Controlling Robots With Two Cameras: How to Do it Properly. *†

Bart LAMIROY, Bernard ESPIAU, Nicolas ANDREFF, Radu HORAUD

INRIA Rhône-Alpes, 655 Avenue de l'Europe, 38330 Montbonnot, FRANCE

e-mail: *Frist.Last@inrialpes.fr*

Abstract

In this paper we address visual servoing through a fixed stereo rig using an image Jacobian. Existing methods are based on the stacking of monocular servo image Jacobians, resulting in largely over-constrained control commands. In this paper we formally show that the epipolar constraint between two images can be taken into account explicitly. We then show that use of stereo significantly increases the quality of the servo task execution, especially where precision, robustness and smoothness of movement is concerned.

1 Introduction

In this paper we consider visual servoing as a task function [10]. Several approaches of robot control using a single camera [3, 5] or stereo rigs [9, 4] exist. In the case where only a single camera is concerned, a certain number of assumptions, such as camera calibration need to be made. In that case, however, the problem is sufficiently constrained and a minimal set of control variables can be used. One has to bear in mind that a number of singularities exists, making visual control impossible near those configurations (*e.g.* 180 degree rotations [2]).

The use of a stereo rig avoids these singularities, and requires less strict camera calibration. The existing approaches are, however, quite *ad hoc* and restricted. In this paper we propose a formal framework for using stereo servoing. We show that the epipolar constraint can be taken into account for all camera configurations and that the use of stereo control has a great number of advantages with respect to monocular servoing.

The paper outline is as follows:

- First, we introduce the notions of monocular and stereo servoing.
- In section § 3, we develop the formalism of constrained stereo servoing, applied to the epipolar constraint.
- Next, we compare our approach to the approach consisting in the stacking of monocular Jacobians. We show that in certain cases both approaches are identical, but that in a number of situations the constrained approach is preferable.
- Finally, we show, by a series of experiments, that stereo servoing behaves better than monocular servoing in a large number of situations.

2 Visual Servoing

It is known [3, 5] that, given the proper knowledge on an observed 3D point set $\{S_i\}_{i=1\dots m}$ and its (fixed) position within the reference frame of a robot end-effector, it is possible to control this robot so to align the projection s of the point set $\{S_i\}$ in an image with a predefined goal position in this image (generally denoted $s^\star$) under condition that $s^\star$ is the projection of an attainable 3D position $S^\star$. This is represented in Figure 1 for the monocular case (*i.e* a single camera is controlling the robot).

In that case, the observed image speed $\dot{s}$ of the considered point set is related to the effector speed, animated with the kinematic screw $\mathcal{T}$, through the image Jacobian $\mathbf{J}$ by the following equation:

$$\dot{s} = \mathbf{J}\mathcal{T} \tag{1}$$

By imposing that the observed point set s moves towards the required goal position $s^\star$ (*i.e.* $\dot{s} = g\,(s^\star - s)$), we can compute the kinematic screw $\mathcal{T}$ to be sent to the robot end-effector within a control loop:

$$\mathcal{T} = g\,\mathbf{J}^\dagger\,(s^\star - s) \tag{2}$$

where g is a scalar gain factor, $\mathbf{J}^\dagger$ the $6 \times 2m$ pseudo-inverse of the image Jacobian and $(s^\star - s)$ the observed $2m$ error vector in the image. The kinematic

*To be presented at the IEEE International Conference on Robotics and Automation (ICRA), April 24-28, 2000, San Francisco, USA.

†This work was supported by VIGOR, Esprit-IV reactive LTR project, number 26247.

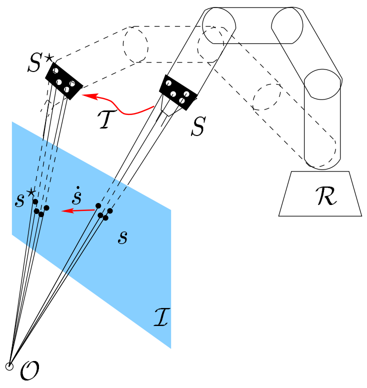

Figure 1: Fixed monocular visual servoing: the observation of a set of well known control points allows to compute the kinematic screw $\mathcal{T}$ that aligns the observed position s with a required goal position $s^\star$.

screw is updated at each control loop iteration with the newly observed error, and, in certain cases with a newly computed Jacobian [5].

Now, the formalism of this approach does in no extent impose that the measured error comes from just one image. It can be rewritten to take into account two images (or more) in a very straightforward manner.

