Visual Servoing Invariant to Changes in Camera-Intrinsic Parameters

Ezio Malis

▶ **To cite this version:**

Ezio Malis. Visual Servoing Invariant to Changes in Camera-Intrinsic Parameters. IEEE Transactions on Robotics and Automation, 2004, 20 (1), pp.72-81. 10.1109/TRA.2003.820847 . hal-04654359

HAL Id: hal-04654359
https://inria.hal.science/hal-04654359v1

Submitted on 19 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Visual servoing invariant to changes in camera intrinsic parameters

Ezio Malis

Abstract— **This paper presents a new visual servoing scheme which is invariant to changes in camera intrinsic parameters. Current visual servoing techniques are based on the learning of a reference image with the same camera used during the servoing. With the new method it is possible to position a camera (with eventually varying intrinsic parameters), with respect to a non-planar object, given a "reference image" taken with a completely different camera. The necessary and sufficient conditions for the local asymptotic stability show that the control law is robust in the presence of large calibration errors. Local stability implies that the system can accurately track a path in the invariant space. The path can be chosen such that the camera follows a straight line in the Cartesian space. Simple sufficient conditions are given in order to keep the tracking error bounded. This promising approach has been successfully tested with an eye-in-hand robotic system.**

Keywords—**Vision, control, stability analysis, projective invariance, zooming camera, path planning.**

I. Introduction

The aim of visual servoing is to control the movement of a robot using the information provided by vision sensors. A typical task is to position an eye-in-hand system with respect to an observed object. Vision-based control has been widely investigated in the last few years [1], [2]. Despite the diversity of the approaches proposed to accomplish a positioning task, most visual servoing techniques are based on a "teaching-by-showing" approach. With this approach, the robot is moved to a goal position, the camera is shown the target view and a "reference image" of the object is stored (i.e. a set of features of the reference image). The position of the camera with respect to the object will be called the "reference position". After the camera and/or the object has been moved, several visual servoing schemes have been proposed in order to position the camera with respect to the object: image-based visual servoing [3], position-based visual servoing [4], hybrid visual servoing [5]. Each scheme has its own advantages and drawbacks [6] [7], but for all schemes if the visual features currently observed in the image coincide with the features extracted from the reference image, the camera is back to the reference position with respect to the object. Generally speaking, whatever is the visual servoing method used to achieve the task, that will be true if and only if the camera intrinsic parameters at the convergence are the same parameters of the camera used for learning. Indeed, if the camera intrinsic parameters change during the servoing (or the camera used during the servoing is different from the camera used to learn the reference image), even if the current image coincide with the reference image, the position of the camera with respect to the object will be completely different from the reference position. A possible solution to the problem could be to use camera self-calibration techniques [8], [9] in order to measure the camera intrinsic parameters. However, such techniques can fail even for simple camera motions which are very frequent in practice. For example, it has been shown in [10] that if the motion of the camera is a simple pure translation it is impossible to measure the camera internal parameters. The aim of this paper is to increase the versatility of visual servoing by extending the teaching-by-showing technique to the case when different cameras are used for learning the reference image and for servoing, without their explicit calibration. Indeed, intrinsic parameters may significantly vary during the life of the vision system and/or they can be changed intentionally when using zooming cameras. If so, with current visual servoing techniques the reference image must be shown again. In some applications, learning again the reference image could be very difficult. On the other hand, the visual servoing technique proposed in this paper allows us to learn the reference image once and for all. The basic idea is to use projective invariance so as to build a task function, from only measured image features, which is invariant on the intrinsic parameters of the camera. Projective invariance has been used in [11] to define set-points for stereo visual control that do not depend on the viewing location. The system proposed in [11] is made up of two cameras (not mounted on the robot) observing both the object and the robot end-effector. Thus, it is possible to move a point on the manipulator to a point on the object independently from the cameras used to accomplish the task. In this paper, the proposed approach is completely different since the system is made up of only one camera mounted on a robot manipulator. Projective invariance is used to define an error which is dependent on the viewing location but it does not depend on the camera intrinsic parameters. The projective transformation proposed in [12] is used to define a projective space only depending on the position of the camera with respect to an object. Thus, the camera can be repositioned with respect to a non-planar object given a reference image learned with a different camera and even if the camera parameters change during the servoing [12]. Even if the measured task function does not depend on the camera intrinsic parameters, they are needed to estimate the interaction matrix which links the camera velocity to the displacements of the features in the invariant space [13]. Thus, calibration errors can affect the stability of the control law. For this reason, it is important to study the stability of the proposed method and formally prove its robustness with respect to camera calibration errors [14]. In particular, the necessary and sufficient conditions for the

Ezio Malis is with I.N.R.I.A. - ICARE Project, 2004, route des Lucioles - B.P. 93, 06902 Sophia Antipolis Cedex, France.

local asymptotic stability in the presence of large calibration errors are obtained. Local stability implies that the system can accurately follow a path in the invariant space. Simple sufficient conditions are given in order to keep the tracking error bounded. The path can be chosen such that the robot follows a straight line even if the camera parameters are unknowns and the camera is not directly controlled in the Cartesian space. Path planning is not used here to solve the problem of the target visibility during the servoing. Indeed, the scheme proposed in this paper can be used with a zooming camera and the visibility problem can be partially solved by decreasing the focal (i.e. zooming out) length in order to obtain a larger field of view. However, in some situations the zoom may not be sufficient to keep the target in the field of view of the camera. In that case, the visibility constraint can be imposed during the path planning as in [15].

II. Theoretical background

A. Notations

Let the point $\mathcal{C}^*$ of the 3D Cartesian space, be the center of projection (see Figure 1). Suppose that $\mathcal{C}^*$ coincide with the origin $\mathcal{O}^*$ of the absolute frame $\mathcal{F}^*$. Let the plane of projection be parallel to the plane $(\overrightarrow{x}, \overrightarrow{y})$. The 3D point, with homogeneous coordinates $\boldsymbol{\mathcal{X}} = (X, Y, Z, 1)$ is projected in $\mathcal{F}^*$ to the point $\mathbf{m}^*$:

$$\zeta^* \mathbf{m}^* = \left[\begin{array}{cc} \mathbf{I}_3 & \mathbf{0} \end{array} \right] \boldsymbol{\mathcal{X}} \tag{1}$$

where ζ^* is the positive depth. Let the point $\mathcal{C}$ be a different center of projection and the origin of frame $\mathcal{F}$. Let $\mathbf{t}$ and $\mathbf{R}$ be respectively the translation and the rotation between $\mathcal{F}^*$ and $\mathcal{F}$. Let $\mathbf{r} = (r_x, r_y, r_z) = \theta \mathbf{u}$ be the (3×1) vector containing the axis of rotation $\mathbf{u}$ and the angle of rotation θ $(0 \leq \theta < 2\pi)$. If $[\mathbf{r}]_\times$ is the skew symmetric matrix associated to vector $\mathbf{r}$, then $\mathbf{R} = \exp([\mathbf{r}]_\times)$. Let $\boldsymbol{\xi} = (\mathbf{t}, \mathbf{r})$ be the (6×1) vector containing global coordinates of an open subset $\mathcal{S} \subset \mathbb{R}^3 \times SO(3)$. Then, the reference position is $\boldsymbol{\xi}^* = 0$. The 3D point projects in $\mathcal{F}$ to the point $\mathbf{m}$:

$$\zeta(\boldsymbol{\xi})\, \mathbf{m}(\boldsymbol{\xi}) = \left[\begin{array}{cc} \mathbf{R} & \mathbf{t} \end{array} \right] \boldsymbol{\mathcal{X}} \tag{2}$$

where $\zeta(\boldsymbol{\xi})$ is the positive depth. Eliminating $\boldsymbol{\mathcal{X}}$ from equations (1) and (2) we obtain:

$$\zeta(\boldsymbol{\xi})\, \mathbf{m}(\boldsymbol{\xi}) = \zeta^*\, \mathbf{R}\, \mathbf{m}^* + \mathbf{t} \tag{3}$$

This fundamental equation links the perspective projections of the same 3D point in two different frames.

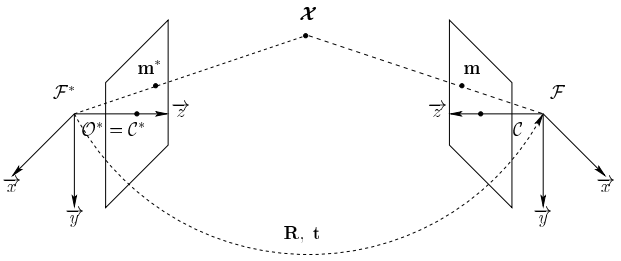


Fig. 1. Perspective projections $\mathbf{m}^*$ and $\mathbf{m}$ of the same 3D point $\boldsymbol{\mathcal{X}}$ from two different points of view $\mathcal{C}^*$ and $\mathcal{C}$. The displacement of current frame $\mathcal{F}$, with respect to the reference frame $\mathcal{F}^*$, is represented by the translation $\mathbf{t}$ and the rotation $\mathbf{R}$.

B. Camera Model

Pinhole cameras perform a perspective projection of a 3D point. However, vectors $\mathbf{m}$ and $\mathbf{m}^*$ are not directly measured by the cameras. Indeed, the point $\mathbf{p} = (u, v, 1)$ (where u and v are the pixel coordinates) observed in the image $\mathcal{I}$, taken at the position $\mathcal{F}$, depends on the camera internal parameters [16]:

$$\mathbf{p}(\boldsymbol{\xi}, \mathbf{K}) = \mathbf{K}\mathbf{m}(\boldsymbol{\xi}) \qquad \mathbf{p} \in \mathcal{I}(\boldsymbol{\xi}, \mathbf{K}) \tag{4}$$

where:

$$\mathbf{K} = \left[\begin{array}{ccc} fk_u & -fk_u/\tan(\theta) & u_0 \\ 0 & fk_v/\sin(\theta) & v_0 \\ 0 & 0 & 1 \end{array} \right] \tag{5}$$

u_0 and v_0 are the coordinates of the principal point (in pixels), f is the focal length (in meters), k_u and k_v are the magnifications respectively in the $\overrightarrow{u}$ and $\overrightarrow{v}$ direction (in pixels/meters), and θ is the angle between these axes. On the other hand, the point $\mathbf{p}^*$ observed in the image $\mathcal{I}$, taken at the reference position $\mathcal{F}^*$, depends on maybe different camera parameters:

$$\mathbf{p}^*(\boldsymbol{\xi}^*, \mathbf{K}^*) = \mathbf{K}^*\mathbf{m}^*(\boldsymbol{\xi}^*) \qquad \mathbf{p}^* \in \mathcal{I}(\boldsymbol{\xi}^*, \mathbf{K}^*) \tag{6}$$

where $\mathbf{K}^*$ has the same form of $\mathbf{K}$ given in equation (5). The image $\mathcal{I}(\boldsymbol{\xi}^*, \mathbf{K}^*)$ will be called the reference image since it is taken at the reference position $\mathcal{F}^*$. The objective of vision-based control is to drive a camera, mounted on the end-effector of a robot, to the reference position using the information provided by the image $\mathcal{I}(\boldsymbol{\xi}, \mathbf{K})$ currently observed. It is clear that both images depend on the intrinsic parameters of the cameras. Thus, even if $\boldsymbol{\xi} = \boldsymbol{\xi}^*$ the reference and the current image will be different if $\mathbf{K} \neq \mathbf{K}^*$.

