Trifocal Tensor and Relative Pose Estimation from 8 Lines and Known Vertical Direction

Banglei Guan, Pascal Vasseur, Cédric Demonceaux

Trifocal Tensor and Relative Pose Estimation from 8 Lines and Known Vertical Direction

Banglei Guan[1], Pascal Vasseur[2] and Cédric Demonceaux[3]

***Abstract*— In this paper, we present a relative pose estimation algorithm based on lines knowing the vertical direction associated to each image. We demonstrate that a closed-form solution requiring only eight lines between three views is possible. As a linear solution, it is shown that our approach outperforms the standard trifocal estimation based on 13 triplets of lines and can be efficiently inserted into an hypothesize-and-test framework such as RANSAC. We also study our approach on different singular configurations of lines. The method is evaluated on both synthetic data and real-world sequences from KITTI and the Zürich Urban Micro Aerial Vehicle datasets. Our method is compared to 13 lines algorithm as well to points based methods such as 7-points, 5-points and 3-points.**

I. INTRODUCTION

Relative pose estimation between views is a fundamental task in computer vision [1], [2], [3], [4], [5], [6], [7] and constitutes an essential step in most of Structure from Motion (SfM) and Simultaneous Localization And Mapping (SLAM) pipelines [8]. This task is mainly carried out by approaches based on points of interest extracted and matched across couples of images [1], [2]. The relative pose estimation of two views has been widely studied. In the non calibrated case, at least seven or eight points are necessary in order to estimate the fundamental matrix and consequently the relative pose up to scale between two views [1]. If intrinsic parameters of the camera are known, the essential matrix is then sufficient to describe the relative pose and can be estimated from five points [9]. Many other cases between two views have been declined according to some prior knowledge such as the vertical direction [10], [11], [12], the kind of motion [13], [14], [15] or the nature of the environment [16], [17], [18].

Trifocal relative pose has long been believed to augment relative pose estimation from two views [19], [20], [21]. When the relative pose estimation of two views fails, the trifocal relative pose can be considered as a fallback option. But the trifocal relative pose estimation is usually considered a more hard problem and requires a more expensive operation. For the uncalibrated case, the trifocal relative pose can be represented by a $3 \times 3 \times 3$ trifocal tensor, which has 18 degrees of freedom (DOFs) [1]. Six points are at least required to estimate the trifocal relative pose [22]. However,

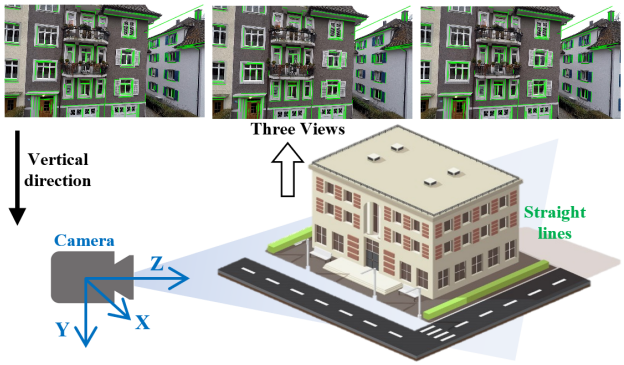


Fig. 1. Trifocal relative pose from a minimal number of eight lines with known vertical direction. Line features are easily obtained in man-made environments. The vertical direction of the camera is provided by an inertial measurement unit (IMU) or detecting vanishing points.

in some scenes with less textures, point features may not be available but line features are visible in large quantities, such as the man-made environments. Even though the relative pose of two views cannot be determined if only line features exist, it has be proved that thirteen lines is sufficient to solve the trinocular relative pose linearly [23]. In the calibrated case, the estimation of trifocal relative pose requires the determination of 11 DOFs, which include six unknowns for each pair of views and less one for metric ambiguity [24]. The trinocular relative pose can be estimated from four points [25], [20] or six lines [26], [27]. Futhermore, the minimal problems for generic arrangements of points and lines have been studied [28], [29], [24].

Reducing the number of requested matched or tracked features between consecutive views allows to apply some robust methods in order to discard outliers while limiting the computation time [30], [31], [32]. The Random Sample Consensus (RANSAC) is probably the most used techniques for selecting the inliers for a final relative pose estimation [33]. It consists in randomly selecting the minimal number of features in order to estimate a solution and to verify how many other features check this solution within a margin error [8]. Consequently, for a probability of success of 0.99 and a rate of outliers equal to 0.5, the number of RANSAC trials is divided by 32, if eight features are used instead of thirteen. In the case of a robust estimation based on thirteen features, 37724 trials are necessary whereas 1177 are sufficient if only eight features are required.

In this paper, we propose to exploit the knowledge of the vertical direction associated to each image into a relative pose estimation algorithm based on lines, as shown in Figure 1. We demonstrate that a closed-form solution requiring only

[1]Banglei Guan is with College of Aerospace Science and Engineering, National University of Defense Technology, China. guanbanglei12@nudt.edu.cn

[2]Pascal Vasseur is with MIS (Modélisation, Information & Systèmes), Université de Picardie Jules Verne, France. pascal.vasseur@u-picardie.fr

[3]Cédric Demonceaux is with ImViA, Université Bourgogne Franche-Comté, France. cedric.demonceaux@u-bourgogne.fr

eight lines between three views is possible rather than the thirteen required in the standard method [23]. The remainder of the paper is organized as follows. First, we review related work in Section II. In Section III, we propose our closed-form solution for three-view relative motion estimation with known vertical direction based on lines and we also study degeneracy cases. We evaluate the performance of our eight lines method using both synthetic and real-world datasets in Section IV. Finally, we conclude with some remarks and comments in Section V.

II. RELATED WORK

Straight lines are particularly well suited for indoor and urban applications where structures are generally man made. In case of low-textured scenarios, the point features are insufficient and the points based pose estimation methods are prone to fail. The pose estimation methods based on line features can serve as ideal complements when there are insufficient point features detected in the scene [34]. In the following text, we focus on the solutions for the relative pose estimation based on points and lines.

