with the CenPRA Research Center – DRVC Division, Rod. Dom Pedro I, km 143,6, Amarais, CEP 13069-901, Campinas/SP, Brazil, Geraldo.Silveira@sophia.inria.fr

E. Malis is with INRIA Sophia-Antipolis – Project ICARE, 2004 Route des Lucioles, BP 93, 06902 Sophia-Antipolis Cedex, France, Ezio.Malis@sophia.inria.fr

To the authors' knowledge, despite that significant progress has been experienced both in theoretical and practical aspects, those relevant issues had not been simultaneously satisfied yet. For example, the technique proposed in [4] (as any other method which relies solely on the Essential matrix [5]), although not requiring nor computing an explicit 3D model of the object, it requires a non-planar target as well as a sufficient amount of translation to be carried out in order to avoid the degenerate cases. As for the image-based visual servoing approaches e.g. [6], minimal knowledge about the depth distribution is necessary to provide a stable control law [7]. The 2.5D visual servoing strategy [8] was then proposed to enlarge that domain of convergence. However, although not requiring that distribution, it requires a coarse 3D estimate of the normal vector of the plane target in order to decide between the two possible reconstruction solutions. Another alternative to augment the domain of convergence is to perform a suitable path planning [9]. However, this latter method, when applied to unknown targets (our objective), also requires that coarse 3D estimate to accomplish its first phase, i.e. to plan the camera trajectory.

Another remarkable difference from the proposed approach to those previously mentioned ones is how image information is exploited: all those 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. Alternatively and inspired by recent work [10], we also use *directly* the intensity of *all* pixels of interest [11]. Therefore, higher accuracy is achieved since much more information is exploited and noise is not introduced (there is no feature extraction process). Moreover, the proposed control error as well as the control law are fully based on image data, which means that no metric 3D measure is either required or estimated. In fact, the computation of the proposed control law, theoretically proved in this article to be locally stable, is extremely simple. With regard to [10], important contributions are given here. First of all, the proposed control error is *independent* on the object's shape, instead of relying on a planar target. Another generalization concerns well-established techniques such as [12], which performs a partial Euclidean reconstruction. Our projective formulation also encompasses naturally this latter solution. In addition to these attractive *generalizations*, other significant improvements are achieved as well. Our proposed control error is *globally* isomorphic to the camera pose. The theoretical proof is provided here. Furthermore, another very important contribution is that it allows for a simple, smooth, physically valid, singularity-free *path planning*. This procedure can enlarge considerably the domain of convergence of the visual

servoing.

Concerning our direct visual tracking method, it is also coherent with the proposed control law in the sense that it makes no distinction about both the object's shape and the motion carried out, i.e. the tracking procedure also does not require any prior knowledge. In fact, the same set of parameters used by the proposed tracking procedure is used by the control law as well. 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 large majority of the existing visual servoing techniques to date, where feature extraction process and control computation are formulated separately. This uncoupled, feature-based framework, although conceptually appealing, presents some relevant drawbacks. For example, global constraints, 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, are not easy to embed into feature correspondence algorithms [13]. As a matter of fact, the robustness of direct methods to lighting variations is out of the scope of this paper, but the interested reader may refer to the efficient solution to generic illumination changes proposed in [14].

The advantages of the approach with respect to existing methods are demonstrated through a diverse set of results, which includes objects of different shapes, large initial displacements, large errors in the camera parameters, as well as path planning examples.

II. THEORETICAL BACKGROUND

A. Notations

Consider a 3D point projected in the reference image $\mathcal{I}^*$ as a pixel with homogeneous coordinates $\mathbf{p}^* \in \mathbb{P}^2$, and whose pixel intensity 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 unit axis of rotation $\mathbf{u}$ and angle of rotation $\theta \in [0, \pi]$, 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 $\widehat{\mathbf{v}}$, $\widetilde{\mathbf{v}}$, $\mathbf{v}'$, $\mathbf{v}^\top$, $\|\mathbf{v}\|$ to represent respectively an estimate, an increment to be found, a modified version, the transpose, and the 2-norm of a variable $\mathbf{v}$. Moreover, $\mathbf{I}$ and $\mathbf{0}$ are respectively the identity matrix and a matrix of zeros of appropriate dimensions.

B. Two-view Projective Geometry

Projective geometry is an extension of Euclidean geometry, which describes a larger class of transformations than just rotations and translations, including in particular the perspective projection performed by a camera [5]. In this general framework, corresponding image points $\mathbf{p} \leftrightarrow \mathbf{p}^*$ are linked 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, $\mathbb{SO}(3) \times \mathbb{R}^3$ in (1) is a subspace of $\mathbb{SL}(3) \times \mathbb{R}(3)$, where an element of the latter is represented 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:

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

C. Direct Visual Servoing w.r.t. Planar Objects

The homography-based 2D visual servoing technique proposed in [10] 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 computation of the projective homography $\mathbf{G}_\Pi$ induced by this plane between two views. Indeed, given an estimate of the camera internal parameters $\mathbf{K}$, $\mathbf{G}_\Pi$ can be transformed to an Euclidean homography through

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

as well as a chosen image 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 to be locally isomorphic (since it is restricted to $\theta \neq \pi$) to the camera pose. Moreover, it is experimentally shown to be robust to errors in $\mathbf{K}$ and to a certain degree of non-planarity.

