Visual servoing based on an analytical homography decomposition

M. Vargas, Ezio Malis

Visual servoing based on an analytical homography decomposition

Manuel Vargas and Ezio Malis

***Abstract*— This paper presents a new vision-based control method for positioning a camera with respect to an unknown planar object. Standard methods use non-linear state observers based on homography decomposition. In the general case, there are two possible solutions to the decomposition problem. Thus, some additional "a priori" information must be used. In this paper, we propose to use an analytical decomposition of the homography matrix in order to define a new control objective that allows to discard the false solution without any "a priori" information. The stability of the control law has been proved.**

I. INTRODUCTION

Visual servoing can be stated as a non-linear output regulation problem [2]. The output is the image acquired by a camera mounted on a dynamic system. The state of the camera is thus accessible via a non-linear map. For this reason, positioning tasks have been defined using the so-called teach-by-showing technique [2]. The camera is moved to a reference position and the corresponding reference image is stored. Then, starting from a different camera position the control objective is to move the camera such that the current image will coincide to the reference one. In this paper, we suppose that the observed object is a plane in the Cartesian space. One solution to the control problem is to build a non-linear observer of the state. This can be done using several output measurements. The problem is that, when considering real-time applications, we should process as few observations as possible. In [3], [4], [6], [5] the authors have built a non-linear state observer using additional information (the normal to the plane, vanishing points, ...). In this case only the current and the reference observations are needed. In this paper, we intend to perform vision-based control without knowing any a priori information. To do this we need more observations. This can be done by moving the camera. If we move the camera and the state is not observable we may have some problems. For this reason, we propose in this paper a different approach. We define a new control objective in order to move the camera by keeping a bounded error and in order to obtain the necessary information for the state observer. The paper is organized as follows: Section II gives the theoretical background. Section III describes the proposed homography-based control strategy and its stability conditions are established in Section IV. Section V describes how to achieve the camera positioning task by switching control law. The control laws have been validated with simulation results in Section VI.

The authors gratefully acknowledge INRIA and MCyT (grant DPI2004-06419) for supporting this work.

M. Vargas is with the Dept. Ingeniería de Sistemas y Automática, University of Seville, Spain, vargas@cartuja.us.es

E. Malis is with the ICARE research team at INRIA, Sophia-Antipolis, France, Ezio.Malis@sophia.inria.fr

II. THEORETICAL BACKGROUND

A. Notation and description of the vision system

We assume that the absolute frame coincides with the reference camera frame $\mathcal{F}^*$. We suppose that the observed object is planar and composed by a set of n 3D points, $\mathcal{P}_i = (X_i, Y_i, Z_i)$ (see Figure 1). The normal and the distance to the plane in the reference frame will be denoted with $\mathbf{n}^*$ and d^*. A calibrated camera measures the reference image homogeneous coordinates $\mathbf{m}_i^* = (x_i^*, y_i^*, 1)$ of the perspective projection of the 3D points: $\mathbf{m}_i^* = (X_i/Z_i, Y_i/Z_i, 1)$. The displacement between the reference and current camera frame $\mathcal{F}$ is represented by the rotation matrix $\mathbf{R}$ and translation vector $\mathbf{t}$. The current image homogeneous coordinates $\mathbf{m}_i = (x_i, y_i, 1)$ are given again by the perspective projection of the 3D points in the current camera frame: $\mathbf{m}_i \propto \mathbf{R}\mathcal{P}_i + \mathbf{t}$.

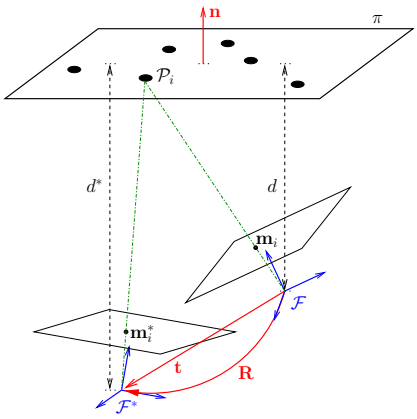

Fig. 1. Desired and current camera frames and involved notation.

B. Vision-based control

We consider the control of the following nonlinear system:

$$\dot{\mathbf{x}} = \mathbf{g}(\mathbf{x}, \mathbf{v}) \quad (1)$$

$$\mathbf{y} = \mathbf{h}(\mathbf{x}, \boldsymbol{\mu}) \quad (2)$$

where $\mathbf{x} \in \mathbb{SE}(3)$ is the state (i.e. the camera pose), $\mathbf{y}$ is the output of the camera and $\mathbf{v}$ is the control input (i.e. the camera velocity). The output of the camera depends on the state and on some parameters $\boldsymbol{\mu}$ (e.g. the normal of the plane,...). Let $\mathbf{x}^*$ be the reference state of the camera. Without loss of generality one can choose $\mathbf{x}^* = e$, where e is the identity element of $\mathbb{SE}(3)$. Then, the reference output is

