Direct Visual Servoing: Vision-Based Estimation and Control Using Only Nonmetric Information

Geraldo Silveira and Ezio Malis

Abstract—**This paper addresses the problem of stabilizing a robot at a pose specified via a reference image. Specifically, this paper focuses on six degrees-of-freedom visual servoing techniques that require neither metric information of the observed object nor precise camera and/or robot calibration parameters. Not requiring them improves the flexibility and robustness of servoing tasks. However, existing techniques within the focused class need prior knowledge of the object shape and/or of the camera motion. We present a new visual servoing technique that requires none of the aforementioned information. The proposed technique directly exploits 1) the projective parameters that relate the current image with the reference one and 2) the pixel intensities to obtain these parameters. The level of versatility and accuracy of servoing tasks are, thus, further improved. We also show that the proposed nonmetric scheme allows for path planning. In this way, the domain of convergence is greatly enlarged as well. Theoretical proofs and experimental results demonstrate that visual servoing can, indeed, be highly accurate and robust, despite unknown objects and imaging conditions. This naturally encompasses the cases of color images and illumination changes.**

Index Terms—**Computer vision, image registration, intensity-based methods, lighting variations, projective information, vision-based control.**

I. Introduction

Visual servoing consists in controlling the motion of a robot from the feedback of images. This paper addresses the problem of stabilizing all six degrees of freedom (DOF) of a holonomic robot at a pose defined by means of a reference image. Of course, the reference image must fully constrain all of its 6 DOF to make that possible. This framework is usually referred to as teach by showing in the visual servoing community [1]. An intuitive strategy to solve that problem is to consider the vision sensor as a 3-D sensor. From the provided pose, standard feedback control laws can be used to perform the stabilization. This approach is called pose-based (or more commonly, position-based) visual servoing [2]. This scheme can be applied even to nonholonomic and underactuated robots. On the other hand, it requires very precise camera and robot calibration parameters, as well as the metric model of the observed object. If this metric model is not available, then the monocular localization problem cannot be solved accurately. Another classical solution to that vision-based stabilization problem is to define the control error in the image space [2]. This approach is commonly called image-based visual servoing. This scheme is remarkably more robust to errors in the camera and robot calibration parameters. Nevertheless, it still requires minimal metric knowledge of the object (a coarse depth distribution) to provide a stabilizing control law [3]. The domain of stability is enlarged by the hybrid visual servoing strategy proposed in [4]. This strategy is hybrid in the sense that part of the control error is defined in the image space and part of it is expressed in the Euclidean space. However, it still needs a metric estimate of the planar object (a coarse normal vector, which is a parameterization of its depths) to decide between the two admissible solutions of the required homography decomposition [5].

Specifically, this paper focuses on visual servoing techniques that do not require metric information of the observed target and can control all 6 DOF of a robot. The fact of not requiring metric information improves the flexibility and robustness of visual servoing tasks [6]. Indeed, recent studies in the domain of biological vision have suggested that the brain processes visual information nonmetrically [6]. Surprisingly, only few works have been conducted on the full 6 DOF nonmetric visual servoing. Moreover, these existing works require prior knowledge of the object shape and/or of the camera motion. A first set of examples consists of the methods solely based on the fundamental or essential matrix (see, e.g., [7]). These methods require both nonplanar objects and a sufficient amount of translation to avoid the degeneracies. They are, hence, problematic near the convergence of the visual servo. Another set of examples are the strategies solely based on the homography matrix (see, e.g., [8]). These techniques are, thus, designed for planar objects and/or pure rotational motions.

In this paper, we present a new 6-DOF visual servoing technique that requires no metric information of the object, regardless of its shape and of the camera motion. In this way, system versatility is further improved. The proposed approach generalizes [8] in the sense that nonplanar objects are now also encompassed. Furthermore, the proposed nonmetric control error is isomorphic to the camera pose around the equilibrium for the largest possible domain of rotations. This property does not hold in [8], e.g., for a camera rotation of 180°. Isomorphism is an extremely important property of systems not defined in the metric space. Indeed, it ensures that the nonmetric control error (and control law) is null if and only if the camera pose corresponds to the desired one. Another strength of the proposed approach is that it allows for path planning, which is not the case in [8]. (Of course, that path is also defined in a nonmetric space.) This represents another improvement since path planning can significantly enlarge the domain of convergence of servoing tasks [9].

The proposed control error is constructed from the projective parameters that geometrically relate the current image with the reference one. The estimation procedure to obtain these parameters exploits the pixel intensities, instead of image features (e.g., points, lines, etc.). This can be performed via direct image registration methods (see, e.g., [10]). Existing visual servoing techniques that exploit the pixel intensities (see, e.g., [11] and [12]) do not fall into the class considered in this paper. Indeed, the work in [11] cannot control all 6 DOF of a robot, and the work in [12] requires metric information. Although feature-based techniques could also be applied to estimate those projective parameters, we argue that intensity-based methods are valuable in the context of visual servoing. First of all, higher accuracy can be achieved since much more infomation is exploited. In fact, even image regions where no feature exists can be used. Additionally, image features getting in and out of the field of view can destabilize the control system more easily. This is due to eventual discontinuities, singularities, and even loss of all tracked features. Hence, the fact of using more visual information leads not only to higher accuracy but to more stable servoing techniques as well. Another advantage is that, within direct methods of estimation, the robustness to general illumination changes can be achieved, even in color images [10].

The term "direct" within the proposed technique are, thus, twofold. On the estimation aspect, it highlights the operation at the signal level, without intermediate measures; on the control aspect, it emphasizes that

Manuscript received August 2, 2011; revised February 9, 2012; accepted March 9, 2012. Date of publication April 11, 2012; date of current version August 2, 2012. This paper was recommended for publication by Associate Editor C. C. Cheah and Editor B. J. Nelson upon evaluation of the reviewers' comments.

G. Silveira is with the Center for Information Technology Renato Archer, Division of Robotics and Computer Vision, CEP 13069-901 Campinas, Brazil (e-mail: Geraldo.Silveira@cti.gov.br).