Suppose we observe the movement of a set of 3D points $\{S_i\}_{i=1\ldots m}$ in two different cameras $\mathcal{C}^l$ and $\mathcal{C}^r$. We suppose also that these points are rigidly fixed to the robot end-effector, and that the latter is animated by a kinematic screw $\mathcal{T}$. Then, because of the rigidity constraint, the following equations hold:

$$\dot{s^l} = \mathbf{J}_l \mathcal{T}$$
$$\dot{s^r} = \mathbf{J}_r \mathcal{T}$$

or, concatenating the two matrices,

$$\begin{aligned} \begin{pmatrix} \dot{s^l} \\ s^r \end{pmatrix} &= \begin{pmatrix} \mathbf{J}_l \\ \mathbf{J}_r \end{pmatrix} \mathcal{T} \\ \dot{s}^{l-r} &= \mathbf{J}_{l-r}\, \mathcal{T} \end{aligned} \qquad (3)$$

which is exactly the same equation as (1), and gives rise to the same solution as the one presented in equation (2). The 3D velocity induced by the command

$$\begin{aligned} \mathcal{T} &= g \left[\left(\mathbf{J}_{l-r}^\mathsf{T} \mathbf{J}_{l-r} \right)^{-1} \mathbf{J}_{l-r}^\mathsf{T} \right] \left(s^{l-r\star} - s^{l-r} \right) \\ \mathcal{T} &= g\, \mathbf{J}_{l-r}^\dagger \left(s^{l-r\star} - s^{l-r} \right) \end{aligned} \qquad (4)$$

now tends to obtain a simultaneous convergence in two images rather than in just one single view, as was the case in equation (2).

Remarks:

1. It is clear that the stereo Jacobian obtained by stacking two monocular Jacobians is largely over constrained, and that the control data s are redundant. Indeed, s^l and s^r are related by the epipolar geometry of the stereo rig [7, 8]. This redundancy should be taken into account when computing the pseudo-inverse of the Jacobian matrix. In this paper we show that the epipolar constraint can be modeled formally, something which has never been done before[1].

2. Another point concerns the 3D trajectory executed by the robot under monocular and stereo servoing. In both cases, the expected image trajectory (2D) is usually a straight line. However, in the monocular case, there is no guarantee that this straight 2D line translates into a straight 3D trajectory, since any planar curve can project into a line. On the other hand, the 3D trajectory in the stereo case projects into a line on either image, imposing *de facto* a straight line movement in space.

In the next section we shall show how the epipolar constraint can be taken into account to compute a new Jacobian which takes into account the relationship between the left and right image points.

3 Constrained Servoing

3.1 Notations

Consider a fixed stereo rig, with known epipolar geometry $\mathbf{F}$. The rig observes a set of m 3D control points $\{S_i\}_{i=1\ldots m}$ projecting into the left (resp. right) image in $\{s_i^l\}_{i=1\ldots m}$ (resp. $\{s_i^r\}_{i=1\ldots m}$).

Let s_i be the 4×1 vector $({s_i^l}^\mathsf{T}\ {s_i^r}^\mathsf{T})^\mathsf{T}$ formed by the concatenation of the left and right projections of the 3D control point S_i, and let s^T be the $1 \times 4m$ vector $(s_1^\mathsf{T} \ldots s_m^\mathsf{T})$ the concatenation of all $s_{i=1\ldots m}$.

The homogeneous coordinate notation of s_i^l (resp. s_i^r) will be referred to as $\tilde{s}_i^l$ (resp. $\tilde{s}_i^r$), such that the epipolar constraint h_i between both points can be expressed as follows [7, 8]:

$$h_i = {\tilde{s}_i^l}^\mathsf{T} \mathbf{F} \tilde{s}_i^r = 0.$$

[1] In the case where the images are rectified, the redundancy is easily expressed as $s_y^l = s_y^r$ and can be treated by simply neglecting the y-component of one of the points. This is what was done by Hager *et al.* in [4]

3.2 Establishing the Control Command

Let $s^\star$ be an attainable visual goal position and s the observed current position of the set of 3D points $\{S_i\}_{i=1\ldots m}$. Then, the visual servoing problem can be expressed as the minimization of the error norm $\frac{1}{2}(s-s^\star)^\mathsf{T}(s-s^\star)$ by interacting with the robot control parameters [3].

