2-1/2-D Visual Servoing

Ezio Malis, François Chaumette, and Sylvie Boudet

***Abstract*—In this paper, we propose a new approach to vision-based robot control, called 2-1/2-D visual servoing, which avoids the respective drawbacks of classical position-based and image-based visual servoing. Contrary to the position-based visual servoing, our scheme does not need any geometric three-dimensional (3-D) model of the object. Furthermore and contrary to image-based visual servoing, our approach ensures the convergence of the control law in the whole task space. 2-1/2-D visual servoing is based on the estimation of the partial camera displacement from the current to the desired camera poses at each iteration of the control law. Visual features and data extracted from the partial displacement allow us to design a decoupled control law controlling the six camera d.o.f. The robustness of our visual servoing scheme with respect to camera calibration errors is also analyzed: the necessary and sufficient conditions for local asymptotic stability are easily obtained. Then, due to the simple structure of the system, sufficient conditions for global asymptotic stability are established. Finally, experimental results with an eye-in-hand robotic system confirm the improvement in the stability and convergence domain of the 2-1/2-D visual servoing with respect to classical position-based and image-based visual servoing.**

***Index Terms*— Eye-in-hand system, scaled Euclidean reconstruction, visual servoing.**

I. Introduction

VISION feedback control loops have been introduced in order to increase the flexibility and the accuracy of robot systems [12], [13]. Consider for example the classical positioning task of an eye-in-hand system with respect to a target. After the image corresponding to the desired camera position has been learned, and after the camera and/or the target has been moved, an error control vector can be extracted from the two views of the target. A zero error implies that the robot end-effector has reached its desired position with an accuracy regardless of calibration errors. However, these errors influence the way the system converges. In many cases, and especially when the initial camera position is far away from its desired one, the target may leave the camera field of view during servoing, which thus leads to failure. For this reason, it is important to study the robustness of visual servoing with respect to calibration errors.

Vision-based robot control using an eye-in-hand system is classified into two groups [12], [13], [19]: *position-based* and *image-based* control systems. In a *position-based* control system, the input is computed in the three-dimensional (3-D) Cartesian space [20] (for this reason, this approach can be called 3-D visual servoing). The pose of the target with respect to the camera is estimated from image features corresponding to the perspective projection of the target in the image. Numerous methods exist to recover the pose of an object (see [6] for example). They are all based on the knowledge of a perfect geometric model of the object and necessitate a calibrated camera to obtain unbiased results. Even if a closed loop control is used, which makes the convergence of the system possible in presence of calibration errors, it seems to be impossible to analyze the stability of the system. On the other hand, in an image-based control system, the input is computed in the 2-D image space (for this reason, this approach can be called 2-D visual servoing) [7]. In general, image-based visual servoing is known to be robust not only with respect to camera but also to robot calibration errors [8]. However, its convergence is theoretically ensured only in a region (quite difficult to determine analytically) around the desired position. Except in very simple cases, the analysis of the stability with respect to calibration errors seems to be impossible, since the system is coupled and nonlinear.

Contrary to the previous approaches, we will see that it is possible to obtain analytical results using a new approach which combines the advantages of 2-D and 3-D visual servoing and avoids their respective drawbacks. This new approach is called 2-1/2-D visual servoing since the used input is expressed in part in the 3-D Cartesian space and in part in the 2-D image space [14]. More precisely, it is based on the estimation of the camera displacement (the rotation and the scaled translation of the camera) between the current and desired views of an object. It must be emphasized that, contrary to the 3-D visual servoing, the partial camera displacement estimation does not need any 3-D model of the target, which increases the versatility and the application area of visual servoing. Since the camera rotation between the two views is computed at each iteration, the rotational control loop is immediately obtained. In order to control the translational camera d.o.f, we introduce *extended image coordinates* of a reference point of the target. We thus obtain a triangular interaction matrix with very satisfactory decoupling properties. It is interesting to note that this Jacobian matrix has no singularity in the whole task space. This allows us to obtain the convergence of the positioning task for any initial camera position if the camera intrinsic parameters are known. If the camera intrinsic parameters are not perfectly known, the estimated control vector can be analytically computed as a function of camera calibration errors. Then, the necessary and sufficient conditions

Manuscript received April 3, 1998; revised February 12, 1999. This paper was supported by INRIA and the National French Company of Electricity Power: EDF. This paper was recommended for publication by Associate Editor H. Zhuang and Editor V. Lumelsky upon evaluation of the reviewers' comments.

E. Malis was with IRISA/INRIA Rennes, Rennes cedex 35042, France. He is now with the University of Cambridge, Cambridge, U.K.

F. Chaumette is with IRISA/INRIA Rennes, Rennes cedex 35042, France.

S. Boudet is with DER-EDF, Chatou cedex 78401, France.

Publisher Item Identifier S 1042-296X(99)03918-X.

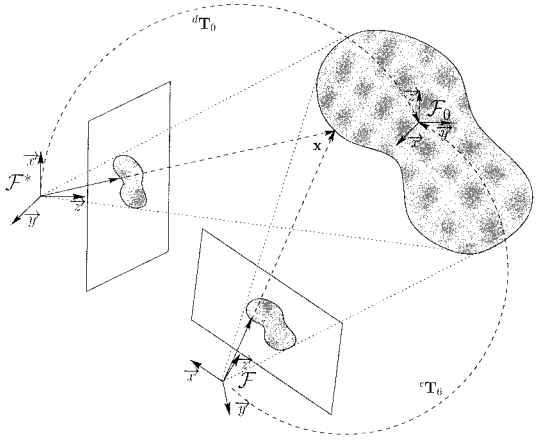


Fig. 1. Modeling of camera displacement for 3-D visual servoing.

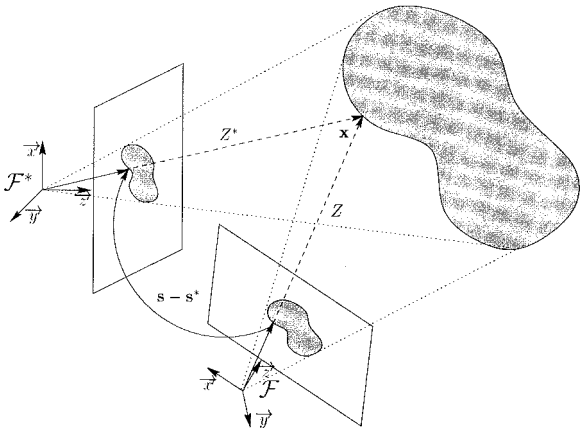


Fig. 2. Two-dimensional (2-D) visual servoing.

for the local asymptotic stability in the presence of camera calibration errors are easily obtained. Moreover, owing to the simple structure of the system, sufficient conditions for global asymptotic stability are presented. Using an adaptive control law, we can finally ensure that the target will always remain in the camera field of view. Experimental results confirm that 2-1/2-D visual servoing is more efficient than existing control schemes.

The paper is organized as follows: in Section II and Section III, we briefly recall 3-D and 2-D visual servoing respectively. In Section IV, we show how to use the information extracted from Euclidean partial reconstruction to design our 2-1/2-D visual servoing scheme. Its robustness with respect to camera calibration errors is analyzed in Section V. The experimental results are given in Section VI. A more robust adaptive control law is presented in Section VII and its robustness with respect to camera and hand-eye calibration errors is experimentally shown.

II. Three-Dimensional visual Servoing

Let $\mathcal{F}_0$ be the coordinate frame attached to the target, $\mathcal{F}^*$ and $\mathcal{F}$ be the coordinate frames attached to the camera in its desired and current position respectively (see Fig. 1).

Knowing the coordinates, expressed in $\mathcal{F}_0$, of at least four points of the target [6] (i.e. the 3-D model of the target is supposed to be perfectly known), it is possible from their projection to compute the desired camera pose and the current camera pose. The camera displacement to reach the desired position is thus easily obtained, and the control of the robot end-effector can be performed either in open loop or, more robustly, in closed-loop. The main advantage of this approach is that it directly controls the camera trajectory in Cartesian space. However, since there is no control in the image, the image features used in the pose estimation may leave the image (especially if the robot or the camera are coarsely calibrated), which thus leads to servoing failure. Also note that, if the camera is coarse calibrated, or if errors exist in the 3-D model of the target, the current and desired camera poses will not be accurately estimated. Finally, since the error made on the pose estimation cannot be computed analytically as a function of the camera calibration errors, it seems to be impossible to analyze the stability of the system [5].

III. Two-Dimensional Visual Servoing

The control error function is now expressed directly in the 2-D image space (see Fig. 2).

Let $\mathbf{s}$ be the current value of visual features observed by the camera and $\mathbf{s}^*$ be the desired value of $\mathbf{s}$ to be reached in the image. The time variation of $\mathbf{s}$ is related to camera velocity $\mathbf{v} = [\boldsymbol{v}^T \ \boldsymbol{\omega}^T]^T$ by [7]

$$\dot{\mathbf{s}} = \mathbf{L}(\mathbf{s}, Z)\mathbf{v} \tag{1}$$

where $\mathbf{L}(\mathbf{s}, Z)$ is the interaction matrix (also called the image Jacobian matrix) related to $\mathbf{s}$. Note that $\mathbf{L}$ depends on the depth Z of each selected feature.

The interaction matrix for a large range of image features can be found in [7]. The vision-based task $\mathbf{e}$ (to be regulated to 0), corresponding to the regulation of $\mathbf{s}$ to $\mathbf{s}^*$, is defined by

$$\mathbf{e} = \mathbf{C}(\mathbf{s} - \mathbf{s}^*) \tag{2}$$

where $\mathbf{C}$ is a matrix which has to be selected such that $\mathbf{CL}(\mathbf{s}, Z) > 0$ in order to ensure the global stability of the control law. The optimal choice is to consider $\mathbf{C}$ as the pseudo-inverse $\mathbf{L}(\mathbf{s}, Z)^+$ of the interaction matrix. The matrix $\mathbf{C}$ thus depends on the depth Z of each target point used in visual servoing. An estimation of the depth can be obtained using, as in 3-D visual servoing, a pose determination algorithm (if a 3-D target model is available), or using a structure from known motion algorithm (if the camera motion can be measured). However, using this choice for $\mathbf{C}$ may lead the system close to, or even reach, a singularity of the interaction matrix. Furthermore, the convergence may also not be attained due to local minima reached because of the computation by the control law of unrealizable motions in the image [5].

