Minimal Solutions for Relative Pose with a Single Affine Correspondence

Banglei Guan[1], Ji Zhao[*], Zhang Li[1], Fang Sun[1] and Friedrich Fraundorfer[2,3]

[1]College of Aerospace Science and Engineering, National University of Defense Technology, China

[2]Institute for Computer Graphics and Vision, Graz University of Technology, Austria

[3]Remote Sensing Technology Institute, German Aerospace Center, Germany

guanbanglei12@nudt.edu.cn   zhaoji84@gmail.com   zhangli_nudt@163.com

sunfang@nudt.edu.cn   fraundorfer@icg.tugraz.at

Abstract

In this paper we present four cases of minimal solutions for two-view relative pose estimation by exploiting the affine transformation between feature points and we demonstrate efficient solvers for these cases. It is shown, that under the planar motion assumption or with knowledge of a vertical direction, a single affine correspondence is sufficient to recover the relative camera pose. The four cases considered are two-view planar relative motion for calibrated cameras as a closed-form and a least-squares solution, a closed-form solution for unknown focal length and the case of a known vertical direction. These algorithms can be used efficiently for outlier detection within a RANSAC loop and for initial motion estimation. All the methods are evaluated on both synthetic data and real-world datasets from the KITTI benchmark. The experimental results demonstrate that our methods outperform comparable state-of-the-art methods in accuracy with the benefit of a reduced number of needed RANSAC iterations.

1. Introduction

Simultaneous localization and mapping (SLAM), visual odometry (VO) and Structure-from-Motion (SfM) have been active research topics in computer vision for decades [34, 36]. These technologies have been used successfully in a wide variety of applications and they play an important role in future technologies like autonomous driving. Relative pose estimation from two views is regarded as a fundamental algorithm, which is an essential part of SLAM and SfM pipelines. Thus, improving the accuracy, efficiency and robustness of relative pose estimation algorithms is still of relevant interest [1, 2, 37, 41].

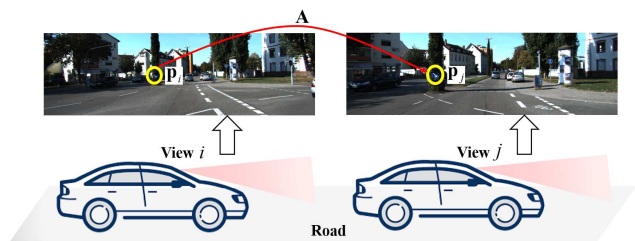Most of the SLAM and SfM pipelines follow the scheme where 2D-2D putative correspondences between subse-

Figure 1. An affine correspondence between two cameras. The local affine transformation $\mathbf{A}$ transforms the patches surrounding of point correspondence $(\mathbf{p}_i, \mathbf{p}_j)$.

quent views are established by feature matching. Then a robust motion estimation framework such as the Random Sample Consensus (RANSAC) [13] is typically adopted to identify and remove matching outliers. Finally, only inlier matches between subsequent views are used to estimate the final relative pose [34]. This outlier removal step is critical for the robustness and reliability of the pose estimation step. Besides, the efficiency of the outlier removal process affects the real-time performance of SLAM and SfM directly, in particular, as the computational complexity of the RANSAC estimator increases exponentially with respect to the number of data points needed. Thus minimal case solutions for relative pose estimation are still of significant importance [5, 2, 42, 11].

The idea of minimal solutions for relative pose estimation ranges back to the work of Hartley and Zisserman with the seven-point method [20]. Other classical works are the five-point method [30] and the homography estimation method [20]. By exploiting motion constraints on camera movements or utilizing an additional sensor like an inertial measurement unit (IMU), the minimal number of point correspondences needed can be further reduced, which makes the outlier removal more efficient and numerically more stable. For instance, two points are sufficient to recover camera motion under the planar motion assumption since the pose only has two degrees of freedom (DOF) [31, 8, 9]. Another

---

[*]Corresponding author.

example is to make use of the Ackermann steering principle which allows us to parameterize the camera motion with only one point correspondence [35, 21]. These scenarios are typical for self-driving vehicles and ground robots. For unmanned aerial vehicles (UAV) and smartphones, a camera is often used in combination with an IMU [17]. The partial IMU measurements can be used to provide a known gravity direction for the camera images. In this case relative pose estimation is thus possible with only three point correspondences. [14, 29, 40, 33].

It is now possible to replace simple point correspondences with affine-covariant feature detectors, such as ASIFT [27] and MODS [26]. Such an affine correspondence (AC) consists of a point correspondence and a 2×2 affine transformation, see Figure 1. It has been proven that 1 AC yields three constraints on the geometric model estimation [7, 32, 3]. In this paper we exploit these additional affine parameters in the process of relative pose estimation which allows to reduce the number of correspondences needed. We propose the following 4 novel minimal solutions for relative pose estimation using a single affine correspondence:

- Three solvers under the planar motion constraint are proposed. We prove that a single affine correspondence is sufficient to recover the planar motion of a calibrated camera (2DOF) and a partially uncalibrated camera for which only the focal length is unknown (3DOF).

- A fourth solver for the case of a known vertical direction is proposed. The egomotion estimation of calibrated camera with a common direction has 3DOF, and we will show that only a single affine correspondence is required to estimate the relative pose for this case.

The remainder of the paper is organized as follows. First we review related work in Section 2. We propose three minimal solutions for planar motion estimation in Section 3. In Section 4, we propose a minimal solution for two-view relative motion estimation with known vertical direction. In Section 5, we evaluate the performance of proposed methods using both synthetic and real-world dataset. Finally, concluding remarks are given in Section 6.

2. Related Work

For non-calibrated cameras, a minimum of 7 point correspondences is sufficient to estimate the fundamental matrix [20]. If the camera is partially uncalibrated such that only the common focal length is unknown, a minimum of 6 point correspondences is required to estimate the relative pose [38, 23]. For calibrated cameras, at least 5 point correspondences are needed to estimate the essential matrix [30]. If all the 3D points lie on a plane, the point correspondences are related by a planar homography and the number of re-

quired point correspondences is reduced to 4 [20]. The relative pose of two views can be recovered by the decomposition of the essential matrix or the homography.

To further improve the computational efficiency and reliability of relative pose estimation, assumptions about the camera motion or additional information can help to reduce the number of required point correspondences across views. For example, if the camera is mounted on ground robots and follows planar motion, the relative pose of two views has only 2DOF and can be estimated by using 2 point correspondences [31, 8, 9]. By taking into account the Ackermann motion model, only 1 point correspondence is sufficient to recover the camera motion [35].

When additional information can be provided by an additional sensor, such as an IMU, the DOF of relative pose estimation can also be reduced. If the rotation of the camera is fully provided by an IMU, only the translation of two views is unknown and can easily be solved with 2 point correspondences [22]. It is more often the case that a common direction of rotation is assumed to be known. This common direction can be determined from an IMU (which provides the known pitch and roll angles of the camera), but as well from vanishing points extracted across the two views. When the common direction of rotation is known, a variety of algorithms have been proposed to estimate the relative pose utilizing this information [14, 29, 40, 33, 16, 10].

