Int. Conf. on Computer Vision, ICCV'01, Vol. 1, pages 704-709, Vancouver, Canada, July 2001.

Visual Servoing Invariant to Changes in Camera Intrinsic Parameters

Ezio Malis
I.N.R.I.A.
2004, route des Lucioles - B.P. 93,
06902 Sophia Antipolis Cedex, France.
Ezio.Malis@sophia.inria.fr

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. The scheme proposed in the paper differs from previous techniques in that it is camera independent. With the new scheme 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.

1. Introduction

The aim of visual servoing is to control the movement of a robot using the information provided by vision sensors. In order to position a camera with respect to an object, visual servoing schemes are generally based on a "teaching-by-showing" approach [3] [6] [7]. With this approach, the robot is moved to a goal position, the camera is shown the target view and a "reference image" is stored (i.e. a set of reference image features). 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, a visual servoing scheme can be used to reposition the camera with respect to the object [4] [12] [1] [9]. Each scheme has its own advantages and drawbacks [2], 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 of the camera used during the learning. Indeed, if the camera intrinsic parameters change during the servoing (or the camera used during the servoing is different from the camera used for learning), 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. For example, if one double the focal length, at the convergence the distance with respect to the object will be a factor of two smaller than the distance at the reference position. Consequently, we will call standard visual servoing techniques "camera-dependent" since they assume the following constraining hypothesis: "*the camera intrinsic parameters at the convergence must be the same camera parameters used during the learning*".

The aim of this paper is to eliminate the above hypothesis in order to increase the versatility and the domain of application of visual servoing techniques. 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. On the other hand, the visual servoing technique proposed in the paper allows us to learn the reference image once and for all. The basic idea is to use projective invariance so as to build an error function, from only measured image features, which is invariant on the intrinsic parameters of the camera. Projective invariance has been widely used in computer vision especially for matching and recognition (see for example [13] [11]). In visual servoing, projective invariance was used in [5] to define setpoints for stereo visual control that are independent of viewing location. The system proposed in [5] 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, observing a non-planar object, mounted on a robot manipulator. The projective transformation proposed in [9] is used to define an error only depending on the position of the camera with respect to the object. Consequently, the new image-based visual servoing technique is able to position a camera with respect to a non-planar object given a reference image observed with a different camera and even if the camera parameters change during servoing.

2. Theoretical background

2.1. Perspective projection

Let $\mathcal{C}^*$ be the center of projection (see Figure 1) coinciding 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 distance f^* between $\mathcal{C}^*$ and the plane of projection is called the *focal length*. A 3D point with homogeneous coordinates $\boldsymbol{\mathcal{X}} = \begin{bmatrix} X & Y & Z & 1 \end{bmatrix}^T$ is projected to the point $\mathbf{m}^*$:

$$\zeta^* \mathbf{m}^* = \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \boldsymbol{\mathcal{X}} \tag{1}$$

where $\zeta^* = \frac{Z}{f^*}$. Let the point $\mathcal{C}$ be a different centre of projection coinciding with the origin of frame $\mathcal{F}$ (see Figure 1).

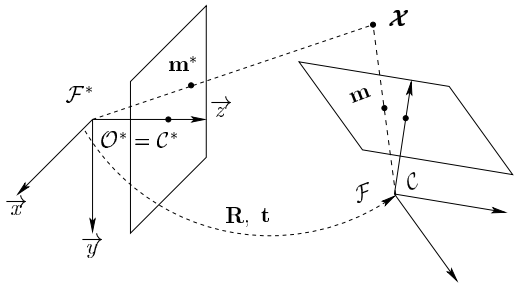


Figure 1. Perspective projections of a 3D point

The 3D point $\boldsymbol{\mathcal{X}}$ projects to the point $\mathbf{m}$ in the frame $\mathcal{F}$:

$$\zeta \mathbf{m} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix} \boldsymbol{\mathcal{X}} \tag{2}$$

where $\zeta = Z/f$, $\mathbf{R}$ and $\mathbf{t}$ are respectively the rotation and the translation between frame $\mathcal{F}^*$ and $\mathcal{F}$. Eliminating $\boldsymbol{\mathcal{X}}$ from equations (1) and (2) we obtain:

$$\zeta \mathbf{m} = \zeta^* \mathbf{R} \mathbf{m}^* + \mathbf{t} \tag{3}$$

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

2.2. Camera model

Pinhole cameras perform a perspective projection of a 3D point. However, the vectors $\mathbf{m}$ and $\mathbf{m}^*$ are not directly measured by the cameras. The information given by the pinhole cameras are two image points $\mathbf{p}$ and $\mathbf{p}^*$. The point $\mathbf{p} = \begin{bmatrix} u & v & 1 \end{bmatrix}^T$ observed in the image $\mathcal{I}$, taken at the position $\mathcal{F}$, depends on the camera internal parameters:

