

ROBUST IMAGE-BASED VISUAL SERVOING SYSTEM USING A REDUNDANT ARCHITECTURE

Rafael Aracil*

N. García, C. Pérez, L. Payá, L. M. Jiménez, O. Reinoso**

* *Dpto. de Automática, Ingeniería Electrónica e
Informática Industrial, Universidad Politécnica de Madrid*

** *Dpto. Ingeniería de Sistemas Industriales, Universidad*

Miguel Hernández

Avd. Universidad s/n Edif. Torreblanca, 03202

Elche(Alicante), Spain

nicolas.garcia@umh.es

Abstract: The reliability and robustness of image-based visual servoing systems have recently received a growing interest. In order to address this issue, a redundant and cooperative 2D visual servoing system based on the information provided by two cameras in *eye-in-hand/eye-to-hand* configurations to control the 6 dof of an industrial robot manipulator (Mitsubishi PA-10) is proposed. Its control law has been defined to assure that the whole system is stable if each subsystem is stable and to allow avoiding typical problems of image-based visual servoing systems like task singularities, features extraction errors, disappearance of image features, local minima, etc. Experimental results are presented which demonstrate the stability, performance and robustness of the system proposed.

Keywords: Computer vision, control, robotics

1. INTRODUCTION

Visual servoing is a good known solution to control the position and motion of an industrial manipulator evolved in unstructured environments. However, visual servoing systems are not complete efficiency due to the numerous problems that are still unresolved.

This paper contributes on making visual servoing techniques more robust. To do this, a particular solution based on a redundant and cooperative 2D visual servoing system to solve its typical

problems like task singularities, features extraction errors, disappearance of image features, etc is proposed. The system proposed is based on the information provided by two cameras in *eye-in-hand/eye-to-hand* configurations to control the 6 dof of an industrial robot manipulator.

The first approximation about the use of two cameras in *eye-in-hand/eye-to-hand* configurations was presented in the work of (Marchand and Hager, 1998). The system described used two tasks which were 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 which integrates a fixed camera and a camera mounted on the robot end-effector is presented. One task is used

¹ This work has been supported by the *Generalitat Valenciana* inside the framework *Ajudes per a projectes d'investigació científica i desenvolupament tecnològic* through project GV04A/667.

to control the translation dof of the robot with the fixed camera while other task is used to control the eye-in-hand camera orientation. Contrary to the two works commented before, in this paper, the redundant image-based visual control proposed can control all the 6 dof of the robot with one of the two cameras or with both at the same time in a cooperative way. In Section 2, the theoretical 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 the cooperative image-based visual servoing system is presented. In the last section, some experimental results of this control scheme are shown.

2. THEORETICAL FUNDAMENT

The objective of this section is to introduce the notation shown along the paper and a short description of the theoretical background of image-based visual servoing approach for *eye-in hand* and *eye-to-hand* configurations.

All image-based visual servoing approaches are based on the selection of a set \mathbf{s} of visual features that has to reach a desired value \mathbf{s}^* which has been learned previously. If $\mathbf{s}(\boldsymbol{\xi}) = \mathbf{s}^*(\boldsymbol{\xi}^*)$ then $\boldsymbol{\xi} = \boldsymbol{\xi}^*$ and the camera is back to the reference position.

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

$$\dot{\mathbf{s}} = \mathbf{L} \mathbf{v} \quad (1)$$

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).

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

$$\mathbf{e} = \mathbf{L}^+ (\mathbf{s} - \mathbf{s}^*) \quad (2)$$

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

$$\mathbf{v} = -\lambda \mathbf{L}^+ (\mathbf{s} - \mathbf{s}^*) \quad (3)$$

where λ may be as simple as a proportional gain (Espiau *et al.*, 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} .

In this kind of systems, the number of cameras used and their configurations are many and very varied. In this paper, one camera in *eye-in-hand* configuration and one camera in *eye-to-hand* configuration are considered (Figure 1).

