Control Engineering Practice 46 (2016) 59–65

Contents lists available at ScienceDirect

Control Engineering Practice

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

An adaptive observer framework for accurate feature depth estimation using an uncalibrated monocular camera

Jishnu Keshavan*, Hector Escobar-Alvarez, J. Sean Humbert

Autonomous Vehicles Laboratory, University of Maryland, College Park 20742, USA

A R T I C L E I N F O

Article history:
Received 16 June 2015
Received in revised form
17 September 2015
Accepted 12 October 2015
Available online 26 October 2015

Keywords:
Feature depth estimation
Uncalibrated camera
Optic flow
Focal length estimation
Lyapunov analysis
Adaptive observer

A B S T R A C T

This paper presents a novel solution to the problem of depth estimation using a monocular camera undergoing known motion. Such problems arise in machine vision where the position of an object moving in three-dimensional space has to be identified by tracking motion of its projected feature on the two-dimensional image plane. The camera is assumed to be uncalibrated, and an adaptive observer yielding asymptotic estimates of focal length and feature depth is developed that precludes prior knowledge of scene geometry and is simpler than alternative designs. Experimental results using real camera imagery are obtained with the current scheme as well as the extended Kalman filter, and performance of the proposed observer is shown to be better than the extended Kalman filter-based framework.

Published by Elsevier Ltd.

1. Introduction

A classical problem in machine vision is the determination of the three-dimensional position of an object from observation of motion of its projected feature on the two-dimensional image plane of a charge coupled-device (CCD) camera. The case of a calibrated camera undergoing known motion has received particular attention (Chen & Kano, 2002; Dixon, Fang, Dawson, & Flynn, 2003; Jankovic & Ghosh, 1995; Ma, Chen, & Moore, 2004), where the problem of determining object coordinates is reduced to the problem of estimating object depth. Alternatively, the problem of estimating the motion parameters from measurement of the three-dimensional coordinates has also been considered (Soatto, Frezza, & Perona, 1996; Tsai & Huang, 1981).

The traditional approach to the problem of depth estimation involves the identification and tracking of a set of distinctive features across consecutive image frames (Azarbayejani & Pentland, 1995; Broida, Chandrashekhar, & Chellappa, 1990). The observed features are usually classified as points, lines or curves in the Euclidean three-dimensional space, and depth estimation is typically accomplished assuming ideal perspective projection (Azarbayejani & Pentland, 1995) – where simple similarity rules are used to project the three-dimensional coordinates of the scene onto the two-dimensional image plane – and the use of the extended Kalman filter (EKF) (Ayache & Faugeras, 1989; Broida & Chellappa, 1986; Chiuso, Favaro, Jin, & Soatto, 2002; Durrant-Whyte & Bailey, 2006; Davison, Reid, Molton, & Stasse, 2007; Dickmanns & Graefe, 1988; Faugeras, Ayache, & Faverjon, 1986; Matthies, Kanade, & Szeliski, 1989; Sridhar, Soursa, & Hussein, 1993; Young & Chellappa, 1990).

Extended Kalman filtering provides a relatively computationally intensive paradigm for the estimation of feature depth. A key step in filter implementation requires the linearization of process and measurements models about the mean and covariance of the current depth estimate, which causes it to suffer from the drawback of lacking a complete guarantee of convergence leading to failure in some cases (Chen & Kano, 2002), especially if the process and measurement models are incorrect, or the degree of nonlinearity is high, or if the initial depth estimates are incorrect. Furthermore, the accuracy of EKF-based depth estimates is contingent on an accurate a priori description of the noise model (Chen & Kano, 2002). Thus, the convergence conditions for the use of the EKF are application-specific, and can only be checked by filter implementation (Reif, Sonnemann, & Unbehauen, 1998).

