Computer Vision and Image Understanding 117 (2013) 603–619

Contents lists available at SciVerse ScienceDirect

Computer Vision and Image Understanding

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

Adjustable linear models for optic flow based obstacle avoidance ☆

Manuela Chessa *, Fabio Solari, Silvio P. Sabatini

Department of Informatics, Bioengineering, Robotics and System Engineering – DIBRIS, University of Genoa, Via all'Opera Pia 13, 16145 Genova, Italy

ARTICLE INFO

Article history:
Received 21 July 2011
Accepted 26 January 2013
Available online 5 February 2013

Keywords:
Motion interpretation
Affine description
Recursive filtering
Kalman filter
Time-to-contact
Surface orientation
Biologically inspired vision

ABSTRACT

An original framework to recover the first-order spatial description of the optic flow is proposed. The approach is based on recursive filtering, and uses a set of linear models that dynamically adjust their properties on the basis of context information. These models are inspired by the experimental evidence about motion analysis in biological systems. By checking the presence of these models in the optic flow through a multiple model Kalman Filter, it is possible to compute the coefficients of the affine description and to use this information for estimating the motion of the observer as well as the three-dimensional orientation of the surfaces in some points of interest in the scene. In order to systematically validate the approach, a set of benchmarking sequences is used, and, finally, the proposed algorithm is successfully applied in real-world automotive situations.

1. Introduction

The problem of visual motion interpretation has been a major concern in computer vision for many years and it remains one of the most challenging problems. Since the ability to move in a structured environment and to estimate impeding collisions is a vital requirement for any organism, both living and artificial, several multidisciplinary studies address the problem, ranging from behavioral studies in many different animal species, to collision avoidance approaches in automotive and robotics applications. To guide a vehicle in an environment we are interested in how it should detect the presence of other objects, both moving and still, that occupy its path. This is related to the movement of the vehicle itself, to the velocity of the other objects in the environment and to the three-dimensional (3D) structure of the environment [1]. In the literature, a wide variety of solutions to the problem of visual guidance have been proposed, based on different measurements and different estimation algorithms. The different approaches can either use sparse sets of corresponding feature points or dense optic flow estimates, then both linear and nonlinear techniques are applied to obtain estimates of the 3D parameters.

The main contributions of this paper, that extends and systematically validates the approach presented in [2], are:

– to recover a robust affine description of the optic flow by using a biologically inspired method, based on a set of adjustable templates that analyze the dense input optic flow in order to compute its linear approximation through a recursive approach based on a multiple-model Kalman Filter (KF) and
– to obtain reliable information about the 3D structure of the environment in terms of the orientation of the surfaces in some points of interest in the scene, such as the heading direction and the independently moving objects (IMOs), and to derive a measure of the nearness of objects in time, also known as time-to-contact (TTC), by combining the affine estimates.

The rest of the paper is organized as follows. We review the related works in Section 2. The affine models of image motion are introduced in Section 3. In Section 4 we describe our recursive approach to compute the affine description of the optic flow. The problem of motion interpretation for obstacle avoidance, by computing the TTC and the surface orientation, is addressed in Section 5. In Section 6 the proposed approach is validated with respect to a set of specifically designed benchmarking sequences, and to real-world situations. In Section 7 we compare the proposed method with respect to some state-of-the-art approaches. In Section 8 we draw the conclusions.

2. Related works

The original formulation of a computational theory of the interpretation of visual motion, caused by the projection of a moving surface, is described in [3,4]. In [5,6], the authors showed that a

☆ This paper has been recommended for acceptance by John K. Tsotsos.

* Corresponding author.
E-mail address: manuela.chessa@unige.it (M. Chessa).
URL: http://www.pspc.unige.it (M. Chessa).

