AVOIDING VISUAL SERVOING SINGULARITIES USING A COOPERATIVE CONTROL ARCHITECTURE

N. García, C. Pérez, L. Payá, R. Ñeco, J.M. Sabater, J. M. Azorín
Dept. Ingeniería de Sistemas Industriales. Universidad Miguel Hernández.
Avd. de la Universidad s/n. Edif. Torreblanca. 03202 Elche (Spain)

Keywords: Visual servoing, control, sensors, computer vision, robotics.

Abstract: To avoid the singularities of an image-based visual control of an industrial robot (Mitsubishi PA-10), a simple and efficient control law which combines the information of two cameras in a cooperative way has been developed and tested. One of this cameras is rigidly mounted on the robot end-effector (eye-in-hand configuration) and the other one observes the robot within its workspace (eye-to-hand configuration). The system architecture proposed allows us to control the 6 dof of an industrial robot when typical problems of image-based visual control techniques are produced.

1 INTRODUCTION

Nowadays, the great majority of robot population operates in factories where the work environment is structured and previously well-known. The application of a robot to carry out a certain task depends, in a high percentage, on the previously knowledge about the work environment and object placement. This limitation is due to inherent lack of sensory capability in contemporary commercial industrial robots. It has been long recognized that sensor integration is fundamental to increase the versatility and application domain of robots. One of these sensor systems is Computer Vision.

Computer Vision is a useful robotic sensor since it mimics the human sense of vision and allows for non contact measurement of the work environment. Industrial robot controllers with fully integrated vision systems are now available from a large number of suppliers. In these systems, visual sensing and manipulation are typically integrated in an open-loop mode, looking then moving. The precision of the resulting operation depends directly on the accuracy of the visual sensor and the robot end-effector.

An alternative solution for the position and motion control of an industrial manipulator evolved in unstructured environments is to use the visual information in a feedback loop. This robot control strategy is called visual servo control or visual servoing. Visual servoing systems have recently received a growing interest, as the computational power of commercially available computers became compatible with real time visual feedback(B. Espiau, 1992)(Hutchinson et al., 1996).

During the last years, position or image based visual servoing systems, systems with different architectures (a camera or multiples cameras), stability problems, calibration limitation, etc, have been studied. In particular, many image based visual servoing systems have been developed basically with two types of architecture: eye-in-hand configuration(Garcia et al., 2002), when the camera is rigidly mounted on the robot end-effector or eye-to-hand configuration(R. Horaud and Espiau, 1998) when the camera observes the robot within its work space.

The first approximation of using two cameras in eye-in-hand/eye-to-hand configurations were presented in the work of (Marchand and Hager, 1998). The system proposed used two task controlled by a camera mounted on the robot and a global camera to avoid obstacles during a 3D task. Then, in the paper reported by (Flandin et al., 2000) a system for integration a fixed camera and a camera mounted on the robot end-effector is presented. One task is used to control the translation dof of the robot with the fixed camera while other task is used to control the in eye-in-hand camera orientation. In this paper, an image based visual servo control of a 6 dof industrial robot manipulator with a cooperative eye-in-hand/eye-to-hand configuration is presented. In Section 2, the the-

García N., Pérez C., Payá L., Ñeco R., Sabater J. and Azorín J. (2004).
AVOIDING VISUAL SERVOING SINGULARITIES USING A COOPERATIVE CONTROL ARCHITECTURE.
In *Proceedings of the First International Conference on Informatics in Control, Automation and Robotics*, pages 162-168
DOI: 10.5220/0001134201620168

oretical background of an image based visual servoing system with eye-in-hand configuration and with eye-to-hand configuration is described. In Section 3, the control architecture of a cooperative image-based visual servoing system is presented. In the last section, some experimental results of this control scheme are shown.

2 THEORETICAL BACKGROUND

In this section a short description of the theoretical background of eye-in hand and eye-to-hand image-based visual servoing approach will be described. Moreover, this section is used to introduce the notation shown along the paper.