Recently, a number of methods have been proposed which reduce the number of required points by exploiting the additional affine parameters between two feature matches. These additional information can come from the feature's rotation and scale estimates when SIFT [25] or SURF [6] feature detectors are used. From five such point correspondences extended by the rotational angles of the features the fundamental matrix can be computed [2]. Similarly, the homography can be estimated by using two correspondences when including the corresponding rotational angles and scales of the features [4]. Of high interest are methods which use affine correspondences obtained by an affine-covariant feature detector, such as ASIFT [27] and MODS [26]. One AC yields three constraints on the geometric model estimation. This allows the estimation of a fundamental matrix from 3 ACs [7]. The estimation of a homography and an essential matrix can be accomplished from 2 ACs [32, 12, 3]. There is an independent work which also uses a single AC to estimate relative planar motion [18]. Furthermore, it is shown in [32] that ACs have benefits as compared to point correspondences for visual odometry in the presence of many outliers.

3. Relative Pose Estimation Under Planar Motion

For planar motion shown in Figure 2, we derive three minimal solvers by exploiting one affine correspondence

only. (1) We develop two minimal solvers for calibrated cameras. Since one AC provides three independent equations and there are two unknowns for the pose, the equation system is over-determined. We propose two variants for this scenario including a closed-form solution and a least-squares solution. (2) For uncalibrated cameras with unknown focal length only, we propose a minimal solver for this scenario as well.
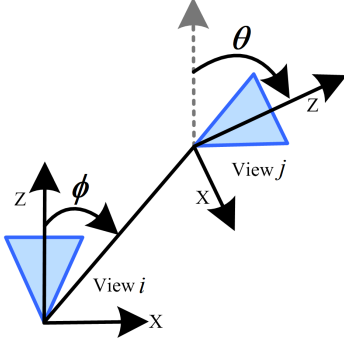

Figure 2. Planar motion between two cameras in top-view. There are two unknowns: yaw angle θ and translation direction ϕ.

3.1. Solver for Planar Motion with Calibrated Camera

With known intrinsic camera parameters, the epipolar constraint between views i to j is given as follows [20]:

$$\mathbf{p}_j^T \mathbf{E} \mathbf{p}_i = 0, \tag{1}$$

where $\mathbf{p}_i = [u_i, v_i, 1]^T$ and $\mathbf{p}_j = [u_j, v_j, 1]^T$ are the normalized homogeneous image coordinates of a feature point in views i and j, respectively. $\mathbf{E} = [\mathbf{t}]_\times \mathbf{R}$ is the essential matrix, where $\mathbf{R}$ and $\mathbf{t}$ represent relative rotation and translation respectively.

For planar motion, we assume that the image plane of the camera is vertical to the ground plane without loss of generality, see Figure 2. There are only a Y-axis rotation and 2D translation between two different views, so the rotation $\mathbf{R} = \mathbf{R}_y$ and the translation $\mathbf{t}$ from views i to j can be written as:

$$\mathbf{R}_y = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix}, \ \mathbf{t} = -\mathbf{R}_y \begin{bmatrix} \rho\sin(\phi) \\ 0 \\ \rho\cos(\phi) \end{bmatrix}. \tag{2}$$

where ρ is the distance between views i and j. Based on Eq. (2), the essential matrix $\mathbf{E} = [\mathbf{t}]_\times \mathbf{R}_y$ under planar motion is reformulated:

$$\mathbf{E} = \rho \begin{bmatrix} 0 & \cos(\theta-\phi) & 0 \\ -\cos(\phi) & 0 & \sin(\phi) \\ 0 & \sin(\theta-\phi) & 0 \end{bmatrix}. \tag{3}$$

By substituting the above equation into Eq. (1), the epipolar constraint can be written as:

$$v_i\sin(\theta-\phi)+v_iu_j\cos(\theta-\phi)+v_j\sin(\phi)-u_iv_j\cos(\phi) = 0. \tag{4}$$

Moreover, the widely-used affine-covariant feature detectors, *e.g.* ASIFT [27], provide affine correspondences between two views directly. Here, we exploit the affine transformation in the relative pose estimation under planar motion, to further reduce the number of required point correspondences. Firstly, we introduce the affine correspondence, which is considered as a triplet: $(\mathbf{p}_i, \mathbf{p}_j, \mathbf{A})$. The local affine transformation $\mathbf{A}$ which relates the patches surrounding $\mathbf{p}_i$ and $\mathbf{p}_j$ is defined as follows [2]:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}. \tag{5}$$

The relationship of essential matrix $\mathbf{E}$ and local affine transformation $\mathbf{A}$ can be described as follows [3]

$$(\mathbf{E}^T\mathbf{p}_j)_{(1:2)} = -(\hat{\mathbf{A}}^T\mathbf{E}\mathbf{p}_i)_{(1:2)}, \tag{6}$$

where $\mathbf{n}_i \triangleq \mathbf{E}^T\mathbf{p}_j$ and $\mathbf{n}_j \triangleq \mathbf{E}\mathbf{p}_i$ are the epipolar lines in the views i and j, respectively. $\hat{\mathbf{A}}$ is a 3×3 matrix:

$$\hat{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}. \tag{7}$$

By substituting Eq. (3) into Eq. (6), two equations which relate the affine transformation to the relative pose are obtained

$$a_{11}v_i\cos(\theta-\phi) + a_{21}\sin(\phi) - (a_{21}u_i + v_j)\cos(\phi) = 0, \tag{8}$$

$$\sin(\theta-\phi) + (a_{12}v_i + u_j)\cos(\theta-\phi) + a_{22}\sin(\phi)- $$
$$a_{22}u_i\cos(\phi) = 0. \tag{9}$$

3.1.1 Closed-Form Solution

For an affine correspondence, the combination of Eqs. (4), (8) and (9) can be expressed as $\mathbf{Cx} = \mathbf{0}$, where $\mathbf{x} = [\sin(\theta-\phi), \ \cos(\theta-\phi), \ \sin(\phi), \ \cos(\phi)]^T$. To facilitate the description of the following method, we denote

$$\begin{cases} x_1 \triangleq \sin(\theta-\phi), & x_2 \triangleq \cos(\theta-\phi) \\ x_3 \triangleq \sin(\phi), & x_4 \triangleq \cos(\phi) \end{cases} \tag{10}$$

By ignoring the implicit constraints between the entries of $\mathbf{x}$, *i.e.*, $x_1^2 + x_2^2 = 1$ and $x_3^2 + x_4^2 = 1$, $\mathbf{x}$ should lie in the null space of $\mathbf{C}$. Thus the solution of the system $\mathbf{x}$ can be obtained directly based on the eigenvector of $\mathbf{C}^T\mathbf{C}$ corresponding to the least eigenvalue. Once $\mathbf{x}$ has been obtained, the angles ϕ and θ are

$$\begin{cases} \phi = \arctan2(x_3, x_4), \\ \theta = \arctan2(x_1, x_2) + \phi. \end{cases} \tag{11}$$

3.1.2 Least-Squares Solution

Eqs. (4), (8) and (9) together with the implicit constraints of the trigonometric functions can be reformulated as:

$$\begin{cases} a_ix_1 + b_ix_2 + c_ix_3 + d_ix_4 = 0, \ i = 1, 2, 3 \\ x_1^2 + x_2^2 = 1 \\ x_3^2 + x_4^2 = 1 \end{cases} \tag{12}$$

