

Position based visual servoing using a non-linear approach

Philippe Martinet, J. Gallice

▶ **To cite this version:**

Philippe Martinet, J. Gallice. Position based visual servoing using a non-linear approach. IROS'99 - IEEE/RSJ International Conference on Intelligent Robots and Systems, Oct 1999, Kyongju, South Korea. pp.531-536, 10.1109/IROS.1999.813058 . hal-02465645

HAL Id: hal-02465645
https://inria.hal.science/hal-02465645

Submitted on 4 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Position based visual servoing using a non-linear approach

Philippe Martinet and Jean Gallice
LASMEA, Université Blaise Pascal, UMR 6602 du CNRS,
63177 Aubière cedex, France
martinet@lasmea.univ-bpclermont.fr

Abstract

Vision based control has retained attention of many authors during the last few years. We first have been interested in Image Based Visual Servoing approach and recently we have focused our attention in Position Based Visual Servoing approach. In this paper, our goal is to study how we can introduce 3D visual features in a closed robot control loop. We consider a camera mounted on the end effector of the manipulator robot to estimate the pose of the target object. The required positioning task is to reach a specific pose between the sensor frame and a target object frame. Knowing the target object model, we can localize the object in the 3D visual sensor frame and estimate the pose between the camera and the target object at each iteration. To perform the visual servoing task, we use a non-linear state feedback. We propose a new exact model for parametrization of the pose (position and the orientation of the frame object in the sensor frame). The main advantage of this approach is that camera translation and camera rotation are separately controlled due to use of a particular choice of frames. Convergence and stability have been proved theoretically, and the tests in simulation and on our experimental site show good behaviour using this type of approach.

1 Introduction

Sanderson and Weiss in [15] introduced an important classification of visual servo structures based on two criteria: space of control and presence of joint feedback. So in this classification we distinguish two main approaches:

- *Position Based Control* : image features are extracted from the image and a model of the scene and the target is used to determine the pose of the target with respect to the frame attached to the camera.
- *Image Based Control* : in image based control, pose estimation is omitted, and the control law is directly expressed in the sensor space (image space).

A bibliographical review on Visual Servoing has been compiled in detail by P. Corke in [4]. A tutorial on Visual Servo Control was organised by G. Hager et *al* [8] in 1996. These surveys of the state of the art in the field of visual servoing show that *Image Based Control* has been retained as an alternative to the *Position Based Control* approach. Generally, many authors consider that the *Image Based Control* approach is the better of the two, according to the criteria of camera calibration, hand-eye calibration, robot modelling, scene and target modelling, and also regarding the processing time required to compute the sensor signal. It is clear that the Image Based Control approach does not need precise calibration and modelling, because of the closed loop defined in the sensor space. Much work, e.g. [1, 2, 7, 9, 10, 12], has been done on the camera sensor and the $2D$ space.

The notion of *Task Function* introduced by Samson et *al* in [14], can be used to elaborate a control law in the sensor space. According to this concept, Martinet et *al* in [11] introduce the notion of a $3D$ visual sensor which delivers a $3D$ sensor signal by monocular vision at video rate. Recent progress in pose estimation, localization and $3D$ modelling [5, 6] shows that it is not unrealistic to introduce $3D$ visual information into a closed loop control. Using this assumption, we can synthesize control laws using this kind of information as we do directly with the camera sensor. Another way to obtain the estimation of the pose is to compute a Kalman filter using several visual features, as described in [16]. In this approach, the control law is defined to reach a particular pose between sensor and object frames with a PD controller.
In fact, little work has been done using a $3D$ sensor signal. However, we can remark that precise calibration and modelling are really useful only where

the task to be achieved is expressed in Cartesian space. If the $3D$ reference signal is learned by means of the 3D sensory apparatus and pose estimation algorithm (used in real conditions), as in the *Image Based Approach*, we obtain the same good results in the $3D$ *sensor space*.

