Vision Based Control Law using 3D Visual Features

Philippe Martinet, Jean Gallice, Khadraoui Djamel

Vision Based Control Law using 3D Visual Features

P. Martinet, J. Gallice, D. Khadraoui

Université Blaise Pascal de Clermont-Ferrand,

Laboratoire des Sciences et Matériaux pour l'Electronique, et d'Automatique.

U.R.A. 1793 du C.N.R.S., F-63177 Aubière Cedex, France.

E-Mail : Philippe.Martinet@lasmea.univ-bpclermont.fr

ABSTRACT

This paper presents the use of 3D visual features in the so-called "Visual Servoing Approach". After having briefly recalled how the task function approach is used in visual servoing, we present the notion of 3D logical vision sensors which permit us to extract visual information. In particular, we are interested in those composed of the estimation of both a 3D point and a 3D attitude. We give the control law expression with regard to the two kinds of visual features performed at video rate with our robotic platform. We present some of the experimental results and show the good convergence of the control laws.

KEYWORDS: visual servoing, sensor based control, 3D modeling, robotics, real time vision system, visual features extraction.

INTRODUCTION

In the last ten years, many people have become interested in how to introduce external sensor signals directly into the control loop. Near the end of the 1980's, C. Samson and B. Espiau proposed a new approach, called "Sensor Based Control" [3] [9]. In this approach, any robot task can be characterized as a regulation to zero of a task function, where it defines the interaction between the sensors and the robot environment. It is expressed with the following relation:

$$\underline{e}(\underline{r},t) = C[\underline{s}(\underline{r},t) - \underline{s}^{\star}] \tag{1}$$

where $\underline{s}^{\star}$ is considered as a reference target sensor feature to be reached in the sensor frame, $\underline{s}(\underline{r},t)$ is the value of sensor information currently observed by the sensors, and C is a constant matrix which permits taking into account more features than the number of robot degrees of freedom.

For a given task, the problem consists of choosing relevant sensor features to achieve the task, and then constructing the constant matrix C. This matrix requires the establishement of the interaction matrix related to the chosen sensor features. The interaction between sensor and scene is given in matrix form by:

$$\frac{\partial \underline{s}}{\partial \underline{r}} = L_{\underline{s}}^{T} \tag{2}$$

This is called the interaction matrix which links the interaction between the robot and its environment. In the case of visual servoing [3], we obtain the following relation:

$$\dot{\underline{s}} = L_{\underline{s}}^{T} T \tag{3}$$

where $\dot{\underline{s}}$ is the time variation of $\underline{s}$ and $T = (\underline{V}, \underline{\Omega})^T$ is the velocity screw of the camera which englobes three translations and three rotations.

By combining (1) and (3), and considering an exponential convergence (for a task positioning) of all components of the task function (i.e. $\dot{e} = -\lambda e$, where λ is a positive scalar constant), we obtain the expression (4) of the control law, where $L_{\underline{s}}^{T+}$ represents the pseudo-inverse matrix of $L_{\underline{s}}^{T}$.

$$T = -\lambda L_{\underline{s}}^{T+}(\underline{s} - \underline{s}^*) \tag{4}$$

The aim of this paper is to study how we can introduce 3D visual features in a robot control loop. In the first section of the paper, we discuss how to extract 3D visual features with a logical sensor. Knowing the object model, the logical sensor uses the De Menthon Algorithm [2] to localize the object in the 3D visual sensor frame. At each iteration we obtain the attitude between the camera and the target object. To validate our approach, we use a specific object composed of four illuminated points. In the second section, with this object as a consideration, we synthesized a specific control law using a 3D point feature and another one with 3D attitude feature. To conclude, we examine the behavior of the robot under various final conditions and the different modeling cases developed. Our experimental site is composed of a six degrees of freedom cartesian robot associated with the parallel vision system "Windis" [1] [8].

LOGICAL VISUAL SENSOR ASPECT

In the case of visual sensor, many works [1] [3] [4] [5] and [7] has been done using 2D visual features extracted from the image plane like point, segment, circle, ellipses,... The main advantages of the vision based control approach is avoidance of camera calibration and reduction of image processing. The desired visual features are extracted from each new image, and do not take into account any calibration operation. At each iteration, visual features are extracted from the image at video rate and compared to the desired configuration directly in the image plane. The main difficulty of this approach is defining the task to achieve in the sensor frame.

In this paper we consider 3D visual features in the robot control loop. The required positioning task is to reach a specific attitude between the sensor frame and a target object frame. At each iteration we need to localize the object in order to evaluate any 3D visual feature. Recent works [2][6] on $3D$ modeling have shown that $3D$ object localization can be done in real time video.