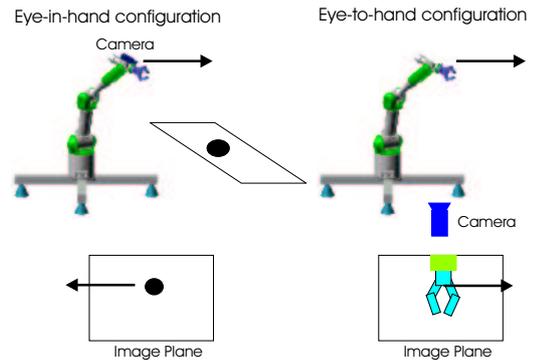


Fig. 1. Relation between eye-in-hand and eye-to-hand configurations.

From the visual control theory point of view, the difference between using one camera in *eye-in-hand* configuration or in *eye-to-hand* configuration to control a robot is in the computation of the interaction matrix \mathbf{L} . The interaction matrix for a camera in *eye-to-hand* configuration (\mathbf{L}_{ETH}) can be calculated from the analytical form of the interaction matrix in the case of *eye-in-hand* configuration (\mathbf{L}_{EIH}) taking 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:

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

where $S(a)$ is the skew symmetric matrix associated with vector a .

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 occur:

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

3. CONTROLLER DESIGN

In this section the design of a redundant and cooperative image-based visual servoing controller is presented. This controller is based on the visual information provided by two cameras located respectively in *eye-in-hand* and *eye-to-hand* configurations.

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 dimensional vector signal \mathbf{s}_i where $n_i > 6$ to be able to control the 6 dof of the robot with anyone of the cameras or with the two cameras at the same time in a cooperative way. 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 (Samson *et al.*, 1991), a total error function $\mathbf{e} = \mathbf{C}(\mathbf{s} - \mathbf{s}^*)$ can be defined as:

$$\mathbf{e} = \begin{bmatrix} \mathbf{e}_{EIH} \\ \mathbf{e}_{ETH} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{EIH} \\ \mathbf{C}_{ETH} \end{bmatrix} \left(\begin{bmatrix} \mathbf{s}_{EIH} \\ \mathbf{s}_{ETH} \end{bmatrix} - \begin{bmatrix} \mathbf{s}_{EIH}^* \\ \mathbf{s}_{ETH}^* \end{bmatrix} \right) \quad (5)$$

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:

$$\begin{aligned} \dot{\mathbf{s}} &= \begin{bmatrix} \dot{\mathbf{s}}_{EIH} \\ \dot{\mathbf{s}}_{ETH} \end{bmatrix} = \begin{bmatrix} \mathbf{L}_{EIH} & 0 \\ 0 & \mathbf{L}_{ETH} \end{bmatrix} \begin{bmatrix} \mathbf{T}_{CEIH} \\ \mathbf{T}_{CEETH} \end{bmatrix} \\ \mathbf{T}_E &= \mathbf{L}_T \cdot \mathbf{T}_{CE} \cdot \mathbf{T}_E \end{aligned} \quad (6)$$

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.

To compute both \mathbf{L}_{ETH} and \mathbf{T}_{CEETH} , the mapping from the camera frame onto the robot control frame (\mathbf{R}, \mathbf{t}) must be estimated. In this paper, a model based pose estimation algorithm are used since the model of the robot gripper is a priori known (DeMenthon and Davis, 1992). To show the accuracy of the pose estimation, a wire model of the robot gripper is drawn at each iteration of the control law (Figure 2).

The time derivative of the task function (5), considering \mathbf{C} and \mathbf{s}^* constant, is:

$$\dot{\mathbf{e}} = \mathbf{C}\dot{\mathbf{s}} = \mathbf{C}\mathbf{L}_T\mathbf{T}_{CE}\mathbf{T}_E \quad (7)$$

The key in designing a task function based controller is to select a suitable constant matrix \mathbf{C} , while ensuring that the matrix $\mathbf{C}\mathbf{L}_T\mathbf{T}_{CE}\mathbf{T}_E$ has a full rank and the system is stable. In this paper, \mathbf{C} is designed as a function of the pseudo-inverse of \mathbf{L}_T and \mathbf{T}_{CE} with the purpose of $(\mathbf{C}\mathbf{L}_T\mathbf{T}_{CE})^{-1}$ to be the identity:

$$\mathbf{C} = [k_1 \mathbf{T}_{CEIH}^{-1} \mathbf{L}_{EIH}^+ \quad k_2 \mathbf{T}_{CEETH}^{-1} \mathbf{L}_{ETH}^+] \quad (8)$$

