Affine Correspondences between Multi-Camera Systems for 6DOF Relative Pose Estimation

Banglei Guan[1] and Ji Zhao ✉

[1]National University of Defense Technology, Changsha 410073, China
✉ corresponding author
guanbanglei12@nudt.edu.cn, zhaoji84@gmail.com

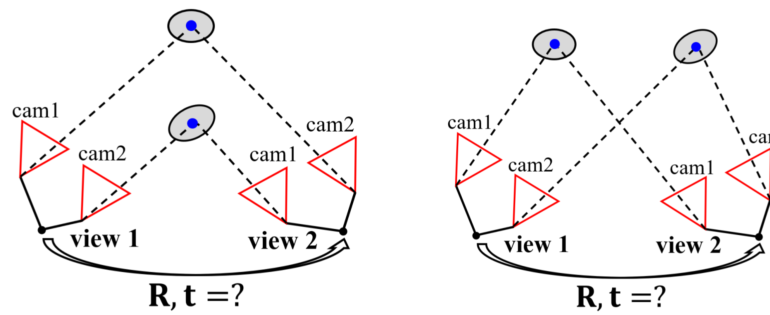
Abstract. We present a novel method to compute the 6DOF relative pose of multi-camera systems using two affine correspondences (ACs). Existing solutions to the multi-camera relative pose estimation are either restricted to special cases of motion, have too high computational complexity, or require too many point correspondences (PCs). Thus, these solvers impede an efficient or accurate relative pose estimation when applying RANSAC as a robust estimator. This paper shows that the relative pose estimation problem using ACs permits a feasible minimal solution, when exploiting the geometric constraints between ACs and multi-camera systems using a special parameterization. We present a problem formulation based on two ACs that encompass two common types of ACs across two views, *i.e.*, inter-camera and intra-camera. Experiments on both virtual and real multi-camera systems prove that the proposed solvers are more efficient than the state-of-the-art algorithms, while resulting in a better relative pose accuracy. Source code is available at https://github.com/jizhaox/relpose-mcs-depth.

Keywords: Relative pose estimation · Multi-camera system · Affine correspondence · Minimal solver

1 Introduction

Estimating the relative poses of a monocular camera, or a multi-camera system is a key problem in computer vision, which plays an important role in structure from motion (SfM), simultaneous localization and mapping (SLAM), and augmented reality (AR) [39,45,42,26,43,20,22]. A multi-camera system refers to a system of individual cameras that are rigidly fixed onto a single body, and it can be set in a configuration that maximizes the field-of-view. Motivated by the fact that multi-camera systems are an interesting choice in the context of robotics applications such as autonomous drones and vehicles, relative pose estimation for multi-camera systems has started to receive attention lately [20,22,1,35,18].

Different from monocular cameras which are modeled by the perspective camera model, the multi-camera systems can be modeled by the generalized camera model [16,46,37]. The generalized camera model does not have a single center of projection. The light rays that pass through the multi-camera system do not



(a) Inter-camera ACs (b) Intra-camera ACs

Fig. 1. Relative pose estimation from two ACs for a multi-camera system. Red triangle represents a single camera, and gray ellipse represents a spatial patch which relates to an AC. Specifically, inter-camera ACs refer to correspondences which are seen by different cameras over two consecutive views. Intra-camera ACs refer to correspondences which are seen by the same camera over two consecutive views.

intersect in a single center of projection, *i.e.*, non-central projection [40]. Thus, the relative pose estimation problem of multi-camera systems [45] is different from the monocular cameras [39], which results in different equations. In addition, since feature correspondences established by feature matching inevitably contain outliers, the relative pose estimation algorithms are typically employed inside a robust estimation framework such as the Random Sample Consensus (RANSAC) [13]. The computational complexity of the RANSAC estimator increases exponentially with respect to the number of feature correspondences needed. Thus, minimal solvers for relative pose estimation are very desirable for RANSAC schemes, which maximizes the probability of picking an all-inlier sample and reduces the number of necessary iterations [45,31,24,33,48,27].

The development of minimal solvers for relative pose estimation of multi-camera systems ranges back to the method of Stewénius *et al.* with six point correspondences (PCs) [45]. Later, some methods have been subsequently proposed, such as the linear method with seventeen PCs [31], iterative optimization method [26] and global optimization method [50]. In recent years, a number of solvers use affine correspondences (ACs), instead of PCs, to estimate the relative pose, which reduces the number of required correspondences [5,41,19,4,11,1,18]. Because an AC carries more information than a PC. However, existing AC-based solvers to the 6DOF relative pose estimation for multi-camera systems are either restricted to pose priors [17] or require at least six ACs [1]. It is desirable to find an AC-based minimal solver for 6DOF relative pose estimation of multi-camera systems, whose efficiency and accuracy are both satisfactory. This allows us to reduce the computational complexity of the RANSAC procedure.

In this paper, we focus on the 6DOF relative pose estimation problem of multi-camera systems from a minimal number of two ACs, see Fig. 1. We propose minimal solvers based on the common configurations of two ACs across two views, *i.e.*, inter-camera and intra-camera. The contributions of this paper are:

- We derive the geometric constraints between ACs and multi-camera systems using a special parameterization, which eliminates the translation parameters by utilizing two depth parameters. Moreover, the implicit constraints about the affine transformation constraints are found and proved.
- We develop two novel minimal solvers for 6DOF relative pose estimation of multi-camera systems from two ACs. Both solvers are designed for two common types of ACs in practice. We obtain practical solvers for totally new settings. In addition, three degenerate cases are proved.
- We exploit a unified and versatile framework for generating the minimal solvers, which uses the hidden variable technique to eliminate the depth parameters. This framework can be extended to solve various relative pose estimation problems, *e.g.*, relative pose estimation for a monocular camera.

2 Related Work

Stewénius *et al.* proposed the first minimal solver based on algebraic geometry, and this solver requires 6 PCs in order to come up with 64 solutions [45]. Kim *et al.* later presented alternative solvers for relative pose estimation with non-overlapping multi-camera systems using second-order cone programming [23] or branch-and-bound technique over the space of all rotations [24]. Clipp *et al.* also derived a solver using 6 PCs for non-overlapping multi-camera systems [9]. Lim *et al.* presented antipodal epipolar constraints on the relative pose by exploiting the geometry of antipodal points, which are available in large field-of-view cameras [33]. Li *et al.* used 17 PCs to solve the relative pose of multi-camera systems linearly, which ignores side-constraints on the generalized essential matrix and the contained essential and rotation matrices [31]. Kneip and Li proposed an iterative approach for the relative pose estimation with an efficient eigenvalue minimization strategy [26]. The above mentioned works are designed for 6DOF relative pose estimation of multi-camera systems.

