Cooperative Visual Servoing System

Nicolás García, M. Asunción Vicente, Carlos Pérez, César Fernández, and Rafael Puerto

Dpto. Ingeniería, Miguel Hernández University, Avda. Ferrocarril s/n, 03202 Elche (Alicante), Spain nicolas.garcia@umh.es
http://lorca.umh.es

Abstract. A simple and efficient control law that combines the information of two cameras in order to realize the positioning task of a robot end-effector is presented in this paper. One of these cameras is rigidly mounted on the robot end-effector(eye-in-hand configuration) the other one observes the robot within its work space(eye-to-hand configuration). The aim of this work is to take advantage of both in a cooperation scheme. The control architecture of this system is presented. Comparison between simulations and experimental results with only one camera and with the two cameras in a cooperative way are shown.

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 fashion, 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 [9], [3].

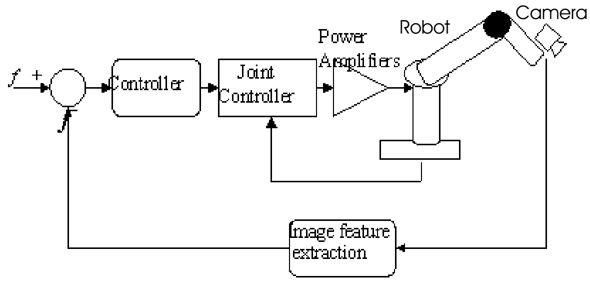

Fig. 1. Image based visual servoing with eye-in-hand configuration.

During the last years, position or image based visual servoing systems, their 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 [4], when the camera is rigidly mounted on the robot end-effector or eye-to-hand [5] configuration when the camera observes the robot within its work space. In these configurations, it's possible to place more than one camera.

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, an image based visual servoing system with eye-in-hand configuration and with eye-to-hand configuration is described. In Section 3, a combined eye-in-hand/eye-to-hand configuration is presented. Finally, some simulations and experimental results, where the advantages of this cooperative configuration, could be seen.

2 Image Based Visual Servoing

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 f of visual features that has to reach a desired value f_d (Fig.1.).

It is well known that the Image Jacobian J_f, also called interaction matrix, relates the image features changes with the camera velocity screw:

$$\dot{f} = J_f \cdot T \tag{1}$$

where $T = \left(V^T \Omega^T\right)^T$ is the camera velocity screw (V and Ω are translational and rotational camera velocity respectively). Using a classical perspective projection model with an intrinsic parameter matrix A, and if x_i, y_i are the image coordinates of the feature selected f, then J_f is obtained from:

$$J_f = A \cdot \begin{bmatrix} \frac{-1}{Z_i} & 0 & \frac{x_i}{Z_i} & x_i \cdot y_i & -1 - x_i^2 & y_i \\ 0 & \frac{-1}{Z_i} & \frac{y_i}{Z_i} & 1 + y_i^2 & -x_i \cdot y_i & -x_i \end{bmatrix} \tag{2}$$

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):

$$T = -\lambda \cdot J_f^+ \cdot \left(\bar{f}(t) - \bar{f}_d\right) \tag{3}$$

where function λ may be as simple as a proportional gain [9], or a more complex function used to regulate f to f_d (optimal control, non-linear control, etc.), and J_f^+ is the pseudoinverse of J_f.

As a general framework for sensor-based control of robots, the task function approach [6] has been used. It is well known that in the task function approach, a sufficient condition to ensure global asymptotic stability of the system is:

$$J_f^+ \cdot J_f \left(f(t), z(t)\right) > 0, \quad \forall t \tag{4}$$

In practice, three different cases of possible choices for J_f^+ have been considered [7]:

1. Image jacobian is numerically estimated during the camera motion without taking into account the analytical form given by (2).
2. Image jacobian is constant and determined during off-line step using the desired value of the visual features and an approximation of the points depth at the desired camera pose. Stability condition is now ensured only in a neighborhood of the desired position. The performed trajectory in the image may be quite unforeseeable, and some visual features may get out of the camera field of view during the servoing, especially if the initial camera position is far away from its desired one.
3. Image jacobian 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 [8]. Each image point is constrained to reach its desired position following a straight line. However, such a control in the image could be implied inadequate camera motion, leading to possible local minima and the nearing of task singularities.

It's 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 [9]. 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.

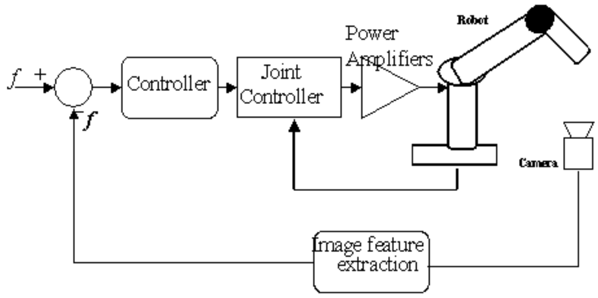

Fig. 2. Image based visual servoing with eye-to-hand configuration.

