

Homography-based 2D Visual Servoing

Selim BENHIMANE

Ezio MALIS

INRIA Sophia Antipolis, FRANCE
Selim.Benhimane@sophia.inria.fr

INRIA Sophia Antipolis, FRANCE
Ezio.Malis@sophia.inria.fr

Abstract—The objective of this paper is to propose a new homography-based approach to image-based visual servoing. The visual servoing method does not need any measure of the 3D structure of the observed target. Only visual information measured from the reference and the current image are needed to compute the task function (isomorphic to the camera pose) and the control law to be applied to the robot. The control law is designed in order to make the task function converge to zero. We provide the theoretical proof of the existence of the isomorphism between the task function and the camera pose and the theoretical proof of the local stability of the control law. The experimental results, obtained with a 6 d.o.f. robot, show the advantages of the proposed method with respect to the existing approaches.

I. INTRODUCTION

Visual servoing is a robotic task that consists in controlling a robot thanks to visual information acquired by one or multiple cameras [11], [12]. This robotic task can be considered as the regulation of a task function $e(\mathbf{q}, t)$ that depends on the robot configuration \mathbf{q} and the time t [20]. In this paper, we consider eye-in-hand visual servoing approaches that use as less as possible 3D information on the observed target. In the literature, the visual servoing methods are generally classified as follows:

- 3D visual servoing: the task function $e(\mathbf{q}, t)$ is expressed in the Cartesian space, i.e. the visual information acquired from the two images (the reference and the current images) are used to reconstruct explicitly the pose (the translation and the rotation in the Cartesian space) of the camera (see for example [23], [18], [1], [22], [14]). The advantage of an explicit estimation of the error in the Cartesian space is the decoupling of the task function, i.e. the camera rotation and the camera translation can be controlled independently one from the other. The camera translation (up to a scale factor) and the camera rotation can be estimated through the Essential matrix [13], [10], [9]. However, the Essential matrix can not be estimated when the target is planar or when the motion done by the camera between the reference and the current pose is a pure rotation. For these reasons, it is better to estimate the camera translation (up to a scale factor) and the camera rotation using a homography matrix [16].

- 2D visual servoing: the task function $e(\mathbf{q}, t)$ is expressed directly in the image, i.e. these visual servoing methods do not need the explicit estimation of the pose error in the Cartesian space (see for example [8], [4]). A task function isomorphic to the camera pose is built. As far as we know, except for some special “ad hoc” target [6], the isomorphism is generally supposed true without any formal proof. The real existence of

the isomorphism avoids situations where the task function is null and the camera is not well positioned [3]. In general, the task function is built using simple image features such as interest points coordinates. Since the control is done in the image, the target has much more chance to remain visible in the image. However, the robot trajectory is not optimal because the task function is not decoupled. Many methods have been proposed in order to obtain a task function as decoupled as possible [5], [21].

- 2D 1/2 visual servoing: the task function $e(\mathbf{q}, t)$ is expressed in the Cartesian space and in the image, i.e. the rotation error is estimated explicitly and the translation error is expressed in the image (see for example [15], [7]). These visual servoing approaches make it possible not only to decouple the rotation and the translation control but also to perform the control in the image. With this approach, it is possible to demonstrate the stability and the robustness of the control law [14].

We notice that, for any of the previous methods, we need a measure (on-line or off-line) of some 3D information concerning the observed target. In the 2D 1/2 visual servoing and 3D visual servoing, the pose reconstruction using the homography estimation is not unique (2 different solutions are possible). In order to choose the right solution, it is necessary to have the normal vector to the target plane. In the 2D visual servoing, when considering for example points as features, the corresponding depths are necessary to have a stable control law [17]. The 3D information can be obtained on-line. However, the price to pay is a time consuming estimation step. For example, when the target is planar, many images are needed to obtain a precise estimation of the normal to the plane.

Our objective is to design a visual servoing method that does not need any measure of the 3D structure of the target and that only needs the reference image and the current image to compute the task function $e(\mathbf{q}, t)$. In this paper, we present a new 2D visual servoing method that makes it possible to control the robot by building a task function isomorphic to the camera pose in the Cartesian space. We have demonstrated that it exists an isomorphism between a task function e (measured using the homography that matches the reference target plane image and the current one) and the camera pose in the Cartesian space (i.e. the task function e is null if and only if the camera is back to the reference pose). Contrarily to the standard 2D visual servoing, we have demonstrated that we do not need to measure any 3D information in order to guarantee the control stability. The computation of the control

law is quite simple (we do not need neither the estimation of an interaction matrix nor the decomposition of a homography) and, similarly to the task function, it does not need any measure of 3D information on the observed target.

For simplicity, in order to introduce our approach, we consider in this paper planar targets with unknown 3D information (i.e. the normal vector to the target plane is unknown). The generalization of the new approach to non-planar targets is straightforward since a homography can also be measured if the target is non-planar [16].

II. THEORETICAL BACKGROUND

As already mentioned in the introduction, we consider visual servoing methods that aim to control a robot thanks to the images acquired by an on-board camera. In other words, the robot is controlled in order to position the current camera frame \mathcal{F} to the reference camera frame \mathcal{F}^* . We suppose that the only available information are an image \mathcal{I}^* of the scene at the reference pose and a current image \mathcal{I} of the observed scene (acquired in real-time).

