Proceedings of ICAR 2003
The 11th International Conference on Advanced Robotics
Coimbra, Portugal, June 30 - July 3, 2003

Vision-based control with respect to planar and non-planar objects using a zooming camera

Selim Benhimane and Ezio Malis

I.N.R.I.A. Sophia Antipolis, France
{first_name}.{last_name}@sophia.inria.fr

Abstract

This paper concerns the visual servoing with a zooming camera. It presents a new visual servoing approach which allows to position a camera, equipped with a motorized zoom lens, with respect to an object independently if the object is planar or not. Indeed, planar objects are singular cases which cannot be handled by previous intrinsics-free visual servoing. The proposed method makes it possible to bring back a camera to its reference position while zooming. The focal length is initially controlled in order to keep the object in the field of view of the camera during the servoing. Then, the control of the focal length allows to solve the singularity for planar objects.

1 Introduction

In order to achieve robotic tasks autonomously, it is usually admitted that information provided by vision sensors is essential. Visual servoing is an efficient method to control robots in unknown and dynamic environments. Typically, a robotic task consists in positioning an eye-in-hand system with respect to an observed object. Many methods have been proposed in the last few years to perform this task [5] [7]. Among the different techniques developed, the most used ones are based on the "teaching-by-showing" approach. This means that the camera is shown the target view corresponding to a goal position of the robot and a reference image is then stored. Starting from another position, for which the target is in the field of view of the camera, the robot is controlled in order to reach the reference position. Obviously, this approach can be used only if the camera has kept, during the servoing, the same intrinsic parameters as in the learning step. Indeed, if these internal parameters change during the servoing or if the camera used during the servoing is different from the one used at the learning step, having an image that coincides with the reference image does not imply that the robot is brought back to the reference position. The constraining hypothesis of keeping invariable the camera intrinsic parameters (which include the focal length) limits the field of use of this visual servoing approach. Indeed, the zoom mechanism can be used to have better results in the visual control. Many studies [3] [1] proved that zooming while servoing can have many advantages since the precision of the positioning task and the accuracy of the extraction of the primitives are highly correlated with the resolution of the image. Standard vision-based control techniques that use the zoom mechanism during the servoing do not permit the positioning of the robot independently on the camera intrinsic parameters. Some of them need to control the size of the target in the image [1] or need partial information about the model of the target [12] in order to be able to control the focal length. Others use the equivalence between the camera positioning and the zoom setting (which can be verified under some conditions) in order to have at the convergence an image that coincides with the reference image without being back to reference position [6]. On the other hand, recent works [8] [9] have shown that 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. Unfortunately, the proposed methods do not work when the target is planar. The aim of this paper is to eliminate the hypothesis of non-planarity (which is very constraining) in order to increase the versatility and enlarge the domain of application of visual servoing techniques. First of all, we will discuss why planar objects are particular cases and why the method described in [9] does not work with such objects. Then, a visual servoing controller independent on the planarity of the target and using a zooming camera is proposed. Furthermore, we propose in this paper a simple focal length control strategy that allows to keep the target in the field of view of the camera during the servoing and recovers at the convergence the focal length value of the reference image without having any previous information about it. Experimental results using the new visual servoing scheme are presented to demonstrate the efficiency of the proposed method. The experiments are done using a 3 degrees of freedom eye-in-hand system with respect to a planar target and then to a non-planar target. Despite the contribution of this paper is mainly a theoretical development, the experiments prove that the method can be used in practical applications.

2 Theoretical Background

2.1 Camera model

In this paper, we suppose that the absolute frame coincides with the reference camera frame $\mathcal{F}^*$. A 3D point $\boldsymbol{\mathcal{X}}$ is projected on a virtual plane parallel to the plane $(\vec{x}, \vec{y})$ to a point $\mathbf{m}^*$:

$$\mathbf{m}^* = (x^*, y^*, 1) \propto \begin{bmatrix} \mathbf{I}_{3\times3} & \mathbf{0}_{3\times1} \end{bmatrix} \boldsymbol{\mathcal{X}} \tag{1}$$