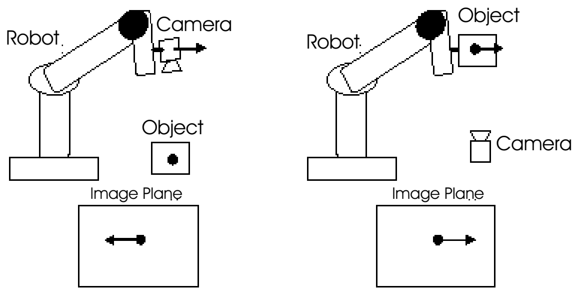

Fig. 3. Relation between eye-in-hand and eye-to-hand configuration.

2.2 Eye-to-hand Configuration

In this section, fundamentals about image based visual servoing with eye-to-hand configuration is 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 f of visual features that has to reach a desired value f_d (Fig. 2).

From the movement of image features point of view, the relation between eye-in-hand and eye-to-hand configurations is shown in Fig. 3.

It's necessary to stress the fact that, in the eye-to-hand configuration, the image jacobian has to take into account the mapping from the camera frame onto the robot control frame. If we denote (R, t) this mapping (R being the rotational matrix and t the traslational vector), the eye-to-hand jacobian J_{eth} is related to the eye-in-hand one J_{eih} by:

$$J_{eth} = -J_{eih} \cdot \begin{bmatrix} R & -R \cdot S\left(-R^T t\right) \\ 0 & R \end{bmatrix} \tag{5}$$

Where $S(a)$ is the skew symmetric matrix associated with vector a. The control law is identical to (3).

3 Combined eye-in-hand / eye-to-hand Configuration

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 approachs will be referred as sensory data fusion and sensory data selection respectivally.

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 sensor.

The approach to cooperative eye-in-hand/eye-to-hand configuration shown in this paper continues the work originally presented in [10] and is a clearly case of multi sensory robot control. 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 T_e representing the end-effector velocity, whose components are supposed to be expressed in the end-efector frame. There are two cameras, one of them rigdly 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 f_i . An image jacobian is attached to each sensor, such that:

$$\begin{aligned} \dot{f}_1 &= J_{eih} \cdot T_{ce_1} \cdot T_e \\ \dot{f}_2 &= J_{eth} \cdot T_{ce_2} \cdot T_e \end{aligned} \tag{6}$$

where T_{ce_i} 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.

Let $f = \begin{bmatrix} f_1^T & f_2^T \end{bmatrix}$ be the two dimensional vector containing the signals provided by the two sensors. The relationship between the time derivative of the global signal vector and the end-effector velocity T_e is:

$$\dot{f} = \begin{bmatrix} J_{eih} & 0 \\ 0 & J_{eth} \end{bmatrix} \cdot \begin{bmatrix} T_{ce_1} \\ T_{ce_2} \end{bmatrix} \cdot T_e = J \cdot T_{ce} \cdot T_e \tag{7}$$

Now, let f_d be the desired value of the sensor signal vector f . A task function [6] of the form $e = C \cdot (f - f_d)$ is used, where C is a full rank constant matrix of $(6 \times n)$ dimensions, which allows to take into account the information redundancy. The matrix C being constant, the time derivate of the task function is:

$$\dot{e} = C \cdot \dot{f} = C \cdot J \cdot T_{ce} \cdot T_e \tag{8}$$

A major concern in designing a task function based controller is to select a suitable constant matrix C, while ensuring that the matrix $C \cdot J \cdot T_{ce}$ has a full rank.

In this paper, C is designed as a function of the matrices J and T_{ce}.

$$C = \begin{bmatrix} k_1 \cdot T_{ce_1}^{-1} \cdot J_{eih}^{+} & k_2 \cdot T_{ce_2}^{-1} \cdot J_{eth}^{+} \end{bmatrix} \tag{9}$$

k_i is a positive weighting factor such that $\sum_{i=1}^{2} k_i = 1$. If for each sensor a task function $e_i = C_i \cdot (f_i - f_{d_i})$ with $C = T_{ce}^{-1} \cdot J_{eih}^{+}$ is considered, then the task function of the entire system is a weighted sum of the task functions relative to each sensor:

$$e = C \cdot (f - f_d) = \sum_{i=1}^{2} k_i \cdot e_i = \sum_{i=1}^{2} k_i \cdot C_i \cdot (f_i - f_{d_i}) \tag{10}$$

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.

