

Camera Pose Estimation from Lines using Plücker Coordinates

Bronislav Příbyl
ipribyl@fit.vutbr.cz
Pavel Zemčik
zemcik@fit.vutbr.cz
Martin Čadík
cadik@fit.vutbr.cz

Graph@FIT, CPhoto@FIT
Department of Computer Graphics and Multimedia
Faculty of Information Technology
Brno University of Technology
Božetěchova 2
612 66 Brno
Czech Republic

Camera pose estimation is the task of determining the position and orientation of a camera in 3D space, based on correspondences between 3D features and their 2D images. When these features are *lines*, the task is called the *Perspective-n-Line* (PnL) problem. A remarkable progress in solving PnL has been achieved in the last years, particularly thanks to the work of Mirzaei and Roumeliotis [5] and more recently to the work of Zhang *et al.* [6]. Both of the methods are accurate, cope well with noisy data, and they are more efficient than the previously known methods.

In this paper, we propose an algebraic solution to the PnL problem which (i) is more than the order of magnitude faster than the state-of-the-art [5, 6], (ii) yields only one solution of the PnL problem, and (iii) similarly to the state-of-the-art, copes well with image noise and is initialization-free. As an alternative to the commonly used RANSAC, Algebraic Outlier Rejection scheme [1] is used to deal with mismatched line correspondences. The proposed method requires at least 9 lines, but it is particularly suitable for large scale and noisy scenarios, where it can be reliably used for 25 and more lines.

Given two distinct 3D points $\mathbf{A} = (a_x a_y a_z a_w)^\top$ and $\mathbf{B} = (b_x b_y b_z b_w)^\top$ in homogeneous coordinates, a line joining them can be represented using Plücker coordinates [3] as a homogeneous 6-vector $\mathbf{L} = (\mathbf{u}^\top \mathbf{v}^\top)^\top = (L_1 L_2 L_3 L_4 L_5 L_6)^\top$, where

$$\begin{aligned} \mathbf{u}^\top &= (L_1 L_2 L_3) = (a_x a_y a_z) \times (b_x b_y b_z) \\ \mathbf{v}^\top &= (L_4 L_5 L_6) = a_w(b_x b_y b_z) - b_w(a_x a_y a_z), \end{aligned} \quad (1)$$

' \times ' denotes a vector cross product. The \mathbf{v} part encodes direction of the line while the \mathbf{u} part encodes position of the line in space. In fact, \mathbf{u} is a normal vector of an interpretation plane – a plane passing through the line and the origin. As a consequence, \mathbf{L} must satisfy a bilinear constraint $\mathbf{u}^\top \mathbf{v} = 0$.

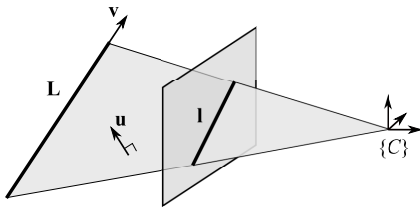


Figure 1: 3D line projection. The 3D line \mathbf{L} is parameterized by its direction vector \mathbf{v} and a normal \mathbf{u} of its interpretation plane, which passes through the origin of the camera coordinate system $\{C\}$. Since the projected 2D line \mathbf{I} lies at the intersection of the interpretation plane and the image plane, it is fully defined just by the normal \mathbf{u} .

A 3D line \mathbf{L} is projected onto the normalized image plane using the 3×6 line projection matrix \mathbf{P} as

$$\mathbf{I} \approx \mathbf{P}\mathbf{L}, \quad (2)$$

where $\mathbf{I} = (l_x l_y l_w)^\top$ is a homogeneous 2D line, ' \approx ' denotes an equivalence of homogeneous coordinates and

$$\mathbf{P} = \begin{pmatrix} \mathbf{R} & \mathbf{R}[-\mathbf{t}]_\times \end{pmatrix}. \quad (3)$$

\mathbf{R} is a 3×3 rotation matrix and $[\mathbf{t}]_\times$ is a 3×3 skew-symmetric matrix constructed from the translation vector \mathbf{t} .

