Paper Type : Full Paper.

Paper title : Self-calibrated Visual Servoing with respect to Axial-Symmetric 3D Objects.

Authors : Duccio Fioravanti, Carlo Colombo and Benedetto Allotta.

Addresses : Duccio Fioravanti (fioravanti@mapp1.de.unifi.it) and Benedetto Allotta (benedetto.allotta@unifi.it) are with the Departmentof Energetics, University of Florence, Carlo Colombo (colombo@dsi.unifi.it) is with Department of Systems and Informatics, University of Florence, Via di S.Marta 3, 50139 Florence, Italy.

abstract: A self-calibrated approach to visual servoing with respect to non planar targets modeled through a pair of coaxial circles plus one point is discussed. Full calibration data (fixed internal parameters) are obtained from two views, and used to recover Euclidean target structure and camera relative pose. Pose disambiguation is achieved without requiring any real third view of the target. The approach benefits of an off-line planning strategy by which the camera follows a 3D helicoidal path around an arbitrarily chosen axis. A convenient choice for the helicoidal axis is found to be that of the target axis itself. Simulation results demonstrate that the approach is robust with respect to noise both in the off-line and on-line control phases.

keywords:Uncalibrated Visual Servoing, Multiple View Geometry, Axial Symmetry, Helicoidal Path Planning.

corresponding author: Carlo Colombo.

Self-calibrated Visual Servoing with respect to Axial-Symmetric 3D Objects

Duccio Fioravanti [a] Carlo Colombo [b,*] Benedetto Allotta [a]

[a] *University of Florence, Department of Energetics*
Via di S.Marta 3, 50139 Italy

[b] *University of Florence, Department of Systems and Informatics*
Via di S.Marta 3, 50139 Italy

Abstract

A self-calibrated approach to visual servoing with respect to non planar targets modeled through a pair of coaxial circles plus one point is discussed. Full calibration data (fixed internal parameters) are obtained from two views, and used to recover Euclidean target structure and camera relative pose. Pose disambiguation is achieved without requiring any real third view of the target. The approach benefits of an off-line planning strategy by which the camera follows a 3D helicoidal path around an arbitrarily chosen axis. A convenient choice for the helicoidal axis is found to be that of the target axis itself. Simulation results demonstrate that the approach is robust with respect to noise both in the off-line and on-line control phases.

Key words: Uncalibrated Visual Servoing, Multiple View Geometry, Axial Symmetry, Helicoidal Path Planning.

1 Introduction

Thanks to its nice stability and convergence properties even in the presence of modeling and calibration errors, image-based visual servoing has become a popular paradigm for robot control [11, 15]. Image-based approaches basically differ by visual measurements exploited, type of target, and control scheme.

* Corresponding author.
Email addresses: fioravanti@mapp1.de.unifi.it (Duccio Fioravanti), colombo@dsi.unifi.it (Carlo Colombo), benedetto.allotta@unifi.it (Benedetto Allotta).

In particular, visual measurements refer to both the 2D features (e.g., points or lines) tracked in the image during task execution, and the 3D data (e.g., point depths, camera displacement etc. extracted from the images as well) which are used for interaction matrix and/or system state estimation, and trajectory planning, according to the adopted control scheme. Planar targets have been widely investigated in the recent literature. Several types of 2D features have been addressed so as to deal with complex planar shapes and increase task robustness and generality. These include: B-spline snakes [6], active contours [19] and moment invariants [3]. Estimation of 3D data is remarkably simple with planar targets, for which the theory of planar homographies can be used to recover camera displacements and plane normals by matrix decomposition based on standard linear techniques. Planar homography decomposition is affected by a twofold ambiguity, which typically requires a third view to be removed—see e.g. [16]. In the presence of non planar targets, 3D data estimation from image features is more complex, and must rely on standard linear computer vision methods such as the "8-point" algorithm or the "plane plus parallax" decomposition [14]. A non linear method for 3D data estimation has also been proposed in [5].

Diverse types of control schemes have also been discussed in the past few years. In [4] it is remarked that pure image-based control schemes may have problems of singularities and local minima in the absence of 2D feature trajectory planning, especially when the initial and final views are very different. To overcome these problems, two different extensions to the basic "pure image-based/no planning" scheme have been proposed recently. The first one, referred to as "2-1/2-D visual servoing," adopts a hybrid state vector composed by both 2D and 3D visual measurements [18]. This scheme, while ensuring global control stability, is both sensitive to noise close to convergence, and unable to compute in advance target visibility conditions. The second solution uses 2D trajectory planning and introduces a feedforward term in the control law [1, 21]. Although unable to ensure a global control stability, trajectory planning has the advantage of letting constraints such as those on camera velocity profiles and target visibility be met, while keeping the robustness of pure image-based control.

Visual servoing methods typically require the a priori knowledge of internal calibration data. A recent research trend is that of the development of strategies that relax this requirement, thus allowing to work with uncalibrated settings. In the approach proposed in [17], image features are mapped onto a special projective space which is invariant to changes in camera intrinsics. Although the control error is invariant with respect to camera intrinsics, these parameters are still required to estimate the Jacobian; for this reason, calibration errors may affect the stability of the control law. A second way to avoid off-line calibration is to rely on self-calibration techniques developed by the computer vision community, and exploit the (partial) knowledge of scene

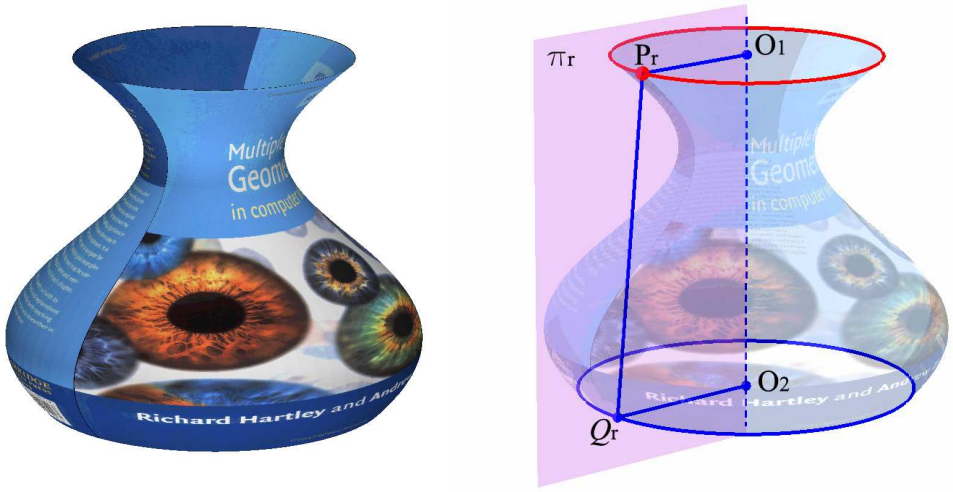
structure and/or camera parameters to perform on-line camera calibration—see e.g. [14]. The main limitation of self-calibration approaches is that they require a priori information that may not be available.

In this paper, a self-calibrated approach to visual servoing with respect to a wide class of non planar targets is presented. Valid targets are those that can be modeled through a pair of coaxial circles plus one point on either circle plane. This a priori assumption on target shape makes it possible to self-calibrate a full perspective camera (fixed internal parameters) from two views, by exploiting the method originally developed in [7] for a single view of a solid of revolution. The special target structure also allows the reformulation of the servoing problem in terms of camera positioning with respect to a virtual planar target. Calibration data are used to recover the scaled Euclidean structure of the virtual plane, together with scaled camera relative pose. Pose disambiguation is achieved by generating a synthetic third view of the target from the second one. The approach includes an off-line 2D trajectory planning strategy by which the camera follows a 3D helicoidal path around an arbitrary axis, which can be conveniently chosen as the target axis itself. Simulations results demonstrate the feasibility of the approach, and its robustness with respect to noisy image data, both in the off-line and on-line control phases. The paper is organized as follows. In Sect. 2.1 the target model is defined; its imaged properties are then discussed in Sect. 2.2 together with calibration and 3D estimation issues. In Sect. 2.3 the control scheme is described, and an algorithm for helicoidal path planning is given. Sect. 3 includes results on simulation tests. Finally, in Sect. 4 conclusions are drawn, and directions for future work are outlined.

2 The Approach

2.1 Modeling

The visual servoing approach proposed in this paper works for a class of targets which can be modeled through a pair of coaxial circles. This geometric structure is shared by a large class of man-made objects, including solids of revolution such as bottles, vases, lamps and any lathe-crafted objects as a special case. An additional requirement for targets is the existence of a well detectable point different from a circle center, and placed anywhere on either circle planes: this point (referred to in this work as *reference point* $\mathbf{P}_r$) can be related either to the visual texture of the object, or to special artifacts on it—e.g., the handle of a cup. Fig. 1(a) provides an example of solid of revolution target.