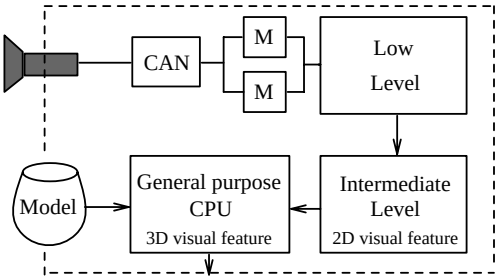

Figure 1. 3D logical visual sensor.

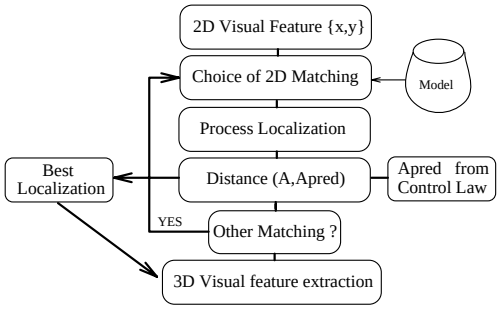

Figure 2. 3D visual features extraction algorithm.

Our approach is to consider a $3D$ logical visual sensor which will be able to provide any of $3D$ visual feature like point (3D coordinates), attitude (position and orientation), line, plane... The logical sensor is composed of a camera, embedded on the end effector of the robot, associated with a general purpose CPU board, which extracts $3D$ visual features. Figure 1 shows the details of the $3D$ visual sensor. At each iteration, $2D$ visual features are extracted from the image plane by a specific hardware (low and intermediate level processors) and sent to the general purpose CPU. With a model of the object to be localize, the CPU processes the De Menthon algorithm [2] in order to extract the object attitude in the sensor frame.

[1] Window distributor

In our experimentation we choose a specific object composed of four illuminated points. Three points are coplanar and the fourth one is fixed on the perpendicular of the plane, which is centered relative to the others. The object frame is placed in the center of the object as shown in figure 3. At each iteration, we compute the algorithm represented in Figure 2. In the image plane, we extract by projection the center of gravity of each illuminated point to obtain a list $\{x_i, y_i\}$ of $2D$ visual features. From this list, we match the $2D$ visual features with $2D$ interesting points of the model to detect. We evaluate the best match by estimating the distance between the processed attitude $\mathcal{A}$, issued from the $2D$ visual feature, and the predicted one $\mathcal{A}_{pred}$ elaborated from the precedent iteration. Then, when all combinations are computed, we obtain the selected attitude and the corresponding $2D$ visual feature. Thereby, with this $3D$ visual information we provide $3D$ visual point, and finally the $3D$ visual attitude.

CONTROL LAWS

We consider a scene with a 3D object and a wrist 3D sensory apparatus mounted on the end effector of the robot. We define three frames as follows: $\mathcal{R}_S$ is a fixed frame attached to the scene, $\mathcal{R}_C$ is a frame attached to the 3D sensor, and $\mathcal{R}_O$ is the object frame.

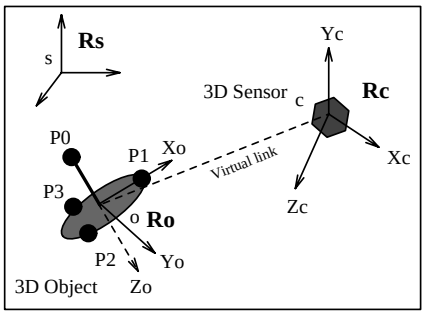

Figure 3: Object and sensor in the scene.

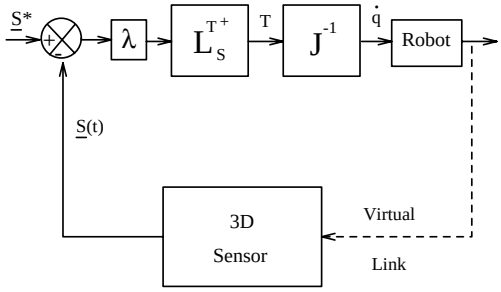

Figure 4: 3D visual servoing scheme

Figure 3 represents the scene with different frames, and figure 4 the visual servoing scheme where J^{-1} is the *Jacobian matrix* which maps the cartesian velocities T of the end effector to differential velocities in joint space $\dot{q}$. We present here two examples of modeling, each concerning the positioning task of the 3D visual sensor.