E. Malis is with the French National Institute for Research in Computer Science and Control, 06902 Sophia-Antipolis, France (e-mail: Ezio.Malis@sophia.inria.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TRO.2012.2190875

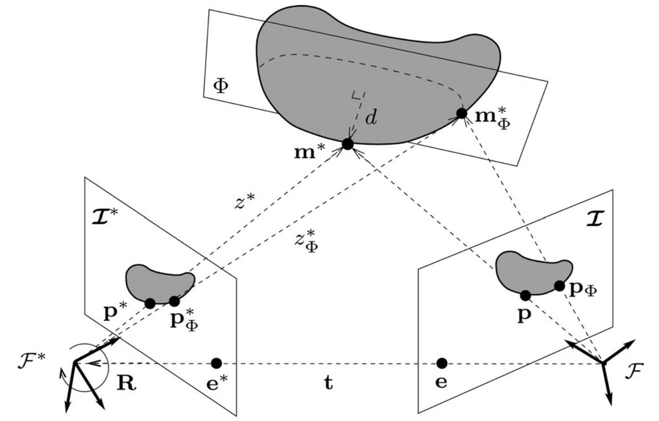

Fig. 1. Geometry of two views of a nonplanar object. Let the dominant (virtual) plane of this object be defined by Φ. The projective parallax ρ^* of the 3-D point $\mathbf{m}^*$ with respect to Φ is proportional to its distance $d(\mathbf{m}^*, \Phi)$ and is inversely proportional to its depth z^*. Hence, the projective parallax ρ^*_Φ of the 3-D point $\mathbf{m}^*_\Phi$ (this object point lies on Φ) relatively to Φ is $\rho^*_\Phi = 0$.

only nonmetric quantities are used, without decompositions or priors. That space is closer to the one where the task is specified (i.e., an image). This paper is built on the theoretical material presented in [13] and [14], adding an experimental validation with a real robot that was not performed in previous works. In addition to the theoretical results, various experiments have been conducted using objects of different shapes, under large initial displacements, large errors in the camera intrinsic parameters, as well as under varying illumination conditions.

II. Preliminaries

Consider a 3-D point $\mathbf{m}^* = [x^*, y^*, z^*]^\top$ defined with respect to the reference frame $\mathcal{F}^*$. This point is projected in the n-channel reference image $\mathcal{I}^*$, $n \geq 1$, as a pixel with homogeneous coordinates $\mathbf{p}^* = [u^*, v^*, 1]^\top \in \mathbb{P}^2$ (see Fig. 1). Of course, $n = 1$ refers to a grayscale image. The intensity value of this pixel is represented by $\mathcal{I}^*(\mathbf{p}^*)$. Let the camera be displaced from $\mathcal{F}^*$ by a translation $\mathbf{t} \in \mathbb{R}^3$ and a rotation $\mathbf{R} = \exp([\boldsymbol{w}]_\times) \in \mathbb{SO}(3)$, where the vector $\boldsymbol{w} = \theta\mathbf{u} \in \mathbb{R}^3$ contains the angle of rotation θ and the unit axis of rotation $\mathbf{u}$. The notations $[\boldsymbol{w}]_\times$ and $\mathrm{vex}([\boldsymbol{w}]_\times)$ represent, respectively, the skew-symmetric matrix associated with the vector $\boldsymbol{w} = [w_1, w_2, w_3]^\top$ and its inverse operator, i.e.,

$$[\boldsymbol{w}]_\times = \begin{bmatrix} 0 & -w_3 & w_2 \\ w_3 & 0 & -w_1 \\ -w_2 & w_1 & 0 \end{bmatrix}, \quad \mathrm{vex}([\boldsymbol{w}]_\times) = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}, \tag{1}$$

where the latter extracts the components of that vector from the skew-symmetric matrix. Given the displacement of the camera from $\mathcal{F}^*$, let the current frame be denoted as $\mathcal{F}$. The coordinates of that same 3-D point $\mathbf{m}^*$ is expressed in the basis of $\mathcal{F}$ as $\mathbf{m} = [x, y, z]^\top$ using the equation of rigid-body motion. From this new pose, the current image $\mathcal{I}$ is acquired. That 3-D point is projected in $\mathcal{I}$ as the pixel $\mathbf{p} = [u, v, 1]^\top \in \mathbb{P}^2$, whose current intensity value is represented by $\mathcal{I}(\mathbf{p})$. Finally, let $\|\mathbf{v}\|$ and $\mathbf{v}'$, respectively, denote the Euclidean norm and a normalized or transformed version of the original variable $\mathbf{v}$.

A. Two-View Geometry

The relation between corresponding points in a pair of images (see Fig. 1) can be formally described in different geometries [5]. This section recalls the two most important ones to this paper.

1) Euclidean Description: Euclidean geometry relates corresponding pixel coordinates $\mathbf{p} \leftrightarrow \mathbf{p}^*$ through

$$\mathbf{p} \propto \mathbf{K}\,\mathbf{R}\,\mathbf{K}^{-1}\,\mathbf{p}^* + (z^*)^{-1}\mathbf{K}\,\mathbf{t}, \tag{2}$$

from the equation of rigid-body motion, the Thales' theorem, and the perspective camera model with intrinsic parameters $\mathbf{K} \in \mathbb{R}^{3\times3}$ (focal lengths $\alpha_u, \alpha_v > 0$, skew, and principal point). The symbol "$\propto$" denotes proportionality up to a nonzero scale factor.

2) Projective Description: Projective geometry is an extension of Euclidean geometry, and naturally includes the perspective projection performed by a camera [5]. In this case, the general relation between corresponding pixel coordinates $\mathbf{p} \leftrightarrow \mathbf{p}^*$ is given by

$$\mathbf{p} \propto \mathbf{G}\,\mathbf{p}^* + \rho^*\mathbf{e}, \tag{3}$$

where $\mathbf{G} \in \mathbb{SL}(3)$ is a projective homography relative to a dominant (virtual) plane, $\mathbf{e} \in \mathbb{R}^3$ denotes the epipole (strictly speaking, $\mathbf{e} \in \mathbb{P}^2$), and $\rho^* \in \mathbb{R}$ is the projective parallax of the 3-D point whose projection in $\mathcal{I}^*$ is $\mathbf{p}^*$, i.e., of $\mathbf{m}^*$, relatively to that virtual plane. This parallax is proportional to the distance of $\mathbf{m}^*$ to that virtual plane and is inversely proportional to the depth of this 3-D point. Hence, $\rho^* = 0$ if and only if the 3-D point $\mathbf{m}^*$ lies on the virtual plane. Of course, if the object is planar, then that plane corresponds to the object itself and all of its 3-D points have zero parallax. A procedure to estimate these parameters is presented in Section IV.

B. Two Visual Servoing Techniques

This section recalls two visual servoing techniques that have greatly inspired the proposed one.

1) Pose-Based Visual Servoing: This technique aims to stabilize the robot pose $(\mathbf{R}, \mathbf{t})$ that is reconstructed from images. To solve the visual localization problem, one needs some knowledge of the metric model of the object, as well as of the camera and robot calibration parameters. The accuracy of this information is crucial to the accuracy of the servoing task. Many solutions to this reconstruction problem have been presented in the literature [5].