In our case, the vector s is obtained through a fixed stereo rig, and is therefore constrained by the epipolar geometry. A particular case, when both images of the stereo rig are rectified, has already been developed in [4]. In the general case, the above minimization problem can be rewritten to take into account the constraints on s, and thus becomes:

$$\mathcal{L} = \frac{1}{2}(s-s^\star)^\mathsf{T}(s-s^\star) + \lambda^\mathsf{T} h$$

λ being the Lagrange multiplier, and $h^\mathsf{T} = (h_1 \ldots h_m)$. If the $4m \times m$ gradient matrix $\mathbf{G}$ corresponding to $\frac{\partial h}{\partial s}$ is of full rank m , then solving for $\frac{\partial \mathcal{L}}{\partial s} = 0$ yields

$$\mathbf{P}(s - s^\star) = 0$$

where $\mathbf{P} = Id - \mathbf{G}^\mathsf{T}\left(\mathbf{G}\mathbf{G}^\mathsf{T}\right)^{-1}\mathbf{G}$. $\mathbf{P}$ is the rank $3m$ symmetric, $4m \times 4m$ orthogonal projection operator onto $\mathrm{Ker}\,(h)$ (See [10] pp. 122–124 for details). We need to prove before, however, that $\mathbf{G}$ is of full rank m.

1. Formalism of $\frac{\partial h_i}{\partial s_i}$.

$$\begin{aligned} h_i &= \tilde{s}_i^{l\,\mathsf{T}}\mathbf{F}\tilde{s}_i^r \\ s_i &= \begin{pmatrix} s_i^l \\ s_i^r \end{pmatrix} \\ \mathbf{F} &= \begin{pmatrix} \mathbf{\Phi}_{[2\times 2]} & \Phi_{1\,[2\times 1]} \\ \Phi_2{}^\mathsf{T}{}_{[1\times 2]} & \phi_3 \end{pmatrix} \end{aligned}$$

Therefore we can write the 4×1 vector

$$\frac{\partial h_i}{\partial s_i} = \begin{pmatrix} s_i^{r\,\mathsf{T}}\mathbf{\Phi}^\mathsf{T} + \Phi_1^\mathsf{T} \\ s_i^{l\,\mathsf{T}}\mathbf{\Phi} + \Phi_2 \end{pmatrix} \tag{5}$$

2. Rank of $\mathbf{G}$.

$\mathbf{G}$ is of full rank m *iff* $\frac{\partial h_i}{\partial s_i} \neq 0\,\forall i$. Suppose $\exists j$: $\frac{\partial h_j}{\partial s_j} = 0$.

$$\frac{\partial h_j}{\partial s_j} = 0$$

$$\Leftrightarrow \begin{pmatrix} s_i^{r\,\mathsf{T}}\mathbf{\Phi}^\mathsf{T} + \Phi_1^\mathsf{T} \\ s_i^{l\,\mathsf{T}}\mathbf{\Phi} + \Phi_2 \end{pmatrix} = 0 \tag{6}$$

$$\Leftrightarrow \begin{cases} \tilde{s}_i^{l\,\mathsf{T}}\mathbf{F} \approx \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \\ \tilde{s}_i^{r\,\mathsf{T}}\mathbf{F}^\mathsf{T} \approx \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \end{cases}$$

$\Leftrightarrow$ The epipolar lines associated to s_i^l and s_i^r are of the form $0x + 0y + c = 0$. (7)

It is clear from equation (6) that there exists at least one s that satisfies the condition *iff* $-\Phi_1 \in \mathrm{Im}\,(\mathbf{\Phi})$ and $-\Phi_2 \in \mathrm{Im}\left(\mathbf{\Phi}^\mathsf{T}\right)$. So $\mathbf{G}$ cannot be guaranteed to be of full rank.

However, the following geometric proof shows that the equivalent constraints, formulated in equation (7), restrict the case to very particular camera positions that can easily be avoided. In a generic setup, $\mathbf{G}$ is always of rank m.

Proof: Consider a fixed stereo rig of cameras $\mathcal{C}_1$ and $\mathcal{C}_2$. $\mathcal{I}_1$ represents the image plane of camera $\mathcal{C}_1$, f_1 its focal point, and $\mathcal{F}_1$ the focal plane (resp. $\mathcal{I}_2$, f_2, $\mathcal{F}_2$ for camera $\mathcal{C}_2$; as represented in Figure 2). Since, for any camera, the locus of the

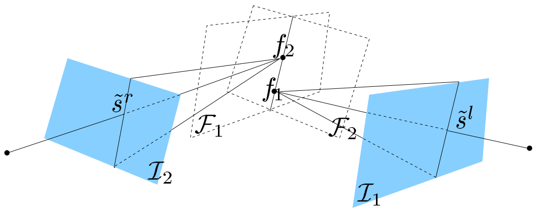

Figure 2: Generic configuration of a stereo rig for which $\mathbf{G}$ may not be of full rank m. The focal points of the cameras lie on the intersection between their respective focal planes.

points projecting at infinity in the image plane is, by construction, its focal plane, any 3D line projecting onto $(\,0\,0\,1\,)$ lies on the focal plane of that camera.

