



**HAL**  
open science

# Uncalibrated active affine reconstruction closing the loop by visual servoing

Ezio Malis, P. Rives

► **To cite this version:**

Ezio Malis, P. Rives. Uncalibrated active affine reconstruction closing the loop by visual servoing. 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453), Oct 2003, Las Vegas, United States. pp.498-503, 10.1109/IROS.2003.1250678 . hal-04647493

**HAL Id: hal-04647493**

**<https://hal.science/hal-04647493v1>**

Submitted on 14 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Uncalibrated Active Affine Reconstruction closing the loop by Visual Servoing

Ezio Malis and Patrick Rives

I.N.R.I.A. Sophia Antipolis  
2004, route des Lucioles - B.P. 93  
06902 Sophia Antipolis Cedex, France.

**Abstract**—This paper presents a new approach to active affine reconstruction without an exact knowledge of the robot's kinematic model nor camera intrinsic parameters. Affine reconstruction from perspective image pairs is easy if the motion of the camera between the two images is a pure translation. If the robot is not well calibrated, a pure translation achieved with an open loop control would lead to a bias in the reconstruction. The problem can be solved by using a 2 1/2 D visual servoing technique in order to close the loop and control the camera trajectory. The affine reconstruction is equivalent to the estimation of the depths of the 3D points of the scene. Thus, affine reconstruction is very useful to implement several visual servoing approaches which need an estimation of the depths.

## I. INTRODUCTION

The design of many vision-based control approaches [1] often demands some a priori knowledge on the geometry of the environment. For example, in order to implement an image-based visual servoing using image points as visual features [2] we need to know the depths of all the points (i.e. the depth distribution). Even if the depths can be approximated, it has been shown in [3] that errors on the depth distribution can make the vision-based control unstable. It is therefore an important issue to obtain a good estimation of the information needed to implement a vision-based control law. Affine reconstruction is a possible solution to the problem since it allows to recover the depth distribution of a set of points. In contrast to full Euclidean reconstruction [4], affine reconstruction need weaker assumptions on the system. Firstly, there is no need to accurately calibrate the robot or the camera. Secondly, only two views of the same set of points are sufficient. Moreover, it is well known that affine reconstruction from perspective image pairs is very easy if the motion of the camera between the two images is a pure translation [5] (the direction and the distance of the translation can be arbitrary and unknown). Obviously, a pure translation can be obtained if the camera is mounted on a well calibrated robot manipulator. On the other hand, if the robot is not well calibrated or mobile (i.e. with unavoidable drifting) it is not possible to perform an open-loop pure translation. The goal of this paper is to propose an active affine reconstruction method in which the pure translation is performed closing the loop with a visual servoing tech-

nique. The robot and the camera are not well calibrated and the model of environment is completely unknown. Thus, standard position-based visual servoing cannot be used [6]. Note that, a reference image is not available to achieve the task. On the other hand, some epipolar geometry constraints can be used in a visual servoing scheme as proposed in [7]. The key idea is to define some constraints that the final (but unknown) image points must verify in order to have a pure translation between the initial and final images. However, the visual servoing method proposed in [7] cannot be used since it also needs the knowledge of the depth distribution. Instead, we use the 2 1/2 D visual servoing [8] since it does not need any knowledge on the depth distribution. Using such a technique, the active affine reconstruction does not need any accurate robot calibration. The visual servoing is performed without any model of the object nor any reference image. Only the current and initial images are used to control the camera. Once the pure translation has been achieved the affine reconstruction is straightforward.

The paper is organized as follows. In Section II we review some theoretical background on projective geometry. In Section III we describe which constraints on the unknown final image define a pure translation of the camera. Section IV presents the control law used to close the loop. Finally, the simulations described in Section V prove the validity of the proposed approach.

## II. THEORETICAL BACKGROUND

We consider in the paper an object which can be described by a set of 3D points. Let  $\mathcal{F}_0$  be a frame attached to the object. The homogeneous coordinates, with respect to  $\mathcal{F}_0$ , of a discrete set of  $n$  3D points are  $\mathcal{X}_i = (X_i, Y_i, Z_i, 1)$  ( $i = \{1, 2, \dots, n\}$ ).

### A. One view geometry

Let  $\mathcal{C}$  be the center of projection coinciding with the origin of the camera frame  $\mathcal{F}$ . Let the image plane be parallel to the plane  $(\vec{x}, \vec{y})$ . A 3D point  $\mathcal{X}_i$  is projected to the image point  $\mathbf{x}_i$ :

$$\mathbf{x}_i = \frac{1}{Z_i} \mathbf{K} \begin{bmatrix} \mathbf{R}_0 & \mathbf{t}_0 \end{bmatrix} \mathcal{X}_i = (x_i, y_i, 1) \quad (1)$$

where  $\mathbf{R}_0$  and  $\mathbf{t}_0$  are respectively the rotation and the translation between frame  $\mathcal{F}_0$  and  $\mathcal{F}$ , and the triangular matrix  $\mathbf{K}$  contains the camera internal parameters:

$$\mathbf{K} = \begin{bmatrix} f & sf & u_0 \\ 0 & rf & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

$f$  is the focal length (pixels),  $u_0$  and  $v_0$  are the coordinates of the principal point (pixels),  $s$  is the skew and  $r$  is the aspect ratio.