1077-3142/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cviu.2013.01.012

closed form solution for the 3D motion and surface orientation can be obtained from the quadratic parameters of the motion field. A closed form solution using only first-order coefficients for two or more distinct planes was described in [7]. However, the main problem of these algorithms is that they are noise sensitive and consequently not sufficiently robust. A robust recursive algorithm for recovering structure from motion, able to provide the normal to the surfaces, as well as the 3D motion, has been proposed in [8]. In that paper, the author estimates the affine motion parameters from the image sequence and then uses an Extended Kalman Filter in order to integrate over time and to obtain robust estimates of the motion parameters. In a recent review [9], several methods to recover motion parameters (i.e. ego-motion) from image sequences are presented and compared.

Among the approaches that aim to solve the problem of motion interpretation, many different approaches also addressed the problem of the estimation of the TTC. They can be classified in the following groups: optic flow based approaches, approaches based on feature (points or lines) correspondences, and approaches based on closed contours.

Many optic flow based approaches work on first-order (affine) approximation of the motion field. In [10] the authors describe a computational scheme based on the computation of first-order spatial properties of optic flow. They applied the proposed approach to compute the TTC by integrating over a number of independent motion measurements, from a real-world controlled situation (natural scenes acquired in their laboratory). Their estimates can be affected by large errors and become point-wise unreliable, but more stable and accurate estimates are obtained by using a Kalman filtering procedure. Although the results presented by the authors prove the validity of the approach, they are not accompanied by quantitative benchmarks or a systematic testing of the method. In [11] the author demonstrates how TTC can be robustly and accurately recovered by using a single-calibrated camera, using a first-order model of the motion field. The affine model of the 2D field is computed with a multiresolution approach, with an approach similar to the one described by [12], instead of directly deriving a dense optic flow field. Furthermore in their approach a standard Kalman Filter is used on the Taylor coefficients to smooth the motion parameters over time. The author applied the proposed approach on real-world sequences, showing both qualitative and quantitative results. The proposed approach, even if it does not require the complete computation of the optic flow field, has a high computational cost and it is quite sensitive to noise and to variations in the illumination. A method to estimate time-to-collision with a biologically motivated population coding of motion energy neuron is presented in [13]. The developed framework is used to estimate both the optic flow and the TTC, and comparisons with the TTC estimated starting from other optic flow methods are presented. Optic flow has been used to design autonomous guidance algorithms for several robotic applications (e.g. see [14–17]). Nevertheless, such approaches are inspired by simple biological systems, thus they are not directly comparable to the method presented in this paper.

Other approaches for the computation of TTC are based on feature (points and lines) correspondences [18], these methods suffer when the localization of corner features is insufficient for a reliable estimation of the differential invariants, as described in [19]. In [20] the authors proposed a method for obstacle detection, by computing the differential invariants (divergence and deformation) from the change in moments in the images. They aimed both to give an interpretation of the scene, in terms of orientation of the surfaces, and to compute TTC, once they have decided if an object could be an obstacle or not. They applied the proposed approach in real-world situations, in an automotive context, even if they have not performed an extensive evaluation of their method.

A completely different approach is proposed in [21], where the authors define the boundary of a patch in the image by B-splines. The variation in time of the boundary of the patch is determined by integration of the normal flow components along its boundary. From the relationships between divergence and deformation, both the orientation of the surfaces and the TTC are estimated, without the need for a dense accurate optic flow fields. A disadvantage of these methods is the need of finding a robust closed contour on the objects toward which one aims to estimate the slant, the tilt or the TTC.

3. Affine models of image motion

3.1. Elementary flow components

The image motion field $\boldsymbol{v} = (v_x, v_y)^T$ can be described in terms of its linear decomposition, on the basis of its first-order (linear) properties. From this perspective, local spatial features around a given location of flow field, can be of two types: (1) the average flow velocity at that location and (2) the structure of the local variation in a neighborhood of that locality. The former relates to the smoothness constraint or structural uniformity. The latter relates to linearity constraint or structural gradients [22]. Velocity gradients provide important cues about the 3D layout of the visual scene. Formally, they can be described as linear deformations by a first-order Taylor decomposition, around the image point $\boldsymbol{x}_0 = (x_0, y_0)^T$: $\boldsymbol{v} = \bar{\boldsymbol{v}} + \bar{\mathbf{T}}\boldsymbol{x}$, where $\bar{\boldsymbol{v}} = \boldsymbol{v}|_{\boldsymbol{x}_0} - \bar{\mathbf{T}}\boldsymbol{x}_0$ and

