

Eye-in-hand / Eye-to-hand Cooperation for Visual Servoing

Grégory Flandin François Chaumette Eric Marchand
IRISA - INRIA Rennes
Campus de Beaulieu,
35042 Rennes Cedex, France
e-mail: {Firstname.Name@irisa.fr}

Abstract

The use of a camera in a robot control loop can be performed with two types of architecture: the camera is said eye-in-hand when rigidly mounted on the robot end-effector and it is said eye-to-hand when it observes the robot within its work space. These two schemes have technical differences and they can play very complementary parts. Obviously, the eye-in-hand one has a partial but precise sight of the scene whereas the eye-to-hand camera has a less precise but global sight of it. The motivation of our work is to take advantage of both, free-standing and robot-mounted sensors, in a cooperation scheme. The system we present in this paper performs two separate tasks: a positioning one that is ensured in the global image and a tracking one performed in the local image. For robustness considerations, the control law stability is proved and several cooperative schemes are studied and compared in experimental results.

1 Overview

For the last fifteen years, eye-in-hand 2D visual servoing has been extensively studied [11, 3, 5, 7]. Central to this approach is the image jacobian L (also called interaction matrix). It relates the variations of some image features s to the robot control (generally expressed as a velocity screw T): $\dot{s} = LT$. Most of the time, the control can be expressed as the regulation of a task function [10]. If s^* is the desired value of s , we aim at controlling the robot in such a way that $s - s^*$ decreases to 0. A simple method consists in applying the following control law:

$$T = -\lambda \hat{L}^+(s - s^*) \quad (1)$$

where \hat{L}^+ is the pseudo-inverse of the estimated jacobian. If $\dim(s) \leq \dim(T)$, we can show that when \hat{L} is exactly computed at each time, (1) ensures an exponential decrease in $s - s^*$. But, for several reasons [1], \hat{L} is often fixed to a constant matrix (generally calculated at the desired position). Anyway, in that case, if we can show positiveness of the eigenvalues of $L\hat{L}^+$, local asymptotic stability of (1) is ensured.

In [6] the results of [3] are extended to the general case of a camera observing the robot being controlled. In [9] and [12], we can find eye-to-hand systems developed for specific tasks. We should stress the fact that, in the eye-to-hand case, the image jacobian

has to take into account the mapping from the camera frame onto the robot control frame. If we note $[R, t]$ this mapping (R being the rotational matrix and t the translation vector), the eye-to-hand jacobian L_g is related to the eye-in-hand one L by:

$$L_g = -L \begin{pmatrix} R & -RS(-R^T t) \\ 0 & R \end{pmatrix} \quad (2)$$

where $S(a)$ is the skew symmetric matrix associated with vector a . In [6] the control law is identical to (1).

For complex tasks in natural or complex environments, we can no longer be limited to the use of one type of camera. On the one hand, there is no possibility for a local (eye-in-hand) camera to interact with its whole work space. Furthermore, it can hardly consider important modification in its environment. On the other hand, a global (eye-to-hand) camera is not maneuverable enough to explore the scene. Many papers deal with the use of several cameras in computer vision: usually for 3D reconstruction with stereo vision, sometimes for visual servoing with a pair of images [4, 6] but rarely making global and local images cooperate [8]. The work that we describe in this paper is a first step towards such a cooperation. We define and show feasibility of a low level visual servoing task achieved by means of a system bringing together a local view of the scene and a global one. This system is able to ensure positioning while keeping visibility of a target.

In Section 2, we precisely describe and model the tasks to be achieved by both eye-to-hand and eye-in-hand cameras and show, in Section 3, how they merge into a single one. While stressing the need of independence between the tasks, we will see that their natural interaction can not be ignored and will show how to take it into account in an efficient and robust way. In Section 4, we present results obtained with a six degrees of freedom cartesian robot.

2 Task description and modeling

In our system (see Figures 1 and 2), the global camera is static and controls the translating degrees of freedom of the robot effector to ensure its correct positioning while the local one controls its orientation to center a static target in its image.