$\mathbf{y}^* = \mathbf{h}(\mathbf{x}^*, \boldsymbol{\mu})$. If we suppose that the camera displacement is not too big, we can choose a state vector $\mathbf{x} = (\mathbf{t}, \mathbf{r})$, where $\mathbf{r} = \tan(\frac{\theta}{2})\,\mathbf{u}$ is the chosen rotation parameterization ($\mathbf{u}$ and θ are the axis and angle of rotation, respectively). The rotation matrix can be written as a function of the vector $\mathbf{r}$ as follows:

$$\mathbf{R} = \mathbf{I} + \frac{2}{1 + \|\mathbf{r}\|^2} \left([\mathbf{r}]_\times + [\mathbf{r}]_\times^2 \right) \tag{3}$$

where the notation $[\mathbf{r}]_\times$ means for the skew-symmetric matrix of a vector $\mathbf{r}$. The derivative of the state vector $\mathbf{x}$ is:

$$\dot{\mathbf{x}} = \begin{bmatrix} \dot{\mathbf{t}} \\ \dot{\mathbf{r}} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & -[\mathbf{t}]_\times \\ \mathbf{0} & \mathbf{J}_\omega \end{bmatrix} \begin{bmatrix} \boldsymbol{\upsilon} \\ \boldsymbol{\omega} \end{bmatrix} \tag{4}$$

where:

$$\mathbf{J}_\omega = \frac{1}{2} \left(\mathbf{I} - [\mathbf{r}]_\times + \mathbf{r}\,\mathbf{r}^\top \right) \tag{5}$$

and $\mathbf{v} = (\boldsymbol{\upsilon}, \boldsymbol{\omega})$ is our input vector containing the velocity of translation $\boldsymbol{\upsilon}$ and the velocity of rotation $\boldsymbol{\omega}$ (see [9]).

C. Homography-based state observer with full information

In [6] a new class of visual servoing methods has been proposed. It is based on the estimation of a (3×3) *homography matrix* $\mathbf{H}$. This matrix transforms the reference image coordinates of a point into the corresponding coordinates in the current image up to a scale factor:

$$\mathbf{m} = \alpha_h\, \mathbf{H}\, \mathbf{m}^*$$

This homography matrix can be written as a function of the state of the camera:

$$\mathbf{H} = \gamma \left(\mathbf{R} + \frac{\mathbf{t}}{d^*}\, \mathbf{n}^{*\top} \right) \tag{6}$$

where the normalization scale factor γ is chosen in such a way that the determinant of matrix $\mathbf{H}$ is unitary (i.e. $\det(\mathbf{H}) = 1$): $\gamma = \sqrt[-3]{1 + \mathbf{n}^{*\top}\mathbf{R}^\top\mathbf{t}/d^*}$. An efficient real-time algorithm for estimating the homography from raw images has been proposed in [7]. Once the homography has been estimated, we can extract $\mathbf{R}$, $\mathbf{t}/d^*$ and $\mathbf{n}^*$ [1]:

$$\mathbf{H} \implies \{\mathbf{R},\, \mathbf{t}/d^*,\, \mathbf{n}^*\}$$

Note that the translation is estimated up to a positive scalar factor. Without loosing generality, we can suppose $d^* = 1$ and include this factor in the translation vector. On the other hand, we will use $\mathbf{n} = \mathbf{n}^*$ for a simpler notation. In [1], the homography is decomposed using a numerical method (involving Singular Value Decomposition). In the general case, there exist 4 solutions, two of them being the "opposites" of the others:

$$Rtn_a = \{\mathbf{R}_a, \mathbf{t}_a, \mathbf{n}_a\}; \quad Rtn_{a-} = \{\mathbf{R}_a, -\mathbf{t}_a, -\mathbf{n}_a\} \tag{7}$$

$$Rtn_b = \{\mathbf{R}_b, \mathbf{t}_b, \mathbf{n}_b\}; \quad Rtn_{b-} = \{\mathbf{R}_b, -\mathbf{t}_b, -\mathbf{n}_b\} \tag{8}$$

These can be reduced to only two solutions applying the constraint that all the reference points must be visible from the camera (*visibility constraint*). Without loss of generality, we will assume along the development that the two solutions verifying this constraint are Rtn_a and Rtn_b and that, among them, Rtn_a is the "true" solution. In practice, in order to determine which one is the good solution, we generally use an approximation of the normal $\mathbf{n}^*$. Thus, having an approximated parameter vector $\widehat{\boldsymbol{\mu}}$ we build a non-linear state observer:

$$\widehat{\mathbf{x}} = \boldsymbol{\varphi}(\mathbf{y}(\mathbf{x}), \mathbf{y}^*, \widehat{\boldsymbol{\mu}})$$

III. A MODIFIED CONTROL OBJECTIVE

A control law based on the Cartesian error in position and orientation is being developed. The error between the current and desired camera poses is obtained from the homography decomposition. Without any a priory knowledge about the true normal to the plane, we are not able to discard one of the two possible solutions as a false one. On the other hand, contrarily to [1], we have found an analytic solution for the homography decomposition problem. The analytic solution has the great advantage of providing a deeper understanding of the decomposition problem. For instance, it allows to obtain the relations between the possible solutions of the problem [9]:

$$\mathbf{t}_b = \frac{\|\mathbf{t}_a\|}{\rho}\, \mathbf{R}_a \left(2\,\mathbf{n}_a + \mathbf{R}_a^\top \mathbf{t}_a \right) \tag{9}$$

$$\mathbf{n}_b = \frac{1}{\rho} \left(\frac{\|\mathbf{t}_a\|}{2}\, 2\,\mathbf{n}_a + \frac{2}{\|\mathbf{t}_a\|} \mathbf{R}_a^\top \mathbf{t}_a \right) \tag{10}$$

$$\mathbf{r}_b = \frac{\left[(2 - \mathbf{n}_a^\top \mathbf{t}_a)\, \mathbf{I} + \mathbf{t}_a\, \mathbf{n}_a^\top + \mathbf{n}_a\, \mathbf{t}_a^\top \right]\, \mathbf{r}_a + (\mathbf{n}_a \times \mathbf{t}_a)}{2 + \mathbf{n}_a^\top \mathbf{t}_a + \mathbf{r}_a^\top (\mathbf{n}_a \times \mathbf{t}_a)} \tag{11}$$