Point-based methods: For two-views relative pose estimation, Hartley *et al.* proposed a minimal solver to estimate the fundamental matrix of non-calibrated cameras by using 7 points [1]. Nistér *et al.* further presented a minimal solver to estimate the essential matrix using 5 points when the cameras are calibrated [9]. If the points are coplanar, 4 points is sufficient to estimate the homography matrix [1]. With assumptions about the camera motion, the number of required points can be reduced. Choi *et al.* estimated the planar motion of the cameras by using 2 points [14]. Scaramuzza *et al.* used a single point to recover the relative pose by taking into account the Ackermann motion model [35]. When the vertical direction of the camera can be provided by an IMU or detecting vanishing points, a variety of algorithms utilizing this information have been proposed. Fraundorfer *et al.* proposed a minimal solver to estimate a simplified essential matrix with 3 points [10]. Sweeney *et al.* formulated the relative pose problem with known vertical direction as quadratic eigenvalue problem, and used 3 points to solve directly for relative rotation and translation [11]. For the planar scene, Saurer *et al.* [16] and Guan *et al.* [17] derived several simplified homographies with known vertical direction, and solved them by using a minimum of 2 points and 1.5 points, respectively. Recently, a number of methods exploited the affine parameters between the feature matches and estimated the relative pose with affine correspondences [36], [37], [38], [39], [40], [41].

For trifocal relative pose estimation, Heyden showed that a minimal number of 6 points is sufficient to solve the trifocal relative pose for the uncalibrated case [22]. Ressl proposed the parameterization of the trifocal tensor based on algebraic constraints of the correlation slices, which involves 20 parameters and 2 constraints [42]. Nordberg parameterized the trifocal tensor by three 3×3 orthogonal matrices which transform the original tensor into a sparse one, but only valid for non-collinear centers of three views [43]. Ponce *et al.* introduced a analytical parameterization of trifocal constraints, which yields a minimal parameterization of trinocular geometry for cameras with non-collinear or collinear pinholes [44]. Quan *et al.* showed that the trifocal relative pose has a unique solution with 4 points in general when the cameras are calibrated [45]. Nistér *et al.* parametrized the relative pose between two views as a curve of degree ten which represents possible epipoles, and selected the epipole that minimizes reprojection errors by using a third view [25]. Since the problem about estimating the calibrated trifocal relative pose from 4 points is notably difficult to solve [45], [25], [46], a lot of theoretical work has been studied [47], [48], [49], [50].

Line-based methods: From two views, lines do not provide any constraints on the relative pose estimation [23], [51]. However, for lines in indoor and urban scenes, many lines satisfy the structural information, such as parallelism, orthogonality and coplanarity. By exploiting the constraints of these structural information, a lot of algorithms have been developed. Elqursh *et al.* presented an algorithm for the computation of the relative pose between two views using a minimal number of three lines, of which two of the lines parallel and orthogonal to the third [52]. Li *et al.* leveraged the structural regularity of lines to estimate the relative pose and proposed an two-step method to compute the rotation. They estimated two DOFs of rotation by two image lines whose associated 3D lines are aligned to two Manhattan frame (MF) axes, and then estimate its third DOF by another image line whose associated 3D line is aligned to any MF axis [53]. Zhang assumed that the matched lines are projected from two overlapped 3D lines, and recovered the camera motion from two views by using only line segments [54]. Montiel *et al.* considered the image segment midpoints as correspondent in the image optimization, and recovered the camera motion from straight segment correspondences [55].

For trifocal relative pose estimation, Hartley *et al.* proposed a linear algorithm to estimate the trifocal tensor with measurements of 13 lines in three views [23]. Kuang *et al.* utilized the nonlinear constraints on the trifocal tensor and presented several linear solvers using 10 to 12 lines [56]. Larsson *et al.* proposed a minimal solver to estimate the trifocal relative pose problem of the uncalibrated camera using 9 lines [57]. Geppert *et al.* estimated the relative pose of three gravity oriented views from vertical lines [4].

III. POSE ESTIMATION FROM LINES AND A KNOWN DIRECTION

A. Trifocal Tensor for Lines

As shown in Figure 2, a straight line $\mathbf{L}$ visible in three views gives the following relation [1]:

$$l_i^1 = \mathbf{l}^{2\mathrm{T}}\mathbf{T}_i\mathbf{l}^3 \quad \text{for } i = 1,2,3, \tag{1}$$

where l_i^1 represents the i^{th} coordinate of the straight line $\mathbf{l}^1 = [l_x^1,\ l_y^1,\ l_z^1]^{\mathrm{T}}$ in the view 1, $\mathbf{l}^2$ and $\mathbf{l}^3$ are the coordinates of the same line in the view 2 and view 3, respectively. $\mathbf{T}_i$ represents the i^{th} matrix of the trifocal tensor.

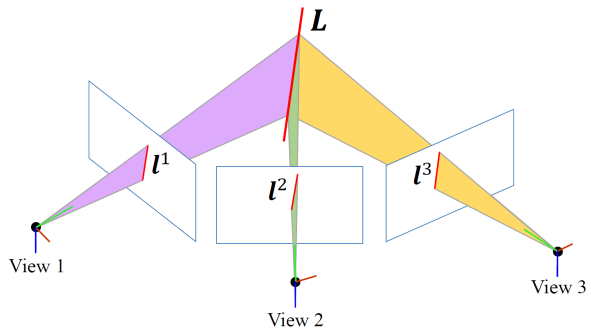


Fig. 2. A straight line $\mathbf{L}$ is imaged as a triplet $\mathbf{l}^1 \leftrightarrow \mathbf{l}^2 \leftrightarrow \mathbf{l}^3$ in three views. Conversely, the planes back-projected from the corresponding image lines in each view all intersect in a single 3D line in space.

In the general case, we also have:

$$\mathbf{T}_i = \mathbf{a}_i \mathbf{b}_4^{\mathrm{T}} - \mathbf{a}_4 \mathbf{b}_i^{\mathrm{T}}, \tag{2}$$

where $\mathbf{a}_4$ and $\mathbf{b}_4$ are respectively the epipoles of the first camera into the second and the third views while $\mathbf{a}_i$ and $\mathbf{b}_i$ are the i^{th} columns of the second and third camera projection matrices. Without loss of generality, if we consider the cameras as intrinsically calibrated and consequently the images as normalized (or expressed on the unitary sphere), the projection matrix $\mathbf{P}_i$ can be expressed only with the extrinsic parameters such as:

$$\mathbf{P}_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R}_i & -\mathbf{R}_i \mathbf{t}_i \\ \mathbf{0} & 1 \end{bmatrix}, \tag{3}$$

where $\mathbf{R}_i$ and $\mathbf{t}_i$ represent the rotations and translations between the first view and the i^{th} view, respectively. The first view is located into an initial reference position such as $\mathbf{R}_1 = \mathbf{I}_{3\times 3}$ and $\mathbf{t}_1 = [0,0,0]^{\mathrm{T}}$.

