SCIENCE CHINA
Information Sciences

• LETTER •

October 2021, Vol. 64 209203:1–209203:3
https://doi.org/10.1007/s11432-019-9925-1

Minimal solution for estimating fundamental matrix under planar motion

Qiang FU[1,2*], Xinxing MU[1] & Yu WANG[3]

[1]*School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing* 100083*, China;*
[2]*Institute of Artificial Intelligence, University of Science and Technology Beijing, Beijing* 100083*, China;*
[3]*Institute of Automation, Chinese Academy of Sciences, Beijing* 100190*, China*

Received 23 January 2019/Revised 2 April 2019/Accepted 19 June 2019/Published online 20 July 2020

Citation Fu Q, Mu X X, Wang Y. Minimal solution for estimating fundamental matrix under planar motion. Sci China Inf Sci, 2021, 64(10): 209203, https://doi.org/10.1007/s11432-019-9925-1

Dear editor,

Fundamental matrix estimation from point correspondences is a basic problem in computer vision and robotics. It is applicable in many areas, such as visual servoing (vision-based control) and structure-from-motion when the camera is uncalibrated. In computer vision, the fundamental matrix under general motion has been studied extensively [1], while in robotics, visual navigation methods are widely used [2], and the simplified fundamental matrix with fewer degrees of freedom than the fundamental matrix under general motion has been investigated. The reason for the simplified fundamental matrix is that mobile robots have their own motion characteristics, e.g., they often operate on floors (indoor) or roads (outdoor); thus, their motions can be considered nearly planar or locally planar [3].

To address outliers (i.e., mismatches) in practice, minimal solvers combined with robust estimation algorithms such as the random sample consensus (RANSAC) algorithm [4] are typically adopted. The number of iterations required for RANSAC increases exponentially with the required number of minimal point correspondences [4]. Therefore, the design of minimal solvers is very important in real systems. The commonly used minimal solver for the fundamental matrix under general motion is the seven-point method [1]. However, this method is not appropriate for planar motion cases (the accuracy and efficiency of the fundamental matrix estimation are degraded). In addition, a linear six-point method for estimating the fundamental matrix under planar motion has been proposed [5]; however, it is not a minimal solver and considers the rank constraint via singular value decomposition (SVD) correction.

In this study, to solve the minimal solver problem under planar motion, a four-point minimal solver is proposed to estimate the fundamental matrix, which directly incorporates all constraints under planar motion. Combining the proposed minimal solver with RANSAC enables much faster and more accurate estimation of the fundamental matrix under planar motion than using existing methods. A detailed comparison of the proposed and existing methods is given in Appendix A.

Methodology. As shown in Figure 1(a), the camera moves on the XZ plane with unknown three-dimensional (3D) feature points. The fundamental matrix $\boldsymbol{F}$ takes the following form [5]:

$$\boldsymbol{F} = \boldsymbol{K}'^{-\mathrm{T}}[\boldsymbol{T}]_{\times}\boldsymbol{R}\boldsymbol{K}^{-1} = \begin{bmatrix} 0 & f_1 & f_2 \\ f_3 & 0 & f_4 \\ f_5 & f_6 & f_7 \end{bmatrix}, \tag{1}$$

where $\boldsymbol{K} = \boldsymbol{K}'$ are the camera intrinsic matrices corresponding to the two views, $\boldsymbol{R}$ is the rotation matrix, $\boldsymbol{T}$ is the translation vector between the two views, and $f_1, f_2, \ldots, f_7$ are unknown scalars.

Here, consider four corresponding image points, i.e., $\{(\boldsymbol{p}_i, \boldsymbol{p}'_i)|i = 1, \ldots, 4\}$, where $\boldsymbol{p}_i = (u_i, v_i, 1)^{\mathrm{T}}$ and $\boldsymbol{p}'_i = (u'_i, v'_i, 1)^{\mathrm{T}}$. Then $\boldsymbol{F}$ satisfies the following [1]:

$$\boldsymbol{p}_i'^{\mathrm{T}}\boldsymbol{F}\boldsymbol{p}_i = 0. \tag{2}$$

According to (1) and (2), the following can be derived:

$$\boldsymbol{A}\boldsymbol{f} = \boldsymbol{0}, \tag{3}$$

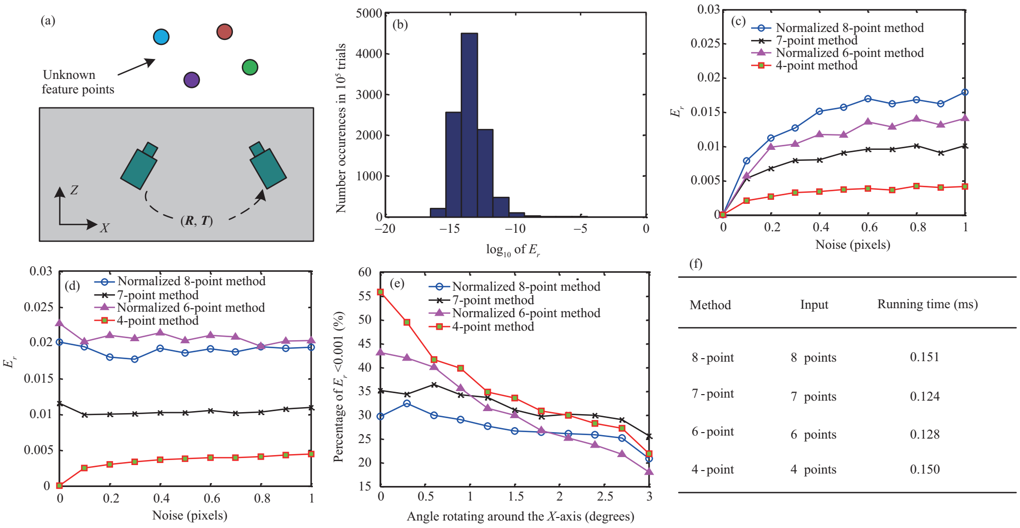
where

$$\boldsymbol{A} = \begin{bmatrix} u_1v'_1 & u_1 & v_1u'_1 & v_1 & u'_1 & v'_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_4v'_4 & u_4 & v_4u'_4 & v_4 & u'_4 & v'_4 & 1 \end{bmatrix} \in \mathbb{R}^{4\times 7}, \tag{4}$$

and $\boldsymbol{f} = [f_1, f_2, \ldots, f_7]^{\mathrm{T}} \in \mathbb{R}^7$. Here, $\boldsymbol{F}'_1$, $\boldsymbol{F}'_2$, and $\boldsymbol{F}'_3$ are vectors that span the right null space of $\boldsymbol{A}$. The three vectors correspond to three 3×3 matrices $\boldsymbol{F}_1$, $\boldsymbol{F}_2$, and $\boldsymbol{F}_3$, respectively, and the fundamental matrix must be of the following form:

$$\boldsymbol{F} = x\boldsymbol{F}_1 + y\boldsymbol{F}_2 + z\boldsymbol{F}_3, \tag{5}$$

* Corresponding author (email: fuqiang@ustb.edu.cn)



Method	Input	Running time (ms)
8 - point	8 points	0.151
7 - point	7 points	0.124
6 - point	6 points	0.128
4 - point	4 points	0.150

Figure 1 (Color online) (a) Illustration of camera movement; (b) $\log_{10}$ of E_r using noise-free data for general 3D scene (planar camera motion); (c) E_r under different noise levels for general 3D scene (planar camera motion); (d) E_r under different noise levels for planar scene (planar camera motion); (e) performance of different methods for increasing non-planar camera motion; (f) runtime of four different methods with eight-, seven-, six-, and four-point correspondences.

where x, y, z are unknown scalars. The fundamental matrix is defined up to a scale factor, and it is assumed that $z = 1$. Thus, we obtain the following:

$$\boldsymbol{F} = x\boldsymbol{F}_1 + y\boldsymbol{F}_2 + \boldsymbol{F}_3. \tag{6}$$

Constraint 1. The fundamental matrix $\boldsymbol{F}$ satisfies the following singularity constraint:

$$\det(\boldsymbol{F}) = 0. \tag{7}$$

Constraint 2. Assume that $\boldsymbol{K}' = \boldsymbol{K}$ (e.g., monocular images). Then, the fundamental matrix $\boldsymbol{F}$ should satisfy the following constraint [1]:

$$\det\left(\frac{\boldsymbol{F} + \boldsymbol{F}^{\mathrm{T}}}{2}\right) = 0. \tag{8}$$

Further, we obtain the following:

$$\det(\boldsymbol{F} + \boldsymbol{F}^{\mathrm{T}}) = 0. \tag{9}$$

Inserting equation (6) into constraint equations (7) and (9) produces two polynomial equations of degree three in unknowns x, y:

$$\boldsymbol{C}[x^3, x^2y, xy^2, y^3, x^2, xy, y^2, x, y, 1]^{\mathrm{T}} = \boldsymbol{0}, \tag{10}$$

where $\boldsymbol{C} \in \mathbb{R}^{2\times 10}$ is the coefficient matrix comprising the elements of $\boldsymbol{F}_1, \boldsymbol{F}_2, \boldsymbol{F}_3$. Note that the polynomial equations (10) can be solved efficiently by the automatic generator method [6]. In total, there are nine complex solutions; however, we are only interested in real solutions. The steps of the proposed four-point method are summarized in Algorithm 1.

Algorithm 1 Fundamental matrix estimation by proposed four-point method

Step 1: Given four-point correspondences $\{(\boldsymbol{p}_i, \boldsymbol{p}'_i) | i = 1, \ldots, 4\}$, obtain $\boldsymbol{A}$ using (2)–(4);
Step 2: Compute right null space of matrix $\boldsymbol{A}$ and obtain the corresponding $\boldsymbol{F}_1, \boldsymbol{F}_2, \boldsymbol{F}_3$;
Step 3: With $\boldsymbol{F}_1, \boldsymbol{F}_2, \boldsymbol{F}_3$, compute coefficient matrix $\boldsymbol{C}$ in (10) (Appendix B);
Step 4: Solve (10) using the automatic generator method [6] and obtain x, y;
Step 5: Recover fundamental matrix $\boldsymbol{F}$ according to (6).