2.1 Eye-in-hand configuration

In this section, fundamentals about image-based visual servoing with eye-in-hand configuration is presented. This category of visual servoing is based on the selection of a set s of visual features that has to reach a desired value s*(Figure 1).

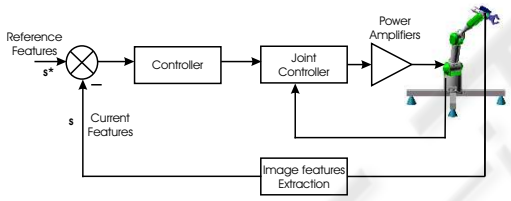


Figure 1: Image-based visual servoing with eye-in-hand configuration.

It is well known that the Image Jacobian $\mathbf{L}$, also called interaction matrix, relates the image features changes with the camera velocity screw:

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

where $\mathbf{v} = (\mathbf{V^T}, \boldsymbol{\omega}^T)$ is the camera velocity screw ($\mathbf{V}$ and $\boldsymbol{\omega}$ represent its translational and rotational component respectively). Using a classical perspective projection model with an intrinsic parameters matrix $\mathbf{A}$, and if x_i, y_i are the image coordinates of the feature selected s_i, then $\mathbf{L}$ is computed from:

$$\mathbf{L} = \mathbf{A} \begin{bmatrix} -\frac{1}{Z_i} & 0 & \frac{x_i}{Z_i} & x_i\,y_i & -(1+x_i^2) & y_i \\ 0 & -\frac{1}{Z_i} & \frac{y_i}{Z_i} & 1+y_i^2 & -x_i\,y_i & -x_i \end{bmatrix}$$

where Z_i is the depth of the corresponding point in the camera frame.

The great majority of references about control schemes compute the camera velocity sent to the robot controller (or directly the robot joints velocity, by introducing the robot jacobian express in the camera frame):

$$\mathbf{v} = -\lambda\,\mathbf{L}^{+}(\mathbf{s} - \mathbf{s}^{*}) \tag{2}$$

where λ may be as simple as a proportional gain (B. Espiau, 1992), or a more complex function used to regulate $\mathbf{s}$ to $\mathbf{s}^{*}$(optimal control, non-linear control, etc.), and $\mathbf{L}^{+}$ is the pseudo-inverse of $\mathbf{L}$.

As a general framework for sensor-based control of robots, the task function approach (C. Samson and Espiau, 1991) has been used:

$$\mathbf{e} = \mathbf{L}^{+}\,(\mathbf{s} - \mathbf{s}^{*}) \tag{3}$$

It is well known that in the task function approach, a sufficient condition to ensure global asymptotic stability of the system is:

$$\mathbf{L}^{+}\mathbf{L}(s_i, Z_i) > 0$$

In practice, three different cases of possible choices for $\mathbf{L}^{+}$ have been considered(Chaumette, 1998):

- It is numerically estimated during the camera motion without taking into account the analytical form given by (2).
- It is constant and determined during off-line step using the desired value of the visual features and an approximation of the depth at the desired camera pose. Stability condition is now ensured only in a neighborhood of the desired position.
- 3. It is now update at each iteration of the control law using in (2) the current measure of the visual features and an estimation of the depth of each considered point. The depth can be obtained from the knowledge of a 3D model of the object(DeMenthon and Davis, 1992).

It is well known that the performance of image visual servoing system is generally satisfactory, even in the presence of important camera or hand-eye calibration errors (Espiau, 1993). However, the following stability and convergence problems may be occurred:

- Image jacobian may become singular during the servoing, which of course leads to unstable behavior.
- Local minima may be reached owing to the existence of unrealizable image motions.
- The image features go out of the image plane during the control task

2.2 Eye-to-hand configuration

In this section, fundamentals about image based visual servoing with eye-to-hand configuration are presented. The camera used in this eye-to-hand configuration observes a moving robot gripper. This category of visual servoing is based on the selection of a set s of visual features that has to reach a desired value

s* (Figure 2). From the movement of image features point of view, the relation between eye-in-hand and eye-to-hand configurations is shown in Figure 3.