Suppose now that the current camera frame $\mathcal{F}$ is in a different position. The 3D point $\boldsymbol{\mathcal{X}}$ is projected to a 2D point $\mathbf{m}$ in the current camera frame:

$$\mathbf{m} = (x, y, 1) \propto \begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix} \boldsymbol{\mathcal{X}} \tag{2}$$

where $\mathbf{R}$ is the rotation matrix and $\mathbf{t}$ the translation vector between $\mathcal{F}$ and $\mathcal{F}^*$. The information given by a pinhole camera (which performs a perspective projection of 3D points) is not directly $\mathbf{m}$ but an image point $\mathbf{p} = (u, v, 1)$:

$$\mathbf{p} = \mathbf{K}\mathbf{m} \tag{3}$$

where $\mathbf{K}$ is the camera internal parameter matrix:

$$\mathbf{K} = \begin{bmatrix} f & s & u_0 \\ 0 & rf & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

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

2.2 Invariance to camera parameters

The vectors $\mathbf{p}_i$ depend on the camera internal parameters. In order to control the robot regardless to the camera used during the visual servoing, we need information which is independent on these parameters. This is possible by using the projective transformation proposed in [9]. Suppose that an object composed by n points is observed by the camera. These points are projected using equation (2) to $\{\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_n\}$ in the current frame and the corresponding image points that can be obtained using equation (3) are $\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n\}$. We define then two matrices $\mathbf{S}_m$ and $\mathbf{S}_p$ as follows:

$$\mathbf{S}_m = \frac{1}{n}\sum_{i=1}^{n} \mathbf{m}_i \mathbf{m}_i^\top \quad \text{and} \quad \mathbf{S}_p = \frac{1}{n}\sum_{i=1}^{n} \mathbf{p}_i \mathbf{p}_i^\top$$

These positive symmetric matrices can be written using the Cholesky decomposition:

$$\mathbf{S}_m = \mathbf{T}_m \mathbf{T}_m^\top \quad \text{and} \quad \mathbf{S}_p = \mathbf{T}_p\, \mathbf{T}_p^\top$$

where $\mathbf{T}_m$ and $\mathbf{T}_p$ are upper triangular matrices. The two matrices are related by:

$$\mathbf{T}_p = \mathbf{K}\mathbf{T}_m \tag{5}$$

Thus, it is possible to compute the invariant vectors $\mathbf{q}_i$ with the following projective transformation:

$$\mathbf{q}_i = \mathbf{T}_p^{-1} \mathbf{p}_i \tag{6}$$

Note that the vectors $\mathbf{q}_i$ are computed only from the image coordinates. Using equations (5) and (6), we can write:

$$\mathbf{q}_i = \mathbf{T}_p^{-1}\mathbf{p}_i = \mathbf{T}_m^{-1}\mathbf{K}^{-1}\mathbf{K}\mathbf{m}_i = \mathbf{T}_m^{-1}\mathbf{m}_i \tag{7}$$

As a consequence, the vectors $\mathbf{q}_i$ are independent on the camera internal parameters. Similarly, from the reference points $\{\mathbf{p}_1^*, \mathbf{p}_2^*, \ldots, \mathbf{p}_n^*\}$, we can compute the reference invariants vectors $\mathbf{q}_i^*$. If $\mathbf{q} = \mathbf{q}^*$ then the current camera frame $\mathcal{F}$ coincide with the reference camera frame $\mathcal{F}^*$.

2.3 Intrinsics-free visual servoing

The aim of visual servoing is to control a robot using images taken by an eye-in-hand camera in order to bring the end-effector back to a reference position. In the case of the intrinsics-free visual servoing [9], this means that a vector $\mathbf{q}$, which contains the information of the current image, must converge to a vector $\mathbf{q}^*$, which contains the information of the reference image. In the vector $\mathbf{q}$, we have the coordinates of the invariant points: $\mathbf{q} = (\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n)$. If we differentiate the vector $\mathbf{q}$, we obtain:

$$\dot{\mathbf{q}} = \mathbf{L}_q \mathbf{v}_c \tag{8}$$

where $\mathbf{L}_q$ is the interaction matrix and $\mathbf{v}_c$ is the camera velocity. In order to make $\mathbf{q}$ converge to $\mathbf{q}^*$, we use the task function approach [10] which consists in minimizing an error vector $\mathbf{e}$:

$$\mathbf{e} = \widehat{\mathbf{L}}_q^+ (\mathbf{q} - \mathbf{q}^*) \tag{9}$$

where $\widehat{\mathbf{L}}_q^+$ is an estimation of the pseudo-inverse of $\mathbf{L}_q$. If we differentiate equation (9) and linearize around the equilibrium point $\mathbf{q} = \mathbf{q}^*$, we obtain:

$$\dot{\mathbf{e}} = \widehat{\mathbf{L}}_q^+ \mathbf{L}_q \mathbf{v}_c \tag{10}$$

