Toward an efficient method to estimate trifocal tensor based on lines

Daniel Braun[1] Pascal Vasseur[2] Cédric Demonceaux[1]

[1] ERL VIBOT CNRS 6000, ImViA, Université Bourgogne Franche-Comté, 71200 Le Creusot, France
[2] Normandie Univ, UNIROUEN, UNIHAVRE, INSA Rouen, LITIS, 76000 Rouen, France

daniel.braun@u-bourgogne.fr

1 Abstract

This paper presents a new approach for the trifocal tensor estimation in three view geometry based on lines. Current state of the art methods using lines require simultaneously 13 visible lines in three views, while point based methods only require 7 points [1]. In order to propose a tractable alternative to the point based method, our objective here is to reduce to 8 the number of required lines by assuming two angles of the camera rotation matrix as known thanks to an IMU for example. We propose a first validation of this minimal solution on simulated data in order to estimate its robustness to noise.

2 Trifocal tensor

2.1 Definition

The trifocal tensor is generally represented by a set of three 3 × 3 matrices $\{T_1, T_2, T_3\}$. It is gathering all geometric relations binding three views that are independent to scene structure. The system is defined in its canonical form by three projective cameras $P_1 = [I \mid 0]$, $P_2 = [R_2 \mid \boldsymbol{t}_2]$ and $P_3 = [R_3 \mid \boldsymbol{t}_3]$ such that :

$$T_i = \boldsymbol{r}_2^i(\boldsymbol{t}_3)^{\mathrm{T}} - \boldsymbol{t}_2(\boldsymbol{r}_3^i)^{\mathrm{T}} \text{ for } i = 1, 2, 3\ , \tag{1}$$

where $\boldsymbol{r}_2^i$ and $\boldsymbol{r}_3^i$ are the i^{th} columns of the rotation matrices R_2 and R_3 respectively.
The trifocal tensor is composed of 27 elements, which are 26 independent ratios up to a common scale. Yet, it only has 18 degrees of freedom (dof), considering that each camera has 6 dof in the projective world frame. It means that the system can be completely described by 18 parameters.

2.2 Trilinearity constraint

In [2], if a line is visible in three camera views, there is a relation linking the three projected lines called the trilinearity constraint :

$$\boldsymbol{l}_i^1 = (\boldsymbol{l}^2)^{\mathrm{T}} T_i \boldsymbol{l}^3 \text{ for } i = 1, 2, 3\ , \tag{2}$$

with $\boldsymbol{l}_i^1$ representing the i^{th} coordinates of the line $\boldsymbol{l}^1$ in the first camera. $\boldsymbol{l}^2$ and $\boldsymbol{l}^3$ are the coordinates of the same line respectively in the second and the third camera and T_i is the i^{th} matrix of the trifocal tensor.

3 Algorithm

3.1 Rotation constraint

In order to reduce our system complexity, we impose a constraint on the camera rotations. In that way, we assume the knowledge of the camera's attitude, i.e. pitch and roll angles, leaving the camera only depending on the yaw rotation. The new constraint system can then be expressed as :

$$P_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, P_i = \begin{bmatrix} \cos(\theta_i^z) & -\sin(\theta_i^z) & 0 & t_i'^x \\ \sin(\theta_i^z) & \cos(\theta_i^z) & 0 & t_i'^y \\ 0 & 0 & 1 & t_i'^z \end{bmatrix} \text{ for } i = 2, 3\ , \tag{3}$$

$$T_1 = \begin{bmatrix} T_1^{11} & T_1^{12} & T_1^{13} \\ T_1^{21} & T_1^{22} & T_1^{23} \\ T_1^{31} & T_1^{32} & 0 \end{bmatrix}, T_2 = \begin{bmatrix} T_2^{11} & T_2^{12} & T_2^{13} \\ T_2^{21} & T_2^{22} & T_2^{23} \\ T_2^{31} & T_2^{32} & 0 \end{bmatrix} \text{ and } T_3 = \begin{bmatrix} 0 & 0 & T_3^{13} \\ 0 & 0 & T_3^{23} \\ T_3^{31} & T_3^{32} & T_3^{33} \end{bmatrix}. \tag{4}$$

The tensor matrices in (4) are deduced from equation (1) with the new constraint system in (3). The number of elements of the trifocal tensor has been reduced from 27 to 21 with only 16 linearly independant elements, plus the overall scale.