(a) (b)

Fig. 1. A solid of revolution target and its geometry. (a): A view of the target. (b): The features relevant for the approach.

Only two views of the target, referred to as *initial* and *desired* view, are required in the proposed approach. Moreover, the approach is self-calibrated, the only condition on the camera internal parameters being that they remain unchanged during task execution. For any target, the plane π_r through the circles' axis and the reference point can be defined. Consequently, the servoing task can be reformulated in terms of initial and final visual appearance of a *reference quadrangle* on π_r, with vertexes $\mathbf{P}_r$, $\mathbf{O}_1$, $\mathbf{O}_2$, $\mathbf{Q}_r$—see Fig. 1(b).

2.2 Vision

2.2.1 Imaged target and related 2D features

Anytime during task execution, perspective projection of the 3D object onto the image plane gives rise to the two ellipses $\mathtt{C}_1$ and $\mathtt{C}_2$ (i.e., the images of the two coaxial circles, represented as 3×3 symmetric homogeneous matrices), and the imaged reference point $\mathbf{p}_r$ (represented as a homogeneous 3-vector). The axial symmetry properties of the 3D target induce a projectively symmetric 2D configuration of the ellipses, which is characterized by the imaged symmetry axis $\mathbf{l}_s$ and the vanishing point $\mathbf{v}_\infty$ of the normal direction of the plane passing through $\mathbf{l}_s$ and the camera center. Two other fundamental entities characterizing the imaged target are the imaged circular points $\mathbf{i}$ and $\mathbf{j}$ of the pencil of planes orthogonal to the target axis. As shown in [7], the image entities $\mathbf{l}_s$, $\mathbf{v}_\infty$, $\mathbf{i}$ and $\mathbf{j}$ can all be computed from the knowledge of the ellipses $\mathtt{C}_1$ and $\mathtt{C}_2$ and their conditions of visibility. Given $\mathtt{C}_1$, $\mathtt{C}_2$ and $\mathbf{p}_r$ (which without loss of generality will be assumed as belonging to $\mathtt{C}_1$), the vanish-

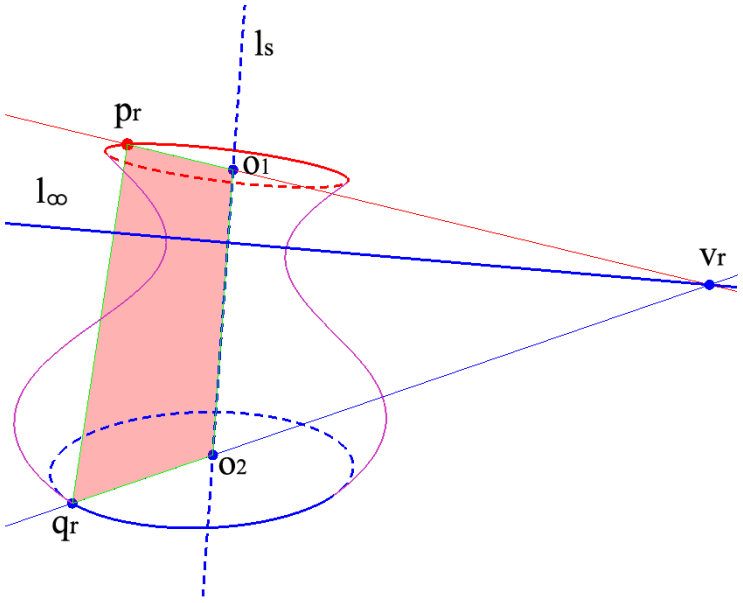


Fig. 2. Construction of the imaged target geometry.

ing line $\mathbf{l}_\infty = \mathbf{i} \times \mathbf{j}$ of the plane pencil above is used at any time during task execution to compute the remaining imaged vertexes of the reference quadrangle, namely $\mathbf{o}_1$, $\mathbf{o}_2$, and $\mathbf{q}_r$—see Fig. 2. Explicitly, the imaged circle centers are computed from the pole-polar relationships $\mathbf{o}_1 = \mathtt{C}_1^{-1}\mathbf{l}_\infty$ and $\mathbf{o}_2 = \mathtt{C}_2^{-1}\mathbf{l}_\infty$. Once these are known, the vanishing point of the line $\mathbf{O}_1 \times \mathbf{P}_r$ is computed as $\mathbf{v}_r = \mathbf{o}_1 \times \mathbf{p}_r \times \mathbf{l}_\infty$. Finally, the fourth vertex $\mathbf{q}_r$ is obtained by first intersecting the line $\mathbf{o}_2 \times \mathbf{v}_r$ with the ellipse $\mathtt{C}_2$, and then choosing, between the two possible solutions, the one farthest from $\mathbf{v}_r$. During task execution, the position of the imaged vertexes of the reference quadrangle are computed at each frame by estimates of both the two ellipses and the reference point. Occlusion or illumination problems occurring during reference point visual extraction can be dealt with at control time, using robust approaches such as [13] or [8].

2.2.2 Self-calibration

The imaged target axis $\mathbf{l}_s$ and the vanishing point $\mathbf{v}_\infty$ are related to each other as [22]:

$$\mathbf{l}_s = \omega\, \mathbf{v}_\infty \ , \tag{1}$$

where $\omega = (\mathtt{K}\mathtt{K}^{\mathrm{T}})^{-1}$ carries information about the camera calibration matrix $\mathtt{K}$ and its internal parameters. Other two constraints on $\mathtt{K}$ which are induced by the special structure of the target are:

$$\mathbf{i}^{\mathrm{T}}\omega\, \mathbf{i} = \mathbf{j}^{\mathrm{T}}\omega\, \mathbf{j} = 0 \ . \tag{2}$$

In [7] it is demonstrated that, for each view, the system composed by eqs. 1 and 2 provides three independent constraints on ω, and hence on the five internal camera parameters. Therefore, under the hypothesis of unchanged $\mathtt{K}$,

a full camera calibration can be obtained from the image data of both the initial and final target views.

2.2.3 Scaled 3D measurements

For the purpose of trajectory planning (addressed in detail in Sect. 2.3), estimates of the 3D structure of the reference quadrangle and the relative camera displacement are required. These can be obtained in three steps. First, the 3×3 planar homography $\mathtt{H}$, mapping the corresponding vertexes of the initial and final imaged quadrangles, is computed as shown in [14] as the solution of

$$[{}^{f}\mathbf{p}_r \;\; {}^{f}\mathbf{o}_1 \;\; {}^{f}\mathbf{o}_2 \;\; {}^{f}\mathbf{q}_r] = \mathtt{H}\,[{}^{i}\mathbf{p}_r \;\; {}^{i}\mathbf{o}_1 \;\; {}^{i}\mathbf{o}_2 \;\; {}^{i}\mathbf{q}_r] \;\; . \tag{3}$$

The calibration matrix $\mathtt{K}$, estimated as above, is then used to compute the so called "Euclidean homography" $\mathtt{G} = \mathtt{K}^{-1}\mathtt{H}\mathtt{K}$. Finally, using the algorithm reported in [16], the Euclidean homography can be decomposed into its rotational and parallax components as

$$\mathtt{G} = {}^{i}\mathtt{R}_f^{\mathrm{T}}(\mathtt{I} - {}^{i}\widetilde{\boldsymbol{\tau}}_f \, {}^{i}\widetilde{\mathbf{n}}_r^{\mathrm{T}}) \;\; , \tag{4}$$

where "$\sim$" denotes inhomogeneous vectors, and

- ${}^{i}\mathtt{R}_f$ is the rotation matrix between the initial ($\mathcal{F}_i$) and final ($\mathcal{F}_f$) camera frames;
- ${}^{i}\widetilde{\boldsymbol{\tau}}_f$ is the scaled translation between $\mathcal{F}_i$ and $\mathcal{F}_f$;
- ${}^{i}\widetilde{\mathbf{n}}_r$ is the unit normal to π_r.

Notice that all the elements are expressed with respect to $\mathcal{F}_i$. It is well known that the planar homography decomposition problem has a twofold ambiguity, which can be removed by the additional information provided by a third view of the plane. In this work, a real third view of the target is actually not needed, being it sufficient to exploit the symmetry properties of the target. To this aim, a *synthetic view* of the quadrangle can be generated by arbitrarily choosing, in the final image, a new point $\mathbf{p}_v \in \mathtt{C}_1$ other than $\mathbf{p}_r$, and using the method expounded above to obtain the corresponding fourth quadrangle vertex (Fig. 3).