Definition 2.1: A *pose-based control error* $\bar{\boldsymbol{\varepsilon}} \in \mathbb{R}^6$ can be defined as

$$\bar{\boldsymbol{\varepsilon}} = \begin{bmatrix} \bar{\boldsymbol{\varepsilon}}_v \\ \bar{\boldsymbol{\varepsilon}}_\omega \end{bmatrix} = \begin{bmatrix} \mathbf{t} \\ \theta\mathbf{u} \end{bmatrix}, \tag{4}$$

where $\bar{\boldsymbol{\varepsilon}}_\omega$ can be computed from the rotation matrix $\mathbf{R}$ via

$$\mathbf{r} = \frac{1}{2}\,\mathrm{vex}\big(\mathbf{R} - \mathbf{R}^\top\big), \tag{5}$$

$$\theta = \begin{cases} \arcsin(\|\mathbf{r}\|), & \text{if } \mathrm{tr}(\mathbf{R}) \geq 1, \\ \pi - \arcsin(\|\mathbf{r}\|), & \text{otherwise,} \end{cases} \tag{6}$$

$$\mathbf{u} = \frac{\mathbf{r}}{\|\mathbf{r}\|}, \tag{7}$$

where θ is the angle of rotation, $\mathbf{u}$ is the unit axis of rotation, and the function $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. If $\|\mathbf{r}\| = 0$, then $\mathbf{u}$ is not determined and, therefore, can be chosen arbitrarily (e.g., $\mathbf{u} = [0, 0, 1]^\top$). The angle-axis representation in (4) is important to this paper due to its link to the Lie algebra [15].

Having constructed the control error (4), standard feedback control strategies can, then, be applied to perform the stabilization.

2) Homography-Based Visual Servoing: The visual servoing technique proposed in [8] aims to control the motion of a camera with respect to a planar object. Let this plane be defined by Π, whose metric information is not required. The method is, in fact, based on the projective homography $\mathbf{G}_\Pi \in \mathbb{SL}(3)$ induced by this plane between two views. It first performs a normalization step

$$\mathbf{H}_\Pi = \mathbf{K}^{-1}\,\mathbf{G}_\Pi\,\mathbf{K}, \quad \mathbf{m}^{*\prime}_\Pi = \mathbf{K}^{-1}\,\mathbf{p}^*_\Pi, \tag{8}$$

where $\mathbf{p}^*_\Pi \in \mathbb{P}^2$ is a chosen image point (not necessarily an interest point) of this object, also called control point.

Definition 2.2: The *plane-based nonmetric control error* $\boldsymbol{\varepsilon}_\Pi \in \mathbb{R}^6$ is defined as

$$\boldsymbol{\varepsilon}_\Pi = \begin{bmatrix} \boldsymbol{\varepsilon}_{\upsilon\,\Pi} \\ \boldsymbol{\varepsilon}_{\omega\,\Pi} \end{bmatrix} = \begin{bmatrix} (\mathbf{H}_\Pi - \mathbf{I})\,\mathbf{m}^{*\prime}_\Pi \\ \boldsymbol{r}_\Pi \end{bmatrix}, \tag{9}$$

with

$$\boldsymbol{r}_\Pi = \operatorname{vex}\big(\mathbf{H}_\Pi - \mathbf{H}_\Pi^\top\big). \tag{10}$$

The control error (9) is proven in [8] to be locally isomorphic to the camera pose at the equilibrium, within a limited domain of rotations. This domain is limited since, e.g., both $\theta = 0$ and $\theta = \pi$ are mapped to by $\boldsymbol{\varepsilon}_{\omega\,\Pi} = \boldsymbol{r}_\Pi = \mathbf{0}$. Local asymptotic stability is proven to be ensured via a proportional control law using (9).

Remark 2.1: It is easy to verify that $\mathbf{G}_\Pi$ is a particular case of the general morphism $\mathbf{G}$ in (3). Indeed, for planar objects and/or pure rotations between views, (3) is simplified using $\rho^*\mathbf{e} = \mathbf{0}$.

III. Direct Visual Servoing: Control Aspects

This section presents a new visual servoing technique to stabilize a robot with respect to a rigid object of unknown shape. The main idea is to extend the plane-based nonmetric control error (9) to encompass nonplanar objects, while having a structure as closely as possible to the pose-based control error (4). In this way, we combine the advantages of both strategies, namely, the nonmetric control of the object in the image, while searching for an optimal camera trajectory.

A. Control Error and Some Properties

The proposed control error uses the set of projective information $\mathbf{g} = \{\mathbf{G}, \mathbf{e}, \rho^*\}$ that geometrically relates the current image with the reference one. This relation is given in (3), and the procedure to estimate that information will be described in Section IV. A first step to construct that error is to perform the following normalization:

$$\mathbf{H} = \mathbf{K}^{-1}\,\mathbf{G}\,\mathbf{K}, \quad \mathbf{e}' = \mathbf{K}^{-1}\,\mathbf{e}, \quad \mathbf{m}^{*\prime} = \mathbf{K}^{-1}\,\mathbf{p}^*. \tag{11}$$

Details on how to choose the control point $\mathbf{p}^*$ will be given in Section III-B. Let us first provide some important definitions.

Definition 3.1: The *proposed nonmetric control error* $\boldsymbol{\varepsilon} \in \mathbb{R}^6$ is defined as

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_\upsilon \\ \boldsymbol{\varepsilon}_\omega \end{bmatrix} = \begin{bmatrix} (\mathbf{H} - \mathbf{I})\,\mathbf{m}^{*\prime} + \rho^*\mathbf{e}' \\ \vartheta\boldsymbol{\mu} \end{bmatrix}, \tag{12}$$

where $\boldsymbol{\varepsilon}_\omega$ can be computed from the homography matrix $\mathbf{H}$ via

$$\boldsymbol{r} = \frac{1}{2}\operatorname{vex}\big(\mathbf{H} - \mathbf{H}^\top\big), \tag{13}$$

$$\vartheta = \begin{cases} \operatorname{real}\big(\arcsin(\|\boldsymbol{r}\|)\big), & \text{if } \operatorname{tr}(\mathbf{H}) \geq 1, \\ \pi - \operatorname{real}\big(\arcsin(\|\boldsymbol{r}\|)\big), & \text{otherwise}, \end{cases} \tag{14}$$

$$\boldsymbol{\mu} = \frac{\boldsymbol{r}}{\|\boldsymbol{r}\|}, \tag{15}$$

where $\boldsymbol{\mu}$ can be viewed as a unit "projective axis of rotation," ϑ can be viewed as a "projective angle of rotation," and $\operatorname{real}(\cdot)$ is used since ϑ is a real-valued scalar. If $\|\boldsymbol{r}\| = 0$, then $\boldsymbol{\mu}$ is not determined and, therefore, can be chosen arbitrarily (e.g., $\boldsymbol{\mu} = [0, 0, 1]^\top$).

Let us state two remarks regarding this control error as follows.

Remark 3.1: It is important to note that the control error (12) is constructed without requiring any metric information of the observed object, regardless of its shape and of the camera motion.