### B. Two view geometry

The fundamental relationship between the  $i$ -th point in two different images is:

$$\frac{Z_{2i}}{Z_{1i}} \mathbf{x}_{2i} = {}^2\mathbf{H}_1^\infty \left( \mathbf{x}_{1i} + \frac{1}{Z_{1i}} \mathbf{c}_1 \right) \quad (3)$$

where  ${}^2\mathbf{H}_1^\infty = \mathbf{K} {}^2\mathbf{R}_1 \mathbf{K}^{-1}$  is the homography of the plane at infinity and  $\mathbf{c}_1 \propto \mathbf{K} \mathbf{t}$  is the epipole in the first image. Equation (3) represents a set of non-linear equations in the unknowns  ${}^2\mathbf{H}_1^\infty$ ,  $\mathbf{c}_1$ ,  $\frac{Z_{2i}}{Z_{1i}}$  and  $\frac{1}{Z_{1i}}$ . If we can estimate  $\frac{1}{Z_{1i}}$  up to a scale factor (i.e. we have done an affine reconstruction), the relationship between the unknowns  ${}^2\mathbf{H}_1^\infty$ ,  ${}^2\mathbf{H}_1^\infty \mathbf{c}_1$  and  $\frac{Z_2}{Z_1}$  becomes linear. Thus, one benefit of estimating an affine reconstruction of the object will be a linear reconstruction algorithm for the camera pose. Note also that equation (3) can be rewritten as:

$$\frac{Z_{1i}}{Z_{2i}} \mathbf{x}_{1i} = {}^1\mathbf{H}_2^\infty \left( \mathbf{x}_{2i} + \frac{1}{Z_{2i}} \mathbf{c}_2 \right) \quad (4)$$

where  ${}^1\mathbf{H}_2^\infty = {}^2\mathbf{H}_1^\infty^{-1}$  and  $\mathbf{c}_2 \propto {}^2\mathbf{H}_1^\infty \mathbf{c}_1$  is the epipole in the second image.

### C. Affine reconstruction

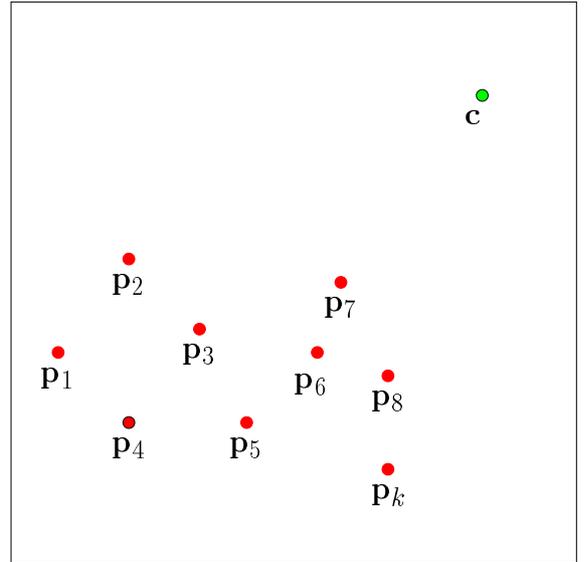
As already mentioned, affine reconstruction is easy if the camera motion is a pure translation (i.e.  ${}^2\mathbf{R}_1 = \mathbf{I}$ ). In that case, equation (3) is simplified since  ${}^2\mathbf{H}_1^\infty = \mathbf{I}$ :

$$\frac{Z_{2i}}{Z_{1i}} \mathbf{x}_{2i} = \mathbf{x}_{1i} + \frac{1}{Z_{1i}} \mathbf{c}_1 \quad (5)$$

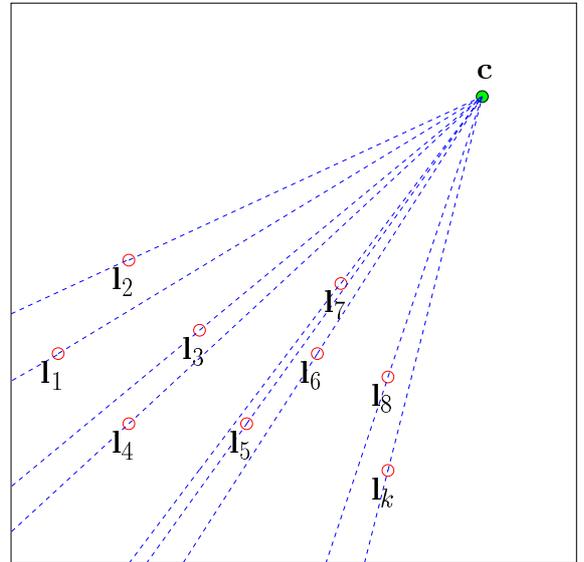
Obviously, if the camera motion is a pure translation then  $\mathbf{c}_2 \propto \mathbf{c}_1$ . If we are able to perform a desired pure translation (i.e.  $\mathbf{c}_1$  is known), then the set of equations becomes linear in the unknowns  $\frac{1}{Z_{1i}}$  and  $\frac{Z_{2i}}{Z_{1i}}$ .