We can impose an exponential decrease of the task function by choosing a control law as follows:

$$\mathbf{v}_c = -\lambda \mathbf{e} \tag{11}$$

where λ is a positive scalar. Consequently, and using equations (10) and (11), the closed-loop equation is:

$$\dot{\mathbf{e}} = -\lambda \widehat{\mathbf{L}}_q^{+} \mathbf{L}_q \mathbf{e}$$

It is well know from control theory [2] that if the matrix $\widehat{\mathbf{L}}_q^{+} \mathbf{L}_q$ is definite positive then the task function $\mathbf{e}$ converges to zero and so does the error $\mathbf{q}-\mathbf{q}^*$. According to [9], this visual servoing method is efficient when the visual servoing is done with respect to non-planar objects.

3 Problems with planar objects

Several problems have been observed with planar objects. In fact, the visual servoing does not converge to the reference position. We explain here the reasons why this happens and we will propose in the next section a solution for these problems. Suppose that all the 3D points $\boldsymbol{\mathcal{X}}_i$ are on a plane. We have:

$$\begin{pmatrix} \mathbf{n}^{*\top} & -d^* \end{pmatrix}^\top \boldsymbol{\mathcal{X}}_i = 0$$

where $\mathbf{n}^* = (a^*, b^*, c^*)$ is the normal vector to the plane that contains the target and d^* is the distance between the plane and the center of projection. Consequently, a vector $\mathbf{m}_i$ in the current frame is related to its homologous $\mathbf{m}_i^*$ in the reference frame by a homography $\mathbf{H}$:

$$\mathbf{m}_i \propto \mathbf{H}\mathbf{m}_i^* \quad (12)$$

where $\mathbf{H} = \mathbf{R} + \mathbf{t}\mathbf{n}^{*\top}/d^*$. In the image space, the current points are related to the reference points by a homography $\mathbf{G}$ such that:

$$\mathbf{p}_i \propto \mathbf{G}\mathbf{p}_i^* \quad (13)$$

Using equation (3), we have: $\mathbf{G} = \mathbf{K}\mathbf{H}\mathbf{K}^{*-1}$. This particular relationship is valid only if all the points are on the same plane.

3.1 Several solutions

One problem with planar targets is that it exists a position $\mathcal{F} \neq \mathcal{F}^*$ such that $\mathbf{q}_i = \mathbf{q}_i^*$. In fact, using equation (3), the points of the reference image can be written as follows: $\mathbf{p}_i^* = \mathbf{K}^*\mathbf{m}_i^*$. Let $\mathbf{p}_i$ be the points of the image of the object taken in a position $\mathcal{F}$. Using equation (13), we can write: $\mathbf{p}_i \propto \mathbf{G}\mathbf{K}^*\mathbf{m}_i^*$. If $\mathbf{G}$ is an upper triangular matrix, we have: $\mathbf{p}_i \propto \mathbf{K}'\mathbf{m}_i^*$ where $\mathbf{K}' = \mathbf{G}\mathbf{K}^*$ is also an upper triangular matrix. Since the vectors $\mathbf{q}_i$ are camera-independent (see equation (7)), we have $\mathbf{q}_i = \mathbf{q}_i^*$ despite $\mathcal{F} \neq \mathcal{F}^*$.

3.2 Particular cases

A solution for the previous problem is to make $\mathbf{G}$ converge to $\mathbf{I}_{3\times3}$. However, having $\mathbf{G} = \mathbf{I}_{3\times3}$ does not imply $\mathbf{H} = \mathbf{I}_{3\times3}$ and $\mathbf{K}\mathbf{K}^{*-1} = \mathbf{I}_{3\times3}$. In fact, it is possible to have : $\mathbf{H} = \mathbf{K}^{-1}\mathbf{K}^* \neq \mathbf{I}_{3\times3}$. There are two solutions $\{\mathbf{R}_1, \mathbf{t}_1, \mathbf{n}_1\}$ and $\{\mathbf{R}_2, \mathbf{t}_2, \mathbf{n}_2\}$ verifying:

$$\mathbf{K}^{-1}\mathbf{K}^* = \mathbf{R}_1 + \mathbf{t}_1\mathbf{n}_1^\top/d_1 = \mathbf{R}_2 + \mathbf{t}_2\mathbf{n}_2^\top/d_2$$