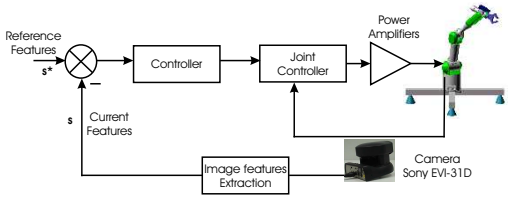


Figure 2: Image-based visual servoing with eye-to-hand configuration.

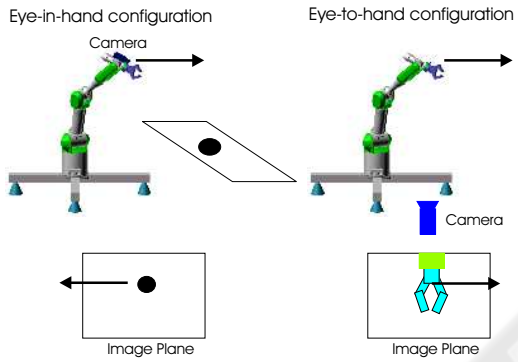


Figure 3: Relation between eye-in-hand and eye-to-hand configurations.

It's necessary to stress the fact that, in the eye-to-hand configuration, the image jacobian or interaction matrix has to take into account the mapping from the camera frame onto the robot control frame. If we denote $(\mathbf{R}, \mathbf{t})$ this mapping ($\mathbf{R}$ being the rotational matrix and $\mathbf{t}$ the translational vector), the eye-to-hand jacobian $\mathbf{L_{ETH}}$ is related to the eye-in-hand one $\mathbf{L_{EIH}}$ by:

$$\mathbf{L_{ETH}} = -\mathbf{L_{EIH}} \begin{bmatrix} \mathbf{R} & -\mathbf{R} \cdot S(-\mathbf{R}^T t) \\ 0 & \mathbf{R} \end{bmatrix} \quad (4)$$

where $S(a)$ is the skew symmetric matrix associated with vector a. The control law is identical to (2).

3 COOPERATIVE EYE-IN-HAND/EYE-TO-HAND SYSTEM

Combining several sensory data is also an important issue that has been studied considering two fundamentally different approaches. In the first one, the different sensors are considered to complementary measure of the same physical phenomena. Thus, a sensory data fusion strategy is used to extract a pertinent information from the multiple sensory data. The second control approach consists of selecting, among the available sensory signals, a set of pertinent data, which is then servoed. The two approaches will be referred as sensory data fusion and sensory data selection respectively.

A typical example of sensory data fusion is stereo vision. With this approach, two images provided by two distinct cameras are used to extract a complete Euclidean information on the observed scene. On the other hand, sensory data selection is used when all the different data no provide the same quality of information. In this case one can use data environment models in order to select the appropriate sensor and to switch control between sensors.

The approach to cooperative eye-in-hand/eye-to-hand configuration shown in this paper is a clearly case of multi sensory robot control(Malis et al., 2000). It does not pertain to sensory data fusion because we assume that the sensors may observe different physical phenomena from which extracting a single fused information does not make sense. It neither pertains to sensory data selection because we consider potential situations for which it is not possible to select a set of data that would be more pertinent than others. Consequently, the proposed approach addresses a very large spectrum of potential applications, for which the sensory equipment may be disparate and complex. As an improvement over previous approaches, there is no need to provide a model of the environment that would be required to design a switching or fusion strategy.