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

where:

$$\mathbf{K}(t) = \begin{bmatrix} fk_u & -fk_u \cot(\theta) & u_0 \\ 0 & fk_v / \sin(\theta) & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5}$$

u_0 and v_0 are the coordinates of the principal point (pixels), f is the focal length (meters), k_u et k_v are the magnifications respectively in the $\overrightarrow{u}$ and $\overrightarrow{v}$ direction (pixels/meters), and θ is the angle between these axes. The matrix $\mathbf{K}(t)$ depends on time since the camera intrinsic parameters may vary. On the other hand, the point $\mathbf{p}^* = \begin{bmatrix} u^* & v^* & 1 \end{bmatrix}^T$ observed in the image $\mathcal{I}$, taken at the reference position $\mathcal{F}^*$, depends on maybe different camera parameters:

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

The image $\mathcal{I}(\mathcal{F}^*, \mathbf{K}^*)$ 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 $\mathcal{I}(\mathcal{F}, \mathbf{K})$ currently observed. Both images $\mathcal{I}(\mathcal{F}, \mathbf{K})$ and $\mathcal{I}(\mathcal{F}^*, \mathbf{K}^*)$ depend on camera intrinsic parameters. Thus, even if $\mathcal{F} = \mathcal{F}^*$ the reference and the current image will be different if $\mathbf{K} \neq \mathbf{K}^*$ (i.e. $\mathcal{I}(\mathcal{F}^*, \mathbf{K}) \neq \mathcal{I}(\mathcal{F}^*, \mathbf{K}^*)$). As already mentioned in the introduction, visual servoing techniques are based on the hypothesis that the camera frame $\mathcal{F}$ will coincide to the reference frame $\mathcal{F}^*$ if $\mathbf{p} = \mathbf{p}^*, \forall\ \mathbf{p}, \mathbf{p}^*$ (supposing that a sufficient number of corresponding points are observed in the image). However, this hypothesis is good *if and only if* $\mathbf{K} = \mathbf{K}^*$ at the convergence. This is a very strong constraint which limit the versatility and the applications of visual servoing. From both theoretical and practical point of views, it would be interesting to eliminate such constraint.

3 An invariant projective space

In order to position a robot with respect to an object 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. This can be obtained using the simple projective transformation proposed in [9].

3.1. Invariance by a projective transformation

Let us consider three non-collinear 3D points $\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3$ of the observed object. These three points project to the points $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3$ in the current image and to the points $\mathbf{m}_1^*, \mathbf{m}_2^*, \mathbf{m}_3^*$ in the reference image. 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:

$$\mathbf{P} = \mathbf{K}(t)\mathbf{M} \tag{7}$$

$$\mathbf{P}^* = \mathbf{K}^* \mathbf{M}^* \tag{8}$$

where $\mathbf{P} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \end{bmatrix}$, $\mathbf{M} = \begin{bmatrix} \mathbf{m}_1 & \mathbf{m}_2 & \mathbf{m}_3 \end{bmatrix}$, and $\mathbf{P}^* = \begin{bmatrix} \mathbf{p}_1^* & \mathbf{p}_2^* & \mathbf{p}_3^* \end{bmatrix}$, $\mathbf{M}^* = \begin{bmatrix} \mathbf{m}_1^* & \mathbf{m}_2^* & \mathbf{m}_3^* \end{bmatrix}$. Let us suppose that the three chosen points are not collinear

in both images. The matrices $\mathbf{P}$ and $\mathbf{P}^*$ 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 respectively:

$$\mathbf{q} = \mathbf{P}^{-1}\mathbf{p} \qquad \mathbf{q} \in \mathcal{Q}(\mathcal{F}) \tag{9}$$

$$\mathbf{q}^* = \mathbf{P}^{*-1}\mathbf{p}^* \qquad \mathbf{q}^* \in \mathcal{Q}(\mathcal{F}^*) \tag{10}$$

