Exploiting the Trifocal Tensor in Dynamic Pose-Estimation for Visual Control

H. M. Becerra[1] and C. Sagüés[2]

***Abstract*—Image-based approaches for visual control are memoryless and they depend on the information extracted from the image plane. We propose the use of dynamic pose estimation into the task of driving a mobile robot to a desired location specified by a target image. This approach reduces the dependence of the control on the quality of visual data and facilitates the planning of complex tasks. The pose estimation exploits the 1D trifocal tensor (TT) as measurement, which allows to obtain a semicalibrated estimation scheme that is valid for any visual sensor obeying a central projection model. The contribution of the paper is a novel observability analysis of the estimation problem from the 1D TT using nonlinear tools, as well as the demonstration of the validity of closed loop control from the estimated pose by showing a separation principle in our nonlinear framework. The overall position-based scheme drives the robot to a desired pose through smooth velocities without the need of a target model neither scene reconstruction nor depth information. The effectiveness of the approach is evaluated via real-world experiments.**

I. INTRODUCTION

Nowadays, the research in mobile robotics is motivated by the introduction of wheeled mobile robots in service applications, where the precise positioning of the robot is an important task. In this context, a vision system is a very good sensor for mobile robots [1]. In this paper we present a visual servoing (VS) approach to drive a wheeled mobile robot to a desired pose (position and orientation), which is specified by a target image previously captured. VS methods can be classified as image-based (IB) when image data is used directly in the control loop, or position-based (PB) if an estimate of pose parameters is carried out [2]. Pioneering works on VS of mobile robots are classical IB schemes using largely overconstrained control commands, for instance [3]. More recently, geometric constraints relating two views have been applied to improve the performance and robustness of IB schemes, e.g., the epipolar geometry [4] and the homography model [5]. Nevertheless, these geometric constraints have both serious drawbacks. The epipolar geometry is ill-conditioned with short baseline and with planar scenes, while the homography model is not well defined if there are no dominant planes in the scene.

Although the trifocal tensor (TT) is more general, more robust and without the drawbacks of other geometric constraints [6], it has been less exploited in visual control. The 2D TT has been introduced for visual control of mobile robots using an overconstrained controller that may suffer of local minima problems [7]. This is overcome in [8] by defining a square control system and by using direct feedback of the elements of the 1D TT into a two-step switching control law. The use of more than two views in VS provides robustness as well as enough information to correct depth from visual feedback, which is not possible from two views.

IBVS schemes are memoryless, which makes this type of approach completely dependent on the information extracted from the image plane. This dependence can be reduced with the use of an estimation strategy, which improves robustness by filtering and smoothing the measurements. Additionally, the pose estimation facilitates the planning of complex tasks, like obstacle avoidance. In this sense, dynamic estimation is of great interest in VS, and has been introduced in [9]. In the context of mobile robots, on one hand, an approach that recovers the robot pose has been proposed using structure from motion in [10]. On the other hand, dynamic estimation has been used in [11], where the authors propose a Kalman filtering approach to match a set of landmarks to a prior map and then to estimate the robot pose from these visual observations. The effectiveness of applying a Kalman filtering approach on PBVS has been particularly studied in [12]. We have introduced preliminary results of a pose estimation scheme from several measurements given by the 1D TT in [13], where the tensor is computed from metric information and a linear observability analysis is presented.

In this paper, a mapless pose estimation scheme using one element of the 1D TT as measurement and its application in visual servoing of mobile robots is presented. The proposed estimation scheme is valid for any type of central vision system. Therefore, when the robot pose is estimated from the adequate visual sensor and used for feedback, the visibility constraint problem is overcome. The core of our contribution is a comprehensive observability analysis of the estimation problem that has been developed using nonlinear tools and a stability analysis of the closed loop, showing the validity of a separation principle between estimation and control in our nonlinear framework. In the context of control, although the orientation is a DOF in the closed loop system, the proposed scheme ensures total correction of both, position and orientation, using a single controller and smooth control inputs. Additionally, the proposed approach for pose estimation does not require a precise camera calibration, neither a target model nor scene reconstruction or depth information.

The paper is organized as follows. Section II describes the motion model of the robot and the 1D TT. Section III presents an extensive observability analysis of the estimation problem.

This work was supported by project DPI 2009-08126 and grants of Banco Santander-Universidad de Zaragoza and Conacyt-México.

[1]H.M. Becerra is with Centro de Investigación en Matemáticas (CIMAT), 36240, Guanajuato, Gto., Mexico, hector.becerra@cimat.mx

[2]C. Sagüés is with Instituto de Investigación en Ingeniería de Aragón, Universidad de Zaragoza, C/ María de Luna 1, E-50018, Zaragoza, Spain, csagues@unizar.es

Section IV details the control law and the stability of the closed loop. Section V shows the performance evaluation real-world experiments using a hypercatadioptric imaging system, and finally, Section VI states the conclusions.

II. MATHEMATICAL MODELING

A. Robot Model

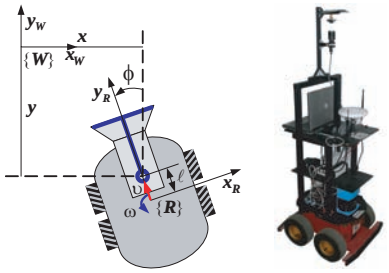

Fig. 1. Configuration of the robot with an on-board central camera. Left: Kinematic model. Right: Experimental platform.

The kinematic motion of many wheeled robotic platforms can be characterized through a differential-drive model. This work focuses on driving a wheeled mobile robot from visual information under the framework that is depicted in Fig. 1(left). A central camera is fixed to the robot in a position ℓ translated along the longitudinal axis y_R. The motion of the robot frame $\{R\}$ with respect to a world frame $\{W\}$ can be expressed using the unicycle model $\dot{x} = -\upsilon \sin\phi$, $\dot{y} = \upsilon\cos\phi$ and $\dot{\phi} = \omega$, where υ and ω represent the translational and rotational velocities, respectively. The kinematic behavior of the on-board camera is described by the following continuous-time model:

$$\begin{aligned} \dot{\mathbf{x}} &= \begin{bmatrix} g_1(\mathbf{x}) & g_2(\mathbf{x}) \end{bmatrix} \mathbf{u}, \\ \mathbf{y} &= h(\mathbf{x}). \end{aligned} \tag{1}$$

This driftless affine system has state vector $\mathbf{x} = [x, y, \phi]^T$, input vector $\mathbf{u} = [\upsilon, \omega]^T$ and input vector fields $g_1(\mathbf{x}) = [-\sin\phi, \cos\phi, 0]^T$ and $g_2(\mathbf{x}) = [-\ell\cos\phi, -\ell\sin\phi, 1]^T$. The measurement vector $\mathbf{y}$, modeled as a nonlinear function of the state $h(\mathbf{x})$, will be defined through information extracted from images. By applying an Euler approximation (forward difference) on the continuous derivatives [11], the discrete versions of the state equations are obtained:

$$\begin{aligned} x_{k+1} &= x_k - T_s\left(\omega_k \ell\cos\phi_k + \upsilon_k \sin\phi_k\right), \\ y_{k+1} &= y_k - T_s\left(\omega_k \ell\sin\phi_k - \upsilon_k \cos\phi_k\right), \\ \phi_{k+1} &= \phi_k + T_s\omega_k \end{aligned} \tag{2}$$

where T_s is the sampling period. In the sequel, we use the notation $s\phi = \sin\phi$, $c\phi = \cos\phi$.

