Vision-based control invariant to camera intrinsic parameters: stability analysis and path tracking

Ezio Malis

Abstract— **This paper concerns the stability analysis of a new vision-based control which is invariant to camera intrinsic parameters. 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. Even if the camera is uncalibrated, 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.**

I. Introduction

Most visual servoing techniques are based on a "teaching-by-showing" approach [1], [2]: the robot is moved to a goal position, the camera is shown the target view and a reference image of the object is stored. With this approach, whatever is the visual servoing method used to drive the robot back to the reference position (image-based [3], position-based [4] or hybrid [5] visual servoing), the camera used for servoing will be able to reach the reference image if and only if its 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) the reference image must be shown again. In [6] I proposed a new vision-based control, invariant to camera intrinsic parameters, which allows us to learn the reference image once and for all and can be used with a zooming camera. 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 [7] for stereo visual control. The system proposed in [7] 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 on the cameras used to realize 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. Experiments have shown that the teaching-by-showing technique can be extended to the case when different uncalibrated cameras are used for learning the reference image and for servoing [8]. Even if the measured task function does not depend on the camera intrinsic parameters, they are needed to estimate the Jacobian matrix which links the camera velocity to the displacements of the features in the invariant space [8]. Thus, calibration errors can affect the stability of the control law. For this reason, this paper concerns the stability analysis of the proposed method and shows its robustness with respect to camera calibration errors. In particular, the necessary and sufficient conditions for the 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. Contrarily to [9], 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 easily solved by decreasing the focal length in order to obtain a larger field of view.

Ezio Malis is with I.N.R.I.A. - ICARE Project, Sophia Antipolis, France. E-mail: Ezio.Malis@sophia.inria.fr

II. Theoretical background

A. Notations

Let the point $\mathcal{C}^*$ of the 3D Cartesian space, be the center of projection. 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 $\mathbf{m}^*$: $\zeta^*\mathbf{m}^* = \left[\begin{array}{cc} \mathbf{I}_3 & \mathbf{0} \end{array}\right] \boldsymbol{\mathcal{X}}$, 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 $\mathbf{r} = \theta\mathbf{u}$ 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 $\mathbf{m}$: $\zeta(\boldsymbol{\xi})\, \mathbf{m}(\boldsymbol{\xi}) = \left[\begin{array}{cc} \mathbf{R} & \mathbf{t} \end{array}\right] \boldsymbol{\mathcal{X}}$, where $\zeta(\boldsymbol{\xi})$ is the depth.

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 $\mathbf{p} = (u, v, 1)$ observed in the image $\mathcal{I}$, taken at the position $\mathcal{F}$, depends on the camera internal parameters:

$$\mathbf{p}(\boldsymbol{\xi}, \mathbf{K}) = \mathbf{K}\mathbf{m}(\boldsymbol{\xi}) \qquad \mathbf{p} \in \mathcal{I}(\boldsymbol{\xi}, \mathbf{K}) \tag{1}$$

where:

$$\mathbf{K} = \begin{bmatrix} fk_u & -fk_u \cot(\theta) & u_0 \\ 0 & fk_v / \sin(\theta) & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

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:

$$\mathbf{p}^*(\boldsymbol{\xi}^*, \mathbf{K}^*) = \mathbf{K}^*\mathbf{m}^*(\boldsymbol{\xi}^*) \qquad \mathbf{p}^* \in \mathcal{I}(\boldsymbol{\xi}^*, \mathbf{K}^*) \tag{3}$$

where $\mathbf{K}^*$ has the same form of $\mathbf{K}$ given in equation (2). 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 is independent on the camera intrinsic parameters. That is possible by using the simple projective transformation proposed in [6]. 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 (1) and equation (3) as follows:

$$\mathbf{Q}(\boldsymbol{\xi}, \mathbf{K}) = \mathbf{K}\mathbf{M}(\boldsymbol{\xi}) \tag{4}$$

$$\mathbf{Q}^*(\boldsymbol{\xi}^*, \mathbf{K}^*) = \mathbf{K}^*\mathbf{M}^*(\boldsymbol{\xi}^*) \tag{5}$$

where $\mathbf{Q} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \end{bmatrix}$, $\mathbf{M} = \begin{bmatrix} \mathbf{m}_1 & \mathbf{m}_2 & \mathbf{m}_3 \end{bmatrix}$, $\mathbf{Q}^* = \begin{bmatrix} \mathbf{p}_1^* & \mathbf{p}_2^* & \mathbf{p}_3^* \end{bmatrix}$ and $\mathbf{M}^* = \begin{bmatrix} \mathbf{m}_1^* & \mathbf{m}_2^* & \mathbf{m}_3^* \end{bmatrix}$. The matrices $\mathbf{Q}$ and $\mathbf{Q}^*$ 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 transformed points $\mathbf{q}$ and $\mathbf{q}^*$ are:

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

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 camera internal parameters. However, $\mathbf{q}$ and $\mathbf{q}^*$ are also invariant on the rotation around the $\overrightarrow{z}$ axis [6]. 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 can be controlled separately, as explained in Section IV. Thus, there are no problems coming from the control of r_z when the displacement is large, as for image-based visual servoing [10], [11].