The robot is supposed to be controlled by a six dimensional vector $\mathbf{T_E}$ representing the end-effector velocity, whose components are supposed to be expressed in the end-effector frame. There are two cameras, one of them rigidly mounted on the robot end-effector(eye-in-hand configuration) and the other one observing the robot gripper(eye-to-hand configuration). Each sensor provides an n_i ($n_i > 6$) dimensional vector signal $\mathbf{s_i}$. Let $\mathbf{s} = [\mathbf{s_{EIH}}\ \mathbf{s_{ETH}}]^T$ be the vector containing the signals provided by the two sensors. Using the task function formalism (C. Samson and Espiau, 1991), an error function $\mathbf{e} = \mathbf{C}(\mathbf{s} - \mathbf{s}^*)$ can be defined as:

$$\begin{aligned} \mathbf{e} &= \begin{bmatrix} \mathbf{e_{EIH}} \\ \mathbf{e_{ETH}} \end{bmatrix} = \\ &= \begin{bmatrix} \mathbf{C_{EIH}} \\ \mathbf{C_{ETH}} \end{bmatrix} \left(\begin{bmatrix} \mathbf{s_{EIH}} \\ \mathbf{s_{ETH}} \end{bmatrix} - \begin{bmatrix} \mathbf{s^*_{EIH}} \\ \mathbf{s^*_{ETH}} \end{bmatrix} \right) \end{aligned} \quad (5)$$

where $\mathbf{C} = [\mathbf{C_{EIH}}\ \mathbf{C_{ETH}}]^T$ is a full rank matrix, of dimension $m \times n_i$(where m must be equal to dof to be controlled in this case $m = 6$), which allows to take into account information redundancy.

An interaction matrix is attached to each sensor, such that:

$$\begin{aligned} \dot{\mathbf{s}}_{\mathbf{EIH}} &= \mathbf{L}_{\mathbf{EIH}}\mathbf{T}_{\mathbf{CE_{EIH}}}\mathbf{T}_{\mathbf{E}} \\ \dot{\mathbf{s}}_{\mathbf{ETH}} &= \mathbf{L}_{\mathbf{ETH}}\mathbf{T}_{\mathbf{CE_{ETH}}}\mathbf{T}_{\mathbf{E}} \end{aligned} \tag{6}$$

where $\mathbf{T_{CE}}$ is the transformation matrix linking sensor velocity and the end effector velocity, in the case of eye-in-hand configuration will be constant and on the other case(eye-to-hand configuration) will be variable. The equation (6) can be expressed as a matrix:

$$\begin{aligned} \dot{\mathbf{s}} &= \begin{bmatrix} \dot{\mathbf{s}}_{\mathbf{EIH}} \\ \dot{\mathbf{s}}_{\mathbf{ETH}} \end{bmatrix} = \begin{bmatrix} \mathbf{L}_{\mathbf{EIH}} & 0 \\ 0 & \mathbf{L}_{\mathbf{ETH}} \end{bmatrix} \begin{bmatrix} \mathbf{T}_{\mathbf{CE_{EIH}}} \\ \mathbf{T}_{\mathbf{CE_{ETH}}} \end{bmatrix} \\ \mathbf{T}_{\mathbf{E}} &= \mathbf{L}_{\mathbf{T}} \cdot \mathbf{T}_{\mathbf{CE}} \cdot \mathbf{T}_{\mathbf{E}} \end{aligned} \tag{7}$$

The time derivative of the task function (5), considering C and s* constant, is:

$$\dot{\mathbf{e}} = \mathbf{C}\dot{\mathbf{s}} = \mathbf{C}\mathbf{L}_{\mathbf{T}}\mathbf{T}_{\mathbf{CE}}\mathbf{T}_{\mathbf{E}} \tag{8}$$

A major concern in designing a task function based controller is to select a suitable constant matrix C, while ensuring that the matrix $\mathbf{C}\mathbf{L}_{\mathbf{T}}\mathbf{T}_{\mathbf{CE}}\mathbf{T}_{\mathbf{E}}$ has a full rank.