B. The 1D Trifocal Tensor as Visual Measurement

Omnidirectional vision in visual servoing provides the important advantage of avoiding problems with the scene leaving the field of view, which requires a special navigation strategy when conventional cameras are used [14]. However, given that omnidirectional vision systems are represented by a nonlinear projection model [15], the use of the radial trifocal tensor (1D TT) simplifies the synthesis of visual servo-controllers.

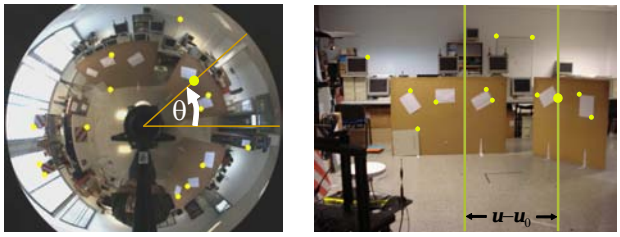

Fig. 2. Extracted bearing measurements from central cameras to estimate the 1D TT. Left: Hypercatadioptric image, the bearing measurement θ is converted to its 1D projection as $\mathbf{p} = [\sin\theta, \cos\theta]^T$. Right: Perspective image, where the 1D projection is obtained from the normalization of the difference u-u_0

The 1D TT is a geometric constraint that encodes the geometry among three views in the frame of planar motion, which is a typical situation in the context of wheeled mobile robots. This constraint behaves better for general scenes in comparison with the epipolar geometry and the homography model [6]. The estimation of the 1D TT is basically the same for any camera that approximately obeys the generic camera model [15], e.g., conventional cameras and its combination with mirrors (catadioptric systems). In order to estimate the 1D TT, point features are converted to their projective formulation in a 1D virtual retina as shown in Fig. 2 for hypercatadioptric systems and conventional cameras.

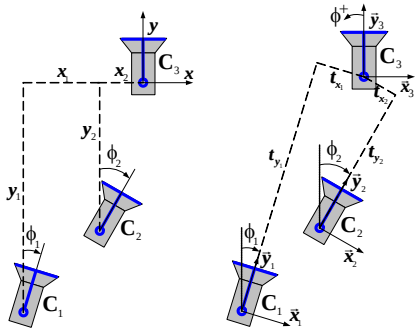

Fig. 3. Complete geometry between three camera locations. A global reference frame is placed in the third view. Left: Absolute locations. Right: Relative locations.

Let us define a global (world) reference frame as depicted in Fig. 3(left) with the origin in the third camera. Then, the camera locations at the initial, current and target views with respect to that global reference are $\mathbf{C}_1 = (x_1, y_1, \phi_1)$, $\mathbf{C}_2 = (x_2, y_2, \phi_2) = \mathbf{x}$ and $\mathbf{C}_3 = (0, 0, 0)$, respectively. The relative locations between cameras are defined by a local reference frame in each camera as shown in Fig. 3(right). The geometry of the three views is encoded in eight tensor elements $T_{l,m,n}$ $(l, m, n = 1, 2)$ that in general can be expressed as:

$$T_{l,m,n} = -\kappa_1 t_x + \kappa_2 t_y + \kappa_3 f_{sc}\left(\phi\right) \tag{3}$$

where $t_x = -xc\phi - ys\phi$, $t_y = xs\phi - yc\phi$, $f_{sc}\left(\phi\right)$ can be $s\phi$ or $c\phi$, and κ_1, κ_2, κ_3 are adequate constants depending on the geometric relationships between the fixed views $\mathbf{C}_1$ and $\mathbf{C}_3$. Thus, the tensor elements are a function of the current camera-robot state $\mathbf{x}$. Details about the expression of each element of

the 1D TT can be verified in [16], [8]. It is worth mentioning that in practice, it is a need to normalize the tensor elements in order to fix a scale, which means to divide each element by a non-null factor. Notice that the bearing measurements are independent on focal length, so that, only the center of projection for omnidirectional images or the principal point for conventional cameras is required to estimate the 1D TT. Thus, the use of bearing information allows to get a semicalibrated measurement for pose estimation and visual control, in contrast to previous approaches [4], [17].

III. DYNAMIC POSE ESTIMATION FROM 1D TT

The high dependence of the IB control schemes on the information extracted from the image plane can be reduced by combining a prediction of the current robot pose with robust visual measurements. The problem of pose estimation exploiting a geometric constraint for visual control has been tackled in the literature in a static fashion ([10], [5]), which means that the pose is extracted by decomposing a mathematical entity at each time instant. In this section, we analyze the use of the information provided by the 1D TT in order to estimate the camera-robot pose dynamically, i.e., using a true state estimator. We prove that the system is not observable from a continuous-time viewpoint, however, a discrete-time analysis shows that with an exciting controller the system achieves observability. This controller is designed in the next section.

Observability is a structural property of a system that may affect the convergence of an estimation scheme. This property specifies if two states are distinguishable by measuring the output, i.e., $\mathbf{x}_1 \neq \mathbf{x}_2 \Longrightarrow h(\mathbf{x}_1) \neq h(\mathbf{x}_2)$. Few works concerned about nonlinear state observability of mobile robots. Some basic results on the observability analysis of the estimation problem from 1D TT measurements using linear theory are reported in [13]. In that work, the observability is ensured by using three different elements of the tensor, which seems trivial to estimate three unknown quantities. Currently, new results on the observability of the robot pose with only one measurement from the 1D TT is reported using nonlinear tools. Firstly, the nonlinear theory for the analysis of continuous systems introduced in [18] is used. According to this theory, the following *observability rank condition* can be enunciated.