REMARK 1. *It is easy to verify that $\mathbf{G}_\Pi$ is a particular case of the general morphism $\mathbf{G}$ in (2). 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 induced parallax from $\mathbf{p}^*$ w.r.t. Π is then $\rho^* = 0$.*

III. THE GENERAL DIRECT VISUAL SERVOING

This section presents an unified framework where the object can be either planar or non-planar. Moreover, we propose a new task function which possesses important properties, such as global isomorphism.

A. 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 exploited. However, it is not defined for planar scenes or if the camera undergoes a pure rotation motion. Hence, a robust method e.g. [15] 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 provide here a brief description of our general tracking technique which exploits all possible image information.

Consider a region of interest in the image, $\mathcal{R}^* \subset \mathcal{I}^*$, having n pixels. Our proposed approach is similar to [16], where those parameters are obtained through an efficient second-order optimization technique which minimizes *directly* the intensity discrepancies. That is, by minimizing

$$\frac{1}{2}\sum_{\mathbf{p}_i^* \in \mathcal{R}^*} \Big[\mathcal{I}\Big(\mathbf{w}\big(\mathbf{p}_i^*; \widehat{\mathbf{Q}}\mathbf{Q}(\widetilde{\mathbf{x}}), \widehat{\rho}_i^* + \rho_i^*(\widetilde{\mathbf{y}})\big)\Big) - \mathcal{I}^*(\mathbf{p}_i^*)\Big]^2, \quad (8)$$

whose parameters are iteratively updated with the rules

$$\widehat{\mathbf{Q}} \longleftarrow \widehat{\mathbf{Q}}\mathbf{Q}(\widetilde{\mathbf{x}}), \quad (9)$$

where $\mathbf{x} \in \mathbb{R}^{12-1}$ (it is defined up to a scale factor) comprises the elements of the base field of the corresponding Lie algebra [17], as well as

$$\widehat{\boldsymbol{\rho}}^* \longleftarrow \widehat{\boldsymbol{\rho}}^* + \boldsymbol{\rho}^*(\widetilde{\mathbf{y}}), \quad (10)$$

where $\boldsymbol{\rho}^* \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ with $m \ll n$. This latter set of parameters allows to regularize the surface. The convergence of the tracking between successive images may be established, for example, when the increments become arbitrarily small, i.e. $\|\widetilde{\mathbf{x}}\| < \epsilon$. The obtained set of parameters $\widehat{\mathbf{Q}}$ and $\widehat{\boldsymbol{\rho}}^*$ is then used as an estimate for the same procedure when a new image is acquired.

With regard to the degeneracies, a conservative solution to recover the parameters is applied in [16]. That is, in every iteration, an initial attempt to explain the image differences is performed by using $\mathbf{Q}$ only. Afterwards, the remainder is corrected with the parallaxes $\boldsymbol{\rho}^*$. Here, we adapt the technique proposed in [18] to our projective formulation. Thus, $\boldsymbol{\rho}^*$ is only used (and whenever used this is made jointly with $\mathbf{Q}$) to explain the image discrepancies if and only if the difference between resulting cost values from consecutive (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 afterwards would perturb the estimate of $\mathbf{Q}$ in the next iteration. Furthermore, once the optimal parallaxes are obtained, there is not 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 motion carried out. In fact, adapting the strategy proposed in [18] to our context leads to leaving unaltered the initial parallax values $\boldsymbol{\rho}^* = \mathbf{0}$ either if the object is planar or if the camera undergoes a pure rotation motion because $\mathbf{Q}$ solely explains all image discrepancies.*

B. The Global Task Function

From the parameters $\mathbf{Q}$ and $\boldsymbol{\rho}^*$ estimated by the visual tracking method (see Section III-A), we construct in this section a suitable task function for driving the robot from an initial pose to a reference (desired) one. Indeed, using Eq. (3) and an estimate of the camera internal parameters $\mathbf{K}$ permit to obtain

$$\mathbf{H} = \mathbf{K}^{-1}\,\mathbf{G}\,\mathbf{K} \quad (11)$$

and

$$\mathbf{e}'_p = \mathbf{K}^{-1}\mathbf{e}_p. \quad (12)$$

Now, define a projective axis of rotation from $\mathbf{H}$ as

$$[\boldsymbol{\mu}]_\times = \frac{1}{2}\big(\mathbf{H} - \mathbf{H}^\top\big) \quad (13)$$

and the projective angle of rotation as

$$\phi = \begin{cases} \arcsin(\|\boldsymbol{\mu}\|), & \text{if } \big(\mathrm{tr}(\mathbf{H}) - 1\big)/2 \geq 0, \\ \pi - \arcsin(\|\boldsymbol{\mu}\|), & \text{otherwise,} \end{cases} \quad (14)$$

with $0 < \|\boldsymbol{\mu}\| \leq 1$ and where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. Of course, if $\|\boldsymbol{\mu}\| = 0$ then $\boldsymbol{\mu}$ can be chosen arbitrarily.