Remark 3.2: The proposed nonmetric control error (12) could be modified in several ways. For example, a decoupled translation error could be devised by using the epipole solely. The rotational error could also be defined differently. In other terms, a modified version $\boldsymbol{\varepsilon}' \in \mathbb{R}^6$ of (12) could simply be defined as

$$\boldsymbol{\varepsilon}' = \begin{bmatrix} \boldsymbol{\varepsilon}'_\upsilon \\ \boldsymbol{\varepsilon}'_\omega \end{bmatrix} = \begin{bmatrix} \mathbf{e}' \\ \boldsymbol{r}' \end{bmatrix}, \tag{16}$$

with

$$\boldsymbol{r}' = \operatorname{vex}\big(\mathbf{H} - \mathbf{H}^\top\big). \tag{17}$$

The translation error in (16) is decoupled from the rotation since $\mathbf{e}' = \mathbf{K}^{-1}\mathbf{e} \propto \mathbf{K}^{-1}\mathbf{K}\,\mathbf{t} \propto \mathbf{t}$. However, if the object is planar, then one is not sure if the recovered epipole corresponds to the correct solution because, in this case, there exist two admissible ones [5]. Furthermore, the coupling present in (12) is not a major concern to the stability of the system because a path planning can be performed (see Section III-C). As for the modified $\boldsymbol{\varepsilon}'_\omega$ in (16), which is equivalent to setting $\vartheta = 2\|\boldsymbol{r}\|$, some improvements are achieved through $\boldsymbol{\varepsilon}_\omega$ in (12) as will be formally stated in Corollary 3.1. In particular, it considers the largest possible domain of rotations.

Before announcing an important property of the proposed control error (local isomorphism in Theorem 3.1), let us state how this error is related to the camera pose. This relation is only used in theoretical demonstrations, i.e., it is not used for servoing the system.

Lemma 3.1 (Control error and camera pose): The proposed nonmetric control error (12) can be expressed as a function of the camera pose (i.e., of the rotation $\mathbf{R} \in \mathbb{SO}(3)$, and of the translation $\mathbf{t} \in \mathbb{R}^3$) between the current frame and the reference one through

$$\boldsymbol{\varepsilon}_\upsilon = \frac{1}{z^*}\,\big((\mathbf{R} - \mathbf{I})\,\mathbf{m}^* + \mathbf{t}\big), \tag{18}$$

and $\boldsymbol{\varepsilon}_\omega$ through

$$\boldsymbol{r} = \sin(\theta)\mathbf{u} + \frac{1}{2}[\mathbf{q}^{*\prime}]_\times\mathbf{t}, \tag{19}$$

where the vector $\mathbf{q}^{*\prime} \in \mathbb{R}^3$ defines the dominant plane of the object.

Proof: The proof is developed in [16]. □

Theorem 3.1 (Local isomorphism): The proposed nonmetric control error (12) is locally isomorphic to the camera pose at the equilibrium $\boldsymbol{\varepsilon} = \mathbf{0}$. Moreover, this holds around the equilibrium for the largest possible domain of rotations, since only $\theta = 0$ is mapped to by $\boldsymbol{\varepsilon}_\omega = \mathbf{0}$.

Proof: The proof is presented in [16]. □

Corollary 3.1 (Generality and improvements): The proposed nonmetric control error (12) is a generalization of (9) to encompass nonplanar objects and unknown motion between views. Indeed, (12) does not assume that $\rho^*\mathbf{e} = \mathbf{0}$, as in (9). Moreover, the proposed one improves the convergence properties of the servoing as well as encompasses the control error presented in [4].

B. Control Law and Stability Analysis

Consider a camera-mounted holonomic robot observing a motionless object of unknown shape.

Definition 3.2: Let $\mathbf{v} = [\boldsymbol{\upsilon}^\top, \boldsymbol{\omega}^\top]^\top \in \mathbb{R}^6$ represent, respectively, the translational and rotational velocities of the camera. The *proposed nonmetric control law*

$$\mathbf{v} = \boldsymbol{\Lambda}\,\boldsymbol{\varepsilon}, \tag{20}$$

with control gain

$$\mathbf{\Lambda} = \operatorname{diag}(\lambda_v \mathbf{I}, \lambda_\omega \mathbf{I}) = \begin{bmatrix} \lambda_v \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \lambda_\omega \mathbf{I} \end{bmatrix}, \quad \lambda_v, \lambda_\omega > 0, \tag{21}$$

uses the control error $\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}_v^\top, \boldsymbol{\varepsilon}_\omega^\top]^\top$ in (12) to compute the input signals (i.e., the camera velocities).

Let us state two remarks regarding this control law as follows.

Remark 3.3: It is important to note that the control law (20) is constructed without requiring any metric information of the observed object, regardless of its shape and of the camera motion.

Remark 3.4: The control law (20) has a positive sign. This is due to the definition of the control error (12), which exploits the geometric set of parameters $\mathbf{g} = \{\mathbf{G}, \mathbf{e}, \rho^*\}$. This set defines a mapping from $\mathcal{F}^*$ to $\mathcal{F}$ and not the opposite. See (3).

The stability analysis is performed in the sequel.

Theorem 3.2 (Local stability): The proposed nonmetric control law (20) ensures local asymptotic stability of the closed-loop system provided that the control point $\mathbf{p}^* \in \mathbb{P}^2$ (11) is chosen such that its parallax $\rho^* \in \mathbb{R}$ is sufficiently small.

Proof: The proof is presented in [16]. □

Corollary 3.2 (Parallax condition): There always exists a control point $\mathbf{p}^* \in \mathbb{P}^2$ (11) with zero parallax (and, thus, can be chosen as the control point). This is due to the fact that, by definition, the dominant plane always crosses the object. Therefore, the closed-loop system is always locally asymptotically stable.

Although the control point can be selected such that its parallax is zero, for robustness reasons, it should be chosen as closely as possible to the center of the object. This reduces the possibility of the object getting out of the field of view due to calibration errors. Indeed, controlling in the image some object's part represents an attractive characteristic of the proposed scheme. Visual servoing techniques that perform this control are known to be more robust to camera and robot calibration errors [1]. An even higher degree of robustness is obtained if path planning is also performed [9].

C. Path Planning

There are two main motivations to perform path planning within visual servoing: 1) to improve its robustness to calibration errors and 2) to deal with large camera displacements. These are both achieved because a large control error is divided into smaller ones, which means always servoing around the equilibrium. Thus, the task can be executed inside the proven domain of stability. This section proposes a straightforward strategy for path planning using the direct visual servoing. To this end, instead of regulating $\boldsymbol{\varepsilon}(t) \to \mathbf{0}$, an appropriate path tracking $\boldsymbol{\varepsilon}(t) \to \boldsymbol{\varepsilon}^*(t)$ is performed.

Definition 3.3: Path tracking is accomplished by regulating the *time-varying nonmetric control error*

$$\boldsymbol{\varepsilon}^r(t) = \boldsymbol{\varepsilon}(t) - \boldsymbol{\varepsilon}^*(t), \quad \forall t \in [0, T], \tag{22}$$

given the control error (12) and a time-varying desired signal

$$t \mapsto \boldsymbol{\varepsilon}^*(t) = [\boldsymbol{\varepsilon}_v^{*\top}(t), \boldsymbol{\varepsilon}_\omega^{*\top}(t)]^\top. \tag{23}$$

Let us state two remarks regarding this control error as follows.