A number of methods estimate the relative pose of multi-camera systems with a prior. Typically, the priors include multi-camera movement prior and known vertical direction prior, which reduce the DOF of the relative pose problem. Lee *et al.* [29] used a minimum of 2 PCs to recover the 2DOF relative pose, while the multi-camera system is mounted on ground robots and the movement follows the Ackermann motion model. In addition, when the vertical direction of the multi-camera system is obtained by vanishing point estimation or sensor fusion with an IMU, Sweeney *et al.* [47], Lee *et al.* [30] and Liu *et al.* [34] proposed several minimal solvers with 4 PCs to solve 4DOF relative pose.

Recently, using ACs to estimate the relative pose of multi-camera systems has drawn much attention. Alyousefi and Ventura [1] proposed a linear solver to recover the 6DOF relative pose using 6 ACs, which generalizes the 17 PCs solver proposed by Li *et al.* [31]. Guan *et al.* [17] used a first-order approximation to relative rotation to estimate the 6DOF relative pose, which generalizes the 6 PCs solver proposed by Ventura *et al.* [48]. They assume that the relative rotation of the multi-camera systems between two consecutive views is small. Furthermore, Guan *et al.* [18] estimated the 3DOF relative pose under planar motion with a single AC and estimated the 4DOF relative pose with known vertical direction with 2 ACs. In this paper, we focus on using a minimal number of 2 ACs to estimate the 6DOF relative pose of multi-camera systems, which does not rely on any motion constraints or pose priors.

3 Relative Pose Estimation for Multi-Camera Systems

In this section, we assume that both the intrinsic and extrinsic parameters of multi-camera systems are known. Aiming at the common configurations of two ACs across two views in Fig. 1, our purpose is to find the minimal solvers for inter-camera ACs and intra-camera ACs. The proposed solvers are the most common ones in practice for multi-camera systems.

3.1 Parameterization

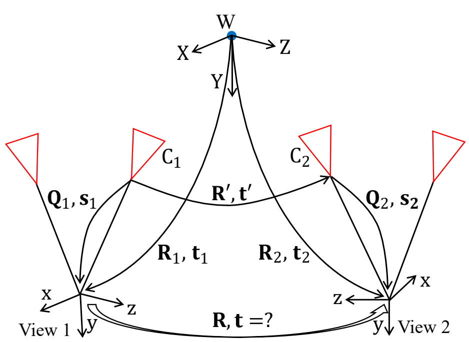


Fig. 2. Relative pose estimation for multi-camera systems.

We first formulate and parameterize the relative pose estimation problem for a multi-camera system. As shown in Fig. 2, the multi-camera system is composed of multiple perspective cameras. The extrinsic parameters of C_i are denoted as $\{\mathbf{Q}_i, \mathbf{s}_i\}$, where $\mathbf{Q}_i$ and $\mathbf{s}_i$ represent relative rotation and translation to the reference of the multi-camera system. Denote the relative pose of multi-camera systems as $\{\mathbf{R}, \mathbf{t}\}$, which represents the relative rotation and translation from view 1 to view 2 of the multi-camera system.

In order to eliminate the translation parameters, we use a special parameterization to formulate the relative pose estimation problem [45]. Take an AC seen by the different cameras for an example. The j-th AC relates the camera C_1 and C_2 across two views, see Fig. 2. Let us denote the j-th AC as $(\mathbf{x}_j, \mathbf{x}'_j, \mathbf{A}_j)$, where $\mathbf{x}_j$ and $\mathbf{x}'_j$ are the normalized homogeneous image coordinates of feature points in the view 1 and view 2, respectively. $\mathbf{A}_j$ is a 2×2 local affine transformation, which relates the infinitesimal patches around $\mathbf{x}_j$ and $\mathbf{x}'_j$ [41,2]. Suppose the j-th AC is chosen to define a world reference system W. The origin of W is set to the position of the j-th AC in 3D space and the orientation of W is consistent with the reference of the multi-camera system in view 1. Denote the relative rotation and translation from reference W to view 1 as $\{\mathbf{R}_1, \mathbf{t}_1\}$. Denote the relative rotation and translation from reference W to view 2 as $\{\mathbf{R}_2, \mathbf{t}_2\}$. It can be seen that $\mathbf{R}_1 = \mathbf{I}$, $\mathbf{R}_2 = \mathbf{R}$. In this paper, the Cayley parameterization is used to parametrized the relative rotation $\mathbf{R}$:

$$\mathbf{R} = \frac{1}{1+q_x^2+q_y^2+q_z^2} \begin{bmatrix} 1+q_x^2-q_y^2-q_z^2 & 2q_xq_y-2q_z & 2q_y+2q_xq_z \\ 2q_xq_y+2q_z & 1-q_x^2+q_y^2-q_z^2 & 2q_yq_z-2q_x \\ 2q_xq_z-2q_y & 2q_x+2q_yq_z & 1-q_x^2-q_y^2+q_z^2 \end{bmatrix}, \tag{1}$$

where $[1, q_x, q_y, q_z]^\mathrm{T}$ is a homogeneous quaternion vector. Note that the Cayley parameterization introduces a degeneracy for 180° rotations, but this is a rare case for consecutive views in the robotics applications [44,26,51,49].

Next, we show that the translation parameters $\mathbf{t}_1$ and $\mathbf{t}_2$ can be removed by using two depth parameters. For a calibrated multi-camera system, each image point corresponds to a unique line in the reference of the multi-camera system. This line in 3D can be represented as a Plücker vector $\mathbf{L} = [\mathbf{p}^\mathrm{T}, \mathbf{q}^\mathrm{T}]^\mathrm{T}$, where the 3D vectors $\mathbf{p}$ and $\mathbf{q}$ represent the unit direction vector and the moment vector, respectively [40]. They satisfy the constraint $\mathbf{p} \cdot \mathbf{q} = 0$. Thus, the set of points $\mathbf{X}(\lambda)$ on the 3D line can be parameterized as

$$\mathbf{X}(\lambda) = \mathbf{q} \times \mathbf{p} + \lambda \mathbf{p}, \quad \lambda \in \mathbb{R}. \tag{2}$$

where λ is the unknown depth parameter of 3D point. Since the origin of W is set to the 3D position $\mathbf{X}_j$ corresponding to j-th AC, the Plücker coordinates of the line connecting the 3D position $\mathbf{X}_j$ and the optical center of camera C_i can be described as $[\mathbf{p}_{ij}^\mathrm{T}, \mathbf{q}_{ij}^\mathrm{T}]^\mathrm{T}$ in the reference of the multi-camera system. The 3D position $\mathbf{X}_j$ in view k satisfies the following constraint:

$$\mathbf{q}_{ij} \times \mathbf{p}_{ij} + \lambda_{jk} \mathbf{p}_{ij} = \mathbf{R}_k \begin{bmatrix} 0, & 0, & 0 \end{bmatrix}^\mathrm{T} + \mathbf{t}_k, \quad k = 1, 2. \tag{3}$$

Based on Eq. (3), the translation $\mathbf{t}_k$ from W to view k is parameterized as the linear expression in the unknown depth parameter λ_{jk}

$$\mathbf{t}_k = \mathbf{q}_{ij} \times \mathbf{p}_{ij} + \lambda_{jk} \mathbf{p}_{ij}, \quad k = 1, 2. \tag{4}$$

where k represents the index of the views, i represents the index of the cameras, and j represents the index of the ACs. It can be seen that λ_{j1} and λ_{j2} are the depth parameters of the origin of W in views 1 and 2, respectively.