Theorem 1 (Task function and global isomorphism). *The task function $\mathbf{e} = \big[\mathbf{e}_\nu^\top, \mathbf{e}_\omega^\top\big]^\top \in \mathbb{R}^6$, with*

$$\begin{cases} \mathbf{e}_\nu = \mathbf{H}\,\mathbf{m}^* + \rho^*\mathbf{e}'_p - \mathbf{m}^* \\ \mathbf{e}_\omega = \dfrac{\boldsymbol{\mu}}{\|\boldsymbol{\mu}\|}\phi, \end{cases} \quad (15)$$

is globally isomorphic to the camera pose. The proof is given in Appendix A. ♢

REMARK 3. *A very important note about the task function defined in (15) is that it is constructed from projective entities only, i.e. without measuring or requiring any metric 3D information about the object. This also means that robustness to camera 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*

$$\mathbf{e}_\nu = \mathbf{e}'_p. \quad (16)$$

The translation error (16) 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 (15) is not a major concern to the stability because a path planning is performed (see Section IV). In addition to the possible modification (16), we could also have defined differently the rotational error, such as:

$$[\mathbf{e}_\omega]_\times = 2[\boldsymbol{\mu}]_\times = \mathbf{H} - \mathbf{H}^\top, \quad (17)$$

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 (15), which are stated in Corollary 1.

Corollary 1 (Generality and improvements). *The proposed task function* $\mathbf{e} = \left[\mathbf{e}_\nu^\top, \mathbf{e}_\omega^\top\right]^\top \in \mathbb{R}^6$ *as defined in (15) is a generalization of the one* $\mathbf{e}_\Pi = \left[\mathbf{e}_{\nu\Pi}^\top, \mathbf{e}_{\omega\Pi}^\top\right]^\top \in \mathbb{R}^6$ *as defined in (7), for coping with objects of arbitrary shape. Moreover, the proposed one is global and permits a straightforward path planning. Furthermore, our projective formulation encompasses naturally the control error* $\mathbf{e}'_\Pi = \left[\mathbf{e}'^\top_{\nu\Pi}, \mathbf{e}'^\top_{\omega\Pi}\right]^\top = \left[\alpha\mathbf{m} - \mathbf{m}^*, \theta\mathbf{u}^\top\right]^\top \in \mathbb{R}^6$ *proposed in [12], which is also based on the planarity of the target for recovering* $\mathbf{e}'_{\omega\Pi}$ *and* $\alpha > 0$*. The proof of those statements is given in Appendix B.* ♢

C. The Control Law and Stability Analysis

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

$$\mathbf{v} = \mathbf{\Lambda}\,\mathbf{e}, \tag{18}$$

where $\mathbf{v} = \left[\boldsymbol{\nu}^\top, \boldsymbol{\omega}^\top\right]^\top \in \mathbb{R}^6$ *comprises the translational and rotational velocities,* $\mathbf{\Lambda} = \mathrm{diag}(\lambda_\nu \mathbf{I}, \lambda_\omega \mathbf{I}) > 0$ *is the gain matrix, and the control error* $\mathbf{e} = \left[\mathbf{e}_\nu^\top, \mathbf{e}_\omega^\top\right]^\top \in \mathbb{R}^6$ *constructed from projective entities as defined in (15), is locally stable provided that the point* $\mathbf{m}^*$ *is chosen such that its parallax relative to the dominant plane of the object is sufficiently small. The proof is given in Appendix C.* ♢

IV. Path Planning

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

$$\mathbf{e}'(t) = \mathbf{e}(t) - \mathbf{e}^*(t). \tag{19}$$

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

Differently from [9], which is composed by three phases and is based on the planarity of the object, a very simple strategy is proven to be sufficient to attain our purposes, despite being independent on the object's shape. This is due to the suitable control error proposed in Section III-B: (i) we need to define the trajectory of only one point, which means that physically valid camera situations are always achieved; (ii) the projective axis-angle parameterization already provides for a smooth trajectory; (iii) given the global isomorphism, no singularity or local minima exist. Therefore, a linear desired path $\mathbf{e}^*(t), \forall t \in [0, T]$, such that $\mathbf{e}^*(0) = \mathbf{e}(0)$ and $\mathbf{e}^*(T) = \mathbf{0}$, can be easily constructed:

$$\mathbf{e}^*(t) = \mathbf{e}^*(0) + \big(\mathbf{e}^*(T) - \mathbf{e}^*(0)\big)\frac{t}{T} = \mathbf{e}(0)\left(1 - \frac{t}{T}\right), \tag{20}$$

with $0 \leq t \leq T$. In this case, considering a motionless target and willing to regulate (19), the control law (18) is transformed into

$$\mathbf{v} = \mathbf{\Lambda}\mathbf{e}'(t) + \frac{\partial \mathbf{e}^*(t)}{\partial t} = \mathbf{\Lambda}\mathbf{e}'(t) - \frac{\mathbf{e}(0)}{T}, \quad \mathbf{\Lambda} > 0, \tag{21}$$