Remark 3.5: It is important to note that the time-varying control error (22) can be constructed without requiring any metric information of the object, regardless of its shape and of the camera motion.

Remark 3.6: The time-varying control error (22) presents interesting characteristics due to the properties of the control error (12): 1) only one image point has its trajectory planned. Therefore, physically valid camera situations are always specified; 2) the "projective axis-angle" parameterization already provides for a smooth trajectory. Hence, there is no need for postprocessing (e.g., interpolation) the divided path to obtain C^2 trajectories; 3) given the local isomorphism, there are no singularities along the entire planned path if its division is made sufficiently high.

Two examples of desired paths are presented in the following.

Case 3.1 (Linear path): A first example of interest consists in specifying the time-varying desired signal (23) as a linear path such that $\boldsymbol{\varepsilon}^*(0) = \boldsymbol{\varepsilon}(0)$ and $\boldsymbol{\varepsilon}^*(T) = \mathbf{0}$, i.e.,

$$\boldsymbol{\varepsilon}^*(t) = \boldsymbol{\varepsilon}^*(0) + \Big(\boldsymbol{\varepsilon}^*(T) - \boldsymbol{\varepsilon}^*(0)\Big)\frac{t}{T} = \boldsymbol{\varepsilon}^*(0)\left(1 - \frac{t}{T}\right). \tag{24}$$

This case is of special interest to the translational part of (23), i.e., to $\boldsymbol{\varepsilon}_v^*(t)$. Indeed, the trajectory of the control point in the image is, thus, specified as a straight line.

Case 3.2 (Geodesic path): A second example of desired path is given here. It is of special interest to the rotational part of (23), i.e., to $\boldsymbol{\varepsilon}_\omega^*(t)$. From the results of the Lemma 3.1, we have that

$$\boldsymbol{\varepsilon}_\omega = \vartheta\boldsymbol{\mu} \to \theta\mathbf{u} = \bar{\boldsymbol{\varepsilon}}_\omega \qquad \text{as} \qquad \mathbf{t} \to \mathbf{0}. \tag{25}$$

In other terms, if $\mathbf{t} = \mathbf{0}$, then geodesic rotations will be induced. This motivates us to replan the desired signal for the rotational part of (24) as the camera approaches the goal. To this end, its rotational part can be slightly changed to

$$\boldsymbol{\varepsilon}_\omega^*(t) = \boldsymbol{\varepsilon}_\omega(t)\left(1 - \frac{t}{T}\right). \tag{26}$$

A stabilizing time-varying control law is defined next.

Definition 3.4: In order to regulate the time-varying control error (22) to zero, the control law (20) is transformed into the *time-varying nonmetric control law*

$$\mathbf{v} = \mathbf{\Lambda}(t)\,\boldsymbol{\varepsilon}^r(t) + \frac{\partial \boldsymbol{\varepsilon}^*(t)}{\partial t}, \tag{27}$$

where the feed-forward term $\partial\boldsymbol{\varepsilon}^*(t)/\partial t$ allows compensation of the tracking error, and

$$\mathbf{\Lambda}(t) = \operatorname{diag}\Big(\lambda_v \mathbf{I}, \lambda_\omega(t)\mathbf{I}\Big) = \begin{bmatrix} \lambda\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \lambda \exp\Big(-\gamma\|\boldsymbol{\varepsilon}_v(t)\|\Big)\mathbf{I} \end{bmatrix}, \tag{28}$$

with $\lambda, \gamma > 0$, is a variable gain matrix also motivated by (25): $\lambda_\omega(t) > 0$ is small for large $\|\boldsymbol{\varepsilon}_v(t)\|$, and $\lambda_\omega(t) \to \lambda$ as $\|\boldsymbol{\varepsilon}_v(t)\| \to 0$.

Two remarks regarding this control law are in order.

Remark 3.7: It is important to note that the time-varying control law (27) is constructed without requiring any metric information of the observed object, regardless of its shape and of the camera motion.

Remark 3.8: The proof that the time-varying control law (27) ensures asymptotic stability of the corresponding closed-loop system is essentially the same as in Theorem 3.2. Furthermore, the planned control error to be regulated to zero, i.e., $\boldsymbol{\varepsilon}^r(t)$ in (27), is usually much smaller than the current $\boldsymbol{\varepsilon}(t)$ in (20) and (22). This means that local operation is even more respected in the case of path planning.

IV. Direct Visual Servoing: Estimation Aspects

This section presents a direct image registration technique to simultaneously estimate the geometric and photometric parameters that relates the current image with the reference one. The geometric parameters are used to construct the proposed nonmetric control error (see Section III), whereas the photometric ones are estimated to achieve effective robustness to general illumination changes.

A. Photometric and Geometric Transformation Models

The general relation (3) allows us to define a geometric transformation model, which is also referred to as warping operator

$$\mathbf{w} : \mathbb{SA}(3) \times \mathbb{R} \times \mathbb{P}^2 \to \mathbb{P}^2; \quad (\mathbf{g}, \mathbf{p}^*) \mapsto \mathbf{p} = \mathbf{w}(\mathbf{g}, \mathbf{p}^*), \qquad (29)$$

where $\mathbf{g} = \{\mathbf{G}, \mathbf{e}, \rho^*\}$, and the Lie group $\mathbb{SA}(3)$ is homeomorphic to $\mathbb{SL}(3) \times \mathbb{R}^3$. This geometric transformation model (29) is general in the sense that it captures both planar and nonplanar objects, as well as any type of camera motion. Furthermore, it is fully defined in the projective space.

With respect to the photometric relation, the current image $\boldsymbol{\mathcal{I}}$ can be transformed such that the resulting $\boldsymbol{\mathcal{I}}'_h$ best matches the reference one $\boldsymbol{\mathcal{I}}^*$ through the general model [10]

$$\boldsymbol{\mathcal{I}}'_h(\mathbf{h}, \boldsymbol{\mathcal{I}}) \, = \, \boldsymbol{\mathcal{S}} \bullet \boldsymbol{\mathcal{I}} + \boldsymbol{\beta}, \qquad (30)$$

where $\mathbf{h} = \{\boldsymbol{\mathcal{S}}, \boldsymbol{\beta}\}$, the set of surfaces $\boldsymbol{\mathcal{S}} = [\mathcal{S}_{ij}]_{i,j=1}^{n}$ captures both local and global lighting variations, $\boldsymbol{\beta} = [\beta_1 \mathbf{1}, \beta_2 \mathbf{1}, \ldots, \beta_n \mathbf{1}]^\top$ models the per-channel shift in the ambient lighting changes and in the camera bias, and the operator "$\bullet$" represents the linear combination of the color channels elementwise multiplied by the corresponding surface. This photometric transformation model (30) is, thus, general in the sense that it compensates for both global and local illumination changes, even in color images.