Another choice is to consider $\mathbf{C}$ as a constant matrix equal to $\mathbf{L}(\mathbf{s}^*, Z^*)^+$, the pseudo-inverse of the interaction matrix computed for $\mathbf{s} = \mathbf{s}^*$ and $Z = Z^*$, where Z^* is an approximate value of Z at the desired camera position. In this simple case, the condition for convergence is satisfied only in the neighborhood of the desired position, which means that the

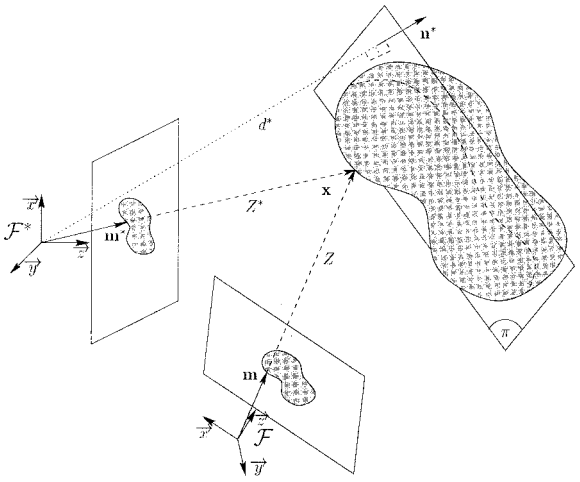


Fig. 3. The 2-1/2-D visual servoing.

convergence may not be ensured if the initial camera position is too far away from the desired one. Once again, several examples leading to failure are given in [5].

IV. 2-1/2-D Visual Servoing

The main drawback of 3-D visual servoing is that there is no control in the image which implies that the target may leave the camera field of view. Furthermore, a model of the target is needed to compute the pose of the camera. 2-D visual servoing does not explicitly need this model. However, a depth estimation or approximation is necessary in the design of the control law. Furthermore, the main drawback of this approach is that the convergence is ensured only in a neighborhood of the desired position (whose domain seems to be impossible to determine analytically). The objective of this section is to present a control scheme avoiding these drawbacks. We will see that a satisfactory solution is to combine image features and 3-D data. This 3-D information can be obtained by performing partial pose estimation as described in the next subsection.

A. Scaled Euclidean Reconstruction

Consider three 3-D target points $\mathcal{P}_i$ defining a reference plane π (see Fig. 3). It is well known that the resulting image points $\mathbf{p}_i$ in the current camera frame $\mathcal{F}$, are related to the corresponding image points $\mathbf{p}_i^*$ in the desired camera frame $\mathcal{F}^*$, by a projective homography $\mathbf{G}$ such that $\mathbf{p}_i = \mathbf{G}\mathbf{p}_i^*$ $(i = 1,2,3)$ [9]. On one hand, if the target is known to be planar, $\mathbf{G}$ can be estimated solving a linear system using at least 4 points of this plane. On the other hand, if the target is not planar, the estimation of $\mathbf{G}$ is a nonlinear problem. If at least 8 points are available (3 to define π and 5 not belonging to π), it is possible to estimate the homography matrix at video-rate using for example the linearized algorithm presented in [16]. Other classical, but less robust, linearized methods for computing the camera displacement through the recovery of the epipolar geometry can be found in [11] and [4]. Such classical methods have recently been used in vision-based control [10], [2]. In [10], a stereovision system is used. The work described in [2], closely related to ours, will be discussed afterward.

Let $\mathbf{A}$ be the intrinsic parameters matrix of the camera

$$\mathbf{A} = \begin{bmatrix} fk_u & fk_u\cot(\phi) & u_0 \\ 0 & \frac{fk_v}{\sin(\phi)} & v_0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \alpha_u & \alpha_{uv} & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

where u_0 and v_0 are the pixels coordinates of the principal point, k_u and k_v are the scaling factors along the $\vec{x}$ and $\vec{y}$ axes (in pixels/meters), ϕ is the angle between these axes and f is the focal length. $\mathbf{A}$ is the transformation matrix between the pixel coordinates $\mathbf{p}$ and the normalized coordinates $\mathbf{m}$ of an image point

$$\mathbf{p} = \mathbf{A}\mathbf{m}. \quad (4)$$

Assuming that the camera calibration is known (i.e., $\mathbf{A}$ is known), the Euclidean homography $\mathbf{H}$ is calculated as follows:

$$\mathbf{H} = \mathbf{A}^{-1}\mathbf{G}\mathbf{A}. \quad (5)$$

After $\mathbf{H}$ is computed, it can be decomposed as the sum of a rotation matrix and of a rank 1 matrix [9]

$$\mathbf{H} = \mathbf{R} + \mathbf{t}_{d^*}\mathbf{n}^{*T} \quad (6)$$

where $\mathbf{R}$ is the rotation matrix between frames $\mathcal{F}$ and $\mathcal{F}^*$ (i.e., the homography of the plane at infinity $\mathbf{R} = \mathbf{H}_\infty$), $\mathbf{n}^*$ is the unit vector normal to π expressed in $\mathcal{F}^*$, and $\mathbf{t}_{d^*}$ is defined as $\mathbf{t}/d^*$, $\mathbf{t}$ being the translation vector between $\mathcal{F}$ and $\mathcal{F}^*$, d^* being the distance of π to $\mathcal{F}^*$. From $\mathbf{H}$ and the image features, it is thus possible to determine the motion parameters and the structure of the reference plane. For example, the distances d and d^* are unknown (where d is the current distance between $\mathcal{F}$ and π), but the ratio $r = d/d^*$ can easily be estimated. Indeed, noting $\mathbf{n} = \mathbf{R}\mathbf{n}^*$ the vector normal to π, expressed in $\mathcal{F}$, we have [15]

$$r = \frac{d}{d^*} = 1 + \mathbf{n}^T\mathbf{t}_{d^*} = \det(\mathbf{H}). \quad (7)$$

Furthermore, ratio ρ_1 between the unknown depth Z of a point lying on π and d^* can be computed as

$$\rho_1 = \frac{Z}{d^*} = \frac{r}{\mathbf{n}^T\mathbf{m}}. \quad (8)$$

Finally, we can also obtain

$$\rho_2 = \frac{Z}{Z^*} = r\frac{\mathbf{n}^{*T}\mathbf{m}^*}{\mathbf{n}^T\mathbf{m}} = \rho_1\mathbf{n}^{*T}\mathbf{m}^*. \quad (9)$$

These parameters are important since they are used in the design of our control scheme. We will see in Section V how it is still possible to obtain an analytical form of the estimated motion parameters in the presence of camera calibration errors.

B. Control Scheme

In order to control the orientation of the camera, we naturally use the 3-D estimated rotation $\mathbf{R}$ between $\mathcal{F}$ and $\mathcal{F}^*$ (which has to reach the identity matrix). Let $\mathbf{u}$ be the rotation axis and θ the rotation angle obtained from $\mathbf{R}$. Instead of the classical control vectors $\mathbf{u}\sin(\theta)$ or $\mathbf{u}\sin(\frac{\theta}{2})$, the vector $\mathbf{u}\theta$ was chosen since, as shown below, this representation has no singularity in the whole workspace. The time derivative of $\mathbf{u}\theta$ can be expressed as a function of the camera velocity screw $\mathbf{v}$ as

$$\frac{d(\mathbf{u}\theta)}{dt} = [\mathbf{0} \quad \mathbf{L}_\omega]\mathbf{v} \tag{10}$$

where the Jacobian matrix $\mathbf{L}_\omega$ is given by [15]

$$\mathbf{L}_\omega(\mathbf{u},\theta) = \mathbf{I}_3 - \frac{\theta}{2}[\mathbf{u}]_\times + \left(1 - \frac{\operatorname{sinc}(\theta)}{\operatorname{sinc}^2\left(\frac{\theta}{2}\right)}\right)[\mathbf{u}]_\times^2 \tag{11}$$

with $\operatorname{sinc}(\theta) = \sin(\theta)/\theta$, $[\mathbf{u}]_\times$ being the antisymmetric matrix associated to vector $\mathbf{u}$. The determinant of $\mathbf{L}_\omega$ is

$$\det(\mathbf{L}_\omega) = 1 \Big/ \operatorname{sinc}^2\left(\frac{\theta}{2}\right) \tag{12}$$

and it is thus singular only for $\theta = 2k\pi$, $\forall k \in Z^*$ (i.e., out of the possible workspace). We can also note that $\mathbf{L}_\omega \approx \mathbf{I}_3$ for small values of θ.

We now design the position control vector. Consider a point $\mathcal{P}$ (called the reference point) lying on the chosen reference plane π of the target. The time derivative of its coordinates, expressed in the current camera frame, can be written as

$$\dot{\mathbf{x}} = [-\mathbf{I}_3 \quad [\mathbf{x}]_\times]\mathbf{v}. \tag{13}$$

Let us define the extended image coordinates $\mathbf{m}_e$ as follows:

$$\mathbf{m}_e = [x \quad y \quad z]^T = \left[\frac{X}{Z} \quad \frac{Y}{Z} \quad \log(Z)\right]^T \tag{14}$$

where $z = \log(Z)$ is a supplementary normalized coordinate. The time derivative of the extended image coordinates can be written as

$$\dot{\mathbf{m}}_e = \frac{1}{Z}\begin{bmatrix} 1 & 0 & -\frac{X}{Z} \\ 0 & 1 & -\frac{Y}{Z} \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} \dot{X} \\ \dot{Y} \\ \dot{Z} \end{bmatrix} = -\frac{1}{d^*}\mathbf{L}_v\dot{\mathbf{x}} \tag{15}$$

where d^* is unknown and $\mathbf{L}_v$ is an upper triangular matrix given by

$$\mathbf{L}_v = \frac{1}{\rho_1}\begin{bmatrix} -1 & 0 & x \\ 0 & -1 & y \\ 0 & 0 & -1 \end{bmatrix} \tag{16}$$

where ρ_1 is given by (8). Then, using (13) and (15), we finally obtain

$$\dot{\mathbf{m}}_e = \left[\frac{1}{d^*}\mathbf{L}_v \quad \mathbf{L}_{(v,\omega)}\right]\mathbf{v} \tag{17}$$

where $\mathbf{L}_{(v,w)}$ is the matrix

$$\mathbf{L}_{v,\omega} = \begin{bmatrix} xy & -(1+x^2) & y \\ (1+y^2) & -xy & -x \\ -y & x & 0 \end{bmatrix}. \tag{18}$$

The positioning task can be described as the regulation to zero of the following task function $\mathbf{e}$:

$$\mathbf{e} = [(\mathbf{m}_e - \mathbf{m}_e^*)^T \quad \theta\mathbf{u}^T]^T \tag{19}$$

where the first two components of $\mathbf{m}_e - \mathbf{m}_e^*$ are directly computed from the current and desired images, and its last component, equal to $\log(\rho_2)$, is estimated using (9). The time derivative of the task function is related to the camera velocity $\mathbf{v}$ by

$$\dot{\mathbf{e}} = \mathbf{L}\mathbf{v} \tag{20}$$

where $\mathbf{L}$ is an upper triangular matrix given by

$$\mathbf{L} = \begin{bmatrix} \frac{1}{d^*}\mathbf{L}_v & \mathbf{L}_{(v,\omega)} \\ \mathbf{0} & \mathbf{L}_\omega \end{bmatrix}. \tag{21}$$

This matrix is singular only if the camera optical center lies on the reference plane π (in that case, $d = r = \rho_1 = 0$). Other degenerate cases occur when $Z = 0$, $1/Z = 0$, $1/d = 0$ or, as already stated, when $\theta = \pm 2\pi$. The workspace free of singularity is thus composed of the area in front of π. In fact, if the target is known to be not planar, it is even possible to increase this workspace by changing the points used to define π at the nearing of a singularity.