## III. CONSTRAINTS FOR A PURE TRANSLATION

In this paper, we propose to perform a pure translation (starting from any initial position) closing the loop with visual servoing. In order to do this, we need to define some constraints since we do not have any reference image. First of all, the rotation between the initial and the final position must be the identity matrix. Then, by choosing the epipole  $\mathbf{c}_1$  in the first image (see Figure 1(a)) one can fix the direction of translation (indeed  $\mathbf{c}_1 \propto \mathbf{K} \mathbf{t}$  and  $\mathbf{t} \propto \mathbf{K}^{-1} \mathbf{c}_1$ ).



(a) Initial image



(b) Unknown final image

Fig. 1. Two view geometry for a pure translation.

As a consequence, the epipole  $\mathbf{c}_2$  in the second image is constrained to be the same as  $\mathbf{c}_1$  if the camera motion is a pure translation (see Figure 1(b)). Note that the red points in Figure 1(b) are not the reference points (we do not have any reference image) but simply the same points of Figure 1(a). The points in the unknown final image must belong to the epipolar lines (the blue lines in Figure 1(b)) defined by:

$$\mathbf{l}_{2i} = \mathbf{x}_{1i} \times \mathbf{c}_1 \propto (\cos(\varphi_i), \sin(\varphi_i), -\gamma_i)$$

The point  $\mathbf{x}_{2i}$  in the final image can be anywhere on the corresponding epipolar line such that  $\mathbf{l}_{2i}^\top \mathbf{x}_{2i} = 0$ . Despite a reference image cannot be defined, one reference point  $\mathbf{x}_{20}^*$  can be arbitrarily fixed on the epipolar line and this will fix the distance of the translation. Indeed, equation (5)

can be written for  $i = 0$  as:

$$\frac{Z_{20}^*}{Z_{10}} \mathbf{x}_{20}^* = \mathbf{x}_{10} + \frac{1}{Z_{10}} \mathbf{c}_1 \quad (6)$$

Setting  $\rho_{20} = \frac{Z_{20}^*}{Z_{10}}$ , it is possible to compute a reference:

$$\rho_{20}^* = \frac{\|\mathbf{c}_1 \times \mathbf{x}_{10}\|}{\|\mathbf{c}_1 \times \mathbf{x}_{20}^*\|} \quad (7)$$

where  $\times$  represents the cross product between two vectors. Therefore, the distance of the translation in the Cartesian space is fixed (even if unknown):  $\frac{1}{Z_{10}} \|\mathbf{c}_1\| = \|\rho_{20}^* \mathbf{x}_{20}^* - \mathbf{x}_{10}\|$ . The reference point  $\mathbf{x}_{20}^*$  can be fixed by computing the epipolar line:

$$\mathbf{l}_{20} = \mathbf{x}_{10} \times \mathbf{c}_1 \quad (8)$$

If  $\mathbf{x}_{20}^*$  belongs to  $\mathbf{l}_{20}$  then:

$$\mathbf{l}_{20} \bullet \mathbf{x}_{20}^* = \mathbf{l}_{20}^\top \mathbf{x}_{20}^* = 0$$

Let vector  $\boldsymbol{\tau}_{20} = (\sin(\varphi), -\cos(\varphi), 0)$  be the tangent vector to the line  $\mathbf{l}_{20}$ . The reference point  $\mathbf{x}_{20}^*$  in the final image is:

$$\mathbf{x}_{20}^* = \mathbf{x}_{10} + \Delta \mathbf{x} \boldsymbol{\tau}_{20}$$

where  $\Delta \mathbf{x}$  represents the displacement along the epipolar line (if  $\Delta \mathbf{x} = 0$  then  $\mathbf{x}_{20}^* = \mathbf{x}_{10}$ ).