C. Invariance to camera parameters

As already mentioned in the introduction, most visual servoing techniques are generally based on the hypothesis that the camera frame $\mathcal{F}$ will coincide to the reference frame $\mathcal{F}^*$ if $\mathbf{p}_k = \mathbf{p}_k^*$, $\forall\ k \in \{1, 2, ..., n\}$ (supposing that a sufficient number of corresponding points are observed in the images). However, this hypothesis is valid *if and only if* $\mathbf{K} = \mathbf{K}^*$ at the convergence. In order to control the robot regardless to the camera used during the visual servoing it is necessary to build an error function which does not depend on the camera intrinsic parameters. That is possible by using the simple projective transformation proposed in [12]. Suppose that n non coplanar points $\boldsymbol{\mathcal{X}}_k$ $(k \in \{1, 2, ..., n\})$ are visible on the observed object and consider three non-collinear 3D points $\boldsymbol{\mathcal{X}}_1$, $\boldsymbol{\mathcal{X}}_2$ and $\boldsymbol{\mathcal{X}}_3$. These three points project to the points $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3$ in the current frame and to the points $\mathbf{m}_1^*, \mathbf{m}_2^*, \mathbf{m}_3^*$ in the reference frame. The corresponding image points in pixel coordinates $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ and $\mathbf{p}_1^*, \mathbf{p}_2^*, \mathbf{p}_3^*$ are obtained using equation (4) and equation (6) as follows:

$$\mathbf{Q}(\boldsymbol{\xi}, \mathbf{K}) = \mathbf{K}\, \mathbf{M}(\boldsymbol{\xi}) \tag{7}$$

$$\mathbf{Q}^*(\boldsymbol{\xi}^*, \mathbf{K}^*) = \mathbf{K}^*\mathbf{M}^*(\boldsymbol{\xi}^*) \tag{8}$$

where $\mathbf{Q} = \left[\begin{array}{ccc} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \end{array} \right]$, $\mathbf{M} = \left[\begin{array}{ccc} \mathbf{m}_1 & \mathbf{m}_2 & \mathbf{m}_3 \end{array} \right]$, $\mathbf{Q}^* = \left[\begin{array}{ccc} \mathbf{p}_1^* & \mathbf{p}_2^* & \mathbf{p}_3^* \end{array} \right]$ and $\mathbf{M}^* = \left[\begin{array}{ccc} \mathbf{m}_1^* & \mathbf{m}_2^* & \mathbf{m}_3^* \end{array} \right]$. The matrices $\mathbf{Q} \in \mathbb{P}^2$ and $\mathbf{Q}^* \in \mathbb{P}^2$ are non-singular (3×3) matrices and thus can be used to define two projective spaces $\mathcal{Q}$ and $\mathcal{Q}^*$ in both the current and reference images. The projective space $\mathcal{Q}$ and $\mathcal{Q}^*$ will be called "invariant" spaces. The transformed points $\mathbf{q}$ and $\mathbf{q}^*$ are:

$$\begin{array}{rcl} \mathbf{q}(\boldsymbol{\xi}) & = & \mathbf{Q}^{-1}\mathbf{p} = \mathbf{M}^{-1}\mathbf{K}^{-1}\mathbf{K}\mathbf{m} = \mathbf{M}^{-1}\mathbf{m} \\ \mathbf{q}^*(\boldsymbol{\xi}^*) & = & \mathbf{Q}^{*-1}\mathbf{p}^* = \mathbf{M}^{*-1}\mathbf{K}^{*-1}\mathbf{K}^*\mathbf{m}^* = \mathbf{M}^{*-1}\mathbf{m}^* \end{array}$$

Both $\mathbf{q}$ and $\mathbf{q}^*$ do not depend on the internal parameter of the cameras but they only depend on the position of the camera with respect to the observed object and on its three-dimensional structure. From now on we will refer to the transformed points $\mathbf{q} \in \mathcal{Q}(\boldsymbol{\xi})$ and $\mathbf{q}^* \in \mathcal{Q}(\boldsymbol{\xi}^*)$ as two "invariant" points where the invariance is related to the affine matrix $\mathbf{K}$. Affine invariants have already been used in computer vision mostly for object recognition [17]. Thus, the method proposed in the paper is closely related to affine invariant based vision. However, in our case $\mathbf{q}$ and $\mathbf{q}^*$ are not only invariant to camera intrinsic parameters. The invariance holds up to a collineation of the form:

$$\mathbf{G} = \left[\begin{array}{ccc} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ 0 & 0 & 1 \end{array} \right] \tag{9}$$

The matrix $\mathbf{Q}' = \mathbf{G}\mathbf{Q}$ is also a possible change of coordinates in the projective space. Indeed, if $\mathbf{Q}'$ is a possible change of coordinates then its last row must be $(1, 1, 1)$ ($\mathbf{Q}'$ must be build from three image points). Any matrix $\mathbf{G}$ with the form of given in equation (9) transforms $\mathbf{Q}$ into a different change of coordinates $\mathbf{Q}'$. Thus, if $\mathbf{p}' = \mathbf{G}\mathbf{p}$ is a new image point then $\mathbf{q}'$ is invariant to $\mathbf{G}$:

$$\mathbf{q}' = \mathbf{Q}'^{-1}\mathbf{p}' = \mathbf{Q}^{-1}\mathbf{G}^{-1}\mathbf{G}\mathbf{p} = \mathbf{q}$$

As a consequence, it will not be possible using only the error $(\mathbf{q} - \mathbf{q}^*)$ to position the camera with respect to a planar target. Indeed, starting from a general collineation between the two images, we will obtain $\mathbf{q} = \mathbf{q}^*$ up to $\mathbf{G}$. If the object is non planar, the only possible collineation is the rotation around the $\overrightarrow{z}$ axis:

$$\mathbf{G} = \mathbf{R}_z(\theta) = \left[\begin{array}{ccc} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{array} \right]$$

Thus, the transformed coordinates $\mathbf{q}$ and $\mathbf{q}^*$ are invariant to a rotation r_z around the $\overrightarrow{z}$ axis. Thus, if $\mathbf{q} = \mathbf{q}^* \; \forall \mathbf{q}, \mathbf{q}^*$ then $t_x = t_y = t_z = r_x = r_y = 0$ but r_z can be any. It turns out that it is not a drawback at all. Indeed, the rotation r_z is decoupled and can be controlled separately, as explained in Section IV.

III. Path planning in the invariant space

The epipolar geometry in the invariant space is equivalent to a plane + parallax factorization [12]. The three points chosen for the projective transformation define a virtual plane attached to the object. Let $\mathbf{n}^*$ be the normal to the plane in the absolute frame $\mathcal{F}^*$ and let d^* be the distance of $\mathcal{C}^*$ from the plane. Let $\mu_k^* = \frac{1}{\zeta_k^*} - \frac{1}{d^*}\mathbf{n}^{*T}\mathbf{m}_k^*$ be a constant scalar such that if $\mu_k^* = 0$ then the 3D point $\boldsymbol{\mathcal{X}}_k$, projecting to $\mathbf{q}_k$ and $\mathbf{q}_k^*$, lies on the plane defined by the points $\boldsymbol{\mathcal{X}}_1$, $\boldsymbol{\mathcal{X}}_2$ $\boldsymbol{\mathcal{X}}_3$. The fundamental equation linking $\mathbf{q}_k$ and $\mathbf{q}_k^*$ can be obtained from equation (3) [12]:

$$\gamma_k \; \mathbf{q}_k = \boldsymbol{\Gamma}(\mathbf{q}_k^* + \mu_k^*\mathbf{e}^*) \tag{10}$$

where $\gamma_k = \frac{\zeta_k}{\zeta_k^*}$, $\boldsymbol{\Gamma} = \mathrm{diag}(\gamma_1, \gamma_2, \gamma_3) = \mathrm{diag}(\frac{\zeta_1}{\zeta_1^*}, \frac{\zeta_2}{\zeta_2^*}, \frac{\zeta_3}{\zeta_3^*})$ and $\mathbf{e}^*$ is the epipole in the reference invariant space. From the current and reference image, one can measure $\boldsymbol{\Gamma}$, $\mathbf{e}^*$, γ_k and μ_k^* $(k \in \{4, 5, ..., n\})$ up to a scale factor. In order to eliminate the scale factor, the equation (10) can be divided by the last entry of the diagonal matrix $\boldsymbol{\Gamma}$ $(\gamma_3 > 0)$ without loss of generality. After estimating all the parameters, one can build a function $\mathbf{q}_k^*(t)$ $(t \in [0; T])$ such that $\mathbf{q}_k^*(0) = \mathbf{q}_k$, $\mathbf{q}_k^*(T) = \mathbf{q}_k^*$ and such that the camera approximatively follows a straight line in the Cartesian space (even if the camera internal parameters are completely unknown). Indeed, if the camera follows the line $(\mathcal{C}, \mathcal{C}^*)$, then the epipole in the reference image is the same at each iteration. Indeed, by definition [16], the epipole in the reference image is the intersection of the line $(\mathcal{C}, \mathcal{C}^*)$ with the reference image plane. The epipole in the reference image is thus set to $\mathbf{e}^*/\|\mathbf{e}^*\|$ and determine the direction of translation (i.e. 2 d.o.f.). The norm of the translation can be fixed by a function $\beta_k(t)$ such that $\beta_k(0) = \frac{\mu_k^*}{\gamma_3}\|\mathbf{e}^*\|$ ($\beta_k(0)$ can be measured from equation (10)) and $\beta_k(T) = 0$:

$$\beta_k(t) = \beta_k(0) + \frac{t}{T}(\beta_k(T) - \beta_k(0)) = \frac{\mu_k^*}{\gamma_3}\|\mathbf{e}^*\| \left(1 - \frac{t}{T}\right)$$