First law: case of 3D visual point feature ($\underline{\mathcal{P}}$)

First, we consider a 3D visual point $\underline{\mathcal{P}} = (X, Y, Z)^T$ attached to the object. We can write the following relation:

$$\vec{V}(P)\big|_{Rc} = \frac{d\overrightarrow{CS}}{dt}\Big|_{Rs} + \frac{d\overrightarrow{SP}}{dt}\Big|_{Rs} + \vec{\Omega}_{Rs/Rc} \times \overrightarrow{CP} \tag{5}$$

Considering a fixed object, we obtain:

$$\vec{V}(P)\big|_{Rc} = -\vec{V}(C)\big|_{Rs} - \vec{\Omega}_{Rc/Rs} \times \overrightarrow{CP} \tag{6}$$

If we detail this expression we have:

$$\begin{pmatrix} \dot{X} \\ \dot{Y} \\ \dot{Z} \end{pmatrix} = \begin{pmatrix} -V_X - Z\Omega_Y + Y\Omega_Z \\ -V_Y - X\Omega_Z + Z\Omega_X \\ -V_Z - Y\Omega_X + X\Omega_Y \end{pmatrix} \tag{7}$$

In our case, we define the elementary sensor signal as $\underline{s} = \underline{\mathcal{P}} = (X, Y, Z)$. Considering equations (3) and (7), such as $T = (V_X, V_Y, V_Z, \Omega_X, \Omega_Y, \Omega_Z)^T$ is the kinematic screw applied to the 3D sensor frame, we find:

$$L_{\underline{s}}^T = L_{\underline{\mathcal{P}}}^T = \begin{pmatrix} -1 & 0 & 0 & 0 & -Z & Y \\ 0 & -1 & 0 & Z & 0 & -X \\ 0 & 0 & -1 & -Y & X & 0 \end{pmatrix} \tag{8}$$

In our example, the object has four illuminated points. For each of them, we define a 3D visual point attached to the object's center of gravity. We establish the global sensor signal vector such as:

$$\underline{S} = (\underline{\mathcal{P}}_0, \underline{\mathcal{P}}_1, \underline{\mathcal{P}}_2, \underline{\mathcal{P}}_3)^T = (X_0, Y_0, Z_0, X_1, Y_1, Z_1, X_2, Y_2, Z_2, X_3, Y_3, Z_3)^T.$$

So, the global interaction matrix becomes $L_{\underline{S}}^T = (L_{\underline{\mathcal{P}}_0}^T, L_{\underline{\mathcal{P}}_1}^T, L_{\underline{\mathcal{P}}_2}^T, L_{\underline{\mathcal{P}}_3}^T)^T$.

Second law: case of 3D visual attitude feature ($\underline{\mathcal{A}}$)

Secondly, we considered a 3D visual sensor which gives the 3D visual attitude $\underline{\mathcal{A}}$ of an object in the 3D sensor frame. We define parameters for the attitude with the following expression $\underline{\mathcal{A}} = (\underline{\mathcal{P}}, \underline{\mathcal{O}})^T$ where $\underline{\mathcal{P}}$ and $\underline{\mathcal{O}}$ are respectively the position and the orientation of the frame object in the sensor frame. In this case, the global sensor signal vector $\underline{S}$ is composed of $\underline{\mathcal{A}} = (\underline{\mathcal{P}}, \underline{\mathcal{O}})^T = (X, Y, Z, \Theta_X, \Theta_Y, \Theta_Z)^T$. For the orientation $\underline{\mathcal{O}}$, we choose the most relevant parameterization regarding the definition of the Jacobian expression of the robot. Using the roll, pitch, and yaw convention, we have:

$$\begin{pmatrix} \dot{\Theta}_X \\ \dot{\Theta}_Y \\ \dot{\Theta}_Z \end{pmatrix} = \begin{pmatrix} -\Omega_X \\ -\Omega_Y \\ -\Omega_Z \end{pmatrix} \Rightarrow L_{\underline{\mathcal{O}}}^T = \begin{pmatrix} 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix} \tag{9}$$

We use the precedent 3D visual point modeling information to compute $L_{\underline{\mathcal{P}}}^T$. The global interaction matrix is obtained using (8) and (10) and is expressed as:

$$L_{\underline{S}}^T = L_{\underline{\mathcal{A}}}^T = (L_{\underline{\mathcal{P}}}^T, L_{\underline{\mathcal{O}}}^T)^T = \begin{pmatrix} -1 & 0 & 0 & 0 & -Z & Y \\ 0 & -1 & 0 & Z & 0 & -X \\ 0 & 0 & -1 & -Y & X & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix} \tag{10}$$

RESULTS

To test our approach, we implemented the $3D$ logical sensor on the "Windis" parallel vision system [8]. This architecture, developed in collaboration with INRIA[2] in Sophia Antipolis (France), implements the concept of active windows. Figure 5 shows this architecture, where several active windows (of varying size and position) are extracted from the image at video rate. Low level processing ($3 * 3$ or $5 * 5$ convolution) is then executed, and results are sent to the DSP modules. The DSP modules provide a geometric description of the required primitive in each window. The window manager controls all the architecture and computes the "$3D$ logical sensor" features resulting from the geometric description. We use a six degrees of freedom cartesian robot, where the camera is embedded on the end effector. The control laws were evaluated with the two types of modeling and were implemented on our experimental site (see Figure 6).