In this paper, $\mathbf{C}$ is designed as a function of the pseudo-inverse of $\mathbf{L_T}$ and $\mathbf{T_{CE}}$.

$$\mathbf{C} = [k_1\mathbf{T}^{-1}_{\mathbf{CE_{EIH}}}\mathbf{L}^{+}_{\mathbf{EIH}} \;\; k_2\mathbf{T}^{-1}_{\mathbf{CE_{ETH}}}\mathbf{L}^{+}_{\mathbf{ETH}}] \tag{9}$$

where k_i is a positive weighting factor such that $\sum_{i=1}^{2} k_i = 1$. If for each sensor a task function (where i = 1 is referred to eye-in-hand configuration and i=2 to eye-to-hand configuration) is considered, then the task function of the entire system is a weighted sum of the task functions relative to each sensor:

$$\begin{aligned} \mathbf{e} &= \mathbf{C}(\mathbf{s} - \mathbf{s}^*) = \sum_{i=1}^{2} k_i \cdot \mathbf{e_i} = \\ &= \sum_{i=1}^{2} k_i \cdot \mathbf{C_i}\,(\mathbf{s_i} - \mathbf{s_i^*}) \end{aligned} \tag{10}$$

The design of the two sensors combination simply consists of selecting the positive weights k_i. This choice is both task and sensor dependent. The weights k_i can be set according to the relative precision of the sensors, or more generally to balance the velocity contribution of each sensor. Also a dynamical setting of k_i can be implemented.

A simple control law can be obtained by imposing the exponential convergence of the task function to zero:

$$\dot{\mathbf{e}} = -\lambda\mathbf{e} \Rightarrow \mathbf{C}\mathbf{L}_{\mathbf{T}}\mathbf{T}_{\mathbf{CE}}\mathbf{T}_{\mathbf{E}} = -\lambda\mathbf{e} \tag{11}$$

where λ is a positive scalar factor which tunes the speed of convergence:

$$\dot{\mathbf{e}} = -\lambda(\mathbf{C}\mathbf{L}_{\mathbf{T}}\mathbf{T}_{\mathbf{CE}})^{-1}\mathbf{e} \tag{12}$$

Taking into account (9), it can be demonstrated that $(\mathbf{C}\mathbf{L}_{\mathbf{T}}\mathbf{T}_{\mathbf{CE}})^{-1}$ is equal to the identity:

$$\begin{aligned} (\mathbf{C}\mathbf{L}_{\mathbf{T}}\mathbf{T}_{\mathbf{CE}})^{-1} &= (\sum_{i=1}^{2} k_i\mathbf{T}^{-1}_{\mathbf{CE_i}}\mathbf{L}^{+}_{\mathbf{i}}\mathbf{L_i}\mathbf{T}_{\mathbf{CE_i}})^{+} = \\ &= (\sum_{i=1}^{2} k_i\mathbf{I_6})^{+} = \mathbf{I_6} \end{aligned} \tag{13}$$

So, if C is setting to (9) and each subsystem is stable, then $(\mathbf{C}\mathbf{L}_{\mathbf{T}}\mathbf{T}_{\mathbf{CE}})^{-1} > 0$ and the task function converge to zero and, in the absence of local minima and singularities, so does the error s - s*. In this case, this control law can be used to drive back the robot to the reference position.

In Figure 4, a control scheme of the general architecture proposed can be seen. To implement it, a software function to give the corresponding values to k_1 and k_2 is used.

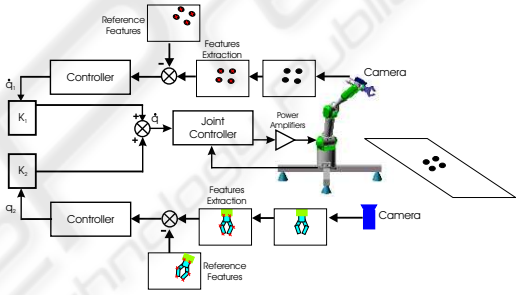


Figure 4: General architecture of the controller proposed.

4 EXPERIMENTAL RESULTS

Experimental results has been carried out using a 7 axis redundant Mitsubishi PA-10 manipulator (only 6 of its 7 dof have been considered). The experimental setup used in this work also include one camera(JAI CM 536) rigidly mounted in robot endeffector, one camera(EVI 31D) observing the robot gripper, some experimental objects and a computer with a Matrox Genesis vision board and other pc with the PA-10 controller board. An RPC link between the robot controller and the computer with the vision board for synchronization tasks and data interchange has been implemented. The whole experimental setup can be seen in Figure 5.

