Vision-based control using different cameras for learning the reference image and for servoing

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

Abstract

Most visual servoing schemes are based on the learning of a reference image with the same camera used for servoing. The scheme proposed in this paper differs from previous ones since it is independent on the camera used for learning. 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. This promising approach has been successfully tested with an eye-in-hand robotic system.

1 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 reposition an eye-in-hand system with respect to an observed object. Vision-based control has been widely investigated in the last few years [4, 5]. Despite the diversity of the approaches proposed to realise 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, a visual servoing scheme can reposition the camera with respect to the object [2, 9, 7]. Each scheme has its own advantages and drawbacks [1], 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. 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. 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 an error function, from only measured image features, which is invariant on the intrinsic parameters of the camera. Projective invariance has been used in [3] to define setpoints for stereo visual control that are independent on the viewing location. In this paper, projective invariance is used to define an error which is dependent on the viewing location but it is independent on the camera intrinsic parameters. The projective transformation proposed in [6] 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 [6]. The visual servoing scheme proposed in the paper has been completely validate by experiment.

2 Theoretical background

The vision-based control scheme presented in the paper is obtained from basic projective geometry. Let the point $\mathcal{C}^*$ be the center of projection (see Figure 1).

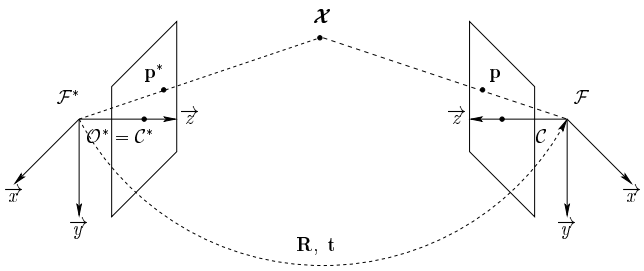


Figure 1: Perspective projections of the same 3D point $\boldsymbol{\mathcal{X}}$ from two different points of view $\mathcal{C}^*$ and $\mathcal{C}$.

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 $(\overrightarrow{x}, \overrightarrow{y})$. A 3D point, with homogeneous coordinates $\boldsymbol{\mathcal{X}} = \begin{bmatrix} X & Y & Z & 1 \end{bmatrix}^T$ is projected to the point $\mathbf{m}^*$: $\zeta^* \mathbf{m}^* = \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \boldsymbol{\mathcal{X}}$, where $\zeta^* = Z$. Let the point $\mathcal{C}$ be a different center of projection and the origin of frame $\mathcal{F}$ (see figure 1). The same 3D point projects to the point $\mathbf{m}$ in the frame $\mathcal{F}$: $\zeta \mathbf{m} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix} \boldsymbol{\mathcal{X}}$, where $\mathbf{R}$ and $\mathbf{t}$ are respectively the rotation and the translation between frames $\mathcal{F}^*$ and $\mathcal{F}$. Pinhole cameras perform a perspective projection of a 3D point. The information given by the pinhole cameras are two image points $\mathbf{p}$ and $\mathbf{p}^*$ (vectors $\mathbf{m}$ and $\mathbf{m}^*$ are not directly measured by the cameras). The point $\mathbf{p} = \begin{bmatrix} u & v & 1 \end{bmatrix}^T$ (where u and v are the image coordinates in pixels) observed in the image $\mathcal{I}$, taken at the position $\mathcal{F}$, depends on the camera internal parameters:

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

where:

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

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:

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

where $\mathbf{K}^*$ has the same form of $\mathbf{K}$ given in equation (2). 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. It is clear that both images depend on the intrinsic parameters of the cameras. Thus, even if $\mathcal{F} = \mathcal{F}^*$ the reference and the current image will be different if $\mathbf{K} \neq \mathbf{K}^*$. 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 $\mathbf{p} = \mathbf{p}^*$, $\forall\, \mathbf{p}, \mathbf{p}^*$ (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. This is a very strong constraint which limits the versatility of visual servoing. From both theoretical and practical point of views, it would be interesting to eliminate such constraint.