Through the above special parameterization, the 6DOF relative pose of multi-camera systems can be described by five unknowns, which consist of three rotation parameters $\{q_x, q_y, q_z\}$ and two depth parameters $\{\lambda_{j1}, \lambda_{j2}\}$.

3.2 Geometric Constraints

It has been shown in Fig. 2 that each AC relates two perspective cameras in view 1 and view 2. The relative pose between two cameras $[\mathbf{R}', \mathbf{t}']$ is determined by the composition of four transformations: (i) from one perspective camera to view 1, (ii) from view 1 to W, (iii) from W to view 2, (iv) from view 2 to the other perspective camera. Among these four transformations, the part (i) and (iv) are determined by known extrinsic parameters. In the part (ii) and (iii), there are unknowns $\mathbf{R}$, $\mathbf{t}_1$ and $\mathbf{t}_2$ which are parameterized as $\{q_x, q_y, q_z, \lambda_{j1}, \lambda_{j2}\}$. Formally, the relative pose $[\mathbf{R}', \mathbf{t}']$ is represented as:

$$\begin{aligned}\begin{bmatrix}\mathbf{R}' & \mathbf{t}' \\ \mathbf{0} & 1\end{bmatrix} &= \begin{bmatrix}\mathbf{Q}_2 & \mathbf{s}_2 \\ \mathbf{0} & 1\end{bmatrix}^{-1} \begin{bmatrix}\mathbf{R} & \mathbf{t}_2 \\ \mathbf{0} & 1\end{bmatrix} \begin{bmatrix}\mathbf{I} & \mathbf{t}_1 \\ \mathbf{0} & 1\end{bmatrix}^{-1} \begin{bmatrix}\mathbf{Q}_1 & \mathbf{s}_1 \\ \mathbf{0} & 1\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{Q}_2^{\mathrm{T}}\mathbf{R}\mathbf{Q}_1 & \mathbf{Q}_2^{\mathrm{T}}(\mathbf{R}\mathbf{s}_1 - \mathbf{R}\mathbf{t}_1 + \mathbf{t}_2 - \mathbf{s}_2) \\ \mathbf{0} & 1\end{bmatrix}.\end{aligned} \tag{5}$$

Once the relative pose $[\mathbf{R}', \mathbf{t}']$ between two perspective cameras for each AC is expressed, the essential matrix $\mathbf{E}' = [\mathbf{t}']_\times \mathbf{R}'$ can be represented as:

$$\mathbf{E}' = \mathbf{Q}_2^{\mathrm{T}} \left(\mathbf{R}[\mathbf{s}_1 - \mathbf{t}_1]_\times + [\mathbf{t}_2 - \mathbf{s}_2]_\times \mathbf{R}\right) \mathbf{Q}_1. \tag{6}$$

By substituting Eq. (4) into Eq. (6), we obtain:

$$\begin{aligned}\mathbf{E}' =& -\lambda_{j1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{R}[\mathbf{p}_{ij}]_\times \mathbf{Q}_1 + \lambda_{j2}\mathbf{Q}_2^{\mathrm{T}}[\mathbf{p}'_{ij}]_\times \mathbf{R}\mathbf{Q}_1 \\ &+ \mathbf{Q}_2^{\mathrm{T}} \big(\mathbf{R}[\mathbf{s}_1 - \mathbf{q}_{ij} \times \mathbf{p}_{ij}]_\times + [\mathbf{q}'_{ij} \times \mathbf{p}'_{ij} - \mathbf{s}_2]_\times \mathbf{R}\big) \mathbf{Q}_1.\end{aligned} \tag{7}$$

It can be verified that each entry in the essential matrix $\mathbf{E}'$ is linear with $\{\lambda_{j1}, \lambda_{j2}\}$. Generally speaking, one AC $(\mathbf{x}_j, \mathbf{x}'_j, \mathbf{A}_j)$ yields three independent constraints on the relative pose estimation of a multi-camera system, which consist of one epipolar constraint derived from PC $(\mathbf{x}_j, \mathbf{x}'_j)$ and two affine transformation constraints derived from local affine transformation $\mathbf{A}_j$. With known intrinsic camera parameters, the epipolar constraint of PC between view 1 and view 2 is given as follows [21]:

$$\mathbf{x}_j'^{\mathrm{T}} \mathbf{E}' \mathbf{x}_j = 0, \tag{8}$$

The affine transformation constraints which describe the relationship of essential matrix $\mathbf{E}'$ and local affine transformation $\mathbf{A}_j$ is formulated as follows [41,2]:

$$(\mathbf{E}'^{\mathrm{T}} \mathbf{x}'_j)_{(1:2)} = -\mathbf{A}_j^{\mathrm{T}} (\mathbf{E}' \mathbf{x}_j)_{(1:2)}, \tag{9}$$

where the subscript (1:2) represents the first two equations.

Even though the perspective cameras are assumed, the geometric constraints can straightforwardly be generalized to generalized camera models as long as local image patches across views are obtained equivalently by arbitrary central camera models [2,12]. Based on Eqs. (8) and (9), two ACs provide six independent constraints. Considering that the relative pose estimation problem of multi-cameras systems has 6DOF, the number of constraints is equal to the number of unknowns. Thus, we explore the minimal solvers using two ACs.

3.3 Equation System Construction

Note that the special parameterization has been adopted by choosing one AC as the origin of world reference system in the subsection 3.1, we found the PC derived from the chosen AC cannot contribute one constraint since the coefficients of the resulting equation are zero. Thus, when j-th AC is chosen to build up the world reference system W, five equations can be provided by two ACs. Specifically, j-th AC provides two equations based on Eq. (9) and the other AC provides three equations based on Eqs. (8) and (9). By substituting Eq. (7) into Eqs. (8) and (9) and using the hidden variable technique [10], the five equations provided by two ACs can be written as:

$$\underbrace{\mathbf{F}_j(q_x, q_y, q_z)}_{5\times 3}\begin{bmatrix}\lambda_{j1}\\ \lambda_{j2}\\ 1\end{bmatrix} = \mathbf{0}. \tag{10}$$