III. Path planning in the invariant space

The epipolar geometry in the invariant space is equivalent to a plane + parallax factorization [6]. 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 $\mathbf{q}_k^*$ is [6]:

$$\gamma_k\, \mathbf{q}_k = \boldsymbol{\Gamma}(\mathbf{q}_k^* + \mu_k^*\mathbf{e}^*) \tag{6}$$

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 (6) 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. 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 (6)) 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 $\Delta(0) = \text{diag}(\frac{\gamma_1}{\gamma_3}, \frac{\gamma_2}{\gamma_3}, 1)$ and $\Delta(T) = \text{diag}(1,1,1)$:

$$\Delta(t) = \Delta(0) + \frac{t}{T}(\Delta(T) - \Delta(0))$$

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:

$$\mathbf{q}_k^*(t) = \frac{\Delta(t)(\mathbf{q}_k^*(T) + \beta_k(t)\mathbf{e}^*/\|\mathbf{e}^*\|)}{[1\ 1\ 1]\,\Delta(t)(\mathbf{q}^*(T) + \beta(t)\mathbf{e}^*/\|\mathbf{e}^*\|)} \quad (7)$$

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 Cartesian space. The 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

The control 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:

$$\dot{\mathbf{s}} = \mathbf{J}(\boldsymbol{\xi})\,\boldsymbol{\mu} \quad (8)$$

where $\boldsymbol{\mu} = (\nu_x, \nu_y, \nu_z, \omega_x, \omega_y)$ and $\mathbf{J}(\boldsymbol{\xi})$ is the Jacobian matrix which can be built from the Jacobian matrix $\mathbf{J}_k(\boldsymbol{\xi})$ relative to each invariant point: $\dot{\mathbf{q}}_k = \mathbf{J}_k(\boldsymbol{\xi})\boldsymbol{\mu}$ [8]. It must be emphasized that each block of the matrix can be written as follows:

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

where:

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

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

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

Therefore, the Jacobian can be decomposed as follows:

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

This matrix depends on the depths distribution $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, ..., \zeta_n)$ and on the camera intrinsic parameters $\mathbf{K}(t)$ which can eventually vary during the servoing. 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 $\kappa(t) > 0$ $\forall t$), as explained in [12]. On the other hand, the camera parameters can eventually vary during the servoing when using a zooming camera. 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 Jacobian matrix. Since $\boldsymbol{\Phi}(\widehat{\boldsymbol{\zeta}}) = \frac{1}{\kappa}\boldsymbol{\Phi}(\boldsymbol{\zeta})$, the estimated Jacobian matrix can be written as:

$$\widehat{\mathbf{J}} = \begin{bmatrix} \frac{1}{\kappa}\boldsymbol{\Phi}(\boldsymbol{\zeta}) & \boldsymbol{\Psi} \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{K}} & 0 \\ 0 & \mathbf{G}(\widehat{\mathbf{K}}) \end{bmatrix} \quad (9)$$

Let us suppose that the Jacobian matrix $\mathbf{J}$ is full rank. If $\mathbf{J}$ is full rank, then also estimated matrix $\widehat{\mathbf{J}}$ is full rank. In order to control the five d.o.f. of the camera I use the task function approach [13]. Consider the following (5×1) task function:

$$\boldsymbol{\varepsilon}(\boldsymbol{\xi}) = \widehat{\mathbf{J}}^{+}(\boldsymbol{\xi})(\mathbf{s}(\boldsymbol{\xi}) - \mathbf{s}^*(t)) \quad (10)$$

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 (7). Obviously, if $\mathbf{s} = \mathbf{s}^*$ then $\boldsymbol{\varepsilon} = 0$. On the other hand, sufficient conditions are given in