The rotational d.o.f. r_x and r_y can be fixed with a diagonal matrix $\Delta(t)$ such that $\Delta(0) = \mathrm{diag}(\frac{\gamma_1}{\gamma_3}, \frac{\gamma_2}{\gamma_3}, 1)$ and $\Delta(T) = \mathrm{diag}(1, 1, 1)$:

$$\begin{array}{rcl} \Delta(t) & = & \Delta(0) + \frac{t}{T}(\Delta(T) - \Delta(0)) \\ & = & \left[\begin{array}{ccc} \frac{\gamma_1}{\gamma_3} + \frac{t}{T}(1 - \frac{\gamma_1}{\gamma_3}) & 0 & 0 \\ 0 & \frac{\gamma_2}{\gamma_3} + \frac{t}{T}(1 - \frac{\gamma_2}{\gamma_3}) & 0 \\ 0 & 0 & 1 \end{array} \right] \end{array}$$

Since by definition $[1\ 1\ 1]\,\mathbf{q}_k(t) = 1$, $\forall t$, the functions $\mathbf{q}_k^*(t)$ $(k \in \{4, 5, ..., n\})$ can be obtained as follows:

$$\mathbf{q}_k^*(t) = \frac{\Delta(t)(\mathbf{q}_k^*(T) + \beta_k(t)\mathbf{e}^*/\|\mathbf{e}^*\|)}{[1\ 1\ 1]\,\Delta(t)(\mathbf{q}^*(T) + \beta(t)\mathbf{e}^*/\|\mathbf{e}^*\|)} \tag{11}$$

It is easy to verify that $\mathbf{q}_k^*(0) = \mathbf{q}_k(0)$ and $\mathbf{q}_k^*(T) = \mathbf{q}_k^*$. Note again that if the current invariant points are such that $\mathbf{q}_k(t) = \mathbf{q}_k^*(t)$ $\forall t$ then the epipole in the reference image is constant and the camera exactly follows a straight line in the 3D space. The rotational d.o.f. r_z has not been considered here since it does not have any influence on the path in the invariant space.

IV. Control in the invariant space

As for standard visual servoing schemes, the control input of the proposed invariant approach is the velocity of the camera. On the other hand, the control of the camera in the invariant space is divided into two different parts since the points $\mathbf{q}_k$ are invariant on the rotation around the $\overrightarrow{z}$ axis and they can only be used to control five d.o.f. of the camera. Another information must be extracted from the images in order to control the last camera d.o.f.

A. Control of five d.o.f. of the camera

Suppose that n matched points are available in both images. Since three points are used to define the projective transformations, only the remaining $n-3$ points can be used to control five d.o.f. of the camera $(t_x, t_y, t_z, r_x, r_y)$. Let $\mathbf{s}(\boldsymbol{\xi}) = (\mathbf{q}_4, \mathbf{q}_5, ..., \mathbf{q}_n)$ be the $(3(n-3)\times 1)$ vector containing the current invariant points. The time derivative of the vector is:

$$\dot{\mathbf{s}} = \mathbf{J}(\boldsymbol{\xi})\ \boldsymbol{\eta} \tag{12}$$

where $\boldsymbol{\eta} = (\nu_x, \nu_y, \nu_z, \omega_x, \omega_y)$ is a vector containing the translation velocity of the camera but only 2 rotation velocities and $\mathbf{J}(\boldsymbol{\xi})$ is the interaction matrix which can be built from the interaction matrix $\mathbf{J}_k(\boldsymbol{\xi})$ relative to each invariant point: $\dot{\mathbf{q}}_k = \mathbf{J}_k(\boldsymbol{\xi})\boldsymbol{\eta}$. Each point $\mathbf{q}_k$ gives a (3×5) interaction matrix $\mathbf{J}_k$ with only 2 independent rows since $q_{1k} + q_{2k} + q_{3k} = 1$ and thus $\dot{q}_{1k} + \dot{q}_{2k} + \dot{q}_{3k} = 0$. Consequently, in order to control five d.o.f. of the camera (except for r_z) we need at least $n = 6$ points. Differentiating $\mathbf{q}_k$, we have:

$$\dot{\mathbf{q}}_k = \frac{d\mathbf{Q}^{-1}}{dt}\mathbf{p}_k + \mathbf{Q}^{-1}\dot{\mathbf{p}}_k = \frac{d\mathbf{Q}^{-1}}{dt}\mathbf{Q}\mathbf{q}_k + \mathbf{Q}^{-1}\dot{\mathbf{p}}_k \tag{13}$$

Note that:

$$\frac{d\mathbf{Q}^{-1}}{dt}\mathbf{Q} = -\mathbf{Q}^{-1}\frac{d\mathbf{Q}}{dt}$$

Plugging this equation in equation (13) we obtain:

$$\dot{\mathbf{q}}_k = \mathbf{Q}^{-1}(\dot{\mathbf{p}}_k - \dot{\mathbf{Q}}\mathbf{q}_k) = \mathbf{Q}^{-1}(\dot{\mathbf{p}}_k - q_{1k}\dot{\mathbf{p}}_1 - q_{2k}\dot{\mathbf{p}}_2 - q_{3k}\dot{\mathbf{p}}_3)$$

The derivative of the current image point is:

$$\dot{\mathbf{p}}_k = \mathbf{L}_k(\zeta_k, \mathbf{K}, \mathbf{p}_k)\begin{bmatrix} \boldsymbol{\eta} \\ \omega_z \end{bmatrix}$$

where $\mathbf{L}_k$ is the (3×6) interaction matrix linking the derivative of vector $\mathbf{p}_k$ to the camera velocity $\mathbf{v} = (\boldsymbol{\eta}, \omega_z)$. Matrix $\mathbf{L}_k$ can be obtained from the standard (2×6) interaction matrix of the image point coordinates [3] adding a third row of zeros. Finally, we obtain:

$$\dot{\mathbf{q}}_k = \mathbf{Q}^{-1}(\mathbf{L}_k - q_{1k}\mathbf{L}_1 - q_{2k}\mathbf{L}_2 - q_{3k}\mathbf{L}_3)\begin{bmatrix} \boldsymbol{\eta} \\ \omega_z \end{bmatrix}$$

Since $\mathbf{q}_k$ is invariant on r_z the last column of the interaction matrix is null and $\mathbf{J}_k$ is obtained from the first five columns of matrix $\mathbf{Q}^{-1}(\mathbf{L}_k - q_{1k}\mathbf{L}_1 - q_{2k}\mathbf{L}_2 - q_{3k}\mathbf{L}_3)$. It must be emphasized that $\mathbf{J}_k$ can be written as follows:

$$\mathbf{J}_k = \begin{bmatrix} \boldsymbol{\Phi}_k(\zeta_1, \zeta_2, \zeta_3, \zeta_k) & \boldsymbol{\Psi}_k \end{bmatrix}\begin{bmatrix} \mathbf{K} & 0 \\ 0 & \mathbf{F}(\mathbf{K}) \end{bmatrix}$$

where:

$$\boldsymbol{\Phi}_k = \mathbf{Q}^{-1}\begin{bmatrix} \sum_{i=1}^{3}\frac{q_{ik}}{\zeta_i} - \frac{q_{ik}}{\zeta_k} & 0 & \frac{u_k}{\zeta_k} - \sum_{i=1}^{3}\frac{u_i q_{ik}}{\zeta_i} \\ 0 & \sum_{i=1}^{3}\frac{q_{ik}}{\zeta_i} - \frac{q_{ik}}{\zeta_k} & \frac{v_k}{\zeta_k} - \sum_{i=1}^{3}\frac{v_i q_{ik}}{\zeta_i} \\ 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{\Psi}_k = \mathbf{Q}^{-1}\begin{bmatrix} u_k v_k - \sum_{i=1}^{3} u_i v_i q_{ik} & \sum_{i=1}^{3} u_i^2 q_{ik} - u_k^2 \\ v_k^2 - \sum_{i=1}^{3} v_i^2 q_{ik} & \sum_{i=1}^{3} u_i v_i q_{ik} - u_k v_k \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{F}(\mathbf{K}) = \begin{bmatrix} \frac{\sin(\theta)}{fk_v} & \frac{-\cos(\theta)}{fk_v} \\ 0 & \frac{1}{fk_u} \end{bmatrix}$$

Thus, the interaction matrix can be decomposed as follows:

$$\mathbf{J}(\boldsymbol{\xi}, \boldsymbol{\zeta}, \mathbf{K}) = \begin{bmatrix} \boldsymbol{\Phi}(\boldsymbol{\zeta}) & \boldsymbol{\Psi} \end{bmatrix}\begin{bmatrix} \mathbf{K} & 0 \\ 0 & \mathbf{F}(\mathbf{K}) \end{bmatrix}$$

This matrix depends on the depths distribution $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, ..., \zeta_n)$ and on the camera intrinsic parameters $\mathbf{K}$. Since the structure of the object is rigid, one can estimate the depth distribution up to a scalar factor using the epipolar geometry $\widehat{\boldsymbol{\zeta}} = \kappa(t)\boldsymbol{\zeta}$ (where $\kappa(t) > 0$ is a positive scalar $\forall t$), as explained in the Appendix A. Note that the camera parameters can eventually vary during the servoing when using a zooming camera: $\mathbf{K} = \mathbf{K}(t)$. The stability analysis will show that it is not necessary to use a self-calibration algorithm to estimate the camera parameters but only a rough approximation $\widehat{\mathbf{K}}$ can be used to compute the interaction matrix. Since $\boldsymbol{\Phi}(\widehat{\boldsymbol{\zeta}}) = \frac{1}{\kappa}\boldsymbol{\Phi}(\boldsymbol{\zeta})$, the estimated interaction matrix can be written as:

$$\widehat{\mathbf{J}} = \begin{bmatrix} \frac{1}{\kappa}\boldsymbol{\Phi}(\boldsymbol{\zeta}) & \boldsymbol{\Psi} \end{bmatrix}\begin{bmatrix} \widehat{\mathbf{K}} & 0 \\ 0 & \mathbf{F}(\widehat{\mathbf{K}}) \end{bmatrix} \tag{14}$$