The mobile landmark from which the global image features will be extracted is mounted at the end of the translating joints of the

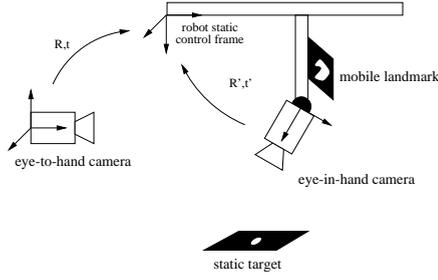


Figure 1: Eye-in-hand / Eye-to-hand cooperation

robot arm. This way, the end effector rotational motions do not influence the global image. This provides us with more independence between both tasks.

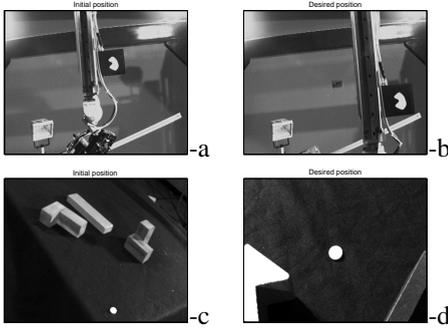


Figure 2: Initial and desired images for eye-to-hand (on the top) and eye-in-hand (on the bottom) systems

2.1 Translation control

The global image must control three degrees of freedom. We thus need to extract at least three independent features from the image. For stability proof convenience, we chose only three of them: the coordinates of the center of gravity (c.o.g.) of the observed target in the image and the projected surface. We further assume that the observed target is planar and parallel to the image plane. First, this results in the fact that the c.o.g of the target can be considered as a physical point of the target and its evolution is governed by:

$$\begin{pmatrix} \dot{X}_G \\ \dot{Y}_G \\ \dot{S} \end{pmatrix} = \begin{pmatrix} \frac{1}{z_G} & 0 & -\frac{X_G}{z_G} \\ 0 & \frac{1}{z_G} & -\frac{Y_G}{z_G} \\ 0 & 0 & -\frac{2S}{z_G} \end{pmatrix} \begin{pmatrix} x_G \\ y_G \\ z_G \end{pmatrix}$$

where (X_G, Y_G) are the coordinates (expressed in meters) of the c.o.g. in the image for a one meter focal length and (x_G, y_G, z_G) are the coordinates of the c.o.g of the target expressed in the camera frame. Second, the evolution of the projected surface S is given by: $\dot{S} = -(2S/z_G)z_G$. Both previous relations amount to

$$\begin{pmatrix} \dot{X}_G \\ \dot{Y}_G \\ \dot{S} \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{1}{z_G} & 0 & -\frac{X_G}{z_G} \\ 0 & \frac{1}{z_G} & -\frac{Y_G}{z_G} \\ 0 & 0 & -\frac{2S}{z_G} \end{pmatrix}}_{L_0} \begin{pmatrix} x_G \\ y_G \\ z_G \end{pmatrix} \quad (3)$$

We chose to express the translation control V in the static robot control frame. It means that the displacement from the eye-to-hand camera frame onto the control frame $[R, t]$ is constant. Besides, if we denote s_T the feature vector, equations (2) and (3) lead to:

$$\dot{s}_T = L_T V = L_0 R V \quad (4)$$

L_T is called the translational jacobian. The associated control law ensuring an exponential decrease in the feature vector is given by:

$$V = -\lambda_T \widehat{L}_T^+ (s_T - s_T^*) \quad (5)$$

where s_T^* is the desired feature vector and \widehat{L}_T is the estimated jacobian. In order to chose a suitable \widehat{L}_T , we now propose to determine the stability domain of the control law (5) under the assumptions presented before.

◆ Stability proof

To deal with internal calibration we remind the meter to pixel transformation. If we consider no radial distortion and assuming that image axes are perfectly orthogonal:

$$\begin{pmatrix} \dot{X}_G \\ \dot{Y}_G \\ \dot{S} \end{pmatrix}_{\text{pixels}} = \underbrace{\begin{pmatrix} \frac{f}{l_x} & 0 & 0 \\ 0 & \frac{f}{l_y} & 0 \\ 0 & 0 & \frac{f^2}{l_x l_y} \end{pmatrix}}_A \begin{pmatrix} \dot{X}_G \\ \dot{Y}_G \\ \dot{S} \end{pmatrix}_m \quad (6)$$

