Robotics and Autonomous Systems 58 (2010) 796–808

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Omnidirectional visual control of mobile robots based on the 1D trifocal tensor

H.M. Becerra *, G. López-Nicolás, C. Sagüés

Dept. Informática e Ingeniería de Sistemas - Instituto de Investigación en Ingeniería de Aragón, Universidad de Zaragoza, 50018 Zaragoza, Spain

ARTICLE INFO

Article history:
Available online 4 March 2010

Keywords:
Visual control
Omnidirectional images
1D trifocal tensor
Mobile robots

ABSTRACT

The precise positioning of robotic systems is of great interest particularly in mobile robots. In this context, the use of omnidirectional vision provides many advantages thanks to its wide field of view. This paper presents an image-based visual control to drive a mobile robot to a desired location, which is specified by a target image previously acquired. It exploits the properties of omnidirectional images to preserve the bearing information by using a 1D trifocal tensor. The main contribution of the paper is that the elements of the tensor are introduced directly in the control law and neither any a priori knowledge of the scene nor any auxiliary image are required. Our approach can be applied with any visual sensor obeying approximately a central projection model, presents good robustness to image noise, and avoids the problem of a short baseline by exploiting the information of three views. A sliding mode control law in a square system ensures stability and robustness for the closed loop. The good performance of the control system is proven via simulations and real world experiments with a hypercatadioptric imaging system.

1. Introduction

Visual servoing is an interesting research field that involves computer vision and control theory in order to command the robot motion. In particular, visual servoing allows mobile robots to improve their navigation capabilities in a single robot task or in cooperative tasks. This paper describes an approach to drive a wheeled mobile robot equipped with an omnidirectional camera on board to a desired location, which is specified by a target image previously acquired, i.e., using a teach-by-showing strategy. In visual servoing, a robot is steered to a desired location by minimizing an error function that relates visual data, typically from two images: the current and the target one. We propose to take advantage of more information by using three views.

Along the years, the research on visual servoing has dedicated important efforts to find suitable error functions in order to obtain a desired behavior of the robotic system in terms of stability and robustness of the closed loop control. The basic approaches are typically separated in image-based visual servoing (IBVS), in which the error function is built from the features available in the image data, and position-based visual servoing (PBVS), in which a set of 3D parameters must be estimated from image measurements [1]. Subsequently, many approaches have been proposed as hybrid schemes [2]. Among the advanced approaches, some schemes that are based on a geometric constraint can be found. A geometric constraint is a robust way to relate features that are observed in different views of the same scene. Nowadays, two geometric constraints have been exploited for the control of mobile robots: epipolar geometry and the homography model. Examples of epipolar visual control are [3,4]. In these works, the epipoles are directly used to compute the control inputs for the robot. The homography model has been used in several visual servoing schemes, for instance [5,6]. In the last, the elements of the homography matrix are used directly in a control law for mobile robots. However, it is known that both of these geometric constraints have serious drawbacks. Epipolar geometry is ill-conditioned with a short baseline and with planar scenes. The homography model is not well defined if there are no dominant planes in the scene.

Most of the visual servoing schemes have the drawback that the target may leave the camera's field of view during the servoing, which leads to failure because the feedback error cannot be computed any more. In this context, the use of wide field of view cameras becomes a very good option to overcome this issue, although some strategies have been proposed for conventional cameras, for instance [7]. One effective way to enhance the field of view is to use mirrors in conjunction with lenses, i.e. catadioptric image formation systems. Some of the pioneer works proposing the use of catadioptric cameras for visual servoing are [8,9]. The approach of exploiting a geometric constraint has been also explored in omnidirectional visual servoing using the epipolar geometry [10] and the homography model [11]. Additionally, there are different approaches for omnidirectional vision-based robot navigation that exploit particular properties of omnidirectional images, for instance [12,13]. More related work on omnidirectional vision can be found in publications of the Omnivis workshop.

* Corresponding author. Tel.: +34 976842472; fax: +34 976761914.
E-mail addresses: hector.becerra@unizar.es (H.M. Becerra), gonlopez@unizar.es (G. López-Nicolás), csagues@unizar.es (C. Sagüés).

0921-8890/$ – see front matter
doi:10.1016/j.robot.2010.02.011

In order to overcome the drawbacks of the typical geometric constraints, we propose a novel approach based on the 1D trifocal tensor (TT) which allows us to exploit the information given by three available images: the initial, the current and the target ones. The 1D TT completely describes the relative geometry of three views and is independent of the observed scene [14]. The first work that proposes a robotic application of a trilinear constraint is [15]. In that work, conventional perspective cameras are converted to 1D virtual cameras through a transformation of bearing measurements. In the context of computer vision, the same idea is introduced to wide-angle cameras as a tool for calibrating the radial distortion in [16]. The same authors present a general hybrid trifocal constraint by representing conventional and omnidirectional cameras as radial 1D cameras in [17]. They assert that the radial 1D camera model is sufficiently general to represent the great majority of omnidirectional cameras under the assumption of knowing the center of radial distortion. The effectiveness of applying the 1D TT to recover the location information has been also proved in [18]. It uses the TT with both conventional and omnidirectional cameras for scene reconstruction, and proposes this approach for the initialization of bearing-only SLAM algorithms. The radial TT has been also proposed for hierarchical localization exploiting omnidirectional images in [19]. A recent work presents a visual control for mobile robots based on the elements of a 2D trifocal tensor constrained to planar motion [20].