Let us suppose that the interaction matrix $\mathbf{J}$ is full rank, thus the estimated matrix $\widehat{\mathbf{J}}$ is also full rank. In order to control the five d.o.f. of the camera I use the task function approach [18]. Consider the following (5×1) task function:

$$\boldsymbol{\varepsilon}(\boldsymbol{\xi}) = \widehat{\mathbf{J}}^{+}(\boldsymbol{\xi})(\mathbf{s}(\boldsymbol{\xi}) - \mathbf{s}^*(t)) \tag{15}$$

where $\mathbf{s}^*(t) = (\mathbf{q}_4^*(t), \mathbf{q}_5^*(t), ..., \mathbf{q}_n^*(t))$ and $\mathbf{q}_k^*(t)$ is given by equation (11). Obviously, if $\mathbf{s} = \mathbf{s}^*$ then $\boldsymbol{\varepsilon} = 0$. On the other hand, sufficient conditions are given in the Appendix B such that if $\|\mathbf{s} - \mathbf{s}^*\|$ is sufficiently small then $\boldsymbol{\varepsilon} = 0$ only if $\mathbf{s} = \mathbf{s}^*$ (i.e. $\mathbf{s} - \mathbf{s}^* \neq 0$ never belongs to $\mathrm{Ker}(\widehat{\mathbf{J}}^{+})$). Differentiating equation (15) we obtain:

$$\dot{\boldsymbol{\varepsilon}} = \frac{d\widehat{\mathbf{J}}^{+}}{dt}(\mathbf{s} - \mathbf{s}^*(t)) + \widehat{\mathbf{J}}^{+}\dot{\mathbf{s}} - \widehat{\mathbf{J}}^{+}\frac{\partial \mathbf{s}^*(t)}{\partial t} \tag{16}$$

Using equation (12) and since the vector $\frac{d\widehat{\mathbf{J}}^{+}}{dt}(\mathbf{s} - \mathbf{s}^*(t))$ can be written as $\frac{d\widehat{\mathbf{J}}^{+}}{dt}(\mathbf{s} - \mathbf{s}^*(t)) = \mathbf{O}(\mathbf{s} - \mathbf{s}^*(t))\boldsymbol{\eta}$ (where $\mathbf{O}(\mathbf{s} - \mathbf{s}^*(t)) \to 0$ if $\mathbf{s} \to \mathbf{s}^*(t)$), we obtain:

$$\dot{\boldsymbol{\varepsilon}} = \left(\mathbf{O}(\mathbf{s} - \mathbf{s}^*(t)) + \widehat{\mathbf{J}}^{+}\mathbf{J}\right)\boldsymbol{\eta} - \widehat{\mathbf{J}}^{+}\frac{\partial \mathbf{s}^*(t)}{\partial t} \tag{17}$$

Consider the control law sent to a low-level controller:

$$\boldsymbol{\eta} = -\lambda\boldsymbol{\varepsilon} + \widehat{\mathbf{J}}^{+}\frac{\partial \mathbf{s}^{*}(t)}{\partial t} \tag{18}$$

where λ is a positive gain. From equations (18) and (17) we obtain the following closed-loop equation:

$$\begin{aligned}\dot{\boldsymbol{\varepsilon}} &= -\lambda\left(\mathbf{O}(\mathbf{s}-\mathbf{s}^{*}(t)) + \widehat{\mathbf{J}}^{+}\mathbf{J}\right)\boldsymbol{\varepsilon} \\ &+ \left(\mathbf{O}(\mathbf{s}-\mathbf{s}^{*}(t)) + \widehat{\mathbf{J}}^{+}\mathbf{J} - \mathbf{I}\right)\widehat{\mathbf{J}}^{+}\frac{\partial \mathbf{s}^{*}(t)}{\partial t}\end{aligned} \tag{19}$$

In order to study the behavior of the task function during the path tacking, consider the system (19) linearized for $\mathbf{s} = \mathbf{s}^{*}(t)$:

$$\dot{\boldsymbol{\varepsilon}} = -\lambda\ \mathbf{A}(t)\ \boldsymbol{\varepsilon} + \mathbf{b}(t) \tag{20}$$

where:

$$\begin{aligned}\mathbf{A}(t) &= \widehat{\mathbf{J}}^{+}\mathbf{J}\Big|_{\mathbf{s}=\mathbf{s}^{*}(t)} = \begin{bmatrix} \kappa(t)\widehat{\mathbf{K}}^{-1}\mathbf{K}(t) & 0 \\ 0 & \widehat{\mathbf{F}}^{-1}\mathbf{F}(t) \end{bmatrix} \\ \mathbf{b}(t) &= (\widehat{\mathbf{J}}^{+}\mathbf{J} - \mathbf{I})\ \widehat{\mathbf{J}}^{+}\Big|_{\mathbf{s}=\mathbf{s}^{*}(t)}\frac{\partial \mathbf{s}^{*}(t)}{\partial t}\end{aligned}$$

Let us consider first the case when the starting point $\boldsymbol{\varepsilon}(0)$ is in a neighborhood of the equilibrium point $\boldsymbol{\varepsilon} = 0$ and suppose that $\partial\mathbf{s}^{*}(t)/\partial t = 0$ and that the camera parameters do not vary during the servoing $\partial\mathbf{K}(t)/\partial t = 0$.

Proposition 1: *The equilibrium point $\boldsymbol{\varepsilon} = 0$, of the time varying linear system (20) when $\mathbf{b}(t) = 0$ and $\partial\mathbf{K}(t)/\partial t = 0$, is locally asymptotically stable if and only if $\kappa(t) > 0$, $\widehat{f} > 0$, $\widehat{k}_u > 0$, $\widehat{k}_v > 0$ and $0 < \widehat{\theta} < \pi$.*

The proof is detailed in the Appendix B. This proposition means that the control law is extremely robust to calibration errors since any positive approximation of the camera parameters is sufficient to stabilize the system. Obviously, this is true only in a small neighborhood of the equilibrium point. In the case when the camera parameters vary during the servoing, we have:

Proposition 2: *The equilibrium point of the time varying linear system (20) when $\mathbf{b}(t) = 0$, is locally asymptotically stable if $\kappa(t) > 0$, $\widehat{\mathbf{K}}^{-1}\mathbf{K}(t) > 0$ and $\widehat{\mathbf{F}}^{-1}\mathbf{F}(t) > 0$.*

The proof is given in the Appendix B. The sufficient conditions are the same found in [5] and they are very large. This means that one could choose once and for all an estimate $\widehat{\mathbf{K}}$ of the camera parameters and keep it constant even if the real parameters $\mathbf{K}(t)$ are changing during the servoing. Suppose that the camera intrinsic parameters are unknown but that, for each parameter $k_{ij} \in \mathbf{K}$, we have a bound on the estimation error $\underline{\delta k}_{ij} \leq k_{ij} - \widehat{k}_{ij} \leq \overline{\delta k}_{ij}$ (where $\underline{\delta k}_{ij}$ and $\overline{\delta k}_{ij}$ are respectively the lower and the upper bound). Then, the true parameter is bounded by $\underline{\delta k}_{ij} + \widehat{k}_{ij} \leq k_{ij} \leq \overline{\delta k}_{ij} + \widehat{k}_{ij}$. Thus, the sufficient conditions given in proposition 2 can be checked $\forall k_{ij} \in \{\underline{\delta k}_{ij} + \widehat{k}_{ij}, \overline{\delta k}_{ij} + \widehat{k}_{ij}\}$.

Consider now the general case when a path computed from equation (11) is tracked (i.e. $\mathbf{b}(t) \neq 0$) and the camera parameters are eventually varying during the servoing.

Proposition 3: *If $\mathbf{A}(t) > 0$ and $\|\mathbf{b}(t)\| \leq \psi/2$ during the path tracking (with $\boldsymbol{\varepsilon}(0) = 0$ since $\mathbf{s}(0) = \mathbf{s}^{*}(0)$) the norm of the task function can be bounded by:*

$$\|\boldsymbol{\varepsilon}(t)\| \leq \frac{\psi}{\varphi} \tag{21}$$

where $\varphi = \lambda\sigma$ (σ is the lower bound of the minimum singular value of the symmetric positive matrix $\mathbf{A}(t) + \mathbf{A}^{\top}(t)$). If $\|\boldsymbol{\varepsilon}(t)\|$ is bounded, then $\|\mathbf{s}(t) - \mathbf{s}^{*}(t)\|$ is also bounded.

The proof is presented in the Appendix B. In order to reduce the tracking error, one can increase the gain λ (i.e. φ increases) or decrease the velocity $\frac{\partial \mathbf{s}^{*}(t)}{\partial t}$ (i.e. ψ decreases). Obviously, the bound on $\|\boldsymbol{\varepsilon}(t)\|$ depends also on calibration errors. Indeed, if all the parameters of th system are perfectly known, then $\widehat{\mathbf{J}} = \mathbf{J}$ and $\psi = 0$. Thus the tracking is perfect since $\mathbf{b}(t) = 0$. On the other hand, even if the bound on the tracking error increases with calibration errors, the misestimation of $\mathbf{b}(t)$ has a small influence on the stability of the servoing (see the experimental results).

B. Control of the last d.o.f. of the camera

As already mentioned, the remaining d.o.f. of the camera cannot be controlled using $\mathbf{q}$. Therefore, it is necessary to find a parameter depending on the rotation around the $\overrightarrow{z}$ axis. Let $\mathbf{T}$ be the following matrix:

$$\mathbf{T} = \mathbf{Q}\mathbf{Q}^{*-1} = \begin{bmatrix} \tau_{11} & \tau_{12} & \tau_{13} \\ \tau_{21} & \tau_{22} & \tau_{23} \\ 0 & 0 & 1 \end{bmatrix} \tag{22}$$

This matrix can be computed directly from the image coordinates of $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_1^{*}, \mathbf{p}_2^{*}, \mathbf{p}_3^{*}$. The matrix $\mathbf{T}$ must be triangular at the convergence (i.e. when the camera is back at the reference position $\mathcal{F} = \mathcal{F}^{*}$). Indeed, if $\boldsymbol{\xi} = \boldsymbol{\xi}^{*}$ then $\mathbf{M} = \mathbf{M}^{*}$ and from equations (8) and (7) we obtain:

$$\mathbf{T} = \mathbf{Q}\mathbf{Q}^{*-1}|_{\boldsymbol{\xi}=\boldsymbol{\xi}^{*}} = \mathbf{K}(t)\mathbf{K}^{*-1} \tag{23}$$

At the convergence, matrix $\mathbf{T}$ must be upper triangular for any matrices $\mathbf{K}(t)$ and $\mathbf{K}^{*}$ (i.e. $\tau_{21} = 0$) and have positive eigenvalues (i.e. $\tau_{11} > 0$ and $\tau_{22} > 0$). Therefore, one must impose the constraints $\tau_{21} = 0$, $\tau_{11} > 0$ and $\tau_{22} > 0$ in order to have $r_z = 0$. Note that $\tau_{21} = 0$ if $r_z = \pm\frac{\pi}{2}$ or $r_z = \pm\pi$. However, in that cases $\tau_{11} < 0$ and/or $\tau_{22} < 0$. For simplicity, I consider now the case when $-\frac{\pi}{2} < r_z < \frac{\pi}{2}$ but few changes have to be made in order to consider the general case. The remaining d.o.f. is thus controlled by a second scalar task function:

$$\epsilon = \det(\mathbf{Q}^{*})\ \tau_{21} = v_1^{*}(v_3 - v_2) + v_2^{*}(v_1 - v_3) + v_3^{*}(v_2 - v_1)$$

