IEEE Int. Conf. on Robotics and Automation, ICRA2000
San Francisco, California, April 2000

2 1/2 D visual servoing: a possible solution to improve image-based and position-based visual servoings

François Chaumette Ezio Malis*
IRISA / INRIA Rennes
Campus de Beaulieu
35042 Rennes-cedex, France

Abstract

We describe in this paper potential problems that may appear in image-based visual servoing when the initial camera position is far away from its desired position. We show by concrete examples that local minima or a singularity of the image Jacobian can be reached during the servoing. We then recall recent results obtained to avoid these drawbacks. It consists in combining visual features obtained directly from the image, and position-based features. This approach, called 2 1/2 D visual servoing, also provides supplementary advantages in function of the mean the features are combined with.

1 Introduction

The two classical approaches of visual servoing (that is image-based control and position-based control) are different in the nature of the inputs used in their respective control schemes [4, 5]. Even if the resulting robot behaviors thus also differ, both approaches generally give satisfactory results: the convergence to the desired position is reached, and, thanks to the closed-loop used in the control scheme, the system is stable, and robust with respect to camera calibration errors, robot calibration errors, and image measurements errors. However, in some cases, convergence and stability problems may occur, especially when the initial camera position is far away from its desired position. In position-based visual servoing [12], the first drawback is that none control is performed in the image, which implies that the target may get out of the camera field of view during the servoing (leading of course to its failure). The second drawback is that strong hypotheses have to be stated in order to demonstrate the stability of the system [1]. In the following section of this paper, we show that image-based visual servoing also suffers from several drawbacks. More precisely, local minima may be reached, which means that the final robot position does not correspond to the desired one. Furthermore, the image Jacobian may become singular during the servoing, which of course leads to an unstable behavior. Finally, if it is possible to exhibit a sufficient stability condition for image-based visual servoing, it is quite impossible to exploit it in practice. To cope with these problems, a promising approach, already described in [6], consists in combining visual features obtained directly from the image, and features expressed in the Euclidean space. As will be described in Section 3, we thus obtain a block-triangular image Jacobian that provides interesting decoupling properties. It is also possible to be sure that the target will remain in the camera field of view whatever the initial camera position. Thanks to recent results in projective geometry, it is not necessary to know any CAD-model of the considered object. It is also possible to obtain analytical conditions to ensure the global stability of the system even in the presence of calibration errors. We finally describe a new control scheme that belongs to the 2 1/2 D visual servoing approach and also allows the camera trajectory to be a straight line in the Cartesian space.

*Ezio Malis is currently at Cambridge University, England.

2 Potential problems in image-based visual servoing

Image-based visual servoing is based on the selection in the image of a set f of visual features that has to reach a desired value f^*. Usually, f is composed of the image coordinates of several points belonging to the considered target. It is well known that the image Jacobian J plays a crucial role in the design of the possible control laws. Using a classical perspective projection model with unit focal length, and if x and y coordinates of image points are selected in f, two successive rows of J are given by:

$$\begin{pmatrix} -1/Z & 0 & x/Z & xy & -(1+x^2) & y \\ 0 & -1/Z & y/Z & 1+y^2 & -xy & -x \end{pmatrix} \quad (1)$$

where Z is the depth of the corresponding point.

Most control schemes that compute the camera velocity $\dot{r}$ sent to the robot controller have the following forms:

$$\dot{r} = g(\widehat{J^+}(f - f^*)) \tag{2}$$

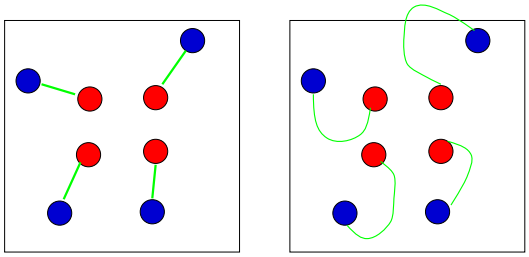
where function g may be as simple as a proportional gain or may be a more complex function used to regulate f toward f^*, and where $\widehat{J^+}$ is a model, an approximation, or an estimation of the pseudo-inverse of J. Indeed, camera calibration errors, noisy image measurements, and unknown depth Z involved in (1) imply the use of such model, since the real value of J remains unknown.