We propose in this paper an image-based approach to perform visual servoing for mobile robots. The visual control is performed using the value of the elements of the 1D TT directly in the control law. The approach is suitable for all central catadioptric cameras and even for fisheye cameras, since all of these imaging systems present a high radial distortion but they preserve the bearing information, which is the only required data in our approach. This paper is an extension of [21], where a visual control based on the 1D TT obtained from metric information is introduced for conventional cameras. However, since there is a constrained field of view with conventional cameras, it is better to use omnidirectional images for this approach. The extension presented here is justified along the paper and as part of the results, realistic simulations with synthetic images, which are generated using the unified model for central catadioptric cameras [22], are reported. We have tested the robustness of the control law under image noise and the general performance is also analyzed through real world experiments with images of a hypercatadioptric system. The approach does not require any a priori knowledge of the scene and does not need any auxiliary image. We propose a two-step control law, the first step performs position correction and the second one corrects the orientation. Our approach ensures the total correction of the robot's pose even for initial locations where epipolar geometry or homography based approaches fail. In comparison with a typical IBVS approach, the proposed scheme allows us to prove the stability of the closed loop on the basis of a square control system. Additionally, from a control theory point of view, we have incorporated robustness properties in the system by using sliding mode control. The field of application of this approach is referred to differential-drive robots constrained to planar motion. It results in a great interest in many areas, specially for service robots, for which our approach could be applied for navigation together with a SLAM scheme like the ones in [23,24].

The paper is organized as follows. Section 2 specifies the mathematical modeling of the mobile robot and the 1D TT geometric constraint. Section 3 details the design procedure of the control law. Section 4 presents the stability analysis. Section 5 shows the performance of the control system via simulations with synthetic images, experimental analysis with real images and real world experiments in a closed loop. Finally, Section 6 provides the conclusions.

2. Mathematical modeling

2.1. Robot and camera modeling

This work focuses on controlling a wheeled mobile robot through the information given by an omnidirectional imaging system mounted onboard as shown in Fig. 1(a) and under the framework that is depicted in Fig. 1(b). The kinematic model of this kind of robot as expressed in state space is the following

$$\begin{bmatrix}\dot{x}\\ \dot{y}\\ \dot{\phi}\end{bmatrix} = \begin{bmatrix}-\sin\phi & 0\\ \cos\phi & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}\upsilon\\ \omega\end{bmatrix}. \tag{1}$$

Thus, $\mathbf{x} = (x, y, \phi)^T$ represents the state of the robotic platform, where x and y are the coordinates of the robot's position in the plane and ϕ is the robot's orientation. Additionally, υ and ω are the translational and rotational input velocities, respectively. From now on, we use the notation $s\beta = \sin\beta, c\beta = \cos\beta$. Note that the model (1) also describes the camera's motion, because of the fact that the robot's frame is defined in such a way that the optical axis coincides with the rotation axis of the robot, i.e. the camera is looking upwards. Fig. 1(c) shows a detailed view of our omnidirectional imaging system used in real experiments. It is a hypercatadioptric system, for which we can assume that there exists a unique effective viewpoint (Fig. 1(d)). This optical arrangement is popular for robotic applications because it is constructed just with a perspective camera and a hyperboloidal mirror. These systems, as well as those using paraboloidal and ellipsoidal mirrors, have been well studied in the field of computer vision [25], and according to this theory, all of them satisfy the fixed view point constraint. In practice, with a careful construction of the system, it is realistic to assume a central configuration and many robotic applications have proven its effectiveness [8–10,18].

2.2. The 1D trifocal tensor

Our approach is based on a direct feedback of the information given by a geometric constraint, the 1D trifocal tensor. A similar idea has been exploited for mobile robots control by using the epipolar geometry relating omnidirectional views [10]. The fundamental epipolar constraint is analogue for conventional perspective as that for central catadioptric cameras if it is formulated in terms of rays which emanate from the effective viewpoint [26]. In a similar way, the 1D TT estimation is basically the same for conventional and central catadioptric cameras. The 1D TT particularly adapts to the property of omnidirectional images to preserve bearing information regardless of the high radial distortion induced by lenses and mirrors. Fig. 1(d) shows the bearing angle of an observed feature, which is measured with respect to a frame centered in the principal point of the image. Thus, a bearing measurement θ can be converted to its projective formulation in a 1D virtual retina as $\mathbf{m} = (\sin\theta, \cos\theta)^T$. By relating this representation for three different views of a feature that is expressed in a 2D projective space, it results in a trifocal constraint

$$\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2} T_{ijk}\mathrm{u}_i\mathrm{v}_j\mathrm{w}_k = 0, \tag{2}$$

where $\mathbf{u} = (\mathrm{u}_1, \mathrm{u}_2)^T, \mathbf{v} = (\mathrm{v}_1, \mathrm{v}_2)^T$ and $\mathbf{w} = (\mathrm{w}_1, \mathrm{w}_2)^T$ are the image coordinates of a feature projected in the 1D virtual retina of the first, second and third cameras respectively, and T_{ijk} are the eight elements of the homogeneous trifocal tensor. The described representation of bearing measurements is sufficiently general to model from pin-hole cameras to omnidirectional ones,