For given observer motion, the dynamics of depth variation is nonlinear, with the rate of change of feature depth being a quadratic function of the current depth estimate (Jankovic & Ghosh, 1995). Thus, numerous studies have naturally attempted to develop depth estimation techniques that rely on nonlinear stability analysis to guarantee convergence (Abdursal, Inaba, & Ghosh,

* Corresponding author.
E-mail addresses: jishnuk@umd.edu (J. Keshavan), hescobar@umd.edu (H. Escobar-Alvarez), humbert@umd.edu (J. Sean Humbert).

http://dx.doi.org/10.1016/j.conengprac.2015.10.005
0967-0661/Published by Elsevier Ltd.

2004; Chen & Kano, 2002, 2004; Dahl, Nyberg, & Heyden, 2007; Dixon et al., 2003; Gupta, Aiken, Hu, & Dixon, 2006; Hu & Ersson, 2004; Jankovic & Ghosh, 1995; Karagiannis & Astol, 2005; Luca, Oriolo, & Giordano, 2007; Ma et al., 2004). All these studies on range identification are based on a calibrated camera undergoing known motion dynamics, and observer performance has typically been demonstrated only in simulation (Tsai & Huang, 1981). Thus, there exists a need for an evaluation of their performance on real world systems, especially as some observers are known to deliver unsatisfactory performance in the presence of noisy velocity measurements (Nath, Braganza, Dawson, & Burg, 2009). Finally, an important limitation of some earlier studies pertains to the requirement of imposing constraints on camera motion (Chaumette, Boukir, Bouthemy, & Juvin, 1996; Conroy & Humbert, 2013; Matthies et al., 1989; Smith & Papanikolopoulos, 1994) accurate depth estimation is typically accomplished only by limiting movement to the camera's image plane.

Optic flow is the apparent motion of object features in the scene as projected on the two-dimensional image plane. This work considers a pinhole camera model that assumes ideal perspective projection to derive equations for optic flow. It is seen that optic flow is a highly nonlinear function of the camera's focal length, rendering EKF-based estimation schemes vulnerable to divergence. To overcome the drawbacks associated with the use of the EKF, a solution to the problem of range identification using an uncalibrated camera is proposed based on a novel adaptive observer design that achieves asymptotic convergence of the observation error to zero. The observer combines online linear and angular velocity estimates with optic flow measurements, obtained by tracking motion of numerous projected features on the image plane over successive frames, to generate accurate focal length and depth estimates in a recursive manner. This approach to depth estimation is validated with experimental results, and departs from most prior studies in the following important aspects. Motion sequences are processed without prior knowledge of the camera and scene geometry. Additionally, depth estimation is accomplished without imposing stringent constraints on camera motion. Furthermore, the current approach provides theoretical stability and convergence guarantee which is in contrast with depth estimation schemes based on the EKF. A performance comparison study with the EKF is undertaken, and the adaptive observer is seen to deliver superior performance compared to the EKF. Finally, it is important to note that the proposed observer is considerably simpler than the higher-order observers proposed in Chen and Kano (2002) and Dixon et al. (2003), as well as the high gain observer proposed in Jankovic and Ghosh (1995).

The rest of the paper is organized as follows. Section 2 deals with the formulation of the visual depth estimation problem and presents the mathematical model used in this study. Section 3 presents the Lyapunov stability analysis framework for the development of the adaptive observer that yields asymptotic estimates of focal length and feature depth. Section 4 presents the experimental setup, including details of camera motion and feature-tracking implementation, and the results using real camera imagery for experimental validation of the proposed approach. Issues of convergence for the uncalibrated camera through a performance comparison study with the EKF are also considered. Section 5 presents the conclusions of this study.