Consequently, the visual servoing converges to the reference position only if $\mathbf{n}_1 \neq \mathbf{n}^*$ and $\mathbf{n}_2 \neq \mathbf{n}^*$. If $s = s^* = 0$ and $r = r^*$ the vector $\mathbf{n}_1 = (0, 0, 1)$ is always a solution. This corresponds to the well known fact that if the plane is perpendicular to the optic axis of the camera in the reference position, zooming and translating in the $\vec{z}$ direction are equivalent. The visual servoing can not converge to the reference position in this case. In conclusion, it exists only one particular case that can be easily avoided by learning the reference image such that $\mathbf{n}^* \neq \mathbf{n}_1$ and $\mathbf{n}^* \neq \mathbf{n}_2$.

3.3 The interaction matrix is not full rank

Another problem with planar targets is that if all the points are on the same plane then the interaction matrix $\mathbf{L}_q$ in equation (8) is not a full rank matrix. In this case, even if the task function $\mathbf{e}$ is null, we are not sure to have $\mathbf{q} = \mathbf{q}^*$. It is possible to show that the rank of the interaction matrix $\mathbf{L}_q$ is equal to 3. In fact, the 1^{st} column is null, the 2^{nd} and the 3^{rd} column can be written as a linear combination of the 3 last columns. This matrix has the following form:

$$\mathbf{L}_q = \begin{bmatrix} 0 & \alpha\mathbf{c}_6 & -\alpha\mathbf{c}_5 + \beta\mathbf{c}_4 & \mathbf{c}_4 & \mathbf{c}_5 & \mathbf{c}_6 \end{bmatrix} \quad (14)$$

where $\mathbf{c}_4$, $\mathbf{c}_5$ and $\mathbf{c}_6$ are independent vectors, $\alpha = a^*/d^*$ and $\beta = b^*/d^*$. Consequently, 3 degrees of freedom remain undetermined. That is why, we need to define new constraints for the servoing.

4 A unified control for planar and non-planar targets

4.1 A Solution for the problems

In the previous section, we have shown that, in order to be able to use this intrinsics-free method, the camera parameters must converge to the values they have in the reference image. This means that we should have at the convergence: $\mathbf{K} = \mathbf{K}^*$. In general, we do not have any information about the reference parameters of the camera. However, we can use the matrix $\mathbf{T}_p$ in order to make the matrix $\mathbf{K}$ converge to $\mathbf{K}^*$. When the servoing converges to the reference position, we have $\mathbf{R} = \mathbf{I}_{3\times3}$ and $\mathbf{t} = \mathbf{0}_{3\times1}$. Thus, using equation (2), we obtain: $\mathbf{m}_i = \mathbf{m}_i^*$. Consequently, we have $\mathbf{T}_m = \mathbf{T}_m^*$. Using equation (5), we can write:

$$\mathbf{T}_p\mathbf{T}_p^{*-1} = \mathbf{K}\mathbf{K}^{*-1} \quad (15)$$

The matrix $\mathbf{T}_p$ is an upper triangular matrix that can be written under the following form:

$$\mathbf{T}_p = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ 0 & t_{22} & t_{23} \\ 0 & 0 & 1 \end{bmatrix}$$

We define a vector $\boldsymbol{\tau}$ that contains the entries of the matrix $\mathbf{T}_p$: $\boldsymbol{\tau} = (t_{11}, t_{12}, t_{13}, t_{22}, t_{23})$. Similarly, $\boldsymbol{\tau}^*$ contains the entries of the matrix $\mathbf{T}_p^*$. If $\boldsymbol{\tau} = \boldsymbol{\tau}^*$, then $\mathbf{T}_p = \mathbf{T}_p^*$ and using (15) we have $\mathbf{K} = \mathbf{K}^*$. As a conclusion, in order to bring back the camera to its reference position, $\mathbf{q}$ must converge to $\mathbf{q}^*$ and $\boldsymbol{\tau}$ must converge to $\boldsymbol{\tau}^*$.