3 Invariance to camera parameters

In order to position the camera with respect to a non-planar 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. That is possible by using the simple projective transformation proposed in [6]. Consider three non-collinear 3D points of the observed object. 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:

$$\mathbf{P} = \mathbf{K}\mathbf{M} \tag{4}$$

$$\mathbf{P}^* = \mathbf{K}^* \mathbf{M}^* \tag{5}$$

where $\mathbf{P} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \end{bmatrix}$, $\mathbf{M} = \begin{bmatrix} \mathbf{m}_1 & \mathbf{m}_2 & \mathbf{m}_3 \end{bmatrix}$, $\mathbf{P}^* = \begin{bmatrix} \mathbf{p}_1^* & \mathbf{p}_2^* & \mathbf{p}_3^* \end{bmatrix}$ and $\mathbf{M}^* = \begin{bmatrix} \mathbf{m}_1^* & \mathbf{m}_2^* & \mathbf{m}_3^* \end{bmatrix}$. 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 (3×1) vectors $\mathbf{q}$ and $\mathbf{q}^*$ containing the projective coordinates of the points in the tranformed projective space are respectively:

$$\mathbf{q} = \mathbf{P}^{-1}\mathbf{p} \qquad \mathbf{q}^* \in \mathcal{Q}(\mathcal{F}) \tag{6}$$

$$\mathbf{q}^* = \mathbf{P}^{*-1}\mathbf{p}^* \qquad \mathbf{q}^* \in \mathcal{Q}(\mathcal{F}^*) \tag{7}$$

From now on we will refer to the transformed spaces $\mathcal{Q}(\mathcal{F})$ and $\mathcal{Q}(\mathcal{F}^*)$ as two "invariant" spaces where the

invariance is related to camera internal parameters. It is easy to show that the transformed vectors $\mathbf{q}$ and $\mathbf{q}^*$ do not depend on the internal parameter of the cameras. Indeed, from equations (4) and (5) we obtain:

$$\begin{aligned} \mathbf{q} &= \mathbf{P}^{-1}\,\mathbf{p} = \mathbf{M}^{-1}\,\mathbf{K}^{-1}\mathbf{K}\,\mathbf{m} = \mathbf{M}^{-1}\,\mathbf{m} \\ \mathbf{q}^* &= \mathbf{P}^{*-1}\mathbf{p}^* = \mathbf{M}^{*-1}\mathbf{K}^{*-1}\mathbf{K}^*\mathbf{m}^* = \mathbf{M}^{*-1}\mathbf{m}^* \end{aligned}$$

The camera parameters are factored out for each image. It must be emphasised that it is impossible to compute the camera parameters for each image since the 3D structure of the observed object is unknown. Camera self-calibration can be obtained from a sequence of several images but it is sensitive to noise and computationally expensive. In this paper, camera calibration is not necessary to obtain $\mathbf{q}$ and $\mathbf{q}^*$ since they are computed from image measurements and they only depend on the position of the camera with respect to the observed object and on its three-dimensional structure. However, $\mathbf{q}$ and $\mathbf{q}^*$ are also invariant on the rotation around the $\vec{z}$ axis. A more detailed description of the invariant space and its epipolar geometry is given in [6].

4 Control in the invariant space

The control the camera in the invariant space is divided into two parts. Since $\mathcal{Q}$ is invariant on the rotation around r_z, it can only be used to control five camera d.o.f. Another information must be extracted from the images in order 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 build the following task function [8]:

$$\mathbf{e}_1 = \mathbf{C}(\mathbf{s} - \mathbf{s}^*) \tag{8}$$

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 $(3(m-3) \times 5)$ matrix. Supposing that $\mathbf{s} - \mathbf{s}^*$ does not belongs to the null space of $\mathbf{C}$, if $\mathbf{e}_1 = 0$ then $\mathbf{s} - \mathbf{s}^* = 0$ and the camera frame $\mathcal{F}$ will coincide to $\mathcal{F}^*$ up to a rotation around the $\vec{z}$ 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 $\mathbf{v}_c = \begin{bmatrix} \nu_x & \nu_y & \nu_z & \omega_x & \omega_y & \omega_z \end{bmatrix}^T$:

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

The interaction matrix $\mathbf{L}_s$ can be built from the interaction matrix relative to each invariant point $\mathbf{q}_k$. The derivative of the point $\mathbf{q}_k = \begin{bmatrix} q_{1k} & q_{2k} & q_{3k} \end{bmatrix}^T$ is:

$$\dot{\mathbf{q}}_k = \mathbf{L}_{qk} \mathbf{v}_c$$

where $\mathbf{L}_{qk}$ is the interaction matrix relative to $\mathbf{q}_k$. Each point $\mathbf{q}_k$ gives a (3×6) interaction matrix $\mathbf{L}_{qk}$ 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 $m = 6$ points. Differentiating equation (6), we have:

$$\dot{\mathbf{q}}_k = \frac{d\mathbf{P}^{-1}}{dt}\mathbf{p}_k + \mathbf{P}^{-1}\dot{\mathbf{p}}_k = \frac{d\mathbf{P}^{-1}}{dt}\mathbf{P}\mathbf{q}_k + \mathbf{P}^{-1}\dot{\mathbf{p}}_k \tag{9}$$

Note that:

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

Plugging this equation in equation (9) we obtain:

$$\dot{\mathbf{q}}_k = \mathbf{P}^{-1}(\dot{\mathbf{p}}_k - \dot{\mathbf{P}}\mathbf{q}_k)$$

which can also be written as:

$$\dot{\mathbf{q}}_k = \mathbf{P}^{-1}(\dot{\mathbf{p}}_k - q_{xk}\dot{\mathbf{p}}_1 - q_{yk}\dot{\mathbf{p}}_2 - q_{zk}\dot{\mathbf{p}}_3)$$

The derivative of the current image point is:

$$\dot{\mathbf{p}}_k = \mathbf{L}_{pk}(\zeta_k, \mathbf{K}, \mathbf{p}_k)\mathbf{v_c}$$

where $\mathbf{L}_{pk}$ is obtained from the standard interaction matrix for an image point [2] adding a third row of zeros. Finally, we obtain:

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

The interaction matrix $\mathbf{L}_s$ is a $(3(m-3) \times 6)$ matrix:

$$\mathbf{L}_s = \begin{bmatrix} \mathbf{P}^{-1}(\mathbf{L}_{p4} - q_{14}\mathbf{L}_{p1} - q_{24}\mathbf{L}_{p2} - q_{34}\mathbf{L}_{p3}) \\ \mathbf{P}^{-1}(\mathbf{L}_{p5} - q_{15}\mathbf{L}_{p1} - q_{25}\mathbf{L}_{p2} - q_{35}\mathbf{L}_{p3}) \\ \vdots \\ \mathbf{P}^{-1}(\mathbf{L}_{pm} - q_{1m}\mathbf{L}_{p1} - q_{2m}\mathbf{L}_{p2} - q_{3m}\mathbf{L}_{p3}) \end{bmatrix}$$

It depends on $\mathbf{K}$ and $\boldsymbol{\zeta} = \begin{bmatrix} \zeta_1 & \zeta_2 & \dots & \zeta_m \end{bmatrix}^T$, 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 generally $\text{rank}(\mathbf{L}_s) = 5$ if the observed object is non-planar. Continuing research will provide more knowledge about eventual singularities of the interaction matrix (for example $\text{rank}(\mathbf{L}_s) = 2$

if the observed object is planar). Differentiating equation (8) 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 $\mathbf{v}_c' = \begin{bmatrix} \nu_x & \nu_y & \nu_z & \omega_x & \omega_y \end{bmatrix}^T$. For simplicity, I suppose here that $\dot{\mathbf{C}}(\mathbf{s}-\mathbf{s}^*) \approx 0$. This is true near the convergence since $(\mathbf{s}-\mathbf{s}^*) \approx 0$. A simple control law can be obtained by imposing the exponential convergence of the task function to zero (i.e. $\dot{\mathbf{e}}_1 = -\lambda_1\mathbf{e}_1$): $\mathbf{C}\mathbf{L}_s'\mathbf{v}_c' \approx -\lambda_1\mathbf{e}_1$. The control law is ($\lambda_1$ being a positive scalar):

$$\mathbf{v}_c' = -\lambda(\mathbf{C}\widehat{\mathbf{L}}_s')^{-1}\mathbf{e}_1 \tag{10}$$

where $\widehat{\mathbf{L}}_s'$ is an approximation of the true (but unknown since $\mathbf{K}$ and $\boldsymbol{\zeta}$ are unknown) matrix $\mathbf{L}_s'$. To obtain a perfect decoupling in the ideal case when calibration is perfect, the combination matrix is set to $\mathbf{C} = \widehat{\mathbf{L}}_s'^{+}$. 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, experiments show that the control law is stable even in the presence of calibration errors. The matrix $\mathbf{L}_s'$ must be computed at each iteration of the control law. Eventually, only $\mathbf{K}$ and $\boldsymbol{\zeta}$ can be chosen constant (for example in the experiments $\mathbf{K} = \widehat{\mathbf{K}}$ and $\widehat{\boldsymbol{\zeta}} = \widehat{\boldsymbol{\zeta}}^*$).