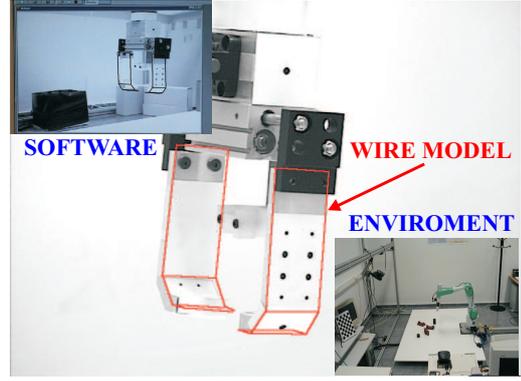


Fig. 2. Wire model of the robot gripper with a software and environment detail.

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:

$$\mathbf{e} = \sum_{i=1}^2 k_i \cdot \mathbf{e}_i = \sum_{i=1}^2 k_i \cdot \mathbf{C}_i (\mathbf{s}_i - \mathbf{s}_i^*) \quad (9)$$

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:

$$\dot{\mathbf{e}} = -\lambda \mathbf{e} \Rightarrow \mathbf{C}\mathbf{L}_T\mathbf{T}_{CE}\mathbf{T}_E = -\lambda \mathbf{e} \quad (10)$$

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

$$\mathbf{T}_E = -\lambda(\mathbf{C}\mathbf{L}_T\mathbf{T}_{CE})^{-1} \mathbf{e} \quad (11)$$

Taking into account (8), it can be demonstrated that $(\mathbf{C}\mathbf{L}_T\mathbf{T}_{CE})^{-1}$ is equal to the identity:

$$\begin{aligned} (\mathbf{C}\mathbf{L}_T\mathbf{T}_{CE})^{-1} &= \left(\sum_{i=1}^2 k_i \mathbf{T}_{CEi}^{-1} \mathbf{L}_i^+ \mathbf{L}_i \mathbf{T}_{CEi} \right)^+ = \\ &= \left(\sum_{i=1}^2 k_i \mathbf{I}_6 \right)^+ = \mathbf{I}_6 \end{aligned} \quad (12)$$

So, if \mathbf{C} is setting to (8) and each subsystem is stable, then $(\mathbf{C}\mathbf{L}_T\mathbf{T}_{CE})^{-1} > 0$ and the task function converges to zero and, in the absence of local minima and singularities, so does the error $\mathbf{s} - \mathbf{s}^*$.

Finally, substituting (8) in (11), the control law to drive back the robot to the reference position is obtained:

$$\mathbf{T}_E = -\lambda(k_1 \mathbf{T}_{CE_{EIH}}^{-1} \mathbf{L}_{EIH}^+ \mathbf{e}_{EIH} + k_2 \mathbf{T}_{CE_{ETH}}^{-1} \mathbf{L}_{ETH}^+ \mathbf{e}_{ETH}) \quad (13)$$

4. CONTROLLER IMPLEMENTATION

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 taken 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. In Figure 3, the flow chart of the checking routine can be seen. Obviously, the system fails if the problems happens in the two configurations at the same time.

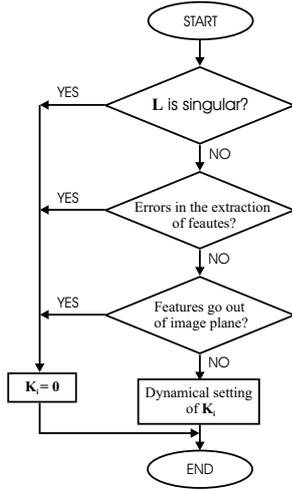


Fig. 3. Flow chart of the routine used to detect the potential problems of image-based visual servoing systems.