4.2 The robot control

We need a single control that takes in consideration both cases: planar and non-planar target. The proposed control uses the approach of primary and secondary task [10]. The minimization of the task function $\mathbf{e}$ is considered as a primary task and, under the constraint of its realization, we will try to minimize the vector $\boldsymbol{\tau} - \boldsymbol{\tau}^*$. The new task function has the following form:

$$\widetilde{\mathbf{e}} = \widehat{\mathbf{L}}_q^+(\mathbf{q} - \mathbf{q}^*) + \lambda_t(\mathbf{I}_{6\times 6} - \widehat{\mathbf{L}}_q^+\widehat{\mathbf{L}}_q)\widehat{\mathbf{L}}_\tau^+(\boldsymbol{\tau} - \boldsymbol{\tau}^*) \quad (16)$$

where λ_t is a positive scalar and $\widehat{\mathbf{L}}_\tau^+$ is an estimation of the pseudo-inverse of the interaction matrix $\mathbf{L}_\tau$ related to the entries of the vector $\boldsymbol{\tau}$ that verifies:

$$\dot{\boldsymbol{\tau}} = \mathbf{L}_\tau \mathbf{v}_c \quad (17)$$

If the target is non-planar, then $\widehat{\mathbf{L}}_q$ is a full rank matrix. Thus, the matrix $\mathbf{I}_{6\times 6} - \widehat{\mathbf{L}}_q^+\widehat{\mathbf{L}}_q$ is null, $\widetilde{\mathbf{e}} = \mathbf{e}$ and the control law is exactly the same as in equation (11). On the contrary, the derivative of equation (16) can be written locally (i.e. $\mathbf{q} \approx \mathbf{q}^*$ and $\boldsymbol{\tau} \approx \boldsymbol{\tau}^*$) as:

$$\dot{\widetilde{\mathbf{e}}} = \widehat{\mathbf{L}}_q^+\dot{\mathbf{q}} + \lambda_t(\mathbf{I}_{6\times 6} - \widehat{\mathbf{L}}_q^+\widehat{\mathbf{L}}_q)\widehat{\mathbf{L}}_\tau^+\dot{\boldsymbol{\tau}} \quad (18)$$

If we replace $\dot{\mathbf{q}}$ and $\dot{\boldsymbol{\tau}}$ using respectively equation (8) and equation (17), we obtain:

$$\begin{aligned}\dot{\widetilde{\mathbf{e}}} &= \widehat{\mathbf{L}}_q^+\mathbf{L}_q\mathbf{v}_c + \lambda_t(\mathbf{I}_{6\times 6} - \widehat{\mathbf{L}}_q^+\widehat{\mathbf{L}}_q)\widehat{\mathbf{L}}_\tau^+\mathbf{L}_\tau\mathbf{v}_c \\ &= \mathbf{W}\mathbf{v}_c \quad (19)\end{aligned}$$

We can impose an exponential decrease of the task function $\widetilde{\mathbf{e}}$ with $\dot{\widetilde{\mathbf{e}}} = -\lambda\widetilde{\mathbf{e}}$ where λ is a positive scalar. The velocity of the camera $\mathbf{v}_c$ is then :

$$\mathbf{v}_c = -\lambda \mathbf{W}^{-1}\widetilde{\mathbf{e}} \quad (20)$$

4.3 A focal length control strategy

We control the focal length in order to have at the convergence the same value as we had in the reference image. When we differentiate (5), we obtain:

$$\dot{\mathbf{T}}_p = \dot{\mathbf{K}}\mathbf{T}_m + \mathbf{K}\dot{\mathbf{T}}_m$$

If we impose an exponential convergence of $\mathbf{T}_p$ to its reference value : $\dot{\mathbf{T}}_p = -\lambda_t^{'}(\mathbf{T}_p - \mathbf{T}_p{}^*)$, where $\lambda_t^{'}$ is a positive scalar, we obtain:

$$\begin{aligned}\dot{\mathbf{K}} &= (\dot{\mathbf{T}}_p - \mathbf{K}\dot{\mathbf{T}}_m)\mathbf{T}_m{}^{-1} \\ &= (-\lambda_t^{'}(\mathbf{T}_p - \mathbf{T}_p{}^*) - \mathbf{K}\dot{\mathbf{T}}_m)\mathbf{T}_p{}^{-1}\mathbf{K} \quad (21)\end{aligned}$$

where the matrix $\mathbf{K}\dot{\mathbf{T}}_m$ can be expressed using $\mathbf{v}_c$. Since $f = \mathbf{K}(1,1)$, the focal length control law is:

$$v_{f1} = \dot{\mathbf{K}}(1,1) \quad (22)$$

and $\dot{\mathbf{K}}(1,1)$ is computed from equation (21). Using this control law, the focal length converges to its reference value without having any information about it. We can also adopt a strategy for the control of the zoom that makes it possible to keep all the points in the field of view of the camera during the servoing. The zoom is controlled in order to keep the nearest point to the image borders in a given distance. The distance δ between that nearest point and the border is:

$$\delta = min_i(u_i, v_i, u_{max} - u_i, v_{max} - v_i)$$

where u_{max} and v_{max} are the image dimensions. We control the zoom in order to have $\dot{\delta} = 0$ when the distance δ is under a given value δ_{min}: $\frac{\partial \delta}{\partial t}|_{\delta<\delta_{min}} \approx 0$. Using equation (4), we have, if $\delta = u_i$ or $\delta = u_{max} - u_i$:

$$\frac{\partial \delta}{\partial t} = \dot{f}x_i + f\dot{x}_i + s\dot{y}_i = 0 \Longrightarrow v_{f2} = \frac{-f\dot{x}_i - s\dot{y}_i}{x_i} \quad (23)$$

if $\delta = v_i$ or $\delta = v_{max} - v_i$, we have :

$$\frac{\partial \delta}{\partial t} = r\dot{f}y_i + rf\dot{y}_i = 0 \Longrightarrow v_{f2} = \frac{-f\dot{y}_i}{y_i} \quad (24)$$

Note that $x_i \neq 0$ and $y_i \neq 0$ since the point is close to the image border. In order to have a focal velocity v_f verifying $v_f = v_{f1}$ when the points are far from the image border and $v_f = v_{f2}$ when one of them is close to a border, we use a weight function Φ continuous, differentiable and verifying:

$$\begin{cases} \lim_{\delta\to 0}\Phi(\delta) = 0 \\ \Phi(\delta) \approx 1 \quad \text{when} \quad \delta_{min} \leq \delta \end{cases}$$

If we impose to the focal length velocity $\mathbf{v}_f$:

$$\mathbf{v}_f = \Phi(\delta)\mathbf{v}_{f1} + (1 - \Phi(\delta))\mathbf{v}_{f2} \quad (25)$$

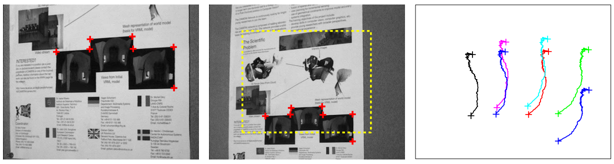
we obtain a control that makes it possible to recover the reference focal length value with keeping all the points in the field of view of the camera during the servoing.

5 Experimental results

The visual servoing scheme proposed in this paper has been tested on a 3 d.o.f. robot. It is a pan-tilt turret that can translate on a rail. The camera mounted on this turret has a motorized zoom. The features used here are corners obtained by the Harris detector [4] and tracked using the KLT algorithm [11]. In the first experiment, we show that the new visual servoing approach works even if the target is planar and the camera is zooming. In the second experiment, we show that the method works also when the target is not planar.

5.1 Planar target

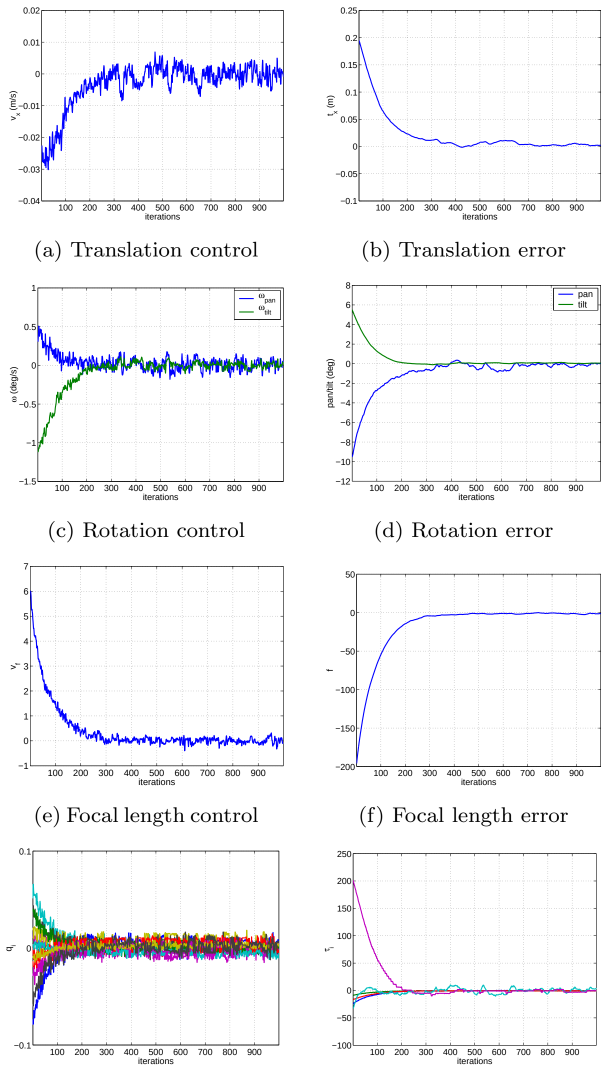
In the first experiment, the camera observes 7 coplanar points on a poster (see the red crosses in Figure 1).