Since $\det(\mathbf{Q}^{*}) \neq 0$, if $\epsilon = 0$ then $\tau_{21} = 0$ and at the convergence $\mathcal{F} = \mathcal{F}^{*}$ (since $\boldsymbol{\varepsilon} \to 0$ thanks to the previous control law). The derivative of ϵ can be written:

$$\dot{\epsilon} = a(\epsilon, t)\omega_z + \mathbf{c}^{\top}(t)\boldsymbol{\eta}(t) \tag{24}$$

where $a(\epsilon,t)$ is a scalar, $\mathbf{c}(t)$ is a (5×1) vector and $\boldsymbol{\eta}(t)$ is the control law (18). Note that if $-\frac{\pi}{2} < r_z < \frac{\pi}{2}$ then $a(\epsilon,t) \neq 0$. Imposing an exponential convergence of ϵ to zero the control law for r_z is:

$$\omega_z = -\frac{\rho\,\epsilon}{\widehat{a}(\epsilon,t)} - \frac{\widehat{\mathbf{c}}^\top \boldsymbol{\eta}(t)}{\widehat{a}(\epsilon,t)} \tag{25}$$

where ρ is a positive scalar tuning the speed of convergence, $\widehat{a}$ and $\widehat{\mathbf{c}}$ are approximations of a and $\mathbf{c}$. The closed-loop equation is the following:

$$\dot{\epsilon} = -\rho\frac{a(\epsilon,t)}{\widehat{a}(\epsilon,t)}\epsilon + \left(\mathbf{c}^\top(t) - \frac{a(\epsilon,t)}{\widehat{a}(\epsilon,t)}\widehat{\mathbf{c}}^\top(t)\right)\boldsymbol{\eta}(t) \tag{26}$$

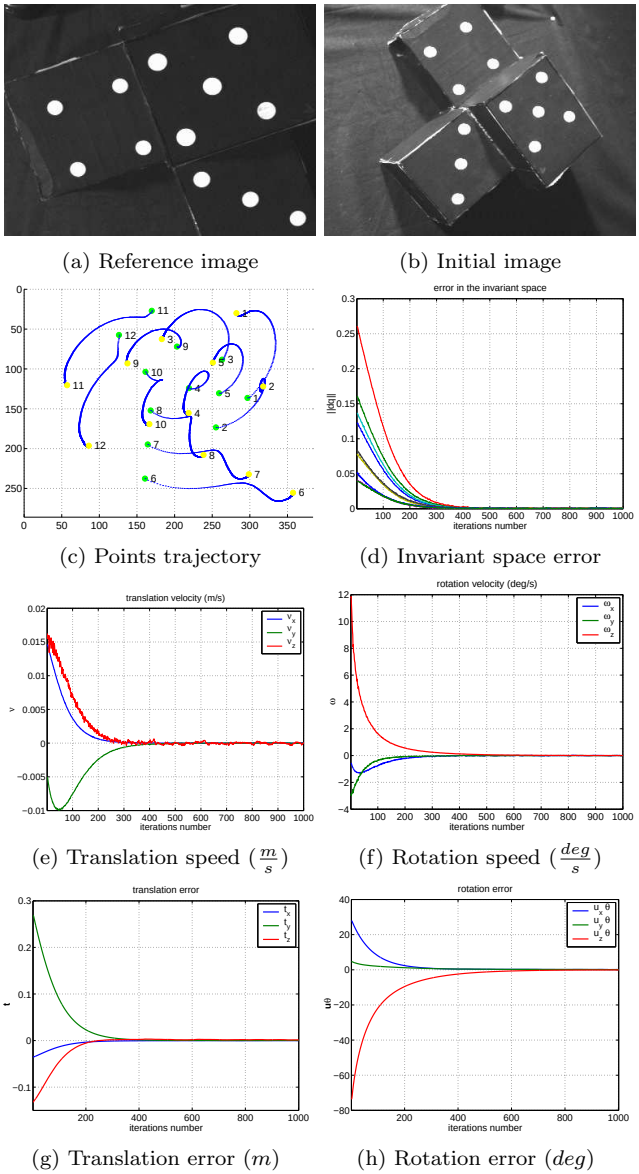
Note that $\boldsymbol{\eta}(t) \to 0$ and $\mathbf{c}(t)$ is bounded. Thus, the control law is stable if and only if $\frac{a(\epsilon,t)}{\widehat{a}(\epsilon,t)} > 0$. In the ideal case when $\widehat{a} = a$ and $\widehat{\mathbf{c}} = \mathbf{c}$ the convergence is exponential. As it is shown by the experiments, the control law is stable even in the presence of calibration errors.

V. Experimental Results

The visual servoing scheme proposed in the paper has been tested on a six d.o.f. Cartesian robot AFMA (at IRISA/INRIA Rennes). The robot is very well calibrated and it provides a ground truth in order to measure the positioning precision of visual servoing. A more detailed description of the experiments can be found in [13], [12].

A. Visual Servoing without path planning

The two following experiments show that the proposed control scheme can work even if the initial error $\mathbf{q}-\mathbf{q}^*$ is large and it is not constrained to follow a path in the invariant space. The reference is fixed to $\mathbf{q}_k^*(t) = \mathbf{q}_k^*\ \forall t, \forall k$, thus $\partial \mathbf{s}^*(t)/\partial t = 0$. For both experiments, a camera (with a 12mm lens) is initially mounted on the robot end-effector and observe 12 points. An approximation $\widehat{\mathbf{K}}$ of the current camera parameters is used in the control law. In order to show that the visual servoing technique is robust with respect to errors on the estimation of the depth of the observed points, the current depth $\boldsymbol{\zeta}$ is fixed to the depth $\boldsymbol{\zeta}^*$ measured in the reference frame (this corresponds to the image-based visual servoing setup proposed in [3]). In the first experiment, the 12mm lens used for learning the reference image (see Figure 2(a)) is the same lens used for servoing. The initial displacement of the robot (Figure 2(b) shows the initial image of the object) is quite large: 300 *mm* for the translation and 80 *degrees* for the rotation. The control laws computed from equations (18) and (25) are plotted in Figures 2(e) and (f). They are stable despite the large initial displacement. Thus, the error in the invariant space (see Figures 2(d)) converges to zero. Consequently, the position error of the robot end-effector converges to zero (see Figures 2(g) and (h)). At the convergence, the positioning error is less than 1 *mm* for the translation and 0.1 *degrees* for the rotation. Figure 2(c) plots the trajectory of the points in the image and shows that the points go from their initial position to their reference position. This experiment proves that the new scheme can perform positioning tasks exactly as standard visual servoing techniques.



(a) Reference image (b) Initial image

(c) Points trajectory (d) Invariant space error

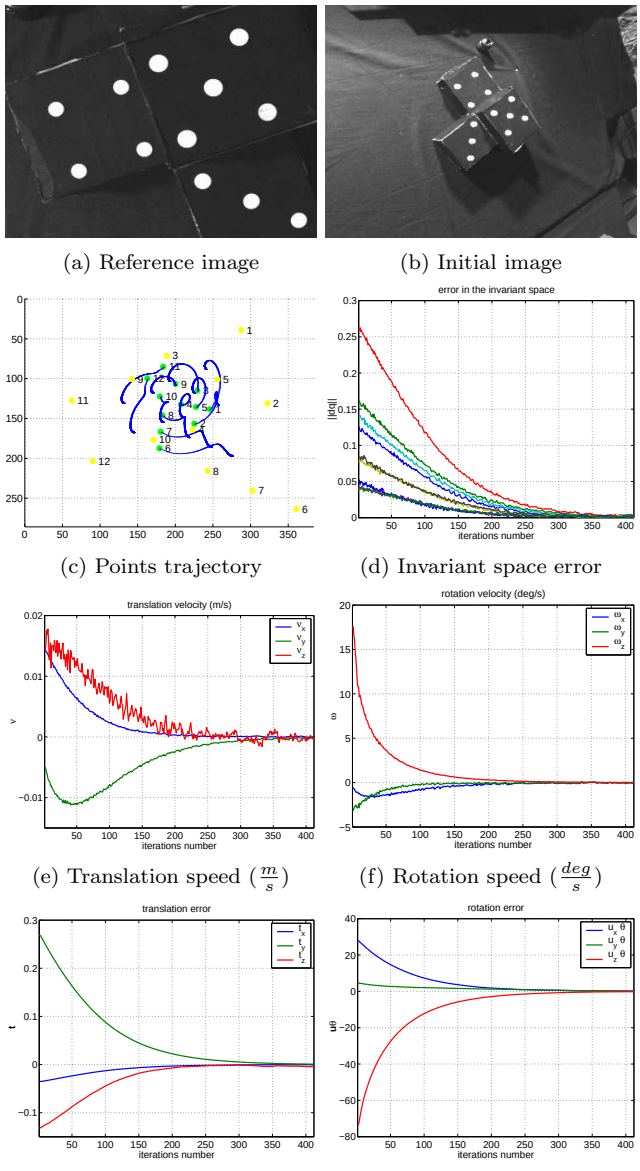
(e) Translation speed ($\frac{m}{s}$) (f) Rotation speed ($\frac{deg}{s}$)

(g) Translation error (m) (h) Rotation error (deg)

Fig. 2. Experiment using a 12mm lens for learning and servoing.

The proof that the camera-independent vision-based control increases the versatility of visual servoing is provided by the second experiment (Figure 3). In this experiment, the reference image is identical to the previous one (see Figure 2(a) and Figure 3(a)). The robot starts from the same initial position (the initial displacement of the robot is the same in Figures 2(g) and (h) and Figures 3(g) and (h)) but now the lens of the camera is changed. A 6mm lens is used for servoing, thus the initial image in Figure 3(b) is completely different from the image in Figure 2(b). On the other hand, the curves plotted in Figure 2(d) and Figure 3(d) have identical initial values since the error in the invariant space does not depend on the camera intrinsic parameters and the initial and reference positions are the same in both experiments. Figure 3(c) shows the trajectory of the points in the image. The final

image is obviously different from the reference image (the yellow points). However, Figures 3(g) and (h) show that the error on the position of the end-effector converges again to zero. Figures 3(e) and (f) plot the control law sent to the robot. The control law is robust to noise and ensures again the convergence of the robot to the reference position.