It's obvious that the performance of the system proposed depends on the selection of the weights k_i. Before giving the corresponding value to k_i some rules have been taking into account to avoid typical problems of image-based visual servoing approaches like task singularities, features extraction errors, disappearance of features from the image plane and so on.

To do this, a checking routine is executed and if one of the problems described before are produced, the corresponding value of k_i will set to zero. Obviously, the system fails if the problems happens in the two configurations at the same time.

Exhaustive number of experiments have been made with different weights(Figure 6). In Figure 7 and Figure 10, the results with ($K_1 = 1$, $K_2 = 0$ only the camera in eye-in-hand configuration is used and $K_1 = 0$, $K_2 = 1$ only the camera in eye-to-hand configuration is used)are presented. Observing them, we can realize that each system is stable and the error is zero excepted by the noise of features extraction.

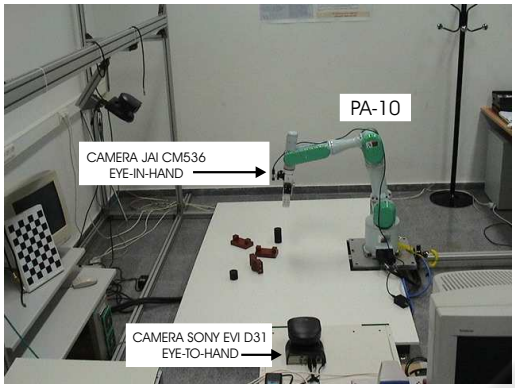


Figure 5: Experimental setup.

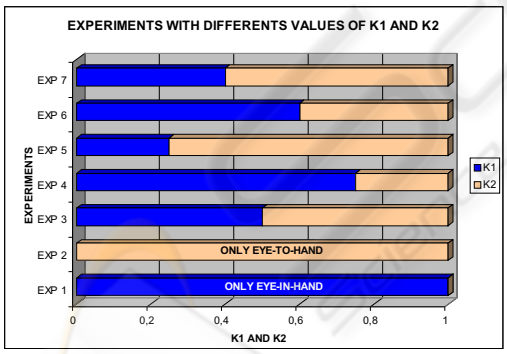


Figure 6: Experiments with different values of k_1 and k_2.

The results of all the experiments show that the control system proposed is stable and independent to the values of k_i. This conclusion corroborates the stability analysis presented at the end of Section 3. Assuring that each system is stable, the cooperative control system allow us to modify the magnitud of k_i without risk of making the system unstable.

For this reason, experiments with variable values of k_i have been carried out. To compute k_i in each sample time, the following function that depends on the relative image error is used:

$$k_1 = \frac{\mathbf{e_{rel_{EIH}}}}{\mathbf{e_{rel_{EIH}}} + \mathbf{e_{rel_{ETH}}}} \tag{14}$$

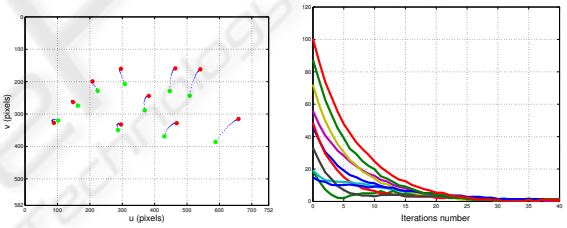
$$k_2 = \frac{\mathbf{e_{rel_{ETH}}}}{\mathbf{e_{rel_{EIH}}} + \mathbf{e_{rel_{ETH}}}} \tag{15}$$

where:

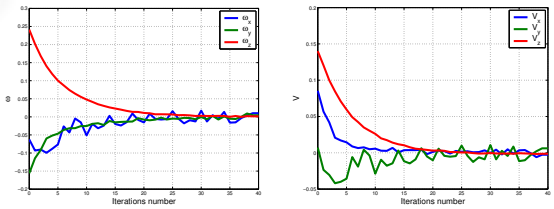
$$e_{rel_i} = \frac{s_i(t) - s_i^*}{s_i(0) - s_i^*} \tag{16}$$

Note that $\mathbf{e_{rel_{EIH}}}$ is computed when $i = 1$ and then is normalized dividing it by the number of image features. In the same way, $\mathbf{e_{rel_{ETH}}}$ is obtained.