Now, from equation (7) we know that $\mathbf{G}$ is not of full rank when both epipolar lines of two observed points (one in each image) $\tilde{s}^l$ and $\tilde{s}^r$ are $(\,0\,0\,1\,)$. This means that:

(a) The line of sight of the point $\tilde{s}^l$ in image $\mathcal{I}_1$ lies in the focal plane $\mathcal{F}_2$. Since the focal point f_1 of camera $\mathcal{C}_1$, by definition, lies on the same line of sight, we can conclude that $f_1 \in \mathcal{F}_2$.

(b) The same reasoning holds for $\tilde{s}^r$ in image $\mathcal{I}_2$: the point lies on the focal plane $\mathcal{F}_1$ as does the focal point f_2.

If using typical ccd cameras with $8-16$mm lenses this setup is impossible to attain, so we can consider that $\mathbf{G}$ is always of rank m. Furthermore, we did not take into account the fact that $\tilde{s}^l$ and $\tilde{s}^r$ are the projections of the same 3D point S, in which case S is constrained to lie on the intersection between $\mathcal{F}_1$ and $\mathcal{I}_2$, defeating the assumption that $\tilde{s}^l$ and $\tilde{s}^r$ are visible points.

□

Since $\mathbf{F}$ is known up to a scale factor, $\mathbf{G}$ is also known up to a scale factor. However, this factor is canceled out in the computation of $\mathbf{P}$, and $\mathbf{P}$ is defined uniquely.

Proof:

let $\mathbf{G} = \gamma \mathbf{G}'$ then

$$\begin{aligned}
\mathbf{P} &= Id - \mathbf{G}^\top \left(\mathbf{G}\mathbf{G}^\top\right)^{-1} \mathbf{G} \\
&= Id - \frac{\gamma^2}{\gamma^2} \mathbf{G}^\top \left(\mathbf{G}\mathbf{G}^\top\right)^{-1} \mathbf{G} \\
&= Id - \frac{1}{\gamma} \mathbf{G}^\top \left(\frac{1}{\gamma}\mathbf{G}\frac{1}{\gamma}\mathbf{G}^\top\right)^{-1} \frac{1}{\gamma}\mathbf{G} \\
&= Id - \mathbf{G}'^\top \left(\mathbf{G}'\mathbf{G}'^\top\right)^{-1} \mathbf{G}'
\end{aligned}$$

□

Now that we have shown that we satisfy the conditions expressed in [10] for constrained redundant task functions we can express our visual servoing task as the minimization of a new control vector e. Our control vector now becomes $e = \mathbf{P}(s - s^\star)$. In order to obtain an exponential convergence rate, ideal behaviour is $\dot{e} = -g\, e$, where g is a positive scalar gain factor. This can be rewritten as

$$\left.\begin{aligned}
\dot{e} &= \mathbf{P}\dot{s} + \dot{\mathbf{P}}\,(s - s^\star) \\
\dot{e} &= -g\,\mathbf{P}\,(s - s^\star) \\
\dot{s} &= \mathbf{J}\mathcal{T}
\end{aligned}\right\} -g\,\mathbf{P}\,(s - s^\star) = \mathbf{P}\mathbf{J}\mathcal{T} + \dot{\mathbf{P}}\,(s - s^\star)$$

In near convergence conditions, $\dot{\mathbf{P}}\,(s - s^\star)$ can be neglected, in analogy with [10], resulting in

$$-g\,\mathbf{P}\,(s - s^\star) = \mathbf{P}\mathbf{J}\mathcal{T} \tag{8}$$

The kinematic screw needed in order to obtain optimal error convergence is obtained by minimizing

$$J = \frac{1}{2}\left(g\,\mathbf{P}\,(s - s^\star) + \mathbf{P}\mathbf{J}\mathcal{T}\right)^\top \left(g\,\mathbf{P}\,(s - s^\star) + \mathbf{P}\mathbf{J}\mathcal{T}\right)$$

which gives

$$\mathcal{T} = -g\,\left(\mathbf{J}^\top\mathbf{P}\mathbf{J}\right)^{-1} \mathbf{J}^\top\mathbf{P}(s - s^\star) \tag{9}$$

It is easily noted that equation (4) in the unconstrained servo version, is a particular case of the above equation (9), with $\mathbf{P} = Id$.

4 Constrained *vs.* Unconstrained Stereo Servoing

In the previous section we developed a constrained stereo servo task function in which we integrated the epipolar constraint over a set of matched points. However, in most setups the 3D control points are part of a rigid object, fixed on the robot end-effector. Due to this rigid link, the epipolar constraint is implicitly verified (noise and deviations due to image treatments set apart). We formally show in this section, that, in the noiseless case and using rigid control points, both approaches are strictly identical.