From now on we will refer to the transformed images $\mathcal{Q}(\mathcal{F})$ and $\mathcal{Q}(\mathcal{F}^*)$ as two "invariant" images where the invariance is related to camera internal parameters. The transformed points $\mathbf{q}$ and $\mathbf{q}^*$ do not depend on the internal parameter of the cameras. Indeed, from equations (7) and (8) we obtain:

$$\mathbf{q} = \mathbf{P}^{-1}\mathbf{p} = \mathbf{M}^{-1}\mathbf{K}^{-1}\mathbf{K}\mathbf{m} = \mathbf{M}^{-1}\mathbf{m}$$

$$\mathbf{q}^* = \mathbf{P}^{*-1}\mathbf{p}^* = \mathbf{M}^{*-1}\mathbf{K}^{*-1}\mathbf{K}^*\mathbf{m}^* = \mathbf{M}^{*-1}\mathbf{m}^*$$

Therefore, $\mathbf{q}$ and $\mathbf{q}^*$ only depend on the position of the camera with respect to the observed object and on its three-dimensional structure. Obviously, the choice of the three reference points is important. This choice is automatically done by selecting three distant points maximizing the surface of the corresponding triangle in both images [9].

3.2. Potential problems due to the invariance

Potential problems can arise from the fact that the tranformed spaces $\mathcal{Q}^*$ and $\mathcal{Q}$ are not only invariant to camera intrinsic parameters. Indeed, invariance holds even if the point $\mathbf{m}$ is tranformed to $\mathbf{p}$ by any collineation of the form:

$$\mathbf{G}' = \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ 0 & 0 & 1 \end{bmatrix} \tag{11}$$

It is clear that $\mathbf{G}' = \mathbf{K}$ is a special case of the above collineation. If $\mathbf{p}' = \mathbf{G}'\mathbf{m} \ \ \forall \mathbf{m}$ then $\mathbf{P}' = \mathbf{G}'\mathbf{M}$ and:

$$\mathbf{q}' = \mathbf{P}'^{-1}\mathbf{p}' = \mathbf{M}^{-1}\mathbf{G}'^{-1}\mathbf{p}' = \mathbf{M}^{-1}\mathbf{m} = \mathbf{q}$$

The matrix $\mathbf{G}'$ must have the form given in equation (11) since $\mathbf{M}$ and $\mathbf{P}'$ must have the same third row. As a consequence, it is possible to reposition the camera with respect to a planar object using the visual servoing scheme proposed in this paper. Indeed, starting from a general collineation $\mathbf{G}$ between the two images, we will obtain a different collineation up to $\mathbf{G}'$. Another particular case is when the collineation is the rotation r_z around the $\overrightarrow{z}$ axis:

$$\mathbf{G}' = \mathbf{R}_z(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Consequently, the tranformed coordinates $\mathbf{q}$ and $\mathbf{q}^*$ are invariant to a rotation r_z around the $\overrightarrow{z}$ axis. Let $\overline{\mathcal{F}}$ and $\overline{\mathcal{F}}^*$ two reference frame defined up to a rotation around the $\overrightarrow{z}$ axis, then $\mathcal{Q}(\mathcal{F}) = \mathcal{Q}(\overline{\mathcal{F}})$ and $\mathcal{Q}(\mathcal{F}^*) = \mathcal{Q}(\overline{\mathcal{F}}^*)$. Consequently, a different information must be obtained from the images in order to control r_z as it is shown in Section 4.2.

3.3. The epipolar geometry in the invariant space

The epipolar geometry in the invariant space is equivalent to a plane + parallax factorisation [8] [9]. 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. The projections of the 3D points $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{P}_3$ are related by a homography matrix $\mathbf{H}$ as follows:

$$\begin{bmatrix} \mathbf{m}_1 & \mathbf{m}_2 & \mathbf{m}_3 \end{bmatrix} \mathbf{\Gamma} = \mathbf{H} \begin{bmatrix} \mathbf{m}_1^* & \mathbf{m}_2^* & \mathbf{m}_3^* \end{bmatrix}$$

where $\mathbf{H} = \mathbf{R} + \mathbf{t}\mathbf{n}^{*T}/d^*$, $\gamma_k = \frac{\zeta_k}{\zeta_k^*} \ \forall k$ and:

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_1 & 0 & 0 \\ 0 & \gamma_2 & 0 \\ 0 & 0 & \gamma_3 \end{bmatrix}$$

In the invariant space, $\mathbf{\Gamma}$ is the diagonal collineation matrix linking $\mathbf{q}_1$, $\mathbf{q}_2$ and $\mathbf{q}_3$ to $\mathbf{q}_1^*$, $\mathbf{q}_2^*$ and $\mathbf{q}_3^*$:

$$\mathbf{\Gamma} = \mathbf{M}^{-1}\mathbf{H}\mathbf{M}^* = \mathbf{P}^{-1}\mathbf{K}\mathbf{H}\mathbf{K}^{*-1}\mathbf{P}^* \tag{12}$$

For all the other points in the invariant space, equation (3) can be written as:

$$\gamma\, \mathbf{q} = \mathbf{\Gamma}(\mathbf{q}^* + \mu^*\mathbf{e}^*) \tag{13}$$

where $\mathbf{e}^*$ is the epipole in the reference image and μ^* a scale factor (if $\mu^* = 0$ then the 3D point $\mathcal{P}$, projecting to $\mathbf{q}$ and $\mathbf{q}^*$, lies on the plane defined by the points $\mathcal{P}_1$, $\mathcal{P}_2$ $\mathcal{P}_3$). From the current and reference image, one can measure γ, $\mathbf{\Gamma}$, μ^* and $\mathbf{e}^*$ up to a scale factor.

4 Control in the invariant space

The control the camera in the invariant space is divided into two different parts. Indeed, the points $\mathbf{q}$ are invariant on the rotation around the $\overrightarrow{z}$ axis and can only be used to control five d.o.f. of the camera. Another information must be extracted from the images to control the last camera d.o.f.

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

Suppose that m matched points are available in both images. Since three points are used to define the projective transformations, only the remaining $m - 3$ points can be used to control the camera. In order to control five d.o.f. of the camera we minimize the following task function [10]:

$$\mathbf{e}_1 = \mathbf{C}(\mathbf{s} - \mathbf{s}^*) \tag{14}$$

where $dim(\mathbf{e}_1) = (5 \times 1)$ and in our case:

$$\mathbf{s} = \begin{bmatrix} \mathbf{q}_4 \\ \mathbf{q}_5 \\ \vdots \\ \mathbf{q}_m \end{bmatrix} \quad \text{and} \quad \mathbf{s}^* = \begin{bmatrix} \mathbf{q}_4^* \\ \mathbf{q}_5^* \\ \vdots \\ \mathbf{q}_m^* \end{bmatrix}$$

the matrix $\mathbf{C}$, called the combination matrix, is a full rank $(3m \times 5)$ matrix. Note that if $\mathbf{e}_1 = 0$ and $\mathbf{s} - \mathbf{s}^*$ does not belong to the null space of $\mathbf{C}$ then $\mathbf{s} - \mathbf{s}^* = 0$. Thus, the camera frame $\mathcal{F}$ will coincide to $\mathcal{F}^*$ up to a rotation around the $\overrightarrow{z}$ axis. Generally, the combination matrix $\mathbf{C}$ is chosen as a function of the interaction matrix $\mathbf{L}_s$ linking the derivative of $\mathbf{s}$ to the velocity of the camera $\mathbf{v}_c = \left[\begin{array}{cccccc} \nu_x & \nu_y & \nu_z & \omega_x & \omega_y & \omega_z \end{array} \right]^T$:

$$\dot{\mathbf{s}} = \mathbf{L}_s \mathbf{v_c}$$

The interaction matrix $\mathbf{L}_s$ is a $(3(m-3) \times 6)$ matrix which depends on $\mathbf{K}$ and $\boldsymbol{\zeta} = \left[\begin{array}{cccc} \zeta_1 & \zeta_2 & ... & \zeta_m \end{array} \right]^T$ (both unknown), on the current image coordinates of the three points $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$ and on the points $\mathbf{q}_k \in \mathcal{Q}$. Since $\mathbf{q}$ is invariant on r_z the last column of $\mathbf{L}_s$ is null and $\max(\text{rank}(\mathbf{L}_s)) = 5$.