[12] such that if $\|\mathbf{s}-\mathbf{s}^*\|$ is sufficiently small then $\boldsymbol{\varepsilon}=0$ only if $\mathbf{s}=\mathbf{s}^*$ (i.e. $\mathbf{s}-\mathbf{s}^*$ never belongs to $\mathrm{Ker}(\widehat{\mathbf{J}}^+)$). Differentiating equation (10) we obtain:

$$\dot{\boldsymbol{\varepsilon}} = \frac{d\widehat{\mathbf{J}}^+}{dt}(\mathbf{s}-\mathbf{s}^*(t)) + \widehat{\mathbf{J}}^+\dot{\mathbf{s}} - \widehat{\mathbf{J}}^+\frac{\partial \mathbf{s}^*(t)}{\partial t} \tag{11}$$

Using equation (8) and since the vector $\frac{d\widehat{\mathbf{J}}^+}{dt}(\mathbf{s}-\mathbf{s}^*(t))$ can be written as $\frac{d\widehat{\mathbf{J}}^+}{dt}(\mathbf{s}-\mathbf{s}^*(t)) = \mathbf{O}(\mathbf{s}-\mathbf{s}^*(t))\boldsymbol{\mu}$ (where $\mathbf{O}(\mathbf{s}-\mathbf{s}^*(t)) \rightarrow 0$ if $\mathbf{s}\rightarrow\mathbf{s}^*(t)$), we obtain:

$$\dot{\boldsymbol{\varepsilon}} = \left(\mathbf{O}(\mathbf{s}-\mathbf{s}^*(t)) + \widehat{\mathbf{J}}^+\mathbf{J}\right)\boldsymbol{\mu} - \widehat{\mathbf{J}}^+\frac{\partial \mathbf{s}^*(t)}{\partial t} \tag{12}$$

Consider the control law:

$$\boldsymbol{\mu} = -\lambda\boldsymbol{\varepsilon} + \widehat{\mathbf{J}}^+\frac{\partial \mathbf{s}^*(t)}{\partial t} \tag{13}$$

where λ is a positive gain. From equations (13) and (12) we obtain the following closed-loop equation:

$$\begin{aligned}\dot{\boldsymbol{\varepsilon}} &= -\lambda\left(\mathbf{O}(\mathbf{s}-\mathbf{s}^*(t)) + \widehat{\mathbf{J}}^+\mathbf{J}\right)\boldsymbol{\varepsilon} \\ &+ \left(\mathbf{O}(\mathbf{s}-\mathbf{s}^*(t)) + \widehat{\mathbf{J}}^+\mathbf{J} - \mathbf{I}\right)\widehat{\mathbf{J}}^+\frac{\partial \mathbf{s}^*(t)}{\partial t}\end{aligned} \tag{14}$$

In order to study the behavior of the task function during the path tracking, consider the system (14) linearized for $\mathbf{s}=\mathbf{s}^*(t)$:

$$\dot{\boldsymbol{\varepsilon}} = -\lambda\,\mathbf{A}(t)\,\boldsymbol{\varepsilon} + \mathbf{b}(t) \tag{15}$$

where:

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

Let us consider first the case when the starting point $\boldsymbol{\varepsilon}(0)$ is in a neighborhood of $\boldsymbol{\varepsilon}=0$ 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 (15) 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 [12]. This proposition means that the control law is extremely robust to calibration errors since any positive approximation of the parameters is sufficient to stabilize the system. When the camera parameters vary during the servoing, we have:

Proposition 2: *The equilibrium point $\boldsymbol{\varepsilon}=0$, of the time varying linear system (15) when $\mathbf{b}(t)=0$, is locally asymptotically stable if $\kappa(t)>0$, $\widehat{\mathbf{K}}^{-1}\mathbf{K}(t)>0$ and $\widehat{\mathbf{G}}^{-1}\mathbf{G}(t)>0$.*

The proof is given in [12]. The sufficient conditions are the same found in [5] and they are very large. This means that one can 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. Consider now the general case when a path computed from equation (7) is tracked (i.e. $\mathbf{b}(t)\neq 0$) and the camera parameters are eventually varying during the servoing.

Proposition 3: *If the system is locally stable (see Proposition 2) and $\mathbf{b}(t)$ is bounded $\|\mathbf{b}(t)\|\leq\psi/2$, 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:*

$$\|\boldsymbol{\varepsilon}(t)\| \leq \frac{\psi}{\varphi} \tag{16}$$

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

The proof is presented in [12]. 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 the 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 when calibration errors increase, 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:

$$\mathbf{T} = \mathbf{Q}\mathbf{Q}^{*-1} = \begin{bmatrix}\tau_{11} & \tau_{12} & \tau_{13} \\ \tau_{21} & \tau_{22} & \tau_{23} \\ 0 & 0 & 1\end{bmatrix} \tag{17}$$