2.3 Control

2.3.1 Control scheme

Given a generic image point $\mathbf{x} = [x \;\; y \;\; 1]^{\mathrm{T}}$ in homogeneous pixel coordinates, the vector $\mathtt{K}^{-1}\mathbf{x} = [\breve{\mathbf{x}}^{\mathrm{T}} \;\; 1]^{\mathrm{T}}$ provides the direction of the optical ray through $\mathbf{x}$, which can also be regarded as an image point viewed by an *ideal camera* with

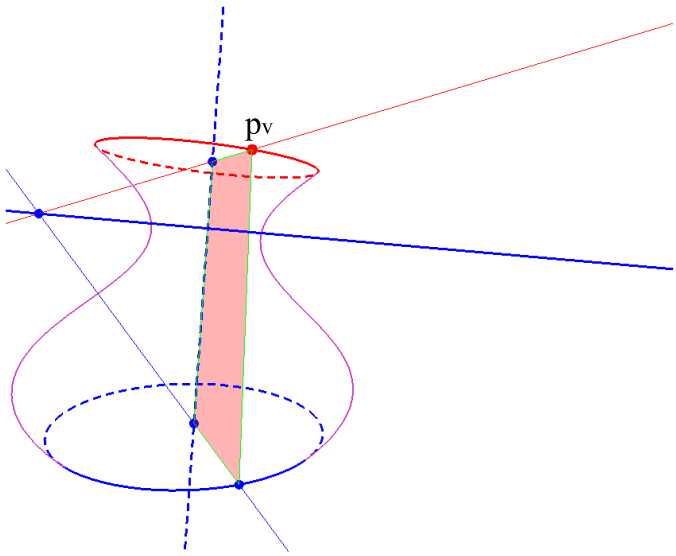


Fig. 3. Construction of the synthetic view for pose disambiguation.

calibration matrix diag(1, 1, 1). Now, let $\boldsymbol{\varphi} = [\breve{\mathbf{p}}_r^{\mathrm{T}} \;\; \breve{\mathbf{o}}_1^{\mathrm{T}} \;\; \breve{\mathbf{o}}_2^{\mathrm{T}} \;\; \breve{\mathbf{q}}_r^{\mathrm{T}}]^{\mathrm{T}}$ be the 8-vector constructed using, in a generic view, the imaged vertexes of the reference quadrangle. The system kinematics can be expressed as

$$\dot{\boldsymbol{\varphi}} = \mathtt{J}\,\mathbf{w} \;\; , \tag{5}$$

where $\mathbf{w} = [\mathbf{v}^{\mathrm{T}} \;\; \boldsymbol{\omega}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{6\times 1}$ is the twist velocity screw of the camera expressed in the camera frame. Defined the image Jacobian as $\mathtt{J} = [\mathtt{J}_1^{\mathrm{T}} \;\; \mathtt{J}_2^{\mathrm{T}} \;\; \mathtt{J}_3^{\mathrm{T}} \;\; \mathtt{J}_4^{\mathrm{T}}]^{\mathrm{T}}$, its generic block $\mathtt{J}_k$ has the form:

$$\begin{bmatrix} \dfrac{-1}{{}^cZ_k} & 0 & \dfrac{\breve{x}_k}{{}^cZ_k} & \breve{x}_k\breve{y}_k & -\left(1+\breve{x}_k^2\right) & \breve{y}_k \\ \\ 0 & \dfrac{-1}{{}^cZ_k} & \dfrac{\breve{y}_k}{{}^cZ_k} & \left(1+\breve{y}_k^2\right) & -\breve{x}_k\breve{y}_k & -\breve{x}_k \end{bmatrix} . \tag{6}$$

$\mathtt{J}_k$ depends on both the ideal image coordinates $(\breve{x}_k, \breve{y}_k)$ and depth ${}^cZ_k$ (camera coordinates) of the associated k-th 3D point.

A control law for the above system that exhibits nice robustness properties with respect to modeling errors and a fast convergence rate [1, 21] is

$$\mathbf{w} = \hat{\mathtt{J}}^{\dagger}[\lambda\mathbf{e} + \dot{\boldsymbol{\varphi}}_d] \;\; , \tag{7}$$

where $\hat{\mathtt{J}}^{\dagger}$ is the pseudo-inverse of the estimated Jacobian obtained through the on-line adaptive depth estimation law proposed in [9], λ is a positive gain, $\mathbf{e} = \boldsymbol{\varphi}_d - \hat{\boldsymbol{\varphi}}$ is the image error (difference between the desired and estimated imaged quadrangle vertexes), and $\dot{\boldsymbol{\varphi}}_d$ is the desired evolution of quadrangle appearance resulting from path planning. By exploiting the kinematic parameters obtained from the homography decomposition of Eq. 4 together with the

planned 3D trajectory, it is possible to estimate all the reference quadrangle points depths up to a scalar factor. Depth ratios are then used to estimate real point depths. Notice that estimated depths don't affect system stability, since they are computed up to a positive scalar factor (see also [20]).

2.3.2 Path planning

In this work, off-line image path planning is performed by generating, at each time step t, a Euclidean homography $\mathtt{G}_d(t)$ by which the desired imaged quadrangle is given by

$$\mathtt{K}^{-1}[\mathbf{p}_r(t)\ \ \mathbf{o}_1(t)\ \ \mathbf{o}_2(t)\ \ \mathbf{q}_r(t)] = \mathtt{G}_d(t)\ \mathtt{K}^{-1}[{}^i\mathbf{p}_r\ \ {}^i\mathbf{o}_1\ \ {}^i\mathbf{o}_2\ \ {}^i\mathbf{q}_r]\ . \tag{8}$$

The planned Euclidean homography is designed so that the camera follows a helicoidal trajectory in the 3D space, whose radius and axis depend on the overall displacement $[{}^i\mathtt{R}_f\ {}^i\widetilde{\boldsymbol{\tau}}_f]$ between the initial and current camera frames. Helicoidal trajectories exhibit the nice property of combining translations and rotations in a natural way. In [2], the special case of helicoidal planning with constant radius and axis identical to the instantaneous axis of motion has been discussed. In the present work, the general case of variable radius and arbitrary axis is addressed.

Following the decomposition of eq. 4, the planning homography can be written as

$$\mathtt{G}_d(t) = {}^i\mathtt{R}_d^{\mathrm{T}}(t)(\mathtt{I} - {}^i\widetilde{\boldsymbol{\tau}}_d(t)\,{}^i\widetilde{\mathbf{n}}_r^{\mathrm{T}})\ , \tag{9}$$

where $[{}^i\mathtt{R}_d(t)\,{}^i\widetilde{\boldsymbol{\tau}}_d(t)]$ denotes the desired 3D displacement of the camera at time t with respect to frame $\mathcal{F}_i$. The desired rotation ${}^i\mathtt{R}_d(t)$ can be expressed through the Rodrigues' formula [16] as

$${}^i\mathtt{R}_d^{\mathrm{T}}(t) = \mathtt{I} + \left[{}^i\mathbf{u}\right]_\times \sin(\vartheta_{if}\,\xi(t)) + \left[{}^i\mathbf{u}\right]_\times^2 [1 - \cos(\vartheta_{if}\,\xi(t))]\ , \tag{10}$$

where $[\,\cdot\,]_\times$ is the skew-symmetric operator, $({}^i\mathbf{u}, \vartheta_{if})$ is the axis-angle representation of the overall camera rotation matrix ${}^i\mathtt{R}_f$, and $\xi(t)$ is any smooth monotonic function of time such that $\xi(t_i) = 0$ and $\xi(t_f) = 1$. In order to plan a helicoidal trajectory around an arbitrary 3D axis L, the desired scaled translation ${}^i\widetilde{\boldsymbol{\tau}}_d(t)$ is computed by the following algorithm—see also Fig. 4.

i) Let π_i and π_f be the planes through the target axis L and the initial and final camera centers, respectively. Compute the unit vectors ${}^i\mathbf{n}_i$ and ${}^i\mathbf{n}_f = {}^i\mathtt{R}_f\,{}^f\mathbf{n}_f$ normal to the planes using the imaged axes $\mathbf{l}_i$ and $\mathbf{l}_f$ and the formula $\mathbf{n} = \mathtt{K}^{\mathrm{T}}\mathbf{l}/\|\mathtt{K}^{\mathrm{T}}\mathbf{l}\|$.

ii) Compute the direction of the helicoidal axis as ${}^i\mathbf{u}_l = ({}^i\mathbf{n}_i \times {}^i\mathbf{n}_f)/\|{}^i\mathbf{n}_i \times {}^i\mathbf{n}_f\|$. Compute the vector component ${}^i\widetilde{\boldsymbol{\tau}}_f^{\perp}$ perpendicular to ${}^i\mathbf{u}_l$ of the overall scaled translation ${}^i\widetilde{\boldsymbol{\tau}}_f$.