Differentiating equation (14) we obtain:

$$\dot{\mathbf{e}}_1 = \dot{\mathbf{C}}(\mathbf{s} - \mathbf{s}^*) + \mathbf{C}\dot{\mathbf{s}} = \dot{\mathbf{C}}(\mathbf{s} - \mathbf{s}^*) + \mathbf{C}\mathbf{L}'_s\mathbf{v}'_c$$

where $\mathbf{L}'_s$ is the matrix composed by the first five columns of $\mathbf{L}_s$ and $\mathbf{v}'_c = \left[\begin{array}{ccccc} \nu_x & \nu_y & \nu_z & \omega_x & \omega_y \end{array} \right]^T$. We will suppose that $\dot{\mathbf{C}}(\mathbf{s} - \mathbf{s}^*) \approx 0$. This is true near the convergence since $(\mathbf{s} - \mathbf{s}^*) \approx 0$ but experimental results show that it is a good approximation even if the initial error is large. By imposing the exponential convergence of the task function to zero (i.e. $\dot{\mathbf{e}}_1 = -\lambda_1 \mathbf{e}_1$) we have:

$$\mathbf{C}\mathbf{L}'_s\mathbf{v}'_c \approx -\lambda_1 \mathbf{e}_1$$

and the control law is (λ_1 being a positive scalar):

$$\mathbf{v}'_c \approx -\lambda (\mathbf{C}\widehat{\mathbf{L}}'_s)^{-1}\mathbf{e}_1$$

where $\widehat{\mathbf{L}}'_s$ is an approximation of the true (but unknown since $\mathbf{K}$ and $\boldsymbol{\zeta}$ are unknown) matrix $\mathbf{L}'_s$. In order to obtain a perfect decoupling in the ideal case when calibration is perfect, the combination matrix is chosen as $\mathbf{C} = \widehat{\mathbf{L}}'^{+}_s$. It can be easily verified that the control law is stable and decoupled in the ideal case when $\widehat{\mathbf{L}}'_s = \mathbf{L}'_s$. However, as it is shown by the experiments, the control law is stable even in the presence of calibration errors.

4.2 Control of the sixth 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:

$$\mathbf{T} = \mathbf{P}\mathbf{P}^{*-1} \tag{15}$$

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, from equation (12) we have:

$$\mathbf{P}\ \mathbf{\Gamma}\ \mathbf{P}^{*-1} = \mathbf{K}\ \mathbf{H}\ \mathbf{K}^{*-1} \tag{16}$$

If $\mathcal{F} = \mathcal{F}^*$ then $\mathbf{H} = \mathbf{I}$ and $\mathbf{\Gamma} = \mathbf{I}$, thus:

$$\mathbf{T} = \mathbf{P}\mathbf{P}^{*-1} = \mathbf{K}\mathbf{K}^{*-1} \tag{17}$$

and $\mathbf{T}$ must be upper triangular for any matrices $\mathbf{K}$ and $\mathbf{K}^*$. The constraints $t_{31} = 0$, $t_{32} = 0$ and $t_{33} = 1$ are always verified $\forall \mathbf{K}, \mathbf{K}^*$. Therefore, the constraint $t_{21} = 0$ is the only one we must impose in the general case. The remaining d.o.f. is controlled by a second task function:

$$e_2 = \det(\mathbf{P}^*)\ t_{21}$$

If $e_2 = 0$ then $t_{21} = 0$ and at the convergence $\mathcal{F} = \mathcal{F}^*$ (since $\mathbf{e}_1 \rightarrow 0$ thanks to the previous control law). Note that $\det(\mathbf{P}^*) \neq 0$ if the three reference points are not collinear. The derivative of e_2 can be written as:

$$\dot{e}_2 = \mathbf{l}_1 \mathbf{v}'_c + l_2 \omega_z$$

where $\mathbf{l}_1$ and l_2 can be obtained similarly to the interaction matrix. Note that from the previous control law $\mathbf{v}'_c = -\lambda_1 \mathbf{e}_1$, thus imposing the exponential convergence of t_{21} to zero (i.e. $\dot{e}_2 = -\lambda_2 e_2$) the control law for r_z is:

$$\omega_z = -(\lambda_2 e_2 - \lambda_1 \widehat{\mathbf{l}}_1 e_1(t))/\widehat{l}_2$$

where $\widehat{\mathbf{l}}_1$ and $\widehat{l}_2$ are approximations of $\mathbf{l}_1$ and l_2. One can easily verify that the control law is stable in the ideal case when $\widehat{\mathbf{l}}_1 = \mathbf{l}_1$ and $\widehat{l}_2 = l_2$. The experiments in the following section prove that the control law is stable even in the presence of calibration and measurement errors.