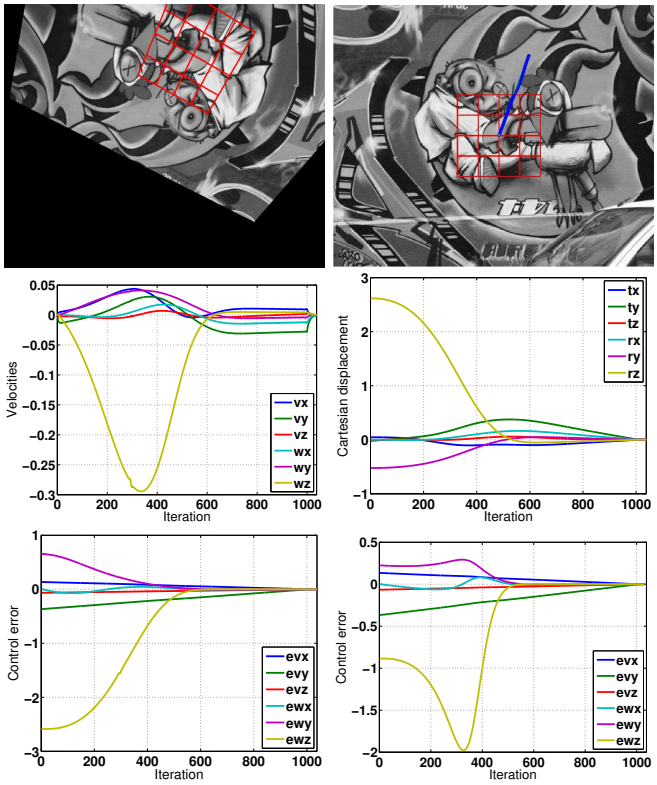

Fig. 1. Direct visual servoing w.r.t. a planar object.

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

V. Results

To use a plane to show the generality and the performances improvements w.r.t. H2VS. Path planning.

To use the sphere showing a different object and a pure rotation motion.

To use the hyperbolic paraboloid to show it is capable of dealing with complex shapes and robustness to large camera parameters. Path planning.

VI. Conclusions

TODO

As a conclusion, we can conclude several concluding conclusions.

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.

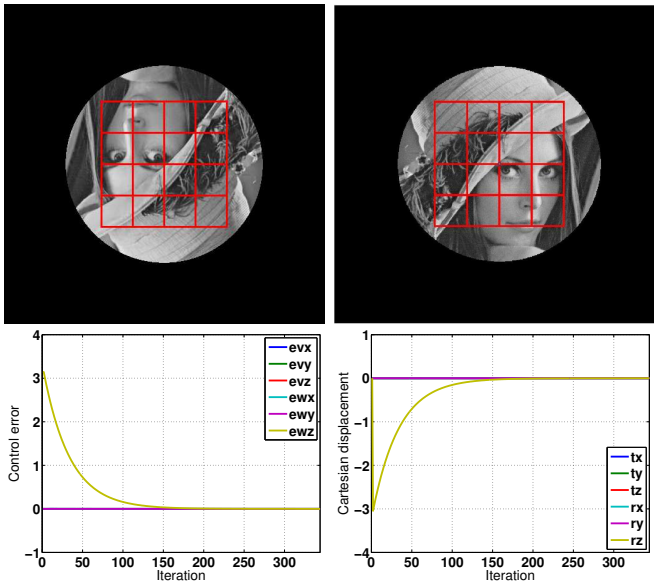

Fig. 2. Direct visual servoing w.r.t. a sphere.

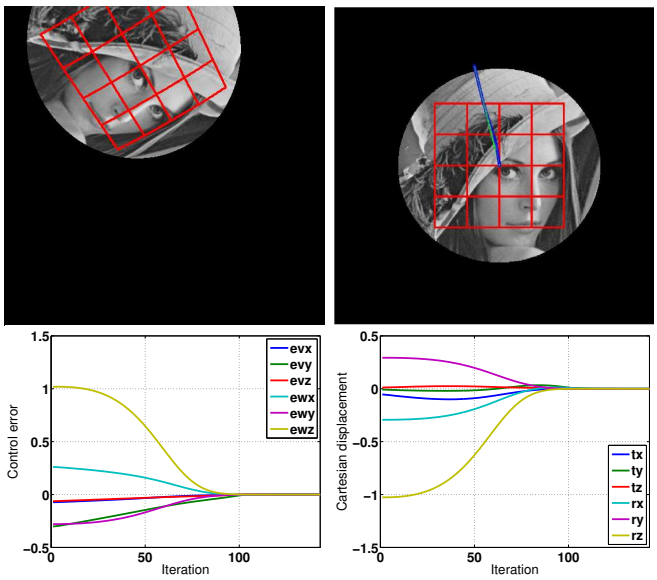

Fig. 3. Direct visual servoing w.r.t. a sphere.

APPENDIX

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

$$\mathcal{X} = \mathbf{R}\mathcal{X}^* + \mathbf{t}. \tag{22}$$

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

$$\mathbf{m}^* = \frac{1}{Z^*}\mathcal{X}^* \quad \text{and} \quad \mathbf{m} = \frac{1}{Z}\mathcal{X}. \tag{23}$$

Before proceeding to the theoretical proofs, let us state an important result which will be used in all subsequent

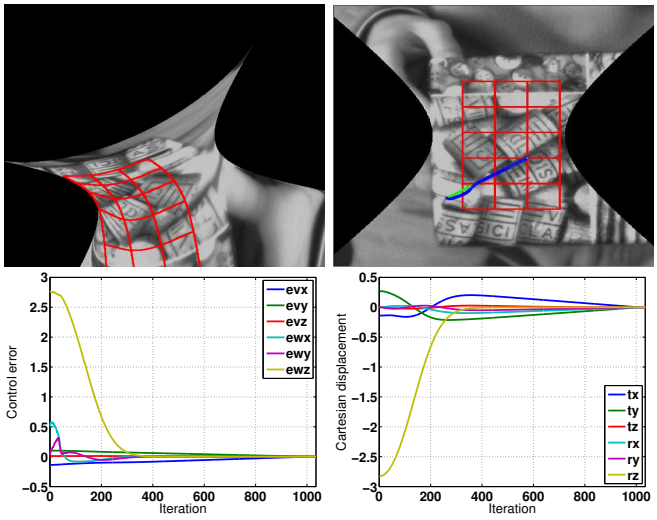

Fig. 4. Direct visual servoing w.r.t. a complex object.

demonstrations. For this, neglect for the moment the errors in the camera internal parameters.