B. Closed-form Solution

We assume that the attitudes (roll and pitch angles) of the views are known, which are obtained directly from the IMU. The camera coordinate system can be aligned with the known vertical direction. Then the rotation only depends on the yaw angle around the known vertical direction as shown in Figure 2. Consequently, we can write:

$$\mathbf{R}_i = \begin{bmatrix} C_y^i & 0 & S_y^i \\ 0 & 1 & 0 \\ -S_y^i & 0 & C_y^i \end{bmatrix}, \tag{4}$$

where C_y^i and S_y^i correspond respectively to cosine and sine of the yaw angles θ^i of the views ($i = 2,3$). The translation can be represented as $\mathbf{t}_i = [t_x^i,\ t_y^i,\ t_z^i]^{\mathrm{T}}$. Thus, equation (3) equals to:

$$\mathbf{P}_i = \begin{bmatrix} C_y^i & 0 & S_y^i & -C_y^i t_x^i - S_y^i t_z^i \\ 0 & 1 & 0 & -t_y^i \\ -S_y^i & 0 & C_y^i & S_y^i t_x^i - C_y^i t_z^i \end{bmatrix}. \tag{5}$$

If we consider the trifocal tensor equations in Eq. (2), we can deduce the following three equalities:

$$\mathbf{T}_1 = \begin{bmatrix} Q_1 & Q_2 & Q_3 \\ Q_4 & 0 & Q_5 \\ Q_6 & Q_7 & Q_8 \end{bmatrix}, \tag{6}$$

$$\mathbf{T}_2 = \begin{bmatrix} 0 & Q_9 & 0 \\ Q_{10} & Q_{11} & Q_{12} \\ 0 & Q_{13} & 0 \end{bmatrix}, \tag{7}$$

$$\mathbf{T}_3 = \begin{bmatrix} Q_{14} & -Q_7 & Q_{15} \\ -Q_5 & 0 & Q_4 \\ Q_{16} & Q_2 & Q_{17} \end{bmatrix}, \tag{8}$$

with

$$\begin{cases} Q_1 = C_y^3 Q_9 + C_y^2 Q_{10}, & Q_2 = -C_y^2 t_y^3, \\ Q_3 = C_y^2 Q_{12} - S_y^3 Q_9, & Q_4 = C_y^3 t_y^2, \\ Q_5 = -S_y^3 t_y^2, & Q_6 = -S_y^2 Q_{10} + C_y^3 Q_{13}, \\ Q_7 = S_y^2 t_y^3, & Q_8 = -S_y^2 Q_{12} - S_y^3 Q_{13}, \\ Q_9 = C_y^2 t_x^2 + S_y^2 t_z^2, & Q_{10} = -C_y^3 t_x^3 - S_y^3 t_z^3, \\ Q_{11} = t_y^2 - t_y^3, & Q_{12} = S_y^3 t_x^3 - C_y^3 t_z^3, \\ Q_{13} = C_y^2 t_z^2 - S_y^2 t_x^2, & Q_{14} = S_y^3 Q_9 + S_y^2 Q_{10}, \\ Q_{15} = S_y^2 Q_{12} + C_y^3 Q_9, & Q_{16} = S_y^3 Q_{13} + C_y^2 Q_{10}, \\ Q_{17} = C_y^2 Q_{12} + C_y^3 Q_{13}. \end{cases} \tag{9}$$

Equation (1) can then be reformulated as:

$$(\mathbf{l}^{2\mathrm{T}} [\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3] \mathbf{l}^3)[\mathbf{l}^1]_\times = \mathbf{0}, \tag{10}$$

where the symbol $[\,]_\times$ represents the skew-symmetric matrix and consequently. Then we obtain:

$$\begin{bmatrix} l_x^2 & l_y^2 & l_z^2 \end{bmatrix} [\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3] \begin{bmatrix} l_x^3 \\ l_y^3 \\ l_z^3 \end{bmatrix} \begin{bmatrix} 0 & -l_z^1 & l_y^1 \\ l_z^1 & 0 & -l_x^1 \\ -l_y^1 & l_x^1 & 0 \end{bmatrix} = \mathbf{0}. \tag{11}$$

Two independent equations can be extracted from Eq. (11), which means that 8 triplets of lines are sufficient to get the 16 equations required to fully solve the system, up to scale. The equation system can be written as $\mathbf{A}\mathbf{q} = 0$ with $\mathbf{A}$ being a 16×17 matrix and $\mathbf{q}$ a 17×1 vector containing all the entries of the trifocal tensor. The scaling factor is imposed with the constraint $\|\mathbf{q}\| = 1$. All the relative poses can then be retrieved from the estimated parameters of $\mathbf{q}$. Similar to the algorithms in [10], [58], the proposed solver can be used to find a least squared solution to an over-constrained system if more than 8 triplets of lines are used. Based on the formula expression of Q_8 and Q_{14} in Eq. (9), we can compute S_y^2 and S_y^3. Meanwhile, C_y^2 and C_y^3 can be also solved by using the formula expression of Q_1 and Q_{17} in Eq. (9). Thus, the unique solution of the trifocal relative pose with known vertical direction can be computed as follows:

- Yaw angle between views 1 and 2:

$$\begin{cases} S_y^2 = (Q_8 Q_9 + Q_{13} Q_{14})/(Q_{10} Q_{13} - Q_9 Q_{12}) \\ C_y^2 = (Q_1 Q_{13} - Q_9 Q_{17})/(Q_{10} Q_{13} - Q_9 Q_{12}) \\ \theta^2 = \arctan2(S_y^2, C_y^2) \end{cases}$$

- Translation between views 1 and 2:

$$\mathbf{t}_2 = [C_y^2 Q_9 - S_y^2 Q_{13},\ C_y^3 Q_4 - S_y^3 Q_5,\ S_y^2 Q_9 + C_y^2 Q_{13}]^{\mathrm{T}}$$

- Yaw angle between views 1 and 3:

$$\begin{cases} S_y^3 = -(Q_8 + S_y^2 Q_{12})/Q_{13} \\ C_y^3 = (Q_1 - C_y^2 Q_{10})/Q_9 \\ \theta^3 = \arctan2(S_y^3, C_y^3) \end{cases}$$

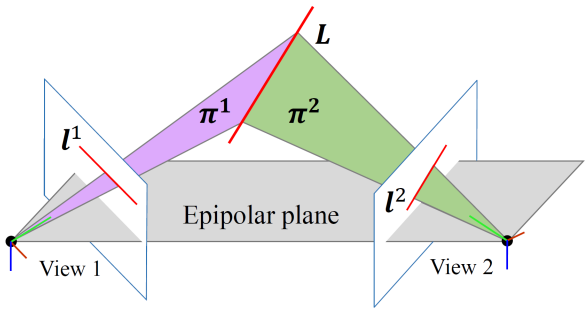


Fig. 3. Degenerate case. When the planes through l^1 and l^2, namely π^1 and π^2, are coincident, i.e. in the case of epipolar planes, the line L in space is clearly undefined in this condition.