#### IV. THE CONTROL LAW

The constraints defined in the previous section can be imposed by using a  $2\frac{1}{2}$ D visual servoing technique. At each iteration of the control law we can compute an estimation of the rotation  ${}^2\mathbf{R}_1$  between the current and the initial image [8]. This rotation must be the identity if we want to perform a pure translation. At the first iteration the current and the initial image coincide and the rotation is indeed the identity. However, due to bad calibration of the robot the rotation must be kept small by minimizing the error of rotation  $\mathbf{u}\theta$ , where  $\mathbf{u}$  and  $\theta$  are respectively the axis and angle of rotation extracted from  ${}^2\mathbf{R}_1$ . It is important to notice that in order to obtain an exact estimation of the rotation the camera parameters are needed. An error on camera calibration will affect the measure of  ${}^2\mathbf{R}_1$  unless we are able to obtain  ${}^2\mathbf{H}_1$ . The reference epipole  $\mathbf{c}_1$  can be fixed by the user. If the camera is perfectly calibrated this is equivalent to fix the direction of the translation since  $\mathbf{t} \propto \mathbf{K}^{-1} \mathbf{c}_1$ . Once the epipole is fixed, we choose one of the points ( $\mathbf{x}_{10}$ ) in the first image. From equation (8) it is possible to compute the epipolar line passing through that point. As already mentioned in the previous section, we can set a reference  $\mathbf{x}_{20}^*$  and  $\rho_{20}^*$ . It is also possible to define a sequence of reference points along the epipolar line by setting  $\Delta \mathbf{x}(t)$  such that  $\Delta \mathbf{x}(0) = 0$  and  $\Delta \mathbf{x}(T) = \Delta \mathbf{x}$ . Thus, we obtain a reference trajectory for the point:

$$\mathbf{x}_{20}^*(t) = \mathbf{x}_{10} + \Delta \mathbf{x}(t) \boldsymbol{\tau}_{20}$$

Similarly, it is possible to define a reference trajectory  $\rho_{20}^*(t)$  by solving equation (6)  $\forall \mathbf{x}_{20}^*(t)$ :

$$\rho_{20}^*(t) = \frac{\|\mathbf{c}_1 \times \mathbf{x}_{10}\|}{\|\mathbf{c}_1 \times \mathbf{x}_{20}^*(t)\|}$$

where  $\rho_{20}^*(0) = 1$ . Note that we must impose  $\rho_{20}^*(t) \neq 0$ , thus  $\mathbf{x}_{20}^*(t) \neq \mathbf{c}_1 \quad \forall t$ . Let us define the current signal  $\mathbf{s} = (\mathbf{x}_{20}, \log(\rho_{20}), \mathbf{u}\theta)$ . The corresponding reference signal is  $\mathbf{s}^*(t) = (\mathbf{x}_{20}^*(t), \log(\rho_{20}^*(t)), \mathbf{0})$ . Thus, the camera can be controlled by defining the following task function:

$$\mathbf{e} = \mathbf{s} - \mathbf{s}^*(t) = (\mathbf{x}_{20} - \mathbf{x}_{20}^*(t), \log(\rho_{20}/\rho_{20}^*(t)), \mathbf{u}\theta)$$

The derivative of the task function is:

$$\dot{\mathbf{e}} = \mathbf{L} \mathbf{v} - \frac{\partial \mathbf{s}^*(t)}{\partial t}$$

where the  $(6 \times 6)$  matrix  $\mathbf{L}$  is the interaction matrix [8] and  $\mathbf{v} = (\boldsymbol{\nu}, \boldsymbol{\omega})$  the velocity of the camera. The interaction matrix depends on the unknown camera intrinsic parameters  $\mathbf{K}$ , and on the unknown depth  $Z_{10}$  in the initial frame. In order to control the movement of the camera we impose an exponential decreasing of the task function  $\dot{\mathbf{e}} = -\lambda \mathbf{e}$ , where  $\lambda$  is a positive scalar tuning the speed of the convergence. Thus, we use the following control law:

$$\mathbf{v} = -\lambda \hat{\mathbf{L}}^{-1} \mathbf{e} + \hat{\mathbf{L}}^{-1} \frac{\partial \mathbf{s}^*(t)}{\partial t} \quad (9)$$

where  $\hat{\mathbf{L}}$  is an approximation of the interaction matrix. Using this control law, the closed-loop equation is:

$$\dot{\mathbf{e}} = -\lambda \mathbf{L} \hat{\mathbf{L}}^{-1} \mathbf{e} + (\mathbf{I} - \mathbf{L} \hat{\mathbf{L}}^{-1}) \frac{\partial \mathbf{s}^*(t)}{\partial t}$$