Finally, the exponential convergence of $\mathbf{m}_e$ toward $\mathbf{m}_e^*$ and $\mathbf{u}\theta$ toward 0 can be obtained by imposing $\dot{\mathbf{e}} = -\lambda\mathbf{e}$ (where λ tunes the convergence rate). If the target is known to be motionless (see [1], [3], and [17] otherwise), the corresponding control law is given by [7]

$$\mathbf{v} = -\lambda\hat{\mathbf{L}}^{-1}\mathbf{e} \tag{22}$$

where $\mathbf{v}$ is the camera velocity sent to the robot controller, and $\hat{\mathbf{L}}^{-1}$ is an approximation of $\mathbf{L}^{-1}$. Since $\mathbf{L}$ depends on the unknown distance d^* (that is the only value which can not be measured or estimated in our method), an approximate value $\hat{d}^*$ has thus to be chosen during the off-line learning stage (when the desired image is acquired), and introduced in $\hat{\mathbf{L}}^{-1}$. However, $\hat{d}^*$ has not to be precisely determined (by hand in the following experiments) since it has a small influence on the stability of the system. More precisely, it influences the time-to-convergence of the translational velocity and the amplitude of the possible tracking error due to a wrong compensation of the rotational motion. As far as the tracking error is concerned, it is proportional to the rotational velocity and thus disappears when the camera is correctly oriented. We will see in the next section that it is possible to determine bounds on $\hat{d}^*/d^*$ such that the global stability of the system is ensured. Finally, the control law is given by

$$\mathbf{v} = -\lambda\begin{bmatrix} \hat{d}^*\mathbf{L}_v^{-1} & -\hat{d}^*\mathbf{L}_v^{-1}\mathbf{L}_{(v,\omega)} \\ \mathbf{0} & \mathbf{I}_3 \end{bmatrix}\begin{bmatrix} \mathbf{m}_e - \mathbf{m}_e^* \\ \mathbf{u}\theta \end{bmatrix}. \tag{23}$$

Indeed, due to the particular form of the matrix $\mathbf{L}_\omega^{-1}$, we can set $\mathbf{L}_\omega^{-1} = \mathbf{I}_3$ since [15]

$$\begin{aligned}\mathbf{L}_\omega^{-1}\mathbf{u}\theta &= \left(\mathbf{I}_3 + \frac{\theta}{2}\operatorname{sinc}^2\left(\frac{\theta}{2}\right)[\mathbf{u}]_\times + (1 - \operatorname{sinc}(\theta))[\mathbf{u}]_\times^2\right)\mathbf{u}\theta \\ &= \mathbf{u}\theta.\end{aligned} \tag{24}$$

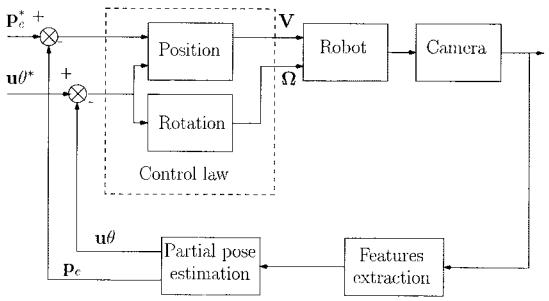


Fig. 4. Block diagram of the 2-1/2-D visual servoing.

The corresponding bloc-diagram is given in Fig. 4. Let us emphasize that $\hat{\mathbf{L}}^{-1}$ is an upper triangular square matrix without any singularity in the whole task space. Such a decoupled system also provides a satisfactory camera trajectory in the Cartesian space. Indeed, the rotational control loop is decoupled from the translational one, and the chosen reference point is controlled by the translational camera d.o.f. such that its trajectory is a straight line in the state space, and thus in the image. If a perfect model is available, the reference point will thus always remain in the camera field of view whatever the initial camera position. Of course, this property does not ensure that all the target points remain visible. However, in practice, it would be possible to change the chosen reference point during servoing, and we could select as reference point the target point nearest the bounds of the image plane. Indeed, it is possible to consider as reference point any point of the target (and not only points lying on π). In fact, for points lying outside π, the only difference in the previous computations is that the value of ρ_1 and ρ_2 are not given by (8) and (9) but have a different form (see [15] for the corresponding equations).

In practice, we have not considered the possibility of changing the reference point during the servoing since it would lead to a discontinuity in the translational components of the camera velocity at each change. Another strategy would be to select the reference point as the nearest of the center of gravity of the target in the image. This would increase the probability that the target remains in the camera field of view, but without any complete assurance. Therefore, we have preferred to use an adaptive control law, described in Section VII, to deal with this problem.

Finally, and contrary to 2-D and 3-D visual servoing, it will be shown in the following section that it is possible to obtain the necessary and sufficient conditions for local asymptotic stability, and sufficient conditions for global asymptotic stability in the presence of camera calibration errors.

Remark it is possible to design a control law directly in the Cartesian space (such that $\mathbf{t}_{d^*}$ has to reach 0, which thus implies achievement of the positioning task). A scheme very similar to classical 3-D visual servoing can hence be performed without knowing the 3-D structure of the target. In [2], such a scheme is used where the direction of translation is obtained from the essential matrix, instead of the homography matrix. However, as for 3-D visual servoing, such a control vector does not ensure that the considered object will always remain in the camera field of view, particularly in the presence of important camera or robot calibration errors.

It is also possible to control the camera position directly in the image space (as is done for 2-D visual servoing, the main difference being that the orientation is controlled using the result of the motion estimation). Contrary to 2-D visual servoing, in the present case, the decoupled control of the camera orientation allows the system to avoid local minima. However, the stability analysis is as difficult as for 2-D visual servoing. Furthermore, at least 2 image points are necessary, and the coupling of the related interaction matrix leads to an unpredictable camera trajectory. Experimental results show that, using this approach when the camera displacement is very important, the robot may unfortunately reach its joint limits, or the target may become so little in the image that visual servoing has to be stopped.

V. System Stability in Presence of Camera Calibration Errors

If the camera is not perfectly calibrated and $\hat{\mathbf{A}}$ is used instead of $\mathbf{A}$ [see (3)], the measured image point $\hat{\mathbf{p}}$ can be written as a function of the real image point $\mathbf{p}$ as

$$\hat{\mathbf{p}} = \hat{\mathbf{A}}\mathbf{m} = \delta\mathbf{A}\mathbf{p} \tag{25}$$

where $\delta\mathbf{A} = \hat{\mathbf{A}}^{-1}\mathbf{A}$. Furthermore, the estimated homography matrix is given by

$$\hat{\mathbf{H}} = \hat{\mathbf{A}}^{-1}\mathbf{G}\hat{\mathbf{A}} = \delta\mathbf{A}\mathbf{H}\delta\mathbf{A}^{-1}. \tag{26}$$

It can be decomposed as the sum of a matrix similar to a rotation matrix and of a rank 1 matrix

$$\hat{\mathbf{H}} = \hat{\mathbf{H}}_\infty + \hat{\mathbf{t}}_{d^*}\hat{\mathbf{n}}^{*T} \tag{27}$$

where $\hat{\mathbf{H}}_\infty = \delta\mathbf{A}\mathbf{R}\delta\mathbf{A}^{-1}$, $\hat{\mathbf{t}}_{d^*} = \|\mathbf{n}^{*T}\delta\mathbf{A}^{-1}\|\delta\mathbf{A}\mathbf{t}_{d^*}$ and $\hat{\mathbf{n}}^{*T} = \frac{\mathbf{n}^{*T}\delta\mathbf{A}^{-1}}{\|\mathbf{n}^{*T}\delta\mathbf{A}^{-1}\|}$ [15]. The eigenvalues of $\mathbf{R}$ depend on the angle of rotation θ, and its eigenvector corresponding to the unit eigenvalue is the axis of rotation $\mathbf{u}$. Matrix $\hat{\mathbf{H}}_\infty$ is not a rotation matrix, but is similar to $\mathbf{R}$, which implies that the two matrices have the same eigenvalues and the eigenvectors of $\hat{\mathbf{H}}_\infty$ are the eigenvectors of $\mathbf{R}$ multiplied by matrix $\delta\mathbf{A}$. The estimated rotation angle $\hat{\theta}$ and the estimated rotation axis $\hat{\mathbf{u}}$, extracted directly from $\hat{\mathbf{H}}_\infty$, can thus be written as a function of the real parameters and of the calibration errors

$$\hat{\theta} = \theta \quad \text{and} \quad \hat{\mathbf{u}} = \frac{\delta\mathbf{A}\mathbf{u}}{\|\delta\mathbf{A}\mathbf{u}\|}. \tag{28}$$

It must be emphasized that, as well as the rotation angle θ, the ratios r and ρ_2 are computed without error

$$\hat{r} = \det(\hat{\mathbf{H}}) = \det(\mathbf{H}) = r, \quad \hat{\rho_2} = \frac{Z}{Z^*} = \rho_2. \tag{29}$$

Finally, since $\hat{\mathbf{n}} = \hat{\mathbf{R}}^{-T}\hat{\mathbf{n}}^* = \frac{\delta\mathbf{A}^{-T}\mathbf{n}}{\|\delta\mathbf{A}^{-T}\mathbf{n}^*\|}$, $\hat{\rho_1}$ is given by

$$\hat{\rho_1} = \frac{\hat{r}}{\hat{\mathbf{n}}^T\hat{\mathbf{p}}} = \frac{r}{\hat{\mathbf{n}}^T\hat{\mathbf{p}}} = \|\delta\mathbf{A}^{-T}\mathbf{n}^*\|\rho_1. \tag{30}$$