(a) Reference image (b) Initial image

(c) Points trajectory (d) Invariant space error

(e) Translation speed ($\frac{m}{s}$) (f) Rotation speed ($\frac{deg}{s}$)

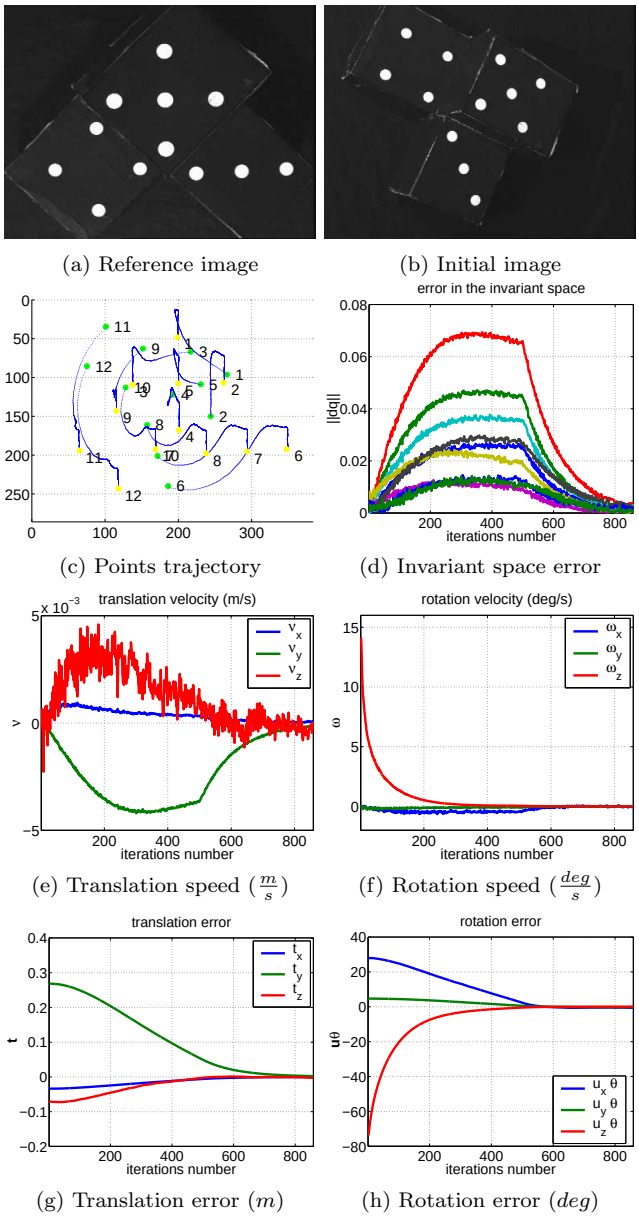
(g) Translation error (m) (h) Rotation error (deg)

Fig. 3. Experiment using a 12mm lens for learning and a 6mm lens for servoing. The intrinsics-free visual servoing approach does not depend on the lens used for servoing.

When compared to the previous experiment, the higher level of noise (especially on ν_z, as is it shown by Figure 3(e)) is explained by the fact that the size of the object is now small in the image due to the change of the lens. Indeed, the extraction of the centroids of the white spots is less precise. Furthermore, the main failure mode for the method occurs when the observed object is planar. The points of the object used in the experiments are close to a planar configuration. Consequently, the precision obtained in this experiment is around 2 mm for the translation and 0.2 degrees for the rotation.

B. Path planning in the invariant space

The local stability analysis presented in the paper is valid if the error $\mathbf{q} - \mathbf{q}^*$ is sufficiently small at each iteration of the control law. As explained in Section III, one can build a path $\mathbf{s}^*(t)$ in the invariant space such that the camera follows a straight line in the Cartesian space. Simple sufficient conditions on calibration errors have been obtained such that the tracking error is bounded. In the following experiments, I show that the sufficient conditions for the local stability found in the theoretical analysis are very large. Indeed, a very bad approximation $\widehat{\mathbf{K}}$ of the current camera parameters is used in the control law (50% error on each parameter). After the reference image (Figure 4(a)) has been learned the robot is displaced to its initial position (Figure 4(b) shows the initial image of the object).



(a) Reference image (b) Initial image

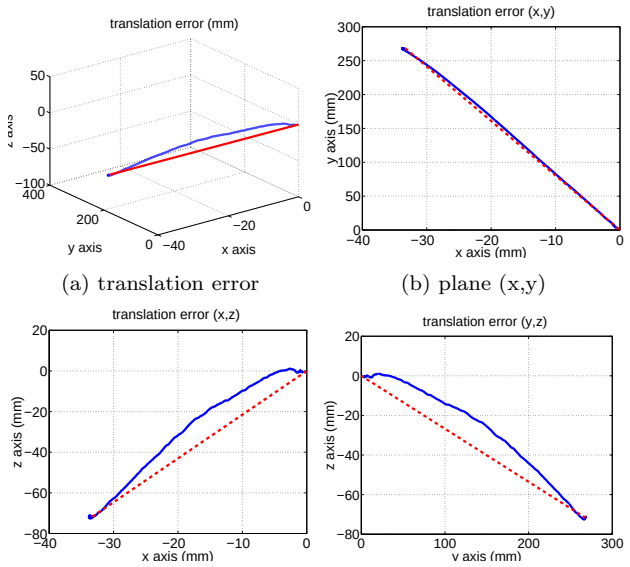
(c) Points trajectory (d) Invariant space error

(e) Translation speed ($\frac{m}{s}$) (f) Rotation speed ($\frac{deg}{s}$)

(g) Translation error (m) (h) Rotation error (deg)

Fig. 4. Path tracking in the invariant space. The error in the invariant space stay small. The invariant vision-based control is stable and, at the convergence, the camera is back to the reference position.

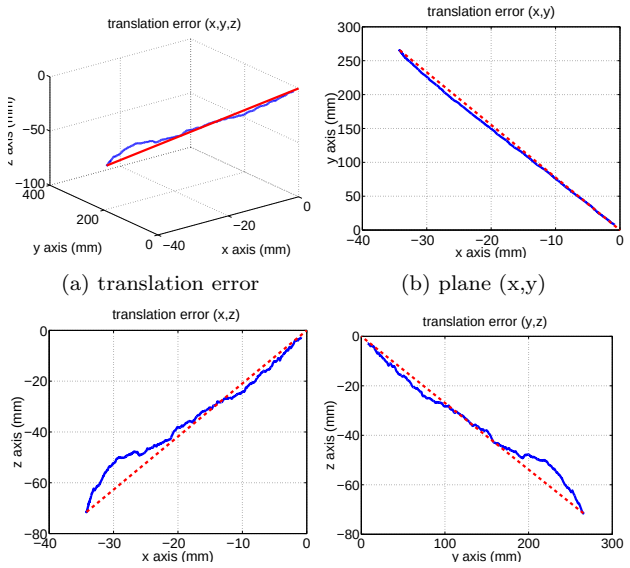
The initial displacement of the robot is approximately the same of previous experiments. As already mentioned the bad estimation of $\mathbf{b}(t)$ has no influence on the stability of the servoing as far as $\mathbf{b}(t)$ is bounded. To show that, I set $\widehat{\mathbf{J}}^{+}\partial\mathbf{s}^{*}(t)/\partial t = 0$ in the control law used in the experiments even if in reality $\partial\mathbf{s}^{*}(t)/\partial t \neq 0$. In the first experiment, the path is sampled $N = 500$ times. The first part of the control law corresponding to the translational velocity $\boldsymbol{\nu}$ is plotted in Figure 4(e) while the second part corresponding to the rotational velocity $\boldsymbol{\omega}$ is plotted in Figure 4(f). The error in the invariant space is bounded and converges to zero as soon as the path tracking ends (see Figures 4(d)). Consequently, the position error of the end-effector converges to zero (see Figures 4(g) and (h)). Figure 4(c) plots the trajectory of the points in the image.



(a) translation error (b) plane (x,y)

(c) plane (x,z) (d) plane (y,z)

Fig. 5. The path is sampled with N=500. The camera approximately follows the dashed straight line.



(a) translation error (b) plane (x,y)

(c) plane (x,z) (d) plane (y,z)

Fig. 6. The path is sampled with N=3000. The tracking error is reduced and the camera follows more accurately the straight line.

Although the camera is controlled in the invariant space, the behavior of the robot in the Cartesian space is very nice. Figure (5) shows that the camera approximately follows a straight line despite the large calibration errors. As expected, the tracking error is bounded. Since the control law ν_z is very noisy, the error is mainly on t_z as is it shown by Figures (5)(c) and (d). In the second experiment, the robot starts approximately from the same initial position. The results of the visual servoing are therefore similar to the results shown in Figure 4. On the other hand, the path is sampled $N = 3000$ times. Thus, the speed of the reference trajectory is reduce by a factor 6. As a consequence, the tracking error presented in Figure 6 is considerably smaller than the tracking error presented in Figure 5.

VI. Conclusion

This work shows how to position a camera, with respect to a non-planar object, even if the intrinsic parameters change. The new visual servoing scheme will be useful especially when a zooming camera is mounted on the end-effector of the robot. The zoom can be used to enlarge the field of view of the camera and to bound the size of the object observed in the image. The stability analysis of the visual servoing invariant to camera intrinsic parameters has shown that the control law proposed in the paper is robust to camera calibration errors. Despite the camera is controlled in a projective space, it is possible to follow a straight line in the Cartesian space. Future work will concern the control of the camera zoom while servoing.

Acknowledgments

I would like to thank Eric Marchand (IRISA/INRIA Rennes) for his help in programming the experiments.