In Figure 3, the dynamical setting of k_i box represents a function to give values to k_i depending on some predefined functions. In this paper, k_i is computed in each sample time by the following function that depends on the relative image error:

$$k_1 = \frac{\mathbf{e}_{rel_{EIH}}}{\mathbf{e}_{rel_{EIH}} + \mathbf{e}_{rel_{ETH}}} \quad (14)$$

$$k_2 = \frac{\mathbf{e}_{rel_{ETH}}}{\mathbf{e}_{rel_{EIH}} + \mathbf{e}_{rel_{ETH}}} \quad (15)$$

where:

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

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 $\mathbf{e}_i = 0 \forall i$.

5. 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 includes 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 computer 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 4.

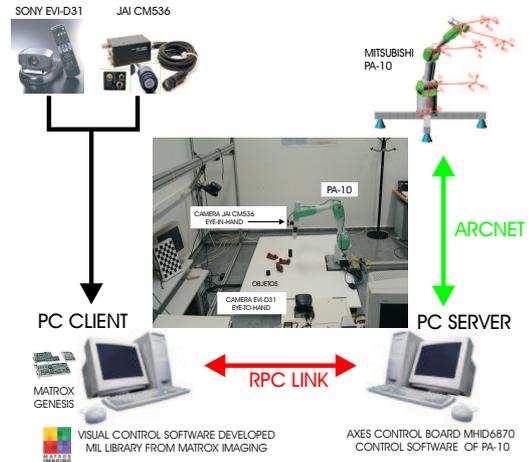


Fig. 4. Experimental setup.

With this experimental setup, exhaustive number of experiments have been made with different constant weights during the control task as it can be consulted in (Garcia *et al.*, 2004). In these experiments, we could verify that each system is stable and the error tends to zero excepted by the noise of features extraction.

In this paper, the dynamical setting of k_i is used to carry out a huge number of experiments. In Figure 5, the values of k_1 and k_2 during the control task in one of the experiments can be seen. In Figure 6,

the results of using a variable value of the weights are shown. Observing them, we can realize that the system is stable and the error tends to zero excepted by the noise of features extraction.

In Figure 7, the values of k_1 and k_2 during a control task with errors in the extraction of image features are presented. In this Figure, the performance of the system with errors in the extraction of features can be seen:

- Iterations 20-22: an error in the extraction of features (*eye-in-hand configuration*) is produced deliberately (Figure 8 a). This error is detected by the checking routine (Section 4) and k_1 is set to zero.
- Iterations 33-36: an error in the extraction of features (*eye-to-hand configuration*) is produced deliberately (Figure 8 b). This error is detected by the checking routine (Section 4) and k_2 is set to zero.

In spite of these forced errors, the system is stable and the robot reaches its reference position accurately.

6. CONCLUSIONS

The redundant and cooperative visual servoing system proposed in this paper has been designed to make more robust the classical imaged based visual servoing systems. In all the experimental results obtained, the positioning accuracy of the architecture presented in this paper is better than the classical one and also problems like local minima, task singularities and features extraction errors are avoided. Moreover, the architecture proposed, permits also to use several kind of sensors like cameras, force sensors, etc. without excessive difficulty. Now, we are testing different functions to give values to k_i .

REFERENCES

- DeMenthon, D. and L. S. Davis (1992). Model-based object pose in 25 lines of code. In: *European Conference on Computer Vision*. pp. 335–343.
- Espiau, B. (1993). Effect of camera calibration errors on visual servoing in robotics.
- Espiau, B., F. Chaumette and P. Rives (1992). A new approach to visual servoing in robotics. *IEEE Trans. Robotics and Automation* 8(3), 313–326.
- Flandin, G., F. Chaumette and E. Marchand (2000). Eye-in-hand / eye-to-hand cooperation for visual servoing. In: *IEEE Int. Conf. on Robotics and Automation*. Vol. 3. San Francisco. pp. 2741–2746.
- Garcia, N., L. Paya, O. Reinoso, R. Aracil and J.M. Sabater (2004). Cooperative image-based visual servoing system using a redun-

dant architecture. In: *International Conference on Intelligent Manipulation and Grasping*. Vol. 1. Italy.