5. Experimental Results

5.1 Stationary zooming camera

To test whether the theory presented in the paper is a reasonable approximation it is important to prove the invariance of space $\mathcal{Q}$ to changes in camera intrinsic parameters. In this experiment, the focal length of a stationary camera changes six times approximatively from 2250 pixels to 1550 pixels (the corresponding six images are given in Figure 2). For each image, 16 points (corresponding to the 4 corners of the border of the 4 calibration grids) are extracted. The three points defining the reference plane are chosen spread in the image (points number 1, 8 and 11). The first image at the top to the left in Figure 2 is chosen as the reference image. Figure 3(a) shows that the error $\|\mathbf{p}_k - \mathbf{p}^*_k\|$ $(k = 1, 2, ..., 16)$ in the image space increases while the camera is zooming. On the other hand, Figure 3(b) shows that the error $\|\mathbf{q}_k - \mathbf{q}^*_k\|$ $(k = 1, 2, ..., 16)$ in the invariant space is not only close to zero but also practically constant. Obviously, the error is not exactly null because of noise in features extraction.

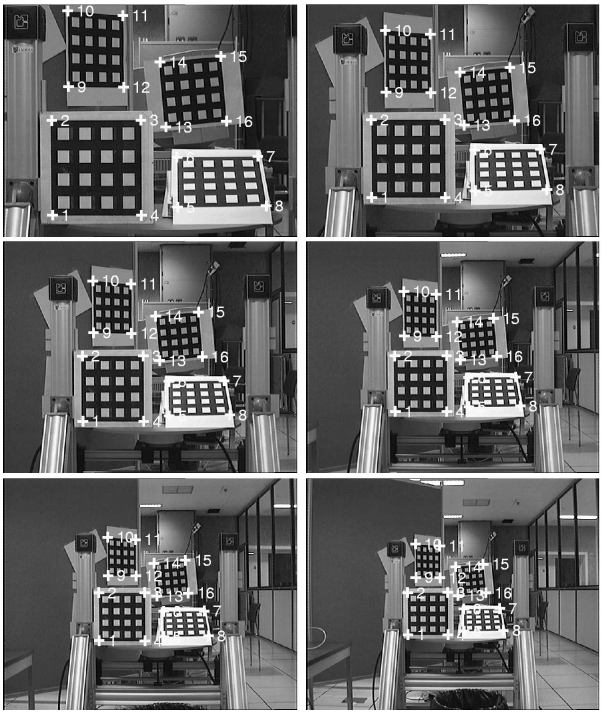
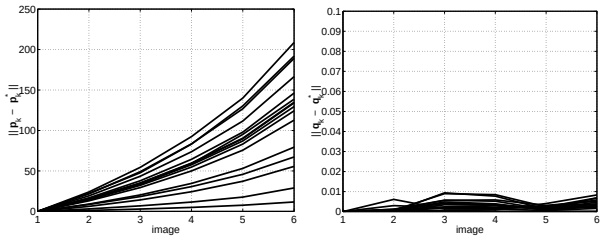


Figure 2. Sequence of six images taken with a stationary camera which is zooming out.

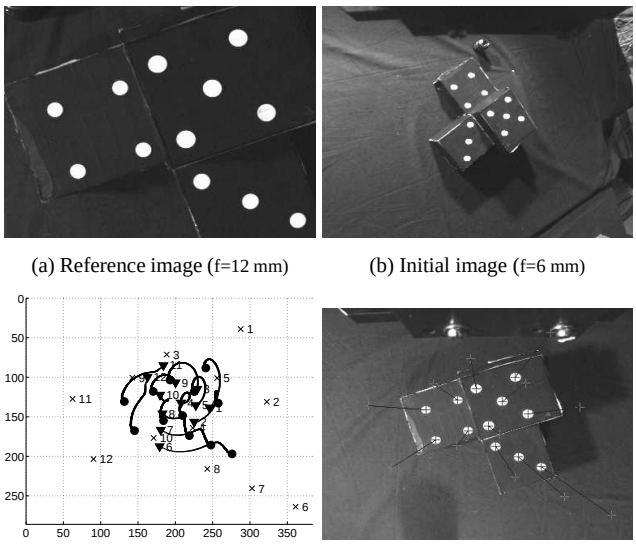


(a) Error $\|\mathbf{p}_k - \mathbf{p}_k^*\|$ (b) Error $\|\mathbf{q}_k - \mathbf{q}_k^*\|$