The key idea of using this function is that the control contribution due to one of the cameras has more effect when its image features are far from their reference position. With this formulation of variable k_i, the local minima problems are avoided since the change in the weights k_i will bring the system away from it. So we can assure that $\mathbf{e} = 0$ *if and only if* $\mathbf{e}_i = 0 \, \forall \, i$. In Figure 9, the values of k_1 and k_2 during the control task can be seen. In Figure 11, the results of using a variable value of the weights are shown.



(a) Image plane (b) Errors $\|s_i\text{-}s_i^*\|$

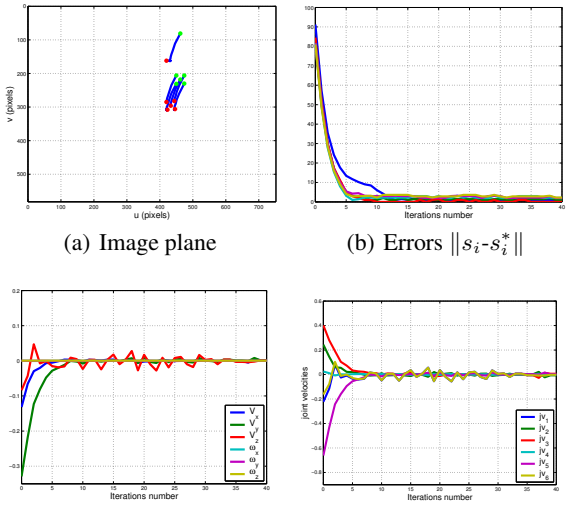


(c) Rotation speed (d) Translation speed

Figure 7: Results with $K_1 = 1$, $K_2 = 0$. Only the results of the camera in eye-in-hand configuration is shown. The translation and rotation speeds are measured in $\frac{m}{s}$ and $\frac{deg}{s}$.

5 CONCLUSION

The cooperative visual servoing proposed in this paper have been designed to make more efficient the classical imaged based visual servoing systems. In all the experimental results presented, the positioning accuracy of the architecture presented in this pa-



(a) Image plane (b) Errors $\|s_i\text{-}s_i^*\|$

(c) Rotation speed (d) Translation speed

Figure 8: Results with $K_1 = 0$, $K_2 = 1$. Only the results of the camera in eye-to-hand configuration is shown. The translation and rotation speeds are measured in $\frac{m}{s}$ and $\frac{deg}{s}$.

per is improved and also problems like local minima, task singularities, features extraction errors are avoided. Moreover, the architecture proposed, permits also to use several sensors (cameras, force sensors, etc.). Now, we are testing different functions to give values to k_i.

ACKNOWLEDGEMENT

This work has been supported by the Spanish Government through the 'Comision Interministerial de Ciencia y Tecnologa' (CICyT) through project "Modelado de espacios virtuales para entrenamiento de sistemas teleoperados en entornos dinmicos" DPI2001-3827-C02-02