In this paper a proportional control law is used by imposing the exponential decreasing of the task function. Considering equation (10), the control law is equivalent to a weighted sum of the control laws of each sub-system:

$$T_e = -\lambda \cdot \sum_{i=1}^{2} k_i \cdot e_i \tag{11}$$

The general architecture of this cooperative eye-in-hand/eye-to-hand visual servoing system is shown in Fig.4.

4 Simulations and Experimental Results

Simulations and experimental results has been obtained 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 D31) observing the robot gripper, some experimental objects and a computer with a Matrox Genesis vision board and the PA-10 controller board. The whole experimental setup can be seen in the Fig. 5.

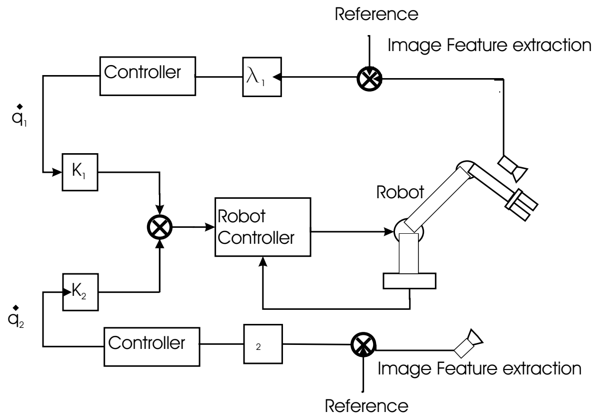

Fig. 4. General architecture of the proposed controller.

Simulations have been developed in Matlab and Simulink [11]. All the simulations have been made based on a block that includes all the robot parameters.

In order to demonstrate the performance of the eye-in-hand/eye-to-hand cooperation for visual servoing, the poorest case of image jacobian updating is chosen. In This case which the image jacobian is updated at each iteration of the control law using the current measure of the visual features, and an estimation of the depth of each considered point. It is well known that this kind of control in the image could implicated inadequate camera motion, leading to possible local minima and the nearing of task singularities.

4.1 Simulation Results

Simulations have been developed in Matlab and Simulink [11]. All the simulations have been made based on a Simulink block that includes all the robot parameters.

The selection of the positive weights k_i is task and sensor dependent. In the simulations, the weights k_i have been set according to the velocity contribution of each sensor. First, the visual servoing task is carried out using only one camera ($k_1 = 1$ and $k_2 = 0$ in the eye-in-hand configuration) or ($k_1 = 0$ and $k_2 = 1$ in the eye-to-hand configuration). Then, starting in the same initial position, the servoing task is performed using the two cameras with the same weight ($k_1 = 0.5$ and $k_2 = 0.5$). Therefore, an initial joint velocity checking is made before the weights are set in order to avoid task singularities ($k_i = 0$ if the control law produce a joint velocity vector bigger that a threshold).

In all the simulations, the positioning accuracy is considerably improved and also local minima and task singularities are avoided. In Fig. 6 a comparison between using only one camera or two cameras is shown.

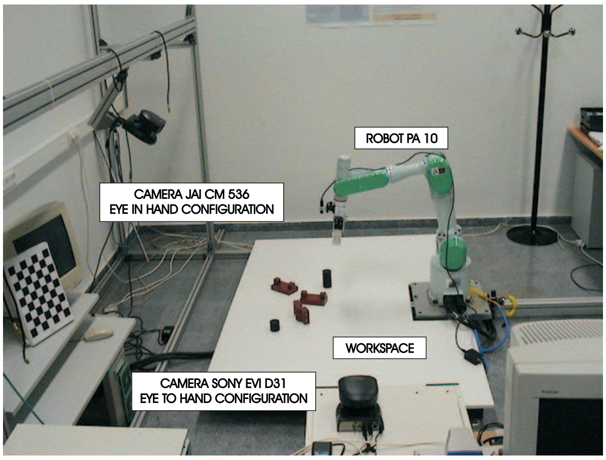

Fig. 5. Experimental setup.

4.2 Experimental Results

Real Experiments with eye-in-hand and eye-to-hand configurations only with one camera have been made. Now, experiments with cooperative visual servoing system with two cameras (Fig. 5) are being developed.

5 Conclusions

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 simulations, the positioning accuracy of the architecture presented in this paper is considerably improved and also local minima and task singularities are avoided. In the first experiments the simulation results are being corroborated. The architecture proposed allows also to use several sensors (cameras, force sensors, etc.).