Figure 3. The error in the image space increases while the error in the invariant space is close to zero and practically constant.

5.2. Vision-based robot control

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 is mounted on the robot end-effector and observes a non-planar object composed by 12 points. A 12 mm lens is used for learning the reference image (see Figure 4(a)). Then, the robot is displaced to its starting position and the lens is replaced with a 6 mm lens (see Figure 4(b)). Finally, the robot is repositioned using the control law presented in the paper. Figure 4(c) shows the trajectory of the points in the image during the servoing. The initial image points are marked with a triangle. The final image points observed at the convergence are marked with a circle. The centroids extracted from the reference image are marked with a cross. Since a 6 mm lens is used for servoing, the final image observed by the camera at the convergence (Figure 4(d)) is different from the reference image.

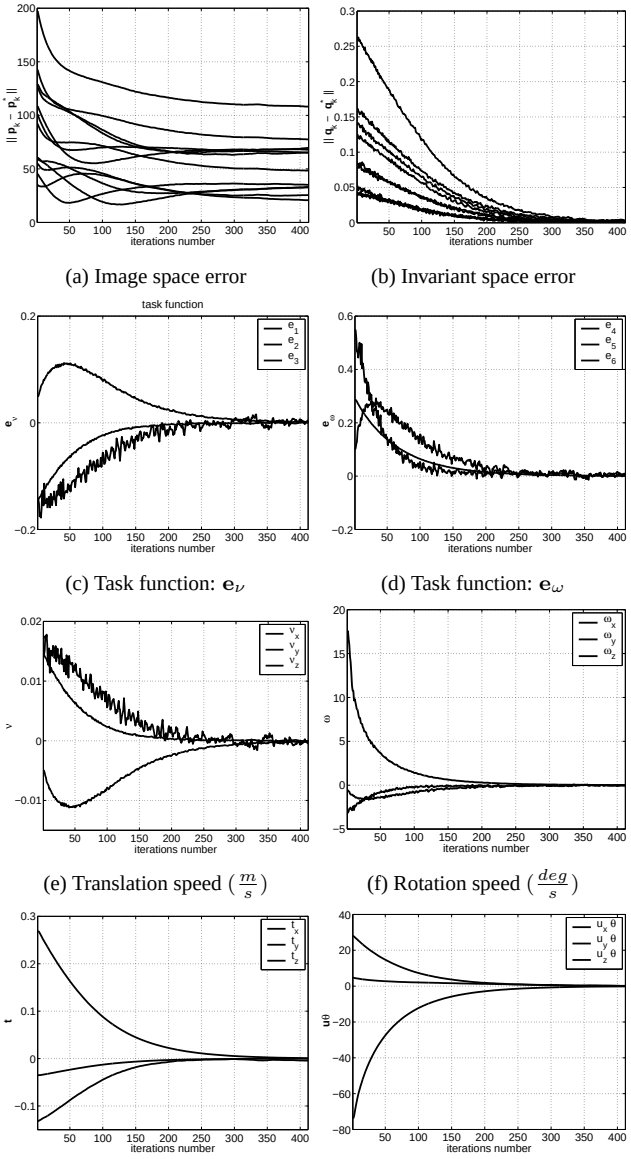


(a) Reference image (f=12 mm) (b) Initial image (f=6 mm)

(c) Points trajectory (d) Final image (f=6 mm)

Figure 4. Vision-based control using a 12mm lens for learning and a 6mm lens for servoing.