Definition 1. *The continuous-time nonlinear system that describes the camera-robot kinematics (1) with some measurement $h(\mathbf{x})$ is locally weakly observable if the observability matrix with rows $\mathbf{O}_c \triangleq \left[\nabla \mathscr{L}^q_{g_i g_j} h(\mathbf{x}) \shortmid i,j=1,2;\, q \in \mathbb{N}\right]^T$ is of full rank n for every $\mathbf{x}$.*

The expression $\mathscr{L}^q_{g_i} h(\mathbf{x})$ denotes the qth order Lie derivative of the scalar function h along the vector field g_i and ∇ is the gradient operator. From the previous observability rank condition, it can be seen that the observability property of the nonlinear camera-robot system depends on the excitations, since it is a driftless system. It suffices to find a set of rows linearly independent in order to fulfill the rank condition. Locally weak observability is a concept stronger than observability, which states that one can instantaneously distinguish each point of the state space from its neighbors, without necessity to travel a considerable distance, as admitted by the observability concept. Let us express the elements of the tensor (3) as a generic measurement model that is a function of the state of the system (1) as follows:

$$h(\mathbf{x}) = \alpha x s\phi + \beta x c\phi + \gamma y s\phi + \delta y c\phi \tag{4}$$

where α, β, γ, δ are suitable constants defined for each tensor element. This expression of the measurement model allows us to generalize the results for any of the eight tensor elements as introduced in the following proposition.

Proposition 1. *The space spanned by all possible Lie derivatives given by the generic measurement model (4) along the vector fields g_1 and g_2 of the continuous system (1) is of dimension three if the measurement accomplishes $\alpha + \delta \neq 0$ or $\beta - \gamma \neq 0$.*

Proof: The proof of this proposition is done by finding the space spanned by all possible Lie derivatives and verifying its dimension. This space is given as:

$$\Omega = \left(h, \mathscr{L}^1_{g_1} h, \mathscr{L}^1_{g_2} h, \mathscr{L}^2_{g_1} h, \mathscr{L}^2_{g_2} h, \mathscr{L}_{g_1}\mathscr{L}_{g_2} h, \mathscr{L}_{g_2}\mathscr{L}_{g_1} h, ...\right)^T. \tag{5}$$

The first order Lie derivatives are as follows:

$$\begin{aligned} \mathscr{L}^1_{g_1} h &= \delta c^2\phi - \alpha s^2\phi + (\gamma - \beta)\, s\phi c\phi = \varphi_a(\phi), \\ \mathscr{L}^1_{g_2} h &= -\ell\left(\beta c^2\phi + \gamma s^2\phi + (\alpha + \delta)\, s\phi c\phi\right) + \tfrac{\partial h}{\partial \phi}. \end{aligned}$$

We have introduced the notation φ for functions depending on ϕ, which emphasizes that some of the Lie derivatives only span in that direction. It is sufficient for the search of independent vectors to find the second order Lie derivatives, which results:

$$\begin{aligned} \mathscr{L}^2_{g_1} h &= 0, \\ \mathscr{L}^2_{g_2} h &= -\ell\left((2\alpha + \delta)\, c^2\phi + 3(\gamma - \beta)\, s\phi c\phi\right) \\ &\quad -\ell\,(2\delta + \alpha)\, s^2\phi - h, \\ \mathscr{L}_{g_1}\mathscr{L}_{g_2} h &= \varphi_b(\phi), \\ \mathscr{L}_{g_2}\mathscr{L}_{g_1} h &= \varphi_c(\phi). \end{aligned}$$

Notice that the Lie derivatives $\mathscr{L}_{g_1}\mathscr{L}_{g_2} h$ and $\mathscr{L}_{g_2}\mathscr{L}_{g_1} h$ span in the same direction of $\mathscr{L}^1_{g_1} h$ and the corresponding gradients of the formers do not contribute to the dimension of the space. The dimension of the observable space is determined by the rank of the matrix (6).

Clearly, by taking the first three rows, this matrix has rank three if $\alpha + \delta \neq 0$ or $\beta - \gamma \neq 0$. Thus, the space spanned by all possible Lie derivatives is of dimension three under such conditions. ■

The interest of the previous result is to conclude about the observability property of the camera-robot pose using one element of the 1D TT. In this sense, the following Lemma is stated.

Lemma 1. *The maximum rank of the nonlinear observability matrix for the continuous camera-robot system (1) with one element of the 1D TT (3) as measurement turns out to be two, and therefore, the system is not locally weakly observable using that information according to Definition 1.*

Proof: This proof results as a derivation of the Proposition 1. From (3), it is seen that, by adding to (4) the term $\kappa_3 f_{sc}(\phi)$

$$\nabla\Omega = \begin{bmatrix} \nabla h \\ \nabla \mathscr{L}^1_{g_1} h \\ \nabla \mathscr{L}^1_{g_2} h \\ \nabla \mathscr{L}^2_{g_2} h \end{bmatrix} = \begin{bmatrix} \alpha s\phi + \beta c\phi & \gamma s\phi + \delta c\phi & \alpha x c\phi - \beta x s\phi + \gamma y c\phi - \delta y s\phi \\ 0 & 0 & -2(\alpha+\delta) s\phi c\phi + (\beta-\gamma)\left(s^2\phi - c^2\phi\right) \\ \alpha c\phi - \beta s\phi & \gamma c\phi - \delta s\phi & \ell\left((\alpha+\delta)\left(s^2\phi - c^2\phi\right) - 2(\beta-\gamma) s\phi c\phi\right) - h \\ -\alpha s\phi - \beta c\phi & -\gamma s\phi - \delta c\phi & \ell\left(6(\alpha+\delta) s\phi c\phi - 3(\beta-\gamma)\left(s^2\phi - c^2\phi\right)\right) - \frac{\partial h}{\partial \phi} \end{bmatrix} \tag{6}$$

$$\mathbf{O}_d = \begin{bmatrix} \kappa_1 c\phi_k + \kappa_2 s\phi_k & \kappa_1 s\phi_k - \kappa_2 c\phi_k & \kappa_1\left(-x_k s\phi_k + y_k c\phi_k\right) + \kappa_2\left(x_k c\phi_k + y_k s\phi_k\right) + \kappa_3 f_{cs}(\phi_k) \\ \kappa_1 c\epsilon_k + \kappa_2 s\epsilon_k & \kappa_1 s\epsilon_k - \kappa_2 c\epsilon_k & \kappa_1\left(-x_k s\epsilon_k + y_k c\epsilon_k\right) + \kappa_2\left(x_k c\epsilon_k + y_k s\epsilon_k\right) + \kappa_3 f_{cs}(\epsilon_k) \\ \kappa_1 c\zeta_k + \kappa_2 s\zeta_k & \kappa_1 s\zeta_k - \kappa_2 c\zeta_k & \kappa_1\left(-x_k s\zeta_k + y_k c\zeta_k\right) + \kappa_2\left(x_k c\zeta_k + y_k s\zeta_k\right) + \kappa_3 f_{cs}(\zeta_k) \end{bmatrix} \tag{7}$$

with κ_3 a constant and $f_{sc}(\phi)$ being $s\phi$ or $c\phi$, the non-normalized elements of the tensor can be expressed in two of the following forms:

1) Four elements of the tensor, T_{121}, T_{122}, T_{221} and T_{222}, can be written as $h_1(\mathbf{x}) = \beta x c\phi + \gamma y s\phi + \kappa_3 f_{sc}(\phi)$. In accordance to the generic measurement model (4), for these tensor elements $\alpha = 0$, $\delta = 0$, $\beta = \gamma$, and consequently the conditions of the previous proposition are not accomplished.
2) The other four elements, T_{111}, T_{112}, T_{211} and T_{212}, can be expressed as $h_2(\mathbf{x}) = \alpha x s\phi + \delta y c\phi + \kappa_3 f_{sc}(\phi)$. In this case, $\beta = 0$, $\gamma = 0$, $\alpha = -\delta$, and the conditions in Proposition 1 to span a space of dimension three are not fulfilled.