iii) Compute the (scaled) distances between L and the initial and final camera centers respectively as $r_i = |\,{}^i\widetilde{\boldsymbol{\tau}}_f^{\perp} \times {}^i\mathbf{n}_i|$ and $r_f = |\,{}^i\widetilde{\boldsymbol{\tau}}_f^{\perp} \times {}^i\mathbf{n}_f|$. Compute the angle between the planes π_i and π_f as

$$\alpha_{if} = \arccos\left(\frac{r_i^2 + r_f^2 - \|{}^i\widetilde{\boldsymbol{\tau}}_f^{\perp}\|^2}{2r_i r_f}\right) \quad . \tag{11}$$

iv) Compute the (scaled) pitch of the planned helix as

$$p_d = \frac{2\pi}{\alpha_{if}}\, {}^i\widetilde{\boldsymbol{\tau}}_f^{\mathrm{T}}\mathbf{u}_l \quad . \tag{12}$$

v) Define the right-handed coordinate frame $\mathcal{F}_h$ with the following characteristics: (a) (scaled) origin on L, at the point at minimum distance from the initial camera center; (b) x-axis orthogonal to L and directed towards the initial camera center; (c) z-axis parallel to $\mathbf{u}_l$. The planned (scaled) helix equations in this frame are

$${}^h\widetilde{\boldsymbol{\tau}}(t)_d = \begin{bmatrix} r(t)\cos(\alpha_{if}\xi(t)) \\ r(t)\sin(\alpha_{if}\xi(t)) \\ (p_d/2\pi)\alpha_{if}\xi(t) \end{bmatrix} \quad , \tag{13}$$

where $r(t) = r_i + (r_f - r_i)\,\xi(t)$ represents the variable helicoidal radius.

vi) Finally, compute the desired scaled translation ${}^i\widetilde{\boldsymbol{\tau}}_d(t)$ from ${}^h\widetilde{\boldsymbol{\tau}}(t)_d$ by the change of frame mapping $\mathcal{F}_h \mapsto \mathcal{F}_i$.

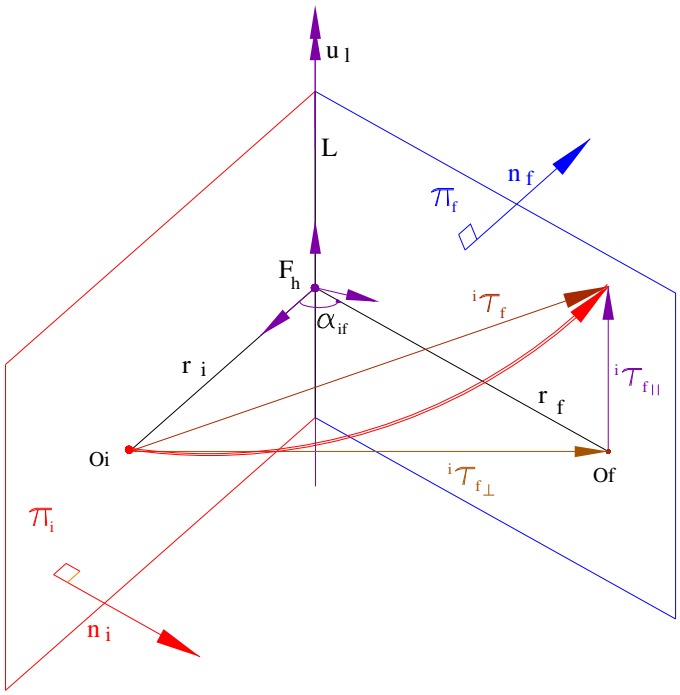


Fig. 4. Geometric construction for the desired scaled translation.

Notice that, as ${}^i\widetilde{\boldsymbol{\tau}}_f$ is defined up to a scale factor, the lack of knowledge about the real distance between the initial and final camera centers does not affect the planning in any way.

Although the algorithm above applies to any arbitrarily chosen axis, the special structure of the target suggests that the most natural choice for the axis is that of the target axis itself. In Sect. 3, this choice will be discussed and compared with the one proposed in [2]. Fig. 5 shows the three planar loci for camera center where singularities occur. In planes π_1 and π_2, one of the ellipses degenerates into a line segment; in π_r the vertexes of the imaged quadrangle are all aligned. Singularities occurring on-line can be pointed out by visual analysis: They can be bypassed by opening the control loop near degeneracy conditions, and using a "reduced" control law including the feedforward term and the last useful estimate of the imaged Jacobian [1]. Choosing the helicoidal axis coincident with the target axis has the further advantage that degenerate ellipses are avoided, provided that both the initial and final camera centers both lie below, in the middle or above π_1 and π_2.

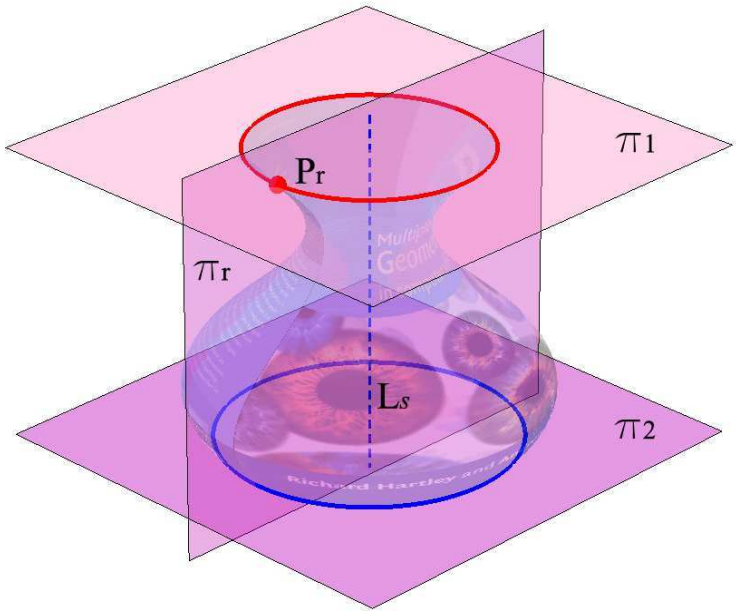


Fig. 5. Vision and control singularity planes for camera center position.

3 Results

Several tests were carried out with a simulation tool developed to study the characteristics and performance of the proposed visual servoing approach. The software allows the definition of the 3D target models, and the specification of visual tasks by their initial and goal views. Off-line path planning can be

defined with respect to an arbitrary axis in space. During on-line task execution, successive views reflecting the actual camera motion and noise conditions are generated, and the corresponding visual parameters are estimated. Target self-occlusions in the feature extraction process are also taken into account. Ellipses are estimated from their visible portions with the least square fitting algorithm proposed in [12]. In all the tests, views are generated using a solid of revolution target, a perspective camera with 640×480 pixels, and the following calibration matrix:

$$\mathtt{K} = \begin{bmatrix} 468.2 & 0.0 & 350.0 \\ & 427.2 & 258.0 \\ & & 1.0 \end{bmatrix} . \tag{14}$$

Tasks executed in the absence of image noise are performed with negligible error—both on camera calibration and positioning accuracy—with respect to ground truth values, thus proving the theoretical correctness of the formulation. In the following, results of several noise sensitivity tests are presented, demonstrating the robustness of the approach.

3.0.3 Off-line noise

A first test session concerned the analysis of the influence of noise on the camera parameters estimated during the self-calibration phase. For this purpose, a zero mean Gaussian noise was added to the visible ellipse points of both the initial and goal images. The noisy ellipse points were then used to estimate the calibration matrix as expounded in Section 2.2.2. Fig. 6 reports the results of a Monte Carlo analysis of self-calibration performance, with ten equally spaced values of the noise standard deviation between 0 to 1 pixel. For each noise level, 2000 trials were run, for a total of 20000 trials. Results for all of the five calibration parameters are reported. The figure shows that, while the standard deviation of the estimated parameters increases linearly with the noise value, the mean values exhibit a more than linear behavior. The figure shows that, for 1 pixel of noise standard deviation, the mean percentage error for both the focal length and principal point is about 7.5 %.

In a second test session devoted to check servoing robustness with respect to calibration errors (see also [10]), all of the above 20000 matrices were used as calibration input for the execution of a typical positioning task. *Notwithstanding the often large calibration errors, all the positioning trials were accomplished with negligible 3D errors.* Figs. 7 through 11 show the results for the positioning trial corresponding to one of the most badly estimated calibration

matrix:

$$\hat{\mathtt{K}} = \begin{bmatrix} 496.3 & 5.7 & 331.4 \\ & 407.5 & 211.6 \\ & & 1.0 \end{bmatrix} . \tag{15}$$

Fig. 7(a) illustrates the initial and goal views used for the test. The evolution of target appearance during task execution is summarized in Fig. 7(b).

Fig. 8(a) shows both the planned and the actual trajectories of the imaged vertexes of the reference quadrangle. Noisy trajectories reach the goal points with negligible error. Fig. 8(b), reporting the time evolution of the eight components of the image error $\mathbf{e}$ (see Section 2.3), provides an insight into task convergence.