In the first part of this paper, we develop the model of the closed loop system using a non-linear state feedback, and analyse the problem of stability and convergence. In the second part, we present results obtained with our experimental robotic platform. We use a specific object composed of four illuminated points in real experimentation. In conclusion, we present some prospects for developments in visual servoing.

2 Modelling the system with the state space formalism

The task to be performed is a positioning task (Pose control) of the end effector frame (sensor frame when the sensor is embedded) relative to the absolute frame attached to the scene.

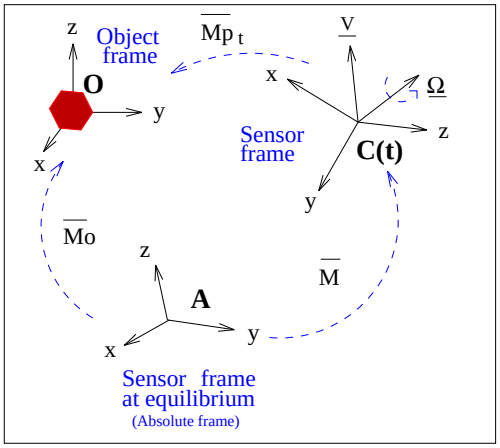


Figure 1: Different frames used in modelling

In the following, 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 homogeneous transformation matrices as follows:

- $\overline{M}_o$ is the homogeneous transformation matrix between an absolute frame $\mathcal{R}_A$ attached to the scene, and the object frame $\mathcal{R}_o$
- $\overline{M}$ is the homogeneous transformation matrix between an absolute frame attached to the scene, and the sensor frame computed at each iteration $\mathcal{R}_{c_t}$
- $\overline{M}_{p_t}$ is the homogeneous transformation matrix between the sensor frame computed at each iteration $\mathcal{R}_{c_t}$, and the object frame $\mathcal{R}_o$

Without loss of generality, we can consider that the absolute frame is equivalent ($\mathcal{R}_A = \mathcal{R}_c^*$) to the sensor frame at the equilibrium situation (see figure 1). The consequences are a simplified representation of the state model.

If we measure the pose of the object $\overline{M}_{p_t}$ in the sensor frame by monocular vision at each iteration, we can deduce the matrix $\overline{M}$ by the following relation:

$$\overline{M} = \overline{M}_o . \overline{M}_{p_t}^{-1} \tag{1}$$

2.1 Principle

The pose parameters of the sensor frame in the absolute frame can be expressed as a rigid transformation matrix $\overline{M}$ as follows:

$$\overline{M} = \begin{bmatrix} R & \underline{x} \\ 0_3 & 1 \end{bmatrix} \tag{2}$$

where R represents the orientation part of the pose, and $\underline{x}$ the position part.

We use the exponential representation for the expression of the rotation. So we write:

$$R = exp(-AS(\underline{\theta}(t))) \tag{3}$$

where $AS(\underline{\theta}(t))$ is the antisymmetric matrix associated with the orientation vector $\underline{\theta}(t)$:

$$\underline{\theta}(t) = \|\underline{\theta}(t)\| . \underline{u}(t) \tag{4}$$

($\|\underline{\theta}(t)\|$ is the norm of the vector $\underline{\theta}(t)$ and $\underline{u}(t)$ is the associated unitary vector).

Deriving relation 2, we obtain:

$$\frac{d}{dt}\overline{M} = \begin{bmatrix} \frac{d}{dt}R & \frac{d}{dt}\underline{x} \\ 0_3 & 0 \end{bmatrix} \tag{5}$$

and after developing, we have:

$$\frac{d}{dt}\underline{x} = [\underline{V}(C)]_A = R.\underline{V} \tag{6}$$

and

$$\frac{d}{dt}R = R.AS(\underline{\Omega}) \tag{7}$$

with the following notations:

- $\underline{V}$ the translation velocity of the camera expressed in the sensor frame $\mathcal{R}_{c_t}$
- $AS(\underline{\Omega})$ the antisymmetric matrix associated with the rotation velocity $\underline{\Omega}$ of the camera expressed in the sensor frame $\mathcal{R}_{c_t}$

In order to transform relations 6 and 7 using the state space formalism, we have to define the state vector $\underline{X}$. We choose a 6-dimension vector $\underline{X} = (\underline{x}^T, \underline{y}^T)^T$ ($\underline{x}^T$ represents the tranpose of $\underline{x}$), where

- $\underline{x}$ is the position of the sensor frame $\mathcal{R}_{c_t}$ expressed in the absolute frame $\mathcal{R}_A$
- $\underline{y}$ is a function of the orientation of the sensor frame $\mathcal{R}_{c_t}$ relative to the absolute frame $\mathcal{R}_A$

We now have to introduce the definition of the variable $\underline{y}$. Developing the exponential representation (Eq. 3) of R with the Rodrigues formulae [13], we can shown that:

$$R^T - R = 2.sin(\|\underline{\theta}(t)\|).AS(\underline{u}(t)) \tag{8}$$

and define the vector $\underline{y}$ as:

$$AS(\underline{y}) = \frac{1}{2}(R^T - R) = sin(\|\underline{\theta}(t)\|).AS(\underline{u}(t)) \tag{9}$$

Then, using Eq. 4 we obtain:

$$\underline{y} = \frac{sin(\|\underline{\theta}(t)\|)}{\|\underline{\theta}(t)\|}.\underline{\theta}(t) \tag{10}$$

Using this representation, we can express the state equation of the system:

$$\frac{d}{dt}\,\underline{X} = B(\underline{X})\,.\,\underline{U} \tag{11}$$

where

- $\underline{U} = (\underline{V}^T, \underline{\Omega}^T)^T$ represents the kinematic screw (the control vector)
- $B(\underline{X}) = \begin{bmatrix} R & O_3 \\ O_3 & A \end{bmatrix}$ with
  - $A = -\frac{1}{2}(trace(R).I_3 - R^T)$
  - $O_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

The state equation of the system is then linear with regard to the control vector $\underline{U}$, and non-linear with regard to the state vector. Controllability of the system is obtained if the control matrix $B(\underline{X})$ is full rank. In our case, this condition is always realized except in the singular case $\|\underline{\theta}\| = \frac{\pi}{2} + k\pi$.

In the conditions used in all theoretical development ($0 \leq \|\underline{\theta}\| < \frac{\pi}{2}$), the inverse of the control matrix $B(\underline{X})$ can be computed, and its expression is the following:

$$B^{-1}(\underline{X}) = \begin{bmatrix} R^T & O_3 \\ O_3 & A^{-1} \end{bmatrix} \tag{12}$$

Developing and using the matrix inversion lemma, we obtain the expression of the matrix A^{-1} defined by:

$$(\frac{1}{cos(\frac{\|\underline{\theta}\|}{2})} - \frac{1}{cos(\|\underline{\theta}\|)})\underline{u}.\underline{u}^T - \frac{1}{cos(\frac{\|\underline{\theta}\|}{2})}.R^T(\frac{1}{2}.\underline{\theta}(t))$$

2.2 Stability analysis

We now proceed to analyse the stability of the control law and then discuss the problems which may be encountered when using a Pose estimation algorithm from image features.

To control the system, we choose a non-linear state feedback which linearizes the closed loop system. In this case, we have:

$$\begin{cases} \underline{U} & = & -B^{-1}(\underline{X}).K.\underline{X} \\ \frac{d}{dt}\underline{X} & = & -K.\underline{X} \end{cases} \tag{13}$$

In these conditions, to stabilize the system it is sufficient to choose the control gain matrix K as a diagonal matrix with positive values. The closed loop system behaves as a set of decoupled integrators, and each component of the state vector has an exponential decrease.