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

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

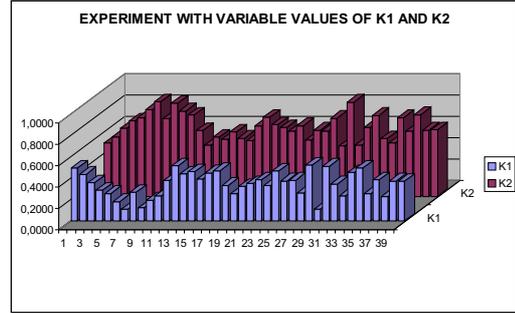


Fig. 5. Experiments with variable values of k_1 and k_2 .

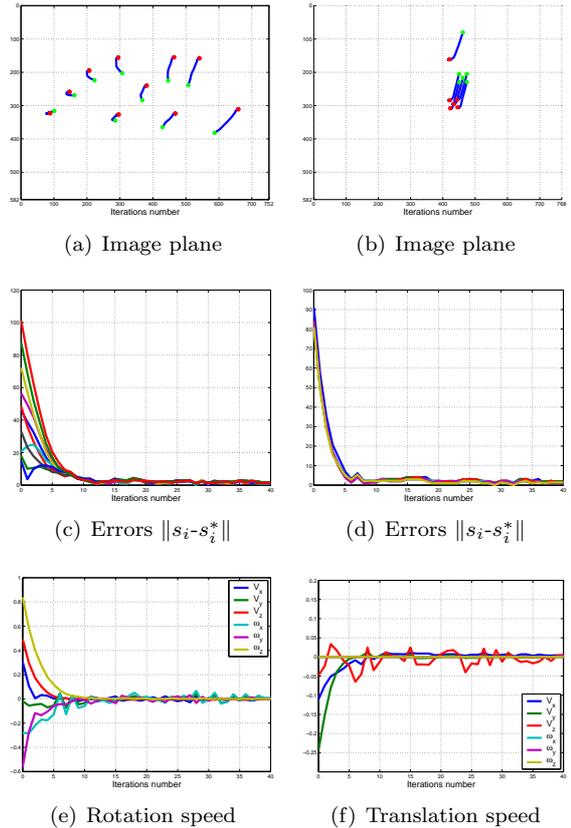


Fig. 6. Results with variable values of K_1 and K_2 . Figures (a,c,e) are the results of the camera in eye-in-hand configuration and figures (b,d,f) are the same for the camera in eye-to-hand configuration. The translation and rotation speeds are measured in $\frac{m}{s}$ and $\frac{deg}{s}$.

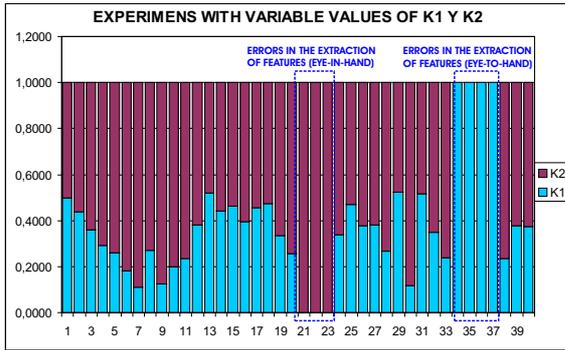
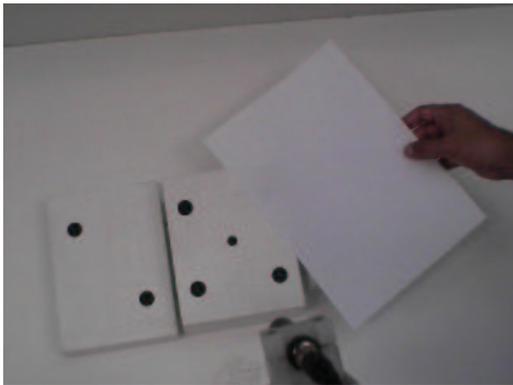


Fig. 7. Values of k_1 and k_2 with forced error in the extraction of features.



(a) Occlusion in *eye-in-hand* configuration



(b) Occlusion in *eye-to-hand* configuration

Fig. 8. Images of the errors which are produced by the occlusion of features during the control task.