2. Problem definition

In this section, the mathematical model describing the motion of an object relative to the camera and the associated optic flow equations are presented. Assuming a stationary environment, for a camera undergoing motion with a translational velocity $\mathbf{v}$ and an angular velocity $\boldsymbol{\omega}$, the relative motion of an object feature $\mathbf{P}$ in three-dimensional space is given by

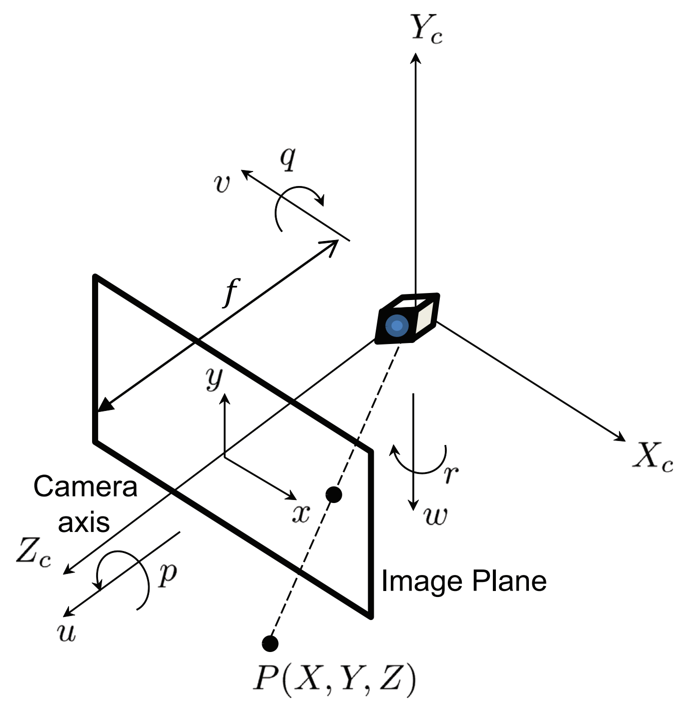

Fig. 1. Pinhole camera model with the linear and angular velocity components as shown.

$$\frac{d\mathbf{P}}{dt} = -\mathbf{v} - \boldsymbol{\omega} \times \mathbf{P} \tag{1}$$

If the translational and angular velocity components of the camera are given by $\mathbf{v} = \{u, v, w\}^T$, $\boldsymbol{\omega} = \{p, q, r\}^T$ (Fig. 1), object feature motion can be described by the affine system

$$\begin{bmatrix} \dot{X} \\ \dot{Y} \\ \dot{Z} \end{bmatrix} = \begin{bmatrix} v \\ w \\ -u \end{bmatrix} + \begin{bmatrix} Y & 0 & Z \\ -X & -Z & 0 \\ 0 & Y & -X \end{bmatrix} \begin{bmatrix} p \\ q \\ r \end{bmatrix} \tag{2}$$

where $(X, Y, Z) \in \mathbb{R}^3$ are the unmeasurable coordinates of the object feature in the camera's reference frame $\{X_c, Y_c, Z_c\}^T$, with Z_c being perpendicular to the camera's image plane, as shown in Fig. 1. The velocity components $\mathbf{v}$, $\boldsymbol{\omega}$ are in m/s and rad/s respectively, possibly time-varying and are assumed known.

Assuming perspective projection (pinhole) model for the camera, the image plane coordinates of the projected feature are given by the transformation π: $\mathbb{R}^3 \to \mathbb{R}^2$

$$[x/f, y/f]^T = \pi[X, Y, Z]^T \triangleq [X/Z, Y/Z]^T \tag{3}$$

Here, the object feature coordinates (X, Y, Z) are in m, and the image plane coordinates (x,y) as well as the focal length f are in pixels. The camera is assumed to be uncalibrated and the focal length unknown.

As in Jankovic and Ghosh (1995) and Chen and Kano (2002), introduce inverse depth $d = 1/Z$. Also, assume inverse focal length $\beta = 1/f$, as considered in Azarbayejani and Pentland (1995). Taking the derivatives of (x, y) with respect to time, and substituting in (2) leads to the well-known equations of optic flow (Matthies et al., 1989):