In these relations the sub-indexes a and b can be exchanged. The coefficients ρ and ν are:

$$\rho = \|2\,\mathbf{n}_e + \mathbf{R}_e^\top \mathbf{t}_e\| = \sqrt{\|\mathbf{t}_e\|^2 + 2\,\nu} > 1$$

$$\nu = 2\,(\mathbf{n}_e^\top \mathbf{R}_e^\top \mathbf{t}_e + 1) > 0\,; \qquad e = \{a, b\}$$

It will be shown that a control law based on the *average of these two solutions* can be used, such that the system will converge in such a way that it is always possible to discard the false solution. Once the true solution has been identified, the camera can be controlled using only this solution.

A. Camera control

The task function [10] to be minimized will be defined as a translation and an orientation error:

$$\mathbf{e} = \begin{bmatrix} \mathbf{e}_t \\ \mathbf{e}_r \end{bmatrix} = \begin{bmatrix} \mathbf{t}_m \\ \mathbf{r}_m \end{bmatrix}$$

Being $\mathbf{t}_m$ and $\mathbf{r}_m$ the translation and orientation means, respectively, computed as follows:

$$\mathbf{t}_m = \frac{\mathbf{t}_a + \mathbf{t}_b}{2} \tag{12}$$

$$\mathbf{r}_m \iff \mathbf{R}_m = \mathbf{R}_a\, (\mathbf{R}_a^\top\, \mathbf{R}_b)^{1/2} \tag{13}$$

The rotation matrix, $\mathbf{R}_m$, average of $\mathbf{R}_a$ and $\mathbf{R}_b$, computed in such way is defined as the *Riemmanian mean of two rotations* (for more details, see [8]).

According to the relations between the true and false solutions, $\mathbf{r}_m$ can be obtained as:

$$\mathbf{r}_m = \frac{\mathbf{r}_a + \mathbf{r}_h + \mathbf{r}_a \times \mathbf{r}_h}{1 - \mathbf{r}_a^\top \mathbf{r}_h} \tag{14}$$

where $\mathbf{r}_h$ describes a required intermediary rotation, related to the true solution by means of:

$$\mathbf{r}_h = \frac{\rho - \left(2 + \mathbf{n}_a^\top \mathbf{t}_a^*\right)}{\|\mathbf{n}_a \times \mathbf{t}_a^*\|^2} \left(\mathbf{n}_a \times \mathbf{t}_a^*\right) \tag{15}$$

where $\mathbf{t}_a^* = \mathbf{R}_a^\top \mathbf{t}_a$ (derivation is detailed in [9]). We compute the input control action from the defined task error:

$$\mathbf{v} = \begin{bmatrix} \boldsymbol{\upsilon} \\ \boldsymbol{\omega} \end{bmatrix} = -\lambda\, \mathbf{e} \tag{16}$$

where λ is a positive scalar, tuning the closed-loop convergence rate. The derivative of the task error is related to the velocity screw, according to some interaction matrix $\mathbf{L}$ to be determined:

$$\dot{\mathbf{e}} = \mathbf{L}\, \mathbf{v}$$

Giving the following closed-loop system:

$$\dot{\mathbf{e}} = -\lambda\, \mathbf{L}\, \mathbf{e} \tag{17}$$

We are now interested in the computation and properties of the interaction matrix $\mathbf{L}$. We will identify the components of this matrix as:

$$\begin{bmatrix} \dot{\mathbf{e}}_t \\ \dot{\mathbf{e}}_r \end{bmatrix} = \begin{bmatrix} \mathbf{L}_{11} & \mathbf{L}_{12} \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\upsilon} \\ \boldsymbol{\omega} \end{bmatrix}$$

$$\begin{array}{ll} \mathbf{L}_{11} = \frac{\partial \mathbf{e}_t}{\partial \mathbf{t}_a} & \mathbf{L}_{12} = -\frac{\partial \mathbf{e}_t}{\partial \mathbf{t}_a}[\mathbf{t}_a]_\times + \frac{\partial \mathbf{e}_t}{\partial \mathbf{r}_a}\mathbf{J}_\omega \\ \mathbf{L}_{21} = \frac{\partial \mathbf{e}_r}{\partial \mathbf{t}_a} & \mathbf{L}_{22} = -\frac{\partial \mathbf{e}_r}{\partial \mathbf{t}_a}[\mathbf{t}_a]_\times + \frac{\partial \mathbf{e}_r}{\partial \mathbf{r}_a}\mathbf{J}_\omega \end{array} \tag{18}$$

Next, we present a very short description providing the expressions of these sub-matrices.