- Translation between views 1 and 3:

$$\mathbf{t}_3 = [S_y^3 Q_{12} - C_y^3 Q_{10},\ S_y^2 Q_7 - C_y^2 Q_2,\ -S_y^3 Q_{10} - C_y^3 Q_{12}]^{\mathrm{T}}$$

Finally, the trifocal relative pose between three views can be further recovered by leveraging IMU measurement [41].

C. Degenerate Case

A degenerate case can appear when the planes passing through the 3D line and the camera centers are coincident, i.e. in the case of epipolar planes, see Figure 3. As expressed in [1], while this degenerate configuration could appear between two views, it is very unlikely that this will happen between the three views. In this way, in the extreme majority of cases, there will be a pair of views that can be used for the trifocal tensor. However, in the case of lines in the trifocal plane, the transfer between views is then always degenerate and consequently undefined.

IV. EXPERIMENTS

In this section, we evaluate the performance of the proposed method using both synthetic and real-world data. The proposed solver in Section III is referred to as `8Lines`. The `8Lines` method is compared with state-of-the-art methods, which include the `13Lines-Hartley` method [23], the `7pt-Hartley` method [1], the `5pt-Nister` method [9] and the `3pt-Sweeney` method [11]. The methods `13Lines-Hartley` and `8Lines` estimate the relative motion of three views. The methods `7pt-Hartley`, `5pt-Nister` and `3pt-Sweeney` estimate the relative motion of two views. Since the points and lines are different features and have their own advantages in different environments, the purpose of the following experiments is not to outperform the points based methods. Instead, we illustrate the feasibility and practicality of the proposed method in real scenarios.

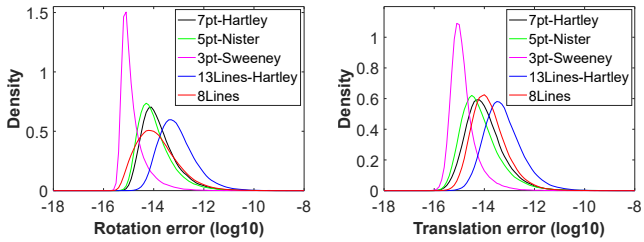
The rotation and translation errors are compared separately in the experiments. The rotation error is computed as the angle difference between the ground truth rotation and the estimated rotation. Since the estimated translation is only known up to scale, the translation error is also computed as the angle difference between the ground truth translation and the estimated translation. Specifically, the errors are computed as follows:

- Rotation error: $\varepsilon_{\mathbf{R}} = \arccos((\mathrm{trace}(\mathbf{R}_{gt}\mathbf{R}^T) - 1)/2)$

TABLE I

RUN-TIME COMPARISON OF RELATIVE POSE SOLVERS (UNIT: *ms*).

Methods	7pt [1]	5pt [9]	3pt [11]	13Lines [23]	**8Lines**
Timings	0.26	0.12	0.18	0.75	0.24



(a) (b)

Fig. 4. Probability density functions over relative pose estimation errors in the noise-free case (100,000 runs). The horizontal axis represents the $\log_{10}$ estimated errors and the vertical axis represents the empirical probability density. (a) reports the rotation estimation error. (b) reports the translation estimation error.

- Translation error: $\varepsilon_{\mathbf{t}} = \arccos((\mathbf{t}_{gt}^T\mathbf{t})/(\|\mathbf{t}_{gt}\| \cdot \|\mathbf{t}\|))$

where $\mathbf{R}_{gt}$ and $\mathbf{R}$ denote the ground truth and estimated rotations, respectively. $\mathbf{t}_{gt}$ and $\mathbf{t}$ denote the ground truth and estimated translations, respectively.

A. Efficiency Comparison and Numerical Stability

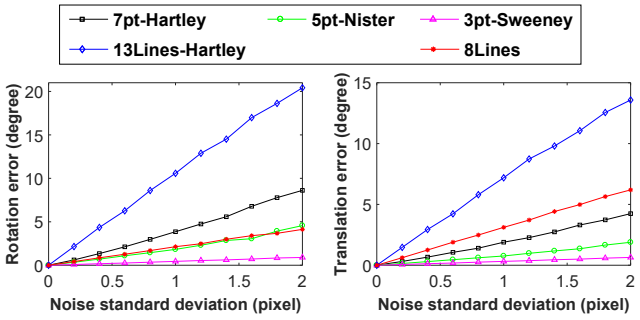
The runtimes of our solver and the comparative solvers are evaluated on an Intel(R) Core(TM) i7-8550U 1.80GHz using MATLAB. All algorithms are implemented in Matlab, except that the `5pt-Nister` method is implemented in C by using mex file. Table I summarizes the average run-times of the solvers over 100,000 runs. The runtime of the `5pt-Nister` method is the lowest, because the mex file is used. The `3pt-Sweeney` method solves the relative pose problem as Quadratic Eigenvalue Problem that is also efficient. It can be seen that the proposed `8Lines` method is more efficient than the methods `7pt-Hartley` and `13Lines-Hartley`.

The numerical stability of the solvers in the noise-free case is shown in Figure 4. We repeat the procedure 100,000 times. The vertical axis represents the empirical probability density functions, which are plotted as the function of the $\log_{10}$ estimated errors. The `3pt-Sweeney` method achieves the best numerical stability. The methods `7pt-Hartley`, `5pt-Nister` and `8Lines` have comparable numerical stability. It can also be seen that the proposed `8Lines` method has better numerical stability than the `13Lines-Hartley` method in both rotation and translation.