A well known sufficient condition to ensure the global asymptotic stability of the system is [11]:

$$\widehat{J^+}J(f(t), Z(t)) > 0 \ , \ \forall t \tag{3}$$

This condition, even if it is difficult to exploit in practice, allows one to set the possible choices for $\widehat{J^+}$. In fact, three different cases have been considered in the literature:

- $\widehat{J^+} = \widehat{J^+}(t)$. In that case, the image Jacobian is numerically estimated without taking into account the analytical form given by (1). This approach seems to be very interesting if any camera and robot models are available. However, it is impossible in that case to demonstrate when condition (3) is ensured. Furthermore, coarse estimation of the image Jacobian may lead to unstable results.
- $\widehat{J^+} = J^+(f(t), \widehat{Z}(t))$. The image Jacobian is now updated at each iteration of the control law using in (1) the current measure of the visual features and an estimation $\widehat{Z}(t)$ of the depth of each considered point. $\widehat{Z}(t)$ is generally obtained from the knowledge of a 3D model of the object [2]. This case seems to be optimal since, ideally, we thus have $\widehat{J^+}J = I\!I, \forall t$. In that case, each image point is constrained to reach its desired position following a straight line (see Figure 1.a). However, we will see that such a control in the image may imply inadequate camera motion, leading to possible local minima and the nearing of task singularities.
- $\widehat{J^+} = J^+(f^*, \widehat{Z^*})$. In this last case, $\widehat{J^+}$ is constant and determined during an off-line step using the desired value of the visual features and an approximation of the points depth at the desired camera pose. Condition (3) is now ensured only in a neighborhood of the desired position, and a decoupled behavior will be achieved only in a smaller neighborhood. Determining analytically the limits of these neighborhoods seems to be out of reach because of the complexity of the involved symbolic computations. 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 if the initial camera position is far away from its desired one (see Figure 1.b).



a) $J^+(f(t), \widehat{Z}(t))$ b) $J^+(f^*, \widehat{Z^*})$

Figure 1: Possible choices for $\widehat{J^+}$ and corresponding behavior: black points represent the initial position of the target in the image, and gray points and lines respectively represent its desired position and a possible trajectory in the image)

2.1 Reaching or nearing a task singularity

It is well known that the image Jacobian is singular if f is composed by the image of three points such that they are collinear, or belong to a cylinder containing the camera optical center [10]. Using more than three points generally allows one to avoid such singularities. However, whatever the number of points and their configuration, the image Jacobian may become singular during the visual servoing if image points are chosen as visual features. A simple example is described below.

Let us consider that the camera motion from its initial to desired poses is a pure rotation of 180 dg around the optical axis. If $J^+(f(t), \widehat{Z}(t))$ is used in the control scheme and perfect measurements and estimations are assumed, we can note that $\widehat{J^+}J = I\!I$ for the initial camera position, which leaves us to expect a correct behavior. However, the obtained image trajectory of each point is a straight line such that all the points lie at the principal point at the same instant (see Figure 2.a). It corresponds to a pure backward translational camera motion along the optical axis (and unfortunately to a zero rotational motion around the optical axis), that moves the camera at infinity. At this unexpected position, the image Jacobian of each point i is given by (see (1)):

$$J_i = \begin{pmatrix} 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \tag{4}$$

Matrices J and $\widehat{J^+}$ are here of rank 2, instead of 6, which of course corresponds to a task singularity, and where condition (3) is no more ensured.

Let us now consider the case where $J^+(f^*, \widehat{Z^*})$ is used in the control scheme. This choice implies that the control law behaves as if the error in the image was as small as possible. It is clear from Figure 2.b (where white points

correspond to such near position) that the obtained camera motion is now a pure forward translational motion along the optical axis (and, once again, without any rotational motion around the optical axis). The camera thus moves directly toward the target, and toward another singularity of J. Indeed, when $Z = 0$, for all points not lying on the optical center, J_i is given by:

$$J_i \rightarrow \begin{pmatrix} \infty & 0 & \infty & \infty & \infty & \infty \\ 0 & \infty & \infty & \infty & \infty & \infty \end{pmatrix} \quad (5)$$

It is interesting to note that, in that case, $J^+(f^*, \widehat{Z^*})$, that is used in the control scheme, is not singular. However, the problem occurs because of the singularity of J, which is involved in condition (3).