Starting from equations (4) and (9) we obtain the following equations:

$$\begin{aligned}
\mathcal{T} &= \left(\mathbf{J}^\top\mathbf{J}\right)^{-1} \mathbf{J}^\top \dot{s} && (10) \\
\mathcal{T}' &= \left(\mathbf{J}^\top\mathbf{P}\mathbf{J}\right)^{-1} \mathbf{J}^\top\mathbf{P}\dot{s} && (11)
\end{aligned}$$

Since s respects the epipolar constraint $h(s) = 0$ by construction, it therefore is invariant under the projection $\mathbf{P}$ onto the kernel of the constraints h: $s = \mathbf{P}\, s$.

$$\begin{aligned}
& h(s) &&= 0 \\
\Rightarrow\; & \tfrac{dh}{dt} &&= 0 \\
\Rightarrow\; & \tfrac{dh}{ds}\dot{s} &&= 0 \\
\Rightarrow\; & \mathbf{G}\mathbf{J}\mathcal{T} &&= 0 \quad \forall \mathcal{T} \\
\Rightarrow\; & \mathbf{G}\mathbf{J} &&= 0 \\
\Rightarrow\; & \mathbf{P}\mathbf{J} &&= \mathbf{J} && (12)
\end{aligned}$$

From which we conclude that

$$\left(\mathbf{J}^\top\mathbf{J}\right)^{-1} \mathbf{J}^\top = \left(\mathbf{J}^\top\mathbf{P}\mathbf{J}\right)^{-1} \mathbf{J}^\top\mathbf{P}$$

□

This shows that the simple least-squares minimization in equation (10) is an as good approximation (and computationally less expensive) as the complete constrained framework developed by equation (11). Indeed, when the the control points used for visual servoing are detected with a sufficiently high precision (in the case of reflecting circular patches, detection precision can go up to 0.05 pixels [1]) the real world conditions are very near noiseless. This will be further shown by the experiments described in section § 5.

The above demonstration also justifies the existing approaches to the subject [9, 4] of which the theoretical soundness had never been proved.

Although both approaches (constrained and unconstrained) are identical in the noiseless case, two fundamental advantages of formally constraining the Jacobian remain:

1. In cases where the task is intrinsically underconstrained (*i.e.* when the DOF controlled by the observed data are too few compared to the DOF controlling the robot) the role of the control points can be reduced to its real dimension, thus allowing the introduction of complementary tasks for filling out the missing degrees of freedom [10].

2. In cases where the task goal $s^\star$ is only roughly computed or when the localization of the control data s is unprecise, the constrained approach will guarantee that the final attained position is correctly constrained with respect to the epipolar geometry and allows for an *a priori* correction of $s^\star$.

5 Experiments

In this section we're showing the results of a series of experiments in a real servoing environment. We're first comparing the improvement that can be expected from using stereo servoing with respect to monocular servoing. In a second series of experiments we show that in a real-world environment where we can detect the 2D control points with sufficient accuracy, the constrained and unconstrained servo control laws give raise to identical behaviour, thus supporting the above demonstration.

5.1 Experimental Setup

In order to observe and quantify the behaviour of a visually controlled robot $\mathcal{R}$ in different circumstances, we used the configuration shown in Figure 3.

We register two 3D positions of the robot, S and $S^\star$. S will serve as the initial position and $S^\star$ as the requested goal position. A rigid stereo rig (rigid in the sense that both cameras are rigidly linked one to another, or, on other terms, that the fundamental matrix of the rig $\mathbf{F}$ is fixed.) observes the robot from a set of different view points $\{P_i\}_{i=1\ldots99}$. The appropriate computer vision algorithms (currently unpublished work, but the reader can refer to [6] for a general outline) allow us to compute the required goal positions $s^\star$ in each of the views at all positions P_i. The precision of these computed goal positions depends of a large number of parameters, and the committed error can be assumed to be Gaussian around the ideal position.

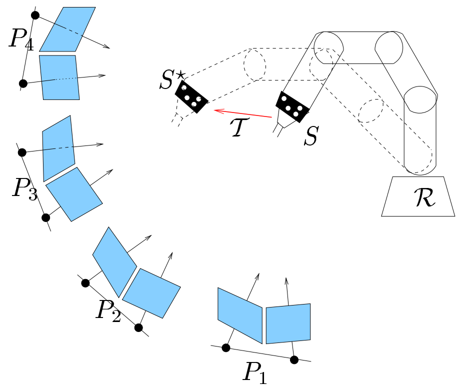

Figure 3: Experimental Platform: the robot $\mathcal{R}$ observed through a rigid stereo rig moves from position S to $S^\star$ under visual servoing for each of the positions P_i of the rig. The absolute 3D positions of S and $S^\star$ being known, it is possible to assess the committed error for each of the positions P_i.