Experiments. The numerical accuracy of the proposed method for a camera under planar motion with randomly-generated 3D scene points is shown in Figure 1(b). Similar to that shown in [7], the estimation error is computed as follows:

$$E_r = \min_i \min\left(\left\|\frac{\boldsymbol{F}}{\|\boldsymbol{F}\|} - \frac{\boldsymbol{F}_i}{\|\boldsymbol{F}_i\|}\right\|, \left\|\frac{\boldsymbol{F}}{\|\boldsymbol{F}\|} + \frac{\boldsymbol{F}_i}{\|\boldsymbol{F}_i\|}\right\|\right), \tag{11}$$

where $\boldsymbol{F}$ is the ground truth and $\{\boldsymbol{F}_i\}$ are the estimated fundamental matrices. It is known that the proposed four-point method deals with the general 3D structure case successfully (similar results can be obtained for the planar structure case).

To demonstrate the superiority of the proposed method, we compare it with the six-point [5], seven-point, and eight-point [1] methods. Gaussian noises with mean value $\mu = 0$ and standard deviation σ varying from 0 to 1 pixel are added to the image points. The results are shown in Figures 1(c) and (d). As can be seen, four-point and seven-point methods outperform the six-point and eight-point methods in the general structure case. Note that the six-point, seven-point and eight-point methods fail in the noise-free planar structure case; however, the proposed four-point method works in this case. In addition, in the presence of noise, the proposed four-point method is more accurate and robust than the other methods.

Next, it is interesting to investigate what occurs if planar motion assumption is invalid. Here, add a varying amount of angular rotation around the X-axis to the camera motion. Gaussian noise with $\mu = 0$ and $\sigma = 0.5$ pixel is added to the image points. Figure 1(e) shows that the performance of the proposed four-point method and the six-point method decreases quickly under non-planar motion because they are designed according to the planar motion assumption. However, the number of good estimations ($E_r < 0.001$) obtained using the proposed four-point method is greater than that obtained using the other methods with about two degrees of non-planar motion. Note that an angular change of one degree means a change of 2% in steepness for roads, which is rather uncommon [8].

The computation efficiency of the four methods is investigated. The experiments are performed using Matlab 2011b running on a desktop computer (Intel Core Processor at 3.2 GHz). Figure 1(f) shows that all four methods run very quickly with similar computation time. Appendix C compares the results of the four methods combined with RANSAC for real images.

Note that the seven-point and eight-point methods are designed for general motion. The four-point method outperforms these methods for planar or near-planar motion because it exploits all available geometric constraints of the problem. In addition, the four-point method outperforms the six-point method because it directly enforces the rank two constraint of the fundamental matrix (not by SVD correction).

Conclusion. We have proposed a four-point minimal solver to estimate the fundamental matrix. The proposed method directly incorporates all constraints under planar motion. The experimental results demonstrate that the proposed method works for both planar and non-planar structure cases. Moreover, in the presence of noise when the camera undergoes planar or near-planar motion, the proposed method is more accurate and robust than state-of-the-art methods. Combining the proposed method with RANSAC enables us to determine the fundamental matrix under planar motion more accurately in the presence of outliers.

Acknowledgements This work was supported by Fundamental Research Funds for the Central Universities (Grant No. FRF-TP-18-100A1), National Natural Science Foundation of China (Grant No. 61803025), Beijing Science and Technology Project (Grant No. Z181100003118006), and Joint Funds of Equipment Pre-Research and Ministry of Education of China (Grant No. 6141A02033339).

Supporting information Appendixes A–C. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Hartley R, Zisserman A. Multiple View Geometry in Computer Vision. 2nd ed. Cambridge: Cambridge University Press, 2004
2 He W, Li Z J, Chen C L P. A survey of human-centered intelligent robots: issues and challenges. IEEE/CAA J Autom Sin, 2017, 4: 602–609
3 Choi S, Kim J H. Fast and reliable minimal relative pose estimation under planar motion. Image Vision Comput, 2018, 69: 103–112
4 Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM, 1981, 24: 381–395
5 Ortín D, Montiel J M M. Indoor robot motion based on monocular images. Robotica, 2001, 19: 331–342
6 Kukelova Z, Bujnak M, Pajdla T. Automatic generator of minimal problem solvers. In: Proceedings of the 10th European Conference on Computer Vision, Marseille, 2008. 302–315
7 Nister D. An efficient solution to the five-point relative pose problem. IEEE Trans Pattern Anal Machine Intell, 2004, 26: 756–770
8 Booij O, Zivkovic Z. The Planar Two Point Algorithm. IAS Technical Report IAS-UVA-09-05. 2009