B. Computation of $\mathbf{L}_{11}$

The translation error has been defined in equation (12) as the average of the two solutions for the translation vector. $\mathbf{L}_{11}$ is then:

$$\mathbf{L}_{11} = \frac{\partial \mathbf{e}_t}{\partial \mathbf{t}_a} = \frac{1}{2}\left(\frac{\partial \mathbf{t}_b}{\partial \mathbf{t}_a} + \mathbf{I}\right)$$

¿From (9), the expression of this matrix can be found:

$$\mathbf{L}_{11} = \frac{1}{2}\left[2\mu_1 \mathbf{n}_a' \mathbf{t}_a^\top + \mu_1 \mathbf{t}_a \mathbf{t}_a^\top - 4\mu_2 \mathbf{n}_a' \mathbf{n}_a'^\top - 2\mu_2 \mathbf{t}_a \mathbf{n}_a'^\top + \mu_3 \mathbf{I}\right] \tag{19}$$

where $\mathbf{n}_a' = \mathbf{R}_a \mathbf{n}_a$ and the scalars μ_i are:

$$\mu_1 = \frac{1}{\rho\|\mathbf{t}_a\|} - \mu_2\,; \qquad \mu_2 = \frac{\|\mathbf{t}_a\|}{\rho^3}\,; \qquad \mu_3 = \frac{\|\mathbf{t}_a\|}{\rho} + 1 \tag{20}$$

C. Computation of $\mathbf{L}_{12}$

For the computation of $\mathbf{L}_{12}$, the Jacobian $\frac{\partial \mathbf{e}_t}{\partial \mathbf{r}_a}$ is needed (see (18)):

$$\frac{\partial \mathbf{e}_t}{\partial \mathbf{r}_a} = \frac{1}{2}\frac{\partial \mathbf{t}_b}{\partial \mathbf{r}_a}$$

Starting from this and after many cumbersome manipulations, we conclude that this matrix can be written as [9]:

$$\mathbf{L}_{12} = -[\mathbf{e}_t]_\times \tag{21}$$

D. Computation of $\mathbf{L}_{21}$

Regarding to the interaction sub-matrix:

$$\mathbf{L}_{21} = \frac{\partial \mathbf{e}_r}{\partial \mathbf{t}_a}$$

a rather involved expression has been obtained for it, and no equivalent closed form has been found yet. Nevertheless, we will see afterwards that we do not need to care about this matrix, whatever its form and complexity.

E. Computation of $\mathbf{L}_{22}$

We consider here as a starting point the time derivative of $\mathbf{e}_r$, which can be written as:

$$\dot{\mathbf{e}}_r = \dot{\mathbf{r}}_m = \frac{\partial \mathbf{r}_m}{\partial \mathbf{r}_h}\, \dot{\mathbf{r}}_h + \frac{\partial \mathbf{r}_m}{\partial \mathbf{r}_a}\, \dot{\mathbf{r}}_a \tag{22}$$

Developing this expression, and matching it with the following one:

$$\dot{\mathbf{r}}_m = \mathbf{L}_{21}\, \boldsymbol{\upsilon} + \mathbf{L}_{22}\, \boldsymbol{\omega}$$

$\mathbf{L}_{22}$ can be written as:

$$\mathbf{L}_{22} = \frac{\partial \mathbf{r}_m}{\partial \mathbf{r}_a}\, \mathbf{J}_\omega$$

Using relations (14) and (5) and after some reductions (see again [9]), we get:

$$\mathbf{L}_{22} = \frac{1 + \|\mathbf{r}_m\|^2}{4}\, \left(\mathbf{I} + \mathbf{R}_m^\top\right) \tag{23}$$

IV. STABILITY ANALYSIS

In this section, the stability of the control law presented in the previous section is studied. First, the stability of the translation error, $\mathbf{e}_t$, is considered.

A. Stability in the translation error $\mathbf{e}_t$

In order to prove the convergence of $\mathbf{e}_t$ to zero, the following Lyapunov function candidate is proposed:

$$V_t = \frac{1}{2}\, \mathbf{e}_t^\top \mathbf{e}_t$$

Its time derivative is:

$$\dot{V}_t = \frac{d\|\mathbf{e}_t\|}{dt} = \mathbf{e}_t^\top \dot{\mathbf{e}}_t$$

The expression of $\dot{\mathbf{e}}_t$ in terms of the components of the interaction matrix is:

$$\dot{\mathbf{e}}_t = \mathbf{L}_{11}\, \boldsymbol{\upsilon} + \mathbf{L}_{12}\, \boldsymbol{\omega}$$

Using the form (21) for $\mathbf{L}_{12}$ and replacing the control inputs $\boldsymbol{\upsilon}$ and $\boldsymbol{\omega}$ using (16):

$$\dot{\mathbf{e}}_t = -\lambda\, \mathbf{L}_{11}\, \mathbf{e}_t - \lambda\, \mathbf{L}_{12}\, \mathbf{e}_r = -\lambda\, \mathbf{L}_{11}\, \mathbf{e}_t + \lambda\, [\mathbf{e}_t]_\times\, \mathbf{e}_r$$

giving

$$\dot{V}_t = -\lambda\, \mathbf{e}_t^\top \mathbf{L}_{11}\, \mathbf{e}_t$$

As $\mathbf{L}_{11}$ is not, in general, a symmetric matrix, it is convenient to write the previous expression as:

$$\dot{V}_t = -\lambda\, \mathbf{e}_t^\top \mathbf{S}_{11}\, \mathbf{e}_t \tag{24}$$

being $\mathbf{S}_{11}$ the symmetric part of matrix $\mathbf{L}_{11}$:

$$\mathbf{S}_{11} = \frac{\mathbf{L}_{11} + \mathbf{L}_{11}^\top}{2}$$

Then, the convergence of $\mathbf{e}_t$ depends only on the positiveness of matrix $\mathbf{S}_{11}$. Given the structure of matrix $\mathbf{L}_{11}$ (19), the eigenvalues of its symmetric part can be easily computed, being:

$$\begin{aligned}
\lambda_1 &= \frac{\rho + \|\mathbf{t}_a\|}{2\,\rho} > 0 \\
\lambda_2 &= \frac{\|\mathbf{t}_a\| + 1}{2\,\rho} + \frac{1}{2} + \frac{\mathbf{t}_a^\top \mathbf{R}_a \mathbf{n}_a}{2\,\rho\,\|\mathbf{t}_a\|} > \lambda_3 \\
\lambda_3 &= \frac{\|\mathbf{t}_a\| - 1}{2\,\rho} + \frac{1}{2} + \frac{\mathbf{t}_a^\top \mathbf{R}_a \mathbf{n}_a}{2\,\rho\,\|\mathbf{t}_a\|} \geq 0
\end{aligned}$$

The condition for the first eigenvalue is clear as we know that ρ is positive. The second eigenvalue is greater than the third one, as their difference is $\lambda_2 - \lambda_3 = \frac{1}{\rho} > 0$. Finally, it can be proved that the third eigenvalue is always non-negative and it only becomes null in a particular configuration [9], namely

$$\frac{\mathbf{R}_a^\top \mathbf{t}_a}{\|\mathbf{t}_a\|} = -\mathbf{n}_a \tag{25}$$

The geometric interpretation of this condition is shown in the left drawing of Figure 2. In this figure, we can see that

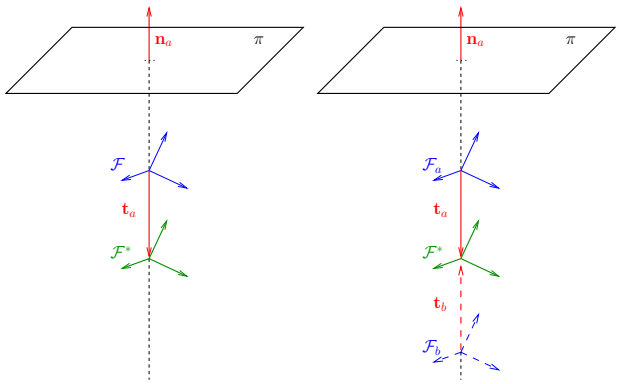

Fig. 2. Geometric configuration in which $\mathbf{L}_{11}$ becomes singular.

$\mathbf{t}_a$, if expressed in the same frame as $\mathbf{n}_a$, that is, frame $\mathcal{F}^*$, is parallel to the latter. According to relation (25), the only possible configuration should be the one depicted in the figure, that is, the current frame between the desired frame and the object plane, so the translation vector points in the opposite direction to $\mathbf{n}_a$. However, it is easy to see that the other configuration, when the current frame is behind the desired frame is also possible. The reason is very simple, if we make notice of the peculiar relation existing between Rtn_a and Rtn_b in the configuration described by (25). In this particular case, (9)-(11) become:

$$\mathbf{t}_b = -\mathbf{t}_a\,; \qquad \mathbf{n}_b = -\mathbf{n}_a\,; \qquad \mathbf{r}_b = \mathbf{r}_a \tag{26}$$

This situation is depicted in the right-most drawing of Figure 2, where the "true" and "false" solutions are shown. During our developments, we have been assuming that Rtn_a corresponds to the true solution, and Rtn_b to the false one. However, in practice we assume there is no way of distinguishing between them. This means that, Rtn_b can be the true solution, instead of Rtn_a. When this happens, we are in the situation when the real solution for the current frame is behind the desired one. Then we need to generalize the geometric configuration (25) by this new one:

$$\frac{\mathbf{R}_a^\top \mathbf{t}_a}{\|\mathbf{t}_a\|} = \pm\mathbf{n}_a \tag{27}$$

One further consideration, derived from the relations (26), is that one of the two solutions Rtn_a, Rtn_b will not verify the visibility constraint in this configuration, as the normals are pointing in opposite directions. Resuming our study of $\dot{V}_t$ (24) and according to the previous paragraphs, we can state that $\dot{V}_t$ is always non-positive and it is null only at the equilibrium point, $\mathbf{e}_t = \mathbf{0}$. The reason is that the only eigenvalue of $\mathbf{S}_{11}$ that can be zero, λ_3, only becomes effectively zero when the relation (27) holds. We know that in this configuration $\mathbf{t}_b = -\mathbf{t}_a$, what implies that the mean, and hence the translation error, are null. As a conclusion for $\mathbf{e}_t$, it always converges to the equilibrium point $\mathbf{e}_t = \mathbf{0}$, that coincides with the geometric configuration (27).