Camera velocities during task execution are shown in Fig. 9. Velocity profiles are smooth and physically feasible (in particular, with zero velocity at the initial and final task times), thanks to the choice of an appropriate quintic polynomial for the planning function $\xi(t)$ of Eq. 10.

Fig. 10(a) shows the 3D task positioning errors as a function of time. The translation error is defined as the magnitude of the 3D vector joining the origins of the current and goal camera frames. The rotation error is defined as the rigid body rotation angle between the current and goal camera frames. Notice that *positioning 3D errors converge to zero, so that the servoing task is exactly fulfilled*. This result is obtained notwithstanding residual errors in the estimates of depth values for the four quadrangle vertexes—see Fig. 10(b). Being initialized by fixing to 1 the unknown scale factor of camera translation, depths are obtained with a gradient-based adaptive estimation scheme, and used for the computation of the image Jacobian in the control law of Eq. 7. Depth errors quickly decrease from their (quite large) initial value, thanks to the properties of the adaptation scheme. As a result of depth adaptation, the unknown scale factor is eventually estimated, thus obtaining a good approximation of the real target size (e.g., in millimeters). This would be impossible to obtain, whatever the number of target images, using standard computer vision multi-view approaches.

Fig. 11 shows the planned and actual 3D trajectories of the camera. Notice that the actual 3D trajectory followed by the camera deviates from the planned one: Indeed, due to calibration errors, the latter would bring the camera away from the correct goal position. Calibration errors also cause an inaccurate estimate of the scaled target axis. However, thanks to the closure of the control loop in the image, convergence to the correct goal position is obtained anyway.

3.0.4 Off-line and on-line noise

In a second test, task accuracy was evaluated in the presence of image noise both in the off-line (self-calibration) and on-line (task execution) phases. As before, a Monte Carlo test was performed for noise standard deviation values ranging from 0 to 1 pixel at steps of 0.1 pixel. For each noise level, 1000 trials were run, for a total of 10000 trials. Fig. 12 shows the mean and standard deviation of the 3D camera positioning error (magnitude of translation, and overall rotation angle). Results are shown in percentage with respect to the ground truth camera displacement between the final and goal views. Differently from the previous test session, positioning errors are now greater than zero. Specifically, the errors almost linearly increase with noise values, translational errors being slightly larger than rotational ones: This is due to the fact that, since rotations do not actually depend on depth estimates—see Eq. 6—, depth estimation errors only affect translational velocities.

Results for one of the Monte Carlo trials with a noise standard deviation of 0.2 pixel are provided hereafter. The estimated calibration matrix for this trial was

$$\hat{\mathtt{K}} = \begin{bmatrix} 517.3 & 3.1 & 322.7 \\ & 470.9 & 212.3 \\ & & 1.0 \end{bmatrix} , \tag{16}$$

which is even worse than that of Eq. 15.

Fig. 13(a) shows the task at hand, and Fig. 13(b) shows the image trajectories for the quadrangle vertexes. Despite the 2D jittering due to the image noise added in the on-line phase, the actual 2D trajectories follow quite well the planned 2D trajectories. Fig. 14 shows the discrepancy between the planned and the actual 3D camera trajectories. The same general comments to Fig. 11 apply here. In addition, a 3D jittering can be noticed, which is due to the use, in the simulation, of ideal actuators with zero response time—real actuators would actually smoothen the 3D trajectory. Differently from the off-line noise case, the 3D error with respect to the overall camera displacement is about 3.6% for translations, and 1.8% for rotations.

3.0.5 Choice of the helicoidal axis

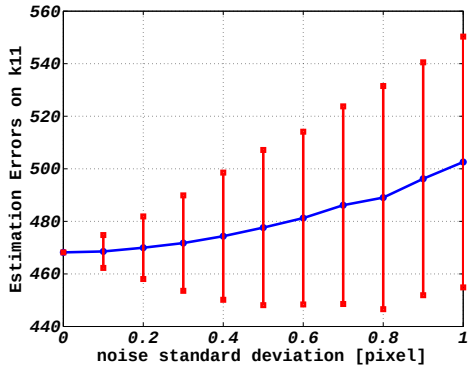
A noiseless test has also been performed in order to compare the two different choices of the helicoidal axis—namely, target axis L_s, and instantaneous rotation axis L_a—in off-line path planning mentioned earlier in the paper. Fig. 15 shows the different 2D imaged trajectories arising in the two cases, while Fig. 16 shows the corresponding 3D camera trajectories. It is worth noticing that, although the choice of the target axis as the helicoidal axis gives rise to a

2D trajectory which is apparently unnatural, the corresponding 3D helix has a larger radius than the other, thus requiring smaller accelerations for the actuators. Moreover, the 3D trajectory generated by the instantaneous rotational axis choice gets closer to the singularity plane.

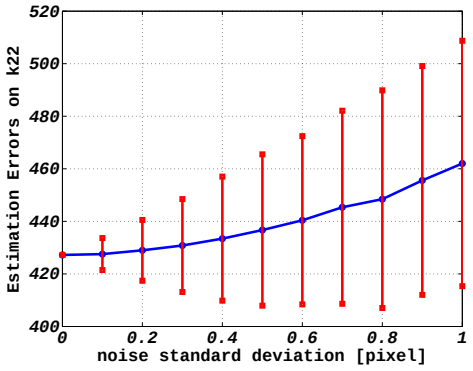
4 Conclusions and Future Work

A self-calibrated approach to visual servoing with respect to non planar targets modeled through a pair of coaxial circles plus one point on either of them has been discussed. Simulations results have demonstrated the theoretical correctness of the approach. Moreover, results have witnessed the robustness of the approach with respect to noisy calibration and image data, thus paving the way to practical applications.

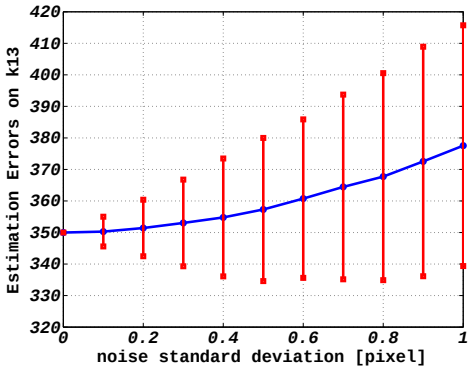
Future work will address three main topics: (*i*) an in-depth study of the singular and degenerate configurations occurring either in vision or control; (*ii*) the extension of the self-calibration theory for zooming cameras, thus making the approach feasible also for larger camera displacements; (*iii*) experiments with practical robotic setups, such as eye-in-hand systems featuring a camera mounted on a 6-DOF manipulator.



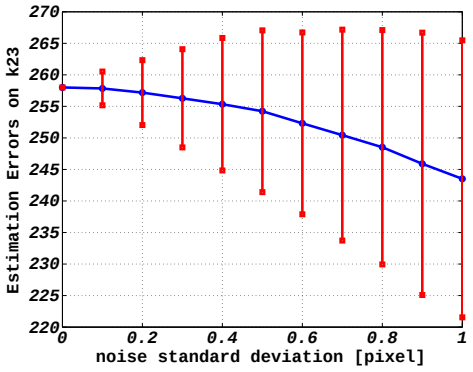
(a)



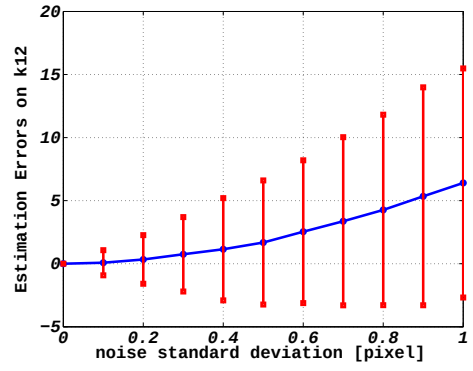
(b)



(c)

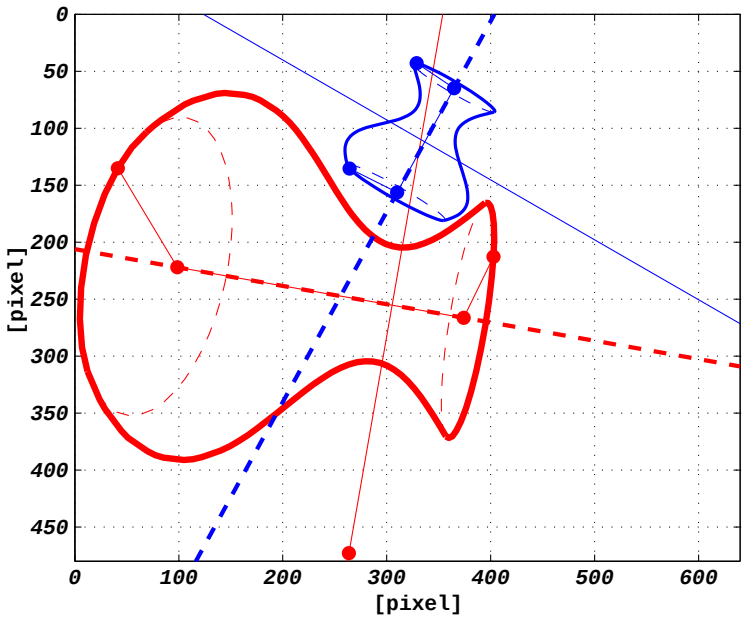


(d)