Hence, since $\alpha + \delta = 0$ and $\beta - \gamma = 0$ in any case, the observability matrix has only two rows linearly independent

$$\mathbf{O}_c = \begin{bmatrix} \nabla h \\ \nabla \mathscr{L}^1_{g_2} h \end{bmatrix} = \begin{bmatrix} \alpha s\phi + \beta c\phi & \gamma s\phi + \delta c\phi & \frac{\partial h}{\partial \phi} \\ \alpha c\phi - \beta s\phi & \gamma c\phi - \delta s\phi & -h \end{bmatrix}. \tag{8}$$

Given that this matrix $\mathbf{O}_c$ has a lower rank than the dimension of the state space ($n = 3$), the observability rank condition is not satisfied and consequently locally weak observability cannot be ensured for the continuous system (1) with any element of the 1D TT (3) as measurement. The same result is obtained by using any linear combination of elements of the tensor. Higher order Lie derivatives are linearly dependent on lower order derivatives, so that, the maximum rank of the observability matrix is two. ■

Notice that in (8) only appears the gradient of the measurement along g_2, in addition to the gradient of h. This means that, the rotational velocity provides an observable direction in the state space. However, this is not enough to cover the three dimensional space. In contrast, the translational velocity does not contribute to gain an observable direction, because the gradient of Lie derivatives related to g_1 provides linearly dependent row vectors.

Actually, the implementation of an estimation and control algorithm must be done in discrete time. Consequently, an observability analysis in discrete time is mandatory. Consider the following observability rank condition for discrete nonlinear systems [19].

Definition 2. *The discrete-time nonlinear system (2) expressed as $\mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{u}_k)$, with measurement $\mathbf{y}_k = h(\mathbf{x}_k)$ is said to be observable at $\mathbf{x}_k$ if the nonlinear observability matrix*

$$\mathbf{O}_d = \begin{bmatrix} \frac{\partial h}{\partial \mathbf{x}}(\mathbf{x}_k) \\ \frac{\partial h}{\partial \mathbf{x}}(\mathbf{x}_{k+1}) \frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}_k) \\ \vdots \\ \frac{\partial h}{\partial \mathbf{x}}(\mathbf{x}_{k+n-1}) \frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}_{k+n-2}) \cdots \frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}_k) \end{bmatrix} \tag{9}$$

is of full rank n at $\mathbf{x}_k$.

Next, an important result on the analysis of the observability of the camera-robot pose is established in the following Lemma.

Lemma 2. *The discrete camera-robot system (2) is said to be observable according to Definition 2 by using only one element of the 1D TT (3) as measurement if rotational velocity is applied during two consecutive time instants and the corresponding velocities are different and no-null for these two consecutive steps.*

Proof: This is proved by constructing the corresponding nonlinear observability matrix and verifying its rank. Let us consider the generic measurement (3) that represents any of the eight elements of the 1D TT in discrete time:

$$h(\mathbf{x}_k) = -\kappa_1 t_{x_k} + \kappa_2 t_{y_k} + \kappa_3 f_{sc}(\phi_k). \tag{10}$$

The required Jacobians in (9) are as follows:

$$\frac{\partial h}{\partial \mathbf{x}_k} = \begin{bmatrix} \kappa_1 c\phi_k + \kappa_2 s\phi_k \\ \kappa_1 s\phi_k - \kappa_2 c\phi_k \\ -\kappa_1 t_{y_k} - \kappa_2 t_{x_k} + \kappa_3 f_{cs}(\phi_k) \end{bmatrix}^T = \mathbf{H}_k, \tag{11}$$

$$\frac{\partial f}{\partial \mathbf{x}_k} = \begin{bmatrix} 1 & 0 & \Delta_y \\ 0 & 1 & -\Delta_x \\ 0 & 0 & 1 \end{bmatrix} = \mathbf{F}_k, \tag{12}$$

being $\Delta_x = T_s(\omega_k \ell c\phi_k + \upsilon_k s\phi_k)$, $\Delta_y = T_s(\omega_k \ell s\phi_k - \upsilon_k c\phi_k)$. The recursive operations of (9) result in the nonlinear observability matrix (7), in which $\epsilon_k = \phi_k + T_s\omega_k$ and $\zeta_k = \phi_k + T_s(\omega_k + \omega_{k+1})$.

It can be seen that this matrix has three row vectors linearly independent if the following conditions are fulfilled:

$$T_s \neq 0,\ \omega_k \neq 0,\ \omega_{k+1} \neq 0,\ \omega_k \neq \omega_{k+1}. \tag{13}$$

Thus, it is proved that the observability matrix is full rank three and the system (2) is observable if rotational velocity is applied during two consecutive time instants and the velocities are different and no-null at each instant. ■

According to Lemma 2, a digital implementation of an estimation scheme for the system (2) with measurement of the type (10) collects enough information along two time instants. In this sense, both observability analysis, the continuous and

discrete, are complementary each other. The no locally weak observability states that the robot pose cannot be distinguished instantaneously; however, the pose can be estimated in two steps in accordance to the discrete analysis. The conditions for the observability in Lemma 2 confirm the dependence of this property on the control inputs, in particular on the rotational velocity. It is worth mentioning that the previous mathematical developments have been shown for the non-normalized tensor for clarity, and they are also valid for normalized measurements, as required for implementation.