A. Modeling and notations

Let \mathcal{P} be a point in the 3D space. Its 3D coordinates are $\mathcal{X}^* = [X^* \ Y^* \ Z^*]$ in the reference frame \mathcal{F}^* . Using a perspective projection model, the point projects on a virtual plane perpendicular to the optical axis and distant one-meter from the projection center in the point $\mathbf{m}^* = [x^* \ y^* \ 1]$ verifying:

$$\mathbf{m}^* = \frac{1}{Z^*} \mathcal{X}^* \quad (1)$$

We call \mathcal{I}_m^* the reference image in normalized coordinates. A pinhole camera performs a perspective projection of the point \mathcal{P} on the image plane \mathcal{I}^* [9]. The images coordinates $\mathbf{p}^* = [u^* \ v^* \ 1]$ can be obtained from the normalized coordinates with an affine transformation:

$$\mathbf{p}^* = \mathbf{K} \mathbf{m}^* \quad (2)$$

where the camera intrinsic parameters matrix \mathbf{K} can be written as follows:

$$\mathbf{K} = \begin{bmatrix} f & fs & u_0 \\ 0 & fr & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

where f is the focal length in pixels, s represents the default of orthogonality between the image frame axis, r is the aspect ratio and $[u_0 \ v_0]$ are the coordinates of the principal point (in pixels).

Let $\mathbf{R} \in \text{SO}(3)$ and $\mathbf{t} \in \mathbb{R}^3$ be respectively the rotation and the translation between the two frames \mathcal{F} and \mathcal{F}^* . In the current frame \mathcal{F} , the point \mathcal{P} has the following coordinates $\mathcal{X} = [X \ Y \ Z]$ and we have:

$$\mathcal{X} = \mathbf{R} \mathcal{X}^* + \mathbf{t} \quad (4)$$

Let $\mathbf{u} = [u_x \ u_y \ u_z]$ be the unit vector corresponding to the rotation axis and θ ($\theta \in [-\pi, \pi]$) be the rotation angle. Setting $\mathbf{r} = \theta \mathbf{u}$, we have:

$$\mathbf{R} = \exp([\mathbf{r}]_{\times}) \quad (5)$$

where \exp is the matrix exponential function and where the skew matrix $[\mathbf{r}]_{\times}$ is defined as follows:

$$[\mathbf{r}]_{\times} = \begin{bmatrix} 0 & -r_z & +r_y \\ +r_z & 0 & -r_x \\ -r_y & +r_x & 0 \end{bmatrix} \quad (6)$$

The point \mathcal{X} projects on the current normalized image \mathcal{I}_m in $\mathbf{m} = [x \ y \ 1]$ where:

$$\mathbf{m} = \frac{1}{Z} \mathcal{X} \quad (7)$$

and projects on the current image \mathcal{I} in $\mathbf{p} = [u \ v \ 1]$ where:

$$\mathbf{p} = \mathbf{K} \mathbf{m} \quad (8)$$

B. Projective transformation between two images of a plane

Let us suppose that the point \mathcal{P} belongs to a plane π . Let \mathbf{n}^* be the normal vector to π expressed in the reference frame \mathcal{F}^* and d^* is the distance (at the reference pose) between the plane π and the center of projection. If we choose \mathbf{n}^* such that $\|\mathbf{n}^*\| = \sqrt{\mathbf{n}^{*\top} \mathbf{n}^*} = 1/d^*$ then, we can write:

$$\mathbf{n}^{*\top} \mathcal{X}^* = 1 \quad (9)$$

By using equations (1), (4), (7) and (9), we obtain a the following relationship between \mathbf{m} and \mathbf{m}^* :

$$\frac{Z}{Z^*} \mathbf{m} = \mathbf{H} \mathbf{m}^* \quad (10)$$

where the homography matrix \mathbf{H} can be written as follows:

$$\mathbf{H} = \mathbf{R} + \mathbf{t} \mathbf{n}^{*\top} \quad (11)$$

Note that $\det(\mathbf{H}) > 0$, otherwise the camera has moved through the 3D plane and the target is not visible in the image any more.

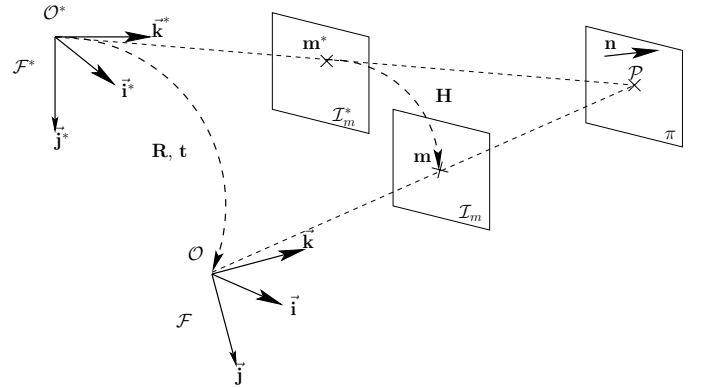


Fig. 1. Projection model and homography between two images of a plane

By using equations (2), (8) and (10), we obtain the following relationship between \mathbf{p} and \mathbf{p}^* :

$$\frac{Z}{Z^*} \mathbf{p} = \mathbf{G} \mathbf{p}^* \quad (12)$$

where the matrix \mathbf{G} can be written as follows:

$$\mathbf{G} = \mathbf{K} \mathbf{H} \mathbf{K}^{-1} \quad (13)$$

Given two images \mathcal{I} and \mathcal{I}^* of a planar target, it is possible to compute the homography matrix. In fact, four non collinear matched points $\{\mathbf{p}_i^*, \mathbf{p}_i\}$, $i \in \{1, 2, 3, 4\}$ suffice to compute \mathbf{G} up to a scale factor. Then, using an approximation of the matrix \mathbf{K} , we compute the matrix \mathbf{H} up to a scale factor. Decomposing the matrix \mathbf{H} to obtain the rotation \mathbf{R} and the translation \mathbf{t} has more than one solution [9]. In general, given the matrix \mathbf{K} , four solutions $\{\mathbf{R}_i, \mathbf{t}_i, \mathbf{n}_i^*\}$, $i \in \{1, 2, 3, 4\}$ are possible but only two are physically admissible. An approximation of the real normal vector \mathbf{n}^* to the target plane makes it possible to choose the right pose.