In the two previous cases, the reaching of the singularity can be avoided if the camera rotation is less important. However, the coupling between translational and rotational camera motion implies a really unsatisfactory camera trajectory, by the nearing (and then the moving away) of the singularity. In fact, the problem relies in the selection in f of the visual features. For the considered example, the choice of image points coordinates is really inadequate. Indeed, for the same initial position, the singularity can be avoided, and a perfect camera trajectory can be achieved (that is a pure rotational camera motion around its optical axis) if (ρ, θ) parameters describing straight lines in the image are used in f [1].

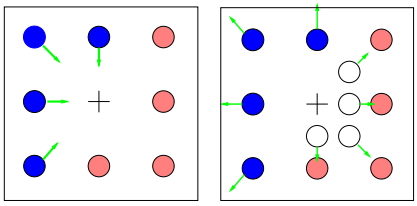


Figure 2: Reaching a singularity: on the left, image motion using $J^+(f(t), \widehat{Z}(t))$; on the right, image motion using $J^+(f^*, \widehat{Z^*})$

2.2 Reaching local minima

We now focus on another potential problem that may appear in practice. By definition, local minima are defined such that $\dot{r} = 0$ and $s \neq s^*$, i.e. such that:

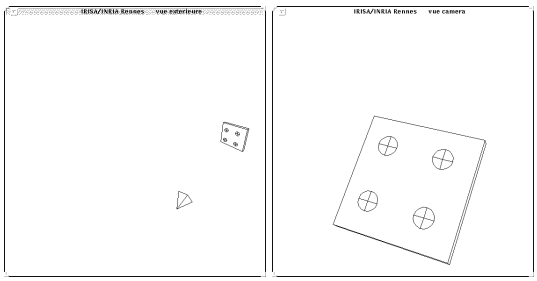
$$f - f^* \in \text{Ker } \widehat{J^+} \quad (6)$$

If f is composed by three image points we have Ker $\widehat{J^+} = 0$ when J is full rank 6. This implies that there is none local minima. However, it is well known that the same image of three points can be seen from four different camera poses. In other words, there exist four camera poses (that is four global minima) such that $s = s^*$). A unique pose can theoretically be obtained by using at least four points. However, in that case, J is of dimension 8×6, which implies that dim Ker $\widehat{J^+} = 2$. This does not demonstrate that local minima always exist. Indeed, the configurations f such that $f - f^* \in$ Ker $\widehat{J^+}$ must be physically coherent (which means that a corresponding camera pose exists). The complexity of the involved symbolic computations seems to make impossible the determination of general results. Particular cases can however be exhibited. In Figure 3 are presented the simulation results for a planar target composed of four points. When $J(f(t), \widehat{Z}(t))$ is used in the control scheme, the visual features simultaneously decrease owing to the used strategy. However, a local minimum is reached since the camera velocity is zero while the final camera position is far away from its desired one. At that position, the error $f - f^*$ in the image does not completely vanish (and is around 1 pixel in the presented example). As explained in [1], reaching local minima is due to the existence of unrealizable motions in the image that are computed by the control law. Using 4 points, the control law indeed enforces 8 constraints on the image trajectory while the system has only 6 dof.

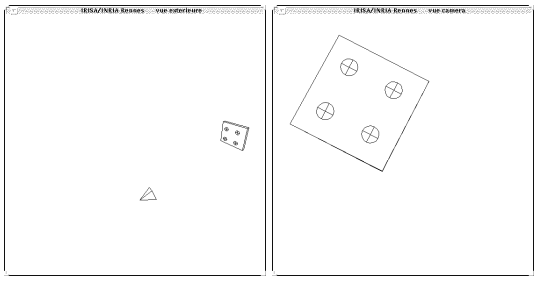
It is interesting to note that the global minimum is reached from the same initial camera position if $J(f^*, \widehat{Z^*})$ is used in the control scheme. In that case, as can be seen on the plots, the trajectory in the image is quite surprising, as well as the computed control law, but this behavior allows the system to avoid the local minima. However, in that case, some points of the target may leave the camera field of view