To estimate the pose parameters for 3D objects by monocular vision, many methods are proposed in the litterature. Some methods give closed form solutions of the inverse perspective problem addressed, others use iterative processes to reach the solution. The problem of unicity for the solution is often omitted, and the authors use spatio-temporal filters to extract the right solution. At the present time, it is clear that we are not able to demonstrate that pose measurement is stable and always converges towards the right solution. However, some authors have addressed this kind of problem and some results are known. We

find similar problems when using the *Image Based Approach* as presented by F. Chaumette in [3].

In our application we use the DeMenthon algorithm [6] and we choose the best matching using a spatio-temporal filter. So far, we have not encountered any problems, but this is not a theoretical proof.

3 Experimental results

To validate this approach, we wrote a simulator in Matlab and used our experimental robotic platform. In this paper, we present only a small sample of our experimental results. We use a regular tetrahedron object in the form of four leds as shown in figure 2

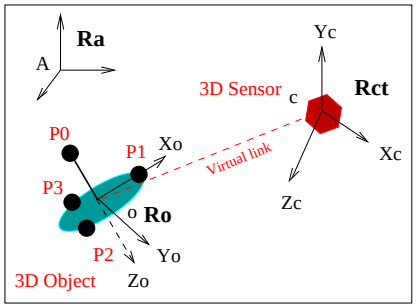


Figure 2: Object and sensor in the scene.

The low level image processing consists of a simple barycenter computation for each projection of the leds. From these four detected points, the DeMenthon algorithm [6] is used to locate the modelled object and thus the control law can be computed. The sampling period is twice the image acquisition time, i.e. 80 ms.

We use a Cartesian robot with six degrees of freedom, the camera being embedded on the end effector (see Figure 3). All software was written in C language and we use the VxWorks real time system environment.

We measure the pose of the object $\overline{M}_{p_t}$ in the sensor frame by monocular vision at each iteration, and we deduce the matrix $\overline{M}$ (Pose of the sensor in the absolute frame) with equation Eq. (1). At the equilibrium situation, the object pose in the sensor frame is represented by the matrix $\overline{M}_o$, and is defined by the programmer using the Cartesian position and the roll, pitch and yaw angles. We have used various initial and final situations between the sensor frame and the object frame and a diagonal control gain



Figure 3: Overview of the robotic platform

matrix $K = 0.5.I_6$.

For this paper, we choose to move back and forward according to the following conditions:

Initial situations $(\mathrm{m},^\circ)$	**Final situations** $(\mathrm{m},^\circ)$
$(0, 0, 0.5, 0, 0, 0)$	$(0, 0, 1.2, 15, -15, -30)$
$(0, 0, 1.2, 15, -15, -30)$	$(0, 0, 0.5, 0, 0, 0)$

For both experiments, in the initial situation the $3D$ sensor give us the following components $(\underline{x}^T_{(0)}, \underline{y}^T_{(0)})$ of the state vector:

$\underline{x}^T_{(0)}$ (m)	$\underline{y}^T_{(0)}$
$(0.173, 0.049, 0.733)$	$(-0.175, 0.303, 0.454)$
$(-0.313, -0.305, -0.619)$	$(0.178, -0.306, -0.453)$

Figure 4 shows the sensor trajectory in Cartesian space during the first experiment. The trajectory is obtained by reconstruction using the joint measurement at each iteration.

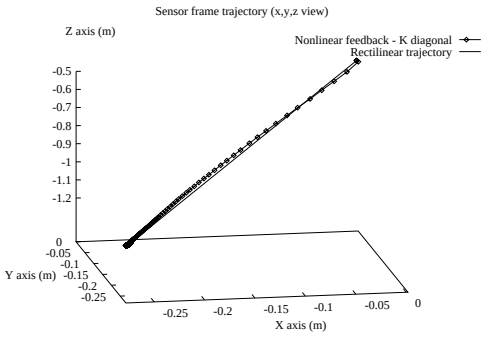


Figure 4: Sensor frame trajectory (scale in m)

The positioning task is well performed, and we can observed that the trajectory of the camera in

the reference frame $\mathcal{R}_c^*$ approximates to a straight line. This result is due to splitting between the control of position and the control of orientation, which represents one of the main advantages of this control law.