The task function can thus be reconstructed as

$$\hat{\mathbf{e}} = \begin{bmatrix} \mathbf{E}_v & 0 \\ 0 & \mathbf{E}_\omega \end{bmatrix}\mathbf{e} \tag{31}$$

with

$$\mathbf{E}_v = \begin{bmatrix} \delta\mathbf{A}_{11} & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{E}_\omega = \mu\delta\mathbf{A} = \mu\begin{bmatrix} \delta\mathbf{A}_{11} & \delta\mathbf{p}_0 \\ 0 & 1 \end{bmatrix}$$

where $\delta\mathbf{A}_{11}$ is the (2×2) sub-matrix of $\delta\mathbf{A}$ containing the error on the pixel lengths [see (3)], $\delta\mathbf{p}_0$ is the (2×1) sub-vector containing the error on the principal point and $\mu = \frac{1}{\|\delta\mathbf{A}\mathbf{u}\|}$. It is interesting to note that we have $\hat{\mathbf{e}} = 0$ if and only if $\mathbf{e} = \mathbf{0}$, since the estimated task function $\hat{\mathbf{e}}$ depends linearly of the real (but unknown) one $\mathbf{e}$. Of course, we here assume that the intrinsic camera parameters do not change after the off-line acquisition of the desired image. In other words, since $\hat{\mathbf{e}}$ is measured from the current and desired images, we have $\hat{\mathbf{e}} = 0$ if and only if the target points have reached their desired position in the image, the function computing $\hat{\mathbf{e}}$ from the image points being bijective. The closed-loop system taking into account the camera calibration errors can thus be written

$$\dot{\mathbf{e}} = \mathbf{f}(\mathbf{e}) = -\lambda\mathbf{Q}(\mathbf{e})\mathbf{e} = -\lambda\mathbf{L}\hat{\mathbf{L}}^{-1}\mathbf{E}\mathbf{e}. \tag{32}$$

Function $\mathbf{f}$ is a C^∞ vector field defined on an open subset S of SE_3. It is easy to show the existence and uniqueness of the equilibrium point:

Proposition 1: The only point of equilibrium for $\mathbf{f}$, i.e., a point $\mathbf{e}^\circ \in S$ such that $\mathbf{f}(\mathbf{e}^\circ) = 0$, is $\mathbf{e}^\circ = 0$.

Proof: The existence of the equilibrium point is evident since if $\mathbf{e} = 0$, then $\mathbf{f}(\mathbf{e}) = \mathbf{Q}(\mathbf{e})\mathbf{e} = 0$. This equilibrium point is unique if and only if $\det(\mathbf{Q}(\mathbf{e})) \neq 0$, $\forall \mathbf{e} \in S$. Since matrix $\mathbf{Q}$ is upper triangular, its determinant can be easily calculated

$$\begin{aligned}\det(\mathbf{Q}) &= \det(\mathbf{L})\det(\hat{\mathbf{L}}^{-1})\det(\mathbf{E}) \\ &= \nu^3\mu^3\frac{\alpha_u^2}{\hat{\alpha}_u^2}\frac{\alpha_v^2}{\hat{\alpha}_v^2}\frac{1}{\operatorname{sinc}^2\left(\frac{\theta}{2}\right)}\end{aligned} \tag{33}$$

where $\nu = \frac{\hat{Z}}{Z} = \frac{\hat{d}^*}{d^*}\|\delta\mathbf{A}^{-T}\mathbf{n}^*\|$. We thus have $\det(\mathbf{Q}) \neq 0$, $\forall \mathbf{e} \in S$ since $\nu \neq 0$ and $\mu \neq 0$ in the non degenerate cases (such that $Z = 0$ or $d = 0$ for example).

Therefore, there is no singularity in the workspace and, if the task function decreases, it decreases toward 0. We now give the conditions to ensure the local asymptotic stability of the system.

Theorem 1 (Local Asymptotic Stability): The differential system (32) is locally asymptotically stable around the equilibrium point $\mathbf{e}^\circ$ if and only if

$$\frac{\hat{\alpha}_u}{\alpha_u} > 0, \quad \frac{\hat{\alpha}_v}{\alpha_v} > 0 \quad \text{and} \quad \frac{\hat{d}^*}{d^*} > 0. \tag{34}$$

Let us note that these conditions are also necessary and sufficient for global asymptotic stability of the orientation control subsystem. The proof of this theorem is given in [15]. In practice, these conditions are of course easily verified. However, it is well known that the error may considerably increase before convergence toward zero. Therefore, it is more interesting to find sufficient conditions to ensure the decreasing of $\|\mathbf{e}\|$ at each iteration, which ensures the global asymptotic stability of the system.

It is well known that such a sufficient condition is the positiveness of matrix $\mathbf{Q}(\mathbf{e}), \forall \mathbf{e}$. Owing to the particular form of $\mathbf{Q}(\mathbf{e})$ (upper triangular matrix), we have been able to obtain the following result:

Theorem 2 (Global Asymptotic Stability): The differential system (32) is globally asymptotically stable, only if conditions (34) are verified, and if

$$\sigma = \frac{\alpha_u}{\hat{\alpha}_u} + \frac{\alpha_v}{\hat{\alpha}_v} - \sqrt{\left(\frac{\alpha_u}{\hat{\alpha}_u} - \frac{\alpha_v}{\hat{\alpha}_v}\right)^2 + \left(\frac{\alpha_{uv}}{\hat{\alpha}_u} - \frac{\hat{\alpha}_{uv}}{\hat{\alpha}_u}\frac{\alpha_v}{\hat{\alpha}_v}\right)^2} > 0 \tag{35}$$

$$\sigma_1 = \sigma + 1 - \sqrt{(\sigma-1)^2 + \|\mathbf{I}_3 - \delta\mathbf{A}\|^2(1+\bar{\gamma}^2)} > 0 \tag{36}$$

$$\sigma_2 = \sigma + 1 - \sqrt{(\sigma-1)^2 + \|\delta\mathbf{p}_0\|^2} > 0 \tag{37}$$

$$\mu(\nu\|\mathbf{I}_3 - \delta\mathbf{A}\| + |1-\nu|)^2\|\delta\mathbf{A}\|^2 g^2(\bar{\gamma}) < \nu\sigma_1\sigma_2 \tag{38}$$

where $\bar{\gamma} = \tan(\psi)$ is the tangent of the vision angle ψ ($\bar{\gamma}$ is the maximum value of $\gamma = \sqrt{x^2+y^2}$) and

$$g^2(\bar{\gamma}) = \frac{2 + \bar{\gamma}^2 + \sqrt{\bar{\gamma}^2(\bar{\gamma}^2+4)}}{2}(1+\bar{\gamma}^2). \tag{39}$$

The proof is given in [15]. Let us note that these sufficient conditions ensure the decreasing of $\|\mathbf{e}\|$ at each iteration, but do not ensure that the target will remain in the camera field of view, since such a constraint is not taken into account in the positiveness of matrix $\mathbf{Q}(\mathbf{e})$. In fact, we can even not be sure that the reference point will always be visible, since the decrease of $\|\mathbf{e}\|$ does not necessarily imply the decrease of $\|\mathbf{m}_e - \mathbf{m}_e^*\|$. As already stated, the visibility constraint will be taken into account in Section VII. The global asymptotic stability is however interesting since it implies that $\|\mathbf{e}\|$, and hence $\|\mathbf{v}\|$, will not have an oscillatory behavior. To our knowledge, theorems 1 and 2 are the first analytical results of the stability of any visual servoing scheme able to control the six camera d.o.f. in presence of camera calibration errors.

Conditions (35–38) can be analyzed in order to have an approximate idea of the convergence domain. However, they depend on $\delta\mathbf{A}$ and μ (which depends on the rotation axis $\mathbf{u}$, and thus on the current camera position). In order to simplify the analysis, we now present the following corollary whose proof is given in [15].

Corollary 1 (Global Asymptotic Stability): The differential system (32) is globally asymptotically stable, only if conditions (34) are verified, and if

$$\sigma > 0 \tag{40}$$

$$\sigma_3 = \sigma + 1 - \sqrt{(\sigma-1)^2 + k_1^2(1+\bar{\gamma}^2)} > 0 \tag{41}$$

$$\frac{\sqrt{2}(\nu k_1 + |1-\nu|)^2 k_2^2 g^2(\bar{\gamma})}{\sqrt{\sigma' + 1 - \sqrt{(\sigma'-1)^2 + 4\|\delta\mathbf{A}_{11}\|^2\|\delta\mathbf{p}_0\|^2}}} < \nu\sigma_2\sigma_3 \tag{42}$$

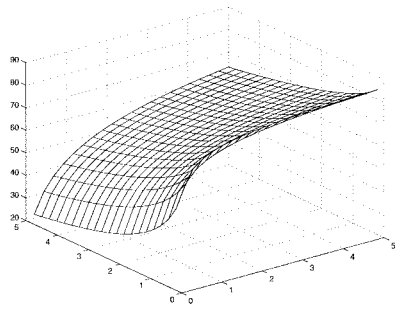


Fig. 5. Stability bounds for $|\delta\phi|$.

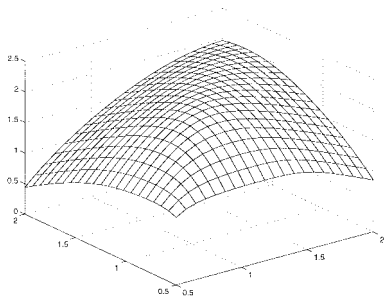


Fig. 6. Stability bounds for $\|\delta\mathbf{p}_0\|$.

where

$$\sigma' = \frac{1}{2}\left(k_3 + k_4 - \sqrt{(k_3+k_4)^2 - 4\frac{\alpha_u^2}{\hat{\alpha}_u^2}\frac{\alpha_v^2}{\hat{\alpha}_v^2}}\right)$$
$$k_1 = \|\mathbf{I_2} - \delta\mathbf{A}_{11}\| + \|\delta\mathbf{p}_0\|$$
$$k_2 = \|\delta\mathbf{A}_{11}\| + \sqrt{1+\|\delta\mathbf{p}_0\|^2}$$
$$k_3 = \frac{\alpha_u^2}{\hat{\alpha}_u^2} + \frac{\alpha_v^2}{\hat{\alpha}_v^2}$$
$$k_4 = \left(\frac{\alpha_{uv}}{\hat{\alpha}_u} - \frac{\hat{\alpha}_{uv}}{\hat{\alpha}_u}\frac{\alpha_v}{\hat{\alpha}_v}\right)^2.$$

These sufficient conditions are more restrictive than the conditions imposed by Theorem 2. However, these new conditions do not depend on the axis of rotation $\mathbf{u}$, and the error $\delta\mathbf{p}_0$ of the principal point is separated from the error $\delta\mathbf{A}_{11}$ of the pixel lengths. The analysis is thus simplified even if the conditions seem more complex.