2.3 Discussion

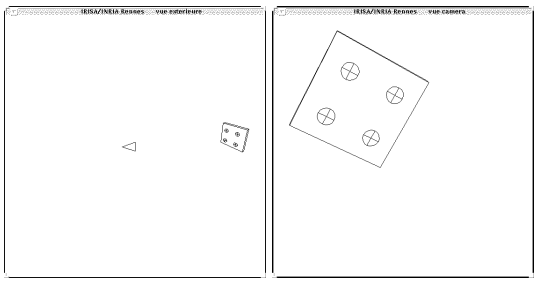
Selecting visual features able to avoid local minima and task singularities whatever the considered target and the initial camera position is a difficult problem that has not been solved yet. Furthermore, other expected properties are that the target always remains in the camera field of view and the camera trajectory is satisfactory in the Cartesian space. When the initial camera position is in the neighborhood of its desired position, using $\widehat{J^+}(f^*, \widehat{Z^*})$ seems to ensure these properties. A solution to cope with the above problems is thus to perform a path planning in the image space and to compute off-line an adequate desired trajectory $f^*(t)$. Ensuring that the error $f(t) - f^*(t)$ always remains small will allow one that condition (3) is also always ensured. This approach has not been investigated yet and we now present another method to improve the behavior of



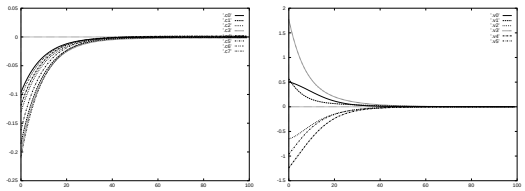
Initial pose and corresponding image



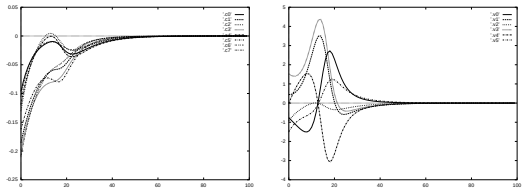
Final pose and image using $J(f(t), \widehat{Z}(t))$



Desired pose and image reached using $J(f^*, \widehat{Z^*})$



$f - f^*$ and $\dot{r}$ using $J(f(t), \widehat{Z}(t))$



$f - f^*$ and $\dot{r}$ using $J(f^*, \widehat{Z^*})$

Figure 3: Reaching (or not) a local minima

image-based (and position-based) visual servoing.

3 2 1/2 D Visual Servoing

2D 1/2 visual servoing consists in combining image features and 3D data. The 3D information can be retrieved either by a classical pose estimation algorithm [2] (if a CAD model of the target is known), either by a projective reconstruction, obtained from the current and desired images [3, 7]. The last case is more interesting, even it is less robust with respect to image measurement errors, since it does not necessitate the knowledge of the 3D shape and dimension of the target. In both cases, the rotation R that the camera has to realize can be computed, as well as any depth ratio.

We select the feature vector as $f = (x, y, z, \theta U^T)^T$ where x and y are the coordinates of an image point, $z = \log Z$ (Z being the depth of the considered point), and where θ and U are the rotation angle and the rotation axis of R. The corresponding image Jacobian is an upper block-triangular matrix given by [6]:

$$J = \begin{pmatrix} \frac{1}{Z} J_v & J_{v\omega} \\ 0_3 & J_\omega \end{pmatrix} \tag{7}$$

where:

$$J_v = \begin{pmatrix} -1 & 0 & x \\ 0 & -1 & y \\ 0 & 0 & -1 \end{pmatrix}$$

$$J_{v\omega} = \begin{pmatrix} xy & -(1+x^2) & y \\ (1+y^2) & -xy & -x \\ -y & x & 0 \end{pmatrix}$$

$$J_\omega = \mathbb{I}_3 - \frac{\theta}{2}\tilde{U} + \left(1 - \frac{\mathrm{sinc}(\theta)}{\mathrm{sinc}^2(\frac{\theta}{2})}\right)\tilde{U}^2 \tag{8}$$

with $\mathrm{sinc}(\theta) = \sin(\theta)/\theta$, $\tilde{U}$ being the antisymmetric matrix associated to U. The determinant of J_ω is

$$\det(J_\omega) = 1/\mathrm{sinc}^2(\frac{\theta}{2}) \tag{9}$$

and it is thus singular only for $\theta = 2k\pi,\ \forall k \in \mathbb{Z}^*$ (i.e. out of the possible workspace). We have also the following nice property:

$$J_\omega^{-1} U\theta = U\theta \tag{10}$$

We can note that the image Jacobian J is singular only in degenerate cases (such as $Z = 0$ and $1/Z = 0$). Finally,

if the the target is known to be motionless and if a simple exponential decrease of $f - f^*$ is specified, we obtain the following control law:

$$\dot{r} = -\lambda \, \widehat{J}^{-1} \, (f - f^*) \tag{11}$$

where λ tunes the convergence rate, the first two components of $f - f^*$ are directly computed from the current and desired images, the last four components of $f - f^*$ are computed from the available 3D data, and $\widehat{J}^{-1}$ is an approximation of J^{-1}.