Acknowledgements

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

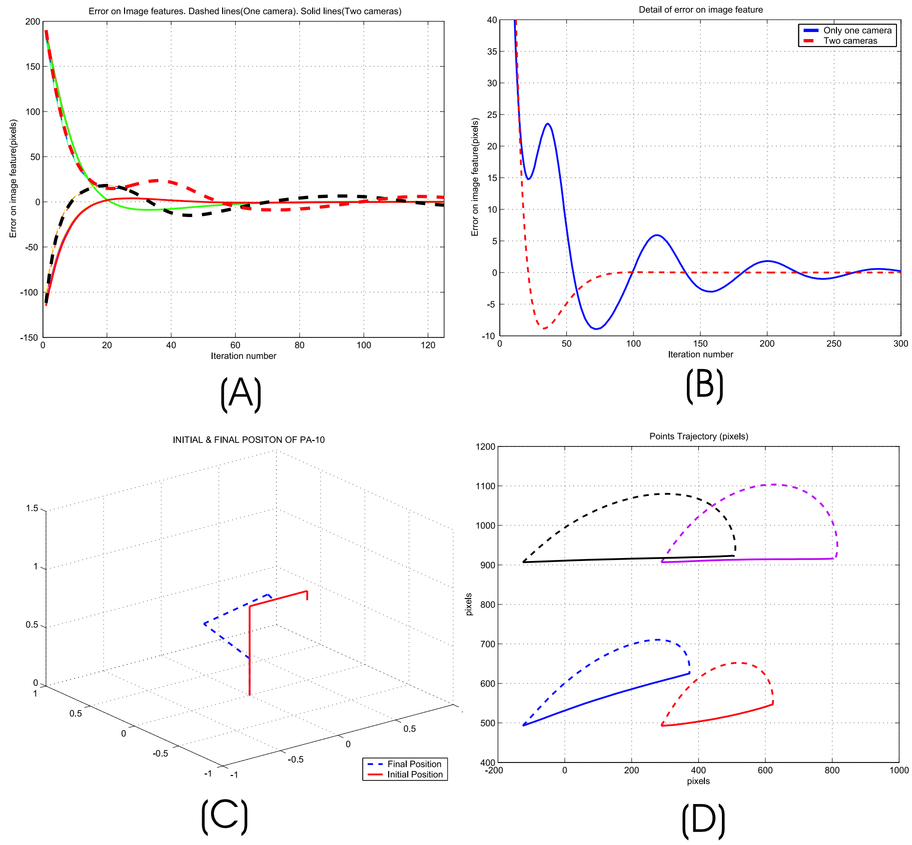

Fig. 6. Simulation results (A)-(B) Error on image features. Comparison between mono and multi-cameras. (C) Initial and final position of the robot (D) Trajectory in the image plane.Solid lines: one camera. Dashed lines: two cameras.

References

1. Shin Ho, E.K., Wan, L.C.: How to Design a Connectionist Holistic Parser. Neural Computation 11, p. 1995-2016 (1999)
2. Espiau, B., Chaumette, F. and Rives, P.: A new approach to visual servoing in robotics, IEEE Transactions on Robotics and Automation, 8, 313-326 (1992)
3. S. Hutchinson, G.D. Hager, P. Corke, I.: A Tutorial on Visual Servo Control.). IEEE Transactions on Robotics and Automation, 12(5) pp. 651-670, (1996)
4. García, N., Mamani G.,Reinoso, O., Nasisi, O., Aracil, R., Carelli, R.: Visual Servo Control of Industrial Robot Manipulator. 15th IFAC World Congress in Automation and Control, (2002)
5. Kragic, D., Miller, A., Allen, P.: Realtime tracking meets online grasp planning, IEEE International Conference on Robotics and Automation, ICRA'01, (2001)

6. Samson, C., Le Borgne, M., and Espiau, B.: Robot Control: the Task Function Approach, vol. 22 of Oxford Engineering Science Series, Clarendon Press, Oxford, England, (1991)
7. Chaumette, F.: Potential problems of stability and convergence in image-based and position-based visual servoing. In The Confluence of Vision and Control, D. Kriegman, G. Hager, A. Morse (eds.), Lectures Notes in Control and Informations Systems, Vol. 237, pp 66-78, Springer Verlag (1988)
8. Dementhon, D. and Davis, L.: Modelbased object pose in 25 lines of code. Int. Journal of Computer Vision, 15(1/2):123–141, (June 1995)
9. Espiau, B.: Effect of camera calibration errors on visual servoing in robotics. In Third Int. Symposium on Experimental Robotics, Kyoto, Japan, (October 1993)
10. Malis, E., Chaumette, F., and Boudet, S.: Multi-cameras visual servoing. IEEE Int. Conf. on Robotics and Automation, Volume 4, pages 3183-3188, San Francisco, (April 2000)
11. Saltarén, R. Azorín, J.M., Almonacid, M., Sabater, J.M.: Prácticas de robótica utilizando Matlab, UMH-Limencop, Spain (2000)