4.2 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 $\vec{z}$ axis. Let $\mathbf{T}$ be the following matrix:

$$\mathbf{T} = \mathbf{P}\mathbf{P}^{*-1} \tag{11}$$

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 $\mathcal{F} = \mathcal{F}^*$ then $\mathbf{M} = \mathbf{M}^*$ and from equations (5) and (4) we obtain:

$$\mathbf{T} = \mathbf{P}\mathbf{P}^{*-1} = \mathbf{K}\mathbf{K}^{*-1} \tag{12}$$

Thus, $\mathbf{T}$ must be upper triangular for any matrices $\mathbf{K}$ and $\mathbf{K}^*$ (i.e. $t_{21} = t_{31} = t_{32} = 0$ and $t_{33} = 1$). The constraints $t_{31} = 0$, $t_{32} = 0$ and $t_{33} = 1$ are always verified $\forall \mathbf{K}, \mathbf{K}^*$. Therefore, one must impose the constraints $t_{21} = 0$, $t_{11} > 0$ and $t_{22} > 0$ in order to have $r_z = 0$. Note that $t_{21} = 0$ if $r_z = \pm\frac{\pi}{2}$ or $r_z = \pm\pi$. However, in that cases $t_{11} < 0$ and/or $t_{22} < 0$. 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 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 $\mathbf{e}_1 \to 0$ thanks to the previous control law). The derivative of e_2 can be written as:

$$\dot{e}_2 = \mathrm{l}_1^T\mathbf{v}_c' + l_2\omega_z$$

where l_1 and l_2 are the interaction matrix . Note that if $-\frac{\pi}{2} < r_z < \frac{\pi}{2}$ then $l_2 \neq 0$. From the previous control law $\mathbf{v}_c' = -\lambda_1\mathbf{e}_1$, thus imposing the exponential convergence of e_2 to zero the control law for r_z is:

$$\omega_z = -(\lambda_2 e_2 - \lambda_1\widehat{\mathrm{l}}_1^T\mathbf{e}_1)/\widehat{l}_2 \tag{13}$$

where $\widehat{\mathrm{l}}_1$ and $\widehat{l}_2$ are approximations of l_1 and l_2. One can easily verify that the control law is stable in the ideal case when $\widehat{\mathrm{l}}_1 = \mathrm{l}_1$ and $\widehat{l}_2 = l_2$. However, as it is shown by the experiments, the control law is stable even in the presence of calibration errors.

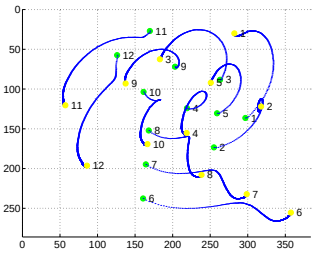
5 Experimental results

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 is mounted on the robot end-effector and observe an "object" composed by 12 points. In the experiments, an approximation $\widehat{\mathbf{K}}$ of the current camera parameters is used in the control law and the current depth $\boldsymbol{\zeta}$ is fixed to the depth $\boldsymbol{\zeta}^*$ measured in the reference frame.