The coefficients a_i, b_i, c_i and d_i denote the problem coefficients in Eqs. (4), (8) and (9). This equation system has

4 unknowns and 5 independent constraints, thus it is over-constrained. We find the least-squared solution by

$$\min_{\{x_i\}_{i=1}^4} \sum_{i=1}^{3}(a_i x_1 + b_i x_2 + c_i x_3 + d_i x_4)^2 \quad (13)$$
$$\text{s.t. } x_1^2 + x_2^2 = 1, \;\; x_3^2 + x_4^2 = 1.$$

The Lagrange multiplier method is used to find all stationary points in problem (13). The Lagrange multiplier is

$$L(x_1, x_2, x_3, x_4, \lambda_1, \lambda_2)$$
$$= \sum_{i=1}^{3}(a_i x_1 + b_i x_2 + c_i x_3 + d_i x_4)^2$$
$$+ \lambda_1(x_1^2 + x_2^2 - 1) + \lambda_2(x_3^2 + x_4^2 - 1). \quad (14)$$

By taking the partial derivatives with $\{x_i\}_{i=1}^4$ and $\{\lambda_i\}_{i=1}^2$ and setting them to be zeros, we obtain an equation system with unknowns $\{x_i\}_{i=1}^4$ and $\{\lambda_i\}_{i=1}^2$, see the supplementary material. This equation system contains 6 unknowns $\{x_1, x_2, x_3, x_4, \lambda_1, \lambda_2\}$, and the order is 2. A Gröbner basis solver with template size 42×50 can be obtained by an automatic solver generator [24]. It also shows that there are at most 8 solutions.

3.2. Solver for Planar Motion and Unknown Focal Length

In this subsection, we assume that there is a camera with known intrinsic parameters except for an unknown focal length. This case is typical to be encounter in practice. For most cameras, it is often reasonable to assume that the cameras have square-shaped pixels and the principal point is well approximated by the image center [19]. By assuming that the only unknown calibration parameter of the camera is the focal length f, the intrinsic matrix of the camera is simplified to $\mathbf{K} = \mathrm{diag}(f, f, 1)$.

Since the intrinsic matrix is unknown, we can not obtain the coordinates of point features in the normalized image plane. Recall that the normalized homogeneous image coordinates of the points in views i and j are $\mathbf{p}_i = [u_i, v_i, 1]^T$ and $\mathbf{p}_j = [u_j, v_j, 1]^T$, respectively. Without loss of generality, we set the principle point as the centre of image plane. Denote coordinates of a point in original image plane i and j as $\bar{\mathbf{p}}_i = [\bar{u}_i, \bar{v}_i, 1]^T$ and $\bar{\mathbf{p}}_j = [\bar{u}_j, \bar{v}_j, 1]^T$, respectively. We also denote $g = f^{-1}$ and obtain the following relations

$$\begin{cases} u_i = f^{-1}\bar{u}_i = g\bar{u}_i, & v_i = f^{-1}\bar{v}_i = g\bar{v}_i, \\ u_j = f^{-1}\bar{u}_j = g\bar{u}_j, & v_j = f^{-1}\bar{v}_j = g\bar{v}_j. \end{cases} \quad (15)$$

By substituting Eq. (15) into Eqs. (4), (8) and (9), we also obtain three equations. To reduce the burden in notation, we substitute Eq. (10) into the three equations. By combining them with two trigonometric constraints, we have a polynomial equation system as follows

$$\begin{cases} \bar{v}_i g x_1 + \bar{v}_i \bar{u}_j g^2 x_2 + \bar{v}_j g x_3 - \bar{u}_i \bar{v}_j g^2 x_4 & = 0 \\ a_1 \bar{v}_i g x_2 + a_3 x_3 - (a_3 \bar{u}_i + \bar{v}_j) g x_4 & = 0 \\ x_1 + (a_2 \bar{v}_i + \bar{u}_j) g x_2 + a_4 x_3 - a_4 \bar{u}_i g x_4 & = 0 \\ x_1^2 + x_2^2 = 1 \\ x_3^2 + x_4^2 = 1 \end{cases} \quad (16)$$

The above equation system contains 5 unknowns $\{x_1, x_2, x_3, x_4, g\}$, and the order is 3. The Gröbner basis solver with template size 20×23 can be obtained by an automatic solver generator [24]. It also shows that there are at most 6 solutions. Note that one trivial solution ($g = x_1 = x_3 = 0$, $x_2 = x_4 = 1$) can be safely removed considering $g = f^{-1}$ must be greater than 0.

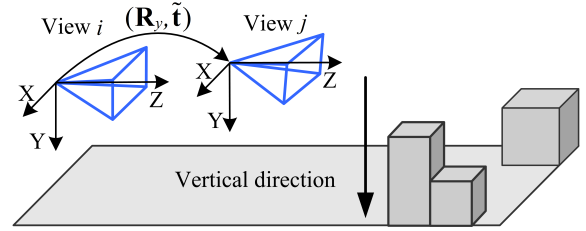## 4. Relative Pose Estimation with Known Vertical Direction

Figure 3. Camera motion with known vertical direction. The unknowns include yaw angle θ and translation $[t_x, \; t_y, \; t_z]^T$.

In this section we present a minimal solution for two-view relative motion estimation with known vertical direction, which uses only one affine correspondence, see Figure 3. In this case, we have an IMU coupled with the camera. Assuming the roll and pitch angles of the camera can be obtained directly from the IMU, we can align every camera coordinate system with the measured gravity direction. The Y-axis of the camera is parallel to the gravity direction and the X-Z-plane of the camera is orthogonal to the gravity direction. The rotation matrix $\mathbf{R}_{\mathrm{imu}}$ for aligning the camera coordinate system to the aligned camera coordinate system can be expressed:

$$\mathbf{R}_{\mathrm{imu}} = \mathbf{R}_x \mathbf{R}_z$$
$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta_x) & \sin(\theta_x) \\ 0 & -\sin(\theta_x) & \cos(\theta_x) \end{bmatrix} \begin{bmatrix} \cos(\theta_z) & \sin(\theta_z) & 0 \\ -\sin(\theta_z) & \cos(\theta_z) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

where θ_x and θ_z represent pitch and roll angle, respectively.