B. Photogeometric Direct Image Registration

This problem consists in searching for the photogeometric parameters $\mathbf{g} = \{\mathbf{G}, \mathbf{e}, \rho^*\}$ and $\mathbf{h} = \{\boldsymbol{\mathcal{S}}, \boldsymbol{\beta}\}$ that best transform the current image such that each pixel intensity $\boldsymbol{\mathcal{I}}(\mathbf{p})$ is matched as closely as possible to the corresponding one in $\boldsymbol{\mathcal{I}}^*(\mathbf{p}^*)$. A photogeometric transformation model can be defined from the model of illumination changes (30), along with the warping model (29). More formally, the action of this general transformation model on pixels is given by

$$\boldsymbol{\mathcal{I}}'_{gh}(\mathbf{g}, \mathbf{h}, \mathbf{p}^*) \, = \, \boldsymbol{\mathcal{S}}(\mathbf{p}^*) \bullet \boldsymbol{\mathcal{I}}\big(\mathbf{w}(\mathbf{g}, \mathbf{p}^*)\big) + \boldsymbol{\beta} \quad \geq \mathbf{0}. \qquad (31)$$

A photogeometric direct image registration system can then be formulated as the following nonlinear optimization problem:

$$\min_{\substack{\mathbf{g} = \{\mathbf{G}, \mathbf{e}, \rho_i^*\} \\ \mathbf{h} = \{\boldsymbol{\mathcal{S}}, \boldsymbol{\beta}\}}} \; \frac{1}{2} \sum_{\mathbf{p}_i^* \in \boldsymbol{\mathcal{T}}^*} \Big[\, \boldsymbol{\mathcal{I}}'_{gh}(\mathbf{g}, \mathbf{h}, \mathbf{p}_i^*) - \boldsymbol{\mathcal{I}}^*(\mathbf{p}_i^*) \,\Big]^2, \qquad (32)$$

which seeks to minimize the vector of all image differences within the reference template $\boldsymbol{\mathcal{T}}^* \subseteq \boldsymbol{\mathcal{I}}^*$. Of course, the cost function can be different, but the sum of square differences is the most widely used one to register images without aberrant measures (e.g., unknown occlusions). If these measures are present, a robust function [17] (e.g., an M-estimator) can be considered in (32). The optimization problem (32) can be solved by standard iterative nonlinear methods such as the Gauss–Newton one. For an improved solution in terms of convergence properties, see [10].

C. Initialization Issues

The estimation system (32) exploits the intensity of all pixels within an image region of interest $\boldsymbol{\mathcal{T}}^*$. Although the entire image could be used, i.e., $\boldsymbol{\mathcal{T}}^* = \boldsymbol{\mathcal{I}}^*$, the quantity of information to be exploited strongly depends on the available computing resources. As an example, for an estimation of ten parameters at 30 Hz using the technique in [10], only 30 000 pixels should be exploited if only a monocore Pentium 4 3.2 GHz is available. Indeed, in this case, that computer runs at 6 ms/iteration (i.e., ≈30 Hz using five iterations per image). Although different strategies can be applied to select the pixels, for the sake of simplicity, the results in Section V are obtained by exploiting all pixels within a manually selected region of $\boldsymbol{\mathcal{I}}^*$ (outlined in blue). Furthermore, the parameters estimated in the registration of $\boldsymbol{\mathcal{I}}^*$ with $\boldsymbol{\mathcal{I}}(t)$ are used as a starting point to align $\boldsymbol{\mathcal{I}}^*$ with $\boldsymbol{\mathcal{I}}(t + \Delta t)$, where t indexes the images, and Δt is the sampling period. Nevertheless, if there is no sufficient overlapping of the object between images, a local minimum can be attained. In this case, a predictor or a global minimization procedure can be used. The former can also be applied to speed up the estimation, whereas the latter is of special interest for the initial image due to a possibly very large displacement between the initial and desired poses.

V. Experimental Results

This section reports typical results from the proposed visual servoing technique. The considered task consists of positioning a camera-mounted holonomic system with respect to a rigid object, regardless of its shape. To this end, a reference image is acquired at the reference pose. After displacing the camera to another pose, which is also called initial one, the control objective is to stabilize it at that reference situation. The control error and control law are both calculated at the signal level as described in Section III, i.e., they do not use either metric information of the object or image feature extraction. The needed projective parameters are obtained using the photogeometric direct image registration method described in Section IV.

A. Synthetic Data

To obtain a ground truth, we constructed synthetic objects of different shapes. We also mapped textured images onto them to simulate realistic situations as closely as possible. The control system is modeled as a pinhole camera mounted on the end-effector of a classical manipulator robot. The motion of the control point and the projection of its planned path are shown in the images as blue and green marks, respectively. The latter is typically composed of $T = 100$ waypoints, and the variable gain matrix (28) uses $\lambda = \gamma = 10$. A challenging scenario is then set up: The object is an hyperbolic paraboloid (*a priori* unknown), i.e., the horse's saddle, whose center is placed 100 cm away from the camera; the focal lengths are set very differently from the true ones, i.e., instead of $\alpha_u = \alpha_v = 500$ pixels, we set $\widehat{\alpha}_u = 900$ and $\widehat{\alpha}_v = 800$ pixels; as well as a large initial displacement (relatively to the scene depths) is carried out, i.e., a translation of $[-20.95, \; 21.77, \; -58.94]$ cm (norm: 66.2 cm) and a rotation of $[12.00, \; -9.60, \; 30.00]^\circ$ (norm: 33.7°) relatively to the reference frame. The proposed technique successfully performs the visual servo, despite all of these large perturbations. See Fig. 2 for the results. Final errors less than 1 mm in translation and less than 0.1° in rotation are obtained. In addition, observe that the Cartesian errors converge to zero smoothly and are nearly decoupled. These results show that the technique 1) can be highly accurate, 2) is robust to large errors in the camera intrinsic parameters, 3) possesses a large domain of convergence, and 4) works with fully nonplanar objects, despite large initial translation with respect to the scene depths. See [13] and [14] for results with different objects, varying illumination conditions, and color cameras.