Remark 1. *It can be demonstrated that a Discrete Extended Kalman Filter is an efficient implementation of an observer that accomplishes Lemma 2. The demonstration can be found in [20].*

IV. NONHOLONOMIC VISUAL SERVOING IN THE CARTESIAN SPACE

In this section, we demonstrate the feasibility of closing the control loop using the estimated pose $\hat{\mathbf{x}}_k = \left[\hat{x}_k, \hat{y}_k, \hat{\phi}_k\right]^T$ given by an estimation scheme, in order to drive the robot to reach a desired position and orientation. This can be referred as a pose regulation problem and in the three-view framework defined by the 1D TT, the goal is to achieve $\mathbf{C}_2 = (0, 0, 0)$. Let us define the output to be controlled as the reduced state vector

$$\mathbf{x}_{r,k} = [x_k, y_k]^T . \tag{14}$$

Hence, the tracking errors are $\xi_k^1 = x_k - x_k^d$, $\xi_k^2 = y_k - y_k^d$, where x_k^d and y_k^d are the discrete values of desired smooth time-varying references. The difference equations of these errors result in the following system:

$$\xi_{k+1} = \xi_k + T_s \mathbf{D}\left(\phi_k, \ell\right) \mathbf{u}_k - T_s \dot{\mathbf{x}}_{r,k}^d \tag{15}$$

where $\xi_k = \mathbf{x}_{r,k} - \mathbf{x}_{r,k}^d$, $\mathbf{D}\left(\phi_k, \ell\right) = \begin{bmatrix} -s\phi_k & -\ell c\phi_k \\ c\phi_k & -\ell s\phi_k \end{bmatrix}$, $\mathbf{u}_k = [\upsilon_k, \omega_k]^T$ and $\dot{\mathbf{x}}_{r,k}^d = \left[\dot{x}_k^d, \dot{y}_k^d\right]^T$. Given that the control inputs appear in the first differentiation of each output, the relative degree of the system is 2, and a first order zero dynamics appears. This represents a DOF of the system, which is the orientation (ϕ). As the control is based on estimation, the static state feedback control law $\mathbf{u}_k$ resulting from the inversion of the error system (15) turns out to be

$$\hat{\mathbf{u}}_k = \hat{\mathbf{D}}^{-1}\left(-\mathbf{k}\hat{\xi}_k + \dot{\mathbf{x}}_{r,k}^d\right). \tag{16}$$

where $\mathbf{k} = \text{diag}(k_1, k_2)$ is a matrix of control gains and $\hat{\mathbf{D}}^{-1} = \mathbf{D}^{-1}\left(\hat{\phi}_k, \ell\right)$. It can be verified that the input velocities achieve global stabilization of the position error system (15) in the case of feedback of the real state $\mathbf{u}_k$. In such a case, the dynamic behavior of the closed loop position error is exponentially stable iff k_1 and $k_2 \in (0,\ 2/T_s)$. Notice that the control law (16) is valid when the camera is shifted from the robot rotational axis ($\ell \neq 0$), which is a common situation.

A. Stability of the Estimation-based Control Loop

In this section, the stability of the closed loop system with feedback $\hat{\mathbf{u}}_k$ of the estimated camera-robot pose is analyzed. It is known that the separation principle between estimation and control is not accomplished in general for nonlinear systems. However, we present an adequate analysis for our case and a separation principle is demonstrated. We consider that the estimation scheme is implemented as a Discrete Extended Kalman Filter for the system $\{\mathbf{F}_k, \mathbf{H}_k\}$, which accomplishes Lemma 2. Given the good behavior of the 1D TT that we have obtained in previous works [8], and given that the robot velocities are computed from the control law rather than measured (unlike an odometric approach), we assume that the effect of the noise can be neglected, in such a way that the deterministic model (2), expressed as $\mathbf{x}_{k+1} = f\left(\mathbf{x}_k, \mathbf{u}_k\right)$, with measurement $\mathbf{y}_k = h\left(\mathbf{x}_k\right)$, is used for the subsequent analysis. The behavior of the a priori estimation error in an EKF is

$$\mathbf{e}_{k+1}^- = \mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1}^- = f\left(\mathbf{x}_k, \hat{\mathbf{u}}_k\right) - f\left(\hat{\mathbf{x}}_k^+, \hat{\mathbf{u}}_k\right). \tag{17}$$

It is worth emphasizing that the same result of the analysis can be obtained using the a posteriori estimation error $\mathbf{e}_k^+$ [21]. Let us introduce an expansion of the functions f and h:

$$f\left(\mathbf{x}_k, \hat{\mathbf{u}}_k\right) - f\left(\hat{\mathbf{x}}_k^+, \hat{\mathbf{u}}_k\right) = \mathbf{F}_k\left(\mathbf{x}_k - \hat{\mathbf{x}}_k^+\right) + \Phi\left(\mathbf{x}_k, \hat{\mathbf{x}}_k^+, \hat{\mathbf{u}}_k\right), \tag{18}$$

$$h\left(\mathbf{x}_k\right) - h\left(\hat{\mathbf{x}}_k^-\right) = \mathbf{H}_k\left(\mathbf{x}_k - \hat{\mathbf{x}}_k^-\right) + \Psi\left(\mathbf{x}_k, \hat{\mathbf{x}}_k^-\right), \tag{19}$$

where Φ and Ψ are the second and higher order terms. By substituting (18) into (17) and using the discrete observer as given by the update stage of the EKF $\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + \mathbf{K}_k\left(h\left(\mathbf{x}_k\right) - h\left(\hat{\mathbf{x}}_k^-\right)\right)$, we have:

$$\mathbf{e}_{k+1}^- = \mathbf{F}_k\left(\mathbf{x}_k - \hat{\mathbf{x}}_k^- - \mathbf{K}_k\left(h\left(\mathbf{x}_k\right) - h\left(\hat{\mathbf{x}}_k^-\right)\right)\right) + \Phi\left(\mathbf{x}_k, \hat{\mathbf{x}}_k^+, \hat{\mathbf{u}}_k\right).$$

By substituting (19) and knowing that the a priori estimation error is given as $\mathbf{e}_k^- = \mathbf{x}_k - \hat{\mathbf{x}}_k^-$, then:

$$\mathbf{e}_{k+1}^- = \mathbf{F}_k\left(\mathbf{I}_3 - \mathbf{K}_k\mathbf{H}_k\right)\mathbf{e}_k^- + \mathbf{\Theta}_k$$

where $\mathbf{\Theta}_k = \Phi\left(\mathbf{x}_k, \hat{\mathbf{x}}_k^+, \hat{\mathbf{u}}_k\right) - \mathbf{F}_k\mathbf{K}_k\Psi\left(\mathbf{x}_k, \hat{\mathbf{x}}_k^-\right)$. Let us denote the first two components of the vector $\mathbf{e}_k^-$ as $\mathbf{e}_{r,k}^-$. Then, the estimated tracking error is $\hat{\xi}_k = \xi_k - \mathbf{e}_{r,k}^-$.