Furthermore, denote $\mathbf{R}_{\mathrm{imu}}^i$ and $\mathbf{R}_{\mathrm{imu}}^j$ as the orientation information delivered by the IMU for views i and j, respectively. Then the aligned image coordinates in views i and j can be expressed by

$$\tilde{\mathbf{p}}_i = \mathbf{R}_{\mathrm{imu}}^i \mathbf{p}_i, \qquad \tilde{\mathbf{p}}_j = \mathbf{R}_{\mathrm{imu}}^j \mathbf{p}_j. \quad (17)$$

By leveraging IMU measurement, the relative pose between original views i and j can be written as

$$\begin{cases} \mathbf{R} = (\mathbf{R}_{\text{imu}}^j)^T \mathbf{R}_y \mathbf{R}_{\text{imu}}^i, \\ \mathbf{t} = (\mathbf{R}_{\text{imu}}^j)^T \tilde{\mathbf{t}}. \end{cases} \tag{18}$$

where $\mathbf{R}_y$ is the rotation matrix between the aligned views i and j, and $\tilde{\mathbf{t}}$ is the translation between the aligned views i and j. Then the essential matrix between the original views i and j can be described as follows

$$\begin{aligned} \mathbf{E} &= [\mathbf{t}]_\times \mathbf{R} = [(\mathbf{R}_{\text{imu}}^j)^T \tilde{\mathbf{t}}]_\times (\mathbf{R}_{\text{imu}}^j)^T \mathbf{R}_y \mathbf{R}_{\text{imu}}^i \\ &= (\mathbf{R}_{\text{imu}}^j)^T \tilde{\mathbf{E}} \mathbf{R}_{\text{imu}}^i. \end{aligned} \tag{19}$$

Note that $\tilde{\mathbf{E}} = [\tilde{\mathbf{t}}]_\times \mathbf{R}_y$ denotes the simplified essential matrix between the aligned views i and j. Now, we substitute Eq. (19) into Eq. (6):

$$((\mathbf{R}_{\text{imu}}^i)^T \tilde{\mathbf{E}}^T \mathbf{R}_{\text{imu}}^j \mathbf{p}_j)_{(1:2)} = -(\hat{\mathbf{A}}^T (\mathbf{R}_{\text{imu}}^j)^T \tilde{\mathbf{E}} \mathbf{R}_{\text{imu}}^i \mathbf{p}_i)_{(1:2)}. \tag{20}$$

The above equation can be reformulated based on Eq. (17):

$$((\mathbf{R}_{\text{imu}}^i)^T \tilde{\mathbf{E}}^T \tilde{\mathbf{p}}_j)_{(1:2)} = -(\hat{\mathbf{A}}^T (\mathbf{R}_{\text{imu}}^j)^T \tilde{\mathbf{E}} \tilde{\mathbf{p}}_i)_{(1:2)}. \tag{21}$$

For further derivation, we denote $\tilde{\mathbf{p}}_i$, $\tilde{\mathbf{p}}_j$, $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{A}}$ as follows

$$\tilde{\mathbf{p}}_i \triangleq [\tilde{u}_i, \tilde{v}_i, \tilde{w}_i]^T, \qquad \tilde{\mathbf{p}}_j \triangleq [\tilde{u}_j, \tilde{v}_j, \tilde{w}_j]^T$$

$$\begin{aligned} \tilde{\mathbf{E}} = [\tilde{\mathbf{t}}]_\times \mathbf{R}_y &= \begin{bmatrix} 0 & -\tilde{t}_z & \tilde{t}_y \\ \tilde{t}_z & 0 & -\tilde{t}_x \\ -\tilde{t}_y & \tilde{t}_x & 0 \end{bmatrix} \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \\ &= \begin{bmatrix} \tilde{t}_y \sin(\theta) & -\tilde{t}_z & \tilde{t}_y \cos(\theta) \\ \tilde{t}_z \cos(\theta) - \tilde{t}_x \sin(\theta) & 0 & -\tilde{t}_x \cos(\theta) - \tilde{t}_z \sin(\theta) \\ -\tilde{t}_y \cos(\theta) & \tilde{t}_x & \tilde{t}_y \sin(\theta) \end{bmatrix} \\ &\triangleq \begin{bmatrix} e_1 & e_2 & e_3 \\ e_4 & 0 & e_5 \\ -e_3 & e_6 & e_1 \end{bmatrix} \end{aligned}$$

$$\tilde{\mathbf{A}} = \hat{\mathbf{A}}^T (\mathbf{R}_{\text{imu}}^j)^T \triangleq \begin{bmatrix} \tilde{a}_1 & \tilde{a}_2 & \tilde{a}_3 \\ \tilde{a}_4 & \tilde{a}_5 & \tilde{a}_6 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{R}_{\text{imu}}^i = \mathbf{R}_x^i \mathbf{R}_z^i \triangleq \begin{bmatrix} \tilde{r}_1 & \tilde{r}_2 & 0 \\ \tilde{r}_3 & \tilde{r}_4 & \tilde{r}_5 \\ \tilde{r}_6 & \tilde{r}_7 & \tilde{r}_8 \end{bmatrix} \tag{22}$$

By substituting Eq. (22) into Eq. (21), we obtain two equations

$$\begin{aligned} &(\tilde{u}_i \tilde{a}_1 + \tilde{w}_i \tilde{a}_3 + \tilde{u}_j \tilde{r}_1 + \tilde{w}_j \tilde{r}_6) e_1 + (\tilde{v}_i \tilde{a}_1 + \tilde{u}_j \tilde{r}_3) e_2 + \\ &(\tilde{w}_i \tilde{a}_1 + \tilde{u}_j \tilde{r}_6 - \tilde{u}_i \tilde{a}_3 - \tilde{w}_j \tilde{r}_1) e_3 + (\tilde{u}_i \tilde{a}_2 + \tilde{v}_j \tilde{r}_1) e_4 + \\ &(\tilde{w}_i \tilde{a}_2 + \tilde{v}_j \tilde{r}_6) e_5 + (\tilde{v}_i \tilde{a}_3 + \tilde{w}_j \tilde{r}_3) e_6 = 0, \end{aligned} \tag{23}$$

$$\begin{aligned} &(\tilde{u}_i \tilde{a}_4 + \tilde{w}_i \tilde{a}_6 + \tilde{u}_j \tilde{r}_2 + \tilde{w}_j \tilde{r}_7) e_1 + (\tilde{v}_i \tilde{a}_4 + \tilde{u}_j \tilde{r}_4) e_2 + \\ &(\tilde{w}_i \tilde{a}_4 - \tilde{u}_i \tilde{a}_6 + \tilde{u}_j \tilde{r}_7 - \tilde{w}_j \tilde{r}_2) e_3 + (\tilde{u}_i \tilde{a}_5 + \tilde{v}_j \tilde{r}_2) e_4 + \\ &(\tilde{w}_i \tilde{a}_5 + \tilde{v}_j \tilde{r}_7) e_5 + (\tilde{v}_i \tilde{a}_6 + \tilde{w}_j \tilde{r}_4) e_6 = 0. \end{aligned} \tag{24}$$

In addition, the epipolar constraint $\tilde{\mathbf{p}}_j^T \tilde{\mathbf{E}} \tilde{\mathbf{p}}_i = 0$ can be written as:

$$\begin{aligned} &(\tilde{u}_i \tilde{u}_j + \tilde{w}_i \tilde{w}_j) e_1 + \tilde{u}_j \tilde{v}_i e_2 + (\tilde{u}_j \tilde{w}_i - \tilde{u}_i \tilde{w}_j) e_3 \\ &\quad + \tilde{u}_i \tilde{v}_j e_4 + \tilde{v}_j \tilde{w}_i e_5 + \tilde{v}_i \tilde{w}_j e_6 = 0. \end{aligned} \tag{25}$$

For an affine correspondence $(\mathbf{p}_i, \mathbf{p}_j, \mathbf{A})$, the combination of equations Eqs. (23)$\sim$(25) can be expressed as $\mathbf{M}\mathbf{x} = \mathbf{0}$, where $\mathbf{x} = [e_1, \ e_2, \ e_3, \ e_4, \ e_5, \ e_6]^T$ is the vector of the unknown elements of the essential matrix. The null space of $\mathbf{M}$ is 3-dimensional. The solution of the polynomial equation system $\mathbf{x}$, which is up to a common scale, can be determined by the linear combination of three null space basis vectors:

$$\mathbf{x} = \beta \mathbf{m}_1 + \gamma \mathbf{m}_2 + \mathbf{m}_3, \tag{26}$$

where the null space basis vectors $\{\mathbf{m}_i\}_{i=1,2,3}$ are computed from the SVD of matrix $\mathbf{M}$, and β and γ are the coefficients.