where f , l_x and l_y are respectively the focal length, the width and height of a pixel in meters. According to (4) and (6), we notice that the global translational jacobian matrix is $L = AL_0R$ where A , L_0 and R are non singular. We have to study positiveness of the eigenvalues of $L\widehat{L}^+$ ($= L\widehat{L}^{-1}$ in the present case). Denoting \widehat{x} the estimate of x , and $R = dR \cdot \widehat{R}$, we obtain: $L\widehat{L}^{-1} = AL_0 dR \widehat{L}_0^{-1} \widehat{A}^{-1}$. We propose to study two different cases:

1. $R = \widehat{R}$ (i.e. $dR = Id$)

In that case, the eigenvalues of $L\widehat{L}^{-1}$ are:

$$\left(\frac{f \widehat{z}_G \widehat{l}_x}{l_x z_G f}, \frac{f \widehat{z}_G \widehat{l}_y}{l_y z_G f}, \frac{f^2 \widehat{l}_y S \widehat{z}_G \widehat{l}_x}{l_x z_G \widehat{S} f^2 l_y} \right)$$

As a consequence, the local asymptotic stability of the system is always ensured since the conditions $f, \widehat{l}_x, \widehat{l}_y, \widehat{l}_x, \widehat{l}_y, z_G, \widehat{z}_G, S$ and $\widehat{S} > 0$ are not restrictive at all and whatever the estimations \widehat{X}_G and \widehat{Y}_G .

2. $A = \widehat{A}$ and $L_0 = \widehat{L}_0$

This time $L\widehat{L}^{-1} = AL_0 dR L_0^{-1} A^{-1}$ so dR and $L\widehat{L}^{-1}$ are similar and have the same eigenvalues. We just need to study positiveness of the real part of the eigenvalues of dR which are $(1, e^{i\theta}, e^{-i\theta})$ where θ is the rotation angle. This means that, for any rotation axis, the system is locally asymptotically stable if

$$-\frac{\pi}{2} < \theta < \frac{\pi}{2} \quad \square$$

Judging from this stability study, we can afford to fix \widehat{L}_0 to a constant value while ensuring the convergence of control. For the application, we chose to fix $\widehat{X}_G, \widehat{Y}_G$ and \widehat{z}_G to their desired values, \widehat{z}_G being coarsely approximated by hand. The constant value of R is also estimated by hand.

2.2 Rotation control

The principle of the second task is to control the eye-in-hand camera orientation such that the center of a static target appears and remains at the center of the image (see Figure 2). This intends to force the visibility of the target during the positioning task.

If the observed object is a point then the movement of the point projection (X, Y) is related to the rotating command $\Omega = [\Omega_X, \Omega_Y, \Omega_Z]$ by the following equation:

$$\begin{pmatrix} \dot{X} \\ \dot{Y} \end{pmatrix} = \underbrace{\begin{pmatrix} XY & -(1+X^2) & Y \\ 1+Y^2 & -XY & -X \end{pmatrix}}_{L_R} \Omega$$

L_R is the rotational jacobian. If $s_R = (X, Y)$ denotes the current eye-in-hand feature vector, s_R^* the desired one and \widehat{L}_R the estimated jacobian, the associated tracking law is given by:

$$\Omega = -\lambda_R \widehat{L}_R^+ (s_R - s_R^*) \quad (7)$$

As previously, we can show that the stability domain of this control law is very large. For future experiments, \widehat{L}_R is computed once using $\widehat{X} = 0$ and $\widehat{Y} = 0$. In that case, a sufficient condition to local asymptotic stability is $f, \widehat{f}, l_x, \widehat{l}_x, l_y, \widehat{l}_y > 0$. Computing \widehat{L}_R at each iteration is also very easy to implement.

2.3 Interaction

Whereas rotational motions do not influence global image features, translating movements of the end effector result in movements in the eye-in-hand image. More precisely, the movement of the point projection is related to the translating control V by the following equation:

$$\dot{s}_R = \underbrace{\begin{pmatrix} -1/z & 0 & X/z \\ 0 & -1/z & Y/z \end{pmatrix}}_{L_{RT}} R' V$$

where z is the depth of the static target from the local camera and R' is the rotation from the local camera frame onto the robot static control frame. When estimating \widehat{L}_{RT} , we will fix \widehat{X}, \widehat{Y} to their desired values (0) and \widehat{z} to a constant value z^* . R' will be computed thanks to odometry.