$$\bar{\mathbf{T}} = \begin{bmatrix} \frac{\partial v_x}{\partial x} & \frac{\partial v_x}{\partial y} \\ \frac{\partial v_y}{\partial x} & \frac{\partial v_y}{\partial y} \end{bmatrix}_{\boldsymbol{x}_0}. \tag{1}$$

By breaking down the tensor in its dyadic components, the motion field can be locally described through two-dimensional maps ($\boldsymbol{g}: R^2 \mapsto R^2$) representing elementary flow components (EFCs):

$$\boldsymbol{v} = \boldsymbol{\alpha}^x \bar{v}_x + \boldsymbol{\alpha}^y \bar{v}_y + \boldsymbol{d}_x^x \frac{\partial v_x}{\partial x}\bigg|_{\boldsymbol{x}_0} + \boldsymbol{d}_y^x \frac{\partial v_x}{\partial y}\bigg|_{\boldsymbol{x}_0} + \boldsymbol{d}_x^y \frac{\partial v_y}{\partial x}\bigg|_{\boldsymbol{x}_0} + \boldsymbol{d}_y^y \frac{\partial v_y}{\partial y}\bigg|_{\boldsymbol{x}_0}, \tag{2}$$

where $\boldsymbol{\alpha}^x:(x,y) \mapsto (1,0)$, $\boldsymbol{\alpha}^y:(x,y) \mapsto (0,1)$ are pure translations and $\boldsymbol{d}_x^x: (x,y) \mapsto (x,0)$, $\boldsymbol{d}_y^x: (x,y) \mapsto (y,0)$, $\boldsymbol{d}_x^y: (x,y) \mapsto (0,x)$, $\boldsymbol{d}_y^y: (x,y) \mapsto (0,y)$ represent cardinal deformations, basis of a linear deformation space. It is worth noting that, by weighting through the coefficients a_i, the components of pure translations could be incorporated in the corresponding deformation components, thus obtaining *generalized deformation components* in which motion boundaries are shifted or totally absent:

$$\begin{aligned} \boldsymbol{v}_x^x &= a_1 \boldsymbol{\alpha}^x + a_2 \boldsymbol{d}_x^x \stackrel{\text{def}}{=} \boldsymbol{m}_1 \\ \boldsymbol{v}_y^x &= a_3 \boldsymbol{\alpha}^x + a_4 \boldsymbol{d}_y^x \stackrel{\text{def}}{=} \boldsymbol{m}_2 \\ \boldsymbol{v}_x^y &= a_5 \boldsymbol{\alpha}^y + a_6 \boldsymbol{d}_x^y \stackrel{\text{def}}{=} \boldsymbol{m}_3 \\ \boldsymbol{v}_y^y &= a_7 \boldsymbol{\alpha}^y + a_8 \boldsymbol{d}_y^y \stackrel{\text{def}}{=} \boldsymbol{m}_4. \end{aligned} \tag{3}$$

In this way, we have four classes of deformation gradients (see Fig. 1): one stretching $(\boldsymbol{v}_i^i)$ and one shearing $\left(\boldsymbol{v}_j^i\right)$ for each cardinal direction. As it will be clear in the following, this choice gives to the model maximum flexibility: every gradient deformation within a single class will be built through the same recurrent network, just by changing its driving inputs on the basis of direct local measures on the input optic flow. The EFCs can be combined to obtain deformation subspaces representing elementary deformations such as expansion, shear and rotation as shown in Fig. 2.

It is worthy to note that Eqs. (2) and (3) describe, in fact, an affine model:

$$\begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} c_1 \\ c_4 \end{bmatrix} + \begin{bmatrix} c_2 & c_3 \\ c_5 & c_6 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}, \tag{4}$$