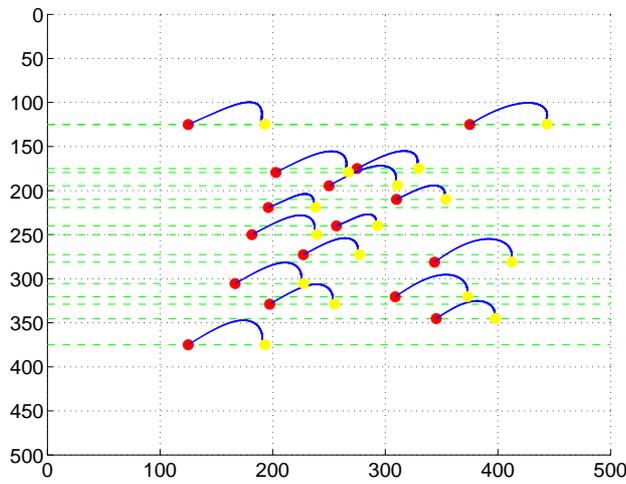
It is well known from control theory that if  $\mathbf{L} \hat{\mathbf{L}}^{-1} > 0$  and  $\lim_{t \rightarrow \infty} \frac{\partial \mathbf{s}^*(t)}{\partial t} = 0$  then the task function  $\mathbf{e}$  converge to zero and the signal  $\mathbf{s}$  converges to  $\mathbf{s}^*$ . Note that, since the reference trajectory  $\mathbf{s}^*(t)$  is fixed by the user, it can always be chosen bounded and such that  $\lim_{t \rightarrow \infty} \frac{\partial \mathbf{s}^*(t)}{\partial t} = 0$ . Control issues (for example the proof of the robustness of the control law) are beyond the aim of this paper and they have already been solved in previous work [8]. However, it must be underlined that we do not need an exact calibration of  $\mathbf{K}$  nor  $Z_{10}$  for the convergence of the servoing. For simplicity, we consider in the experiments only the case when the references  $\mathbf{x}_{20}^*$  and  $\rho_{20}^*$  are constant. In that case,  $\frac{\partial \mathbf{s}^*(t)}{\partial t} = 0$  in equation (9). The final results will be the same for the final affine reconstruction.

#### V. EXPERIMENTAL RESULTS

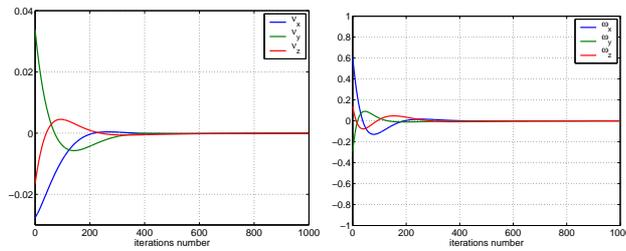
The objective of the experiments is to achieve pure translations of a camera mounted on an uncalibrated 6 d.o.f. manipulator arm. The camera observe a set of  $n = 16$  points randomly distributed in the 3D space. We suppose that we do not know the model of the object (thus, the translation cannot be achieved by a standard position-based control law [6]). Moreover, the robot and the camera are both badly calibrated. Thus, an open loop control of the rotation and the translation would lead to incorrect results. We use instead the control law proposed in the previous section. For all experiments, the red circles in the image represents the points in the initial position while the yellow circles represents the points in the final (and initially unknown) image. The dashed green lines represents the epipolar lines.

### A. Simple translations

In order to achieve a simple pure translation along the  $\vec{x}$  axis one must fix the epipole in the initial image at infinity:  $\mathbf{c} = (1, 0, 0)$ . Once the direction of translation has been defined, the amplitude of the displacement can be fixed in the image. In this experiment, we fixed the displacement of the reference point to 80 pixels. Had the robot and the camera been perfectly calibrated the trajectory of the points in the image would have followed the green epipolar lines in Figure 2. Instead, the trajectory of the points in the image follows the blue lines in Figure 2. However, at the end of the servoing, the current point  $\mathbf{x}_{20}$  coincide with  $\mathbf{x}_{20}^*$  and all the points belongs to the epipolar lines. The rotational error in Figures 2(d) converge to zero. Thus, the camera motion is a pure translation and the depth distribution can be easily computed up to a scalar factor.

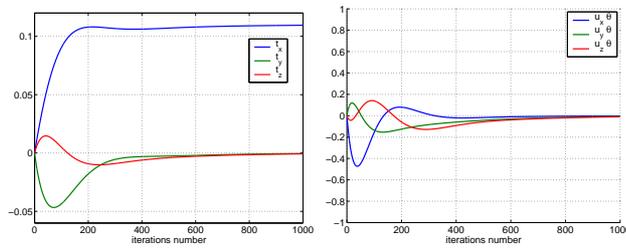


(a) Trajectory of the points in the image



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

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

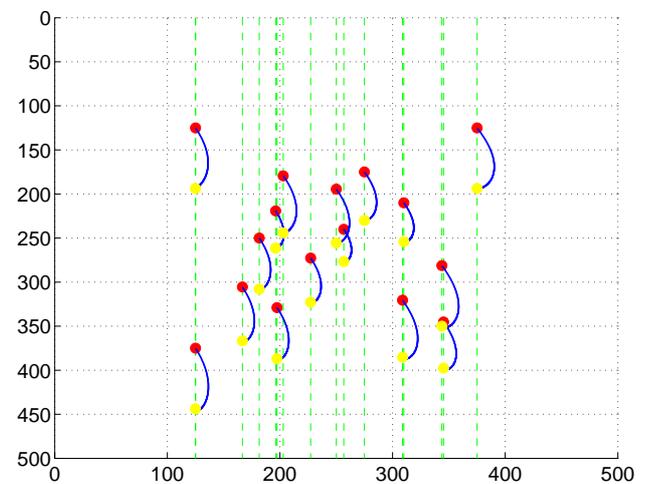


(d) Translation error ( $m$ )

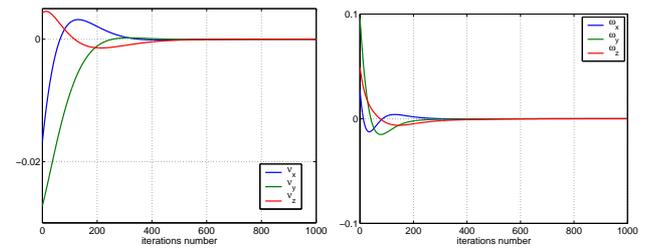
(e) Rotation error ( $deg$ )

Fig. 2. Translation along  $\vec{x}$ .