If the 3D data are computed using a pose estimation algorithm, all terms involved in J are available and the system is globally stable (and has no singularity, nor local minima) under the same strong hypotheses performed in position-based visual servoing (perfect camera calibration, perfect 3D model of the target, perfect image measurement, and perfect pose algorithm). If the 3D data are computed from a projective reconstruction, the depth Z involved in J can be estimated by ρd^* where ρ is available, but not d^*. An approximate value $\widehat{d}^*$ has thus to be chosen before the servoing and introduced in $\widehat{J}^{-1}$. The control law is thus given by:

$$\dot{r} = -\lambda \begin{pmatrix} \rho\widehat{d}^* J_v^{-1} & -\rho\widehat{d}^* J_v^{-1} J_{v\,\omega} \\ 0_3 & \mathbf{I}_3 \end{pmatrix} (f - f^*) \tag{12}$$

Value $\widehat{d}^*$ has not to be precisely determined since it has a small influence on the stability of the system. More precisely, it influences the time-to-convergence of the translational velocity and the amplitude of the possible tracking error due to a wrong compensation of the rotational motion. As far as the tracking error is concerned, it is proportional to the rotational velocity and thus disappears when the camera is correctly oriented. Let us also emphasize that $\widehat{J}^{-1}$ is an upper triangular square matrix without any singularity in the whole task space. Such a decoupled system provides a satisfactory camera trajectory in the Cartesian space. Indeed, the rotational control loop is decoupled from the translational one, and the chosen reference point is controlled by the translational camera d.o.f. such that its trajectory is a straight line in the state space, and thus in the image. If a correct calibration is available, the reference point will thus always remain in the camera field of view whatever the initial camera position. Of course, this property does not ensure that all the target remain visible. In practice, it is possible to change the chosen reference point during servoing, and we can select as reference point the target point nearest the bounds of the image plane. However, this solution leads to a discontinuity in the translational components of the camera velocity at each change of point. Another strategy is to select the reference point as the nearest of the center of gravity of the target in the image. This would increase the probability that the target remains in the camera field of view, but without any complete assurance. In [6], an adaptive control law is proposed to deal with this problem.

Finally, when a projective reconstruction is performed, it is possible to determine, thanks to the nice form of J and $\widehat{J}^{-1}$, the necessary and sufficient conditions for local asymptotic stability, and sufficient conditions for global asymptotic stability in the presence of camera calibration errors. For example, we can determine bounds on $\widehat{d}^*/d^*$ such that the global stability of the system is ensured.

In [9], a similar method is presented. In fact, only the third component of f is different from the one that has been previously presented. This component explicitly takes into account that all the target points have to remain, as far as possible, in the camera field of view, but the triangular form of J_v and $\widehat{J}^{-1}$ is lost. When the target is large in the image, it is impossible to ensure with this scheme that the visibility of all target points will be respected. To deal with this problem, the visual servoing is decomposed in several steps, which implies a quite complex and discontinuous camera trajectory. We now present a simple control scheme, described in [7], that is also in the 2 1/2 D approach, and has interesting practical properties.

If we now select f as $(T^T, x, y, \theta U_z)^T$ where T, expressed in the desired camera frame, is the translation that the camera has to realize, x and y are the coordinates of an image point, and θU_z is the third component of vector θU, we obtain as image Jacobian [12, 7]:

$$J = \begin{pmatrix} R & 0_3 \\ \frac{1}{Z} J_{\omega v} & J_\omega \end{pmatrix} \tag{13}$$

where R is the rotation matrix from current to desired camera frames, and:

$$J_{\omega v} = \begin{pmatrix} -1 & 0 & x \\ 0 & -1 & y \\ 0 & 0 & 0 \end{pmatrix}$$

$$J_\omega = \begin{bmatrix} xy & -(1+x^2) & y \\ (1+y^2) & -xy & -x \\ l_1 & l_2 & l_3 \end{bmatrix}$$

(l_1, l_2, l_3) being the third row of matrix given in (8).