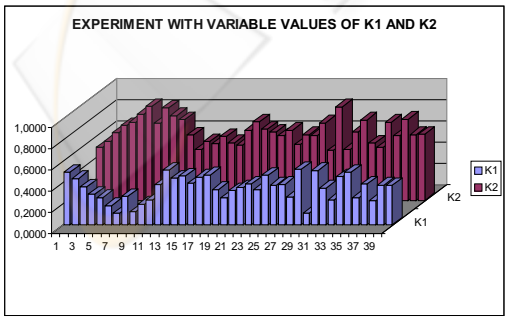
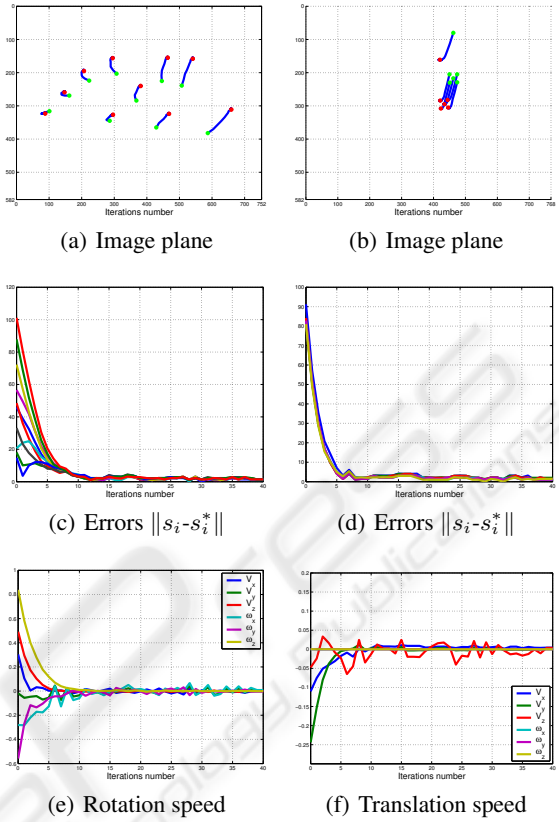


Figure 9: Experiments with variable values of k_1 and k_2.



(a) Image plane (b) Image plane

(c) Errors $\|s_i\text{-}s_i^*\|$ (d) Errors $\|s_i\text{-}s_i^*\|$

(e) Rotation speed (f) Translation speed

Figure 10: Results with variable values of K_1 and K_2. Figures (a,c,e) are the results of the camera in eye-in-hand configuration and figures (b,d,f) are the same for the camera in eye-to-hand configuration. The translation and rotation speeds are measured in $\frac{m}{s}$ and $\frac{deg}{s}$.

REFERENCES

B. Espiau, F. Chaumette, P. R. (1992). A new approach to visual servoing in robotics. *IEEE Trans. Robotics and Automation*, 8(3):313–326.

C. Samson, M. L. and Espiau, B. (1991). *Robot Control: the Task Function Approach.* volume 22 of Oxford Engineering Science Series.Clarendon Press., Oxford, UK, 1st edition.

Chaumette, F. (1998). Potential problems of stability and convergence in image-based and position-based visual servoing. In Kriegman, D., Hager, G. ., and Morse, A., editors, *The Confluence of Vision and Control*, pages 66–78. LNCIS Series, No 237, Springer-Verlag.

DeMenthon, D. and Davis, L. S. (1992). Model-based object pose in 25 lines of code. In *European Conference on Computer Vision*, pages 335–343.

Espiau, B. (1993). Effect of camera calibration errors on visual servoing in robotics.

Flandin, G., Chaumette, F., and Marchand, E. (2000). Eye-in-hand / eye-to-hand cooperation for visual servoing.

In *IEEE Int. Conf. on Robotics and Automation*, volume 3, pages 2741–2746, San Francisco.

Garcia, N., Mamani, G., Reinoso, O., Nasisi, O., Aracil, R., and Carelli, R. (2002). Visual servo control of industrial robot manipulator. In *15th IFAC World Congress in Automation and Control*, volume 1, Barcelona, Spain.

Hutchinson, S. A., Hager, G. D., and Corke, P. I. (1996). A tutorial on visual servo control. *IEEE Trans. Robotics and Automation*, 12(5):651–670.

Malis, E., Chaumette, F., and Boudet, S. (2000). Multi-cameras visual servoing. In *IEEE International Conference on Robotics and Automation*, volume 4, pages 3183–3188, San Francisco, USA.

Marchand, E. and Hager, G. (1998). Dynamic sensor planning in visual servoing. In *IEEE Int. Conf. on Robotics and Automation*, volume 3, pages 1988–1993, Leuven,Belgium.

R. Horaud, F. D. and Espiau, B. (1998). Visually guided object grasping. *IEEE Trans. Robotics and Automation*, 14(4):525–532.