To determine the coefficients of β and γ, note that there are two internal constraints for the essential matrix, i.e., the singularity of the essential matrix and the trace constraint:

$$\det(\tilde{\mathbf{E}}) = 0, \tag{27}$$

$$2\tilde{\mathbf{E}} \tilde{\mathbf{E}}^T \tilde{\mathbf{E}} - \text{trace}(\tilde{\mathbf{E}} \tilde{\mathbf{E}}^T) \tilde{\mathbf{E}} = 0. \tag{28}$$

By substituting Eq. (26) into Eqs. (27) and (28), a polynomial equation system with unknowns β and γ can be generated. A straightforward method to solve the equation system is using a general automatic solver generator [24]. Inspired by [14], we use a more simpler method to convert the equation system to a univariate quartic equation, see supplementary material for details. Once the coefficients β and γ have been obtained, the simplified essential matrix $\tilde{\mathbf{E}}$ is determined by Eq. (26) and can be decomposed into $\mathbf{R}_y$ and $\tilde{\mathbf{t}}$ by exploiting Eq. (22). Finally, the relative pose between views i and j can be obtained by Eq. (18).

5. Experiments

The performance of the proposed methods is evaluated using both synthetic and real scene data. To deal with outliers, the minimal solvers can be integrated into a robust estimator using RANSAC or used for histogram voting. For the RANSAC, the relative pose which produces the highest number of inliers is chosen. For the histogram voting, we estimate the relative pose by selecting the peak of the histogram, which is formed by estimating poses from all the affine correspondences.

For relative pose estimation under planar motion, the proposed solvers in Section 3.1 are referred to as `1AC-Voting` (which uses histogram voting with the closed-form solution), `1AC-CS` (which uses RANSAC with the closed-form solution), and `1AC-LS` (which uses RANSAC with the least-squares solution). The solver for planar motion with unknown focal length in Section 3.2 is referred to as the `1AC-UnknownF`, which also uses RANSAC. The comparative methods include `6pt-Kukelova`[1] [23], `5pt-Nister` [30], `2AC-Barath` [3] and `2pt-Choi` [8]. All comparative methods are integrated into a RANSAC scheme.

---

[1] f+E+f relative pose solver.

For relative pose estimation with known vertical direction, our solver proposed in Section 4 is referred to as the `1AC` method. The proposed solver is compared against `5pt-Nister` [30], `3pt-Sweeney` [40], `3pt-Saurer` [33], `2pt-Saurer` [33] and `2AC-Barath` [3]. All of these minimal solvers are integrated into a RANSAC scheme.

Due to space limit, the efficiency comparison is provided in supplementary material. To demonstrate the suitability of our methods in real scenarios, the `KITTI` dataset [15] is used to validate the performance.

5.1. Experiments on Synthetic Data

The synthetic scene consists of a ground plane and 50 random planes, which are randomly distributed in the range of -5 to 5 meters (X-axis direction), -5 to 5 meters (Y-axis direction), and 10 to 20 meters (Z-axis direction). 50 points are randomly generated in the ground plane. We choose a point in each random plane randomly, so there are also 50 points in the random planes. The corresponding affine transformation related to each point correspondence is calculated from the homography, which is estimated by using four projected image points from the same plane [4]. The baseline between two views is set to be 2 meters. The resolution of the camera is 640×480 pixels. The focal length is set to 400 pixels and the principal point is set to (320, 240).

The rotation and translation error are assessed by the root mean square error (RMSE) of the errors. We report the results on the data points within the first two intervals of a 5-quantile partitioning[2] (Quintile) of 1000 trials. The relative rotation and translation between views i and j are compared separately in the synthetic experiments. The rotation error compares the angular difference between the ground truth rotation and the estimated rotation. The translation error also compares the angular difference between the ground truth translation and the estimated translation since the estimated translation between views i and j is only known up to scale. Specifically, we define:

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

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

where $\mathbf{R}_{gt}$ and $\mathbf{t}_{gt}$ denote the ground truth rotation and translation, respectively. $\mathbf{R}$ and $\mathbf{t}$ denote the corresponding estimated rotation and translation, respectively.

5.1.1 Planar Motion Estimation

In this scenario the motion of the camera is described by (θ, ϕ), see Figure 2. Both angles vary from $-10°$ to $10°$. Figure 4(a) and (b) show the performance of the proposed methods with respect to the magnitude of added image noise. All of our proposed methods for planar mo-

---

[2]k-quantiles divide an ordered dataset into k regular intervals

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

(c) $\varepsilon_{\mathbf{R}}$ with non-planar motion noise   (d) $\varepsilon_{\mathbf{t}}$ with non-planar motion noise
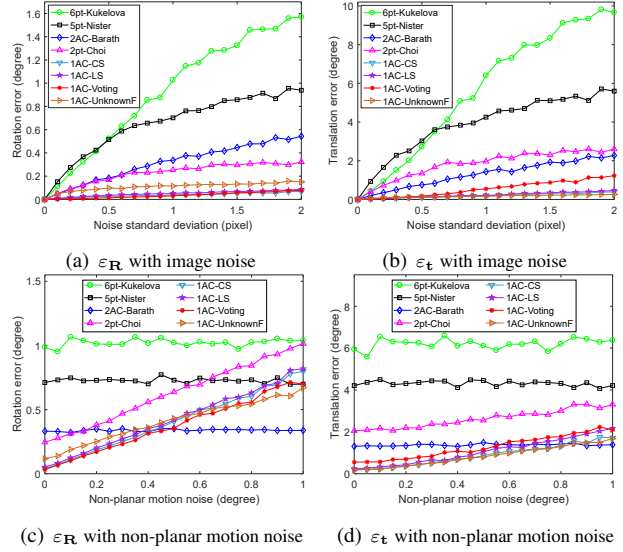
Figure 4. Rotation and translation error with planar motion estimation (unit: degree). (a)(b): vary image noise under perfect planar motion. (c)(d): vary non-planar motion noise and fix the image noise as 1.0 pixel standard deviation. The left column reports the rotation error. The right column reports the translation error.

tion provide better results than comparative methods under perfect planar motion. It is worth to mention that our `1AC-UnknownF` method performs better than comparative methods even when the ground truth of the focal length is not used.