B. Experiments on Synthetic Data

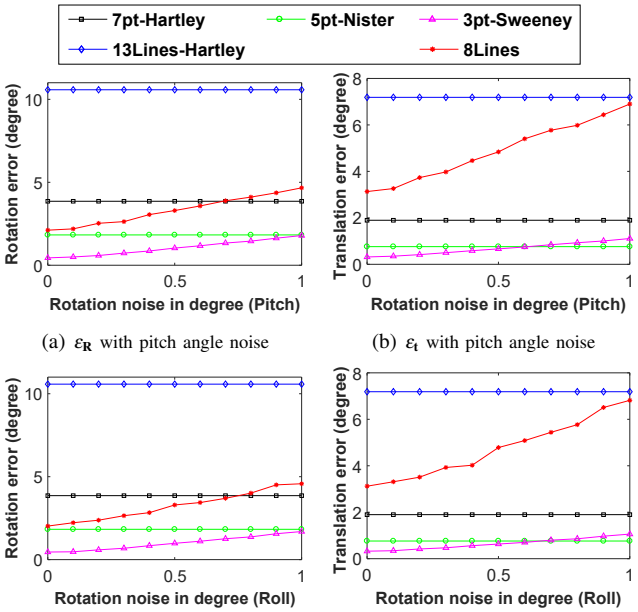
To evaluate the algorithms on synthetic data we choose the following setup. We simulate a monocular perspective camera with a resolution of 640 $\times$ 480 pixels. The focal length of the camera is set to 400 pixels and the principal point is set to (320, 240).

In the synthetic experiments, we evaluate the performance of the proposed `8Lines` method with respect to the image noise and IMU noise. A total of 10000 trials are carried out in per noise step. The rotation and translation errors are



(a) $\varepsilon_{\mathbf{R}}$ with image noise (b) $\varepsilon_{\mathbf{t}}$ with image noise

Fig. 5. Rotation and translation error with varying image noise (unit: degree). The left column reports the rotation error. The right column reports the translation error.



(a) $\varepsilon_{\mathbf{R}}$ with pitch angle noise (b) $\varepsilon_{\mathbf{t}}$ with pitch angle noise

(c) $\varepsilon_{\mathbf{R}}$ with roll angle noise (d) $\varepsilon_{\mathbf{t}}$ with roll angle noise

Fig. 6. Rotation and translation error with varying IMU angle noise (unit: degree). The image noise is set to 1.0 pixel standard deviation. (a)(b): vary pitch angle noise. (c)(d): vary roll angle noise. The left column reports the rotation error. The right column reports the translation error.

assessed by the median of the errors in 10000 trials. In each trial, the direction of the camera motion is set to random motion and three views are captured by the camera. The optical centers of the camera is in a cube of $10\times10\times10$m. The magnitudes of three rotation angles between views vary from -10° to 10°. A set of random 2D pairs of points is generated on the image planes. Each pair defines a 2D line which is at least 70 pixel long [59].

1) Accuracy with the magnitude of image noise: In this scenario, we test the performance of our method for increasing image noise. All tests are run with a set of random image lines in three views, and 10000 tests per noise step are performed. The endpoints of each image line are disturbed by Gaussian noise with a standard deviation ranging from 0 to 2 pixels. This allows to modify both the position and direction of the image lines. The proposed `8Lines` method is compared against the methods `13Lines-Hartley` [23], `7pt-Hartley` [1], `5pt-Nister` [9] and `3pt-Sweeney` [11]. For lines based methods, 8 and 13 image lines in three views are generated for the trifocal relative pose estimation methods `8Lines` and `13Lines-Hartley`, respectively. For points based methods, the 2D endpoints of the image lines are used as input. Thus, 4, 3 and 2 image lines in two views are generated for the methods `7pt-Hartley`, `5pt-Nister` and `3pt-Sweeney`, respectively.

Figure 5 shows the performance of the proposed method with respect to the magnitude of image noise with perfect IMU data. We show the rotation and translation errors, respectively. The `3pt-Sweeney` method has better performance than the other methods. It can be seen that our `8Lines` method performs obviously better than the `13Lines-Hartley` method in both rotation and translation estimation. Moreover, the proposed method also provides better results than the `7pt-Hartley` method in rotation estimation.

2) Accuracy with the magnitude of IMU angle noise: In this set of experiments, we evaluate the performance of our method for increasing IMU angle noise. For the relative pose estimation methods with known vertical direction, the roll angle and pitch angle are assumed to be known with the IMU sensor. Then the camera coordinate system can be aligned with the known vertical direction. However, the roll angle and pitch angle of the IMU sensor are prone to inaccuracies. Currently, the angular accuracy of vertical direction in low cost IMUs is about 0.5°, and in high accuracy IMUs is less than 0.02° [60], [61]. Thus, we add Gaussian noise with a standard deviation ranging from 0° to 1° to both the roll angle and the pitch angle.

Figure 6 shows the performance of the proposed method with respect to the magnitude of IMU noise, while the image noise is set to 1.0 pixel standard deviation. Since the calculation of the methods `13Lines-Hartley`, `7pt-Hartley` and `5pt-Nister` do not use the known vertical direction as prior, these methods are not influenced by the noise of the pitch angle and roll angle. The methods `5pt-Nister` and `3pt-Sweeney` have better performance than the other methods. It is interesting to see that the proposed `8Lines` method performs better than the `13Lines-Hartley` method, even though the noise of the rotation angles is 1.0°. In addition, the accuracy of our method is also better than the `7pt-Hartley` method in rotation estimation, when the rotation angle noise stays below 0.7°.

C. Experiments on Real Data

We evaluate the proposed method on both an autonomous driving and unmanned aerial vehicle environment, which include KITTI dataset [62] and AGZ dataset [63]. There are two popular modern robot applications.

1) Experiments on KITTI Dataset: We evaluate the performance of our `8Lines` method using the KITTI dataset [62] collected on outdoor autonomous vehicles. There are various sequences in the visual odometry section of KITTI dataset.

(a) View 1 (b) View 2 (c) View 3

Fig. 7. Matching lines of triple images on KITTI dataset. There are 69 matching lines in three consecutive images: image 1223 to image 1225.

TABLE II

ROTATION AND TRANSLATION ERROR OVER KITTI DATASET.

Methods	7pt [1]		5pt [9]		3pt [11]		13Lines [23]		**8Lines**	
	ε_R	ε_t	ε_R	ε_t	ε_R	ε_t	ε_R	ε_t	ε_R	ε_t
Median error	0.37°	1.38°	0.26°	1.32°	0.10°	0.75°	1.40°	1.77°	0.22°	0.92°
Mean error	0.50°	1.97°	0.38°	1.56°	0.19°	1.45°	2.01°	2.06°	0.36°	1.03°

The KITTI dataset provides ground truth poses for the sequences, which is directly provided from the output of the built-in GPS/IMU units. A challenging subsequence with two consecutive sharp turns are selected, which starts from image 1223 to image 1276 in the sequence 00 [59]. These 54 images are taken place in an urban environment, which contains a large number of line features. The matching lines of triple images are obtained by LineSfM [64], see Figure 7.