Lemma 1 (Task function and camera pose). *The task function defined in (15) is expressed as a function of the camera pose by using:*

$$\mathbf{e}_\nu = \big((\mathbf{R} - \mathbf{I})\mathcal{X}^* + \mathbf{t}\big)\frac{\beta}{Z^*}, \quad \beta > 0, \tag{24}$$

together with

$$\boldsymbol{\mu} = \sin(\theta)\mathbf{u} + \frac{1}{2}[\mathbf{q}^{*\prime}]_\times \mathbf{t} \tag{25}$$

and Eq. (14). ♢

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. (11) and (12), we have

$$\mathbf{G} = \mathbf{K}\,\mathbf{H}\,\mathbf{K}^{-1} \quad \text{and} \quad \mathbf{e}_p = \mathbf{K}\,\mathbf{e}'_p. \tag{26}$$

By injecting (26) in (2), multiplying both terms of the result by $\mathbf{K}^{-1}$, and then using (6), yields

$$\mathbf{m} \propto \mathbf{H}\,\mathbf{m}^* + \rho^*\mathbf{e}'_p = \alpha\,\mathbf{m}, \quad \alpha > 0. \tag{27}$$

This permits to write

$$\mathbf{e}_\nu = \mathbf{H}\,\mathbf{m}^* + \rho^*\mathbf{e}'_p - \mathbf{m}^* \tag{28}$$

$$= \alpha\,\mathbf{m} - \mathbf{m}^*. \tag{29}$$

From the scale factor

$$\alpha = \beta\frac{Z}{Z^*}, \quad \beta > 0, \tag{30}$$

and Eq. (23), we can rewrite (29) as

$$\mathbf{e}_\nu = (\mathcal{X} - \mathcal{X}^*)\frac{\beta}{Z^*}. \tag{31}$$

Finally, Eq. (22) permits to obtain the desired relation:

$$\mathbf{e}_\nu = \big((\mathbf{R} - \mathbf{I})\mathcal{X}^* + \mathbf{t}\big)\frac{\beta}{Z^*}. \tag{32}$$

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

$$\mathbf{G}_\infty = \mathbf{K}\,\mathbf{R}\,\mathbf{K}^{-1} \quad \text{and} \quad \mathbf{e}_p = \mathbf{K}\,\mathbf{t} \tag{33}$$

and using (11) yield

$$\mathbf{H} = \mathbf{R} + \mathbf{t}\mathbf{q}^{*\prime\top} \tag{34}$$

with

$$\mathbf{q}^{*\prime} = \mathbf{K}^\top \mathbf{q}^*. \tag{35}$$

Then, the projective axis of rotation (13) can be written

$$[\boldsymbol{\mu}]_\times = \frac{1}{2}\big(\mathbf{H} - \mathbf{H}^\top\big) \tag{36}$$

$$= \frac{1}{2}\big(\mathbf{R} + \mathbf{t}\mathbf{q}^{*\prime\top} - \mathbf{R}^\top - \mathbf{q}^{*\prime}\mathbf{t}^\top\big). \tag{37}$$

Using Rodrigues' formula

$$\mathbf{R} = \mathbf{I} + \sin(\theta)[\mathbf{u}]_\times + \big(1 - \cos(\theta)\big)[\mathbf{u}]_\times^2 \tag{38}$$

we have

$$\mathbf{R} - \mathbf{R}^\top = 2\sin(\theta)[\mathbf{u}]_\times, \tag{39}$$

and together with the property

$$\mathbf{t}\mathbf{q}^{*\prime\top} - \mathbf{q}^{*\prime}\mathbf{t}^\top = \big[\,[\mathbf{q}^{*\prime}]_\times \mathbf{t}\,\big]_\times, \tag{40}$$

we obtain the relation:

$$\boldsymbol{\mu} = \sin(\theta)\mathbf{u} + \frac{1}{2}[\mathbf{q}^{*\prime}]_\times \mathbf{t}. \tag{41}$$

The projective angle of rotation follows directly by injecting the norm of (25) in (14):

$$\phi = \begin{cases} \arcsin\big(\|\sin(\theta)\mathbf{u} + \frac{1}{2}[\mathbf{q}^{*\prime}]_\times \mathbf{t}\|\big), \text{ if } \big(\mathrm{tr}(\mathbf{H}) - 1\big)/2 \geq 0, \\ \pi - \arcsin\big(\|\sin(\theta)\mathbf{u} + \frac{1}{2}[\mathbf{q}^{*\prime}]_\times \mathbf{t}\|\big), \qquad \text{otherwise.} \end{cases} \tag{42}$$

with $0 < \|\boldsymbol{\mu}\| \leq 1$ and $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. Finally, by using Eqs. (25), (42) and (34) the desired relation between $\mathbf{e}_\omega$ and the camera pose is achieved. ∎

A. Demonstration of the Theorem 1

This demonstration uses the results from Lemma 1.