(a) (b) (c)

Figure 1: *(a) Reference image, (b) Initial image, (c) Points trajectory*

After the reference image (Figure 1(a)) has been learned and the depth distribution of the points has been estimated, the robot is displaced to its initial position (Figure 1(b)) and the zoom is changed.



(a) Translation control (b) Translation error

(c) Rotation control (d) Rotation error

(e) Focal length control (f) Focal length error

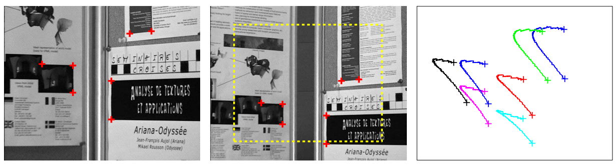
(g) Error $\mathbf{q} - \mathbf{q}^*$ (h) Error $\boldsymbol{\tau} - \boldsymbol{\tau}^*$

Figure 2: *Positioning with respect to a planar object.*

The initial displacement is approximately 20 *cm* for the translation, -10 *deg* for the pan and 6 *deg* for the tilt. With this initial displacement, if we do not zoom out, one point goes out of the field of view of the camera, as shown with the dashed rectangle in Figure 1(b). This is an illustrative case where a zoom is necessary to have the whole target in the image. The focal length changes then from 2758 *pixels* to 1671 *pixels*. Although a bad approximation of the camera intrinsic parameters ($\widehat{f} = 600$ *pixels*) is used, the control law in this case is stable (Figures 2(a), (c) and (e)) and the camera converges to the reference position (Figures 2(b), (d) and (f)). This experiment shows that the zoom control allows, firstly, to keep all the points in the field of view of the camera during the servoing (Figure 1(c)) and, secondly, f to converge to f^* (Figure 2(f)). The camera is back to its reference position when all the entries of the vector $\boldsymbol{\tau}$ (Figure 2(h)) converge to their reference values. The errors in the invariant space (Figure 2(g)) converge also to zero and the accuracy in the image is less than one pixel. This experiment shows that the visual servoing method described in this paper is efficient with planar targets contrarily to the method proposed in [9].

5.2 Non-planar target

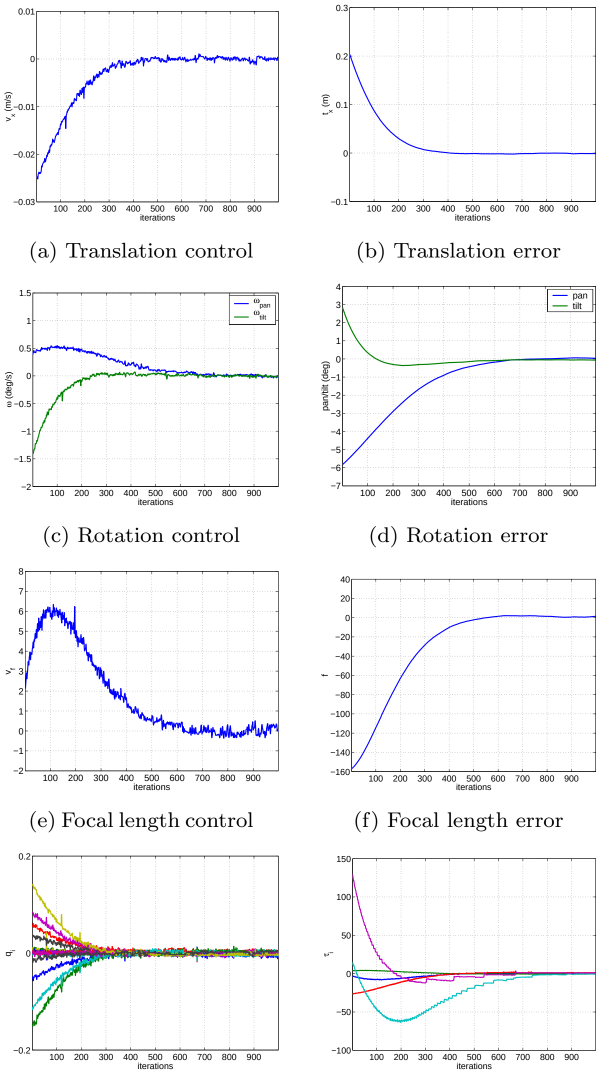
In the second experiment, the 7 points are on three non coplanar posters (see Figure 3(a)).