1) $\|\mathbf{t}_a\|$ never increases using the mean-based control law:

We need to complete the previous analysis to make sure that, during the convergence of $\mathbf{e}_t$, the current camera frame does not go away, at the risk of losing visibility of the object. Recalling the expression:

$$\dot{\mathbf{t}}_a = \boldsymbol{\upsilon} - [\mathbf{t}_a]_\times\,\boldsymbol{\omega}$$

we analyze if the time derivative of the norm of $\mathbf{t}_a$ can be positive:

$$V_{t_a} = \frac{1}{2}\,\|\mathbf{t}_a\|^2 \quad \Longrightarrow \quad \dot{V}_{t_a} = \mathbf{t}_a^\top \dot{\mathbf{t}}_a = \mathbf{t}_a^\top \boldsymbol{\upsilon}$$

Using our mean-based control law $\boldsymbol{\upsilon} = -\lambda\,\mathbf{e}_t$:

$$\dot{V}_{t_a} = -\lambda\,\mathbf{t}_a^\top \mathbf{t}_m = -\frac{\lambda}{2}\,\mathbf{t}_a^\top\,(\mathbf{t}_a + \mathbf{t}_b) \tag{28}$$

¿From relation (9) we can write $\mathbf{t}_b$ as the product of a matrix and $\mathbf{t}_a$:

$$\mathbf{t}_b = \mathbf{M}\,\mathbf{t}_a\,; \qquad \mathbf{M} = \frac{1}{\rho}\left(\frac{2}{\|\mathbf{t}_a\|}\,\mathbf{R}_a\,\mathbf{n}_a\,\mathbf{t}_a^\top + \|\mathbf{t}_a\|\,\mathbf{I}\right)$$

Using this expression in (28), we can write:

$$\dot{V}_{t_a} = -\frac{\lambda}{2}\,\mathbf{t}_a^\top \mathbf{A}\,\mathbf{t}_a\,; \qquad \mathbf{A} = \mathbf{I} + \mathbf{M}$$

where the symmetric part of matrix $\mathbf{A}$, denoted by $\mathbf{S}_A$, can be introduced:

$$\dot{V}_{t_a} = -\frac{\lambda}{2}\,\mathbf{t}_a^\top \mathbf{S}_A\,\mathbf{t}_a$$

Computing the eigenvalues of matrix $\mathbf{S}_A$, we find out that they are exactly double of the eigenvalues of matrix $\mathbf{S}_{11}$, for which we concluded they were always positive, except at the equilibrium point, $\mathbf{e}_t = \mathbf{0}$, where $\lambda_3 = 0$. This confirms

that $\|\mathbf{t}_a\|$ is always non-increasing using the proposed mean-based control law.

B. Stability in the orientation error $\mathbf{e}_r$

We define the following Lyapunov function candidate:

$$V_r = \frac{1}{2}\,\mathbf{e}_r^\top \mathbf{e}_r$$

being its time derivative:

$$\dot{V}_r = \frac{d\|\mathbf{e}_r\|}{dt} = \mathbf{e}_r^\top \dot{\mathbf{e}}_r$$

The expression for the derivative of $\mathbf{e}_r$ is:

$$\dot{\mathbf{e}}_r = \mathbf{L}_{21}\,\boldsymbol{\upsilon} + \mathbf{L}_{22}\,\boldsymbol{\omega} = -\lambda\,\mathbf{L}_{21}\,\mathbf{e}_t - \lambda\,\mathbf{L}_{22}\,\mathbf{e}_r$$

Considering that $\mathbf{e}_t$ always converges to zero, as we have just seen, the first addend goes to zero. This is why the particular form of $\mathbf{L}_{21}$ does not matter, as said before. Regarding to the second addend, the positiveness of matrix $\mathbf{L}_{22}$ has to be proved. Even being simple, we can avoid it if we do not consider $\mathbf{L}_{22}$ alone, but as part of the product:

$$\mathbf{L}_{22}\,\mathbf{e}_r = \mathbf{L}_{22}\,\mathbf{r}_m$$

Replacing $\mathbf{L}_{22}$ using (23) and considering that, as $\mathbf{r}_m$ is in the direction of the rotation axis of $\mathbf{R}_m$, it does not change under this rotation: $\mathbf{R}_m\,\mathbf{r}_m = \mathbf{r}_m$, this product reduces to:

$$\mathbf{L}_{22}\,\mathbf{r}_m = \frac{1+\|\mathbf{r}_m\|^2}{2}\,\mathbf{r}_m$$

Thus, we obtain:

$$\frac{d\|\mathbf{e}_r\|}{dt} \;\rightarrow\; -\lambda \mathbf{e}_r^\top \mathbf{L}_{22}\,\mathbf{e}_r = -\lambda\frac{1+\|\mathbf{r}_m\|^2}{2}\mathbf{e}_r^\top \mathbf{e}_r$$

Since this is always non-positive, the conclusion for $\mathbf{e}_r$ is that it always converges to $\mathbf{e}_r = \mathbf{0}$, that is, $\mathbf{R}_a = \mathbf{R}_b = \mathbf{I}$.