Trifocal relative poses of the subsequence are estimated from these matching lines between three views. The pitch and roll angles provided by the GPS/IMU units are used to obtain the known vertical direction and pre-rotate the matching lines of the images. To ensure the fairness of the experiment, the pitch and roll angles are also provided for the `3pt-Sweeney` method [11], which uses the known vertical direction as a prior. We compute rotation and translation using our proposed `8Lines` method and compare it to the ground truth. The proposed method is also compared with the methods `13Lines` [23], `7pt-Hartley` [1], `5pt-Nister` [9] and `3pt-Sweeney` [11]. All the solvers are integrated within RANSAC to deal with outliers. The median error and mean error for this subsequence are used to evaluate the performance of all the methods. Table II shows the results for the rotation and translation estimation over KITTI dataset. It is shown that the proposed `8Lines` method provides better results than the `13Lines` method, the `7pt-Hartley` method and the `5pt-Nister` method. Even though the `3pt-Sweeney` method outperforms our method, the `8Lines` method has the advantage of robustness over illumination changes. Because the line matching appears more robust than point matching in the practical application.

To visualize the comparison results, Fig. 8 shows the estimated trajectory of the `8Lines` method and the KITTI ground truth poses, as well as the estimated trajectory of `13Lines` method which also estimate the trifocal relative pose based on lines. Since the estimated translation is only known up to scale with a monocular perspective camera, the ground truth scale is used to plot the estimated trajectories. It is to be noted that all the relative poses are independently estimated. The estimated trajectories are plotted by directly concatenating frame-to-frame relative pose measurements, *i.e.* no post-refinement is not applied over the trajectory. Compared with the `13Lines` method, it can be seen that

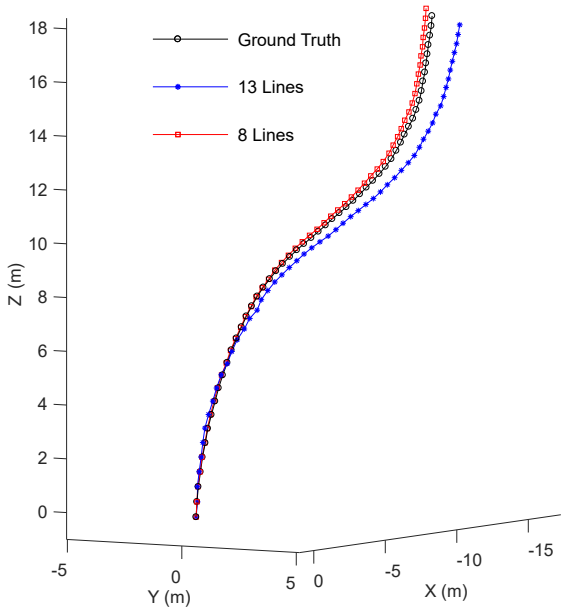
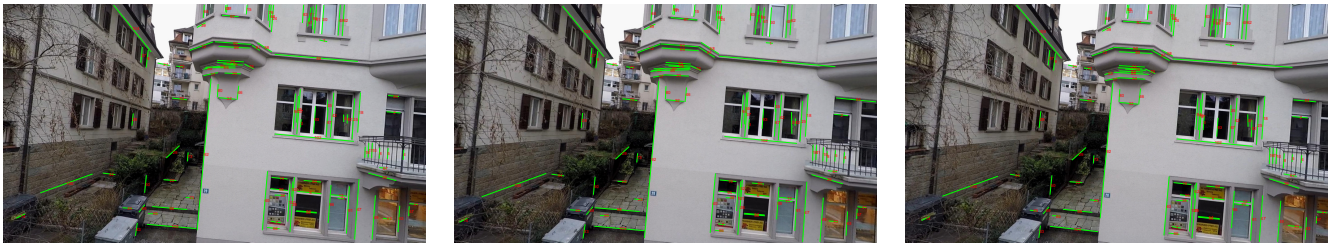


Fig. 8. Comparison between estimated trajectory and ground truth on KITTI dataset. Relative pose measurements between consecutive images are directly concatenated without any post-refinement (Best viewed in color).

the estimated trajectory obtained by our `8Lines` method is more consistent with the ground truth poses.

2) Experiments on AGZ Dataset: To validate the proposed method in unmanned aerial vehicle environment, we test the performance of our method using Zürich Urban Micro Aerial Vehicle (AGZ) dataset [63]. The AGZ dataset is recorded by a camera-equipped UAV, which flies within the urban streets. For the evaluation, we utilize all the available images which have ground truth poses and together consist of 2706 images. The ground truth poses are obtained by using an photogrammetric 3D reconstruction. Fig. 9 shows the matching lines of triple images obtained by LineSfM [64].

Table III shows the results for the rotation and translation estimation over AGZ dataset. The `3pt-Sweeney` method achieves the best accuracy. It is shown that the proposed `8Lines` method provides better results than the `13Lines` method and the `7pt-Hartley` method. Meanwhile, our method performs better than the `5pt-Nister` method in rotation estimation, and has slightly worse performance in translation estimation. Fig. 10 compares the estimated trajectory of the methods `8Lines` and `13Lines` method. Even though both trajectories have a significant accumulation of drift without any post-refinement, it can still be seen that the estimated trajectory of our `8Lines` method is more consistent than the `13Lines` method in comparison with the ground truth poses.



(a) View 1 (b) View 2 (c) View 3

Fig. 9. Matching lines of triple images on AGZ dataset. There are 104 matching lines in three images: 00181, 00211 and 00241.

TABLE III

ROTATION AND TRANSLATION ERRORS OVER AGZ DATASET.

Methods	7pt [1]		5pt [9]		3pt [11]		13Lines [23]		**8Lines**	
	ε_R	ε_t	ε_R	ε_t	ε_R	ε_t	ε_R	ε_t	ε_R	ε_t
Median error	0.78°	3.37°	0.76°	3.29°	0.46°	2.39°	3.86°	5.45°	0.53°	3.31°
Mean error	1.34°	4.13°	1.32°	4.02°	1.15°	3.72°	4.72°	7.46°	1.10°	4.48°

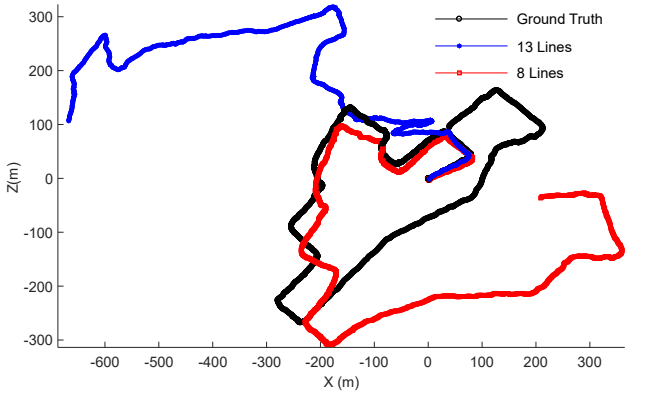


Fig. 10. Comparison between estimated trajectory and ground truth on AGZ dataset. The relative pose measurements between consecutive images are directly concatenated without any post-refinement (Best viewed in color).