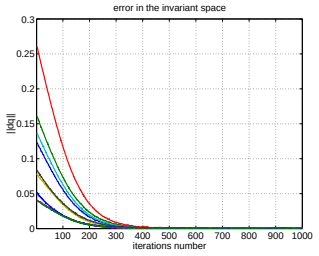
In the first experiment (Figure 2), the 12mm lens used for learning the reference image (see Figure 2(a)) is the same lens used for servoing. After the reference image has been learned the robot is displaced to its initial position (Figure 2(b) shows the initial image of the object). The initial displacement of the robot is quite large (see the initial values of the curves in Figures 2(i) and (j)). The control law computed from equations (10) and (13) is plotted in Figures 2(g) and (h). The task function converges to zero (see Figures 2(e) and (f)) as well as the error in the invariant space (see Figures 2(d)). Consequently, the position error of the robot end-effector converges to zero (see Figures 2(i) and (j)). Figure 2(c) plots the trajectory of the points in the image (the blue curves) and shows that the points go from their initial position (the green points) to their reference position (the yellow points). The first experiment proves that the new scheme can perform positioning tasks exactly as standard visual servoing techniques. The servoing is stopped when the error for each point is smaller than 0.5 pixels. In this experiment the positioning precision is less than 1 mm for the translation and 0.1 degrees for the rotation.

(a) Reference image

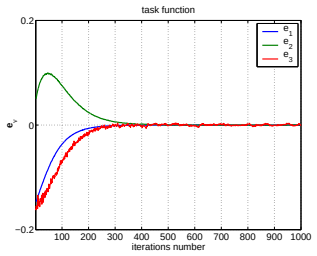
(b) Initial image



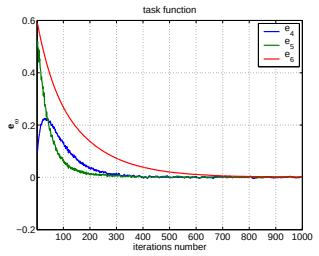
(c) Points trajectory



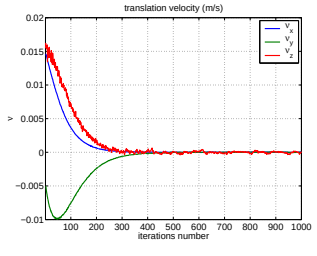
(d) Invariant space error



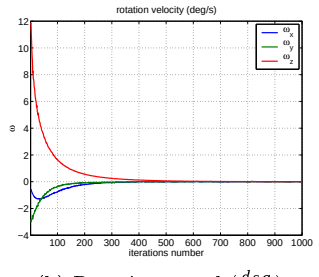
(e) Task function: $\mathbf{e}_\nu$



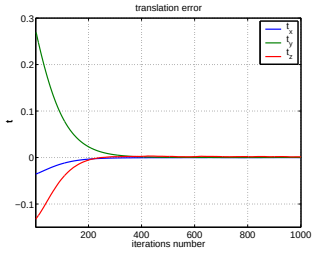
(f) Task function: $\mathbf{e}_\omega$



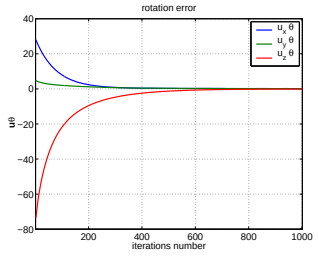
(g) Translation speed ($\frac{m}{s}$)



(h) Rotation speed ($\frac{deg}{s}$)



(i) Translation error (m)

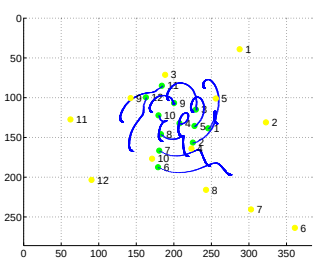


(j) Rotation error (deg)

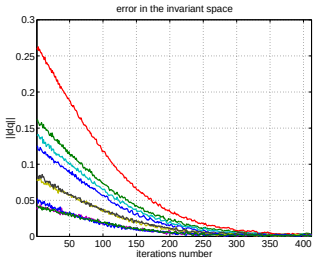
Figure 2: Experiment using the same 12mm lens for learning and servoing. The camera-independent vision-based control works as well as standard visual servoing.