At each of the rig positions P_i we now dispose of sufficient information to servo the robot from S to $S^\star$ using either one of both cameras using a monocular servo loop, or both, using a stereo servo loop.

5.2 Monocular *vs.* Stereo Servoing

In this section we are concerned by quantifying the difference in servo quality between mono and stereo servoing. This quality can be measured through different observations: 2D trajectories, 3D trajectories, 2D convergence quality, 3D convergence quality and movement smoothness.

5.2.1 2D Trajectories

In theory, both approaches, mono and stereo, should result in straight line image trajectories. We can observe, however, that in some cases, especially when the image Jacobian is roughly estimated, image trajectories deviate from this norm. The reasons for this will be discussed in section § 5.2.4. Our experiments have shown that stereo servoing tends to reduce these deviations, as shown in Figure 4.

Figure 4 shows the image trajectories in left and right images, resulting from two monocular servoing experiments and one stereo. Both full and zoomed images are shown.

The stereo trajectories are straight lines (represented for the four control points in *cyan, yellow, magenta* and *bright green*), while the monocular trajectories (in *blue, dark green, cyan* and *red*) present some severe deviations from the straight line image path.

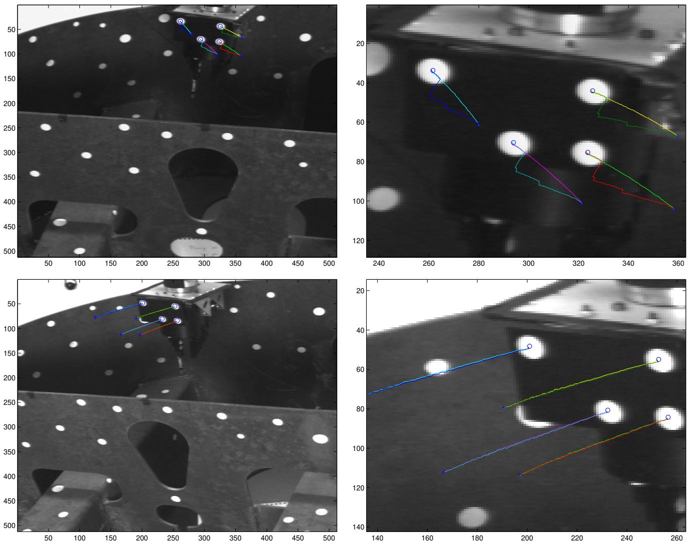

Figure 4: Examples of deviating image trajectories in the image. Upper images come from the left camera, lower images from the right camera. Straight line trajectories come from stereo servoing, skew trajectories (principally left image) come from mono servoing.

These results correspond to the experiment reported in section § 5.2.4 where we show that stereo control naturally smoothes out control commands sent to the robot. The principal reasons for this behaviour will be given there.

5.2.2 2D Convergence Quality

The density plots in Figure 5 give strong indications that the image convergence error is Gaussian for each of the three servo methods. This error is measured with respect to the computed $s^\star$ at convergence.

Note, however, that the 2D convergence error is smaller in the case of monocular servoing. This is due to the fact that the computed goal position $s^\star$ is subject to errors. In most cases, the erroneous position is attainable in either of the images (small 2D convergence error) however it is not *simultaneously* attainable in both. The stereo servoing 2D convergence error is therefore larger, and consists of a minimization compromise between both images.

The measurements, concerning the 3D convergence, presented in the next section, will confirm this, since stereo servoing, due to this minimization, has far better global results where 3D positioning is concerned.

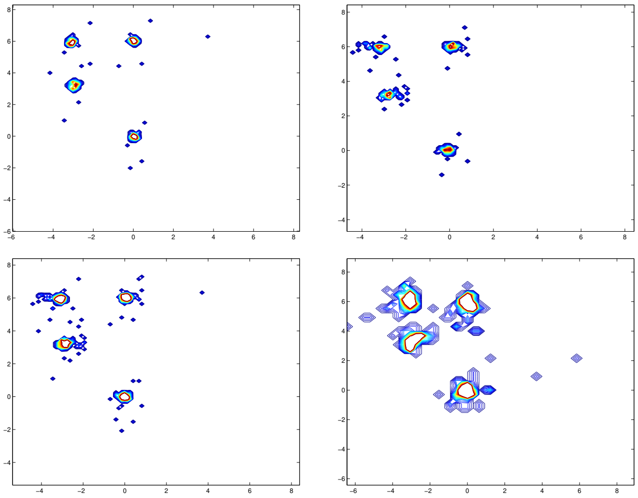

Figure 5: Committed error densities between computed image goal position $s^\star$ and s (representing 4 control points) at convergence. From left to right and top to bottom: left image, right image, cumulated left and right image, cumulated stereo.