III. HOMOGRAPHY-BASED 2D VISUAL SERVOING

In this paper, we present a new visual servoing method that does not need any measure of the structure of the observed target. In order to do that, we have to define an isomorphism between the camera pose and the visual information extracted from the reference image and the current image only. Given this isomorphism, we compute a stable control law which also rely on visual information only.

A. Isomorphism between task function and camera pose

The two frames \mathcal{F} and \mathcal{F}^* coincide, if and only if, the matrix \mathbf{H} is equal to the identity matrix \mathbf{I} . Using the homography matrix \mathbf{H} , we build a task function $\mathbf{e} \in \mathbb{R}^6$ locally isomorphic to the camera pose (since we have restricted $\theta \neq \pm\pi$). The task function \mathbf{e} is null, if and only if the camera is back to the reference pose.

Theorem 1: Task function isomorphism.

Let \mathbf{R} be the rotation matrix and \mathbf{t} be the translation vector between \mathcal{F}^* et \mathcal{F} , where $\mathbf{R} = \exp(\theta [\mathbf{u}]_{\times})$, $\theta \in]-\pi, \pi[$ and let $\mathcal{X}^* = [X^* \ Y^* \ Z^*]$ be the coordinates of a certain point $\mathcal{P} \in \pi$ in the reference frame \mathcal{F}^* . We define the task function \mathbf{e} as follows:

$$\mathbf{e} = \begin{bmatrix} e_{\nu} \\ e_{\omega} \end{bmatrix} = \begin{bmatrix} (\mathbf{t} + (\mathbf{R} - \mathbf{I})\mathcal{X}^*)/Z^* \\ 2 \sin(\theta)\mathbf{u} + [\mathbf{n}^*]_{\times} \mathbf{t} \end{bmatrix} \quad (14)$$

where \mathbf{n}^* is the normal vector to the plane π expressed in the reference frame \mathcal{F}^* . The function \mathbf{e} is isomorphic to the camera pose, i.e. $\mathbf{e} = \mathbf{0}$, if and only if, $\theta = 0$ et $\mathbf{t} = \mathbf{0}$.

The proof of the theorem is given in the Appendix. We can demonstrate also that the task function \mathbf{e} can be computed using the two images \mathcal{I} and \mathcal{I}^* only, i.e. without directly measuring the 3D structure of the target (\mathbf{n}^* et Z^*). Given the homography matrix \mathbf{H} , we can write:

$$\mathbf{e}_{\nu} = (\mathbf{H} - \mathbf{I})\mathbf{m}^* \quad (15)$$

$$[\mathbf{e}_{\omega}]_{\times} = \mathbf{H} - \mathbf{H}^{\top} \quad (16)$$

See the Appendix for the proof of these equations. If we have $\mathbf{e}_{\nu} = \mathbf{0}$, then the two projections \mathcal{X}^* and \mathcal{X} of the same 3D point \mathcal{P} coincide. And if we have $\mathbf{e}_{\omega} = \mathbf{0}$, then the homography matrix \mathbf{H} is symmetric.

In this paper, for simplicity reasons, we will consider only this isomorphism. However, it exists a group of isomorphisms

that can be build using the homography matrix \mathbf{H} . For example, we can choose the task function \mathbf{e} as follows:

$$\begin{aligned} \mathbf{e}_{\nu} &= \frac{\overline{\mathbf{m}}^{\top} \mathbf{H} \overline{\mathbf{m}}^*}{\overline{\mathbf{m}}^{\top} \overline{\mathbf{m}}} \overline{\mathbf{m}} - \overline{\mathbf{m}}^* \\ [\mathbf{e}_{\omega}]_{\times} &= \mathbf{H} - \mathbf{H}^{\top} \end{aligned}$$

where $\overline{\mathbf{m}} = \frac{1}{n} \sum_{i=1}^n \mathbf{m}_i$ and $\overline{\mathbf{m}}^* = \frac{1}{n} \sum_{k=1}^n \mathbf{m}_k^*$, and where $\forall i \in [1, n]$, we have that \mathbf{m}_i^* and \mathbf{m}_i are corresponding points. We demonstrate also that this function is isomorphic to the camera pose.