3 Cooperation

We now aim at making both previous tasks cooperate in a single control scheme. Judging from previous modeling, the estimated global task jacobian is:

$$\widehat{L} = \begin{pmatrix} \widehat{L}_T & 0 \\ \widehat{L}_{RT} & \widehat{L}_R \end{pmatrix} = \begin{pmatrix} \widehat{L}_d \\ \widehat{L}_e \end{pmatrix} \begin{matrix} \text{eye-to-hand} \\ \text{eye-in-hand} \end{matrix}$$

On the one hand, using a classical control (1) where \widehat{L} has a coupling term ($\widehat{L}_{RT} \neq 0$) will lead to unexpected behaviors such as sweeping movements in the images and loss of feature if \widehat{L} is not perfectly estimated. These fears were observed during experiments; that is why we aim at controlling both tasks independently. In addition, if independence is preserved, global stability will be ensured as long as each task is stable.

But on the other hand, if we fix $\widehat{L}_{RT} = 0$, we assume that both tasks are independent. This inaccuracy in modeling the system leads to tracking error. This error can be seen in Figure 3-b. The problem appears all the more dangerous since sometimes the point is about to vanish from the image. In order to suppress the tracking error while keeping independence between tasks, we compare two kinds of methods.

3.1 Estimating the perturbation

The translation of the effector can be seen as an unknown perturbation acting on the eye-in-hand features velocity measures. Because of this unknown perturbation, we must write:

$$\dot{s}_R = \frac{\partial s_R}{\partial r} \frac{dr}{dt} + \frac{\partial s_R}{\partial t} = L_R \Omega + \frac{\partial s_R}{\partial t}$$

where r denotes the effector position. If we still want to ensure an exponential decrease in $s_R - s_R^*$, the corresponding control law is given by [2]:

$$\Omega = \underbrace{-\lambda_R \widehat{L}_R^+ (s_R - s_R^*)}_{\text{classical control law}} \underbrace{-\widehat{L}_R + \frac{\partial s_R}{\partial t}}_{\text{tracking error compensation}} \quad (8)$$

Thus we must estimate the part of image features velocity due to the perturbation, that is $\widehat{\partial s_R / \partial t}$. Several ways have been investigated to compute this estimation:

- **Odometry:** If we have access to the robot odometry, V can be measured and the estimation is computed by $\widehat{\partial s_R / \partial t} = L_{RT} V$. However, we should note that only few robots can afford an accurate measure of their translational motion. For example mobile robots, which constitute an important application field of our work, rarely have such equipments.
- **Iterative scheme:** This method directly comes from control theory and consists in inserting an integrator into the control loop. The integrator is simply achieved by the following iterative scheme:

$$\left(\frac{\widehat{\partial s_R}}{\partial t} \right)_{k+1} = \left(\frac{\widehat{\partial s_R}}{\partial t} \right)_k + \mu (s_{Rk} - s_{Rk}^*) \quad (9)$$

where μ is a compensation gain. Let us note that, when stationarity is reached, $s_{Rk} = s_{Rk}^*$.

- **Observation / Prediction:** An other method consists in comparing the velocity we observe in the image and the velocity we can predict from the computed control law. The difference between both terms should correspond to the part of the movement due to perturbation. At time k , the estimate is:

$$\left(\frac{\widehat{\partial s_R}}{\partial t} \right)_k = \underbrace{\frac{s_{Rk}}{\Delta t}}_{\text{observation}} - \underbrace{\frac{s_{Rk/k-1}}{\Delta t}}_{\text{prediction}}$$

where $s_{Rk/k-1} = s_{Rk-1} + L_R \Omega_{k-1} \Delta t$. Thus:

$$\left(\frac{\widehat{\partial s_R}}{\partial t} \right)_k = \frac{s_{Rk} - s_{Rk-1}}{\Delta t} - L_R \Omega_{k-1} \quad (10)$$

At time k , we must have access to Ω_{k-1} . This is done either assuming that Ω_{k-1} is the previous computed value of the control law or measuring it. The last technique is the one we implemented since the first one assumes an ideal time response of the system. Let us note that rotational motions are always easier to measure accurately than translational ones.