The entries in $\mathbf{F}_j$ are quadratic in unknowns q_x, q_y, and q_z. Since Eq. (10) has non-trivial solutions, the rank of $\mathbf{F}_j$ satisfies $\text{rank}(\mathbf{F}_j) \leq 2$. Thus, all the 3×3 sub-determinants of $\mathbf{F}_j$ must be zero. This gives 10 equations about three unknowns $\{q_x, q_y, q_z\}$. Moreover, we can choose the other AC to build up the world reference system, and its orientation is also consistent with the reference of the multi-camera system in view 1. Suppose the j'-th AC is chosen, we build a new equation system about the same rotation parameters $\{q_x, q_y, q_z\}$, which is similar to Eq. (10):

$$\underbrace{\mathbf{F}_{j'}(q_x, q_y, q_z)}_{5\times 3}\begin{bmatrix}\lambda_{j'1}\\ \lambda_{j'2}\\ 1\end{bmatrix} = \mathbf{0}. \tag{11}$$

Note that Eq. (11) provides new constraints which is different from Eq. (10). We use the computer algebra system `Macaulay 2` [15] to find that there are one dimensional families of extraneous roots if only Eq. (10) or Eq. (11) is used. This phenomenon has also been observed in [45,35]. Based on Eqs. (10) and (11), we have 20 equations with three unknowns $\{q_x, q_y, q_z\}$:

$$\begin{gathered}\det(\mathbf{N}(q_x, q_y, q_z)) = 0, \\ \mathbf{N} \in \{3\times 3 \text{ submatrices of } \mathbf{F}_j\} \cup \{3\times 3 \text{ submatrices of } \mathbf{F}_{j'}\}.\end{gathered} \tag{12}$$

These equations have a degree of 6, *i.e.*, the highest of the degrees of the monomials with non-zero coefficients is 6.

Moreover, we derive extra implicit constraints in our problem, *i.e.*, the rank of $(\mathbf{F}_j)_{(1:2,1:3)}$ is 1. The proof is provided as follows:

Theorem 1. *When j-th AC is chosen to build up the world reference system, the corresponding affine transformation constraints satisfy* $\text{rank}((\mathbf{F}_j)_{(1:2,1:3)}) = 1$.

Proof. To achieve this goal, we need to prove that $(\mathbf{F}_j)_{(1:2,1:3)}$ has two linearly independent null space vectors $\mathbf{v}_1$ and $\mathbf{v}_2$. Based on Eq. (10), $\mathbf{v}_1 = [\lambda_{j1}, \lambda_{j2}, 1]^{\mathrm{T}}$ is obviously a null space vector. Then we suppose that the second null space vector can be expressed as $\mathbf{v}_2 = [\lambda_{z1}, \lambda_{z2}, 0]^{\mathrm{T}}$, where λ_{z1} and λ_{z2} are two unknown depth parameters of the origin of world reference system W in camera 1 (view 1) and camera 2 (view 2), respectively.

For the multi-camera system in Fig. 2, we parameterize the transformation of cameras with respect to the world reference system W. Denote the transformation between camera 1 in view 1 and W as $[\mathbf{Q}_1^{\mathrm{T}}, \lambda_{z1}\mathbf{Q}_1^{\mathrm{T}}\mathbf{p}_{ij}]$, and the transformation between camera 2 in view 2 and W as $[\mathbf{Q}_2^{\mathrm{T}}\mathbf{R}, \lambda_{z2}\mathbf{Q}_2^{\mathrm{T}}\mathbf{p}'_{ij}]$. The transformation between camera 1 in view 1 and camera 2 in view 2 $\{\tilde{\mathbf{R}}, \tilde{\mathbf{t}}\}$ can be computed. Thus, the corresponding essential matrix $\tilde{\mathbf{E}} = [\tilde{\mathbf{t}}]_\times \tilde{\mathbf{R}}$ is represented as

$$\tilde{\mathbf{E}} = -\lambda_{z1}\mathbf{Q}_2^{\mathrm{T}}\mathbf{R}[\mathbf{p}_{ij}]_\times \mathbf{Q}_1 + \lambda_{z2}\mathbf{Q}_2^{\mathrm{T}}[\mathbf{p}'_{ij}]_\times \mathbf{R}\mathbf{Q}_1. \tag{13}$$

Note that the coefficients of the unknowns λ_{z1} and λ_{z2} in Eq. (13) are the same as the coefficients of the unknowns λ_{j1} and λ_{j2} in Eq. (7). Based on the Eq. (9), the affine transformation constraints can be written as follows

$$(\mathbf{G}_j)_{(1:2,1:2)} \begin{bmatrix} \lambda_{z1} \\ \lambda_{z2} \end{bmatrix} = \mathbf{0}. \tag{14}$$

In comparison with Eq. (10), $(\mathbf{G}_j)_{(1:2,1:2)}$ is the first 2×2 sub-matrix of $(\mathbf{F}_j)_{(1:2,1:3)}$. We can see that the null space vector $\mathbf{v}_2 = [\lambda_{z1}, \lambda_{z2}, 0]^{\mathrm{T}}$ is also the null space vector of $(\mathbf{F}_j)_{(1:2,1:3)}$. Thus, the rank of $(\mathbf{F}_j)_{(1:2,1:3)}$ is 1.

Based on Theorem 1, the affine transformation constraints provide extra equations for our problem. Only if j-th AC is chosen to build up the world reference system W, two affine transformation constraints of j-th AC are used in the equation system construction. Similarly, when we choose the j'-th AC to build up the world reference, the extra equations of the corresponding affine transformation constraints can also be provided. Thus, there are 6 extra equations for the relative pose estimation using ACs:

$$\det(\mathbf{M}(q_x, q_y, q_z)) = 0, \tag{15}$$
$$\mathbf{M} \in \{2 \times 2 \text{ submatrices of } (\mathbf{F}_j)_{(1:2,1:3)}\} \cup \{2 \times 2 \text{ submatrices of } (\mathbf{F}_{j'})_{(1:2,1:3)}\}.$$

These extra equations have a degree of 4. Note that the extra implicit constraints Eq. (15) are independent of Eqs. (10) and (11). For geometric explanation, the extra constraints encode that the affine transformation constraints come from a perspective camera of two viewpoints. As we will see later, using the extra constraints from Theorem 1 reduces the number of solutions.

3.4 Polynomial System Solving

We propose two minimal solvers based on the common configurations of two ACs in multi-camera systems, including an inter-camera solver and an intra-camera solver. The inter-camera solver uses inter-camera ACs which are seen by different cameras over two consecutive views. It is suitable for multi-camera systems with large overlapping of views. The intra-camera solver uses intra-camera ACs which are seen by the same camera over two consecutive views. It is suitable for multi-camera systems with non-overlapping or small-overlapping of views.