C. Conclusions on the stability of the mean-based control

At this point, the conclusion for the stability of the complete position-based control scheme is that global asymptotic stability can be achieved using the mean of the true and the false solutions. The only limitation is due to the use of $\tan(\theta_m/2)$ in $\mathbf{r}_m$, that may produce saturation as the mean angle θ_m goes to $\pm\pi$. Finally, it must be noticed that the achieved equilibrium point, $\mathbf{e} = \mathbf{0}$, is not the desired configuration in which the current camera frame coincides with the desired one:

$$\mathbf{e} = \mathbf{0} \;\not\Rightarrow\; \begin{bmatrix}\mathbf{t}_a \\ \mathbf{r}_a\end{bmatrix} = \mathbf{0}$$

Instead, it corresponds to a line in the Cartesian space defined by:

$$\frac{\mathbf{R}_a^\top \mathbf{t}_a}{\|\mathbf{t}_a\|} = \pm\mathbf{n}_a\,; \qquad \mathbf{R}_a = \mathbf{I} \tag{29}$$

$$\mathbf{e} = \mathbf{0} \;\Rightarrow\; \begin{bmatrix}\frac{\mathbf{t}_a}{\|\mathbf{t}_a\|} \\ \mathbf{r}_a\end{bmatrix} = \begin{bmatrix}\pm\mathbf{n}_a \\ \mathbf{0}\end{bmatrix}$$

That is, the current camera frame is properly oriented according to the reference frame, but the translation error always converges reaching a configuration parallel to the reference-plane normal. This will be overcome using a switching control law as we will see in the next section.

V. SWITCHING CONTROL LAW

As it has been shown in the previous sections, the mean-based control law always takes the system to an equilibrium where:

$$\mathbf{t}_b = -\mathbf{t}_a\,; \qquad \mathbf{n}_b = -\mathbf{n}_a\,; \qquad \mathbf{R}_b = \mathbf{R}_a = \mathbf{I}$$

This is not completely satisfactory, as $\mathbf{t}_a$ can be different from zero, as would be required. Now, we want to improve this control law so the desired equilibrium:

$$\begin{bmatrix}\mathbf{t}_a \\ \mathbf{r}_a\end{bmatrix} = \mathbf{0} \implies \begin{cases}\mathbf{t}_a = \mathbf{t}_b = \mathbf{0} \\ \mathbf{R}_a = \mathbf{R}_b = \mathbf{I}\end{cases} \tag{30}$$

is reached (as we know that the norm $\mathbf{t}_b$ is always equal to the norm of $\mathbf{t}_a$, both solutions must have simultaneously null translation). As said before, in practice we choose two solutions that verify the visibility constraint at the beginning and use their average for controlling the system, until it converges to a particular configuration in the Cartesian space (29). During this convergence the true normal $\mathbf{n}_a$ does not change, since the object does not move and the reference frame, $\mathcal{F}^*$, is also motionless. On the other hand, the false normal, $\mathbf{n}_b$, will change from its original direction until it becomes opposite to $\mathbf{n}_a$. During this continuous evolution, it is clear that, at some point, $\mathbf{n}_b$ no longer verifies the visibility constraint. This means that the control based on the average of the two solutions drives the current frame in such a way that it is always possible to detect the false solution, among the two that verified the visibility constraint at the beginning. Then, it could be possible to control the system from this instant on, using just the true solution, that takes the system to the desired equilibrium (30). According to this, a switching control strategy can be proposed, in such a way that when one of the two solutions comes out to be a false one, we start making a smooth transition from the mean control to the control using only the true solution. A smooth transition is preferred to immediately discarding the false solution, in order to avoid any abrupt changes in the evolution of the control signals. Then, we can replace the average control law (12)-(13) by a weighted-average control law:

$$\begin{cases}\mathbf{t}_m \;=\; \dfrac{\alpha_a\,\mathbf{t}_a + \alpha_b\,\mathbf{t}_b}{2} \\ \mathbf{r}_m \;\Longleftarrow\; \mathbf{R}_m = \mathbf{R}_a\,(\mathbf{R}_a^\top \mathbf{R}_b)^{\frac{\alpha_b}{2}}\end{cases}$$

The weighting coefficients α_a and α_b can be defined according to an exponentially decreasing time-function:

$$f(t) = e^{-\lambda_f\,(t-t_\perp)} \tag{31}$$

being $\alpha_a = 2 - f(t)$ and $\alpha_b = f(t)$. The proposed switching strategy can also be replaced by a more complex one [11], [12].

VI. SIMULATION RESULTS

A. Mean-based control

We have validated the proposed control laws with the simulation depicted in Figure 3. The initial orientation error is 36 degrees and the scalar λ used in the control law (16) is chosen $\lambda = 1$. We can notice in the figure that, as expected, $\mathbf{t}_a$ does not converge to zero. In particular, as in the experiment $\mathbf{n}_a = [0, 0, 1]^\top$, we can see that only the third component of $\mathbf{t}_a$ is different from zero at the equilibrium. At the convergence we obtain $\mathbf{n}_b = -\mathbf{n}_a$ so we know which is the true solution. Thus, we can switch the control law in order to achieve the positioning task.

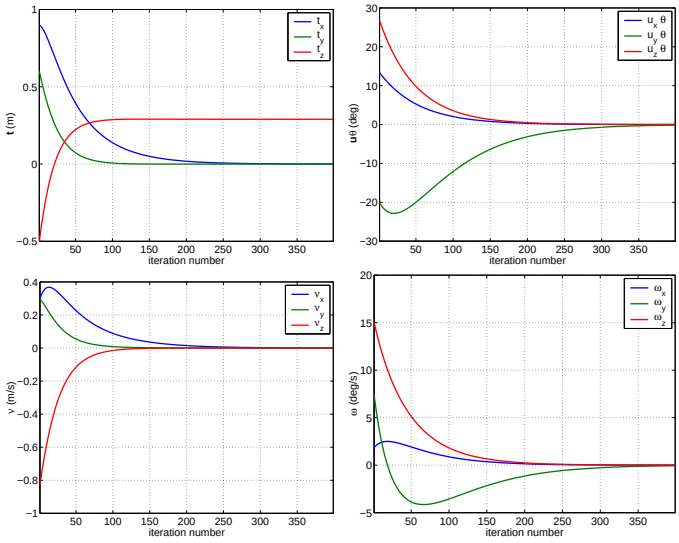

Fig. 3. Simulation experiment using the mean of true and false solutions.