Estimating scheme (10) would be perfect assuming that data are not noisy. As not iterative, this one does not filter noise due to the inaccuracy in the measure of s_R and in the measure of Ω at each time. In order to filter it, we can use either a simple filter

$$\widehat{\left(\frac{\partial s_R}{\partial t}\right)}_k = \alpha \cdot \left(\frac{s_{Rk} - s_{Rk-1}}{\Delta t} - L_R \cdot \Omega_{k-1}\right) + (1 - \alpha) \cdot \widehat{\left(\frac{\partial s_R}{\partial t}\right)}_{(k-1)} \quad (11)$$

where α is a forgetting factor, or a Kalman filter as the one proposed in [2]. Obviously, this last technique implies that we should know an approximation of the evolution model of $\frac{\partial s_R}{\partial t}$ but it is well-known that Kalman filtering is quite robust to this knowledge. We chose a model with a constant velocity state and a constant correlation.

Let us note that observation / prediction techniques can also deal with a mobile target. In that case, no difference can be made between perturbation due to the unknown translation of the effector and the one related to the unknown target motion, but it does not matter in the system behavior.

3.2 Task redundancy

An other method is to consider the tracking task as the main task. As it does not constrain all the robot degrees of freedom, we can use the task redundancy approach [10, 3]. A secondary task g_s can thus be performed regulating the task function:

$$e = W^+(s_R - s_{R^*}) + (I - W^+W)\alpha_2 g_s^T$$

where W is a full rank matrix such that $\text{Ker } W = \text{Ker } \widehat{L}_e$ if we set $\widehat{L}_e = [\widehat{L}_{RT} \quad \widehat{L}_R]$. Matrix $I - W^+W$ is thus a projection operator onto the kernel of \widehat{L}_e . It means that the secondary task realization will have no effect on the main one as long as \widehat{L}_e is a good approximation of L_e . In our case, the main task jacobian $L_e = (L_{RT} \quad L_R)$ is full rank so we can set $W = \widehat{L}_e$. Besides, the secondary task is a positioning one in the eye-to-hand image $g_s = C(s_T - s_T^*)$ where C is a combination matrix. Applying $T = -\lambda e$, the evolution of s_T if s_R has converged is given by $\dot{s}_T = -\lambda \alpha_2 L_d (I - \widehat{L}_e^+ \widehat{L}_e) C (s_T - s_T^*)$. A necessary condition for an exponential decrease is $C = (\widehat{L}_d (I - \widehat{L}_e^+ \widehat{L}_e))^+$. As a conclusion, the task function is given by:

$$e = \widehat{L}_e^+(s_R - s_{R^*}) + (I - \widehat{L}_e^+ \widehat{L}_e) (\widehat{L}_d (I - \widehat{L}_e^+ \widehat{L}_e))^+ \alpha_2 (s_T - s_T^*)$$

4 Experimental results

This section and Figures 3,4 and 5 show the experimental results we obtained for all the presented methods. The gains have been chosen as follows: $\lambda_T = \lambda_R = 0.1$. The mobile landmark where

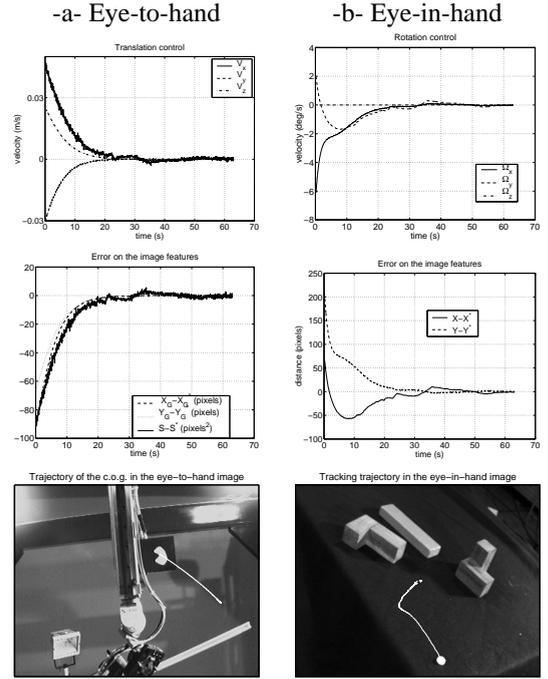


Figure 3: Independent control of both tasks

eye-to-hand features s_T are extracted from, is the white unknown and complex shape that can be seen on figure 2-a. s_T^* was previously learned (figure 2-b) but the value of s_T^* could also be fixed by the operator.