As the projection of 3D lines is defined by Eq. (2), the problem of camera pose estimation resides in estimation of the line projection matrix \mathbf{P} , which encodes all the six camera pose parameters $t_x, t_y, t_z, \alpha, \beta, \gamma$. We solve this problem using the Direct Linear Transformation (DLT) algorithm similarly to Hartley [2], who works with points. The system of linear equations (2) can be transformed into a homogeneous system

$$\mathbf{M}\mathbf{p} = \mathbf{0}. \quad (4)$$

This forms a $2n \times 18$ measurement matrix \mathbf{M} which contains coefficients of equations generated by correspondences between 3D lines and their projections $\mathbf{L}_i \leftrightarrow \mathbf{l}_i$ ($i = 1 \dots n, n \geq 9$). The DLT then solves (4) for \mathbf{p} which is a 18-vector containing the entries of the line projection matrix \mathbf{P} . Eq. (4), however, holds only in the noise-free case. If a noise is present in the measurements, an inconsistent system is obtained.

$$\mathbf{M}\hat{\mathbf{p}} = \boldsymbol{\varepsilon} \quad (5)$$

Only an approximate solution $\hat{\mathbf{p}}$ may be found minimizing a $2n$ -vector of measurement residuals $\boldsymbol{\varepsilon}$. Since DLT algorithm is sensitive to the choice of coordinate system, it is crucial to prenormalize the data to get properly conditioned \mathbf{M} .

Once the system of linear equations given by (5) is solved in the least squares sense, *e.g.* by SVD of \mathbf{M} , the estimate $\hat{\mathbf{P}}$ of the 3×6 line projection matrix can be recovered from the 18-vector $\hat{\mathbf{p}}$. $\hat{\mathbf{P}}$ is obtained as a least squares solution of Eq. (5) so it does not satisfy the constraints imposed on \mathbf{P} by Eq. (3). We propose a method to extract the camera pose parameters from the estimate $\hat{\mathbf{P}}$, which is based on an observation that the right 3×3 submatrix of \mathbf{P} has exactly the same structure as the essential matrix [4] used in multi-view computer vision. Details of the method are described in the full paper.

Simulations and experiments with real data show that computational efficiency of the proposed method is superior to the state-of-the-art, achieving speed-ups of more than one order of magnitude for high number of lines. At the same time, accuracy and robustness of the method are on par with the state-of-the-art.

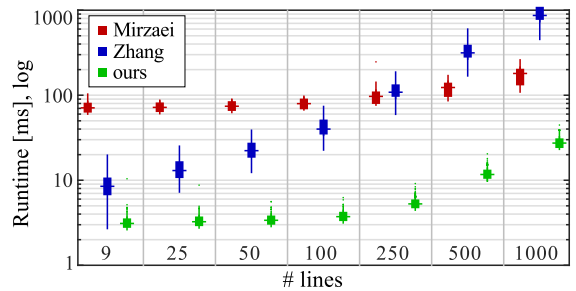


Figure 2: The distribution of runtimes as a function of the number of lines.

- [1] L. Ferraz, X. Binefa, and F. Moreno-Noguer. Very fast solution to the pnp problem with algebraic outlier rejection. In *IEEE Conference on Computer Vision and Pattern Recognition 2014*, pages 501–508. IEEE, 2014.
- [2] R. I. Hartley. Minimizing algebraic error. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 356(1740):1175–1192, 1998.
- [3] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004. ISBN 0521540518.
- [4] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.
- [5] F. M. Mirzaei and S. I. Roumeliotis. Globally optimal pose estimation from line correspondences. In *IEEE International Conference on Robotics and Automation 2011*, pages 5581–5588. IEEE, 2011.
- [6] L. Zhang, C. Xu, K.-M. Lee, and R. Koch. Robust and efficient pose estimation from line correspondences. In *Asian Conference on Computer Vision 2012*, pages 217–230. Springer, 2013.