Let us first analyze condition (40) in more details. This condition is equivalent to the following one:

$$\left(\frac{\alpha_{uv}}{\alpha_u}\frac{\alpha_u}{\hat{\alpha}_u} - \frac{\hat{\alpha}_{uv}}{\hat{\alpha}_u}\frac{\alpha_v}{\hat{\alpha}_v}\right)^2 < 4\frac{\alpha_u}{\hat{\alpha}_u}\frac{\alpha_v}{\hat{\alpha}_v}. \tag{43}$$

Since $\cot(\phi) = \frac{\alpha_{uv}}{\alpha_u}$ [see (3)], we have

$$\left(\cot(\phi)\frac{\alpha_u}{\hat{\alpha}_u} - \cot(\hat{\phi})\frac{\alpha_v}{\hat{\alpha}_v}\right)^2 < 4\frac{\alpha_u}{\hat{\alpha}_u}\frac{\alpha_v}{\hat{\alpha}_v}. \tag{44}$$

Let us note $\delta\phi = \hat{\phi} - \phi$ the error made of the estimation of the angle ϕ. In general, the estimated angle $\hat{\phi}$ is set to $\pi/2$. We show now that this choice does not have a great influence on the stability of the system. Indeed, we thus have $\cot(\hat{\phi}) = 0$, $\cot(\phi) = \cot(\hat{\phi} - \delta\phi) = \cot(\pi/2 - \delta\phi) = \tan(\delta\phi)$ and condition (44) can be written as

$$\tan^2(\delta\phi) < 4\frac{\hat{\alpha}_u}{\hat{\alpha}_v}\frac{\alpha_v}{\alpha_u} \Rightarrow |\delta\phi| < \arctan\left(2\sqrt{\frac{\alpha_v}{\hat{\alpha}_v}\Big/\frac{\alpha_u}{\hat{\alpha}_u}}\right). \tag{45}$$

In Fig. 5, the bounds for $|\delta\phi|$ are plotted versus the ratio $\alpha_u/\hat{\alpha}_u$ on the x axis and versus the ratio $\alpha_v/\hat{\alpha}_v$ on the y axis. From this figure, we obtain $|\delta\phi| < 63°$ if the ratios are well estimated. In the worst considered case, when the estimated ratio is $\hat{\alpha}_u/\hat{\alpha}_v = 1/5$ and the real ratio is $\alpha_u/\alpha_v = 5$ (which corresponds to a rectangular pixel with the x-length five times the y-length), we have $|\delta\phi| < 22°$. For a more common camera with $\alpha_u/\alpha_v = 3/4$, if $\hat{\alpha}_u/\hat{\alpha}_v = 4/3$ then $|\delta\phi| < 56°$, which seems impossible to not satisfy.

Let us now analyze condition (41). This condition can be written as

$$\|(\mathbf{I_2} - \delta\mathbf{A}_{11})\| + \|\delta\mathbf{p}_0\| < 2\sqrt{\sigma}/\sqrt{1+\bar{\gamma}^2} \tag{46}$$

which implies

$$\|\delta\mathbf{p}_0\| < 2\sqrt{\sigma}/\sqrt{1+\bar{\gamma}^2} - \|(\mathbf{I_2} - \delta\mathbf{A}_{11})\|. \tag{47}$$

In Fig. 6, the bounds for $\|\delta\mathbf{p}_0\|$ are plotted versus the ratio $\alpha_u/\hat{\alpha}_u$ on the x axis and versus the ratio $\alpha_v/\hat{\alpha}_v$ on the y axis (in order to obtain these results, we have set $\bar{\gamma} = 0.364$, which corresponds to a camera with a 20° vision angle, and $\delta\phi = 20°$). In the worst considered case, when $\alpha_u/\hat{\alpha}_u = 0.5$ and $\alpha_v/\hat{\alpha}_v = 2$ then $\|\delta\mathbf{p}_0\| < 0.44$. Since $\|\delta\mathbf{p}_0\| = \sqrt{\hat{x}_0^2 + \hat{y}_0^2}$, this means that the 3-D vector associated with the estimated principal point must be in a cone of angle $\psi_0 < 24°$.

The previous conditions are quite easy to satisfy. However, the stability domain is not so large since condition (42) must also be verified. This last condition can be written as a second degree inequality:

$$c_2\nu^2 + c_1\nu + c_0 < 0 \tag{48}$$

where

$$c_2 = \sqrt{2}\frac{(k_1 - \mathrm{sign}(1-\nu))^2 k_2^2 g^2(\bar{\gamma})}{\sqrt{\sigma'+1-\sqrt{(\sigma'-1)^2 + 4\|\delta\mathbf{A}_{11}\|^2\|\delta\mathbf{p}_0\|^2}}}$$
$$c_1 = \frac{2\sqrt{2}\,\mathrm{sign}(1-\nu)(k_1 - \mathrm{sign}(1-\nu))k_2^2 g^2(\bar{\gamma})}{\sqrt{\sigma'+1-\sqrt{(\sigma'-1)^2 + 4\|\delta\mathbf{A}_{11}\|^2\|\delta\mathbf{p}_0\|^2}}} - \sigma_2\sigma_3$$
$$c_0 = \sqrt{2}\frac{(\|\delta\mathbf{A}_{11}\| + \sqrt{1+\|\delta\mathbf{p}_0\|^2})^2 g^2(\bar{\gamma})}{\sqrt{\sigma'+1-\sqrt{(\sigma'-1)^2 + 4\|\delta\mathbf{A}_{11}\|^2\|\delta\mathbf{p}_0\|^2}}}.$$

The solution of (48) is

$$\frac{-c_1 - \sqrt{c_1^2 - 4c_0c_2}}{2c_0} < \nu < \frac{-c_1 + \sqrt{c_1^2 - 4c_0c_2}}{2c_0}. \tag{49}$$

Example 1: We first consider a perfect camera calibration and obtain the possible bounds for $\hat{d}^*/d^*$. In that simple case, we have $\delta\mathbf{A}_{11} = \mathbf{I}_2$, $\delta\mathbf{p}_0 = \mathbf{0}$, $\sigma = \sigma_1 = \sigma_2 = 2$, $\sigma' = 1$ and $\nu = \frac{\hat{d}^*}{d^*}$. Condition (42) can thus be written

$$\left(1 - \frac{\hat{d}^*}{d^*}\right)^2 g^2(\bar{\gamma}) < 4\frac{\hat{d}^*}{d^*}. \tag{50}$$

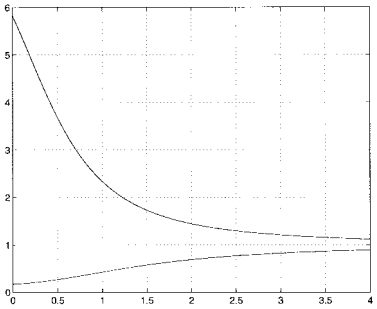


Fig. 7. Stability bounds for relative depth $\hat{d}^*/d^*$.

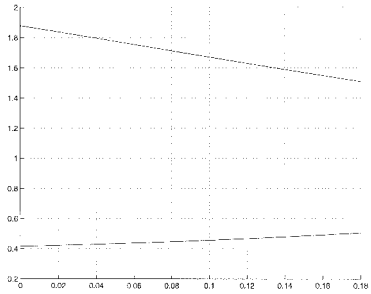


Fig. 8. Stability bounds for ν versus $\|\delta \mathbf{p}_0\|$.

The corresponding solution of (49) is

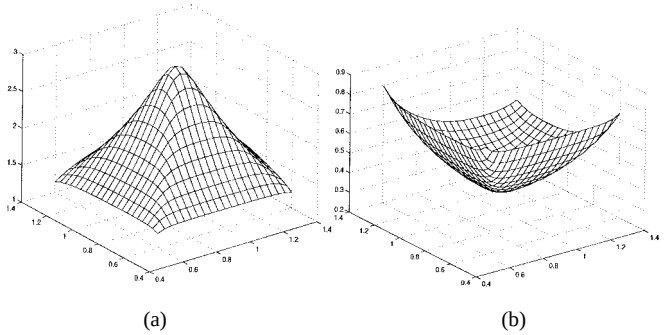
$$1+2\frac{1-\sqrt{g^2(\bar{\gamma})+1}}{g^2(\bar{\gamma})} < \frac{\hat{d}^*}{d^*} < 1+2\frac{1+\sqrt{g^2(\bar{\gamma})+1}}{g^2(\bar{\gamma})}. \quad (51)$$

The two bounds are plotted in Fig. 7 versus $\bar{\gamma}$. From this figure, if we consider for example a camera with a 20° vision angle (then $\bar{\gamma} = 0.364$), the stability condition is verified if $0.24 < \hat{d}^*/d^* < 4.22$. If the real distance d^* is 50 cm, the system will asymptotically converge for any initial position in the task space if $\hat{d}^*$ is chosen between 12 and 211 cm. This result definitively validates the robustness of our control scheme in absence of camera calibration errors.

Moreover, similar results can be obtained by considering camera calibration errors. Since condition (42) depends on the five camera intrinsic parameters, we first study the stability with a fixed $\delta \mathbf{A}_{11}$ and a variable $\delta \mathbf{p}_0$, and after, with a variable $\delta \mathbf{A}_{11}$ and a fixed $\delta \mathbf{p}_0$. It must be noted that, if $\mathbf{n}^* \approx [0\ 0\ 1]^T$, then $\nu \approx \hat{d}^*/d^*$, $\forall \delta \mathbf{A}$.

Example 2: If we consider $\alpha_u/\hat{\alpha}_u = \alpha_v/\hat{\alpha}_v = 1.5$ (which means 50% error on each pixel length) and $\delta\phi = 5°$, the two corresponding bounds are plotted in Fig. 8. For example, if $\|\delta \mathbf{p}_0\| = 0.0875$ (which corresponds to a cone with a 5° angle), then $0.45 < \nu < 1.7$. In order to obtain a simpler interpretation of this condition, we suppose now that $\nu \approx \hat{d}^*/d^*$ (which means that the normal to the reference plane is $\mathbf{n}^* \approx [0\ 0\ 1]^T$). If the real distance d^* is again 50 cm, the system will asymptotically converge for any initial camera position if $\hat{d}^*$ is chosen between 23 and 85 cm.