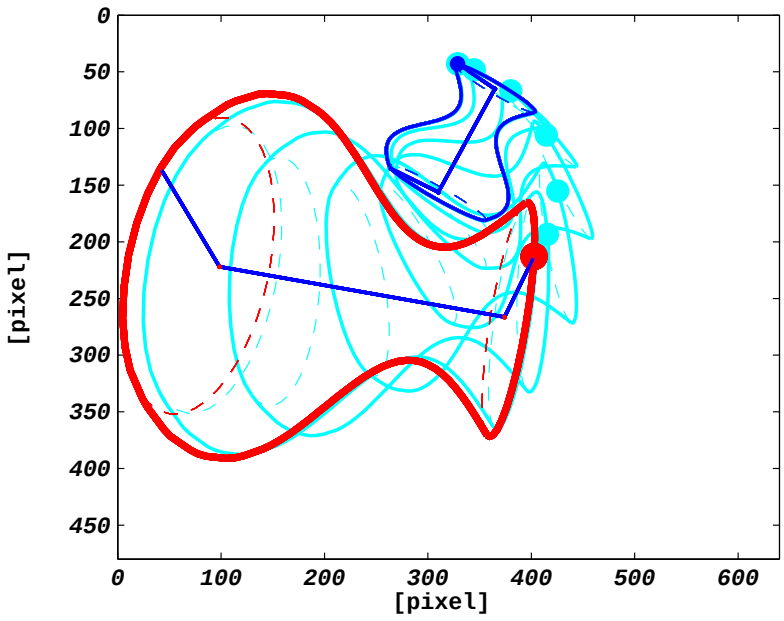


(e)

Fig. 6. The influence of image noise on the estimates of internal calibration parameters (2000 Monte Carlo trials for each noise step). The ground truth is the calibration matrix of Eq. 14. For each estimated parameter, the mean is represented by a polygonal line, and the standard deviation σ by vertical bars of width 2σ centered around the mean value. (a),(b): Focal lengths $\mathtt{k}_{11}$ and $\mathtt{k}_{22}$. (c),(d): Principal point components $\mathtt{k}_{13}$ and $\mathtt{k}_{23}$. (e): Skew factor $\mathtt{k}_{12}$.

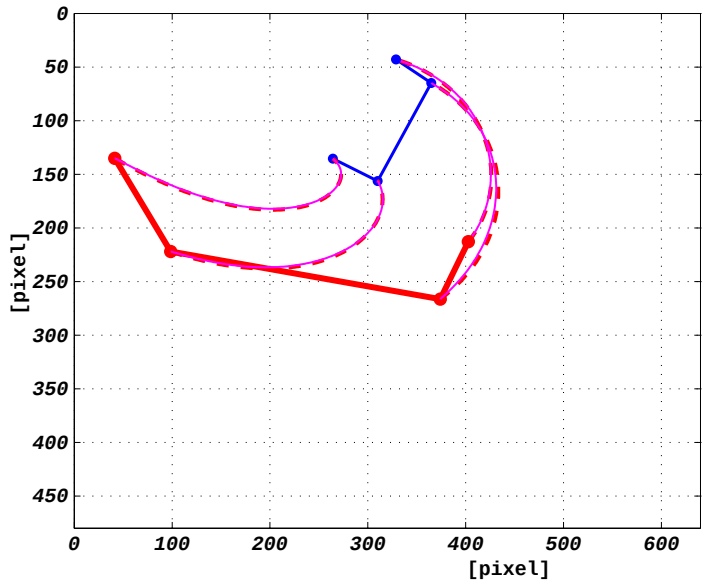


(a)

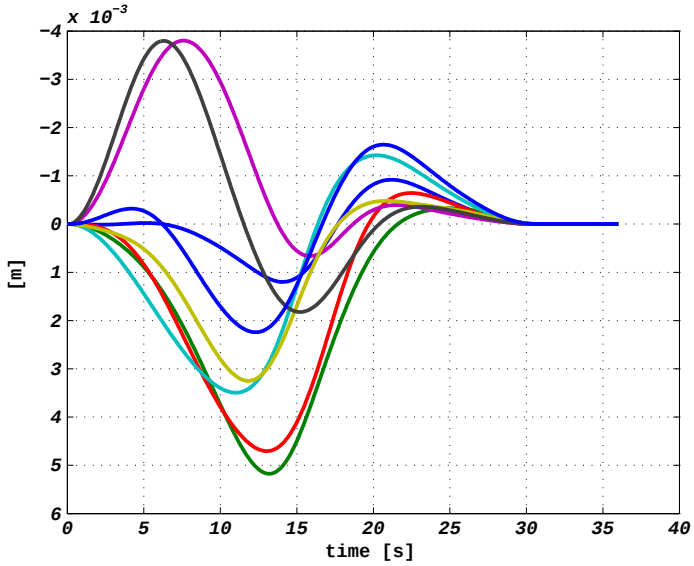


(b)

Fig. 7. (a): Initial (thicker line) and final views for the servoing task considered in the off-line noise case. (b): Evolution of target appearance.

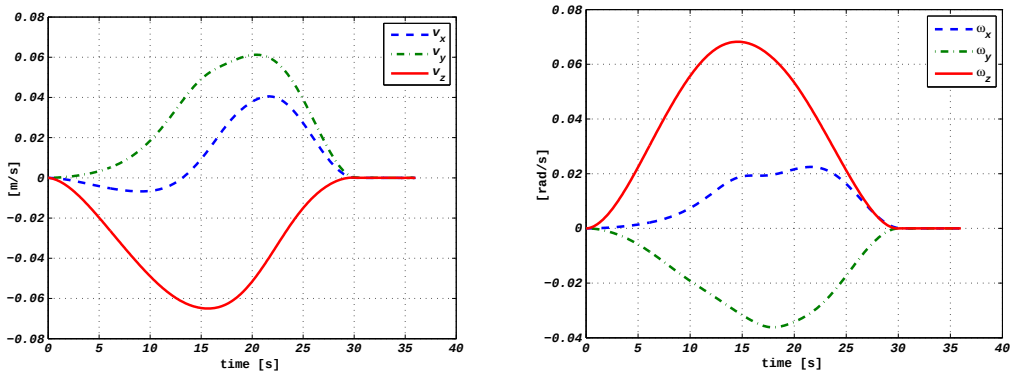


(a)



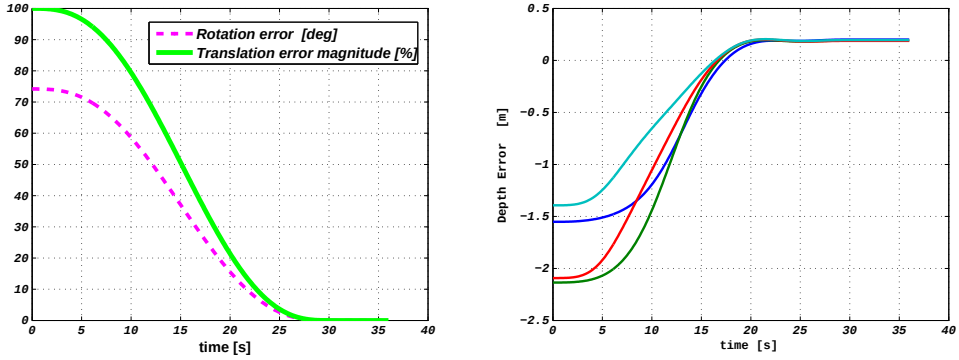
(b)

Fig. 8. (a): Planned *vs* actual trajectories of the vertexes of the imaged quadrangle in the off-line noise case. Thicker lines refer to the initial view. (b): The eight components of the image error $\mathbf{e}$.



(a) (b)

Fig. 9. Camera velocities. (a): Translational velocities. (b): Angular velocities.



(a) (b)

Fig. 10. (a): Task positioning errors for camera translation and rotation. Both of them converge to zero, even in the presence of depth estimation errors for the quadrangle vertexes. (b): Adaptive depth estimation errors for the four quadrangle vertexes exploited in the computation of the interaction matrix J.