The control law (16) can be written using the estimated tracking error as:

$$\hat{\mathbf{u}}_k = \hat{\mathbf{D}}^{-1}\left(-\mathbf{k}\left(\xi_k - \mathbf{e}_{r,k}^-\right) + \dot{\mathbf{x}}_{r,k}^d\right).$$

By introducing this control input into the tracking error system (15), the closed loop difference equation with estimated state feedback results as:

$$\begin{aligned} \xi_{k+1} &= \left(\mathbf{I}_2 - T_s\mathbf{D}\hat{\mathbf{D}}^{-1}\mathbf{k}\right)\xi_k + T_s\mathbf{D}\hat{\mathbf{D}}^{-1}\mathbf{k}\mathbf{e}_{r,k}^- \\ &\quad + T_s\left(\mathbf{D}\hat{\mathbf{D}}^{-1} - \mathbf{I}_2\right)\dot{\mathbf{x}}_{r,k}^d. \end{aligned}$$

The product of matrices

$$\mathbf{D}\hat{\mathbf{D}}^{-1} = \begin{bmatrix} s\phi_k s\hat{\phi}_k + c\phi_k c\hat{\phi}_k & -s\phi_k c\hat{\phi}_k + s\hat{\phi}_k c\phi_k \\ -s\hat{\phi}_k c\phi_k + s\phi_k c\hat{\phi}_k & c\phi_k c\hat{\phi}_k + s\phi_k s\hat{\phi}_k \end{bmatrix}$$

turns out to be a definite positive matrix assuming that the initial estimation of ϕ_k is no $\frac{\pi}{2}$ rad off, according to a TT-based initialization procedure [20]. Finally, the overall closed loop control system with estimated state feedback is expressed as follows:

$$\begin{aligned} \begin{bmatrix} \xi_{k+1} \\ \mathbf{e}_{k+1}^- \end{bmatrix} &= \begin{bmatrix} \mathbf{I}_2 - T_s\mathbf{D}\hat{\mathbf{D}}^{-1}\mathbf{k} & \left[T_s\mathbf{D}\hat{\mathbf{D}}^{-1}\mathbf{k}, 0\right] \\ \mathbf{0} & \mathbf{F}_k\left(\mathbf{I}_3 - \mathbf{K}_k\mathbf{H}_k\right) \end{bmatrix} \begin{bmatrix} \xi_k \\ \mathbf{e}_k^- \end{bmatrix} \\ &\quad + \begin{bmatrix} T_s\left(\mathbf{D}\hat{\mathbf{D}}^{-1} - \mathbf{I}_2\right)\dot{\mathbf{x}}_{r,k}^d \\ \mathbf{\Theta}_k \end{bmatrix}. \end{aligned}$$

Notice that each dynamics ξ_k and $\mathbf{e}_k^-$ is subject to a perturbation. The tracking error is subject to a perturbation depending on the derivative of the desired references and the estimation error is subject to a bounded perturbation $\mathbf{\Theta}_k$ depending on the dynamics of the estimation error. The more important effect comes from the second one, because the former perturbation depends on the accuracy of the orientation estimation and it can be neglected considering the smoothness of the reference signals. Thus, the stability of the overall control scheme is determined by the estimation error dynamics. According to Lemma 4 of [21], the perturbation $\mathbf{\Theta}_k$ can be bounded as $\|\mathbf{\Theta}_k\| \leq \lambda \left\|\mathbf{e}_k^-\right\|^2$, where λ is a real positive number and it is independent on the robot velocities. Given that in our case the EKF fulfills the additional conditions stated in Theorem 7 of [21] (the boundness of the matrices of the linear approximation, the boundness of the estimation covariances and the nonsigularity of the matrix $\mathbf{F}_k$), the perturbation eventually will vanish and the state estimator will behave as an exponential observer. The triangular form of the estimator-control system shows that the stability property is achieved by ensuring the stability of each one of the dynamics ξ_k and $\mathbf{e}_k^-$, i.e., a separation principle holds for the system. This is accomplished in our case and exponential stability of the overall control system is achieved.

B. Pose Regulation through adequate Reference Tracking

Notice that up to now, the proposed controller drives to zero the lateral and longitudinal errors through a smooth evolution, but the orientation evolves freely. In order to obtain also orientation correction for nonholonomic mobile robots, a good option is to define an adequate path for the robot position. The following time-differentiable references are proposed to achieve this goal:

$$\begin{aligned} y_k^d &= \frac{y_0}{2}\left(1+\cos\left(\frac{\pi}{\tau}kT_s\right)\right), \quad 0 \leq kT_s \leq \tau, \\ x_k^d &= \frac{x_0}{y_0^2}\left(y_k^d\right)^2, \quad 0 \leq kT_s \leq \tau \end{aligned} \tag{20}$$

where (x_0, y_0) is the estimated initial position, obtained from an initialization procedure [20], and τ is a user-defined temporal horizon to reach the target position. These references depict a parabolic path on the $x-y$ plane from the point $(x_0,\, y_0)$ to $(0, 0)$. Notice that the robot always starts over the desired path, so that the controller has to maintain the tracking in the path. In case that the robot would start off the path, the stability property of the controller will drive the robot position to asymptotically converge to the desired path. The tracking of the reference drives the robot to perform an initial rotation autonomously when the robot heading is not tangent to the path.

As mentioned previously, when the controlled outputs reach zero at the time τ the so-called *zero dynamics* is achieved in the robot system. Zero dynamics is described by a subset of the state space which makes the output to be identically zero [22]. In the particular case of the robot system (2) with outputs $s_1 = x_k$, $s_2 = y_k$, this set is given as:

$$Z^* = \left\{ \left[0, 0, \bar{\phi}_k\right]^T,\ \bar{\phi}_k = \text{constant} \in \mathbb{R} \right\}.$$

Thus, zero dynamics in this control system means that when lateral and longitudinal positions of the camera-robot system are corrected, the orientation may be different to zero. Next, it is proved that orientation correction is also achieved by tracking the proposed references, in such a way that pose regulation is obtained.

Proposition 2. *The proposed visual servoing scheme with control inputs (16), using feedback of the estimated state* $\hat{\mathbf{x}}_k = \left[\hat{x}_k, \hat{y}_k, \hat{\phi}_k\right]^T$ *provided by the 1D TT-based estimation scheme and the reference signals (20), drives the camera-robot system (2) to reach the location* $(x = 0, y = 0, \phi = 0)$*, i.e., orientation is also corrected.*

Proof: In the previous section we have proved the stability of the position error dynamics with feedback of the estimated state, in such a way that correction of the lateral and longitudinal errors is ensured in τ seconds. It only remains to prove that the orientation is also zero when the target location is reached. From the decomposition of the translational velocity vector given by the kinematic behavior of the robot and using the difference equations $x_{k+1} - x_k = -\delta \hat{v}_k \sin\phi_k$, $y_{k+1} - y_k = \delta \hat{v}_k \cos\phi_k$, we have that

$$\phi_k = \arctan\left(-\frac{x_{k+1} - x_k}{y_{k+1} - y_k}\right).$$

Let us define the parabolic relationship between Cartesian coordinates $x = \frac{x_0}{y_0^2} y^2$ according to the desired trajectories (20). Its corresponding discrete time-derivative results in $x_{k+1} - x_k = 2\frac{x_0}{y_0^2} y_k \left(y_{k+1} - y_k\right)$. Thus, when the x and y-coordinates track the desired trajectories, the robot orientation is related to the current lateral position as follows:

$$\phi_k = \arctan\left(-2\frac{x_0}{y_0^2} y_k\right).$$

As mentioned, when the robot has followed the reference path and $kT_s = \tau$ the position reaches zero ($x = 0$, $y = 0$), and consequently $\phi = \arctan(0) = 0$. This proves that although the orientation is a DOF for the control system, the location $(x = 0, y = 0, \phi = 0)$ is reached in τ seconds by tracking the defined profile (20) for the position coordinates. ■

This behavior can be obtained whenever the tangent of the path is zero at the origin, as in (20). Thus, it is possible to use different functions besides a parabolic one in order to ensure that the robot reaches the target with the desired orientation, for instance, $x^d = x_0\left(1 - \cos\left(y^d \pi / 2 y_0\right)\right)$. However, a smoother performance of the robot motion is obtained using the parabolic path.