A suitable way to find algebraic solutions to the polynomial equation system Eqs. (12) and (15) is to use the Gröbner basis technique. To keep numerical stability and avoid large number arithmetic during the calculation of Gröbner basis, a random instance of the original equation system is constructed in a finite prime field $\mathbb{Z}_p$ [32]. The relations between all observations are appropriately preserved. Then, we use `Macaulay 2` [15] to calculate Gröbner basis. Finally, the solver is found with the automatic Gröbner basis solver [28]. We denote these polynomial equations in Eq. (12) and Eq. (15) as $\mathcal{E}_1$ and $\mathcal{E}_2$, respectively. Note that the polynomial equations $\mathcal{E}_1$ and $\mathcal{E}_2$ can be extended to solve various relative pose estimation problems, such as with known rotation angle and unknown focal lengths. In this paper, $\mathcal{E}_1$ is sufficient to solve the relative pose with inter-camera ACs. For intra-camera ACs, there are one-dimensional families of extraneous roots if only $\mathcal{E}_1$ is used. Moreover, using both $\mathcal{E}_1$ and $\mathcal{E}_2$ can reduce the number of solutions in the inter-camera case.

Table 1. Minimal solvers for the multi-camera systems. `#sol` indicates the number of solutions. `1-dim` indicates one dimensional families of extraneous roots.

AC type	$\mathcal{E}_1$		$\mathcal{E}_1 + \mathcal{E}_2$	
	#sol	template	#sol	template
Inter-camera	56	56×120	48	64×120
Intra-camera	1-dim	–	48	72×120

Table 1 shows the resulting inter-camera and intra-camera solvers. We have the following observations. (1) If $\mathcal{E}_1$ is used, the inter-camera solver maximally has 56 complex solutions and the elimination template of size 56×120. But the intra-camera case has one-dimensional families of extraneous roots. (2) If both $\mathcal{E}_1$ and $\mathcal{E}_2$ are used, the number of complex solutions obtained by the inter-camera solver can be reduced to 48. The number of complex solutions obtained by the intra-camera solver is also 48. The elimination template of the inter-camera solver and intra-camera solver is 64×120 and 72×120, respectively. (3) For the inter-camera case, using equations from $\mathcal{E}_1$ results in smaller eliminate templates than using $\mathcal{E}_1 + \mathcal{E}_2$. Meanwhile, the solver resulting from $\mathcal{E}_1$ has better numerical stability than the solver resulting from $\mathcal{E}_1 + \mathcal{E}_2$. This phenomenon has also been observed in previous literature [7], which shows that the number of basis might affect the numerical stability.

Once the rotation parameters $\{q_x, q_y, q_z\}$ are obtained, $\mathbf{R}$ can be obtained immediately. Then $\{\lambda_{jk}\}_{k=1,2}$ and $\{\lambda_{j'k}\}_{k=1,2}$ are determined by finding the null space of $\mathbf{F}_j$ and $\mathbf{F}_{j'}$, respectively. Note that the translations estimated by $\{\lambda_{jk}\}$ and $\{\lambda_{j'k}\}$ are theoretically the same in minimal problems. Take the translation estimation using $\{\lambda_{jk}\}$ for an example. We can calculate $\mathbf{t}_1$ and $\mathbf{t}_2$ by Eq. (4). Finally we calculate the relative pose by compositing the transformations $[\mathbf{R}_1, \mathbf{t}_1]$ and $[\mathbf{R}_2, \mathbf{t}_2]$. Moreover, our minimal solver generation framework can be easily extended to recover the relative pose of a monocular camera. See supplementary material for details.

3.5 Degenerated Configurations

We prove three cases of critical motions for relative pose estimation from ACs.

Proposition 1. *For inter-camera ACs, if a multi-camera system undergoes pure translation and the baseline of two camera is parallel with the translation direction, the metric scale of translation cannot be recovered.*

Proposition 2. *For intra-camera ACs, when a multi-camera system undergoes pure translation or constant rotation rate, both cases are degenerate motions. Specifically, the metric scale of translation cannot be recovered.*

Due to space limitations, the proof of degenerate cases and the methods to overcome the degenerate cases are provided in the supplementary material.

4 Experiments

The performance of our solvers is validated using both synthetic and real-world data. The proposed `2AC method` are referred to as `2AC-inter` for inter-camera ACs, and `2AC-intra` for intra-camera ACs. To further distinguish two solvers for inter-camera ACs, `2AC-inter-56` and `2AC-inter-48` are used to refer the solvers resulting from $\mathcal{E}_1$ and $\mathcal{E}_1 + \mathcal{E}_2$, respectively. The proposed solvers are implemented in C++. The `2AC-inter` solver and the `2AC-intra` solver are compared with state-of-the-art methods including `17PC-Li` [31], `8PC-Kneip` [26], `6PC-Stewénius` [45] and `6AC-Ventura` [1]. The methods which estimate the relative pose with a prior are not compared in this paper [47,48,17,18]. All the solvers are integrated into RANSAC in order to remove outlier matches of the feature correspondences. The relative pose which produces the most inliers is used to measure the relative pose error. This also allows us to select the best candidate from multiple solutions by counting their inliers.

The relative rotation and translation of the multi-camera systems are compared separately in the experiments. The rotation error compares the angular difference between the ground truth rotation and the estimated rotation: $\varepsilon_{\mathbf{R}} = \arccos((\mathrm{trace}(\mathbf{R}_{gt}\mathbf{R}^{\mathrm{T}}) - 1)/2)$, where $\mathbf{R}_{gt}$ and $\mathbf{R}$ denote the ground truth rotation and the corresponding estimated rotation, respectively. We evaluate the translation error by following the definition in [30]: $\varepsilon_{\mathbf{t}} = 2\left\|(\mathbf{t}_{gt} - \mathbf{t})\right\|/(\|\mathbf{t}_{gt}\| + \|\mathbf{t}\|)$, where $\mathbf{t}_{gt}$ and $\mathbf{t}$ denote the ground truth translation and the corresponding estimated translation, respectively. $\varepsilon_{\mathbf{t}}$ denotes both the metric scale error and the direction error of the translation. The translation direction error is also evaluated separately by comparing the angular difference between the ground truth translation and the estimated translation: $\varepsilon_{\mathbf{t},\mathrm{dir}} = \arccos((\mathbf{t}_{gt}^{\mathrm{T}}\mathbf{t})/(\|\mathbf{t}_{gt}\| \cdot \|\mathbf{t}\|))$.

4.1 Experiments on Synthetic Data

A simulated multi-camera system is made to evaluate the inter-camera and intra-camera solvers simultaneously [17,18]. The baseline length between two simulated cameras is set to 1 meter, and the movement length of the multi-camera system is set to 3 meters. The resolution of cameras is 640 $\times$ 480 pixels with a focal length of 400 pixels. The principal points are set to the image center (320,

Table 2. Runtime comparison of relative pose estimation solvers (unit: μs).