Once again, the image Jacobian is never singular except in degenerate cases and we can apply the following control scheme:

$$\dot{r} = \begin{pmatrix} R^T & 0_3 \\ -\frac{1}{\rho\widehat{d}^*} J_\omega^{-1} J_{\omega v} R^T & J_\omega^{-1} \end{pmatrix} (f - f^*) \tag{14}$$

In that case, the camera translation is specified such that it is a straight line in the cartesian space (which is a particularly satisfactory trajectory), and camera pan and tilt are constrained such that all target points remain in the camera field of view if the selected point is chosen as the nearest of the image limits. These components of the camera velocity will thus be discontinuous at each change of point, but it does not seem to be a crucial aspect. In some particular cases (such that two points are near opposite image limits), it will be impossible to be sure that the visibility constraint will be satisfied, but since these cases are known and not common, they can be easily avoided. A more critical theoretical drawback is that, as in [9], the triangular form of $\widehat{J^{-1}}$ is lost, which makes very difficult, if not impossible, the determination of analytical conditions ensuring the global stability of the system in the presence of camera calibration errors. The experimental results reported in [7] are however satisfactory for both schemes presented in this paper, even if very bad calibration parameters are chosen.

4 Conclusion

In this paper, we have presented the current drawbacks of image-based visual servoing and described a new approach to cope with these drawbacks. A very interesting aspect in 2 1/2 D visual servoing is that, thanks to projective reconstruction, the knowledge of the 3D structure of the considered targets are no more necessary. However, this lack of knowledge implies that the corresponding control laws are more sensitive to image noise than classical image-based visual servoing. Indeed, this latter scheme directly uses visual features as input of the control law without any supplementary estimation step. Path planning in the image is thus one of our current work [8], as well as determining visual features expressed directly in the image and leading to a similar form of the corresponding image Jacobian.

5 Acknowledgment

The authors would like to thank the DER-EDF Chatou for their financial support on part of the work presented in this paper.

References

[1] F. Chaumette, "Potential problems of stability and convergence in image-based and position-based visual servoing", in *The confluence of vision and control*, vol. 237 of *LNCIS Series*, pp. 66-78. Springer Verlag, 1998.

[2] D Dementhon and L. S. Davis, "Model-based object pose in 25 lines of code", *Int. Journal of Computer Vision*, 15(1/2):123-141, June 1995.

[3] O. Faugeras, *Three-dimensional computer vision: a geometric viewpoint*. MIT Press, Cambridge, Massachusetts, 1993.

[4] K. Hashimoto, Ed., *Visual Servoing: Real Time Control of Robot manipulators based on visual sensory feedback*, World Scientific Press, Singapore, 1993.

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

[6] E. Malis, F. Chaumette, and S. Boudet, "2 1/2 D visual servoing", *IEEE Trans. on Robotics and Automation*, 15(2):238-250, April 1999.

[7] E. Malis, *Contributions à la modélisation et à la commande en asservissement visuel*, Ph.D. thesis, Université of Rennes I, IRISA, November 1998.

[8] Y. Mezouar, F. Chaumette, "Path planning in image space for robust visual servoing", *IEEE Int. Conf. on Robotics and Automation, ICRA'00*, San Fransisco, California, April 2000.

[9] G. Morel, T. Leibezeit, J. Szewczyk, S. Boudet, J. Pot, "Explicit incorporation of 2D constraints in vision based control of robot manipulators", *Int. Symp. on Experimental Robotics*, vol. 250 of *LNCIS Series*, pp. 99-108. Springer Verlag, 2000.

[10] N. Papanikolopoulos. "Selection of features and evaluation of visual measurements during robotic visual servoing tasks", *Journal of Intelligent and Robotics Systems*, 13:279–304, 1995.

[11] C. Samson, M. Le Borgne, and B. Espiau, *Robot Control: the Task Function Approach*, Clarendon Press, Oxford, England, 1991.

[12] W. Wilson, C. Hulls, G. Bell,. "Relative end-effector control using cartesian position-based visual servoing", *IEEE Trans. on Robotics and Automation*, 12(5):684-696, October 1996.