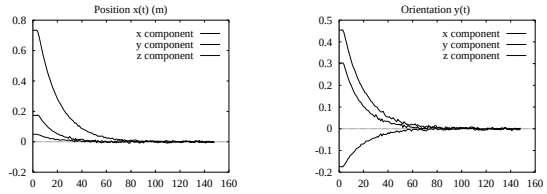


Figure 5: Position $\underline{x}$ and Orientation $\underline{y}$

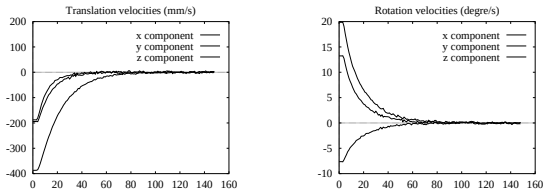


Figure 6: Control vector $(\underline{V}, \underline{\Omega})$

In Figure 5, the position and orientation of the object are presented. The behavior is like a first order system. We obtain exponential decay of the sensor signals and convergence at around 100 iterations. Figure 6 represents the evolution of the control vector.

In the second experiment, from the final position obtained in the first experiment we return to the first initial position. Figure 7 shows the corresponding sensor trajectory in Cartesian space.

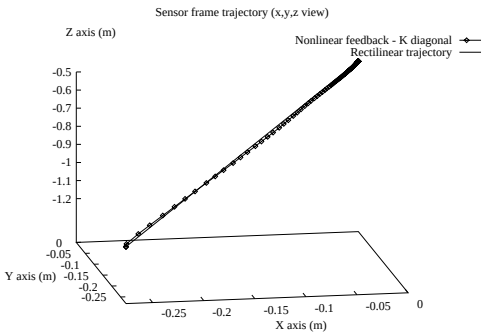


Figure 7: Sensor frame trajectory

Figures 8 and 9 show the exponential decay of the sensor signal and control signal respectively. We

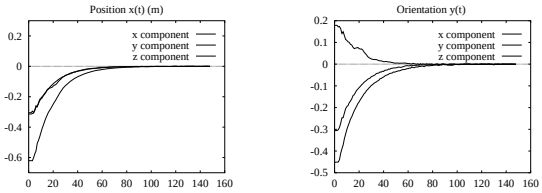


Figure 8: Position $\underline{x}$ and Orientation $\underline{y}$

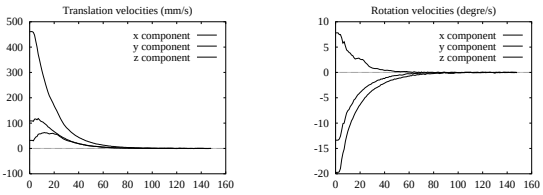


Figure 9: Control vector $(\underline{V}, \underline{\Omega})$

notice a little disturbance on the curves, due to the sensitivity of the $3D$ sensor. This disturbance is due to the difficulty of extracting the Z position and the orientation $\underline{y}$ of the sensor frame in a dynamic sequence, and the sensitivity is greater when the experiment involves forward motion. To improve the results of the control laws we think that one solution is to use a filter on the pose parameters, as Wilson proposes in [16].

4 Conclusion

Many people are interested in visual servoing. Up to now, Image Based visual servoing has principally been considered. In this paper, we show that a 3D visual sensor extracting 3D features at video rate can be used (twice video rate in our implementation). We have developed theoretically an original model for the Pose parameters which simplifies the control synthesis.

We have shown the controllability and stability of this control law, and found the optimal control law which allows the sensor frame to follow a straight line during servoing. Results seem to be satisfactory in regard to primitive extraction (image led barycenter computation) and pose estimation (DeMenthon algorithm) algorithms.