B. Real Data

This section presents the servoing results using a unicycle-type mobile robot. We mounted a pan-tilt camera away from its rear wheel axle so that the mounting point is not subjected to the nonholonomic constraint. In other terms, the resulting control system is holonomic (see the demonstration in [18]). The location of that mounting point relative to that axle is only roughly known. We assumed it as 8 cm away

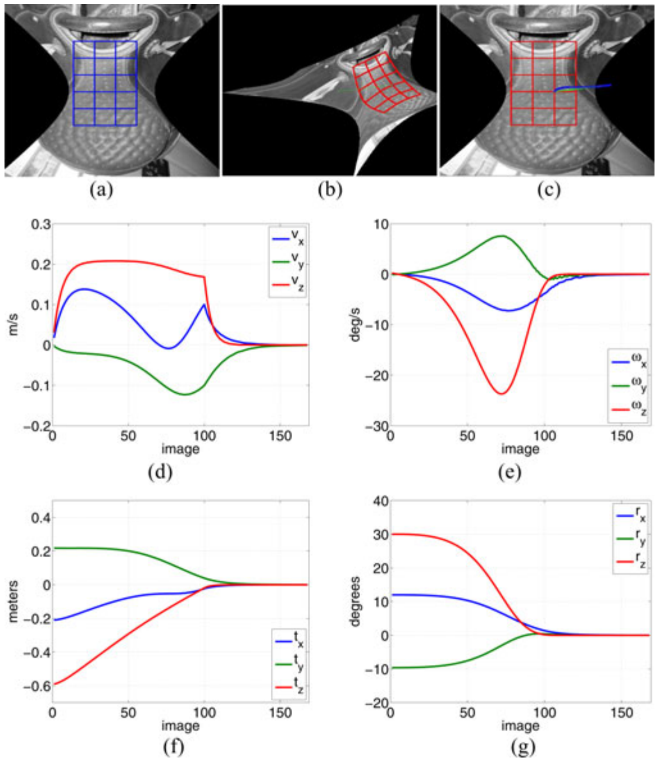

Fig. 2. Direct visual servoing with respect to a hyperbolic paraboloid (*a priori* unknown) using a coarsely calibrated camera. Final errors less than 1 mm in translation and less than 0.1° in rotation are obtained. (a) Reference image. (b) Initial image. (c) Final image. (d) Translation velocity. (e) Rotation velocity. (f) Translation error. (g) Rotation error.

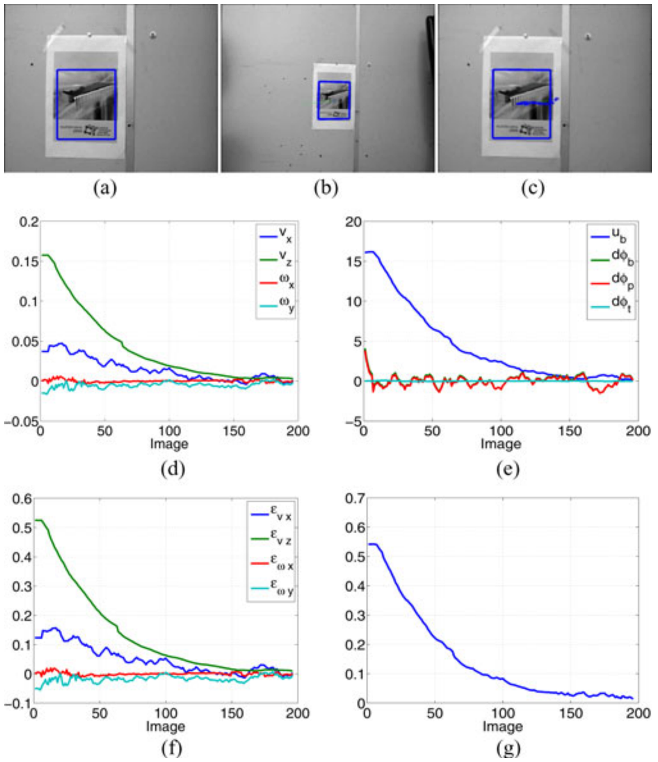

Fig. 3. Direct visual servoing using a camera-mounted robot. Both camera and robot are only coarsely calibrated. The RMS error between the reference template and the final one is of 9.68 levels of grayscale (over 256). (a) Reference image. (b) Initial image. (c) Final image. (d) Camera velocities. (d) Robot velocities. (f) Control error. (g) Norm of the control error.

from that axle and in the middle of it, which is of approximately 40 cm long. The camera calibration parameters are also coarsely known. We assumed them as $\widehat{\alpha}_u = \widehat{\alpha}_v = 440$ pixels, zero skew, and the principal point in the middle of the image, which has 320×240 pixels. No path planning is performed in this experiment and the control gains in (21) are set to $\lambda_v = \lambda_\omega = 0.3$. The results are shown in Fig. 3. Differently from the synthetic case, no ground truth is available here. However, the accuracy can be verified through the RMS error between the reference template and the final one. In this case, it is of 9.68 levels of grayscale (over 256), which corresponds to a termination condition for the servoing of $\|\boldsymbol{\varepsilon}_v\| < 0.01$. See [10] for numerous other experimental results concerning the photogeometric aspects of the estimation. In particular, one can observe in [10] that the robustness of the system to arbitrary illumination changes, even in color images.

VI. Conclusion

This paper has proposed a new approach to visual servoing all 6 DOF of a holonomic robot, given a reference image. The approach does not require any metric information of the observed object, regardless of its shape and of the camera motion. In this way, system versatility and robustness to errors in the calibration parameters are both achieved. The used projective parameters are obtained via a photogeometric registration method that can exploit all image information, even from areas where no image feature exists. Therefore, high levels of accuracy and robustness to illumination changes, even in color images, can both be attained. Finally, the proposed nonmetric control error allows for path planning. Hence, a large domain of convergence is also obtained.

In spite of all these improvements, several problems and analysis issues still remain open. A possible analysis to be conducted within the proposed strategy concerns its robustness to errors in the camera intrinsic parameters. Although a large degree of robustness has been observed in the experiments, no theoretical analysis has been performed. Additionally, a promising research direction concerns the generalization of the proposed nonmetric technique to critical nonlinear systems, such as nonholonomic and underactuated robots. This will certainly be an active topic of research in the near future.

References

[1] F. Chaumette and S. Hutchinson, “Visual servo control part I: Basic approaches,” *IEEE Robot. Autom. Mag.*, vol. 13, no. 4, pp. 82–90, Dec. 2006.

[2] L. E. Weiss and A. C. Anderson, “Dynamic sensor-based control of robots with visual feedback,” *IEEE J. Robot. Autom.*, vol. RA-3, no. 5, pp. 404–417, Oct. 1987.

[3] E. Malis, Y. Mezouar, and P. Rives, “Robustness of image-based visual servoing with a calibrated camera in the presence of uncertainties in the three-dimensional structure,” *IEEE Trans. Robot.*, vol. 26, no. 1, pp. 112–120, Feb. 2010.

[4] E. Malis and F. Chaumette, “Theoretical improvements in the stability analysis of a new class of model-free visual servoing methods,” *IEEE Trans. Robot. Autom.*, vol. 18, no. 2, pp. 176–186, Apr. 2002.

[5] O. Faugeras, Q.-T. Luong, and T. Papadopoulo, *The Geometry of Multiple Images*. Cambridge, MA: MIT Press, 2001.

[6] L. Thaler and M. A. Goodale, “Beyond distance and direction: The brain represents target locations non-metrically,” *J. Vis.*, vol. 10, no. 3, pp. 1–27, 2010.

[7] R. Basri, E. Rivlin, and I. Shimshoni, “Visual homing: Surfing on the epipoles,” *Int. J. Comput. Vis.*, vol. 33, pp. 22–39, 1999.

[8] S. Benhimane and E. Malis, "Homography-based 2D visual servoing," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2006, pp. 2397–2402.