To test the performance of our method under non-planar motion, we generate the non-planar components of a 6DOF relative pose randomly and add them to the camera motion, which include X-axis rotation, Z-axis rotation, and direction of YZ-plane translation [8]. The magnitude of non-planar motion noise is set to Gaussian noise with a standard deviation ranging from $0°$ to $1°$. The image noise is set to 1.0 pixel standard deviation. Figure 4(c) and (d) show the performance of the proposed methods with respect to the magnitude of non-planar motion noise. The methods `6pt-Kukelova`, `5pt-Nister` and `2AC-Barath` do not have an obvious trend with non-planar motion noise levels, because these methods estimate 6DOF relative pose of two views. The proposed four methods perform better than the methods `6pt-Kukelova`, `5pt-Nister` and `2pt-Choi` at the maximum magnitude for the non-planar motion noise up to $1.0°$. Meanwhile, the accuracy of these four methods is also better than the `2AC-Barath` method when the non-planar motion noise is less than $0.3°$.

5.1.2 Motion with Known Vertical Direction

In this set of experiments the directions of the camera motion are set to forward, sideways and random motions, respectively. The second view is rotated around every axis, three rotation angles vary from $-10°$ to $10°$. The roll angle and pitch angle are known and

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

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

(c) $\varepsilon_{\mathbf{R}}$ with pitch angle noise

(d) $\varepsilon_{\mathbf{t}}$ with pitch angle noise

(e) $\varepsilon_{\mathbf{R}}$ with roll angle noise
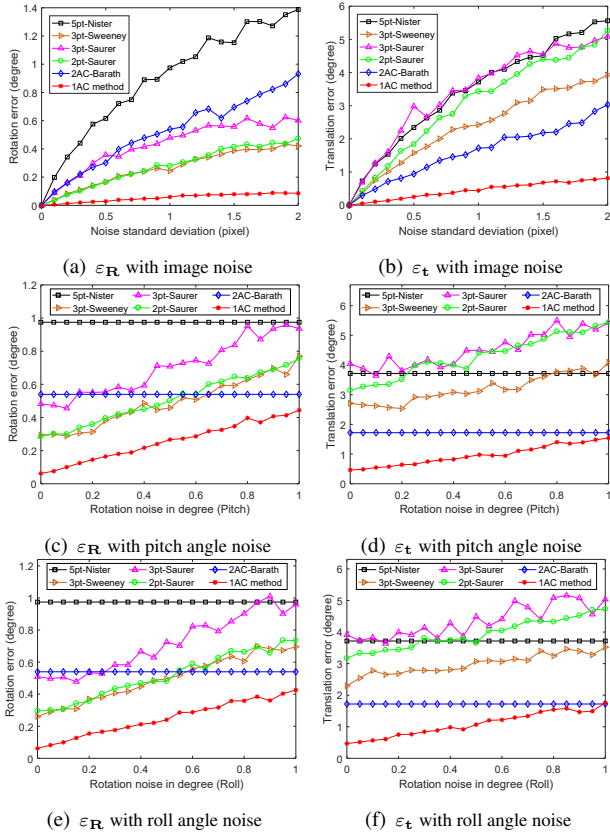
(f) $\varepsilon_{\mathbf{t}}$ with roll angle noise

Figure 5. Rotation and translation error under random motion (unit: degree). (a)(b): vary image noise with perfect IMU data. (c)∼(f): vary IMU angle noise and fix the image noise as 1.0 pixel standard deviation. The left column reports the rotation error. The right column reports the translation error.

used to align the camera coordinate system with the gravity direction. The proposed `1AC method` is compared with `5pt-Nister` [30], `3pt-Sweeney` [40], `3pt-Saurer` [33], `2pt-Saurer` [33] and `2AC-Barath` [3]. Due to space limitations, we only show the results under random motion. The results under forward and sideways motions are available in the supplementary material. Figure 5(a) and (b) show the performance of the proposed method with respect to the magnitude of image noise with perfect IMU data. Our method is robust to the increasing image noise and provides obviously better results than the previous methods.

Figure 5(c)∼(f) show the performance of the proposed method for increasing noise on the IMU data, while the image noise is set to 1.0 pixel standard deviation. The `1AC method` basically outperforms the methods `3pt-Sweeney`, `3pt-Saurer` and `2pt-Saurer`. The methods `5pt-Nister` and `2AC-Barath` are not influenced by the pitch error and the roll error, because their calculation does not utilize the known vertical direction as prior. It is interesting to see that our method performs bet-

ter than the methods `5pt-Nister` and `2AC-Barath` in the random motion case, even though the rotation noise is around $1.0°$. Under forward and sideways motion, the accuracy of our method is also better than the methods `5pt-Nister` and `2AC-Barath`, when the rotation noise stays below $0.3°$.

5.2. Experiments on Real Data

The performance of our methods on real image data is evaluated on the `KITTI` dataset [15]. All the sequences which provide ground truth data are utilized in this experiments. There are about 23000 images in total and are available as sequence 0 to 10.

5.2.1 Pose Estimation from Pairwise Image Pairs

Two settings of experiments are performed with the `KITTI` dataset, including planar motion estimation and relative pose estimation with known vertical direction. The ASIFT feature extraction and matching [27] is performed to obtain the affine correspondences between consecutive frames. Both the histogram voting and the RANSAC schemes are tested in this experiment. An inlier threshold of 2 pixels and a fixed number of 100 iterations are set in RANSAC.

In the first experiment, we test the relative pose estimation algorithms under planar motion. The motion estimation results between two consecutive images (θ, ϕ) are compared to the corresponding ground truth. The median error for each individual sequence is used to evaluate the performance. The proposed methods are compared with `2pt-Choi` [8]. The results of the rotation and translation error under planar motion assumption are shown in Table 1. Table 1 demonstrates that all of our planar motion methods provide better results than the `2pt-Choi` method. The overall performance of the `1AC-Voting` method is best among all the methods, particularly the rotation accuracy of the `1AC-Voting` method is significantly high than other methods.

Seq.	2pt-Choi [8]		1AC-CS		1AC-LS		1AC-Voting	
	$\varepsilon_{\mathbf{R}}$	$\varepsilon_{\mathbf{t}}$	$\varepsilon_{\mathbf{R}}$	$\varepsilon_{\mathbf{t}}$	$\varepsilon_{\mathbf{R}}$	$\varepsilon_{\mathbf{t}}$	$\varepsilon_{\mathbf{R}}$	$\varepsilon_{\mathbf{t}}$
00	0.203	5.169	0.133	**1.335**	0.155	1.345	**0.016**	1.493
01	0.150	3.617	0.117	**1.135**	0.134	1.149	**0.010**	1.165
02	0.154	3.364	0.062	1.152	0.082	1.191	**0.017**	**1.029**
03	0.177	6.441	0.084	1.157	0.100	**1.152**	**0.013**	1.225
04	0.115	2.871	0.029	1.132	0.041	1.155	**0.012**	**1.018**
05	0.143	4.407	0.071	**1.276**	0.085	1.304	**0.011**	1.614
06	0.152	3.379	0.051	**1.302**	0.068	1.340	**0.008**	1.655
07	0.127	4.764	0.059	1.487	0.074	**1.462**	**0.014**	1.769
08	0.137	4.312	0.064	1.428	0.081	**1.427**	**0.014**	1.591
09	0.141	3.508	0.062	**1.215**	0.081	1.218	**0.021**	1.221
10	0.145	3.829	0.067	**1.299**	0.090	1.299	**0.018**	1.464

Table 1. Rotation and translation error for `KITTI` sequences under planar motion assumption (unit: degree).