3.2 Trifocal tensor estimation

The equation (2) can be reformulated as $\left((\boldsymbol{l}^2)^{\mathrm{T}}\left[T_1, T_2, T_3\right]\boldsymbol{l}^3\right)[\boldsymbol{l}^1]_\times = \mathbf{0}^{\mathrm{T}}$. Two independent equations can be extracted from it, which means that 8 triplets of lines are necessary to get the 16 equations required to fully solve the system, up to scale. The equations system can be written as $A\boldsymbol{t} = \mathbf{0}$ with A being a 16×17 matrix and $\boldsymbol{t}$ a 17×1 vector containing all the entries of the trifocal tensor. The scaling factor is imposed with the constraint $\|\boldsymbol{t}\| = 1$.

4 Experiments

The validation of our method is performed on synthetic data composed of a set of lines automatically generated and defined by their two end points.
To evaluate the robustness of our method, we apply a white gaussian noise on the image points composing the lines with a standard deviation σ varying from 0 to 2 pixels with a step of 0.25. For each level of noise, the error is averaged over 100 executions.
In figure 1, we compare the behaviour to image point noise of our method based on 8 lines with the 13 lines trifocal, the 7 points trifocal and 8 points fundamental algorithms. Each method has been optimized with a basic RANSAC to filter out outliers from the input lines or points. The results demonstrate that our method competes with standard methods and highly outperforms the 13 lines algorithm.

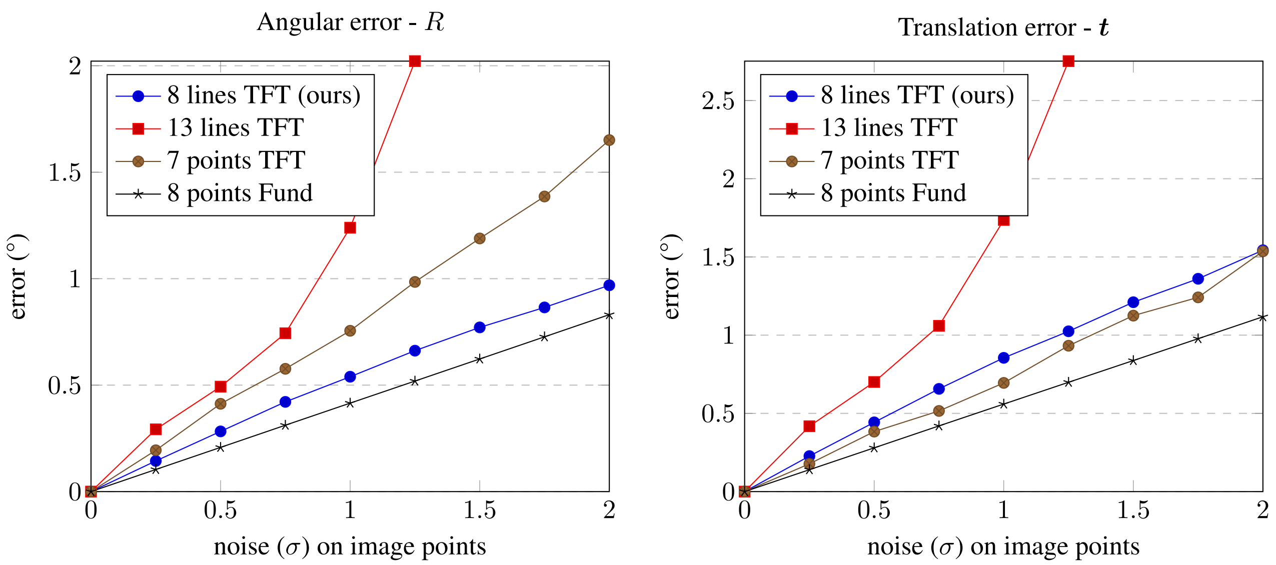

Figure 1: Average angular error for the camera rotation (R) on the left and for the translation direction ($\boldsymbol{t}$) on the right, depending on the gaussian noise added to the image points. Performance comparison of the 8 lines trifocal (ours), the 13 lines trifocal, the 7 points trifocal and the 8 points fundamental methods.

5 Conclusion and Future Work

Considering the first experimental results, the estimation of the trifocal tensor with 8 lines outperforms the 13 lines algorithm and shows promising results with respect to classical approaches. Besides, since there are fewer lines than points visible in the image, this drop in performance is an acceptable compromise compared to the possible gain in optimization complexity.
As future works, we aim at extending this study to noise robustness on angles considered as known and at obtaining experimental results on a real world dataset.

References

[1] L. F. Julià and P. Monasse, “A critical review of the trifocal tensor estimation,” in *Image and Video Technology* (M. Paul, C. Hitoshi, and Q. Huang, eds.), (Cham), pp. 337–349, Springer International Publishing, 2018.

[2] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second ed., 2004.