[2]Institut National de Recherche en Informatique et Automatique

All software was written in C language and we use the VxWorks real time system environment. The first version of the logical sensor processes at twice video rate.

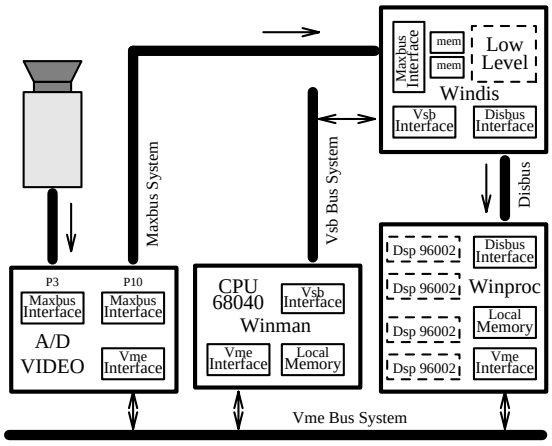

Figure 5 Overview of Windis architecture.

Figure 6 Overview of the robotic platform.

We have tested the two different control laws with various initial and final situations of the visual sensor frame relating to the $3D$ object frame. During our experiments, the results show the same behavior of the two control laws. So we decided to present only the results concerning the $3D$ visual attitude feature. The initial position of the sensor is always $\underline{\mathcal{P}} = (-253, -101, 1500)$ (mm) in front of the object and the orientation $\underline{\mathcal{O}} = (0, 0, 0)$ (deg). For the final attitude $\mathcal{A}_f = (\underline{\mathcal{P}}, \underline{\mathcal{O}})$ we choose a specific position $\underline{\mathcal{P}} = (0, 0, 750)$ (mm) and different orientation $\underline{\mathcal{O}} = (\Theta_X, 0, 0)$ depending on the Θ_X angle $(10, -10, 15, -15, -20)$ (deg). In each case, we estimate the interaction matrix at the equilibrium situation ($L_{\underline{S}}^T = L_{\underline{S}=\underline{S}^*}^T$). We obtained an exponential decay of the error signal and a convergence at around 200 iterations during all tests. We present the results for $\Theta_X = 15$ deg. Figure 7 shows the exponential decay of the error signal and figure 8 the $3D$ points trajectories during the visual servoing (which traduces the object trajectory).

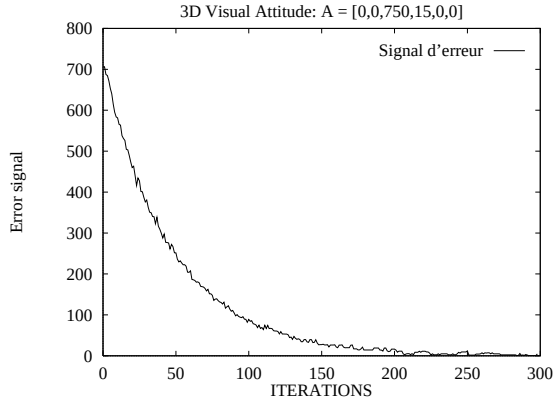

Figure 7 Error signal.

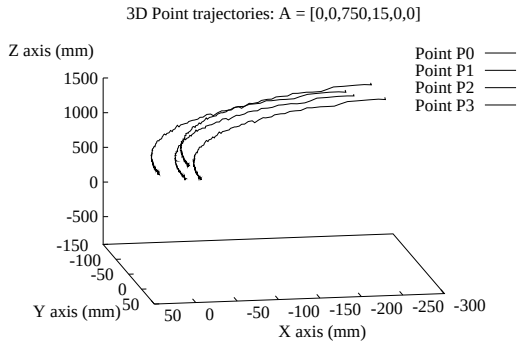

Figure 8 Object trajectory.