Seq.	5pt-Nister [30]		3pt-Sweeney [40]		3pt-Saurer [33]		2pt-Saurer [33]		2AC-Barath [3]		1AC method	
	$\varepsilon_\mathbf{R}$	$\varepsilon_\mathbf{t}$	$\varepsilon_\mathbf{R}$	$\varepsilon_\mathbf{t}$	$\varepsilon_\mathbf{R}$	$\varepsilon_\mathbf{t}$	$\varepsilon_\mathbf{R}$	$\varepsilon_\mathbf{t}$	$\varepsilon_\mathbf{R}$	$\varepsilon_\mathbf{t}$	$\varepsilon_\mathbf{R}$	$\varepsilon_\mathbf{t}$
00	.137	2.254	.065	2.165	.153	2.231	.336	7.675	.196	4.673	**.038**	**2.006**
01	.120	1.988	.082	2.342	.091	2.211	.186	9.806	.111	4.198	**.050**	**1.507**
02	.134	1.787	.059	**1.658**	.113	1.723	.293	6.034	.251	4.694	**.039**	1.861
03	.109	2.507	.067	2.723	.161	2.620	.316	9.249	.175	6.064	**.041**	**2.143**
04	.111	1.692	.048	1.558	.043	1.616	.141	4.816	.184	4.036	**.033**	**1.538**
05	.116	2.059	.054	1.895	.115	1.961	.253	7.238	.162	4.481	**.031**	**1.725**
06	.130	1.783	.068	1.615	.111	1.658	.232	5.750	.176	4.026	**.046**	**1.538**
07	.113	2.434	.052	2.183	.159	2.217	.378	8.293	.161	4.649	**.033**	**2.009**
08	.122	2.335	.053	2.216	.102	2.266	.241	7.556	.182	5.044	**.036**	**2.201**
09	.133	1.843	.059	**1.701**	.176	1.812	.409	6.606	.224	4.924	**.045**	1.799
10	.131	1.839	.059	**1.750**	.145	2.004	.308	7.324	.216	4.520	**.037**	1.935

Table 2. Rotation and translation error for `KITTI` sequences with known vertical direction (unit: degree).

In the second experiment, we test the relative pose estimation algorithm with known vertical direction, *i.e.*, `1AC method`. To simulate IMU measurements which provide a known gravity vector for the views of the camera, the image coordinates are pre-rotated by $\mathbf{R}_x\mathbf{R}_z$ obtained from the ground truth data. Table 2 lists the results of the rotation and translation estimation. The proposed methods are also compared against `5pt-Nister` [30], `3pt-Sweeney` [40], `3pt-Saurer` [33], `2pt-Saurer` [33] and `2AC-Barath` [3]. Table 2 demonstrates that our method is significantly more accurate than the other methods, except for the translation error of sequences 02, 09 and 10.
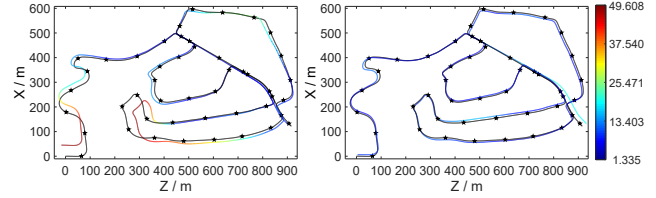
5.2.2 Visual Odometry

We demonstrate the usage of the `1AC method` in a monocular visual odometry pipeline to evaluate its performance in a real application. Our monocular visual odometry is based on `ORB-SLAM2` [28]. The affine correspondences extracted by ASIFT feature matching are used to replace the ORB features. The relative pose between two consecutive frames is estimated based on the combination of the `1AC method` using RANSAC, and is used to replace the original map initialization and the constant velocity motion model. The estimated trajectories after alignment with ground truth are illustrated in Figure 6. The color along the trajectory encodes the absolute trajectory error (ATE) [39]. Due to space limit, we show the trajectories of two sequences only. The results of other sequences can be found in supplementary materials[3]. It can be seen that the proposed `1AC method` method has the smallest ATE among the compared trajectories.

Moreover, we also evaluate the Relative Pose Error (RPE) between the estimated trajectory and the ground truth trajectory, which measures the relative accuracy of the trajectory over fixed time intervals [39]. The RMSE for rotation and translation using the RPE metric is illustrated in Table. 3. Our monocular visual odometry generally has smaller rotation and translational errors than `ORB-SLAM2`.

---

[3]Both `ORB-SLAM2` and our monocular visual odometry fail to produce a valid result for sequence 01, because it is a highway with few tractable close objects.
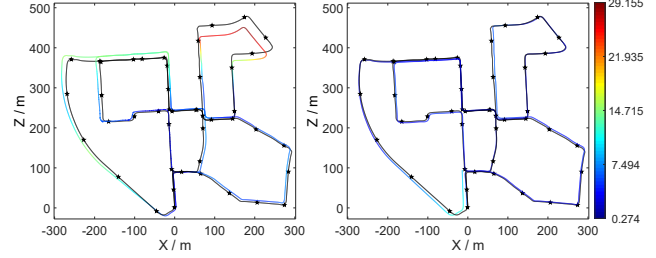
(a) Seq.00

(b) Seq.02

Figure 6. Estimated visual odometry trajectories. The left column reports the results of `ORB-SLAM2`. The right column reports the results of our monocular visual odometry. Colorful curves are estimated trajectories, and black curves with stars are ground truth trajectories. Best viewed in color.

Seq.	ORB-SLAM2 [28]		1AC-SLAM		Seq.	ORB-SLAM2 [28]		1AC-SLAM	
	$\varepsilon_\mathbf{R}$	$\varepsilon_\mathbf{t}$	$\varepsilon_\mathbf{R}$	$\varepsilon_\mathbf{t}$		$\varepsilon_\mathbf{R}$	$\varepsilon_\mathbf{t}$	$\varepsilon_\mathbf{R}$	$\varepsilon_\mathbf{t}$
00	0.821	0.923	**0.803**	**0.421**	06	0.142	1.478	**0.126**	**0.995**
02	0.200	1.052	**0.156**	**0.686**	07	0.149	0.879	**0.137**	**0.330**
03	**0.113**	0.244	0.118	**0.185**	08	0.177	1.778	**0.159**	**0.659**
04	0.151	0.417	**0.097**	**0.307**	09	0.221	0.777	**0.172**	**0.502**
05	0.264	0.681	**0.254**	**0.306**	10	0.129	0.633	0.238	1.008

Table 3. RMSE for rotation and translation using the RPE metric for `KITTI` sequences. Rotation error unit: degree. Translation error unit: meter.