5.2.3 3D Convergence Quality

Since the attained 2D goal positions follow a Gaussian distribution, we can assume that the 3D positioning error at convergence is equally Gaussian. Therefore, the error norm follows a χ^2_3 distribution.

Figure 6 shows the 3D Cartesian positioning errors observed for the three experiments over 99 positions P_i. We note that there is a notable improvement in the positioning of the effector when servoing is done with two cameras.

We equally observe the difference in performance between the left and right cameras. This difference is due to the fact that the 3D position of the control points and the lighting conditions during the experiment were in favour of the left camera (less skew, less specularities and better contrast due smaller reflection angles). In order to take into account the bias introduced by these external parameters we should cumulate the committed errors with both left and right cameras when comparing them to those committed in the stereo case. In that case, we obtain de error distribution given in Figure 7.

5.2.4 Movement Smoothness

In this section we analyze the kinematic screw sent to the robot under visual servoing. Figure 8 shows that the resulting kinematic screws for the same experiment as reported in section § 5.2.1 represent a number of irregularities in the monocular case that

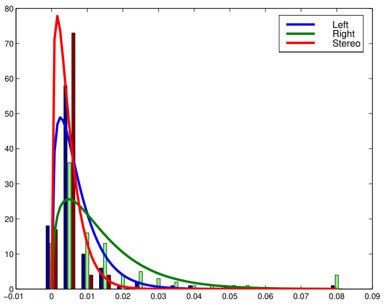

Figure 6: 3D Cartesian Convergence Quality: histogram of committed 3D positioning errors at convergence with, superposed, the corresponding estimated χ^2 distribution. x-axis is expressed in meters.

The mean and variance for the corresponding experiments are given in the following table:

	μ	σ
Left	0.0074	$1.1\ 10^{-4}$
Right	0.0142	$3.6\ 10^{-4}$
Stereo	0.0047	$2.9\ 10^{-5}$

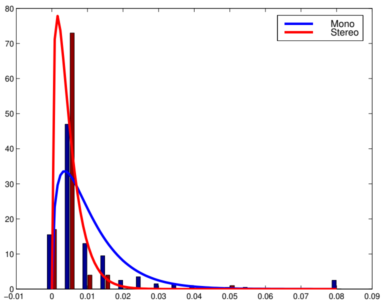

Figure 7: 3D Cartesian Convergence Quality: histogram of committed 3D positioning errors at convergence with, superposed, the corresponding estimated χ^2 distribution. x-axis is expressed in meters.

The mean and variance for the corresponding experiments are given in the following table:

	μ	σ
Mono	0.0108	$2.5\ 10^{-4}$
Stereo	0.0047	$2.9\ 10^{-5}$

are absent in the stereo case.

The following remarks need to be made in order to correctly interpret the graphs in Figure 8:

1. We used a variable Jacobian approach instead of a fixed Jacobian. This means that at each loop iteration $\mathbf{J}$ is re-estimated according to the image data. This guarantees, in the noiseless case, an absolute straight image trajectory for all approaches.

2. We implemented a variable gain for speed up at convergence. This explains why, between iterations 60–70, the kinematic screw increases again. This also explains the amplitude of the oscillations that occur near convergence.

3. No filtering has been used on the screw values, in order to easily observe the differences between the experimented methods.

The most notable difference between the monocular and stereoscopic approaches is the absence of sharp peaks in the latter. These peaks are due to numerical instabilities that occur when re-estimating the Jacobian in the monocular case. This re-estimation requires a pose computation that can sometimes be fooled into believing that the observed points have flipped around a plane parallel to the image plane. Use of redundant information in the stereo case filters this out, resulting in smoother velocities and trajectories.

5.3 Constrained *vs.* Unconstrained Stereo Servoing

This section briefly shows how we can observe the equivalence between the constrained and unconstrained case when the image control points can be extracted with sufficient precision.

Figure 9 compares the 3D cartesian error commited over the same 99 experiments reported previously. The left part superposes the constrained and unconstrained absolute errors. For readability, the right part plots the difference between the two superposed graphs. Note the order of magnitude, which is below the millimeter (y-scale is in meters).

Figure 10 plots the kinematic screw applied over one particular experiment (the same as the one reported in § 5.2.1 and § 5.2.4). One clearly observes the near perfect superposition of the curves.