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 (5) and (4) we obtain:

$$\mathbf{T} = \mathbf{Q}\mathbf{Q}^{*-1}|_{\boldsymbol{\xi}=\boldsymbol{\xi}^*} = \mathbf{K}(t)\mathbf{K}^{*-1} \tag{18}$$

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:

$$\dot{\epsilon} = a(\epsilon, t)\omega_z + \mathbf{c}^T(t)\boldsymbol{\mu}(t) \tag{19}$$

where $a(\epsilon, t)$ is a scalar, $\mathbf{c}(t)$ is a (5×1) vector and $\boldsymbol{\mu}(t)$ is the control law (13). 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:

$$\omega_z = -\frac{\eta\ \epsilon}{\widehat{a}(\epsilon, t)} - \frac{\widehat{\mathbf{c}}^T \boldsymbol{\mu}(t)}{\widehat{a}(\epsilon, t)} \tag{20}$$

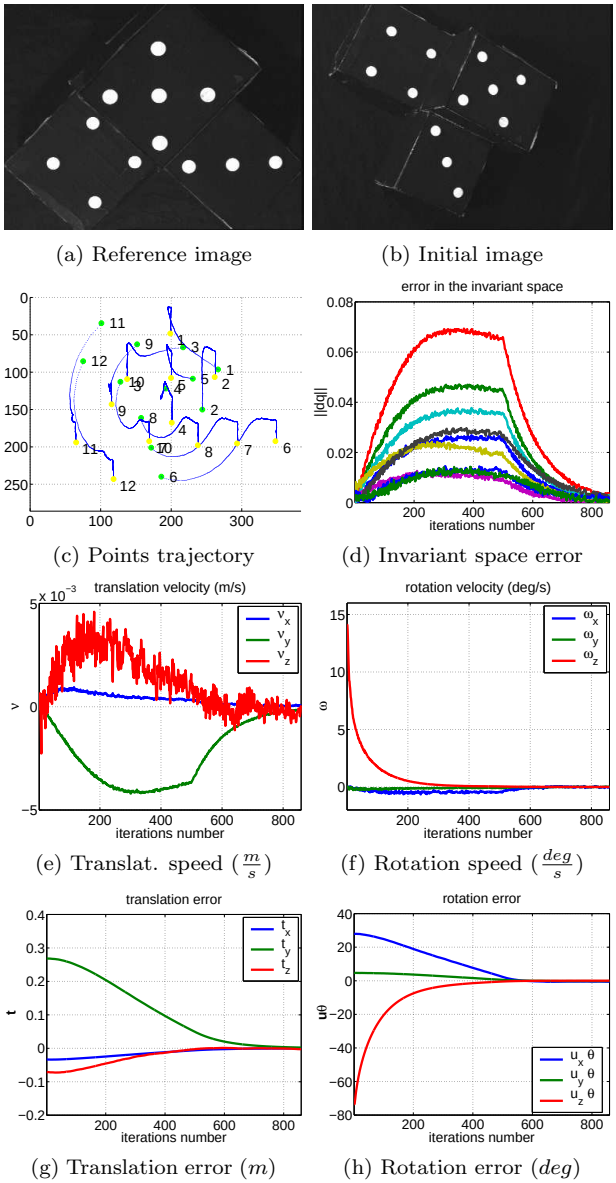
where η is a positive scalar tuning the speed of convergence, $\widehat{a}$ and $\widehat{\mathbf{c}}$ are approximations of a and $\mathbf{c}$. The closed loop equation is the following:

$$\dot{\epsilon} = -\eta \frac{a(\epsilon, t)}{\widehat{a}(\epsilon, t)}\epsilon + \left(\mathbf{c}^T(t) - \frac{a(\epsilon, t)}{\widehat{a}(\epsilon, t)}\widehat{\mathbf{c}}^T(t)\right)\boldsymbol{\mu}(t) \tag{21}$$

Note that $\boldsymbol{\mu}(t) \to 0$ and $\mathbf{c}(t)$ is bounded. Thus 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). The robot is very well calibrated and it provides a ground truth in order to measure the positioning precision of visual servoing. A camera (with a 6mm lens) is mounted on the robot end-effector and observe an "object" composed by 12 points. In the experiments, 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 1(a)) has been learned the robot is displaced to its initial position (Figure 1(b)). The initial displacement of the robot is approximately 30 *cm* for the translation and 80 *deg* for the rotation (see the initial values of the curves in Figures 1(i) and (j)). A path $\mathbf{s}^*(t)$ is planned in the invariant space such that the camera follows a straight line in the 3D space. 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 $\partial \mathbf{s}^*(t)/\partial t \neq 0$. The first part of the control law corresponding to the translational velocity ν_x, ν_y, ν_z is plotted in Figure 1(g) while the second part corresponding to the translational velocity $\omega_x, \omega_y, \omega_z$ is plotted in Figure 1(h). The error in the invariant space converges to zero (see Figure 1(d)). Consequently, the position error of the robot end-effector converges to zero (see Figures 1(i) and (j)). Figure 1(c) plots the trajectory of the points in the image and shows that the points go from their initial position (green points) 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