6. Conclusion

In this paper, we showed that by exploiting the affine parameters it is possible to estimate the relative pose of a camera with only one affine correspondence under the planar motion assumption. Three minimal case solutions have been proposed to recover the planar motion of camera, amongst which is a solver which can even deal with an unknown focal length. In addition, a minimal case solution has been proposed to estimate the relative pose of a camera for the case of a known vertical direction. The assumptions in these methods are common to scenes in which self-driving cars and ground robots operate. By evaluating our algorithms on synthetic data and real-world image data sets, we demonstrate that our method can be used efficiently for outlier removal and for initial motion estimation in visual odometry.

References

[1] Sameer Agarwal, Hon-Leung Lee, Bernd Sturmfels, and Rekha R. Thomas. On the existence of epipolar matrices. *International Journal of Computer Vision*, 121(3):403–415, 2017.

[2] Daniel Barath. Five-point fundamental matrix estimation for uncalibrated cameras. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 235–243, 2018.

[3] Daniel Barath and Levente Hajder. Efficient recovery of essential matrix from two affine correspondences. *IEEE Transactions on Image Processing*, 27(11):5328–5337, 2018.

[4] Daniel Barath and Zuzana Kukelova. Homography from two orientation- and scale-covariant features. In *IEEE International Conference on Computer Vision*, pages 1091–1099, 2019.

[5] Daniel Barath, Tekla Toth, and Levente Hajder. A minimal solution for two-view focal-length estimation using two affine correspondences. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 6003–6011, 2017.

[6] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[7] Jacob Bentolila and Joseph M Francos. Conic epipolar constraints from affine correspondences. *Computer Vision and Image Understanding*, 122:105–114, 2014.

[8] Sunglok Choi and Jong-Hwan Kim. Fast and reliable minimal relative pose estimation under planar motion. *Image and Vision Computing*, 69:103–112, 2018.

[9] Chih-Chung Chou, YoungWoo Seo, and Chieh-Chih Wang. A two-stage sampling for robust feature matching. *Journal of Field Robotics*, 35(5):779–801, 2018.

[10] Yaqing Ding, Jian Yang, Jean Ponce, and Hui Kong. An efficient solution to the homography-based relative pose problem with a common reference direction. In *IEEE International Conference on Computer Vision*, pages 1655–1664, 2019.

[11] Timothy Duff, Kathlen Kohn, Anton Leykin, and Tomas Pajdla. PLMP - point-line minimal problems in complete multi-view visibility. In *IEEE International Conference on Computer Vision*, pages 1675–1684, 2019.

[12] Iván Eichhardt and Dmitry Chetverikov. Affine correspondences between central cameras for rapid relative pose estimation. In *European Conference on Computer Vision*, pages 482–497, 2018.

[13] Martin A Fischler and Robert C Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[14] Friedrich Fraundorfer, Petri Tanskanen, and Marc 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*, pages 269–282. Springer, 2010.

[15] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the KITTI vision benchmark suite. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361, 2012.

[16] Banglei Guan, Pascal Vasseur, Cédric Demonceaux, and Friedrich Fraundorfer. Visual odometry using a homography formulation with decoupled rotation and translation estimation using minimal solutions. In *IEEE International Conference on Robotics and Automation*, pages 2320–2327, 2018.

[17] Banglei Guan, Qifeng Yu, and Friedrich Fraundorfer. Minimal solutions for the rotational alignment of IMU-camera systems using homography constraints. *Computer Vision and Image Understanding*, 170:79–91, 2018.

[18] Levente Hajder and Daniel Barath. Relative planar motion for vehicle-mounted cameras from a single affine correspondence. In *IEEE International Conference on Robotics and Automation*, 2020.

[19] Richard Hartley and Hongdong Li. An efficient hidden variable approach to minimal-case camera motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(12):2303–2314, 2012.

[20] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2003.

[21] Kun Huang, Yifu Wang, and Laurent Kneip. Motion estimation of non-holonomic ground vehicles from a single feature correspondence measured over n views. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 12706–12715, 2019.

[22] Laurent Kneip, Margarita Chli, and Roland Y Siegwart. Robust real-time visual odometry with a single camera and an IMU. In *British Machine Vision Conference*, 2011.

[23] Zuzana Kukelova, Joe Kileel, Bernd Sturmfels, and Tomas Pajdla. A clever elimination strategy for efficient minimal solvers. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4912–4921, 2017.

[24] Viktor Larsson, Kalle Aström, and Magnus Oskarsson. Efficient solvers for minimal problems by syzygy-based reduction. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 820–828, 2017.

[25] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[26] Dmytro Mishkin, Jiri Matas, and Michal Perdoch. MODS: Fast and robust method for two-view matching. *Computer Vision and Image Understanding*, 141:81–93, 2015.

[27] Jean-Michel Morel and Guoshen Yu. ASIFT: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2(2):438–469, 2009.

[28] Raúl Mur-Artal and Juan D Tardós. ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras. *IEEE Transactions on Robotics*, 33(5):1255–1262, 2017.

[29] Oleg Naroditsky, Xun S Zhou, Jean Gallier, Stergios I Roumeliotis, and Kostas Daniilidis. Two efficient solutions for visual odometry using directional correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4):818–824, 2012.

[30] David Nistér. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):0756–777, 2004.

[31] Diego Ortín and José María Martínez Montiel. Indoor robot motion based on monocular images. *Robotica*, 19(3):331–342, 2001.

[32] Carolina Raposo and Joao P Barreto. Theory and practice of structure-from-motion using affine correspondences. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5470–5478, 2016.

[33] Olivier Saurer, Pascal Vasseur, Rémi Boutteau, Cédric Demonceaux, Marc Pollefeys, and Friedrich Fraundorfer. Homography based egomotion estimation with a common direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(2):327–341, 2016.

[34] Davide Scaramuzza and Friedrich Fraundorfer. Visual odometry: The first 30 years and fundamentals. *IEEE Robotics & Automation Magazine*, 18(4):80–92, 2011.

[35] Davide Scaramuzza, Friedrich Fraundorfer, and Roland Siegwart. Real-time monocular visual odometry for on-road vehicles with 1-point RANSAC. In *IEEE International Conference on Robotics and Automation*, pages 4293–4299, 2009.

[36] Johannes L Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4104–4113, 2016.

[37] Thiago L. T. da Silveira and Claudio R. Jung. Perturbation analysis of the 8-point algorithm: A case study for wide FoV cameras. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 11757–11766, 2019.

[38] Henrik Stewénius, David Nistér, Fredrik Kahl, and Frederik Schaffalitzky. A minimal solution for relative pose with unknown focal length. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 789–794, 2005.

[39] Jürgen Sturm, Nikolas Engelhard, Felix Endres, Wolfram Burgard, and Daniel Cremers. A benchmark for the evaluation of RGB-D SLAM systems. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 573–580, 2012.

[40] Chris Sweeney, John Flynn, and Matthew Turk. Solving for relative pose with a partially known rotation is a quadratic eigenvalue problem. In *International Conference on 3D Vision*, 2014.

[41] Ji Zhao. An efficient solution to non-minimal case essential matrix estimation. *arXiv:1903.09067*, 2019.

[42] Ji Zhao, Laurent Kneip, Yijia He, and Jiayi Ma. Minimal case relative pose computation using ray-point-ray features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(5), 2020.