6 Conclusions

In this paper we have shown both formally and experimentally that the pseudo-inverse of a stereo Jacobian matrix, obtained by the stacking of two monocular Jacobian matrices implicitly projects onto the epipolar-constraint space. This means that, as long as the ob-

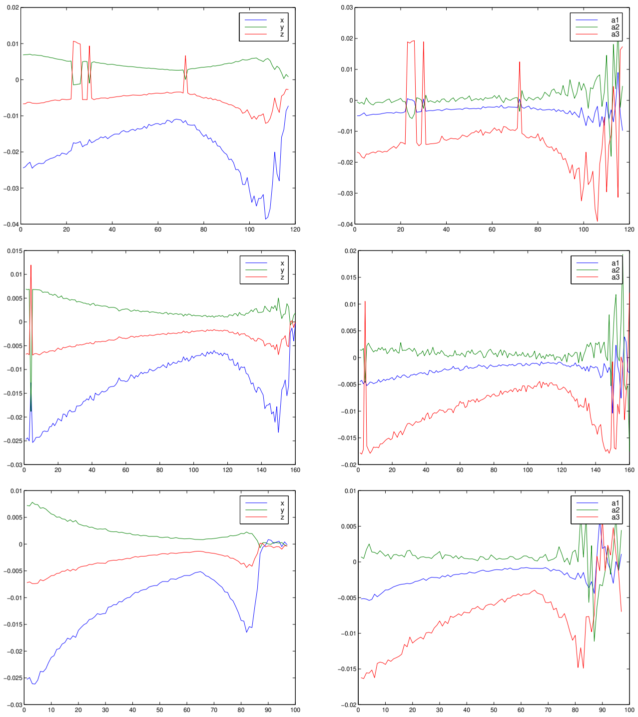

Figure 8: Kinematic screws for left (*top*), right (*middle*) and stereo (*bottom*) servoing. Translational speed is given on the left, in $m.s^{-1}$, rotational speed, on the right, is in $deg.s^{-1}$.

served control points are sufficiently precise, it is unnecessary to explicitly take into account constrained minimization.

We have also shown that the use of stereoscopic servoing holds a real advantage over monocular servoing in all domains: 3D trajectory, movement smoothness and 3D convergence.

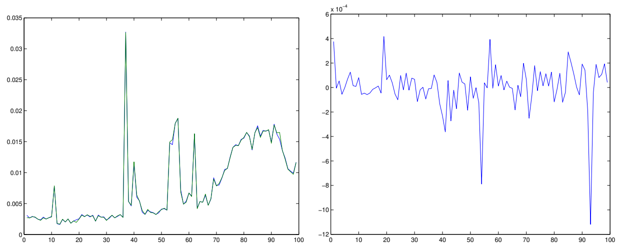

Figure 9: 3D Cartesian convergence error for constrained and unconstrained stereo servoing (*Left*) and their difference (*Right*) over 99 experiments. y-scale is in meters.

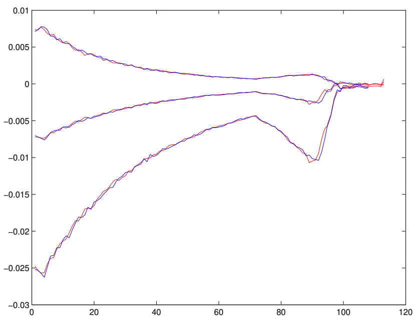

Figure 10: Translational part of the kinematic screw applied over one particular experiment: unconstrained (*red*) and constrained (*blue*) servoing.

References

[1] P. Brand. *Reconstruction tridimensionnelle d'une scène à partir d'une caméra en mouvement : de l'influence de la précision*. Thèse de doctorat, Université Claude Bernard, Lyon I, October 1995.

[2] F. Chaumette. Potential problems of stability and convergence in image-based and position-based visual servoing. In *Workshop on Vision and Control, Block Island, Rhode Island,*, June 1997.

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

[4] G.D. Hager, W. Chang, and A.S. Morse. Robot hand-eye coordination based on stereo vision. IEEE *Control Systems*, page 30, 1995.

[5] R. Horaud, F. Dornaika, and B. Espiau. Visually guided object grasping. IEEE *Transactions on Robotics and Automation*, 1997.

[6] B. Lamiroy. Description of demonstrators, INRIA uncalibrated visual servoing prototype. Public Deliverable D4.1.c, INRIA Rhône-Alpes, VIGOR, Esprit-IV reactive LTR project, number 26247, 1999.

[7] H.C. Longuet-Higgins. A computer program for reconstructing a scene from two projections. *Nature*, 293:133–135, September 1981.

[8] Q.T. Luong and O. Faugeras. The fundamental matrix: Theory, algorithms and stability analysis. *International Journal of Computer Vision*, 17(1):43–76, 1996.

[9] N. Maru, H. Kase, S. Yamada, and A. Nishikawa. Manipulator control by visual servoing with the stereo vision. In *International Conference on Intelligent Robots and Systems, Yokohama, Japan*, pages 1866–1870. IEEE, July 1993.

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