A simple pure translation along the  $\vec{y}$  axis can be obtained by fixing the epipole at infinity:  $\mathbf{c} = (0, 1, 0)$ . In this experiment, we fixed again the displacement of the reference point to 80 pixels. Figure 3 shows similar results to those of the previous experiment. The starting position is the same but now the pure translation is along the  $\vec{y}$  axis. Finally, in order to achieve a simple pure translation along the  $\vec{z}$  axis one must fix the epipole to be the principal point of the image:  $\mathbf{c} = (u_0, v_0, 1)$ . If the camera is not calibrated, the translation will not be exactly in the  $\vec{z}$  direction. In this experiment, we suppose to exactly know the principal point and we fixed the displacement of the reference point to 40 pixels. The effect of bad calibration on the position error visible in Figures 4(d) and (e) are compensated by the control law given in Figure 4(b) and (c).

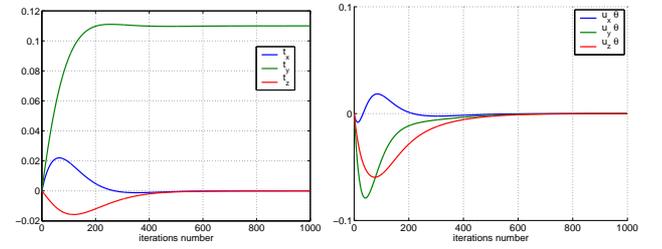


(a) Trajectory of the points in the image



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

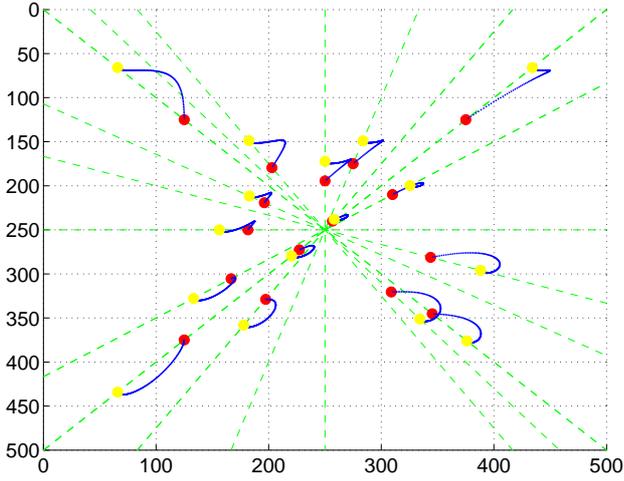
(c) Rotation speed ( $\frac{deg}{s}$ )



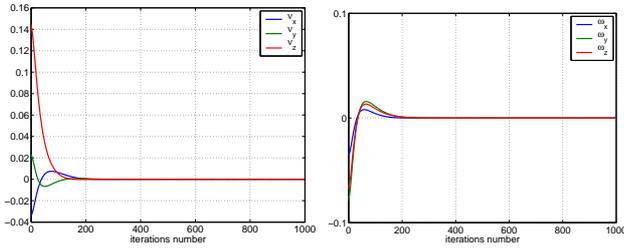
(d) Translation error ( $m$ )

(e) Rotation error ( $deg$ )

Fig. 3. Translation along  $\vec{y}$ .

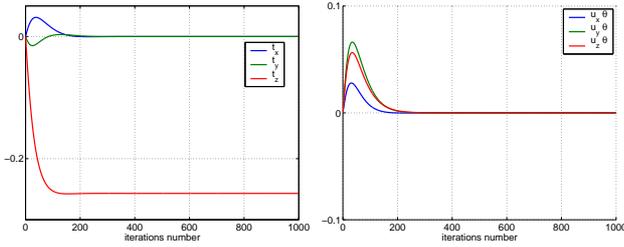


(a) Trajectory of the points in the image



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

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



(d) Translation error ( $m$ )

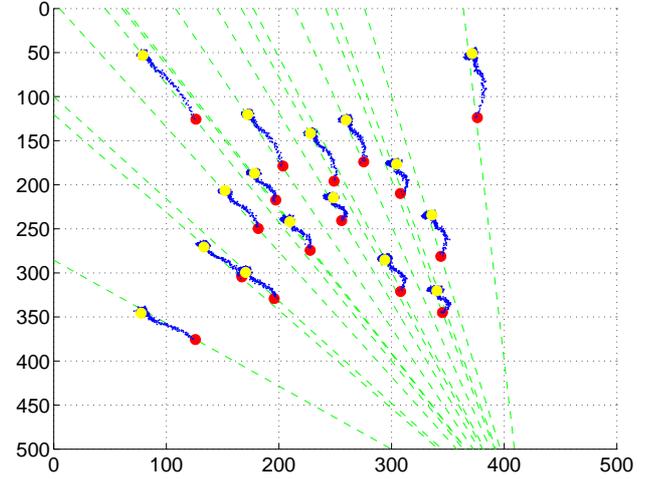
(e) Rotation error ( $deg$ )

Fig. 4. Translation along  $\vec{z}$ .

