Available online at www.sciencedirect.com
ScienceDirect
Image and Vision Computing 26 (2008) 529–549
www.elsevier.com/locate/imavis

3D Motion from structures of points, lines and planes

Andrea Dell'Acqua *, Augusto Sarti, Stefano Tubaro

Dipartimento di Elettronica e Informazione – Politecnico di Milano, Piazza Leonardo Da Vinci 32, 20133 Milano, Italy

Received 10 November 2004; received in revised form 6 July 2007; accepted 6 July 2007

Abstract

In this article we propose a method for estimating the camera motion from a video-sequence acquired in the presence of general 3D structures. Solutions to this problem are commonly based on the tracking of point-like features, as they usually back-project onto viewpoint-invariant 3D features. In order to improve the robustness, the accuracy and the generality of the approach, we are interested in tracking and using a wider class of structures. In addition to points, in fact, we also simultaneously consider lines and planes. In order to be able to work on all such structures with a compact and unified formalism, we use here the Conformal Model of Geometric Algebra, which proved very powerful and flexible.

As an example of application of our approach, we propose a causal algorithm based on an Extended Kalman Filter, for the estimation of 3D structure and motion from 2D observations of points, lines and coplanar features, and we evaluate its performance on both synthetic and real sequences.

Keywords: Camera tracking; Geometric Algebra; Extended Kalman Filter

1. Introduction

Recovering camera parameters and 3D scene structure through the analysis of a set of acquired images is a classical Computer Vision problem that has attracted a great deal of attention of researchers from various fields in the past two decades. The literature is, in fact, rich with solutions to this problem, which exploit all sources of geometric and radiometric information.

If we focus just on the techniques based on the geometry of image features, we can roughly classify the solutions available in the literature in two broad categories: those that exploit the multi-view geometric constraints that are involved in the analysis of small subsets of images, and those that are based on the estimation of a dynamical system that captures the motion of the camera in the scene.

The former class of solutions is based on the analysis of pairs or triplets of views of the sequence, and recovers camera motion through the estimation of the two- or three-view Projective geometry (fundamental matrix or trifocal tensor [1–7]). These algorithms are known as batch techniques, and are characterised by a first step of motion and structure estimation from small subsets of views (typically 2–3) subsequently merged through bundle-adjustment or other approaches [8–10]. The most important results of this sort are nicely collected in [11,12] and, more recently, in [13]. The latter category of solutions concerns those causal and recursive algorithms that estimate camera motion and 3D structure through Extended Kalman Filtering (see [14–19]). Such solutions are particularly useful for real-time camera tracking applications, as they do not need to consider all views at the same time, but they can use only information on the past. The state vector is, in fact, upgraded every time a new measurement is available.

Although the theory behind structure and motion estimation methods is well-established, camera tracking is still considered a rather challenging problem in many practical situations. One major source of problems is the quality of feature localization and tracking (accuracy, lifespan, track-

* Corresponding author. Tel.: +39 339 5225679; fax: +39 02 2399 9611.
E-mail address: andrea.dellacqu@gmail.com (A. Dell'Acqua).

0262-8856/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2007.07.001

ing stability, etc.), as features look different from different viewpoints and occlusions often prevent them from being visible at all. The problem of estimating structure and motion in the presence of missing data (occlusions) has been thoroughly analysed in the literature (see for example [20]). Furthermore, due to the large number of unknowns (structure of the scene, motion parameters as well as camera focal length), in order to overcome the ill-conditioning of the problem it is usually desirable to use as many features and exploit as many constraints as possible in the reconstruction algorithm. For this reason we are interested in considering not just point-like features but also lines and planes in 3D space.

Solutions based on multi-view Projective constraints are available for a variety of image features and structures. An example is [21], which proposes a batch approach that deals with points, lines and planes for the affine camera case. However, no contributions that we are aware of uses mixed features in an EKF-based method.

3D scenes (particularly those of human-made environments) are usually rich with locally straight edges and locally planar surfaces. Linear image features can be easily extracted using (subpixel) edge detectors, while planar surfaces can be identified by structures of points and/or lines or through the analysis of the surface texture over a number of frames [22].

One nice property of line features is that they are less 'local' than points. This means that a line is less likely to be totally occluded and, when it happens, it usually happens quite progressively over a number of frames. As we can expect, this is also true also for planar structures, as long as the plane is directly estimated, and is not just a coplanarity constraint that we force on a set of independently estimated features. We will discuss this again with more detail in Section 6 when describing the EKF-based camera tracking application.

Although the use of line features in camera motion estimation methods has been quite extensively addressed using Projective constraints [11,12,23], lines are usually seen as a source of difficulties. As pointed out in [24], linear algorithms for camera calibration based on line features tend to perform worse than point-based algorithms. Furthermore, using tools from linear algebra it is usually quite difficult to combine mixed features, therefore points and lines are seldom combined. More recently an iterative algorithm that is able to handle the hybrid case was proposed [24], based on the so-called Multiple View matrix. Another recent algorithm exploiting lines is [25], where lines are handled through the so-called Line Motion Matrix.

As far as causal recursive schemes are concerned, some solutions based on line features were recently proposed [26,27], but the representation of lines based on linear algebra causes the algorithm to be quite cumbersome (a more compact representation has been proposed in [25]). A recent algorithm exploiting lines in a real-time visual tracking application is [28].

As far as feature coplanarity is concerned, the literature is rich with solutions that exploit this constraint [29–32]. Such solutions are usually concerned with degeneracy aspects of the multi-view geometry that arise from the constraint itself. Here we are less interested in degeneracy and more focused in exploiting such constraints in a flexible fashion (e.g. several coplanar configurations of features). An effective algorithm for multi-view case exploiting the coplanarity constraint of points is [32], which relies on bundle adjustment techniques. In the case of causal motion estimation, modified versions of [14,15] were recently proposed, which are able to accommodate coplanar point structures [33–35].

In this article we propose a causal camera motion estimation method that is able to use points, lines and planes, and allows us to exploit coplanarity conditions where applicable. This level of flexibility, however, cannot be easily achieved using tools of Linear Algebra (LA), as they would not allow us to handle all the above features and constraints with the same formalism and with a unified framework. For example, rotating sets of points, lines and planes, would require different LA tools, and there would be problems in handling coplanarity conditions for mixed sets of features. One way to overcome these difficulties is to make use of tools of Geometric Algebra (GA) [36–39]. In Geometric Algebra, in fact, all sorts of linear subspaces (lines and planes in Projective space, and also circles and spheres in Conformal space) are easily handled using the same formalism. In particular, GA allows us to adopt a unique formalism for rotating such elements with a minimal parameterization. This is well known to be a crucial issue in minimization algorithms, where the management of constraints between mutually dependent parameters is always a problem. As we will see, GA also allows us to handle intersections between geometric primitives elegantly and compactly. This will allow us to embed coplanarity constraints for points and lines in an EKF structure through a direct estimation of the plane where the features lie. This last issue is quite important as it constitutes another step forward in improving the robustness and the efficiency of motion estimation. Thanks to its capability to easily deal with Euclidean transformations of geometric primitives such as lines or planes, GA has been recently applied with success to a number of Computer Vision tasks, for example the Pose Estimation problem from line observations [40].

2. A brief GA introduction

Geometric Algebra provides us with a unified framework that allows us to exploit all the algebraic properties that are known and commonly exploited in Linear Algebra and in Lie groups/algebras, while retaining a insightful geometric interpretation of all the operations. This is made possible by the fact that the elements in GA are not point coordinates but combinations of subspaces, and by the fact