B. Switching control

The second simulation shows the performance of the switching control strategy (see Figure 4). In particular, the control system starts switching at the 14th iteration since the visibility constraint allows to find which is the true normal even before the convergence of the mean-based control law. The chosen value for the switching-rate parameter is $\lambda_f = 0.25$. It has been implemented as a discrete-time switching, as $(t - t_\perp)$ in (31) has been replaced by $(k - k_\perp)$, being k the current iteration number and $k_\perp$ the iteration number when the false solution was detected. As expected, it can be seen that the system converges to the equilibrium (the rotation and translation errors go to zero) where the desired camera pose is reached.

VII. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed a new method for the vision-based control of a camera observing a planar object. The task of positioning the camera with respect to the plane cannot be achieved without any additional information due to the existence of two possible solutions in the homography decomposition problem. Taking advantage of a new analytic formulation for this decomposition problem, we have been able to prove the stability of a control law which moves the camera so that it enables to find the true solution of the homography decomposition problem. Then, a switching control law has been proposed to accomplish the positioning task. Future work will be focused on the study of the effects of camera calibration errors on the proposed vision-based control.

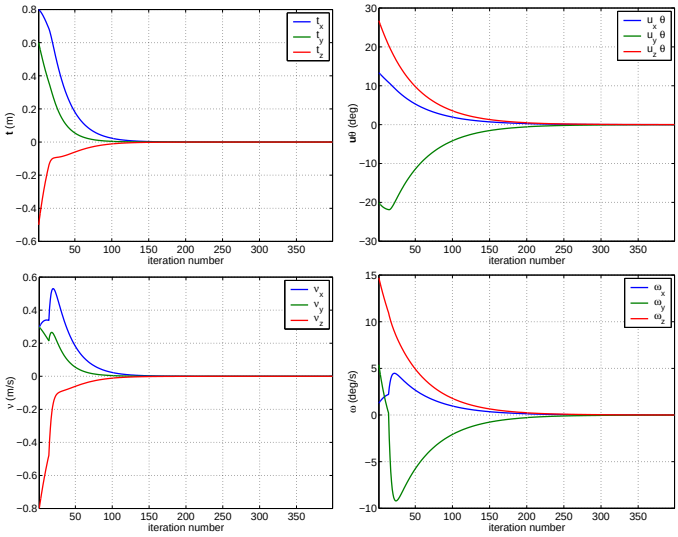

Fig. 4. Simulation experiment using switching control.

REFERENCES

[1] O. Faugeras and F. Lustman, "Motion and structure from motion in a piecewise planar environment", *in International Journal of Pattern Recognition and Artificial Intelligence*, 2(3):485–508, 1988.

[2] S. Hutchinson, G. D. Hager and P. I. Corke, "A tutorial on Visual Servo Control", *in IEEE Transaction on Robotics and Automation*, 12(5):651–670, October 1996.

[3] E. Malis, F. Chaumette and S. Boudet, "2 1/2 D Visual Servoing", *in IEEE Transaction on Robotics and Automation*, 15(2):234–246, April 1999.

[4] E. Malis, "Hybrid vision-based robot control robust to large calibration errors on both intrinsic and extrinsic camera parameters", *in European Control Conference*, pp. 2898-2903, Porto, Portugal, September 2001.

[5] Y. Fang, D.M. Dawson, W.E. Dixon, and M.S. de Queiroz, "Homography-Based Visual Servoing of Wheeled Mobile Robots", *in Proc. IEEE Conf. Decision and Control*, pp. 2866-282871, Las Vegas, NV, Dec. 2002.

[6] E. Malis and F. Chaumette, "Theoretical improvements in the stability analysis of a new class of model-free visual servoing methods", *in IEEE Transaction on Robotics and Automation*, 18(2):176–186, April 2002.

[7] S. Benhimane and E. Malis, "Real-time image-based tracking of planes using efficient second-order minimization", *in IEEE/RSJ International Conference on Intelligent Robots Systems*, Sendai, Japan, October 2004.

[8] M. Moakher., "Means and averaging in the group of rotations", *in SIAM Journal on Matrix Analysis and Applications*, Vol. 24, Number 1, pp. 1-16, 2002.

[9] E. Malis and M. Vargas, "Deeper understanding pf the homography decomposition for vision-based control", *Research Report* INRIA Sophia-Antipolis, number 6303, September 2007.

[10] C. Samson, M. Le Borgne, and B. Espiau. *Robot control: the task function approach*, Vol. 22 of *Oxford Engineering Science Series*. Clarendon Press, Oxford, UK, 1991.

[11] P. Soueres, V. Cadenat, and M. Djeddou. "Dynamical sequence of multi-sensor based tasks for mobile robots navigation" , *in SYROCO*, vol. 2, pp. 423-428, September 2002.

[12] N. Mansard, F. Chaumette. "Tasks sequencing for visual servoing"., *In IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, vol. 1, pp. 992-997, Sendai, Japan, September 2004.