Example 3: We fix now $\|\delta \mathbf{p}_0\|$ =0.0875 (which corresponds to a cone with a 10° angle) and again $\delta\phi = 5°$. The upper and lower bounds for ν are plotted in Fig. 9 versus the ratio $\alpha_u/\hat{\alpha}_u$ on the x axis and versus the ratio $\alpha_v/\hat{\alpha}_v$ on the y axis. For a common camera with $\alpha_u/\alpha_v = 3/4$, we obtain



(a) (b)

Fig. 9. Stability bounds for ν versus $\frac{\alpha_u}{\hat{\alpha}_u}$ and $\frac{\alpha_v}{\hat{\alpha}_v}$. (a) Upper bound. (b) Lower bound.

$0.53 < \nu \approx d^*/\hat{d}^* < 1.51$ if $\hat{\alpha}_u/\hat{\alpha}_v = 4/3$. If the real distance d^* is again 50 cm, the system will asymptotically converge for any initial position if $\hat{d}^*$ is chosen between 26 and 76 cm.

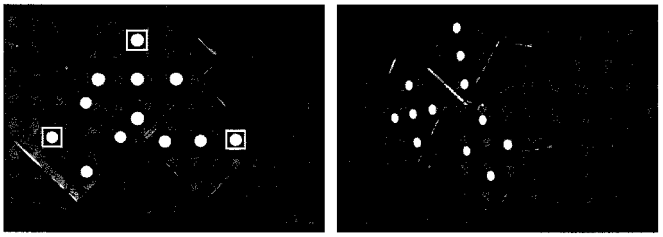
A more complete analysis is given in [15]. Let us emphasize that conditions (40)–(42) are more restrictive than conditions (35)–(38). When they are ensured, error $\|\mathbf{e}\|$ decreases at each iteration whatever the initial camera position in the whole task space. If this initial position is always in a known region, the stability analysis can be made from conditions (35)–(38) taking into account the restriction on the task space, and thus a larger stability domain will be obtained. More generally, all these conditions are only sufficient, and the convergence can be realized even for larger errors. In the next section, we will see that our method is also robust in presence of hand-eye calibration errors (the sufficient conditions for global asymptotic stability of the system in presence of such supplementary errors can be found in [15]).

VI. Experimental Results

The control law has been tested on a seven d.o.f. industrial robot MITSUBISHI PA10 (at EDF DER Chatou) and a six d.o.f. Cartesian robot AFMA (at IRISA). The camera is mounted on the robot end-effector. In the presented experiments, $\hat{d}^*$ is set to 50 cm while its real value d^* is 60 cm. As far as calibration is concerned, two different set of parameters have been used:

1) coarse calibration: the pixel and focal lengths given by the camera manufacturer are used. The image center has been used for the principal point. The transformation matrix between the camera and the robot end-effector frames is set with an accuracy to within 1 cm for translation and 5° for rotation.
2) bad calibration: a supplementary error is added to the camera intrinsic parameters (20%), as well as to the translation (5 cm on each axis) and to the rotation (5° on each axis) of the transformation matrix between the camera and the robot end-effector.

We present first the results obtained using coarse calibration for 3-D visual servoing and 2-D visual servoing. Then, the 2-1/2-D visual servoing results are presented using coarse and bad calibration. The images corresponding to the desired and initial camera position are given in Fig. 10(a) and (b), respectively. As can be seen on Table I, the corresponding



(a) (b)

Fig. 10. Images of the target for the desired and the initial camera position. (a) Desired image. (b) Initial image.

TABLE I
MOTION PARAMETERS

calibration	$\mathbf{u}\theta$ (deg)			$\mathbf{t}/\|\mathbf{t}\|$		
exact	28.1	−33.8	96.1	−0.11	0.99	0.10
coarse	33.0	−32.8	96.3	−0.20	0.97	0.13
bad	26.9	−26.6	99.0	−0.25	0.96	0.04

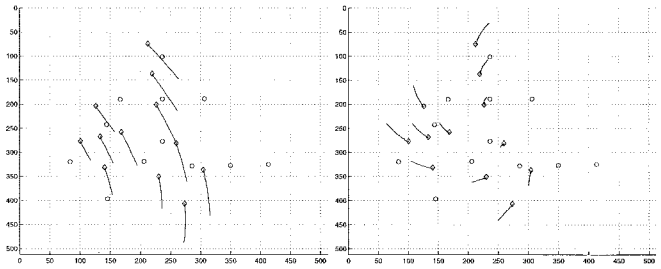
camera displacement is very important. The target is composed by twelve white marks lying on three different planes (see Fig. 10). The extracted visual features are the image coordinates of the center of gravity of each mark. With such simple images, the control loop can easily be realized at video rate (i.e., 25 Hz). For large camera displacements, such as the one considered, point matching between initial and desired images is an important computer vision problem. This problem is not considered here, because of the simplicity of the considered target. Of course, we can note that it also occurs for 2-D visual servoing, and similarly for 3-D visual servoing since point matching between the image and the 3-D model of the target is needed in that case.

A. 3-D Visual Servoing

In this experiment, the camera position is controlled in the Cartesian space. As can be seen in Fig. 11(a), the target leaves the camera field of view (for security reasons, the control scheme is stopped as soon as one of the target points is no longer visible in the image. Of course, the servoing could continue with less than 12 points and, if the system succeeds in converging, it implies that the lost points will come back into the image. However, we have considered that the loss of at least one point shows an unsatisfactory behavior). This failure is encountered because, using this scheme, there is absolutely no control in the image. The probability of failure increases considerably when a bad camera calibration is used or in the presence of hand-eye calibration errors.

B. 2-D Visual Servoing

In this experiment, the camera is fully controlled using classical image-based visual servoing. Great robustness with respect to calibration errors can thus be expected. However, the camera trajectory in the Cartesian space is not satisfactory because of coupling between the different visual features. This, once again, causes failure of servoing, since, as can be seen in Fig. 11(b), one target point leaves the image. This is due to the too large camera displacement from initial to desired poses.



(a) (b)

Fig. 11. Trajectory of target points in the image for 3-D and 2-D visual servoing. (a) 3-D visual servoing. (b) 2-D visual servoing.

C. 2-1/2-D Visual Servoing

We now present the results obtained using 2-1/2-D visual servoing. The three points of the target defining the reference plane are marked with a square in Fig. 10(a), and the chosen reference point is the nearest to the top of the image. Similar results may be obtained using another selection of the reference point. From the estimated homography, we get a partial estimation of the camera displacement. For example, the estimated rotation $\mathbf{R}$ and direction of translation $\mathbf{t}/\|\mathbf{t}\|$, using the initial and desired images, are given in Table I as a function of the camera calibration. Despite the coarse calibration which has been used, the estimation is quite precise (maximal rotational error is around 5°, as well as the angle error on the direction of translation). If a bad calibration is used, the rotational and translational errors may approximatively reach 7° and 9°, respectively.

In the first two experiments, the gain λ involved in the control law (23) was chosen constant. This explains that the convergence is very slow (approximatively 100 s). This is due to the fact that λ has to be set to a small value in order that the camera velocity be not too big at the beginning of the servoing. As it will be shown in the third experiment, λ can be automatically adapted in order to reduce the time to convergence.

1) Coarse Calibration: The error on $\mathbf{m}_e - \mathbf{m}_e^*$ and the estimated rotation are plotted in Figs. 12(a) and (b) respectively. The computed control law is given in Figs. 12(c) and (d). We can now observe the convergence of the task function toward 0. The error on the coordinates of each target point is given in Fig. 12(e). We can note the convergence of the coordinates to their desired value, which demonstrates the correct realization of the task. Finally, the corresponding trajectory in the image is given in Fig. 12(f). The reference point trajectory can be easily identified since it looks like a straight line in the image.

2) Bad Calibration: We now test the robustness of our approach with respect to a bad calibration, as described previously. The obtained results are given in Fig. 13. As can be seen in Fig. 13(a), the convergence of the error is no longer perfectly exponential. This is due to the bad calibration of the camera and the rough approximation of $\hat{d}^*$ (which had a very low influence using a coarse calibration). However, even in this worse case, we can note the stability and the robustness of the control law. Contrary to the previous experiment, the trajectory of the reference point in the image is no longer a

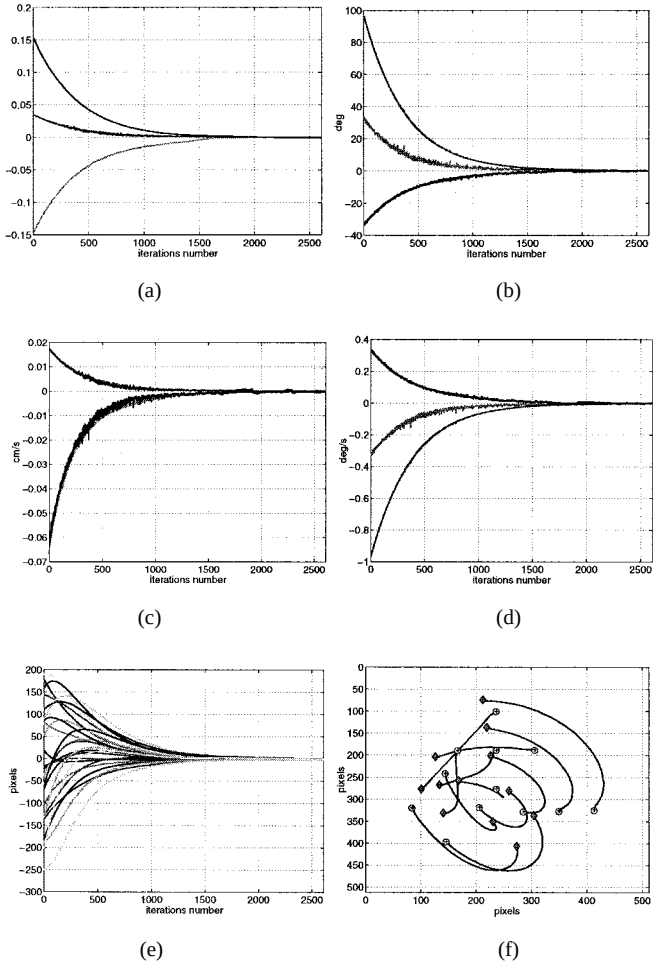


Fig. 12. Results of 2-1/2-D visual servoing with coarse camera and hand-eye calibration: (a) error in extended image coordinates, (b) rotation $u\theta$, (c) translational velocity, (d) rotational velocity, (e) error in image points coordinates, and (f) trajectories in the image of the target points.