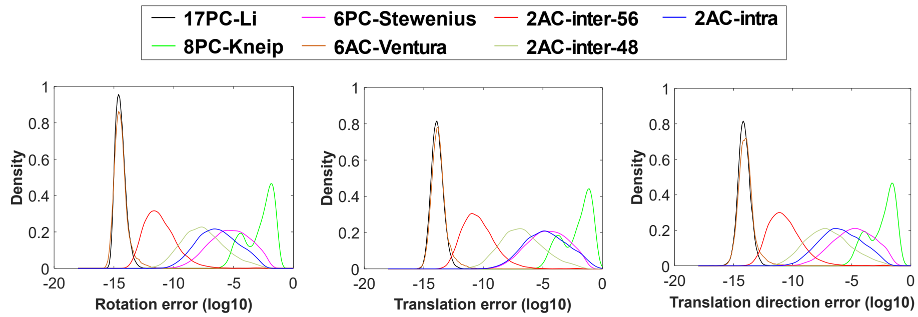
Methods	17PC-Li [31]	8PC-Kneip [26]	6PC-Stew. [45]	6AC-Vent. [1]	**2AC-inter-56**	**2AC-inter-48**	**2AC-intra**
Runtime	43.3	102.0	3275.4	38.1	1084.8	842.3	871.6

240). We carry out a total of 1000 trials and assess the rotation and translation error by the median of errors in the synthetic experiment.

In each test, 100 ACs are generated randomly, including 50 ACs from a ground plane and 50 ACs from 50 random planes. The synthetic scene is randomly generated in a cubic region of size $[-5, 5]\times[-5, 5]\times[10, 20]$ meters. For each AC, the PC is obtained by reprojecting a random 3D point from a plane into two cameras. The associated affine transformation is obtained as follows: First, four additional image points are chosen as the vertices of a square in view 1, where its center is the PC of AC. The side length of the square is set to 30 or 40 pixels. A larger side length means the larger support regions for generating the ACs, which causes smaller noise of affine transformation. The support region is used for AC noise simulation only. Second, the ground truth homography is used to calculate the four corresponding image points in view 2. Third, Gaussian noise is added to the coordinates of four sampled image point pairs. Fourth, the noisy affine transformation is calculated from the first-order approximation of the noisy homography, which is estimated by using four noise image point pairs. This procedure promises an indirect but geometrically interpretable way of noising the affine transformation [3]. The Gaussian noise with a standard deviation is added to the PCs, and, also, to the sampled image point pairs which are used to estimate the affine transformations. In the experiments, the required ACs are selected randomly for the solvers within the RANSAC scheme. For the PC-based solvers, only the PCs derived from the ACs are used.

Efficiency Comparison and Numerical Stability The proposed solvers are evaluated on an Intel(R) Core(TM) i7-7800X 3.50GHz. All comparison solvers are implemented in C++. The `17PC-Li`, `8PC-Kneip` and `6PC-Stewénius` are provided by OpenGV library [25]. The `6AC-Ventura` is publicly available from the code of [1]. Table 2 shows the average processing times of the solvers over 10,000 runs. The methods `17PC-Li` and `6AC-Ventura` have low runtime, because they solve for the multi-camera motion linearly. Among the minimal solvers, all the proposed solvers `2AC-inter-56`, `2AC-inter-48` and `2AC-intra` are significantly more efficient than the `6PC-Stewénius` solver.

Figure 3 reports the numerical stability comparison of all the solvers in noise-free cases. We repeat the procedure 10,000 times and plot the empirical probability density functions as the function of the $\log_{10}$ estimated errors. Numerical stability represents the round-off error of solvers in noise-free cases. The solvers `17PC-Li` and `6AC-Ventura` have the best numerical stability, because the linear solvers with smaller computation burden have less round-off error. Since the `8PC-Kneip` solver uses the iterative optimization, it is susceptible to falling into local minima. Among the minimal solvers, all the proposed solvers



(a) $\varepsilon_{\mathbf{R}}$ (b) $\varepsilon_{\mathbf{t}}$ (c) $\varepsilon_{\mathbf{t},\mathrm{dir}}$

Fig. 3. Probability density functions over relative pose estimation errors on noise-free observations for multi-camera systems. The horizontal axis represents the $\log_{10}$ errors, and the vertical axis represents the density.

`2AC-inter-56`, `2AC-inter-48` and `2AC-intra` have better numerical stability than the `6PC-Stewénius` solver. Moreover, `2AC-inter-56` has better numerical stability than `2AC-inter-48`, which shows that adding the extra equations $\mathcal{E}_2$ is not helpful in improving the numerical stability of the `2AC-inter` solver. Even though `2AC-inter-48` produces the less solutions, we prefer to perform `2AC-inter-56` for the sake of numerical accuracy in the follow-up experiments. In addition to efficiency and numerical stability, another important factor for a solver is the minimal number of needed feature correspondences between two views. Since the proposed solvers require only two ACs, the number of RANSAC iterations is obviously lower than PC-based methods. Thus, our solvers have an advantage in detecting the outlier and estimating the initial motion efficiently when integrating them into the RANSAC framework. See supplementary material for details. Due to space limitations, the performance of the proposed solvers with different image noise is also shown in the supplementary material. As we will see later, the proposed solvers have better overall efficiency than the comparative solvers in the experiments on real-world data.

4.2 Experiments on Real Data

We evaluate the performance of the proposed solvers on three public datasets in popular modern robot applications. Specifically, the `KITTI` dataset [14] and `nuScenes` dataset [8] are collected on an autonomous driving environment. The `EuRoc` MAV dataset [6] is collected on an unmanned aerial vehicle environment. These datasets provide challenging image pairs, such as large motion and highly dynamic scenes. We compare the proposed solvers against state-of-the-art 6DOF relative pose estimation techniques. The rotation error $\varepsilon_{\mathbf{R}}$ and the translation direction error $\varepsilon_{\mathbf{t},\mathrm{dir}}$ are used to evaluate the accuracy of the proposed solvers [1,26,34]. We tested on a total of 30,000 image pairs. Our solvers focus on relative pose estimation, *i.e.*, integrating the minimal solver with RANSAC.

Table 3. Rotation and translation error on `KITTI` sequences (unit: degree).