B. The control law

The derivative of the task function with respect to time $\dot{\mathbf{e}}$ can be written as follows:

$$\dot{\mathbf{e}} = \mathbf{L} \begin{bmatrix} \nu \\ \omega \end{bmatrix} \quad (17)$$

where ν is the camera translation velocity, ω is the camera rotation velocity and \mathbf{L} is the (6×6) interaction matrix. The matrix \mathbf{L} can be written as follows:

$$\mathbf{L} = \begin{bmatrix} 1/Z^* & -[\mathbf{e}_{\nu} + \mathbf{m}^*]_{\times} \\ [\mathbf{n}^*]_{\times} & -[\mathbf{n}^*]_{\times} [\mathbf{t}]_{\times} + 2\mathbf{L}_{\omega} \end{bmatrix} \quad (18)$$

where the (3×3) matrix \mathbf{L}_{ω} can be written as follows:

$$\mathbf{L}_{\omega} = \mathbf{I} - \frac{\sin(\theta)}{2} [\mathbf{u}]_{\times} - \sin^2\left(\frac{\theta}{2}\right) (2\mathbf{I} + [\mathbf{u}]_{\times}^2) \quad (19)$$

Theorem 2: Local stability.

The control law:

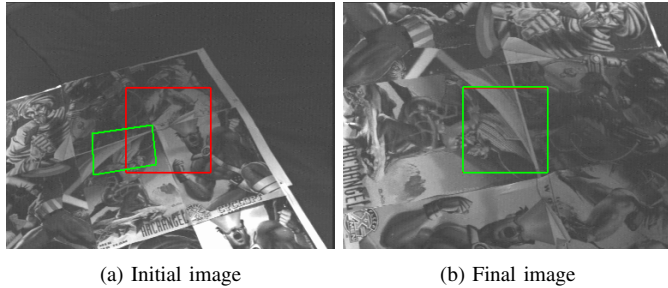
$$\begin{bmatrix} \nu \\ \omega \end{bmatrix} = - \begin{bmatrix} \lambda_{\nu} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \lambda_{\omega} \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{e}_{\nu} \\ \mathbf{e}_{\omega} \end{bmatrix} \quad (20)$$

where $\lambda_{\nu} > 0$ and $\lambda_{\omega} > 0$ is locally stable.

See the Appendix for the proof. This control law only depends on the task function. Consequently, it can be computed using the two images \mathcal{I} and \mathcal{I}^* . The interaction matrix \mathbf{L} does not need to be estimated. It is only useful to analytically prove the stability of the control law. With such control law, the task function \mathbf{e} converges exponentially to $\mathbf{0}$. The local stability of the control law is guaranteed for all \mathbf{n}^* and for all \mathcal{X}^* . By choosing $\lambda_{\nu} > 0$ and $\lambda_{\omega} > 0$ such that $\lambda_{\nu} \neq \lambda_{\omega}$, one can make \mathbf{e}_{ν} and \mathbf{e}_{ω} converge at different speeds.

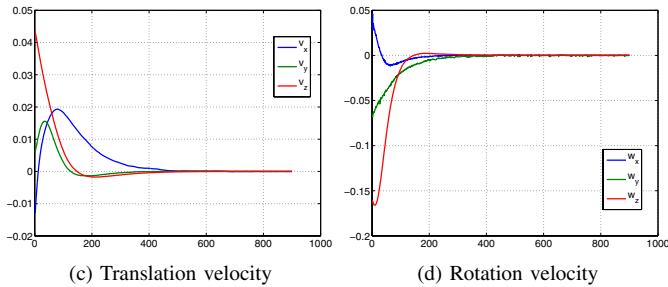
IV. EXPERIMENTAL RESULTS

We have tested the proposed visual servoing on the 6 d.o.f. robot of the LAGADIC research team at IRISA/INRIA Rennes. The robot is accurately calibrated and it provides a ground truth for measuring the accuracy of the positioning task. A calibrated camera is mounted on the end-effector of the robot. A reference image is captured at the reference pose. The positioning task is done with respect to a planar target. Starting from another pose (the initial pose) that makes it possible to see the object from a different angle, the robot is controlled using the control law (20) with $\lambda_{\nu} = \lambda_{\omega} = 0.1$ in order to get back to reference pose. At the initial pose (the translation displacement is 0.68 meters and the rotation displacement



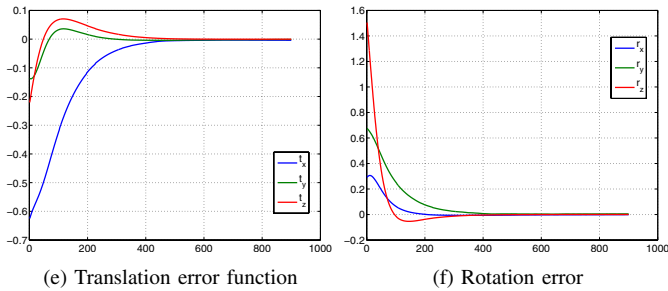
(a) Initial image

(b) Final image



(c) Translation velocity

(d) Rotation velocity



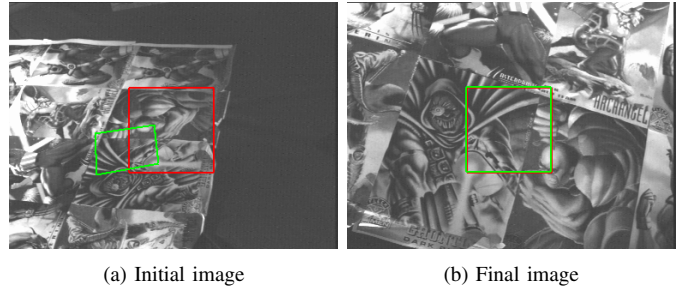
(e) Translation error function

(f) Rotation error

Fig. 2. Experiment 1: Camera positioning with respect to a planar object without approximating the normal vector to the object plane.