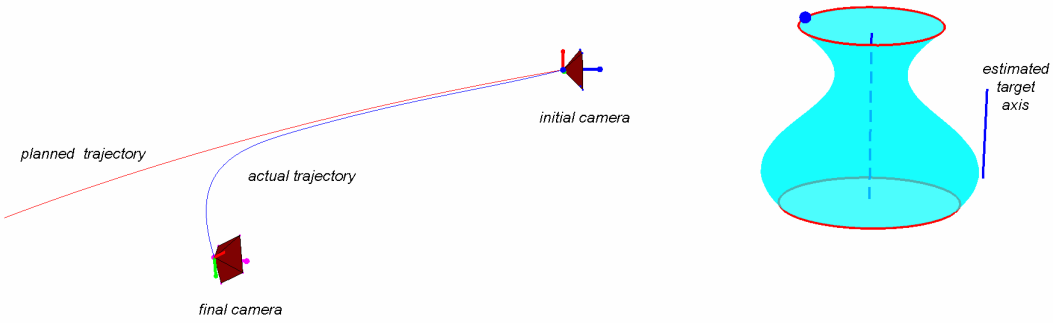
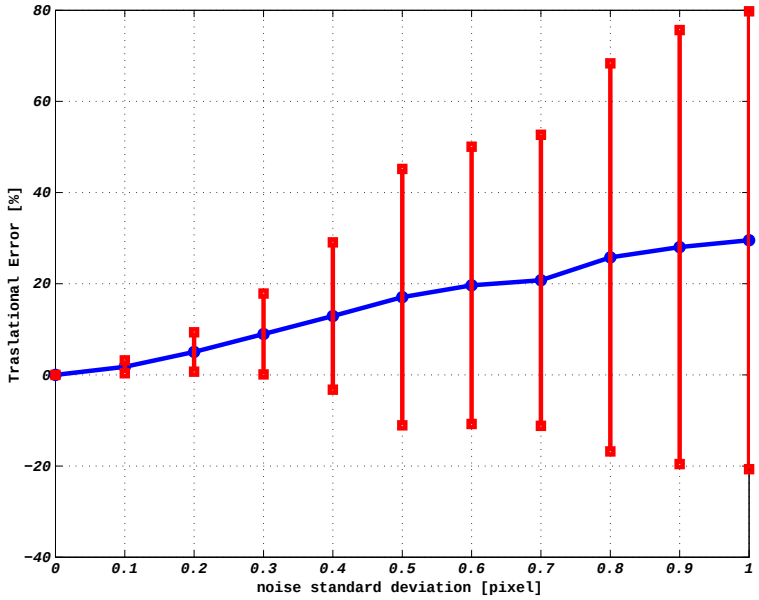
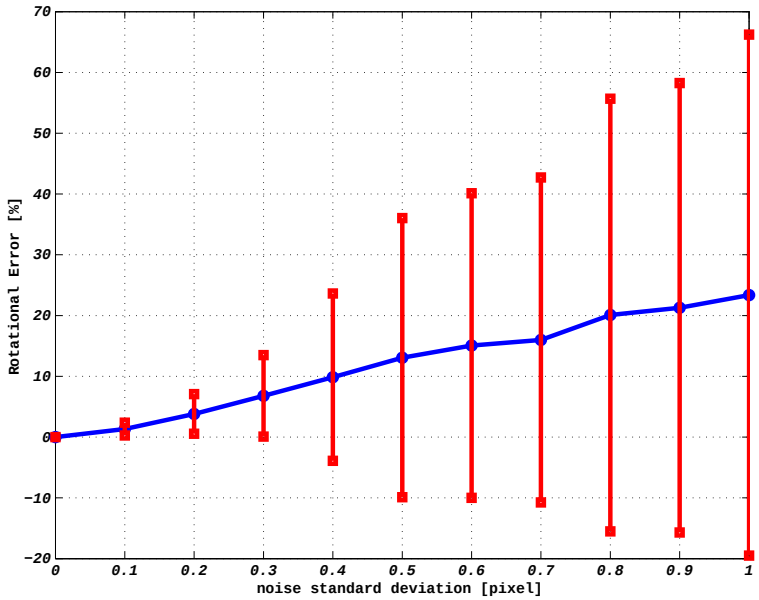


Fig. 11. Planned *vs* actual 3D camera trajectories in the off-line noise case.

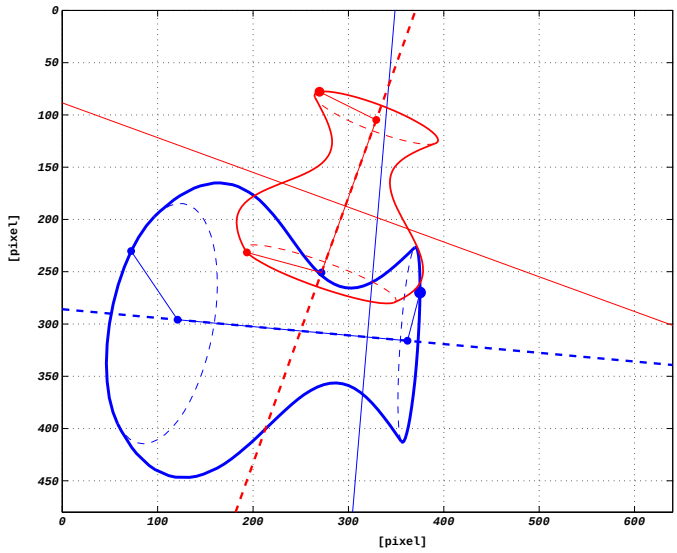


(a)

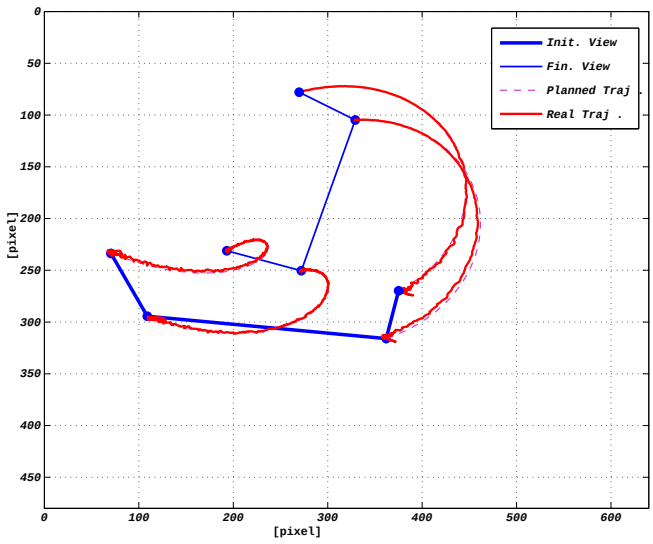


(b)

Fig. 12. Positioning error *vs* noise standard deviation. Errors are reported in percentage with respect to ground truth values. (a): Translational error. (b): Rotational error.



(a)



(b)

Fig. 13. (a): Initial and final views for the servoing task in the off-line/on-line noise case. (b): Planned *vs* actual trajectories of the vertexes of the imaged quadrangle. Thicker lines refer to the initial view.

