Computer Vision and Image Understanding 000 (2017) 1–15

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Absolute pose estimation from line correspondences using direct linear transformation[☆]

Bronislav Přibyl[*], Pavel Zemčík, Martin Čadík

Brno University of Technology, Faculty of Information Technology, Centre of Excellence IT4Innovations, Božetěchova 2, 612 66, Brno Czech Republic

ARTICLE INFO

Article history:
Received 25 August 2016
Revised 7 February 2017
Accepted 2 May 2017
Available online xxx

2010 MSC:
68T45

Keywords:
Camera pose estimation
Perspective-n-Line
Line correspondences
Direct linear transformation

ABSTRACT

This work is concerned with camera pose estimation from correspondences of 3D/2D lines, i. e. with the Perspective-n-Line (PnL) problem. We focus on large line sets, which can be efficiently solved by methods using linear formulation of PnL. We propose a novel method “DLT-Combined-Lines” based on the Direct Linear Transformation (DLT) algorithm, which benefits from a new combination of two existing DLT methods for pose estimation. The method represents 2D structure by lines, and 3D structure by both points and lines. The redundant 3D information reduces the minimum required line correspondences to 5. A cornerstone of the method is a combined projection matrix estimated by the DLT algorithm. It contains multiple estimates of camera rotation and translation, which can be recovered after enforcing constraints of the matrix. Multiplicity of the estimates is exploited to improve the accuracy of the proposed method. For large line sets (10 and more), the method is comparable to the state-of-the-art in accuracy of orientation estimation. It achieves state-of-the-art accuracy in estimation of camera position and it yields the smallest reprojection error under strong image noise. The method achieves top-3 results on real world data. The proposed method is also highly computationally effective, estimating the pose of 1000 lines in 12 ms on a desktop computer.

1. Introduction

Absolute pose estimation is the task of determining the relative position and orientation of a *camera* and an *object* to each other in 3D space. It has many applications in computer vision: 3D reconstruction, robot localization and navigation, visual servoing, and augmented reality are just some of them. The task can be formulated either as object pose estimation (with respect to camera coordinate frame) or as camera pose estimation (with respect to object or world coordinate frame). The latter formulation is used in this paper.

To estimate the camera pose, correspondences between known real world features and their counterparts in the image plane of the camera are needed. The features can be e. g. points, lines, or combinations of both (Kuang and Astrom, 2013). The task has been solved using point correspondences first (Fischler and Bolles, 1981; Lowe, 1987). This is called the *Perspective-n-Point* (PnP) problem and it still enjoys attention of researchers (Ferraz et al., 2014; Lepetit et al., 2009; Valeiras et al., 2016). Camera pose can also be estimated using line correspondences, which is called the *Perspective-n-Line* (PnL) problem. The PnP approach has been studied first, as points are easier to handle mathematically than lines. PnP however is limited only to cases with enough distinctive points, i. e. mainly to well textured scenes. Conversely, the PnL approach is suitable for texture-less scenes, e. g. for man-made and indoor environments. Moreover, line features are more stable than point features and are robust to (partial) occlusions.

When estimating camera pose “from scratch”, the following pipeline is typically used: (i) obtain tentative feature correspondences, (ii) filter out outliers, (iii) compute a solution from all inliers, and (iv), optionally, iteratively refine the solution, e. g. by minimizing reprojection error. Task (ii) is usually carried out by iterative solving of a problem with a minimal number of line correspondences (i. e. P3L) in a RANSAC loop. Task (iii), on the other hand, requires solving a problem with high number of lines. In some applications, the correspondences are already known and thus only task (iii) is to be solved.

In recent years, versatile PnL methods have been developed which are suitable for both of these tasks. Remarkable progress has been achieved (Ansar and Daniilidis, 2003; Mirzaei and Roumeliotis, 2011; Zhang et al., 2013), mainly in accuracy of the methods, in their robustness to image noise, and in their effectiveness. These

---

[☆] Matlab code and supplementary material are available at http://www.fit.vutbr.cz/~ipribyl/DLT-based-PnL/.

[*] Corresponding author.

E-mail addresses: ipribyl@fit.vutbr.cz (B. Přibyl), zemcik@fit.vutbr.cz (P. Zemčík), cadik@fit.vutbr.cz (M. Čadík).

http://dx.doi.org/10.1016/j.cviu.2017.05.002

Please cite this article as: B. Přibyl et al., Absolute pose estimation from line correspondences using direct linear transformation, Computer Vision and Image Understanding (2017), http://dx.doi.org/10.1016/j.cviu.2017.05.002

methods are outperformed in task (iii) however, by LPnL methods – methods based on a linear formulation of the PnL problem (Přibyl et al., 2015; Xu et al., 2016). LPnL methods are superior in terms of both accuracy and computational speed in camera pose estimation from many (∼ tens to thousands) line correspondences. The oldest LPnL method is that proposed by Hartley and Zisserman (2004, p. 180), followed recently by the method of Přibyl et al. (2015). Even more recently, Xu et al. (2016) introduced a series of LPnL methods generated by the use of Cartesian or barycentric coordinates, and by alternating whether the solution is retrieved in closed form or by optimization. As we show in this paper, space for improving accuracy of the methods still exists.