The distinguishing characteristic of this kind of method appears in the simplicity of the formalism. The control law depends only on the desired and

current situations of the observed object. Then, from one application to another, only the pose estimation algorithm has to be modified.

We are now studying ways to introduce a constraint into the control law to be sure that the object is always in the camera field during servoing. Further development should focus on the evaluation of the robustness of the control law with regard to noise estimation in pose extraction, modelling errors, and particularly hand-eye calibration error.

References

[1] P. K. Allen and A. Timcenko and B. Yoshimi and P. Michelman "Automated tracking and grasping of a moving object with a hand-eye system", *IEEE Transactions on Robotics and Automation*, Vol. 9(2), pp. 152-165, 1993.

[2] Espiau B., F. Chaumette, P Rives, "A new approach to visual servoing in robotics", *IEEE Transactions on Robotics and Automation*, vol. 8(3), pp. 313-326, June 1992.

[3] F. Chaumette, "Potentiel problems of stability and convergence in image based and position-based visual servoing", *The confluence of Vision and Control*, LNCIS series, Springer Verlag, 1998.

[4] Corke P., "Visual control of robot manipulators - A review", *in "Visual Servoing"*, Hashimoto K., World Scientific, pp. 1-31, 1993.

[5] S. Christy and R. Horaud, "Iterative pose computation from lines correspondences", *to appear in Computer Vision and Image Understanding*, Academic Press INC., 1998

[6] Dementhon D.F., L.S. Davis, "Model-Based Object Pose in 25 Lines of Code", *International Journal of Computer Vision*, vol. 15(1-2), pp. 123-141, June 1995.

[7] Feddema J.T. and O.R. Mitchell, "Vision-guided servoing with feature-based trajectory generation", *IEEE Transactions on Robotics and Automation*, vol. 5(5), pp. 691-700, October 1989.

[8] Hager G.D., S. Hutchinson, P. Corke, "Tutorial on Visual Servo Control", *IEEE International Conference on Robotics and Automation*, Minneapolis, Minnesota, USA, 22-28 April, 1996.

[9] Khadraoui D., G. Motyl, P. Martinet, J. Gallice , F. Chaumette, "Visual Servoing in Robotics Scheme Using a Camera/Laser-Stripe Sensor", *IEEE Transactions on Robotics and Automation*, vol. 12(5), pp. 743-749, October 1996.

[10] Martinet P., F. Berry, J. Gallice. "Use of first derivative of geometric features in Visual Servoing", *Proceedings of the IEEE International Conference on Robotics and Automation*, Minneapolis, Minnesota, USA, vol.4, pp. 3413-3419, April 1996.

[11] Martinet P., D. Khadraoui, J. Gallice. "Vision Based Control Law using 3D Visual Features", *World Automation Congress*, Montpellier, France, Vol. 3, pp. 497-502, May 1996.

[12] 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 Transactions on Robotics and Automation*, vol. 9(1), pp. 14-35, February 1993.

[13] Rodrigues O. *"Des lois géométriques qui régissent les déplacements d'un système solide dans l'espace, et de la variation des coordonnées provenant de ces déplacements considérés indépendamment des causes qui peuvent les produire"*, Journal de Mathématiques pures et appliquées, Tome 5, pp.380-440, 1840.

[14] Samson C., M. Le Borgne, B. Espiau. *"Robot Control : The Task Function Approach"*, Oxford University Press, 1991.

[15] Sanderson A.C., L.E. Weiss. "Image-based visual servo control using relational graph error signals", *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1074-1077, 1980.

[16] Wilson W.J., C. C. Williams Hulls, G.S. Bell. "Relative End-Effector Control Using Cartesian Position Based Visual Servoing", *IEEE Transactions on Robotics and Automation*, vol. 12(5), pp. 684-696, October 1996.