When we control both tasks independently and add no compensation (see Figure 3), we can notice that, as really independent, the translation control has a good exponential behavior whereas the centering one can not be achieved until the translation one has converged. This is exactly what we call tracking error.

Estimation and compensation of the tracking error

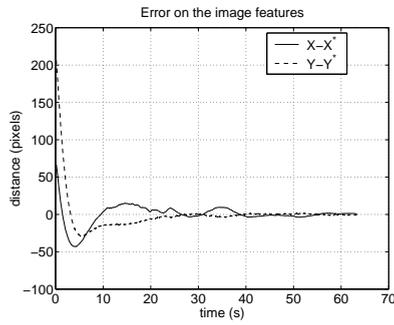
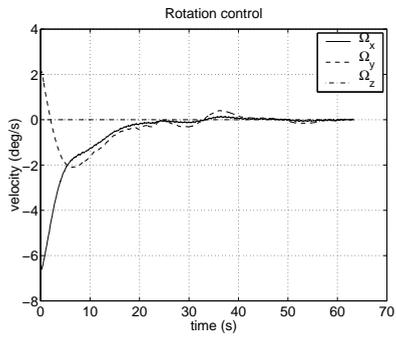
For these methods, the translation control does not change and its performances are exactly the one presented in Figure 3.

The results obtained using an integrator (see (9)) are presented on Figure 4-a. The gain was fixed to $\mu = 0.01$. This method increases the convergence rate but we observe that μ is quite difficult to adjust. Performances are not optimal due to non-constant velocity. The use of an estimator for $\frac{\partial s_R}{\partial t}$ leads to better performances (see Figure 4-b). Using a simple filtering given by (11) for which $\alpha = 0.7$, the noise on the control is significant. Its effect on the errors in the image is naturally filtered by the robot dynamic. The Kalman filter shows really good performances (see Figure 4-c). Finally, using odometry also appears very efficient (see Figure 4-d) but the corresponding performances mainly depend on the robot internal sensors accuracy.

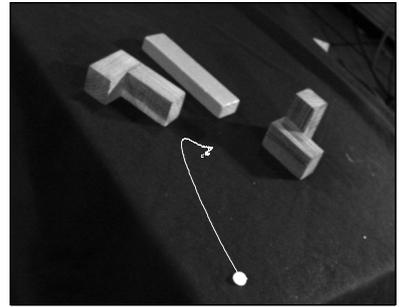
Redundancy approach

The redundancy formalism presents the advantage of specifying the importance of one task against the other. Figure 5 shows how

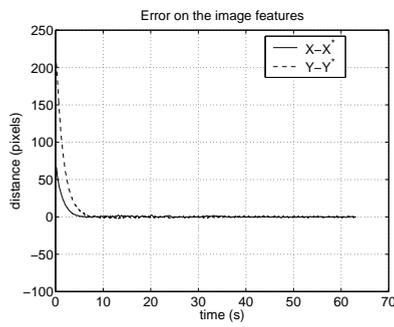
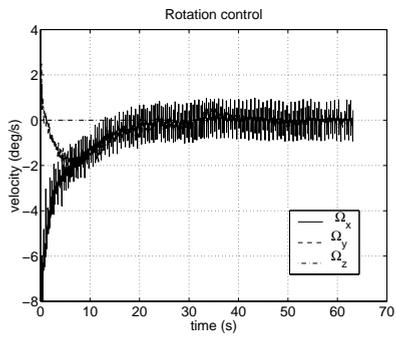
-a- Compensation with integrator