[9] M. Kazemi, K. Gupta, and M. Mehrandezh, "Path-planning for visual servoing: A review and issues," in *Visual Servoing via Advanced Numerical Methods*. (ser. LNIS), vol. 401, New York: Springer, 2010, pp. 189–207.

[10] G. Silveira and E. Malis, "Unified direct visual tracking of rigid and deformable surfaces under generic illumination changes in grayscale and color images," *Int. J. Comput. Vis.*, vol. 89, pp. 84–105, 2010.

[11] V. Kallem, M. Dewan, J. Swensen, G. Hager, and N. Cowan, "Kernel-based visual servoing," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2007, pp. 1975–1980.

[12] C. Collewet, E. Marchand, and F. Chaumette, "Visual servoing set free from image processing," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2008, pp. 81–86.

[13] G. Silveira and E. Malis, "Direct visual servoing with respect to rigid objects," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2007, pp. 1963–1968.

[14] G. Silveira and E. Malis, "Visual servoing from robust direct color image registration," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2009, pp. 5450–5455.

[15] F. W. Warner, *Foundations of Differential Manifolds and Lie Groups*. New York: Springer-Verlag, 1987.

[16] G. Silveira and E. Malis, "Direct visual servoing with respect to rigid objects," French Nat. Inst. Res. Comput. Sci. Control, Sophia-Antipolis, France, Research Rep. 6265, 2007.

[17] P. J. Huber, *Robust Statistics*. Hoboken, NJ: Wiley, 1981.

[18] P. Morin and C. Samson, "Motion control of wheeled mobile robots," in *Springer Handbook of Robotics*. New York: Springer-Verlag, 2008, pp. 799–826.

An Alternative to the Mahalanobis Distance for Determining Optimal Correspondences in Data Association

Jose-Luis Blanco, Javier González-Jiménez,
and Juan-Antonio Fernández-Madrigal

***Abstract*—The most common criteria to determine data association rely on minimizing the squared Mahalanobis distance (SMD) between observations and predictions. We hold that the SMD is just a heuristic, while the alternative *matching likelihood* is the optimal statistic to be maximized. Thorough experiments undoubtedly confirm this idea, with false positive reductions of up to 16%.**

***Index Terms*—Data association (DA), simultaneous localization and mapping (SLAM).**

Manuscript received June 18, 2011; revised January 12, 2012; accepted March 30, 2012. Date of publication April 25, 2012; date of current version August 2, 2012. This paper was recommended for publication by Associate Editor E. Marchand and Editor D. Fox upon evaluation of the reviewers' comments. This work was supported in part by the Spanish Government under research contracts DPI2008-03527 and DPI2011-25483.

The authors are with the Department of System Engineering and Automation, University of Málaga, 29071 Malaga, Spain (e-mail: joseluisblancoc@gmail.com; jgonzalez@ctima.uma.es; jafma@ctima.uma.es).

The supplementary file of size 17.5 MB contains documentation in HTML format which can be seen with any standard browser, as well as images in standard formats (JPEG, PNG). It also contains scripts which require MATLAB/Octave, the bash shell, and some tools from MRPT. Detailed instructions are given for each experiment.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TRO.2012.2193706

I. Introduction

The problem of simultaneous localization and mapping (SLAM) has been one of the most studied topics in mobile robotics during the past decade [1]–[4]. Maps of the robot environment learned by SLAM can be roughly classified as either continuous representations (such as occupancy grid maps [5], elevation maps [6], or gas concentration maps [7]) or discrete (object-based) representations, comprised of a variable number of natural or artificial *landmarks* present in the environment. In this paper, we focus on this latter approach. Using a map of discrete elements has some advantages such as the relative efficiency of graph SLAM [8] and extended Kalman filter (EKF)-like [9] algorithms in comparison with alternatives for continuous maps—e.g., grid mapping with particle filters [10].

However, discrete maps introduce two hurdles: First, in most cases, sensors do not directly detect landmarks; thus, an additional *detection* step must be introduced whose failure would severely degrade the overall mapping performance. Second, once a set of observed landmarks is available from the sensor, they must be paired with those already in the map. This is the *data-association* (DA) problem, which is the central concern of this paper.

The DA problem can be stated as follows: At some time step t, and given the vector with N landmark observations $\mathbf{z}_t$, compute the N-length association vector $\mathbf{n}_t$ which states to which map landmark does each observation correspond (or whether it is a new landmark not observed earlier). Each landmark observation is a point in the *observation space* whose dimensionality depends on the specific problem, e.g., 2-D in planar range-bearing SLAM with point features [4], [11] and in monocular SLAM [12], or 1-D in range-only SLAM [13], [14]. Each of these observation points must be paired with either one or none of a set of *predictions* or expected observation for each known landmark in the map. Given that the sensor model is stochastic and both the vehicle pose and the map are represented as probability densities, these predictions are probability distributions as well—typically, Gaussians.

As we will discuss in Section II, the most popular methods to solve DA are the nearest neighbor (NN) [4] and the joint-compatibility branch and bound (JCBB) [11] algorithms. As described in the literature, these methods aim to establish the most likely pairings by minimizing the squared Mahalanobis distance (SMD) between the observations and their associated predictions.

The central claim of this paper is that minimizing the SMD does not always lead to the most likely pairings, as can be easily demonstrated. Consider the probability mass function over all the possible associations $\mathbf{n}_t$ for a time step t, given the knowledge about the joint vehicle-map state vector $\mathbf{s}_t$ and the latest observation $\mathbf{z}_t$, which follows the conditional distribution $P(\mathbf{n}_t|\mathbf{s}_t, \mathbf{z}_t)$. By definition, the most likely set of associations is the value of $\mathbf{n}_t$ that maximizes this distribution. Applying the Bayes rule over the observation $\mathbf{z}_t$

$$\begin{aligned} P(\mathbf{n}_t|\mathbf{s}_t, \mathbf{z}_t) &\propto \cancelto{\eta}{P(\mathbf{n}_t|\mathbf{s}_t)} p(\mathbf{z}_t|\mathbf{s}_t, \mathbf{n}_t) \\ &\propto p(\mathbf{z}_t|\mathbf{s}_t, \mathbf{n}_t) \end{aligned} \tag{1}$$

using the fact that the *a priori* distribution of the associations not conditioned to any observation must be uniform, leading to an irrelevant constant term η. A natural and expected consequence of the aforementioned equation is that optimal correspondences are those that "best explain" the observations.

If we let $\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the evaluation at $\boldsymbol{x}$ of the probability density function of a multivariate Gaussian with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, the observation likelihood in EKF-based or graph SLAM can be denoted as

$$p(\mathbf{z}_t|\mathbf{s}_t, \mathbf{n}_t) = \mathcal{N}(\mathbf{z}_t; h(\mathbf{s}_t, \mathbf{n}_t), \mathbf{S}(\mathbf{n}_t)) \tag{2}$$