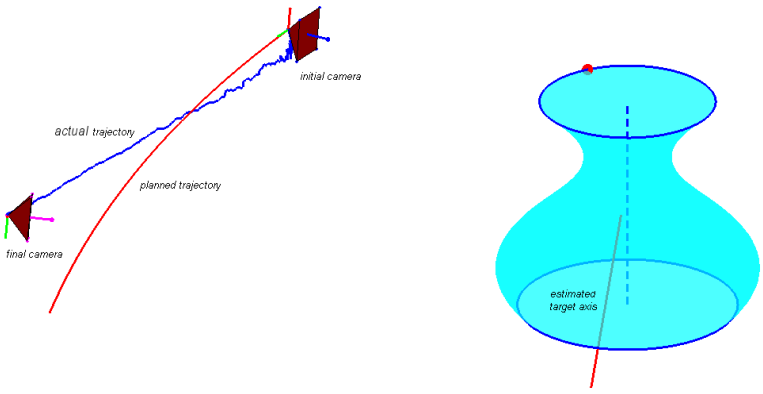
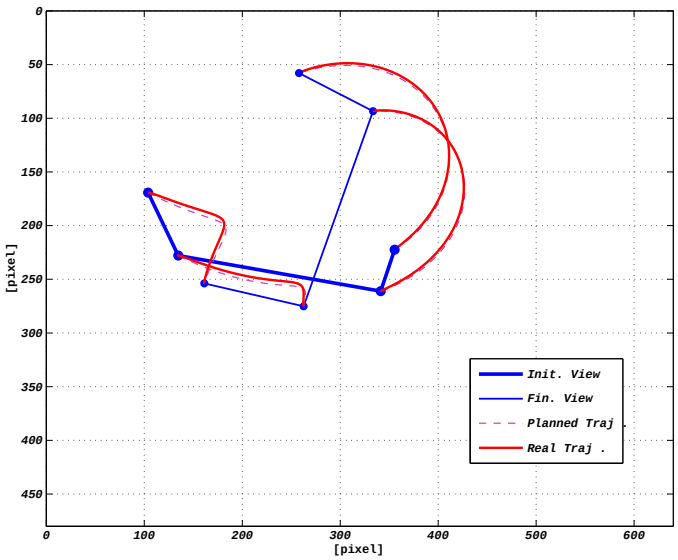
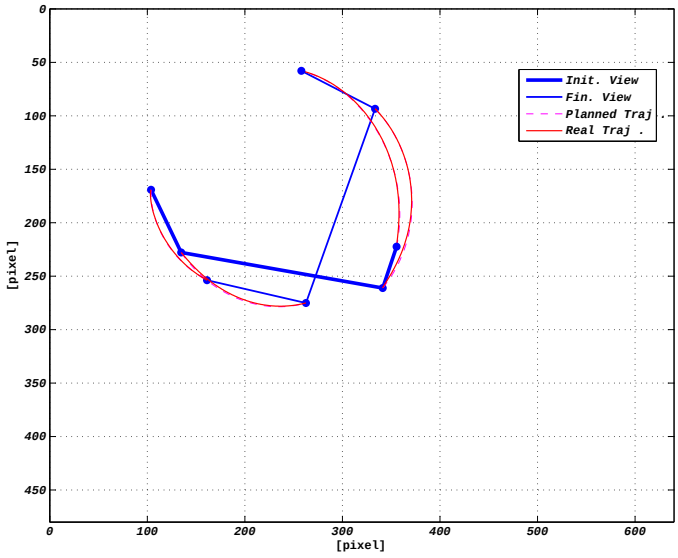


Fig. 14. Planned *vs* actual 3D camera trajectories in the off-line/on-line noise case.



(a)



(b)

Fig. 15. Actual *vs* planned imaged trajectories (thicker lines refer to the initial view) using different helicoidal axes. (a): Target axis. (b): Instantaneous rotational axis.

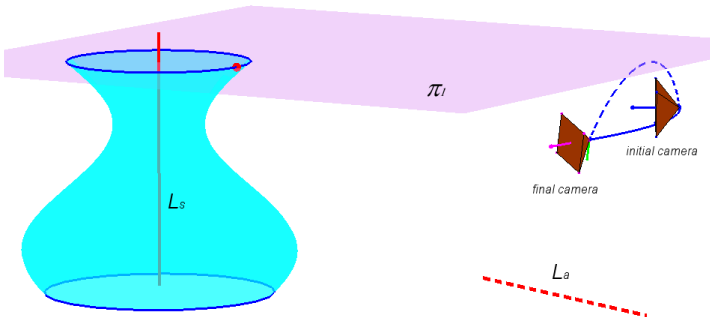


Fig. 16. Helicoidal trajectories for two different axis choices. The solid line refers to the target axis choice.

References

[1] B. Allotta, C. Colombo, On the use of linear camera-object interaction models in visual servoing, IEEE Trans. Robot. Automat. 15 (2) (1999) 350–357.

[2] B. Allotta, D. Fioravanti, 3D motion planning for image-based visual servoing tasks., in: Proc. IEEE International Conference on Robotics and Automation (ICRA'05), Barcelona, Spain, 2005.

[3] F. Chaumette, Image moments: A general and useful set of features for visual servoing, IEEE Trans. Rel. 20 (4) (1999) 713–723.

[4] F. Chaumette, Potential problems of stability and convergence in image-based and position-based visual servoing, The confluence of vision and control (D. Kriegman, G. Hager and A. Morse Eds.) 237 (1999) 66–78, lecture Notes in Control and Information Systems. Springer.

[5] G. Chesi, K. Hashimoto, A simple technique for improving camera displacement estimation in eye-in-hand visual servoing, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (9) (2004) 1239–1242.

[6] C. Colombo, B. Allotta, Image based robot task planning and control using a compact visual representation, IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans 29 (1) (2004) 92–100.

[7] C. Colombo, A. Del Bimbo, F. Pernici, Metric 3D reconstruction and texture acquisition of surfaces of revolution from a single uncalibrated view, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (1) (2005) 99–114.

[8] A. Comport, M. Pressigout, E. Marchand, F. C. and, A visual servoing control law that is robust to image outliers, in: Proc. ot The IEEE/RJS International Conference on Intelligent Robots ansd Systems, Las Vegas, USA, 2003.

[9] F. Conticelli, B. Allotta, Discrete-time robot visual feedback in 3D positioning tasks with depth adaptation., IEEE/ASME Transaction on Mechatronics 6 (3) (2001) 356–363.

[10] B. Espiau, Effect of calibration errors on visual servoing in robotics, in: Proc. 3rd International Workshop on Experimental Robotics, Kyoto, Japan, 1993.

[11] B. Espiau, F. Chaumette, P. Rives, A new approach to visual servoing in robotics, IEEE Trans. Robot. Automat. 8 (3) (1992) 313–326.

[12] A. W. Fitzgibbon, M. Pilu, R. B. Fisher, Direct least square fitting of ellipses, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (5) (1999) 476–480.

[13] N. Garcia-Aracil, E. Malis, R. Aracil-Santonja, C. Perez-Vidal, Continuous visual servoing despite the changes of visibility in image features, IEEE Transactions on Robotics 21 (6) (2005) 147–161.

[14] R. I. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, 2003.

[15] S. Hutchinson, G. D. Hager, P. I. Corke, Tutorial on visual servo control, IEEE Trans. Robot. Automat. 12 (5) (1996) 651–670.

[16] Y. Ma, S. Soatto, J. Košecká, S. Sastry, An Invitation to 3D Vision, 2003.

[17] E. Malis, Visual servoing invariant to changes in camera-intrinsic parameters, IEEE Trans. Robot. Automat. 20 (1) (2004) 72–81.

[18] E. Malis, F. Chaumette, S. Boudet, 2-1/2-d visual servoing, IEEE Trans. Robot. Automat. 15 (2) (1999) 238–250.

[19] E. Malis, G. Chesi, R. Cipolla, 2-1/2-d visual servoing with respect to planar contours having complex and unknown shapes, International Journal of Robotic Research 22 (10-11) (2003) 841–853.

[20] E. Malis, P.Rives, Robustness of imaged based visual servoing with respect to depth distribution errors, in: Proc. of the IEEE International Conference on Robotics and Automation, Taipei, Taiwan, 2003.

[21] Y. Mezouar, F. Chaumette, Path planning for robust image-based control, IEEE Trans. Robot. Automat. 18 (4) (2002) 534–549.

[22] K.-Y. K. Wong, P. Mendonça, R. Cipolla, Camera calibration from surfaces of revolution, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (2) (2003) 147–161.

Duccio Fioravanti received in 2004 the laurea degree in Mechanical Engineering from the University of Florence. Since 2004 he is working towards his PhD on Mechanics of Machines and Mechanisms in the Department of Energetics at the University of Florence. His current research interests are visual servoing, robot control and multi-body dynamics in railway systems. He is author of conference papers concerning both vision based control and railway dynamics. He has been reviewer for journals and international conferences focusing on robotics and visual servoing applications.

Carlo Colombo graduated with Honors In Electronic Engineering at the University of Florence, Italy, in 1992, and earned a PhD in Robotics from the Sant'Anna School of University Studies and Doctoral Research, Pisa, Italy, in 1996. He is currently Associate Professor at the Department of Systems and Informatics of the University of Florence. His main research interests include computer vision and its applications to human-computer interaction, advanced robotics, and multimedia systems. On these topics he has authored or co-authored about 100 papers. He is presently Associate Editor of the international journals *Robotics and Autonomous Systems* and *Journal of Multimedia* and serves on the programme committee of several international conferences in the area of visual information processing.

Benedetto Allotta received the laurea degree in Mechanical Engineering from the School of Engineering of the University of Pisa in 1987. In 1992 he earned a PhD degree in Robotics from the Scuola Superiore Sant'Anna, Pisa where he has been an assistant professor from 1993 to 2001. From 2001 to 2005 he has served as an associate professor in the Department of Energetics of the University of Florence, Italy, where in 2005 he has been appointed full professor. His current research interests are automation in transport systems, hardware-in-the-loop (HIL) simulation, control of robots, mechatronics, active control of vibrations in time-varying systems. He is author of about 80 publications, including more than 20 papers on international journals, and 2 granted international patents. He is responsible of several research grants and contracts coming from both public agencies and private companies.

*** Photo of Dr. Fioravanti**

*** Photo of Prof. Colombo**

*** Photo of Prof. Allotta**