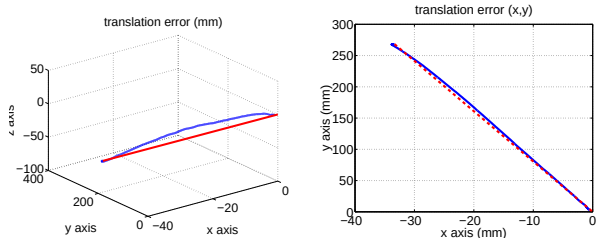
(c) Points trajectory (d) Invariant space error

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

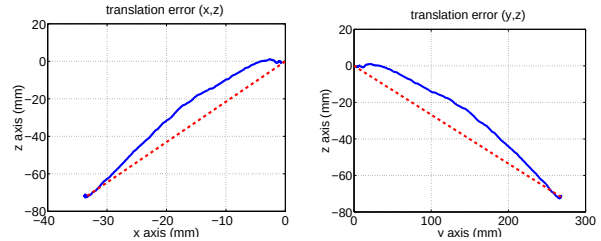
(g) Translation error (m) (h) Rotation error (deg)

Fig. 1. Path tracking in the invariant space. The error in the invariant space stay small. The invariant vision-based control is stable and the camera is back to the reference position.

Although the camera is controlled in the invariant space, the behavior of the robot in the Cartesian space is very nice. Figure (2) 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 (2)(c) and (d). In the second experiment, the robot starts approximately from the same initial position. The visual servoing results are therefore similar to the results shown in Figure 1. On the other hand, the path is sampled $N = 3000$ times. Thus, the speed of the reference trajectory is reduced by a factor 6.

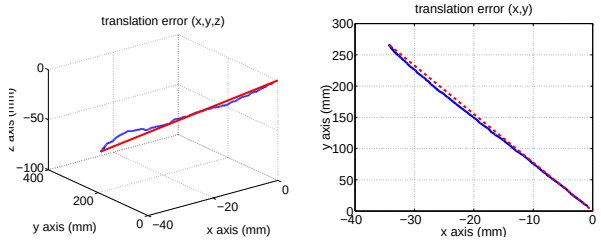


(a) translation error (b) error in plane (x,y)

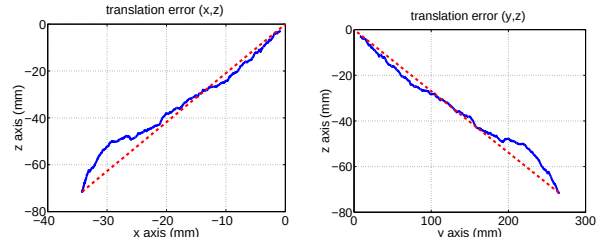


(c) error in plane (x,z) (d) error in plane (y,z)

Fig. 2. The path is sampled with N=500. The camera approximately follows a straight line.



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



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

Fig. 3. The path is sampled with N=3000. The tracking error is reduced and the camera follows the line more accurately.

As a consequence, the tracking error presented in Figure 3 is considerably smaller than the tracking error presented in Figure 2.

VI. Conclusion

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. Even if there is not direct control in the image, a zooming camera can be used to enlarge the field of view of the camera and/or to bound the size of the object observed in the image.

Acknowledgments

I would like to thank Eric Marchand (IRISA, VISTA project) for his help in programming the experiments.

References

[1] K. Hashimoto, *Visual Servoing: Real Time Control of Robot manipulators based on visual sensory feedback*, 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] E. Malis, "Visual servoing invariant to changes in camera intrinsic parameters," in *Int. Conf. on Computer Vision*, Vancouver, Canada, July 2001, vol. 1, pp. 704–709.

[7] 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.

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

[9] 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, April 2000, vol. 3, pp. 2759–2764.

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

[11] P. I. Corke and S. A. Hutchinson, "A new hybrid image-based visual servo control scheme," in *IEEE Conference on Decision and Control*, Sydney, NSW, Australia, December 2000, vol. 3, pp. 2521–2526.

[12] E. Malis, "Visual servoing invariant to changes in camera intrinsic parameters," Tech. Rep. 4309, INRIA, November 2001.

[13] 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.