In Figures 9 and 10, the translation and rotation velocities are presented. The behavior is like a first order system. We noticed a little disturbance on the curves, due to the $3D$ logical sensor sensibility (less than 1 degree per second on the rotation velocities). This disturbance is due to the difficulty to extract the Z position and the orientation $\underline{\mathcal{O}}$ of the object in the sensor frame in a dynamic sequence. To improve the results of the $3D$ logical sensor we think that we have to use a filter on the attitude parameters.

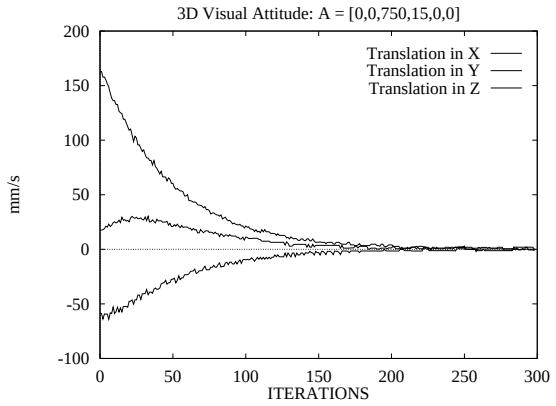

Figure 9 Translation velocities.

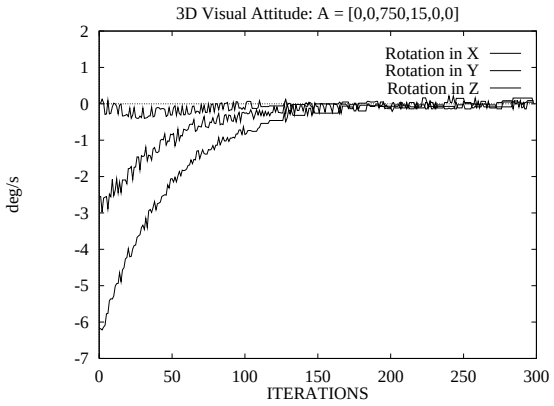

Figure 10 Rotation velocities.

CONCLUSION

Many people are interested in visual servoing. Until now, only $2D$ visual features has been considered. In this paper, we show that 3D visual sensor elaborating 3D features at video rate is not yet perfect. We have tested two control laws based on $3D$ visual features. Results are satisfactory even though we observed a slight disturbance due to the $3D$ logical sensor. As we remarked in this work, the main difficulties are the noise introduced by the logical sensor, and the realization of this one. The main focus of this approach is to control the robot in the task space ($3D$ in this case) where it moves. For the future, we want to analyze more precisely the sensibility of the $3D$ visual logical sensor and improve its efficiency. Also, we want hope to compare the $2D$ and $3D$ approaches in visual servoing.

REFERENCES

1. Allen P.K., B. Yoshimi, A. Timcenko. "Real time visual servoing", Int. Proc. of the IEEE Conference on Robotics and Automation, 1991.
2. De Menthon D., Davis L.S. "Inverse perspective of a triangle : new exact and approximate solutions", IEEE Trans. on Robotics and Automation, vol. 14, n.11, pp.1100-1105, November 1992.
3. Espiau B., F. Chaumette, P Rives. "A new approach to visual servoing in robotics", IEEE Trans. on Robotics and Automation, vol. 8, n.3, pp.313-326, June 1992.
4. Khadraoui D., G. Motyl, P. Martinet, J. Gallice , F. Chaumette. "Visual Servoing in Robotics Scheme Using a Camera/Laser-Stripe Sensor", to appear in IEEE Transactions on Robotics and Automation.
5. Martinet P., F. Berry, J. Gallice. "Use of first derivative of geometric features in Visual Servoing", IEEE International Conference on Robotics and Automation, ICRA'96, Minneapolis, USA, April 1996.
6. Montagne E., J. Alizon, P. Martinet and J. Gallice. "Real Time 3D Location of a car from three characteristic points observed in a video image sequence", 7th IFAC/IFORS Symposium on Transportation Systems, pp.385-390, Tianjin, China, August 1994.
7. Papanikolopoulos N., P.K. Khosla, T. Kanade. "Visual tracking of a moving target by a camera mounted on a robot: A combination of control and vision", IEEE Trans. on Robotics and Automation, vol. 9, n.1, pp.14-35, Sacramento, USA, February 1993.
8. Rives P., J.J. Borrelly, J. Gallice, and P. Martinet. "A Versatile Parallel Architecture for Visual Servoing Applications", Workshop on Computer Architecture for Machine Perception, CAMP'93, News Orleans, USA, 1993
9. Samson C., M. Le Borgne, B. Espiau. "Robot Control : The Task Function Approach", Oxford University Press, 1991.