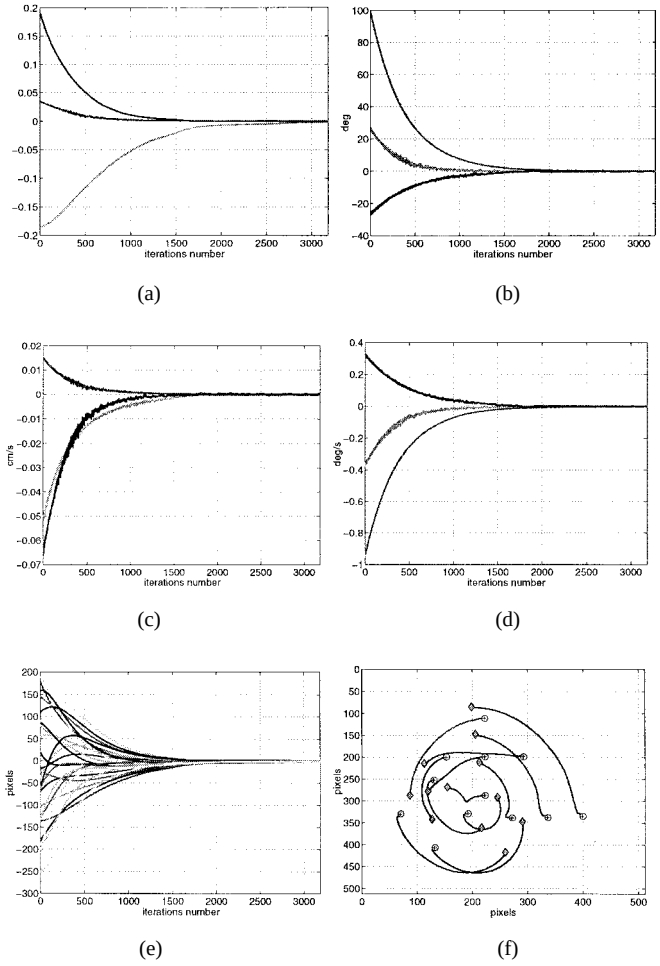


Fig. 13. Results of 2-1/2-D visual servoing with bad camera and hand-eye calibration: (a) error in extended image coordinates, (b) rotation $u\theta$, (c) translational velocity, (d) rotational velocity, (e) error in image points coordinates, and (f) trajectories in the image of the target points.

straight line since the camera is badly calibrated as well as the homogeneous transformation matrix between the camera and the robot end-effector frame. However, the convergence of the image points coordinates to their desired value demonstrates the correct realization of the task.

3) Bad Calibration and Adaptive Gain λ: In this experiment, gain λ is automatically increased when the error decreases. As can be seen in Fig. 14, the convergence rate has been divided by a factor of three while the initial camera position was very far away from its desired position (compare the initial value in Figs. 13(a), (b) and 14(a) and (b). The convergence has been reached in approximately 30 s, which is not so important on account of the large displacement to realize and the limited camera motion necessary for correct image tracking of the visual features. Of course, we can note that the convergence of the system is no longer exponential, and that the gain increase has added some noise to the control law (but not to the image features) near the convergence of the system.

From the numerous experiments that have been realized (see [15]), we can conclude that, when 2-D or 3-D visual servoing succeeds, convergence is also reached with our scheme, but with a more satisfactory behavior. Furthermore, the convergence domain of the 2-1/2-D visual servoing is indeed more large than for the two other schemes. We have to note however that our scheme sometimes fails. It occurs when some parts of the target occlude one or more image points (this problem is not encountered by considering a planar target) or when some points leave the image plane (once again, we always stop an experiment as soon as one point is no longer visible, even if it is not necessary).

VII. Adaptive Gains

We now present an adaptive control law which takes into account the constraint that the target has to remain in the camera field of view. Another approach would consist in determining off line a specified trajectory in the Cartesian frame ensuring this constraint in the image plane. This problem of path planning seems to be very complex, and has not been considered here. As for adaptive control, two kinds of methods can be distinguished [18]:

1) gains adaptation: this kind of control has the purpose of improving the robustness to the calibration errors. The adaptation is based on the stability analysis of the

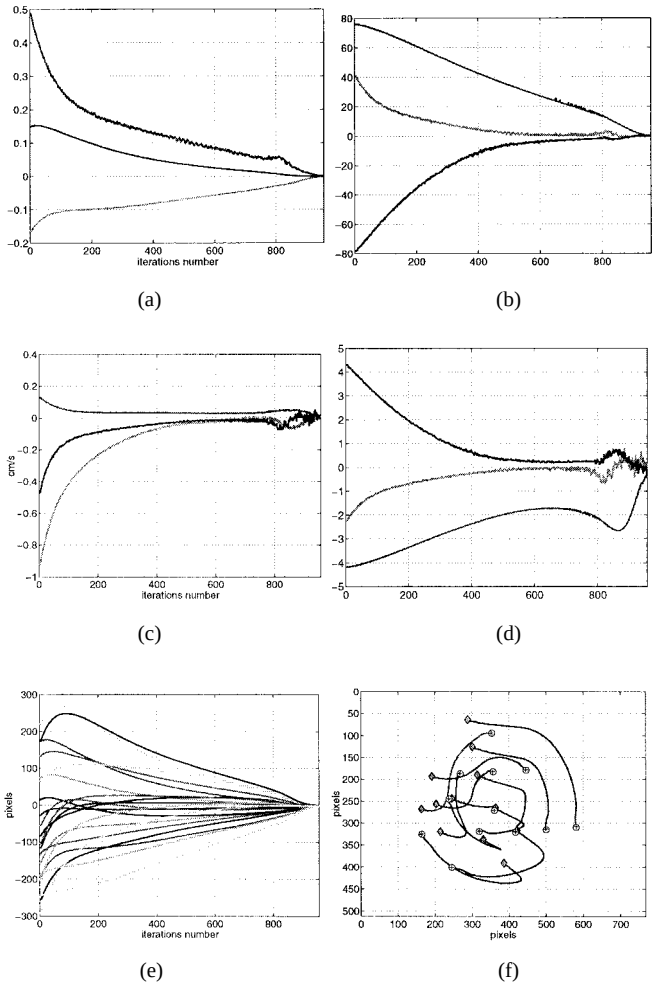


Fig. 14. Same experiment using an adaptive gain λ: (a) error in extended image coordinates, (b) rotation $u\theta$, (c) translational velocity, (d) rotational velocity, (e) error in image points coordinates, and (f) trajectories in the image of the target points.

closed-loop system and does not give any supplementary knowledge on the geometric parameters involved in the system (these parameters remain constant).

2) system parameters adaptation: this kind of control has the purpose of improving not only the stability, but also the dynamic behavior (and then the performance) of the system by the estimation of the involved geometric parameters.

We have used the first kind of adaptive control. The exponential decrease of the task function is again imposed

$$\dot{\mathbf{e}} = -\lambda \mathbf{D}\mathbf{e} \tag{52}$$

but using a positive diagonal matrix $\mathbf{D}$ whose elements are function of the position of the image points. More precisely, thanks to the particular form of the interaction matrix, $\mathbf{D}$ can be chosen as

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & f_u(u) & 0 & 0 \\ 0 & 0 & f_u(u)f_v(v) & 0 \\ 0 & 0 & 0 & \mathbf{I}f_u(u)f_v(v)f_w(w) \end{bmatrix} \tag{53}$$

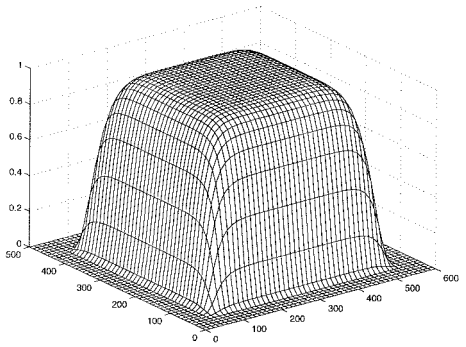


Fig. 15. Function $f(u)f(v)$.

where $0 \leq f(x) \leq 1$ is a bell-curve, symmetric with respect to $x_m = \frac{x_1+x_2}{2}$

$$f(x) = \begin{cases} e^{-\frac{(x-x_m)^{2n}}{(x-x_1)^m(x_2-x)^m}}, & \text{if } x_1 < x < x_2 \\ 0, & \text{otherwise} \end{cases} \tag{54}$$

m and n being two parameters used to design the form of the bell-curve (see for example the function $f(u)f(v)$ in Fig. 15).

In our case, the system is constrained such that

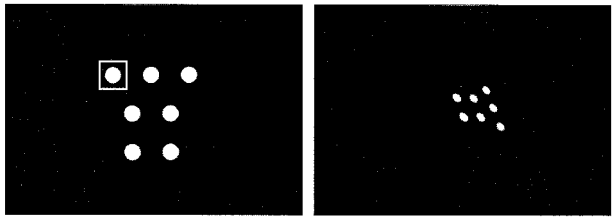
$$\begin{aligned} u &\in [u_{\min}, u_{\max}] \\ v &\in [v_{\min}, v_{\max}] \\ w &\in [w_{\min}, w_{\max}] \\ \mathbf{u}\theta &\in SO_3 \end{aligned} \tag{55}$$

where only the first part of the state is constrained since the rotation is free to evolve in SO_3. The bounds $[u_{\min}, u_{\max}]$ and $[v_{\min}, v_{\max}]$ are defined by the CCD size, while u and v are respectively the x and y coordinates of the image points nearest these bounds. Similarly, $[w_{\min}, w_{\max}]$ can be determined experimentally, such that, for example, the image is not blurred.

The elements of matrix $\mathbf{D}$ work like bandpass filters. For example, if the error on w grows because of the bad estimation of the rotation, then the last three elements of $\mathbf{D}$ become smaller and the rotational control law decreases. The same thing happens if the error on v (or on u) grows, then the last four (or five) elements of $\mathbf{D}$ become smaller, and only the stable part of the control law is considered. Using this control law, we have proved in [15] that the reference point never leaves the image even in presence of large camera and hand-eye calibration errors (providing the analysis of the robustness domain). This control law could also be used in 3-D visual servoing, but without any possible theoretical stability analysis. Furthermore, our technique cannot be used in 2-D visual servoing, since the rotational control loop is not decoupled from the translational one using this scheme.