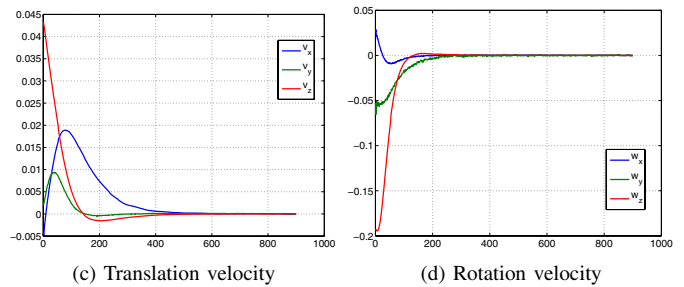
is 96 degrees), we can see the projective transformation of the area of interest (see the red rectangle in the figures 2(b) and 2(a)). We use the ESM¹ visual tracking algorithm [2] to track the zone of interest and to estimate at the same time the homography matrix \mathbf{H} . Given the matrix \mathbf{H} , the control law is computed. We use as control point (\mathbf{m}^* in the equation (15)) the center of gravity of the zone. At the convergence, the robot is back to its reference pose and the visual information coincide with the visual information of the reference pose (see figure 2(b)). The control law is stable: the translation figures 2(c) and the rotation 2(d) velocities converge to zero. As shown in figures 2(e) and 2(f), the camera displacement converge to zero very accurately (less than 1 mm error for the translation and less that 0.1 degrees for the rotation). A second experiment is performed under similar conditions (the same initial camera displacement, an unknown normal vector to the plane, an unknown camera/object distance...). Contrarily to the previous experiment, the positioning task is done with respect to a different target (see figure 3(a)). We also use a very bad estimation of the camera parameters: $\hat{f} = 800$,

¹the ESM visual tracking software can be downloaded from the following web-page: <http://www-sop.inria.fr/icare/personnel/malis/software/ESM.html>



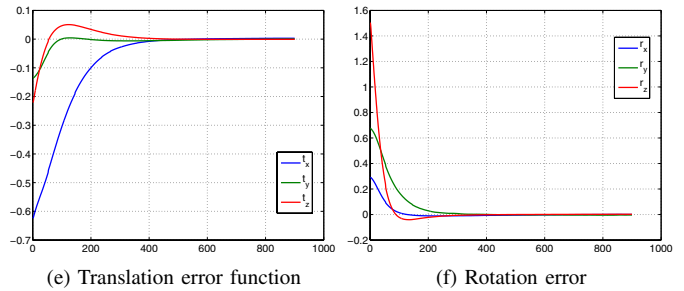
(a) Initial image

(b) Final image



(c) Translation velocity

(d) Rotation velocity



(e) Translation error function

(f) Rotation error

Fig. 3. Experiment 2: Camera positioning with an uncalibrated camera without approximating the normal vector to the object plane.

$\hat{r} = 0.5$, $\hat{u}_0 = 100$, $\hat{v}_0 = 200$ (the calibrated parameters were $f = 592$, $r = 0.96$, $u_0 = 198$, $v_0 = 140$). Figures 3(c) and 3(d) show that the control law is robust to camera calibration errors: the translation and the rotation velocities converge to zero. At the convergence, the visual information coincide with the visual information of the reference image (see figure 3(b)). Again, figures 3(e) and 3(f) show that the camera displacement converges to zero.

V. CONCLUSIONS

In this paper, we have presented for the first time a homography-based 2D approach to visual servoing that do not need any measure of the 3D structure of the observed target. We have presented a simple and stable control law. We think that this approach can open new research directions in the field of vision-based robot control. Indeed, as far as we know, none of the existent methods make it possible to position a robot with respect to an object without measuring, on-line or off-line, some information on its 3D structure. Many improvements of the proposed method can be studied. For example, a great robustness to errors on the camera intrinsic parameters has been observed in the experiments. However, this robustness has not been analytically proved

yet. In addition, the experiments have shown that the stability region is very big, but the true stability region is unknown at the moment. Similarly to [19], we can use trajectory planning in order to have a bigger stability region and to take into account visibility constraints.

APPENDIX

A. The task function is a function of image measures only

Using the equation (14), the vector \mathbf{e}_ν can be written as follows:

$$\mathbf{e}_\nu = (\mathbf{t} + (\mathbf{R} - \mathbf{I})\mathcal{X}^*)/Z^* = (\mathbf{R}\mathcal{X}^* + \mathbf{t} - \mathcal{X}^*)/Z^*$$

Using the equation (4), \mathbf{e}_ν becomes:

$$\mathbf{e}_\nu = (\mathcal{X} - \mathcal{X}^*)/Z^*$$

Plugging equations (1) and (7) gives:

$$\mathbf{e}_\nu = \frac{Z}{Z^*}\mathbf{m} - \mathbf{m}^*$$

Thanks to (10), \mathbf{e}_ν can be written using \mathbf{H} and \mathbf{m}^* only:

$$\mathbf{e}_\nu = \mathbf{H}\mathbf{m}^* - \mathbf{m}^* = (\mathbf{H} - \mathbf{I})\mathbf{m}^*$$

Thanks to equation (11), we have:

$$\mathbf{H} - \mathbf{H}^\top = \mathbf{R} + \mathbf{t}\mathbf{n}^{*\top} - \mathbf{R}^\top - \mathbf{n}^*\mathbf{t}^\top$$

Using the Rodriguez formula for the rotation matrix \mathbf{R} :

$$\mathbf{R} = \mathbf{I} + \sin(\theta) [\mathbf{u}]_\times + 2 \cos^2\left(\frac{\theta}{2}\right) [\mathbf{u}]_\times^2$$

we can write:

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

Given the following property:

$$\mathbf{t}\mathbf{n}^{*\top} - \mathbf{n}^*\mathbf{t}^\top = [[\mathbf{n}^*]_\times \mathbf{t}]_\times$$

The antisymmetric part of the matrix \mathbf{H} can be written as:

$$\mathbf{H} - \mathbf{H}^\top = [2 \sin(\theta)\mathbf{u} + [\mathbf{n}^*]_\times \mathbf{t}]_\times$$

Consequently, given equation (14), we have:

$$\mathbf{H} - \mathbf{H}^\top = [\mathbf{e}_\omega]_\times$$

B. The task function is isomorphic to the camera pose

In order to simplify the proof of Theorem (1), we proof three simpler propositions.

Proposition 1:

The matrix $\mathbf{H}\mathbf{H}^\top$ has one eigenvalue equal to 1. The eigenvector corresponding to the eigenvalue is $\mathbf{v} = [\mathbf{R}\mathbf{n}^*]_\times \mathbf{t}$.

Proof of proposition 1:

Using equation (11), we have:

$$\mathbf{H}\mathbf{H}^\top = (\mathbf{R} + \mathbf{t}\mathbf{n}^{*\top})(\mathbf{R}^\top + \mathbf{n}^*\mathbf{t}^\top)$$

Since we have $\mathbf{R} \in \mathbb{S}\mathbb{O}(3)$ then $\mathbf{R}\mathbf{R}^\top = \mathbf{I}$. Thus, we have:

$$\mathbf{H}\mathbf{H}^\top = \mathbf{I} + \mathbf{t}(\mathbf{R}\mathbf{n}^*)^\top + (\mathbf{R}\mathbf{n}^* + \|\mathbf{n}^*\|^2\mathbf{t})\mathbf{t}^\top$$

The matrix $\mathbf{H}\mathbf{H}^\top$ is the sum of \mathbf{I} and a rank 2 matrix. Thus, one eigenvalue of $\mathbf{H}\mathbf{H}^\top$ is equal to 1. Setting $\mathbf{v} = [\mathbf{R}\mathbf{n}^*]_\times \mathbf{t}$, we have:

$$(\mathbf{R}\mathbf{n}^*)^\top \mathbf{v} = 0 \quad \text{and} \quad \mathbf{t}^\top \mathbf{v} = 0$$

showing that \mathbf{v} is an eigenvector of $\mathbf{H}\mathbf{H}^\top$:

$$\mathbf{H}\mathbf{H}^\top \mathbf{v} = \mathbf{v}$$

Proposition 2:

If $\mathbf{H} = \mathbf{H}^\top$ and $\sin(\theta) \neq 0$, then $\mathbf{n}^{*\top} \mathbf{u} = 0$, $\mathbf{t}^\top \mathbf{u} = 0$ and $\mathbf{n}^{*\top} \mathbf{v} = 0$ (where $\mathbf{v} = [\mathbf{R}\mathbf{n}^*]_\times \mathbf{t}$).

Proof of proposition 2:

If we have $\mathbf{H} = \mathbf{H}^\top$, then we have:

$$2 \sin(\theta)\mathbf{u} + [\mathbf{n}^*]_\times \mathbf{t} = \mathbf{0} \quad (21)$$

By multiplying each side of the equation (21) by $\mathbf{n}^{*\top}$, we obtain:

$$2 \sin(\theta)\mathbf{n}^{*\top} \mathbf{u} = \mathbf{0}$$

Since we have supposed that $\sin(\theta) \neq 0$, we have:

$$\mathbf{n}^{*\top} \mathbf{u} = 0$$

Similarly, by multiplying each side of the equation (21) by \mathbf{t}^\top , we obtain:

$$\mathbf{t}^\top \mathbf{u} = 0$$

Finally, using the Rodriguez formula for the rotation matrix, we have:

$$\begin{aligned} \mathbf{R}\mathbf{n}^* &= \left(\mathbf{I} + \sin(\theta) [\mathbf{u}]_\times + 2 \cos^2\left(\frac{\theta}{2}\right) [\mathbf{u}]_\times^2 \right) \mathbf{n}^* \\ &= \mathbf{n}^* + \sin(\theta) [\mathbf{u}]_\times \mathbf{n}^* + 2 \cos^2\left(\frac{\theta}{2}\right) [\mathbf{u}]_\times^2 \mathbf{n}^* \\ &= \mathbf{n}^* + \sin(\theta) [\mathbf{u}]_\times \mathbf{n}^* + 2 \cos^2\left(\frac{\theta}{2}\right) (\mathbf{u}\mathbf{u}^\top - \mathbf{I}) \mathbf{n}^* \end{aligned}$$

If we have $\mathbf{n}^{*\top} \mathbf{u} = 0$, then we have:

$$\mathbf{R}\mathbf{n}^* = \mathbf{n}^* + \sin(\theta) [\mathbf{u}]_\times \mathbf{n}^* - 2 \cos^2\left(\frac{\theta}{2}\right) \mathbf{n}^* \quad (22)$$