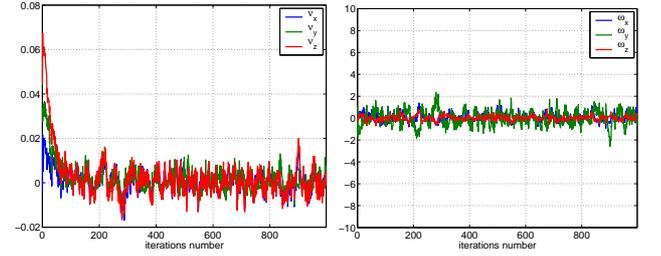
### B. Generic translation

In this experiment, the translation is generic and the epipole is chosen out of the image  $\mathbf{c}_1 = (417, 583, 1)$ . The corresponding true direction of translation is  $\mathbf{t} \propto (0.27, 0.53, 0.80)$ . Obviously, we can exactly choose the true direction of translation only if camera intrinsic parameters are perfectly known. In this experiment, we use a rough approximation of them:  $\hat{f} = 800$ ,  $\hat{u}_0 = 275$ ,  $\hat{v}_0 = 225$  instead of  $f = 500$ ,  $u_0 = 250$ ,  $v_0 = 250$ . Thus the robot move only approximately in the desired direction. However, since the goal is to perform any pure translation, we can fix a direction even if the camera intrinsic parameters are completely unknown. In this experiment, a Gaussian noise with standard deviation  $\sigma = 0.1$  pixels is added to image point coordinates. As in the previous experiments, the rotation is controlled in order to perform a pure translation. Without such a control, noise and robot

calibration errors would cause large estimation errors on the depth distribution. The servoing is stopped when the estimated rotation is less than 0.1 degrees, the error on the image coordinates of the reference point is less than 0.1 pixels and the error on the depth ratio is less than 0.001. At the end of the servoing, the camera displacement is practically a pure translation. At the end of the servoing, the rotational error in Figures 5(d) is closed to zero. Thus, the motion of the camera is practically a pure translation.

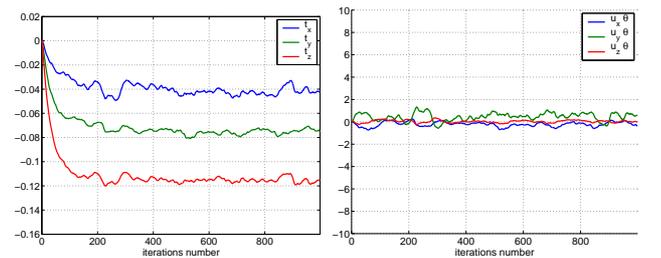


(a) Trajectory of the points in the image



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

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



(d) Translation error ( $m$ )

(e) Rotation error ( $deg$ )

Fig. 5. Generic translation with image noise.

### C. Affine reconstruction

The experiment in the previous section has been repeated 10 times with image noise standard deviation increasing from  $\sigma = 0$  to  $\sigma = 1$  pixels. At the end of the servoing, the depth distributions for the initial and final positions are computed solving the linear system:

$$Z_{2i}\mathbf{x}_{2i} = Z_{1i}\mathbf{x}_{1i} + \mathbf{c}_1 \quad i \in \{1, 2, 3, \dots, n\} \quad (10)$$

From equation (10) we compute the depth distribution  $\mathbf{z}_1$  and  $\mathbf{z}_2$  up to scale in the initial and final camera frames:

$$\mathbf{z}_1 = k_1(Z_{11}, Z_{12}, \dots, Z_{1n}) \quad (11)$$

$$\mathbf{z}_2 = k_2(Z_{21}, Z_{22}, \dots, Z_{2n}) \quad (12)$$

The scalar factors  $k_1$  and  $k_2$  are eliminated by computing “normalized” depth distributions:

$$\mathbf{d}_1 = (1, Z_{12}/Z_{11}, \dots, Z_{1n}/Z_{11}) \quad (13)$$

$$\mathbf{d}_2 = (1, Z_{22}/Z_{21}, \dots, Z_{2n}/Z_{21}) \quad (14)$$

where  $Z_{11}$  and  $Z_{21}$  are the estimated depth of the first point (any point can be used to normalize the distribution). The normalized distributions are computed 1000 times in order to average the estimation errors due to image noise. Finally, we compute the relative error:

$$e_1 = \frac{100}{n} \sum_{i=1}^n \frac{|\hat{\mathbf{d}}_{1i} - \mathbf{d}_{1i}|}{\mathbf{d}_{1i}}$$

$$e_2 = \frac{100}{n} \sum_{i=1}^n \frac{|\hat{\mathbf{d}}_{2i} - \mathbf{d}_{2i}|}{\mathbf{d}_{2i}}$$

and its mean and standard deviation over all trials. Figure 6 plots the results for a displacement of 40 pixels in the image. The reconstruction is very accurate since the mean error is less than 5% in the worst case  $\sigma = 1$ . The mean relative error (the blue line) increases (almost linearly) with image noise as well as the standard deviation (the red dashed lines). The accuracy of the reconstruction also depends on the displacement in the image. Figure 7 plots the results for a displacement of 80 pixels in the image. In this experiment, the mean error is less than 4% in the worst case  $\sigma = 1$ . The bigger is the disparity in the image the more accurate is the estimation of the depth distribution. However, the disparity in the image is bounded by the size of the CCD. In order to improve the estimation of the depth distribution, it is possible to update the affine reconstruction to an Euclidean reconstruction. In that case, several views of the same points are needed and an online camera self-calibration must be computed.

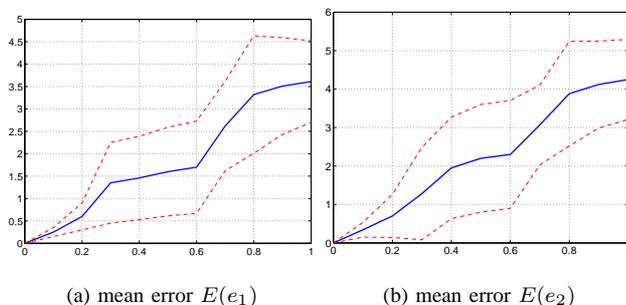
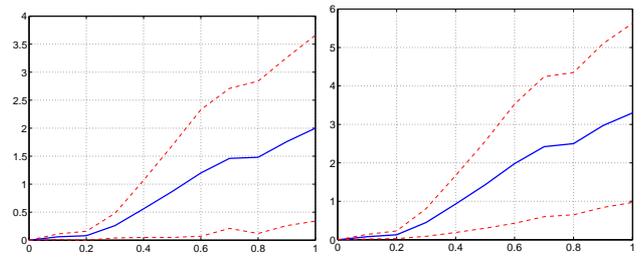


Fig. 6. Mean relative error on the depths in the initial (a) and final (b) positions (blue lines) for a displacement of 40 pixels. The red dashed lines represent the standard deviation.



(a) mean error  $E(e_1)$

(b) mean error  $E(e_2)$

Fig. 7. Mean relative error on the depths in the initial (a) and final (b) positions (blue lines) for a displacement of 80 pixels. The red dashed lines represent the standard deviation.

## VI. CONCLUSION

This paper has shown that it is possible to perform pure translation by visual servoing without any reference image nor any knowledge on the structure of the scene. When the camera displacement is a pure translation the affine reconstruction of the scene is straightforward. Once the depth of the points in the scene has been estimated once and for all, it can be used to improve visual servoing schemes or as an initial guess for full 3D reconstruction of the scene. Ongoing work is devoted to implementing the algorithm on a real robot.

## VII. REFERENCES

- [1] S. Hutchinson, G. D. Hager, and P. I. Corke, “A tutorial on visual servo control” *IEEE Trans. on Rob. and Aut.*, vol. 12, no. 5, pp. 651–670, October 1996.
- [2] B. Espiau, F. Chaumette, and P. Rives, “A new approach to visual servoing in robotics” *IEEE Trans. on Robotics and Automation*, vol. 8, no. 3, pp. 313–326, June 1992.
- [3] E. Malis and P. Rives, “Robustness of image-based visual servoing with respect to depth distribution errors” in *IEEE International Conference on Robotics and Automation*, Taipei, Taiwan, September 2003.
- [4] E. Marchand and F. Chaumette, “Active vision for complete scene reconstruction and exploration” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 1, pp. 65–72, January 1999.
- [5] T. Moons, L. Van Gool, M. Proesmans, and E. Pauwels, “Affine reconstruction from perspective image pairs with a relative object-camera translation in between”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, no. 1, pp. 77–83, 1996.
- [6] W. J. Wilson, C. C. W. Hulls, and G. S. Bell, “Relative end-effector control using cartesian position-based visual servoing” *IEEE Trans. on Robotics and Automation*, vol. 12, no. 5, pp. 684–696, 1996.
- [7] P. Rives, “Visual servoing based on epipolar geometry” in *IEEE Int. Conf. on Intelligent Robots and Systems*, Takamatsu, Japan, November 2000, vol. 1, pp. 602–607.
- [8] E. Malis, F. Chaumette, and S. Boudet, “2 1/2 d visual servoing,” *IEEE Trans. on Robotics and Automation*, vol. 15, no. 2, pp. 234–246, April 1999.