Appendix A

Several approaches have been proposed in order to estimate the depths distribution on-line [19] [20]. These approaches requires some knowledge on the parameters of the system (camera calibration, robot motion,...). A different solution to the problem is to use projective geometry. Consider two images of the same 3D object taken with the same camera having intrinsic parameters $\overline{\mathbf{K}}$. Suppose that the displacement between the two images taken at the positions $\boldsymbol{\xi}^{I}$ and $\boldsymbol{\xi}^{II}$ is an unknown pure translation. A pure translation of the camera frame is very easy to obtain (for example one can mount the camera on a rail). In that case, the equation linking the point $\mathbf{p}_k^{II}$ in the second image to the point $\mathbf{p}_k^{I}$ in the first image is:

$$\frac{\zeta_k^{II}}{\zeta_k^{I}}\mathbf{p}_k^{II} = \mathbf{p}_k^{I} + \frac{\|\overline{\mathbf{e}}\|}{\zeta_k^{I}}\frac{\overline{\mathbf{e}}}{\|\overline{\mathbf{e}}\|} \qquad \forall k \in \{1,2,...,n\} \tag{27}$$

where $\overline{\mathbf{e}} = \overline{\mathbf{K}}\overline{\mathbf{t}}$ is the epipole and $\overline{\mathbf{t}}$ is the translation between frames $\mathcal{F}^{I}$ and $\mathcal{F}^{II}$. If at least 8 points are available (i.e. $n \geq 8$), from these equations one can measure ζ_k^{II}/ζ_k^{I}, $\|\mathbf{e}\|/\zeta_k^{I}$ and $\mathbf{e}/\|\mathbf{e}\|$. Thus, without knowing the translation of the camera $\overline{\mathbf{t}}$ and without knowing its intrinsic parameters $\overline{\mathbf{K}}$, one can measure the depths distribution $\boldsymbol{\zeta}^{I} = (\zeta_1^{I}, \zeta_2^{I}, ..., \zeta_n^{I})$ in frame $\mathcal{F}^{\mathcal{I}}$ up to an unknown

scale factor (i.e. the measured distribution $\widehat{\boldsymbol{\zeta}}^I$ is proportional to the true distribution $\widehat{\boldsymbol{\zeta}}^I = \kappa^I \boldsymbol{\zeta}^I$). Similarly, one can measure the depths distribution $\boldsymbol{\zeta}^{II} = (\zeta_1^{II}, \zeta_2^{II}, ..., \zeta_n^{II})$ in frame $\mathcal{F}^{\mathcal{II}}$ up to a different unknown scale factor (i.e. $\widehat{\boldsymbol{\zeta}}^{II} = \kappa^{II} \boldsymbol{\zeta}^{II}$). This can be done once and for all off-line since the depths distribution depends only on the structure of the object. Suppose now that the camera used to estimate the depths distribution is completely different from the parameters used to learn $\mathbf{K}^* \neq \overline{\mathbf{K}}$ or those used during the servoing $\mathbf{K} \neq \overline{\mathbf{K}}$. Knowing the depths distribution $\widehat{\boldsymbol{\zeta}}^I$ (which corresponds to an affine reconstruction [16]), it is possible to use the epipolar geometry to measure the current depths distribution $\boldsymbol{\zeta}$ and the reference depth distribution $\boldsymbol{\zeta}^*$. For example, between the image taken at frame $\mathcal{F}^{\mathcal{I}}$ and the current image taken at frame $\mathcal{F}$ we have:

$$\frac{\zeta_k^I}{\zeta_k} \mathbf{p}_k^I = \mathbf{G}_\infty \mathbf{p}_k + \frac{\|\mathbf{e}\|}{\zeta_k} \frac{\mathbf{e}}{\|\mathbf{e}\|} \qquad \forall k \in \{1, 2, ..., n\} \tag{28}$$

where $\mathbf{G}_\infty = \overline{\mathbf{K}} \mathbf{R}^I \mathbf{K}^{-1}$ is the collineation of the plane at infinity [16], $\mathbf{e} = \overline{\mathbf{K}} \mathbf{t}^I$ is the epipole, $\mathbf{R}^I$ and $\mathbf{t}^I$ are respectively the rotation and the translation between frames $\mathcal{F}^{\mathcal{I}}$ and $\mathcal{F}$. From equations (28), it is possible to measure the ratio $\frac{\widehat{\zeta}_k^I}{\widehat{\zeta}_k} \propto \frac{\zeta_k^I}{\zeta_k}$. Thus, $\widehat{\zeta}_k \propto \frac{\widehat{\zeta}_k^I}{\zeta_k^I} \zeta_k$ and since $\widehat{\zeta}_k^I = \kappa^I \zeta_k^I$, one can measure all the ζ_k up to the same scale factor: $\widehat{\zeta}_k \propto \kappa^I \zeta_k$, thus $\widehat{\zeta}_k = \kappa \zeta_k$. Note that, the scale factor κ depends on time since it changes with the position of the current camera frame. Note also that, all the points are in front of the camera and all depths ζ_k must be positives, therefore $\kappa(t) > 0 \ \forall t$. Finally, the vector containing the current depth distribution can be written as $\widehat{\boldsymbol{\zeta}} = \kappa(t) \boldsymbol{\zeta}$.

Appendix B

Proof: [**Proposition 1**] The time varying linear system (20) when $\mathbf{b}(t) = 0$ and $\partial \mathbf{K} / \partial t = 0$ is:

$$\dot{\boldsymbol{\varepsilon}} = -\lambda \mathbf{A}(t) \boldsymbol{\varepsilon} = -\lambda \begin{bmatrix} \kappa(t) \widehat{\mathbf{K}}^{-1} \mathbf{K} & 0 \\ 0 & \widehat{\mathbf{F}}^{-1} \mathbf{F} \end{bmatrix} \boldsymbol{\varepsilon}$$

The equilibrium point $\boldsymbol{\varepsilon} = 0$ is locally asymptotically stable if and only if each block on the diagonal of $\mathbf{A}(t)$ has eigenvalues with positive real part. Indeed, if $\widehat{\mathbf{K}}^{-1} \mathbf{K} = \mathbf{C}_1^{-1} \mathbf{D}_1 \mathbf{C}_1$ and $\widehat{\mathbf{F}}^{-1} \mathbf{F} = \mathbf{F}_2^{-1} \mathbf{D}_2 \mathbf{C}_2$ then:

$$\mathbf{A}(t) = \begin{bmatrix} \mathbf{C}_1^{-1} & 0 \\ 0 & \mathbf{C}_2^{-1} \end{bmatrix} \begin{bmatrix} \kappa(t) \mathbf{D}_1 & 0 \\ 0 & \mathbf{D}_2 \end{bmatrix} \begin{bmatrix} \mathbf{C}_1 & 0 \\ 0 & \mathbf{C}_2 \end{bmatrix}$$

The nonsingular constant matrix $\mathbf{C} = \text{diag}(\mathbf{C}_1, \mathbf{C}_2)$ defines a change of coordinates $\mathbf{x} = \mathbf{C} \boldsymbol{\varepsilon}$. In the new coordinate system, the differential equation becomes:

$$\dot{\mathbf{x}} = -\lambda \begin{bmatrix} \kappa(t) \mathbf{D}_1 & 0 \\ 0 & \mathbf{D}_2 \end{bmatrix} \mathbf{x}$$

where $\mathbf{D}_1 = \text{diag}\left(\frac{fk_u}{\widehat{f}\widehat{k}_u}, \frac{fk_v \sin(\widehat{\theta})}{\widehat{f}\widehat{k}_v \sin(\theta)}, 1\right)$ and $\mathbf{D}_2 = \text{diag}\left(\frac{\widehat{f}\widehat{k}_u}{fk_u}, \frac{\widehat{f}\widehat{k}_v \sin(\theta)}{fk_v \sin(\widehat{\theta})}\right)$. The new linear system is fully decoupled and consists of 5 independent differential equations:

$$\dot{x}_i = -\alpha_i \, x_i \qquad i \in \{1, 2, ..., 5\}$$

where $\alpha_1 = \lambda \kappa(t) \frac{fk_u}{\widehat{f}\widehat{k}_u}$, $\alpha_2 = \lambda \kappa(t) \frac{fk_v \sin(\widehat{\theta})}{\widehat{f}\widehat{k}_v \sin(\theta)}$, $\alpha_3 = \lambda \kappa(t)$, $\alpha_4 = \lambda \frac{\widehat{f}\widehat{k}_u}{fk_u}$, $\alpha_5 = \lambda \frac{\widehat{f}\widehat{k}_v \sin(\theta)}{fk_v \sin(\widehat{\theta})}$. This system is asymptotically stable if and only if $\alpha_i > 0 \ \forall i$. This means that $\lambda > 0$, $\kappa(t) > 0$, $\widehat{f} > 0$, $\widehat{k}_u > 0$, $\widehat{k}_v > 0$ and $0 < \widehat{\theta} < \pi$.

Proof: [**Proposition 2**] The time varying linear system (20) when $\mathbf{b}(t) = 0$ is:

$$\dot{\boldsymbol{\varepsilon}} = -\lambda \mathbf{A}(t) \boldsymbol{\varepsilon} = -\lambda \begin{bmatrix} \kappa(t) \widehat{\mathbf{K}}^{-1} \mathbf{K}(t) & 0 \\ 0 & \widehat{\mathbf{F}}^{-1} \mathbf{F}(t) \end{bmatrix} \boldsymbol{\varepsilon}$$

If $\kappa(t) > 0$, $\widehat{\mathbf{K}}^{-1} \mathbf{K}(t) > 0$ and $\widehat{\mathbf{F}}^{-1} \mathbf{F}(t) \ \forall t$ then $\mathbf{A}(t) > 0 \ \forall t$. Consider the following positive Lyapounov function:

$$\mathcal{V}(\boldsymbol{\varepsilon}) = \frac{1}{2} \boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon}$$

The derivative of this function is:

$$\dot{\mathcal{V}}(\boldsymbol{\varepsilon}) = \boldsymbol{\varepsilon}^\top \dot{\boldsymbol{\varepsilon}} = -\lambda \boldsymbol{\varepsilon}^\top \mathbf{A}(t) \boldsymbol{\varepsilon}$$

which is always negative if $\boldsymbol{\varepsilon} \neq 0$, since $\lambda > 0$ and $\mathbf{A}(t) > 0$. Thus the system is locally stable. To prove the asymptotic convergence of $\boldsymbol{\varepsilon}$ to zero, we need also to show that it exists a neighborhood $\mathcal{U}$ of $\boldsymbol{\xi}^* = 0$ such that $\boldsymbol{\varepsilon} = \widehat{\mathbf{J}}^+(\boldsymbol{\xi}^*)(\mathbf{s} - \mathbf{s}^*) \neq 0$, $\forall \boldsymbol{\xi} \in \mathcal{U}$ (i.e. $\boldsymbol{\varepsilon} = 0$ only if $\mathbf{s}(\boldsymbol{\xi}) = \mathbf{s}^*$). Let us suppose that $\mathbf{s}(\boldsymbol{\xi}) \neq \mathbf{s}^*$ and therefore $\boldsymbol{\xi} \neq \boldsymbol{\xi}^* = 0$. The Taylor development of $\mathbf{s}(\boldsymbol{\xi})$ in a neighborhood of $\boldsymbol{\xi}^* = 0$ is:

$$\mathbf{s} - \mathbf{s}^* = \mathbf{J}(\boldsymbol{\xi}^*) \, \boldsymbol{\xi} + O^2(\boldsymbol{\xi}) \tag{29}$$