The antisymmetric matrix associated to the vector $\mathbf{R}\mathbf{n}^*$ is:

$$[\mathbf{R}\mathbf{n}^*]_\times = [\mathbf{n}^*]_\times + \sin(\theta) [[\mathbf{u}]_\times \mathbf{n}^*]_\times - 2 \cos^2\left(\frac{\theta}{2}\right) [\mathbf{n}^*]_\times$$

and since $[[\mathbf{u}]_\times \mathbf{n}^*]_\times = \mathbf{n}^*\mathbf{u}^\top - \mathbf{u}\mathbf{n}^{*\top}$, we can write:

$$[\mathbf{R}\mathbf{n}^*]_\times = [\mathbf{n}^*]_\times + \sin(\theta) (\mathbf{n}^*\mathbf{u}^\top - \mathbf{u}\mathbf{n}^{*\top}) - 2 \cos^2\left(\frac{\theta}{2}\right) [\mathbf{n}^*]_\times$$

By multiplying both sides of the equation by $\mathbf{n}^{*\top}$, we obtain:

$$\mathbf{n}^{*\top} [\mathbf{R}\mathbf{n}^*]_\times = \|\mathbf{n}^*\|^2 \sin(\theta)\mathbf{u}^\top \quad (23)$$

By multiplying both sides of the equation by \mathbf{t} , we obtain:

$$\mathbf{n}^{*\top} [\mathbf{R}\mathbf{n}^*]_\times \mathbf{t} = \|\mathbf{n}^*\|^2 \sin(\theta)\mathbf{u}^\top \mathbf{t}$$

Since $\mathbf{u}^\top \mathbf{t} = 0$, then we prove that:

$$\mathbf{n}^{*\top} \mathbf{v} = 0$$

Proposition 3:

If $\mathbf{H} = \mathbf{H}^\top$, $\mathbf{v} = [\mathbf{Rn}^*]_\times \mathbf{t} = \mathbf{0}$ and $\sin(\theta) \neq 0$ then $\det(\mathbf{H}) = -1$.

Proof of proposition 3:

If $\mathbf{v} = [\mathbf{Rn}^*]_\times \mathbf{t} = \mathbf{0}$ then it exists $\alpha > 0$ such that:

$$\mathbf{t} = \alpha \mathbf{Rn}^*$$

From equation (23), we obtain:

$$[\mathbf{n}^*]_\times \mathbf{Rn}^* = (\mathbf{n}^{*\top} [\mathbf{Rn}^*]_\times)^\top = \|\mathbf{n}^*\|^2 \sin(\theta) \mathbf{u} \quad (24)$$

Then, from equation (21) and equation (24), we obtain:

$$2 \sin(\theta) \mathbf{u} = -[\mathbf{n}^*]_\times \mathbf{t} = -\alpha [\mathbf{n}^*]_\times \mathbf{Rn}^* = -\alpha \|\mathbf{n}^*\|^2 \sin(\theta) \mathbf{u}$$

By multiplying both sides of this equation by \mathbf{u}^\top , we obtain:

$$2 \sin(\theta) = -\alpha \sin(\theta) \|\mathbf{n}^*\|^2$$

Since we supposed $\sin(\theta) \neq 0$, then we can write:

$$\alpha = -\frac{2}{\|\mathbf{n}^*\|^2}$$

and finally the determinant of the matrix \mathbf{H} verifies:

$$\det(\mathbf{H}) = 1 + \mathbf{n}^{*\top} \mathbf{R}^\top \mathbf{t} = 1 + \alpha \|\mathbf{n}^*\|^2 = -1$$

Having a matrix \mathbf{H} with negative determinant means that current frame \mathcal{F} is on the opposite side of the target plane. This is impossible since it means that we cannot see the target in the image any more. This is the reason why $\det(\mathbf{H}) > 0$.

Proof of theorem 1:

It is evident that if $\theta = 0$ and $\mathbf{t} = \mathbf{0}$ then $\mathbf{e} = \mathbf{0}$. We must prove now that if $\mathbf{e} = \mathbf{0}$, then $\theta = 0$ and $\mathbf{t} = \mathbf{0}$. Let us suppose that $\mathbf{e} = \mathbf{0}$. It is evident that if $\theta = 0$ then $\mathbf{t} = \mathbf{0}$ and if $\mathbf{t} = \mathbf{0}$ then $\theta = 0$. Now, let us suppose that $\mathbf{e} = \mathbf{0}$ and $\mathbf{t} \neq \mathbf{0}$ and $\theta \neq 0$. If $\mathbf{e}_\nu = \mathbf{0}$ then $\mathbf{Hm}^* = \mathbf{m}^*$. Thus, \mathbf{H} has an eigenvalue equal to 1 and the vector \mathbf{m}^* is the corresponding eigenvector. The vector \mathbf{m}^* is also eigenvector corresponding to the eigenvalue 1 of the matrix \mathbf{H}^2 . Since $\mathbf{e}_\omega = \mathbf{0}$ then $\mathbf{H} = \mathbf{H}^\top$ and $\mathbf{H}^2 = \mathbf{H}\mathbf{H}^\top$. Given Proposition 1, \mathbf{m}^* is then collinear to the vector $\mathbf{v} = [\mathbf{Rn}^*]_\times \mathbf{t}$. Since $\det(\mathbf{H}) > 0$, this vector is different from zeros (see Proposition 3). On the other hand, Proposition 2 shows that in this case $\mathbf{n}^{*\top} \mathbf{m}^* = Z^* = 0$. This is impossible since by definition $Z^* > 0$. Thus, it is impossible that $\mathbf{e} = \mathbf{0}$ and $\mathbf{t} \neq \mathbf{0}$, $\theta \neq 0$.