Seq.	17PC-Li [31]		8PC-Kneip [26]		6PC-Stew. [45]		6AC-Vent. [1]		**2AC method**	
	$\varepsilon_{\mathbf{R}}$	$\varepsilon_{\mathbf{t},\text{dir}}$	$\varepsilon_{\mathbf{R}}$	$\varepsilon_{\mathbf{t},\text{dir}}$	$\varepsilon_{\mathbf{R}}$	$\varepsilon_{\mathbf{t},\text{dir}}$	$\varepsilon_{\mathbf{R}}$	$\varepsilon_{\mathbf{t},\text{dir}}$	$\varepsilon_{\mathbf{R}}$	$\varepsilon_{\mathbf{t},\text{dir}}$
00 (4541 images)	0.139	2.412	0.130	2.400	0.229	4.007	0.142	2.499	**0.121**	**2.184**
01 (1101 images)	0.158	5.231	0.171	4.102	0.762	41.19	0.146	3.654	**0.136**	**2.821**
02 (4661 images)	0.123	1.740	0.126	1.739	0.186	2.508	0.121	1.702	**0.120**	**1.696**
03 (801 images)	0.115	2.744	0.108	2.805	0.265	6.191	0.113	2.731	**0.097**	**2.428**
04 (271 images)	0.099	1.560	0.116	1.746	0.202	3.619	0.100	1.725	**0.090**	**1.552**
05 (2761 images)	0.119	2.289	0.112	2.281	0.199	4.155	0.116	2.273	**0.103**	**2.239**
06 (1101 images)	0.116	2.071	0.118	1.862	0.168	2.739	0.115	1.956	**0.106**	**1.788**
07 (1101 images)	0.119	3.002	**0.112**	3.029	0.245	6.397	0.137	2.892	0.123	**2.743**
08 (4071 images)	0.116	2.386	0.111	2.349	0.196	3.909	0.108	2.344	**0.089**	**2.235**
09 (1591 images)	0.133	1.977	0.125	1.806	0.179	2.592	0.124	1.876	**0.116**	**1.644**
10 (1201 images)	0.127	1.889	**0.115**	1.893	0.201	2.781	0.203	2.057	0.184	**1.687**

Table 4. Runtime of RANSAC averaged over `KITTI` sequences (unit: s).

Methods	17PC-Li [31]	8PC-Kneip [26]	6PC-Stew. [45]	6AC-Vent. [1]	**2AC method**
Mean time	52.82	10.36	79.76	6.83	4.87
Standard deviation	2.62	1.59	4.52	0.61	0.35

To ensure the fairness of the experiments, the PCs derived from the ACs are used in the PC-based solvers. Due to space limitations, the experiment results on the `EuRoc` MAV dataset are shown in the supplementary material.

Experiments on KITTI Dataset All the solvers are evaluated on `KITTI` dataset [14] collected on outdoor autonomous vehicles with a forward facing stereo camera. We treat it as a general multi-camera system by ignoring the overlap in their fields of view. The `2AC-intra` solver is tested on all the available 11 sequences, which consist of 23000 image pairs in total. The ground truth is directly given by the output of the GPS/IMU localization unit [14]. For consecutive views in each camera, the ASIFT [38] is used to establish the ACs. There are also strategies to speed up the extraction of ACs, such as MSER [36], GPU acceleration, or approximating ACs from SIFT features [18]. To deal with outlier matches, all the solvers are integrated into a RANSAC framework. To select the right solution from multiple solutions, we counted their inliers in a RANSAC-like procedure and the solution with the most inliers is chosen.

Table 3 shows the rotation and translation error of the proposed `2AC method` for `KITTI` sequences. The median error is used to evaluate the performance. It is seen that the overall performance of the `2AC method` outperforms the comparative methods in almost all cases. Moreover, to compare the advantage of computation efficiency, the RANSAC runtime averaged over all the `KITTI` sequences for the solvers is shown in Table 4. The reported runtimes represent the relative pose estimation by RANSAC combined with a minimal solver, which mainly includes hypothesis generation and best candidate selection from multiple solutions by counting their inliers. Even though some solvers are faster than

Table 5. Rotation and translation error on nuScenes sequences (unit: degree).

Part	17PC-Li [31]		8PC-Kneip [26]		6PC-Stew. [45]		6AC-Vent. [1]		**2AC method**	
	$\varepsilon_{\mathbf{R}}$	$\varepsilon_{\mathbf{t},\text{dir}}$	$\varepsilon_{\mathbf{R}}$	$\varepsilon_{\mathbf{t},\text{dir}}$	$\varepsilon_{\mathbf{R}}$	$\varepsilon_{\mathbf{t},\text{dir}}$	$\varepsilon_{\mathbf{R}}$	$\varepsilon_{\mathbf{t},\text{dir}}$	$\varepsilon_{\mathbf{R}}$	$\varepsilon_{\mathbf{t},\text{dir}}$
01 (3376 images)	0.161	2.680	0.156	2.407	0.203	2.764	0.143	2.366	**0.114**	**2.017**

the proposed `2AC method` in Table 2, our method has better overall efficiency than all the comparative methods when integrating them into the RANSAC framework. The detailed analysis is presented in the supplementary material.

Experiments on nuScenes Dataset The performance of the solvers is also tested on the `nuScenes` dataset [8], which consists of consecutive keyframes from 6 cameras. This multi-camera system provides full 360 degree field of view. We utilize all the keyframes of Part 1 for the evaluation, and there are 3376 images in total. The ground truth is given by a lidar map-based localization scheme. Similar to the experiments on `KITTI` dataset, the ASIFT detector is used to establish the ACs between consecutive views in six cameras. The proposed `2AC method` is compared with state-of-the-art methods including `17PC-Li` [31], `8PC-Kneip` [26], `6PC-Stewénius` [45] and `6AC-Ventura` [1]. All the solvers are integrated into RANSAC in order to remove outlier matches of the feature correspondences.

Table 5 shows the rotation and translation error of the proposed `2AC method` for the Part1 of `nuScenes` dataset. The median error is used to evaluate the estimation accuracy. It is demonstrated that the proposed `2AC method` offers the best performance among all the methods. In comparison with experiments on `KITTI` dataset, this experiment also demonstrates that our `2AC method` can be directly used to the relative pose estimation for the systems with more cameras.

5 Conclusion

By exploiting the geometric constraints using a special parameterization, we estimate the 6DOF relative pose of a multi-camera system using a minimal number of two ACs. The extra implicit constraints about the affine transformation constraints are found and proved. Two minimal solvers are designed for two common types of ACs across two views, *i.e.*, inter-camera and intra-camera. Moreover, three degenerate cases are proved. The framework for generating the minimal solvers is unified and versatile, and can be extended to solve various problems, *e.g.*, relative pose estimation for a monocular camera. Compared with existing minimal solvers, our solvers require fewer feature correspondences and are not restricted to special cases of multi-camera motion. Based on a series of experiments on synthetic data and three real-world image datasets, we demonstrate that our solvers can be used efficiently for ego-motion estimation and outperform the state-of-the-art methods in both accuracy and efficiency.

Acknowledgments. This work has been partially funded by the National Natural Science Foundation of China (Grant Nos. 11902349 and 11727804).