Multiplying by $\boldsymbol{\xi}^\top \widehat{\mathbf{J}}^+(\boldsymbol{\xi}^*)$ (where $\boldsymbol{\xi}^\top \widehat{\mathbf{J}}^+(\boldsymbol{\xi}^*) \neq 0$ since $\boldsymbol{\xi} \neq 0$ and $\widehat{\mathbf{J}}^+(\boldsymbol{\xi}^*)$ is full rank) both sides of equation (29) we obtain:

$$\boldsymbol{\xi}^\top \widehat{\mathbf{J}}^+(\boldsymbol{\xi}^*)(\mathbf{s} - \mathbf{s}^*) = \boldsymbol{\xi}^\top \widehat{\mathbf{J}}^+(\boldsymbol{\xi}^*) \mathbf{J}(\boldsymbol{\xi}^*) \boldsymbol{\xi} + O^3(\boldsymbol{\xi})$$

remember that if $\mathbf{A}(t) = \widehat{\mathbf{J}}^+(\boldsymbol{\xi}^*) \mathbf{J}(\boldsymbol{\xi}^*) > 0$ then $\boldsymbol{\xi}^\top \mathbf{A}(t) \boldsymbol{\xi} \geq 2\sigma \|\boldsymbol{\xi}\|^2$, where $\sigma > 0$ is the minimum singular value of the positive matrix $\mathbf{A}(t) + \mathbf{A}^\top(t)$. If $\widehat{\mathbf{J}}^+(\boldsymbol{\xi}^*)(\mathbf{s} - \mathbf{s}^*) = 0$ then:

$$0 \geq 2\sigma \|\boldsymbol{\xi}\|^2 + O^3(\boldsymbol{\xi})$$

this means that, since $2\sigma \|\boldsymbol{\xi}\|^2 > 0$, it exists a neighborhood $\mathcal{V}$ of $\boldsymbol{\xi} = \boldsymbol{\xi}^* = 0$ such that:

$$\|\boldsymbol{\xi}\|^2 \leq \frac{|O^3(\boldsymbol{\xi})|}{2\sigma} \qquad \forall \boldsymbol{\xi} \in \mathcal{V} \tag{30}$$

On the other hand, by definition of $O^3(\boldsymbol{\xi})$:

$$\lim_{\boldsymbol{\xi} \to 0} \frac{1}{2\sigma} \frac{|O^3(\boldsymbol{\xi})|}{\|\boldsymbol{\xi}\|^2} = 0$$

Thus, it exists a neighborhood $\mathcal{U}$ of $\boldsymbol{\xi} = \boldsymbol{\xi}^* = 0$ in which:

$$\|\boldsymbol{\xi}\|^2 > \frac{|O^3(\boldsymbol{\xi})|}{2\sigma} \qquad \forall \boldsymbol{\xi} \in \mathcal{U} \tag{31}$$

In the intersection of the two neighborhoods $\mathcal{W} = \mathcal{U} \cap \mathcal{V}$ both inequalities (30) and (31) must be satisfied, which is

impossible. Thus, $\boldsymbol{\varepsilon} = \widehat{\mathbf{J}}^{+}(\boldsymbol{\xi}^{*})(\mathbf{s}-\mathbf{s}^{*}) \neq 0$ in a neighborhood of $\boldsymbol{\xi}^{*}$ and the system is locally asymptotically stable.

Proof: [**Proposition 3**] The time derivative of the norm of the task function is:

$$\frac{d\|\boldsymbol{\varepsilon}\|^2}{dt} = \frac{d(\boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon})}{dt} = \boldsymbol{\varepsilon}^\top \dot{\boldsymbol{\varepsilon}} + \dot{\boldsymbol{\varepsilon}}^\top \boldsymbol{\varepsilon}$$

Using equation (20), we obtain:

$$\frac{d\|\boldsymbol{\varepsilon}\|^2}{dt} = -\lambda \boldsymbol{\varepsilon}^\top (\mathbf{A}(t) + \mathbf{A}^\top(t))\boldsymbol{\varepsilon} + 2\mathbf{b}(t)^\top \boldsymbol{\varepsilon} \tag{32}$$

From Proposition 2, if the system is locally stable when $\mathbf{b}(t) = 0$, then $\mathbf{A}(t) > 0$. If $\sigma > 0$ is the constant lower bound of the minimum singular value of $(\mathbf{A}(t) + \mathbf{A}^\top(t))$, then $\boldsymbol{\varepsilon}^\top(\mathbf{A}(t) + \mathbf{A}^\top(t))\boldsymbol{\varepsilon} \geq \sigma\|\boldsymbol{\varepsilon}\|^2$. Suppose that $\frac{\partial \mathbf{s}^*(t)}{\partial t}$ is bounded (i.e. $\|\mathbf{b}(t)\| \leq \frac{1}{2}\psi$), equation (32) can be bounded as follow:

$$\frac{d\|\boldsymbol{\varepsilon}\|^2}{dt} \leq -\varphi \, \|\boldsymbol{\varepsilon}\|^2 + \psi \, \|\boldsymbol{\varepsilon}\| \tag{33}$$

where $\varphi = \lambda\sigma$. Setting $\boldsymbol{\chi}(t) = \|\boldsymbol{\varepsilon}\|^2$, consider the following Bernoulli differential equation:

$$\dot{\chi} = -\varphi \, \chi + \psi \chi^{\frac{1}{2}} \tag{34}$$

whose solution is:

$$\chi(t) = e^{-\int_0^t \varphi d\tau} \left(\chi(0) + \frac{1}{2} \int_0^t \psi e^{\frac{1}{2}\int_0^t \varphi d\tau} d\tau \right)^2 \tag{35}$$

Remember that $\|\boldsymbol{\varepsilon}(0)\| = 0$, thus $\chi(0) = 0$ and:

$$\chi(t) = \left(\frac{\psi}{\varphi}\right)^2 \left(1 - e^{-\frac{1}{2}\varphi t}\right)^2 \tag{36}$$

Finally, since $\|\boldsymbol{\varepsilon}(t)\| = \sqrt{\boldsymbol{\chi}(t)}$ we obtain:

$$\|\boldsymbol{\varepsilon}(t)\| \leq \frac{\psi}{\varphi} \tag{37}$$

From proposition 2, we know that in a neighborhood of $\boldsymbol{\xi}^*$ we have $\|\boldsymbol{\varepsilon}\| = \|\widehat{\mathbf{J}}^{+}(\mathbf{s}-\mathbf{s}^{*})\| \neq 0$ if $\mathbf{s} \neq \mathbf{s}^{*}$. Then, $\exists \delta > 0$ such that $\|\boldsymbol{\varepsilon}\| = \|\widehat{\mathbf{J}}^{+}(\mathbf{s}-\mathbf{s}^{*})\| \geq \delta\|\mathbf{s}-\mathbf{s}^{*}\|$. Thus, from equation (37) one obtain:

$$\delta\|\mathbf{s}-\mathbf{s}^{*}\| \leq \|\boldsymbol{\varepsilon}(t)\| \leq \frac{\psi}{\varphi} \tag{38}$$

from which we can deduce:

$$\|\mathbf{s}-\mathbf{s}^{*}\| \leq \frac{\psi}{\delta\varphi} \tag{39}$$

If $\|\boldsymbol{\varepsilon}\|$ is bounded, then error $\|\mathbf{s}-\mathbf{s}^{*}\|$ is also bounded.

References

[1] K. Hashimoto, *Visual Servoing*, vol. 7 of *World Scientific Series in Robotics and Automated Systems*, World Scientific Press, Singapore, 1993.

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

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

[4] 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*, vol. 12, no. 5, pp. 684–696, October 1996.

[5] E. Malis, F. Chaumette, and S. Boudet, "2 1/2 d visual servoing," *IEEE Trans. on Robotics and Automation*, vol. 15, no. 2, pp. 234–246, April 1999.

[6] F. Chaumette, "Potential problems of stability and convergence in image-based and position-based visual servoing," in *The confluence of vision and control*, vol. 237 of *LNCIS Series*, pp. 66–78. Springer Verlag, 1998.

[7] F. Chaumette and E. Malis, "2 1/2 d visual servoing: a possible solution to improve image-based and position-based visual servoings," in *IEEE Int. Conf. on Robotics and Automation*, San Francisco, USA, April 2000, vol. 1, pp. 630–635.

[8] S. Maybank and O. Faugeras, "A theory of self-calibration of a moving camera," *International Journal of Computer Vision*, vol. 8, no. 2, pp. 123–151, 1992.

[9] M. Pollefeys, R. Koch, and L. Van Gool, "Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters," in *Int. Journal of Computer Vision*, 32(1):7–25, August 1999.

[10] P. Sturm, "Critical motion sequences for the self-calibration of cameras and stereo systems with variable focal length," *Image and Vision Computing*, 20(5-6):415–426, March 2002.

[11] G. D. Hager, "Calibration-free visual control using projective invariance," in *IEEE Int. Conf. on Computer Vision*, MIT, Cambridge (USA), June 1995, pp. 1009–1015.

[12] E. Malis, "Visual servoing invariant to changes in camera intrinsic parameters," in *International Conference on Computer Vision*, Vancouver, Canada, July 2001, vol. 1, pp. 704–709.

[13] E. Malis, "Vision-based control using different cameras for learning the reference image and for servoing," in *IEEE/RSJ Int. Conf. on Intelligent Robots Systems*, Maui, Hawaii, vol. 3, pp. 1428-1433, November 2001.

[14] E. Malis, "Vision-based control invariant to camera intrinsic parameters: stability analysis and path tracking," in *IEEE Int. Conference on Robotics and Automation*, Washington, D.C., USA, vol. 1, pp. 217-222, May 2002.

[15] Y. Mezouar and F. Chaumette, "Path planning in image space for robust visual servoing," in *IEEE Int. Conf. on Robotics and Automation*, San Francisco, CA, vol. 3, pp. 2759-2764, April 2000.

[16] O. Faugeras, *Three-dimensional computer vision: a geometric viewpoint*, MIT Press, Cambridge, MA, 1993.

[17] J. Mundy and A. Zisserman, *Geometric invariance in computer vision*, MIT Press, 1992.

[18] C. Samson, M. Le Borgne, and B. Espiau, *Robot Control: the Task Function Approach*, vol. 22 of *Oxford Engineering Science Series*, Clarendon Press, Oxford, UK, 1991.

[19] J. T. Feddema and O. R. Mitchell, "Vision-guided servoing with feature-based trajectory generation," *IEEE Trans. on Robotics and Automation*, vol. 5, no. 5, pp. 691–700, October 1989.

[20] N. P. Papanikolopoulos and P. K. Kosla, "Adaptive robot visual tracking: Theory and experiments," *IEEE Trans. on Robotics and Automation*, vol. 38, no. 3, pp. 429–445, March 1993.

Ezio Malis was born in Gorizia, Italy, in 1970 and graduated both from the University Politecnico di Milano and from the Ecole Supérieure d'Electricité (Suplec), Paris, in 1995. After a three years joint research work with IRISA/INRIA Rennes and with the national French company of electricity power (EDF), he received the Ph.D degree from the University of Rennes in 1998. He spent two years as research associate at the University of Cambridge (UK), where he participated to the Esprit project VIGOR (Visually guided robots using uncalibrated cameras). He joined the ICARE team at INRIA Sophia-Antipolis in 2000 as research scientist. His research interests include automatics, robotics, computer vision, and in particular vision-based control.