Note that pose regulation is achieved using a single controller and smooth control inputs, in contrast to [8], where a two-step control law that produces discontinuous robot velocities is proposed. Additionally, the proposed approach herein takes into account the nonholomicity of wheeled mobile robots unlike [4].

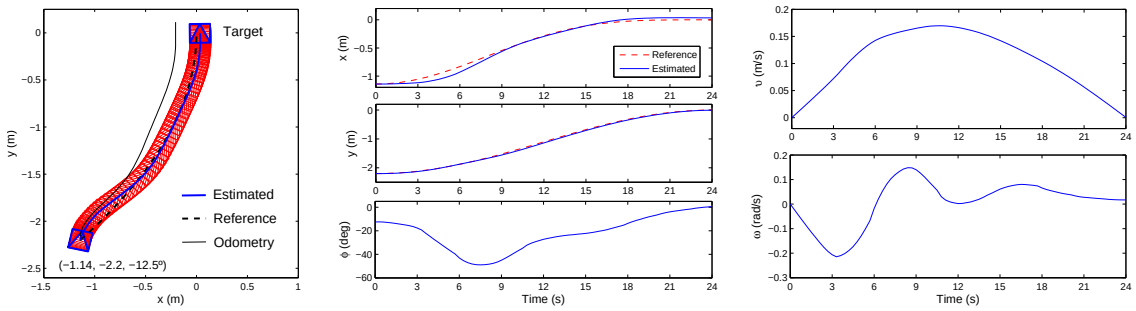

Fig. 4. Experimental evaluation of the closed loop control system. Left: Resultant path plotted using the estimated camera-robot state, but also the reference path and the odometry are shown. Center: Estimated camera-robot state. Right: Computed velocities.

V. EXPERIMENTAL EVALUATION

An extensive evaluation of our proposal through simulations is reported in [20]. This section presents the evaluation of the implemented estimation scheme and the feasibility of closing the loop using the estimated pose for visual servoing purposes through real-world experiments. The experiments herein use the robot Pioneer 3-AT shown in Fig. 1(right), which is equipped with a hypercatadioptric imaging system that consists of a camera Sony XCD-X7101CR and a mirror Neovision H3S.

The omnidirectional images are captured at a size of 1024×768 pixels using the free software Player. This is an adequate image size in order to facilitate the detection of image features in omnidirectional images, which in general present a low resolution. The low resolution is accentuated around the center of projection, however, features in such region are not used. The camera is connected to a laptop onboard the robot (Intel® Core™ 2 Duo CPU at 2.50 GHz with Debian Linux), in which an EKF-based pose estimation and the control law are implemented in C++. The observed scene has been set up with features on different planes in order to ensure a sufficient number of points in the scene. The experiments have been carried out using tracking of features (14 points) as implemented in the OpenCV library. The tracking has a low computational cost and leads to a good behavior of the 1D TT estimation [8].

The 1D TT is estimated using the five-point method [16], with the projection center ($u_0 = 513$, $v_0 = 409$) as the only required information of the imaging system. This point is previously estimated using a RANSAC approach from 3D vertical lines, which project in radial lines in central omnidirectional images. Thus, for this type of images, it is enough to find the point where radial lines join, which avoids the need to obtain the complete camera calibration parameters. The sampling period T_s is set to 0.3 s, which is adequate to obtain a good estimation of the 1D TT and good closed loop frequency. The distance from the camera to the rotation axis of the robot has been roughly set to $\ell = 10$ cm.

We report an experiment where the robot moves directly toward the target. Fig. 4(left) presents the resultant path given by the estimated robot pose for one of the experimental runs. This figure also shows the reference path and the one given by odometry. It can be seen that the estimated path is closer to the reference than the path obtained from odometry. Thus, we assert that the estimation of the camera-robot pose is sufficiently accurate and then, the estimated pose is suitable for feedback control. The duration of the positioning task is fixed to 24 s through the time τ in the references, which is the termination condition of the control law. Fig. 4(center) shows a good behavior in the tracking of references for the position coordinates, which indicates that the robot reaches the target with good precision.

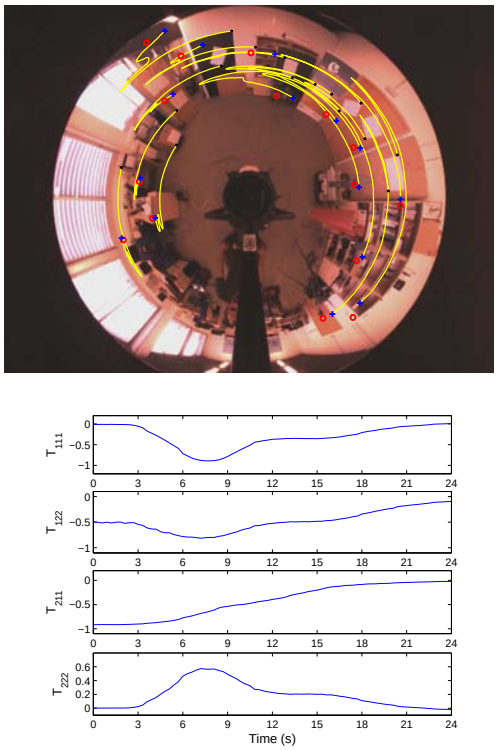

Fig. 5. Behavior of the extracted information from the images. Top: Motion of the point features on the initial image, where the marker "." corresponds to the initial point features, the marker "O" to the target points and the marker "+" are the points in the image at the end of the motion. Bottom: Four normalized tensor elements.