C. Proof of the local stability of the control law

Proof of theorem 2:

After linearizing equation (17) about $\mathbf{e} = \mathbf{0}$, we obtain the following linear system:

$$\dot{\mathbf{e}} = - \begin{bmatrix} \lambda_t/Z^* & -\lambda_r [\mathbf{m}^*]_\times \\ \lambda_t [\mathbf{n}^*]_\times & 2\lambda_r \mathbf{I} \end{bmatrix} \mathbf{e} = -\mathbf{L}_0 \mathbf{e}$$

The eigenvectors of the constant matrix \mathbf{L}_0 are: 2λ , $4Z^*$, $2Z^* + \lambda + \sqrt{\lambda^2 + 4Z^{*2}}$ (twice), $2Z^* + \lambda - \sqrt{\lambda^2 + 4Z^{*2}}$

(twice), where $\lambda = \lambda_\nu/\lambda_\omega$. Since $\lambda > 0$ and $Z^* > 0$, the eigenvalues of matrix \mathbf{L}_0 are always positives. Consequently, the control law defined in equation (20) is always locally stable for any \mathbf{n}^* and any \mathbf{m}^* .

REFERENCES

- [1] R. Basri, E. Rivlin, and I. Shimshoni. Visual homing: Surfing on the epipoles. In *IEEE Int. Conf. on Computer Vision*, pages 863–869, 1998.
- [2] S. Benhimane and E. Malis. Real-time image-based tracking of planes using efficient second-order minimization. In *IEEE/RSJ Int. Conf. on Intelligent Robots Systems*, pages 943–948, 2004.
- [3] F. Chaumette. Potential problems of stability and convergence in image-based and position-based visual servoing. In D. Kriegman, G. Hager, and A. Morse, editors, *The confluence of vision and control*, volume 237 of *LNCIS Series*, pages 66–78. Springer Verlag, 1998.
- [4] F. Chaumette. Image moments: a general and useful set of features for visual servoing. *IEEE Trans. on Robotics*, 20(4):713–723, August 2004.
- [5] P. Corke and S. Hutchinson. A new partitioned approach to image-based visual servo control. *IEEE Trans. on Robotics and Automation*, 14(4):507–515, 2001.
- [6] N. J. Cowan and D. E. Chang. Toward geometric visual servoing. In A. Bicchi, H. Christensen, and D. Prattichizzo, editors, *Control Problems in Robotics*, volume 4 of *STAR, Springer Tracks in Advanced Robotics*, pages 233–248. Springer Verlag, 2002.
- [7] K. Deguchi. Optimal motion control for image-based visual servoing by decoupling translation and rotation. In *IEEE Int. Conf. on Intelligent Robots and Systems*, volume 2, pages 705–711, 1998.
- [8] B. Espiau, F. Chaumette, and P. Rives. A new approach to visual servoing in robotics. *IEEE Trans. on Rob. and Aut.*, 8(3):313–326, 1992.
- [9] O. Faugeras. *Three-dimensional computer vision: a geometric viewpoint*. MIT Press, Cambridge, MA, 1993.
- [10] R. Hartley. Estimation of relative camera positions for uncalibrated cameras. In G. Sandini, editor, *European Conf. on Computer Vision*, volume 588 of *Lecture Notes in Computer Science*, pages 579–587. Springer-Verlag, 1992.
- [11] K. Hashimoto. *Visual Servoing: Real Time Control of Robot manipulators based on visual sensory feedback*, volume 7 of *World Scientific Series in Robotics and Automated Systems*. World Scientific Press, Singapore, 1993.
- [12] S. Hutchinson, G. D. Hager, and P. I. Corke. A tutorial on visual servo control. *IEEE Trans. on Robotics and Automation*, 12(5):651–670, 1996.
- [13] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.
- [14] 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*, 18(2):176–186, 2002.
- [15] E. Malis, F. Chaumette, and S. Boudet. 2 1/2 d visual servoing. *IEEE Trans. on Robotics and Automation*, 15(2):234–246, April 1999.
- [16] E. Malis, F. Chaumette, and S. Boudet. 2 1/2 d visual servoing with respect to unknown objects through a new estimation scheme of camera displacement. *Int. Journal of Computer Vision*, 37(1):79–97, 2000.
- [17] 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.
- [18] P. Martinet, N. Daucher, J. Gallice, and M. Dhome. Robot control using monocular pose estimation. In *Workshop on New Trends In Image-Based Robot Servoing*, pages 1–12, Grenoble, France, September 1997.
- [19] Y. Mezour and F. Chaumette. Path planning for robust image-based control. *IEEE Trans. on Robotics and Automation*, 18(4):534–549, 2002.
- [20] C. Samson, M. Le Borgne, and B. Espiau. *Robot Control: the Task Function Approach*, volume 22 of *Oxford Engineering Science Series*. Clarendon Press, Oxford, UK, 1991.
- [21] O. Tahri and F. Chaumette. Image moments: Generic descriptors for decoupled image-based visual servo. In *IEEE Int. Conf. on Robotics and Automation, ICRA'04*, pages 1185–1190, 2004.
- [22] C. Taylor, J. Ostrowski, and S. Jung. Robust vision-based pose control. In *IEEE Int. Conf. on Robotics and Automation*, pages 2734–2740, 2000.
- [23] W. J. Wilson, C. C. W. Hulls, and G. S. Bell. Relative end-effector control using Cartesian position-based visual servoing. *IEEE Trans. on Robotics and Automation*, 12(5):684–696, 1996.