(a) (b) (c)

Figure 3: *(a) Reference image, (b) Initial image, (c) Points trajectory*

Again, the robot is displaced from its reference position (Figure 1(a)) to its initial position (Figure 1(b)). The displacement is approximately 20 *cm* for the translation, -6 *deg* for the pan and 3.5 *deg* for the tilt. One point goes out if we do not zoom out. That is why the focal length changes from 2758 *pixels* to 1913 *pixels* in order to have the whole target in the field of view of the camera. This experiment shows that the control law is stable (Figures 4(a), (c) and (e)) and the camera converges to the reference position (Figures 4(b), (d) and (f)). At the convergence, the error of translation is less than 1 *mm* and the error of rotation is less than 0.05 *deg*. The zoom allows to the camera to recover the reference value of the focal length (Figure 4(f)). Figure 3(c) plots the trajectory of the points in the image and shows that the points go from their initial position to the final position and remain in the field of view of the camera. The invariant error converges to zero for all the points (Figure 4(g)) and all the entries of the vector $\boldsymbol{\tau}$ converge to their reference value (Figure 4(h)). The new visual servoing scheme is then efficient with both planar and non-planar objects.



(a) Translation control (b) Translation error

(c) Rotation control (d) Rotation error

(e) Focal length control (f) Focal length error

(g) Error $\mathbf{q} - \mathbf{q}^*$ (h) Error $\boldsymbol{\tau} - \boldsymbol{\tau}^*$

Figure 4: *Positioning with respect to a non-planar object while the camera is zooming.*

6 Conclusion

In this paper, we have proposed an improvement of visual servoing with varying focal length. The proposed method makes it possible to position an eye-in-hand zooming camera with respect to a planar or a non-planar object given a reference image. At the beginning, the zoom is used to have all the points in the image. During the servoing, the points are kept in the field of view of the camera by controlling the focal length. At the convergence, the camera and the focal length are back to their reference position.

Acknowledgments

This work has been funded by the project AEROB of ROBEA: "Robotique et Entités Artificielles".

References

[1] B. Espiau. Asservissement visuel et commande de la distance focale. Technical Report 2613, INRIA, 1995.

[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, 1992.

[3] F. Gaspard. *Localisation et Calibration avec un Zoom.* PhD thesis, Nice Sophia Antipolis University, 2001.

[4] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, p. 147–151, 1988.

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

[6] K. Hosoda, H. Moriyama, and M. Asada. Visual servoing utilizing zoom mechanism. In *International Conference on Robotics and Automation*, p. 178–183, Japan, 1995.

[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, 1996.

[8] E. Malis. Visual servoing invariant to changes in camera intrinsic parameters. In *IEEE Int. Conf. on Computer Vision*, p. 704–709, Canada, 2001.

[9] E. Malis. A unified approach to model-based and model-free visual servoing. In *European Conf. on Computer Vision*, p. 433–447, Denmark, 2002.

[10] C. Samson, M. Le Borgne, and B. Espiau. *Robot Control: the Task Function Approach.* Oxford Engineering Science Series. Clarendon Press, UK, 1990.

[11] Jianbo Shi and Carlo Tomasi. Good features to track. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, p. 593–600, USA, 1994.

[12] M. Werman, S. Banerjee, S. Dutta Roy, and M. Qiu. Robot localization using uncalibrated camera invariants. In *IEEE International Conference on Computer Vision and Pattern Recognition*, p. 353–359, USA, 1999.