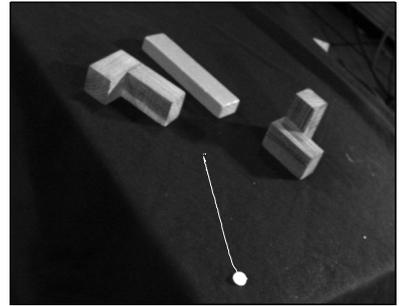
Tracking trajectory in the eye-in-hand image



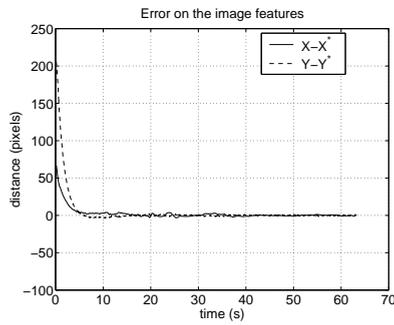
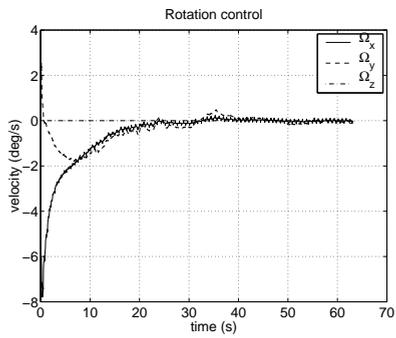
-b- Compensation with estimation + simple filtering



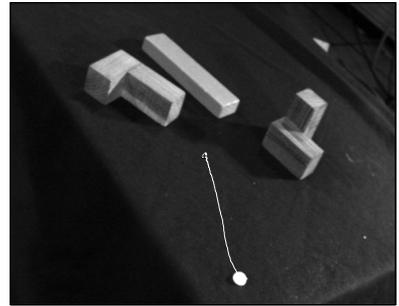
Tracking trajectory in the eye-in-hand image



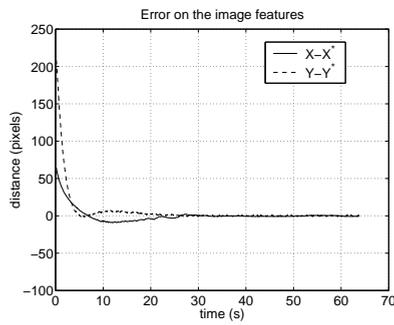
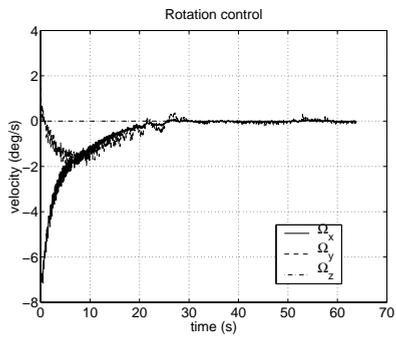
-c- Compensation with estimation + kalman filtering



Tracking trajectory in the eye-in-hand image



-d- Compensation with odometry



Tracking trajectory in the eye-in-hand image

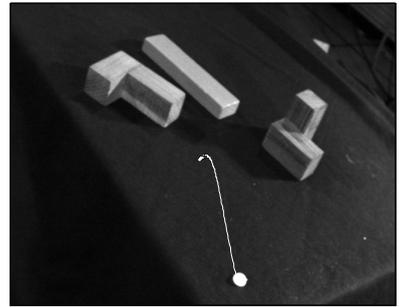


Figure 4: Tracking task using an estimate of $\frac{\partial s_B}{\partial t}$

the priority of the main task delays convergence of the secondary one. But since it has converged, the secondary task follows an exponential decrease. We also note that the primary task is disturbed by the secondary one only during the first iterations. This is due to the inexact estimation of the kernel.