PROOF (GLOBAL 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, 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

$$\begin{cases} (\mathbf{R} - \mathbf{I})\mathcal{X}^* + \mathbf{t} = \mathbf{0} \\ \sin(\theta)\mathbf{u} + \dfrac{1}{2}[\mathbf{q}^{*\prime}]_\times \mathbf{t} = \mathbf{0} \\ \big(\mathrm{tr}(\mathbf{H}) - 1\big)/2 \geq 0. \end{cases} \tag{43}$$

The first equation of (43) was constructed using Eqs. (24) together with the facts that $\beta > 0$ and $Z^* > 0$. Thus, one obtains directly

$$\mathbf{t} = (\mathbf{I} - \mathbf{R})\mathcal{X}^* = \big(\mathbf{I} - \exp([\mathbf{u}\theta]_\times)\big)\mathcal{X}^*. \tag{44}$$

The second equation of (43) was constructed by injecting Eqs. (25) and (42) in (15) together with the following facts. The statement $\mathbf{e}_\omega = \boldsymbol{\mu}\phi/\|\boldsymbol{\mu}\| = \mathbf{0}$ implies that $\|\boldsymbol{\mu}\| = 0$ and/or $\phi = 0$. In fact, the latter also implies $\|\boldsymbol{\mu}\| = 0$, as well as the third equation $\big(\mathrm{tr}(\mathbf{H}) - 1\big)/2 \geq 0$, since $0 < \|\boldsymbol{\mu}\| \leq 1$ which implies $0 < \arcsin(\|\boldsymbol{\mu}\|) \leq \pi/2$. Thus, $\phi = 0$ if and only if $\|\boldsymbol{\mu}\| = 0$.

Pre-multiplying the second equation of (43) by $[\mathbf{t}]_\times$ and then injecting (44), one obtains

$$\big[\big(\mathbf{I} - \exp([\mathbf{u}\theta]_\times)\big)\mathcal{X}^*\big]_\times \sin(\theta)\mathbf{u} = \mathbf{0}. \tag{45}$$

Given that by definition $Z^* > 0$, $\big[\big(\mathbf{I} - \exp([\mathbf{u}\theta]_\times)\big)\mathcal{X}^*\big]_\times = \mathbf{0}$ if and only if $\theta = 0$. We need then to evaluate $\sin(\theta)\mathbf{u} = \mathbf{0}$, which clearly gives only two possible solutions, $\theta = 0$ or $\theta = \pi$, since $\mathbf{u}$ is a unit axis of rotation. Both cases imply $\mathbf{q}^{*\prime} = \mathbf{0}$ or $\mathbf{t} = \mathbf{0}$ from the second equation of (43), which in turn permit to make conclusions from the third equation of (43) by using (34):

$$\big(\mathrm{tr}(\mathbf{H}) - 1\big)/2 \geq 0 \implies \big(\mathrm{tr}(\mathbf{R}) - 1\big)/2 \geq 0 \tag{46}$$

$$\implies \cos(\theta) \geq 0 \implies |\theta| \leq \pi/2. \tag{47}$$

Therefore, the only solution to (45) is $\theta = 0$, $\forall \mathbf{q}^*$ and $\forall \mathcal{X}^*$ such that $Z^* > 0$. In addition, by applying this solution to (44), we must have $\mathbf{t} = \mathbf{0}$. ∎

B. Demonstration of the Corollary 1

PROOF (GENERALITY AND IMPROVEMENTS). By generality, we mean that, besides also coping with non-planar objects, the proposed task function (15) comprises the control error as defined in (7) as well, which deals with planar objects. This is easy to show provided the following facts. From Remark 2, our visual tracking method provides the parallax $\rho^* = 0$ if the object is planar and, 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 (35) gives

$$\mathbf{q}^{*\prime} = \mathbf{K}^\top \mathbf{q}^* = \mathbf{K}^\top \mathbf{K}^{-\top}\mathbf{n}^* = \mathbf{n}^*. \tag{48}$$

Then, the proposed task function (15) can be rewritten as

$$\begin{cases} \mathbf{e}_\nu = \mathbf{H}\,\mathbf{m}^* - \mathbf{m}^* = \mathbf{H}_\Pi \mathbf{m}^* - \mathbf{m}^* \\ \mathbf{e}_\omega = \dfrac{\boldsymbol{\mu}}{\|\boldsymbol{\mu}\|}\phi = \dfrac{\boldsymbol{\mu}_\Pi}{\|\boldsymbol{\mu}_\Pi\|}\phi_\Pi, \end{cases} \tag{49}$$

where

$$\big[\boldsymbol{\mu}_\Pi\big]_\times = \Big[\sin(\theta)\mathbf{u} + \frac{1}{2}[\mathbf{n}^*]_\times \mathbf{t}\Big]_\times = \frac{1}{2}\big(\,\mathbf{H}_\Pi - \mathbf{H}_\Pi^\top\big), \tag{50}$$

by injecting (48) in (41), and ϕ_Π can be found accordingly from (42).