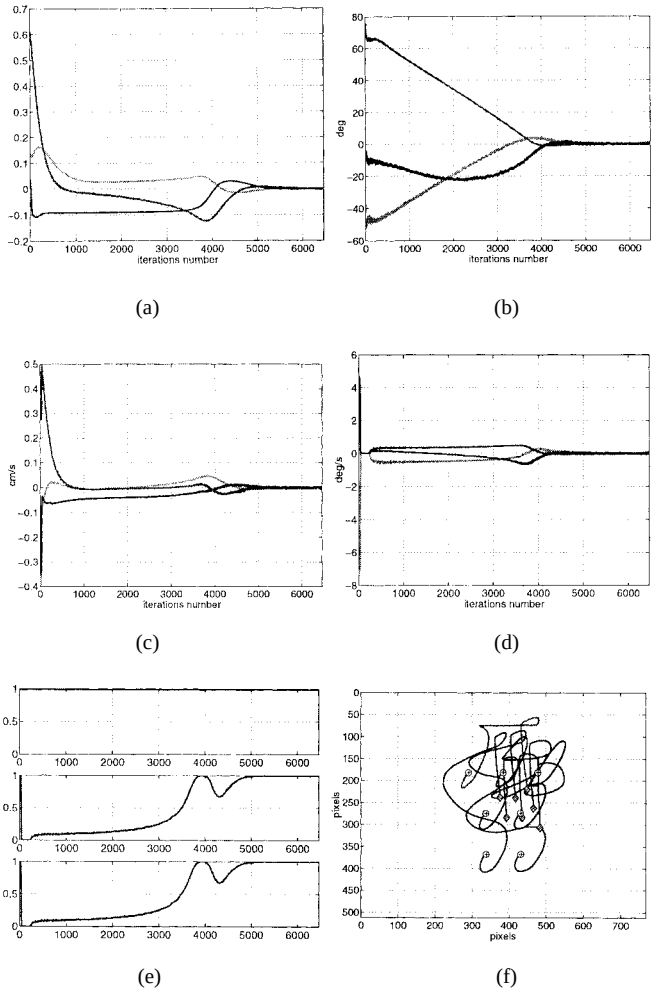
We present now an experiment showing the behavior of the system using the adaptive control law. The used target is now planar (see Fig. 16) and the chosen reference point is marked with a square in the image. Similar results can be obtained with other targets [15].

In order to prove that the adaptive control law allows a larger robustness domain, a supplementary error was added to



(a) (b)

Fig. 16. Images of the target for the desired and the initial camera position. (a) Desired image. (b) Initial image.



(a) (b) (c) (d) (e) (f)

Fig. 17. Robust control law: (a) error in extended image coordinates, (b) rotation $u\theta$, (c) translational velocity, (d) rotational velocity, (e) gains f_u, f_v, and $f_u f_v$, and (f) trajectories in the image of the target points.

the translation (20 cm on each axis) and to the rotation (20° on each axis) of the transformation matrix between the camera and the robot end-effector. With such a bad calibration and a large initial camera displacement, the target leaves the camera field of view using classical 2-D and 3-D visual servoing, and using 2-1/2-D visual servoing with $\mathbf{D} = \mathbf{I}_6$. The distance $\hat{d}^*$ is set again to 50 cm while its real value d^* is 60 cm.

The results obtained using the adaptive control law are shown in Fig. 17. At the beginning of the servoing, the gains are equal to 1 since the target is in the center of the image. When the visual servoing starts, the target moves rapidly at the top of the image. The gain f_v decreases during the first 50 iterations. The rotational velocity thus decreases, and the target does not leave the image.

Then, the target starts to move on the left (since the control on u is stable), and thus, comes back to the center. The gains increase progressively to 1 (iteration 4000). At this moment, the two points on the bottom of the image, which have gone afterward their desired position since the system is badly calibrated, move in the bottom of the image. The gain f_v starts again to decrease until about 0.6. At the same time, the rotation continues to decrease toward zero. Therefore, the error decreases and the gain can increase to 1 until the convergence.

In this experiment, the convergence rate was slow since we only wanted to show the behavior of the system when only the elements of $\mathbf{D}$ were changed. Once again, the convergence rate can be improved by increasing the gain λ when the error decreases.

VIII. Conclusion

In this paper, we have proposed a new approach to vision-based robot control which presents many advantages with respect to classical position-based and image-based visual servoing. This new method does not need any 3-D target model, nor a precise camera calibration and presents very interesting decoupling and stability properties. Thanks to its simple structure, analytical results on its robustness with respect to calibration errors have been obtained. Experimental results show the validity of our approach and its robustness not only with respect to camera calibration errors, but also to hand-eye calibration errors. More experimental results can be found in [15]. One of the drawbacks of our method is that, for a non planar target, at least eight points are necessary to estimate the homography matrix, while at least four points are theoretically needed in the other schemes. Another drawback is that our method is more sensitive to image noise than 2-D visual servoing, since this scheme directly uses visual features as input of the control law, without any supplementary estimation step. Future work will thus be devoted to improve the robustness of our method with respect to image noise, and to the use of 2-1/2-D visual servoing on real objects and complex images.

Acknowledgment

The authors would like to thank the team manager and the researchers of the Teleoperation/Robotics group, DER Chatou, for their participation and help, C. Samson and the anonymous reviewers for their valuable comments, and T. Drummond for careful reading.

References

[1] P. K. Allen, A. Timcenko, B. Yoshimi, and P. Michelman, "Automated tracking and grasping of a moving object with a robotic hand-eye system," *IEEE Trans. Robot. Automat.*, vol. 9, pp. 152–165, Apr. 1993.

[2] R. Basri, E. Rivlin, and I. Shimshoni, "Visual homing: Surfing on the epipoles," in *IEEE Int. Conf. Comput. Vision, ICCV'98*, Bombay, India, Jan. 1998, pp. 863–869.

[3] F. Bensalah and F. Chaumette, "Compensation of abrupt motion changes in target tracking by visual servoing," in *IEEE/RSJ Int. Conf. Intell. Robots Syst., IROS'95*, Pittsburgh, PA, Aug. 1995, vol. 1, pp. 181–187.

[4] B. Boufama and R. Mohr, "Epipole and fundamental matrix estimation using the virtual parallax property," in *IEEE Int. Conf. Comput. Vision, ICCV'95*, Cambridge, MA, 1995, pp. 1030–1036.
[5] F. Chaumette, "Potential problems of stability and convergence in image-based and position-based visual servoing," in *The Confluence of Vision and Control, LNCIS Series*, D. Kriegman, G. Hager, and A. Morse, Eds. New York: Springer Verlag, 1998, vol. 237, pp. 66–78.
[6] D. Dementhon and L. S. Davis, "Model-based object pose in 25 lines of code," *Int. J. Comput. Vision*, vol. 15, nos. 1/2, pp. 123–141, June 1995.
[7] B. Espiau, F. Chaumette, and P. Rives, "A new approach to visual servoing in robotics," *IEEE Trans. Robot. Automat.*, vol. 8, pp. 313–326, June 1992.
[8] B. Espiau, "Effect of camera calibration errors on visual servoing in robotics," in *Proc. 3rd Int. Symp. Experimental Robot.*, Kyoto, Japan, Oct. 1993.
[9] O. Faugeras and F. Lustman, "Motion and structure from motion in a piecewise planar environment," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 2, no. 3, pp. 485–508, 1988.
[10] G. D. Hager, "A modular system for robust positioning using feedback from stereo vision," *IEEE Trans. Robot. Automat.*, vol. 13, pp. 582–595, Aug. 1997.
[11] R. I. Hartley, "In defense of the eight-point algorithm," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 580–593, June 1997.
[12] K. Hashimoto, Ed, *Visual Servoing: Real Time Control of Robot Manipulators Based on Visual Sensory Feedback*, of *World Scientific Series in Robotics and Automated Systems*. Singapore: World Scientific, 1993, vol. 7.
[13] S. Hutchinson, G. D. Hager, and P. I. Corke, "A tutorial on visual servo control," *IEEE Trans. Robot. Automat.*, vol. 12, pp. 651–670, Oct. 1996.
[14] E. Malis, F. Chaumette, and S. Boudet, "Positioning a coarse-calibrated camera with respect to an unknown planar object by 2-D 1/2 visual servoing," in *Proc. 5th IFAC Symp. Robot Contr. (SYROCO'97)*, Nantes, France, Sept. 1997, vol. 2, pp. 517–523.
[15] E. Malis, "Contributions à la modélisation et à la commande en asservissement visuel," Ph.D. Thesis, Univ. Rennes I, IRISA, France, Nov. 1998.
[16] E. Malis, F. Chaumette, and S. Boudet, "Camera displacement through the recovery of a homography: Application to 2-1/2-D visual servoing," *Int. J. Comput. Vision*, 1999.
[17] N. P. Papanikolopoulos, P. K. Kosla, and T. Kanade, "Visual tracking of a moving target by a camera mounted on a robot: A combination of control and vision," *IEEE Trans. Robot. Automat.*, vol. 9, pp. 14–35, Feb. 1993.
[18] C. Samson, M. L. Borgne, and B. Espiau, *Robot Control: The Task Function Approach, Oxford Engineering Science Series*. Oxford, U.K.: Clarendon, 1991, vol. 22.
[19] L. E. Weiss, A. C. Sanderson, and C. P. Neuman, "Dynamic sensor-based control of robots with visual feedback," *IEEE J. Robot. Automat.*, vol. 3, pp. 404–417, Oct. 1987.
[20] W. J. Wilson, C. C. W. Hulls, and G. S. Bell, "Relative end-effector control using Cartesian position-based visual servoing," *IEEE Trans. Robot. Automat.*, vol. 12, pp. 684–696, Oct. 1996.

Ezio Malis was born in Gorizia, Italy, in 1970. He graduated from the University Politecnico di Milano, Italy, and from the Ecole Supérieure d'Electricité (Supélec), Paris, France, in 1995 and received the Ph.D. degree from the University of Rennes, Rennes, France, in 1998.

He is a Research Associate with the Department of Engineering, University of Cambridge, Cambridge, U.K. His research interests include robotics, computer vision, and vision-based control.

François Chaumette was born in Nantes, France, in 1963 and graduated from Ecole Nationale Supérieure de Mécanique, Nantes, in 1987. He received the Ph.D. degree and "Habilitation à Diriger des Recherches" in computer science from the University of Rennes in 1990 and 1998 respectively. Since 1990, he has been with IRISA/INRIA, Rennes. His research interests include robotics, computer vision, and especially the coupling of these two research domains (vision-based control, active vision and purposive vision).

Dr. Chaumette received the AFCET/CNRS Prize for the best french thesis in automatic control, in 1991.

Sylvie Boudet was born in Grenoble, France, in 1971. She graduated from Ecole Supérieure d'Electricité (Supélec), Paris, France, in 1994.

She works as a Research Engineer at the Research Center of Electricité de France (EDF), Chatou. EDF is the French company that produces, transports, and delivers electricity. Her research interests are to develop robotics controllers including force-control and vision-based control, in order to make it easier to have robotics maintenance in nuclear power plant or in any other hostile environment. She has also led a project on a medical robot, Hippocrate, to improve medical diagnosis on heart diseases.