In this paper, we introduce a novel method based on linear formulation of the PnL problem, which is a combination of the DLT-Lines method of Hartley and Zisserman (2004) and the DLT-Plücker-Lines method of Přibyl et al. (2015). The former represents the 3D structure by 3D points, while the latter represents it by 3D lines parameterized by Plücker coordinates. The proposed method exploits the redundant representation of 3D structure by both 3D points and 3D lines, which leads to the reduction of the minimum required line correspondences to 5. A cornerstone of the method is a combined projection matrix recovered by the DLT algorithm. It contains multiple estimates of camera orientation and translation, enabling a more accurate estimation of the final camera pose. The proposed method achieves state-of-the-art accuracy for large line sets under strong image noise, and it performs comparably to state-of-the-art methods on real world data. The proposed method also keeps the common advantage of LPnL methods – being very fast.

The rest of this paper is organized as follows. We present a review of related work on PnL in Section 2. Then we introduce mathematical notation and Plücker coordinates of 3D lines, and show how points and lines transform and project onto the image plane in Section 3. In Section 4, we explain the application of DLT algorithm to the PnL problem in general, and we describe the existing methods DLT-Lines and DLT-Plücker-Lines. In Section 5, we propose the novel method DLT-Combined-Lines. We evaluate the performance of the proposed method using simulations and real-world experiments in Section 6, and we conclude in Section 7.

2. Related work

The task of camera pose estimation from line correspondences has been receiving attention for more than a quarter of century. Some of the earliest works are those by Dhome et al. (1989) and Liu et al. (1990). They introduce two different ways to deal with the PnL problem – algebraic and iterative approaches – both of which have different properties and thus also different uses. A specific subset of algebraic approaches are the methods based on linear formulation of the PnL problem.

2.1. Iterative methods

The iterative approaches consider pose estimation as a nonlinear least squares problem by iteratively minimizing specific error function, which usually has a geometrical meaning. In the early work of Liu et al. (1990), the authors attempted to estimate the camera position and orientation separately developing a method called R_then_T. Later on, Kumar and Hanson (1994) introduced a method called R_and_T for simultaneous estimation of camera position and orientation, and proved its superior performance to R_then_T. Recently, Zhang et al. (2016) proposed two modifications of the R_and_T algorithm exploiting the uncertainty properties of line segment endpoints. Several other iterative methods are also capable of *simultaneous* estimation of pose parameters and line correspondences, e. g. David et al. (2003) and Zhang et al. (2012). They pose an orthogonal approach to the common RANSAC-based correspondence filtering and consecutive separate pose estimation.

Iterative algorithms suffer from two common major issues when not initialized accurately: They converge slowly, and more severely, the estimated pose is often far from the true camera pose, finding only a local minimum of the error function. This makes iterative approaches suitable for final refinement of an initial solution, provided by some other algorithm.

2.2. Algebraic methods

The algebraic approaches estimate the camera pose by solving a system of (usually polynomial) equations, minimizing an algebraic error. Their solutions are thus not necessarily geometrically optimal; on the other hand, no initialization is needed.

Among the earliest efforts in this field are those of Dhome et al. (1989) and Chen (1990). Both methods solve the minimal problem of pose estimation from 3 line correspondences in a closed form. Solutions of the P3L problem are multiple: up to 8 solutions may exist (Chen, 1990). Unfortunately, neither method is able to exploit more measurements to remove the ambiguity, and both methods are sensitive to presence of image noise.

Ansar and Daniilidis (2003) developed a method that is able to handle 4 or more lines, limiting the number of possible solutions to 1. Lifting is employed to convert a polynomial system to linear equations in the entries of a rotation matrix. This approach may, however, fail in cases of singular line configurations (e. g. lines in 3 orthogonal directions – Navab and Faugeras, 1993) as the underlying polynomial system may have multiple solutions. The algorithm has quadratic computational complexity ($O(n^2)$, where n is the number of lines), which renders it impractically slow for processing higher numbers of lines. The method also becomes unstable with increasing image noise, eventually producing solutions with complex numbers.

Recently, two major improvements of algebraic approaches have been achieved. First, **Mirzaei** and Roumeliotis (2011) proposed a method, which is more computationally efficient ($O(n)$), behaves more robustly in the presence of image noise, and can handle the minimum of 3 lines, or more. A polynomial system with 27 candidate solutions is constructed and solved through the eigendecomposition of a multiplication matrix. Camera orientations having the least square error are considered to be the optimal ones. Camera positions are obtained separately using linear least squares. A weakness with this algorithm is that it often yields multiple solutions. Also, despite its linear computational complexity, the overall computational time is still high due to slow construction of the multiplication matrix, which causes a high constant time penalty: 78 ms / 10 lines.

The second recent improvement is the Robust PnL (**RPnL**) algorithm of Zhang et al. (2013). Their method works with 4 or more lines and is more accurate and robust than the method of Mirzaei and Roumeliotis. An intermediate model coordinate system is used in the method of Zhang et al., which is aligned with a 3D line of longest projection. The lines are divided into triples, for each of which a P3L polynomial is formed. The optimal solution of the polynomial system is selected from the roots of its derivative in terms of a least squares residual.

The RPnL algorithm was later modified by Xu et al. (2016) into the Accurate Subset based PnL (**ASPnL**) algorithm, which acts more accurately on small line sets. However, it is very sensitive to outliers, limiting its performance on real-world data. This algorithm is compared to other state-of-the-art methods in Section 6. A drawback of both RPnL and ASPnL is that their computational time increases strongly for higher number of lines – from 8 ms / 10 lines to 630–880 ms / 1000 lines.

Please cite this article as: B. Přibyl et al., Absolute pose estimation from line correspondences using direct linear transformation, Computer Vision and Image Understanding (2017), http://dx.doi.org/10.1016/j.cviu.2017.05.002