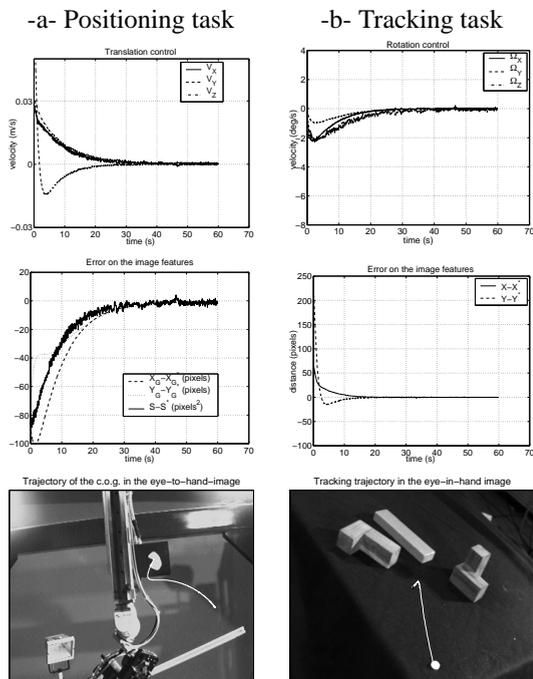


Figure 5: Redundancy approach

5 Conclusion

We have investigated a cooperation scheme successfully integrating a fixed camera and a robot-mounted one. Keeping independence between the tasks allowed us to prove stability of the control law. Most of the available techniques for visual servoing were adapted to the use of our two sensors and experimentally compared. All of them showed their ability to solve our task but some methods like Kalman filtering or odometry proved to be very efficient as regards their robustness.

The application we developed constitutes a good basis to build higher level tasks. Our future work will be dedicated to the development of exploration strategies with both global and local cameras. Very basic issues will arise: finding a suitable knowledge representation that should allow integration and comparison of new, uncertain and partial sensor measures, planning new sensor placement to improve knowledge while avoiding obstacles, making decisions according to current knowledge and the task to achieve.

Acknowledgments. This study was partly supported by INRIA LARA project and by Brittany County Council.

References

- [1] F. Chaumette. Potential problems of stability and convergence in image-based and position-based visual servoing. In D. Kriegman, G. Hager, and A.S. Morse, editors, *The Confluence of Vision and Control*, pp 66–78. LNCIS Series, No 237, Springer-Verlag, 1998.
- [2] F. Chaumette and A. Santos. Tracking a moving object by visual servoing. In *12th IFAC World Congress*, volume 3, pp 643–648, Sidney, Australia, July 1993.
- [3] B. Espiau, F. Chaumette, and P. Rives. A new approach to visual servoing in robotics. *IEEE Trans. on Robotics and Automation*, 8(3):313–326, June 1992.
- [4] G.D. Hager. A modular system for robust positioning using feedback from stereo vision. *IEEE Trans. on Robotics and Automation*, 13(4):582–595, August 1997.
- [5] K. Hashimoto. *Visual Servoing : Real Time Control of Robot Manipulators Based on Visual Sensory Feedback*. World Scientific Series in Robotics and Automated Systems, Vol 7, World Scientific Press, Singapore, 1993.
- [6] R. Horaud, F. Dornaika, and B. Espiau. Visually guided object grasping. *IEEE Trans. on Robotics and Automation*, 14(4):525–532, August 1998.
- [7] S. Hutchinson, G.D. Hager, and P.I. Corke. A tutorial on visual servo control. *IEEE Trans. on Robotics and Automation*, 12(5):651–670, October 1996.
- [8] E. Marchand and G. Hager. Dynamic sensor planning in visual servoing. In *IEEE Int. Conf. on Robotics and Automation*, volume 3, pp 1988–1993, Leuven, Belgium, May 1998.
- [9] F. Reyes and R. Kelly. Experimental evaluation of fixed-camera direct visual controllers on a direct-drive robot. In *IEEE Int. Conf. on Robotics and Automation*, Leuven, Belgium, May 1998.
- [10] C. Samson, M. Le Borgne, and B. Espiau. *Robot Control: the Task Function Approach*. Clarendon Press, Oxford, United Kingdom, 1991.
- [11] L.E. Weiss. Dynamic visual servo control of robots. An adaptive image based approach. Technical Report CMU-RI-TR-84-16, Carnegie-Mellon University, April 1984.
- [12] S.W. Wijesoma, D.F.H. Wolfe, and R.J. Richards. Eye-to-hand coordination for vision-guided robot control applications. *Int. Journal of Robotics Research*, 12(1):65–78, February 1993.