V. CONCLUSION

In this paper, we derived the linear solver for the trifocal relative pose with the known vertical direction from line correspondences. We showed that a closed-form solution requiring only eight lines between three views is possible, and gives on to a unique solution. The proposed solver is well-suited for robust estimation with RANSAC framework. We verified our method with both synthetic data and real-world image data sets. The experimental results showed that the proposed method provides better accuracy for trifocal relative pose estimation in comparison to the standard trifocal estimation based on thirteen triplets of lines. Moreover, our method has better or comparable performance than the points based methods. We demonstrated that the proposed method is suitable for the scenes in which self-driving cars and unmanned aerial vehicles operate.

ACKNOWLEDGMENT

This work has been partially funded by the ANR CLARA project ANR-18-CE33-0004 and the National Natural Science Foundation of China (Grant Nos. 11902349 and 11727804).

REFERENCES

[1] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge University Press, 2003.

[2] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An invitation to 3-d vision: from images to geometric models*. Springer Science & Business Media, 2012, vol. 26.

[3] S. Agarwal, H.-L. Lee, B. Sturmfels, and R. R. Thomas, "On the existence of epipolar matrices," *International Journal of Computer Vision*, vol. 121, no. 3, pp. 403–415, 2017.

[4] M. Geppert, V. Larsson, P. Speciale, J. L. Schönberger, and M. Pollefeys, "Privacy preserving structure-from-motion," in *European Conference on Computer Vision*. Springer, 2020, pp. 333–350.

[5] J. Zhao, "An efficient solution to non-minimal case essential matrix estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[6] B. Guan, J. Zhao, D. Barath, and F. Fraundorfer, "Minimal cases for computing the generalized relative pose using affine correspondences," in *IEEE International Conference on Computer Vision*, 2021, pp. 6068–6077.

[7] T. Huang, Y. Zheng, Z. Yu, R. Chen, Y. Li, R. Xiong, L. Ma, J. Zhao, S. Dong, L. Zhu *et al.*, "1000× faster camera and machine vision with ordinary devices," *Engineering*, 2022.

[8] D. Scaramuzza and F. Fraundorfer, "Visual odometry: The first 30 years and fundamentals," *IEEE Robotics & Automation Magazine*, vol. 18, no. 4, pp. 80–92, 2011.

[9] D. Nistér, "An efficient solution to the five-point relative pose problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 756–777, 2004.

[10] F. Fraundorfer, P. Tanskanen, and M. Pollefeys, "A minimal case solution to the calibrated relative pose problem for the case of two known orientation angles," in *European Conference on Computer Vision*. Springer, 2010, pp. 269–282.

[11] C. Sweeney, J. Flynn, and M. Turk, "Solving for relative pose with a partially known rotation is a quadratic eigenvalue problem," in *International Conference on 3D Vision*, 2014, pp. 483–490.

[12] B. Guan, J. Zhao, Z. Li, F. Sun, and F. Fraundorfer, "Relative pose estimation with a single affine correspondence," *IEEE Transactions on Cybernetics*, pp. 1–12, 2021.

[13] D. Ortín and J. M. M. Montiel, "Indoor robot motion based on monocular images," *Robotica*, vol. 19, no. 3, pp. 331–342, 2001.

[14] S. Choi and J.-H. Kim, "Fast and reliable minimal relative pose estimation under planar motion," *Image and Vision Computing*, vol. 69, pp. 103–112, 2018.

[15] C.-C. Chou, Y. Seo, and C.-C. Wang, "A two-stage sampling for robust feature matching," *Journal of Field Robotics*, 2018.

[16] O. Saurer, P. Vasseur, R. Boutteau, C. Demonceaux, M. Pollefeys, and F. Fraundorfer, "Homography based egomotion estimation with a common direction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 2, pp. 327–341, 2016.

[17] B. Guan, P. Vasseur, C. Demonceaux, and F. Fraundorfer, "Visual odometry using a homography formulation with decoupled rotation and translation estimation using minimal solutions," in *IEEE International Conference on Robotics and Automation*, 2018, pp. 2320–2327.

[18] Y. Ding, J. Yang, J. Ponce, and H. Kong, "Homography-based minimal-case relative pose estimation with known gravity direction," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 1, pp. 196–210, 2020.

[19] A. Alzati and A. Tortora, "A geometric approach to the trifocal tensor," *Journal of Mathematical Imaging and Vision*, vol. 38, no. 3, pp. 159–170, 2010.

[20] R. Fabbri and B. B. Kimia, "Multiview differential geometry of curves," *International Journal of Computer Vision*, vol. 120, no. 3, pp. 324–346, 2016.

[21] E. V. Martyushev, "Necessary and sufficient polynomial constraints on compatible triplets of essential matrices," *International Journal of Computer Vision*, vol. 128, no. 12, pp. 2781–2793, 2020.

[22] A. Heyden, "Reconstruction from image sequences by means of relative depths," *International Journal of Computer Vision*, vol. 24, no. 2, pp. 155–161, 1997.

[23] R. Hartley, "A linear method for reconstruction from lines and points," in *IEEE International Conference on Computer Vision*, 1995, pp. 882–887.

[24] R. Fabbri, T. Duff, H. Fan, M. H. Regan, D. d. C. d. Pinho, E. Tsigaridas, C. W. Wampler, J. D. Hauenstein, P. J. Giblin, B. Kimia, A. Leykin, and T. Pajdla, "Trplp - Trifocal relative pose from lines at points," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 073–12 083.

[25] D. Nistér and F. Schaffalitzky, "Four points in two or three calibrated views: Theory and practice," *International Journal of Computer Vision*, vol. 67, no. 2, pp. 211–231, 2006.

[26] T. S. Huang and A. N. Netravali, "Motion and structure from feature correspondences: a review," *Proceedings of the IEEE*, vol. 82, no. 2, pp. 252–268, 1994.

[27] R. J. Holt and A. N. Netravali, "Motion and structure from line correspondences: Some further results," *International Journal of Imaging Systems and Technology*, vol. 5, no. 1, pp. 52–61, 1994.