The input velocities given by the proposed control law with feedback of the estimated pose are shown in Fig. 4(right). They behave smoothly along the task, in contrast to the control inputs in previous approaches, e.g., [8]. Fig. 5(top) presents

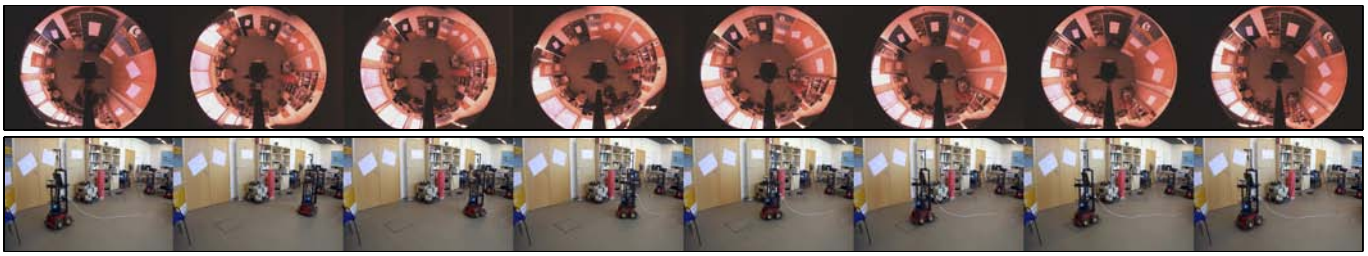

Fig. 6. Sequence of some of the omnidirectional images captured by the hypercatadioptric robot camera during the real experiments (first row). The first is the target image, the second is the initial and the last is the image at the end of the motion. In the second row, the sequence of images taken from an external camera for the same experimental run is shown.

the motion of the image points along the sequence, where the points at the end of the motion (marker "+") are close to the points in the target image (marker "O"). Notice that the point features move smoothly, in such a way that the evolution of the tensor elements is also smooth during the task, as presented in Fig. 5(bottom). Also, it is worth noting that the tensor estimation is not affected when the robot is reaching the target, i.e., there is no problem with the short baseline. Fig. 6 shows a sequence of some images taken by the camera onboard and an external video camera respectively. The robot motion in this experiment is shown in the attached Video[1], together with the view of the on-board camera and evolution of the recorded data. Thus, these results validate the effectiveness of the proposed approach to reach a desired position and orientation using feedback of the estimated camera-robot pose from the 1D TT.

VI. CONCLUSIONS

In this paper we have presented a new generic pose estimation scheme and its application for position-based visual servoing (PBVS) to drive mobile robots to a desired location. This approach reduces the dependence of the control on the visual data and may facilitate the planning of complex tasks. By exploiting the 1D TT estimated from bearing measurements, a semicalibrated estimation scheme that is valid for any visual sensor obeying a central projection model is obtained, so that the visibility constraint problem of visual servoing can be overcome using the appropriate central sensor. Additionally, this approach does not need a target model neither scene reconstruction nor depth information. The core of our contribution is a novel observability study of the nonlinear estimation problem from the 1D TT, as well as the demonstration of the validity of closed loop control from the estimated pose by showing a separation principle in our nonlinear framework. The overall PBVS scheme corrects position and orientation simultaneously using smooth input velocities. The scheme is evaluated via real-world experiments using images from a hypercatadioptric imaging system.

REFERENCES

[1] T.A. Riggs, T. Inanc, and W. Zhang. An autonomous mobile robotics testbed: Construction, validation, and experiments. *IEEE Transactions on Control Systems Technology*, 18(3):757–766, 2010.

[2] F. Chaumette and S. Hutchinson. Visual servo control part I: Basic approaches. *IEEE Robotics and Autom. Magazine*, 13(14):82–90, 2006.

[3] Y. Masutani, M. Mikawa, N. Maru, and F. Miyazaki. Visual servoing for non-holonomic mobile robots. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 1133–1140, 1994.

[4] G. L. Mariottini and D. Prattichizzo. Image-based visual servoing with central catadioptric cameras. *The International Journal of Robotics Research*, 27(1):41–56, 2008.

[5] J. Chen, D.M. Dawson, W.E. Dixon, and A. Behal. Adaptive homography-based visual servo tracking for a fixed camera configuration with a camera-in-hand extension. *IEEE Transactions on Control Systems Technology*, 13(5):814–825, 2005.

[6] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.

[7] G. López-Nicolás, J.J. Guerrero, and C. Sagüés. Visual control through the trifocal tensor for nonholonomic robots. *Robotics and Autonomous Systems*, 58(2):216–226, 2010.

[8] H. M. Becerra, G. López-Nicolás, and C. Sagüés. Omnidirectional visual control of mobile robots based on the 1D trifocal tensor. *Robotics and Autonomous Systems*, 58(6):796–808, 2010.

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

[10] E. Royer, M. Lhuillier, M. Dhome, and J. M. Lavest. Monocular vision for mobile robot localization and autonomous navigation. *International Journal of Computer Vision*, 74(3):237–260, 2007.

[11] A. K. Das, R. Fierro, V. Kumar, B. Southall, J. Spletzer, and C. J. Taylor. Real-time vision-based control of a nonholonomic mobile robot. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 1714–1718, 2001.

[12] A. Shademan and F. Janabi-Sharifi. Sensitivity analysis of EKF and iterated EKF pose estimation for position-based visual servoing. In *IEEE Conference on Control Applications*, pages 755–760, 2005.

[13] H. M. Becerra and C. Sagüés. Pose-estimation-based visual servoing for differential-drive robots using the 1D trifocal tensor. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pages 5942–5947, 2009.

[14] G. Morel, P. Zanne, and F. Plestan. Robust visual servoing: bounding the task function tracking errors. *IEEE Transactions on Control Systems Technology*, 13(6):998–1009, 2005.

[15] C. Geyer and K. Daniilidis. An unifying theory for central panoramic systems and practical implications. In *European Conference on Computer Vision*, pages 445–461, 2000.

[16] J.J. Guerrero, A.C. Murillo, and C. Sagüés. Localization and matching using the planar trifocal tensor with bearing-only data. *IEEE Transactions on Robotics*, 24(2):494–501, 2008.

[17] H. H. Abdelkader, Y. Mezouar, N. Andreff, and P. Martinet. Image-based control of mobile robot with central catadioptric cameras. In *IEEE Int. Conf. on Robotics and Automation*, pages 3522–3527, 2005.

[18] R. Hermann and A. J. Krener. Nonlinear controlability and observability. *IEEE Transactions on Automatic Control*, 22(5):728–740, 1977.

[19] H. Nijmeijer. Observability of autonomous discrete time nonlinear systems: A geometric approach. *International Journal of Control*, 36(5):867–874, 1982.

[20] H. M. Becerra. Unifying vision and control for mobile robots. http://webdiis.unizar.es/~bhecerra/ThesisHMBecerra.pdf, April 2011. PhD Thesis, University of Zaragoza.

[21] K. Reif and R. Unbehauen. The extended kalman filter as an exponential observer for nonlinear systems. *IEEE Transactions on Signal Processing*, 47(8):2324–2328, 1999.

[22] S. Sastry. *Nonlinear Systems: Analysis, Stability and Control*. Springer, New York, 1999.

[1]This video is attached as a multimedia complement.