(a) Reference image

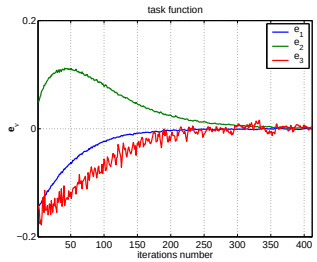
(b) Initial image



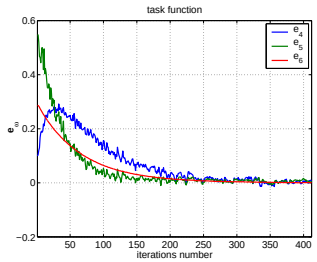
(c) Points trajectory



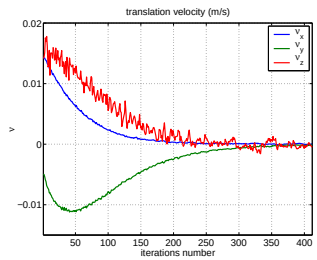
(d) Invariant space error



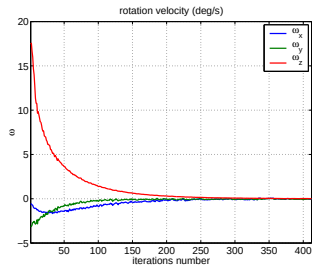
(e) Task function: $\mathbf{e}_\nu$



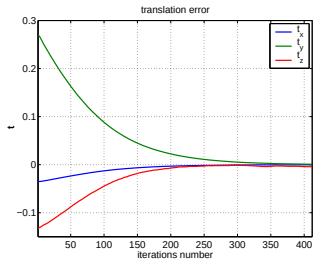
(f) Task function: $\mathbf{e}_\omega$



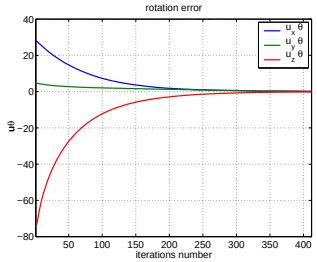
(g) Translation speed ($\frac{m}{s}$)



(h) Rotation speed ($\frac{deg}{s}$)



(i) Translation error (m)



(j) Rotation error (deg)

Figure 3: Experiment using a 12mm lens for learning and a 6mm lens for servoing. The visual servoing scheme is independent on the camera used for 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(i) and (j) and Figures 3(i) and (j)) but now the lens is changed. Since a 6mm lens is used for servoing 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 is independent on the camera intrinsic parameters and the initial and reference positions are the same in both experiments. Even if the task function has similar behaviours in Figures 2(e) and (f) and in Figures 3(e) and (f), the curves are slightly different since different approximations of the camera parameters were used to compute the combination matrix. 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(i) and (j) show that the error on the position of the end-effector converges to zero. With respect to the first experiment, the number of iterations is reduced by a factor of two. Indeed, since the error in the image is not meaningful, the servoing is now stopped when the error in the invariant space converges to zero. Figures 3(i) and (j) plot the control law sent to the robot. The control law is robust to noise and ensures again the convergence of the end-effector to the reference position. When compared to the previous experiment, the higher level of noise (especially on e_3, 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. The experiment was repeated several times with similar results. The error is mainly on the translation along the z axis since a change of the lens causes a small translation of the center of projection. Let us remark that switching the lens in the opposite manner (from a 6mm to a 12mm lens) slightly worsen the results because the reference image is at a lower resolution and some points are not visible in the final image. In that case, a zooming camera should be used in order to zoom out near the converge enlarging the camera field of view. Finally, similar results have been obtained changing the lens from 25mm to 12mm.

6 Conclusion

This paper shows how to position a camera, with respect to a non-planar object, even if a reference image is taken with a camera which is different from the camera used for servoing. The new visual servoing scheme will be useful especially when using a zooming camera (see the experimental result described in [6]). Indeed, 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). Future work will focus on analytical proof of the stability analysis of the control law when the calibration is not known exactly.

Acknowledgements

The experiments in the paper were carried out at IRISA. I would like to thank Eric Marchand (VISTA project) for his help in programming the experiments.

References

[1] 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.

[2] 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.

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

[4] K. Hashimoto. *Visual Servoing: Real Time Control of Robot manipulators based on visual sensory feedback*, volume 7 of *World Scient. Series in Rob. and Aut.Systems*. World Scientific Press, Singapore, 1993.

[5] 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.

[6] E. Malis. Visual servoing invariant to changes in camera intrinsic parameters. In *Int. Conf. on Comp. Vision*, vol. 1, pp. 704–709, Vancouver, BC, July 2001.

[7] E. Malis, F. Chaumette, and S. Boudet. 2 1/2 d visual servoing. *IEEE Trans. on Robotics and Automation*, 15(2):234–246, April 1999.

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

[9] 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.