[28] J. Kileel, "Minimal problems for the calibrated trifocal variety," *SIAM Journal on Applied Algebra and Geometry*, vol. 1, no. 1, pp. 575–598, 2017.

[29] T. Duff, K. Kohn, A. Leykin, and T. Pajdla, "PLMP - point-line minimal problems in complete multi-view visibility," in *IEEE International Conference on Computer Vision*, 2019.

[30] Z. Kukelova, M. Bujnak, and T. Pajdla, "Automatic generator of minimal problem solvers," in *European Conference on Computer Vision*. Springer, 2008, pp. 302–315.

[31] M. Bujnak, Z. Kukelova, and T. Pajdla, "Making minimal solvers fast," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1506–1513.

[32] V. Larsson, M. Oskarsson, K. Aström, A. Wallis, Z. Kukelova, and T. Pajdla, "Beyond Gröbner bases: Basis selection for minimal solvers," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3945–3954.

[33] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[34] H. Li, J. Yao, J.-C. Bazin, X. Lu, Y. Xing, and K. Liu, "A monocular slam system leveraging structural regularity in manhattan world," in *IEEE International Conference on Robotics and Automation*. IEEE, 2018, pp. 2518–2525.

[35] D. Scaramuzza, F. Fraundorfer, and R. Siegwart, "Real-time monocular visual odometry for on-road vehicles with 1-point ransac," in *IEEE International Conference on Robotics and Automation*, 2009, pp. 4293–4299.

[36] J. Bentolila and J. M. Francos, "Conic epipolar constraints from affine correspondences," *Computer Vision and Image Understanding*, vol. 122, pp. 105–114, 2014.

[37] C. Raposo and J. P. Barreto, "Theory and practice of structure-from-motion using affine correspondences," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5470–5478.

[38] I. Eichhardt and D. Chetverikov, "Affine correspondences between central cameras for rapid relative pose estimation," in *European Conference on Computer Vision*, 2018, pp. 482–497.

[39] D. Barath and L. Hajder, "Efficient recovery of essential matrix from two affine correspondences," *IEEE Transactions on Image Processing*, vol. 27, no. 11, pp. 5328–5337, 2018.

[40] L. Hajder and D. Barath, "Relative planar motion for vehicle-mounted cameras from a single affine correspondence," in *IEEE International Conference on Robotics and Automation*, 2020, pp. 8651–8657.

[41] B. Guan, J. Zhao, Z. Li, F. Sun, and F. Fraundorfer, "Minimal solutions for relative pose with a single affine correspondence," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1929–1938.

[42] C. Ressl, "A minimal set of constraints and a minimal parameterization for the trifocal tensor," in *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2002, pp. 277–282.

[43] K. Nordberg, "A minimal parameterization of the trifocal tensor," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 1224–1230.

[44] J. Ponce and M. Hebert, "Trinocular geometry revisited," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 17–24.

[45] L. Quan, B. Triggs, and B. Mourrain, "Some results on minimal euclidean reconstruction from four points," *Journal of Mathematical Imaging and Vision*, vol. 24, no. 3, pp. 341–348, 2006.

[46] R. Fabbri, "Multiview differential geometry in application to computer vision," Ph.D. dissertation, Brown University, 2011.

[47] C. Aholt and L. Oeding, "The ideal of the trifocal variety," *Mathematics of Computation*, vol. 83, no. 289, pp. 2553–2574, 2014.

[48] S. Leonardos, R. Tron, and K. Daniilidis, "A metric parametrization for trifocal tensors with non-colinear pinholes," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 259–267.

[49] E. V. Martyushev, "On some properties of calibrated trifocal tensors," *Journal of Mathematical Imaging and Vision*, vol. 58, no. 2, pp. 321–332, 2017.

[50] L. Oeding, "The quadrifocal variety," *Linear Algebra and Its Applications*, vol. 512, pp. 306–330, 2017.

[51] J. Weng, T. Huang, and N. Ahuja, "Motion and structure from line correspondences; closed-form solution, uniqueness, and optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 3, pp. 318–336, 1992.

[52] A. Elqursh and A. Elgammal, "Line-based relative pose estimation," in *IEEE Conference on Computer Vision and Pattern Recognition*, ser. IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2011, pp. 3049–3056.

[53] H. Li, J. Zhao, J.-C. Bazin, W. Chen, K. Chen, and Y.-H. Liu, "Line-based absolute and relative camera pose estimation in structured environments," in *IEEE International Conference on Intelligent Robots and Systems*. IEEE, 2019, pp. 6914–6920.

[54] Z. Zhang, "Estimating motion and structure from correspondences of line segments between two perspective images," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 17, no. 12, pp. 1129–1139, 1995.

[55] J. Montiel, J. D. Tardós, and L. Montano, "Structure and motion from straight line segments," *Pattern Recognition*, vol. 33, no. 8, pp. 1295–1307, 2000.

[56] Y. Kuang, M. Oskarsson, and K. Aström, "Revisiting trifocal tensor estimation using lines," in *IEEE International Conference on Pattern Recognition*. IEEE, 2014, pp. 2419–2423.

[57] V. Larsson, K. Aström, and M. Oskarsson, "Efficient solvers for minimal problems by syzygy-based reduction," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 820–828.

[58] Y. Ding, J. Yang, and H. Kong, "An efficient solution to the relative pose estimation with a common direction," in *IEEE International Conference on Robotics and Automation*. IEEE, 2020, pp. 11 053–11 059.

[59] L. Lecrosnier, R. Boutteau, P. Vasseur, X. Savatier, and F. Fraundorfer, "Vision based vehicle relocalization in 3d line-feature map using perspective-n-line with a known vertical direction," in *IEEE Intelligent Transportation Systems Conference*. IEEE, 2019, pp. 1263–1269.

[60] Z. Kukelova, M. Bujnak, and T. Pajdla, "Closed-form solutions to minimal absolute pose problems with known vertical direction," in *Asian Conference on Computer Vision*. Springer, 2010, pp. 216–229.

[61] Y. Ding, J. Yang, J. Ponce, and H. Kong, "Minimal solutions to relative pose estimation from two views sharing a common direction with unknown focal length," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7045–7053.

[62] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[63] A. L. Majdik, C. Till, and D. Scaramuzza, "The Zurich urban micro aerial vehicle dataset," *International Journal of Robotics Research*, vol. 36, no. 3, p. 269–273, 2017.

[64] Y. Salaün, R. Marlet, and P. Monasse, "Robust and accurate line- and/or point-based pose estimation without manhattan assumptions," in *European Conference on Computer Vision*. Springer, 2016, pp. 801–818.