TODO: Besides that generality, improvements concerning the rotational control error is attained by determining explicitly in which quadrant the projective angle of rotation operates, instead of not constraining it. This can be disastrous when performing path planning because large rotation motions may be induced and thus, leading possibly to system failure. To see this clearly, let us consider for example the situation where $\mathbf{t} = \mathbf{0}$ and $\theta > \pi/2$. If a rotational control error as defined in (17) (or Eq. (7) for a planar target), then

That is, the second quadrant $\phi = \pi - \arcsin(\|\boldsymbol{\mu}\|)$ is never a possibility.

In this situation, the error in fact increases exponentially. Secondly, the task function (7) and the corresponding (17) are both not defined for $\theta = \pi$, i.e. they do not provide for a global isomorphism.

Furthermore, notice that the knowledge of $\mathbf{q}^{*\prime}$ is neither required nor estimated in our projective framework whilst, in an Euclidean reconstruction framework, $\mathbf{q}^{*\prime} = \mathbf{0}$ and only if perfect camera parameters are available. This statement can be shown directly by injecting Eq. (23) in (22), together with Eqs. (6), (33) and afterward dividing each term by Z^*:

$$\mathbf{p} \propto \mathbf{G}_\infty \mathbf{p}^* + \rho^* \mathbf{e}_p, \tag{51}$$

which gives $\mathbf{G} = \mathbf{G}_\infty$ and $\rho^* = 1/Z^*$ in the Euclidean geometry. This demonstrates that our control error is also a generalization of the hybrid control error $\mathbf{e}'_\Pi = \left[\alpha\mathbf{m} - \mathbf{m}^*, \theta\mathbf{u}^\top\right]^\top \in \mathbb{R}^6$ proposed in [12], without not requiring a coarse 3D estimate of the normal vector to perform the Euclidean reconstruction to determine $\theta\mathbf{u}$. That is, injecting $\mathbf{q}^{*\prime} = \mathbf{0}$ in (42), we arrive at the equivalences $\boldsymbol{\mu}/\|\boldsymbol{\mu}\| = \mathbf{u}$, $\phi = \theta$, as well as it has been shown that our proposed translational error $\mathbf{e}_\nu$ is equivalent to (29). ∎

C. Demonstration of the Theorem 2

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

$$\dot{\mathbf{e}} = \begin{bmatrix} \dot{\mathbf{e}}_\nu \\ \dot{\mathbf{e}}_\omega \end{bmatrix} = \mathbf{L} \begin{bmatrix} \boldsymbol{\nu} \\ \boldsymbol{\omega} \end{bmatrix} = \mathbf{L}\,\mathbf{v} = \mathbf{L}\,\boldsymbol{\Lambda}\,\mathbf{e}, \tag{52}$$

with $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_\nu \mathbf{I}, \lambda_\omega \mathbf{I}) > 0$ and where $\mathbf{L}$ represents the interaction matrix, which is necessary for analysis purposes only.

PROOF (LOCAL STABILITY). The proof consists in analyzing the behavior of the closed-loop system (52) around the equilibrium. First, we need to obtain $\mathbf{L}$ and then to linearize it. That can be achieved by using the results from Lemma 1. Let us start with $\mathbf{e}_\nu$:

$$\dot{\mathbf{e}}_\nu = (\dot{\mathbf{R}}\mathcal{X}^* + \dot{\mathbf{t}})\frac{\beta}{Z^*}. \tag{53}$$

By injecting the well-known relations

$$\dot{\mathbf{t}} = -\boldsymbol{\nu} - [\boldsymbol{\omega}]_\times \mathbf{t} \tag{54}$$

$$\dot{\mathbf{R}} = -[\boldsymbol{\omega}]_\times \mathbf{R} \tag{55}$$

in (53), using (24) and the fact that $[\mathbf{a}]_\times \mathbf{b} = -[\mathbf{b}]_\times \mathbf{a}$, we obtain

$$\dot{\mathbf{e}}_\nu = -\frac{\beta}{Z^*}\boldsymbol{\nu} + \left[\mathbf{e}_\nu + \frac{1}{\beta}\mathbf{m}^*\right]_\times \boldsymbol{\omega}. \tag{56}$$

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

$$\dot{\mathbf{e}}_\omega = \frac{\partial \sin(\theta)\mathbf{u}}{\partial t} + \frac{1}{2}[\mathbf{q}^{*\prime}]_\times \dot{\mathbf{t}}. \tag{57}$$

By using the relation $\frac{\partial \sin(\theta)\mathbf{u}}{\partial t} = -\mathbf{L}_\omega \boldsymbol{\omega}$ with

$$\mathbf{L}_\omega = \mathbf{I} - \frac{\sin(\theta)}{2}[\mathbf{u}]_\times - \sin^2\left(\frac{\theta}{2}\right)(2\mathbf{I} + [\mathbf{u}]_\times^2) \tag{58}$$

and (54), we obtain:

$$\dot{\mathbf{e}}_\omega = -\frac{1}{2}[\mathbf{q}^{*\prime}]_\times \boldsymbol{\nu} - \left(\mathbf{L}_\omega - \frac{1}{2}[\mathbf{q}^{*\prime}]_\times [\mathbf{t}]_\times\right)\boldsymbol{\omega}. \tag{59}$$