Thus, the error in the invariant space converges to zero (Figure 5(b)) but not the error in the image (Figure 5(a)). This experiment shows that the proposed visual servoing scheme is able to position a camera with respect to an object even if $\mathbf{K} \neq \mathbf{K}^*$ and $\mathbf{K}$ and $\boldsymbol{\zeta}$ are roughly known. An approximation $\widehat{\mathbf{K}}$ of the current camera parameters is used in the control law and the current depth $\widehat{\boldsymbol{\zeta}}$ is fixed to the depth $\boldsymbol{\zeta}^*$ measured in the reference frame. In spite of the noise and calibration errors, the task function converges to zero (Figures 5(c) and (d)) and the control law is stable (Figures 5(e) and (f)). Figures 5(g) and (h) plot respectively the error of translation and rotation between the current camera frame $\mathcal{F}$ and the reference frame $\mathcal{F}^*$. The initial displacement of the robot is quite large: $300\ mm$ for the translation and $80\ degrees$ for the rotation. At the convergence, the positioning precision is about $2\ mm$ for the translation and $0.2\ degrees$ for the rotation. The error is mainly on the translation along the z axis since a change of the focal length causes a small translation of the center of projection. Obviously, the positioning precision depends also on the level of noise in the image and on the distance of the camera from the object. A more robust control law can be obtained by discarding the points for which the residual error computed imposing the epipolar constraint is bigger than a threshold.



(a) Image space error (b) Invariant space error

(c) Task function: $\mathbf{e}_\nu$ (d) Task function: $\mathbf{e}_\omega$

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

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

Figure 5. The control law is stable despite noise and calibration errors. The precision is good even if camera parameters are changed.

6. Conclusions

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 (this can improve the robustness of features extraction from the images). Some control issues (for example the proof of the robustness of the control law) have been left unresolved since they are beyond the aim of this paper and they will be addressed in a further work.

Acknowledgements

I would like to thank Eric Marchand (VISTA project at IRISA) for his help in programming the experiments with the robot and François Gaspard (ROBOTVIS project at INRIA) for providing the images of the zooming camera.

References

[1] R. Basri, E. Rivlin, and I. Shimshoni. Visual homing: Surfing on the epipoles. In *IEEE Int. Conf. on Computer Vision*, pages 863–869, Bombay, India, January 1998.

[2] F. Chaumette. Potential problems of stability and convergence in image-based and position-based visual servoing. In D. Kriegman, G. Hager, and A. Morse, editors, *The confluence of vision and control*, volume 237 of *LNCIS Series*, pages 66–78. Springer Verlag, 1998.

[3] P. I. Corke. Visual control of robot manipulators - A review. In K. Hashimoto, editor, *Visual servoing*, volume 7 of *World Scientific Series in Robotics and Automated Systems*, pages 1–31. World Scientifi c Press, 1993.

[4] B. Espiau, F. Chaumette, and P. Rives. A new approach to visual servoing in robotics. *IEEE Trans. on Robotics and Automation*, 8(3):313–326, June 1992.

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

[6] K. Hashimoto. *Visual Servoing: Real Time Control of Robot manipulators based on visual sensory feedback*, volume 7 of *World Scientific Series in Robotics and Automated Systems*. World Scientifi c Press, Singapore, 1993.

[7] S. Hutchinson, G. D. Hager, and P. I. Corke. A tutorial on visual servo control. *IEEE Trans. on Robotics and Automation*, 12(5):651–670, October 1996.

[8] M. Irani, P. Anadan, and D. Weinshall. From reference frames to reference planes: multi-view parallax geometry and applications. In *European Conf. on Computer Vision*, volume 2, pages 829–845, June 1998.

[9] E. Malis, F. Chaumette, and S. Boudet. 2 1/2 d visual servoing with respect to unknown objects through a new estimation scheme of camera displacement. *International Journal of Computer Vision*, 37(1):79–97, June 2000.

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

[11] L. Van Gool, T. Moons, E. Pauwels, and A. Oosterlinck. Vision and lie's approach to invariance. *Image and vision computing*, 13(4):259–277, May 1995.

[12] W. J. Wilson, C. C. W. Hulls, and G. S. Bell. Relative end-effector control using cartesian position-based visual servoing. *IEEE Trans. on Robotics and Automation*, 12(5):684–696, October 1996.

[13] A. Zisserman, D. Forsyth, J. Mundy, C. Rothwell, J. Liu, and N. Pillow. 3D object recognition using invariance. *Artificial Intelligence*, 78:239–288, 1995.