By using Eqs. (56), (59) and setting $\beta = 1$ without loss of generality, the interaction matrix is finally given as

$$\mathbf{L} = \begin{bmatrix} -\frac{1}{Z^*}\mathbf{I} & [\mathbf{e}_\nu + \mathbf{m}^*]_\times \\ -\frac{1}{2}[\mathbf{q}^{*\prime}]_\times & -\mathbf{L}_\omega + \frac{1}{2}[\mathbf{q}^{*\prime}]_\times [\mathbf{t}]_\times \end{bmatrix} \tag{60}$$

Then, we may proceed to the evaluation of (52) around $\mathbf{e} = \mathbf{0}$:

$$\dot{\mathbf{e}} = \mathbf{L}\big|_{\mathbf{e}=\mathbf{0}}\,\boldsymbol{\Lambda}\,\mathbf{e} = -\begin{bmatrix} \lambda_\nu \frac{1}{Z^*}\mathbf{I} & -\lambda_\omega [\mathbf{m}^*]_\times \\ \lambda_\nu \frac{1}{2}[\mathbf{q}^{*\prime}]_\times & \lambda_\omega \mathbf{I} \end{bmatrix} \mathbf{e}, \tag{61}$$

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

$$\begin{bmatrix} -\lambda_\omega \\ -\frac{\lambda_\nu}{Z^*} \\ -\frac{\lambda_\omega Z^* + \lambda_\nu - \sqrt{\lambda_\omega^2 Z^{*2} + \lambda_\nu^2 - 2\lambda_\nu\lambda_\omega Z^*(1 - Z^*\mathbf{q}^{*\prime\top}\mathbf{m}^*)}}{2Z^*} \\ -\frac{\lambda_\omega Z^* + \lambda_\nu + \sqrt{\lambda_\omega^2 Z^{*2} + \lambda_\nu^2 - 2\lambda_\nu\lambda_\omega Z^*(1 - Z^*\mathbf{q}^{*\prime\top}\mathbf{m}^*)}}{2Z^*} \\ -\frac{\lambda_\omega Z^* + \lambda_\nu - \sqrt{\lambda_\omega^2 Z^{*2} + \lambda_\nu^2 - 2\lambda_\nu\lambda_\omega Z^*(1 - Z^*\mathbf{q}^{*\prime\top}\mathbf{m}^*)}}{2Z^*} \\ -\frac{\lambda_\omega Z^* + \lambda_\nu + \sqrt{\lambda_\omega^2 Z^{*2} + \lambda_\nu^2 - 2\lambda_\nu\lambda_\omega Z^*(1 - Z^*\mathbf{q}^{*\prime\top}\mathbf{m}^*)}}{2Z^*} \end{bmatrix} \tag{62}$$

Therefore, provided that $\lambda_\nu > 0$, $\lambda_\omega > 0$, $Z^* > 0$ and

$$\lambda_\omega^2 Z^{*2} + \lambda_\nu^2 - 2\lambda_\nu\lambda_\omega Z^*\left(1 - Z^*\mathbf{q}^{*\prime\top}\mathbf{m}^*\right) < (\lambda_\omega Z^* + \lambda_\nu)^2, \tag{63}$$

whose simplification gives

$$\mathbf{q}^{*\prime\top}\mathcal{X}^* < 2, \tag{64}$$

all eigenvalues of $\mathbf{L}\big|_{\mathbf{e}=\mathbf{0}}\,\boldsymbol{\Lambda}$ have negative real part. As a matter of fact, the condition (64) expresses the distance between the chosen point and the reference plane. Given that this reference plane represents the dominant plane[1] of the object in our projective formulation, this condition can be easily satisfied if the point is chosen such that its parallax 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 (52) is always locally 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] G. Silveira, E. Malis, and P. Rives, "Visual servoing over unknown, unstructured, large-scale scenes," in *Proc. of the IEEE International Conference on Robotics and Automation*, USA, 2006, pp. 4142–4147.

[4] 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.

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

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

[6] 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.
[7] E. Malis and P. Rives, "Robustness of image-based visual servoing with respect to depth distribution errors," in *Proc. of the IEEE International Conference on Robotics and Automation*, 2003.
[8] 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.
[9] Y. Mezouar and F. Chaumette, "Path planning for robust image-based control," *IEEE Trans. on Rob. and Autom.*, vol. 18, pp. 534–549, 2002.
[10] S. Benhimane and E. Malis, "Homography-based 2D visual servoing," in *Proc. of the IEEE Int. Conf. on Robotics and Autom.*, USA, 2006.
[11] M. Irani and P. Anandan, "All about direct methods," in *Workshop on Vision Algorithms: Theory and practice*, 1999.
[12] 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.
[13] H. Jin, P. Favaro, and S. Soatto, "A semi-direct approach to structure from motion," *The Visual Computer*, vol. 6, pp. 377–394, 2003.
[14] G. Silveira and E. Malis, "Real-time visual tracking under arbitrary illumination changes," in *Proc. of the IEEE Computer Vision and Pattern Recognition*, USA, 2007.
[15] 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.
[16] E. Malis, "An efficient unified approach to direct image registration of rigid and deformable surfaces," INRIA Sophia-Antipolis, France, Tech. Rep. 6089, 2007.
[17] F. W. Warner, *Foundations of differential manifolds and Lie groups*. Springer Verlag, 1987.
[18] G. Silveira, E. Malis, and P. Rives, "An efficient direct method for improving visual SLAM," in *Proc. of the IEEE International Conference on Robotics and Automation*, Italy, 2007.