References

1. Alyousefi, K., Ventura, J.: Multi-camera motion estimation with affine correspondences. In: International Conference on Image Analysis and Recognition. pp. 417–431 (2020)
2. Barath, D., Hajder, L.: Efficient recovery of essential matrix from two affine correspondences. IEEE Transactions on Image Processing **27**(11), 5328–5337 (2018)
3. Barath, D., Kukelova, Z.: Homography from two orientation-and scale-covariant features. In: IEEE International Conference on Computer Vision. pp. 1091–1099 (2019)
4. Barath, D., Polic, M., FÃűrstner, W., Sattler, T., Pajdla, T., Kukelova, Z.: Making affine correspondences work in camera geometry computation. In: European Conference on Computer Vision. pp. 723–740 (2020)
5. Bentolila, J., Francos, J.M.: Conic epipolar constraints from affine correspondences. Computer Vision and Image Understanding **122**, 105–114 (2014)
6. Burri, M., Nikolic, J., Gohl, P., Schneider, T., Rehder, J., Omari, S., Achtelik, M.W., Siegwart, R.: The EuRoC micro aerial vehicle datasets. The International Journal of Robotics Research **35**(10), 1157–1163 (2016)
7. Byröd, M., Josephson, K., Aström, K.: Fast and stable polynomial equation solving and its application to computer vision. International Journal of Computer Vision **84**(3), 237–256 (2009)
8. Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O.: nuScenes: A multimodal dataset for autonomous driving. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 11621–11631 (2020)
9. Clipp, B., Kim, J.H., Frahm, J.M., Pollefeys, M., Hartley, R.: Robust 6dof motion estimation for non-overlapping, multi-camera systems. In: IEEE Workshop on Applications of Computer Vision. pp. 1–8. IEEE (2008)
10. Cox, D.A., Little, J., O'Shea, D.: Using algebraic geometry. Springer Science & Business Media (2006)
11. Eichhardt, I., Barath, D.: Relative pose from deep learned depth and a single affine correspondence. In: European Conference on Computer Vision. pp. 627–644 (2020)
12. Eichhardt, I., Chetverikov, D.: Affine correspondences between central cameras for rapid relative pose estimation. In: European Conference on Computer Vision. pp. 482–497 (2018)
13. Fischler, M.A., Bolles, R.C.: 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. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: The KITTI dataset. The International Journal of Robotics Research **32**(11), 1231–1237 (2013)
15. Grayson, D.R., Stillman, M.E.: Macaulay 2, a software system for research in algebraic geometry. `https://faculty.math.illinois.edu/Macaulay2/` (2002)
16. Grossberg, M.D., Nayar, S.K.: A general imaging model and a method for finding its parameters. In: IEEE International Conference on Computer Vision. vol. 2, pp. 108–115. IEEE (2001)
17. Guan, B., Zhao, J., Barath, D., Fraundorfer, F.: Efficient recovery of multi-camera motion from two affine correspondences. In: IEEE International Conference on Robotics and Automation. pp. 1305–1311 (2021)
18. Guan, B., Zhao, J., Barath, D., Fraundorfer, F.: Minimal cases for computing the generalized relative pose using affine correspondences. In: IEEE International Conference on Computer Vision. pp. 6068–6077 (2021)

19. Hajder, L., Barath, D.: Relative planar motion for vehicle-mounted cameras from a single affine correspondence. In: IEEE International Conference on Robotics and Automation. pp. 8651–8657 (2020)
20. Häne, C., Heng, L., Lee, G.H., Fraundorfer, F., Furgale, P., Sattler, T., Pollefeys, M.: 3D visual perception for self-driving cars using a multi-camera system: Calibration, mapping, localization, and obstacle detection. Image and Vision Computing **68**, 14–27 (2017)
21. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge University Press (2003)
22. Heng, L., Choi, B., Cui, Z., Geppert, M., Hu, S., Kuan, B., Liu, P., Nguyen, R., Yeo, Y.C., Geiger, A., Lee, G.H., Pollefeys, M., Sattler, T.: Project AutoVision: Localization and 3D scene perception for an autonomous vehicle with a multi-camera system. In: IEEE International Conference on Robotics and Automation. pp. 4695–4702 (2019)
23. Kim, J.H., Hartley, R., Frahm, J.M., Pollefeys, M.: Visual odometry for non-overlapping views using second-order cone programming. In: Asian Conference on Computer Vision. pp. 353–362 (2007)
24. Kim, J.H., Li, H., Hartley, R.: Motion estimation for nonoverlapping multicamera rigs: Linear algebraic and L_∞ geometric solutions. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(6), 1044–1059 (2009)
25. Kneip, L., Furgale, P.: OpenGV: A unified and generalized approach to real-time calibrated geometric vision. In: IEEE International Conference on Robotics and Automation. pp. 12034–12043 (2014)
26. Kneip, L., Li, H.: Efficient computation of relative pose for multi-camera systems. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 446–453 (2014)
27. Kneip, L., Sweeney, C., Hartley, R.: The generalized relative pose and scale problem: View-graph fusion via 2D-2D registration. In: IEEE Winter Conference on Applications of Computer Vision. pp. 1–9 (2016)
28. Larsson, V., Aström, K., Oskarsson, M.: Efficient solvers for minimal problems by syzygy-based reduction. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 820–828 (2017)
29. Lee, G.H., Faundorfer, F., Pollefeys, M.: Motion estimation for self-driving cars with a generalized camera. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2746–2753 (2013)
30. Lee, G.H., Pollefeys, M., Fraundorfer, F.: Relative pose estimation for a multi-camera system with known vertical direction. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 540–547 (2014)
31. Li, H., Hartley, R., Kim, J.h.: A linear approach to motion estimation using generalized camera models. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8 (2008)
32. Lidl, R., Niederreiter, H.: Finite Fields. Cambridge University Press (1997)
33. Lim, J., Barnes, N., Li, H.: Estimating relative camera motion from the antipodal-epipolar constraint. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(10), 1907–1914 (2010)
34. Liu, L., Li, H., Dai, Y., Pan, Q.: Robust and efficient relative pose with a multi-camera system for autonomous driving in highly dynamic environments. IEEE Transactions on Intelligent Transportation Systems **19**(8), 2432–2444 (2017)
35. Martyushev, E., Li, B.: Efficient relative pose estimation for cameras and generalized cameras in case of known relative rotation angle. Journal of Mathematical Imaging and Vision **62**, 1076–1086 (2020)

36. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions. Image and Vision Computing **22**(10), 761–767 (2004)
37. Miraldo, P., Araujo, H., Queiró, J.: Point-based calibration using a parametric representation of the general imaging model. In: IEEE International Conference on Computer Vision. pp. 2304–2311. IEEE (2011)
38. Morel, J.M., Yu, G.: ASIFT: A new framework for fully affine invariant image comparison. SIAM Journal on Imaging Sciences **2**(2), 438–469 (2009)
39. Nistér, D.: An efficient solution to the five-point relative pose problem. IEEE Transactions on Pattern Analysis and Machine Intelligence **26**(6), 756–777 (2004)
40. Pless, R.: Using many cameras as one. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–7 (2003)
41. Raposo, C., Barreto, J.P.: Theory and practice of structure-from-motion using affine correspondences. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 5470–5478 (2016)
42. Scaramuzza, D., Fraundorfer, F.: Visual odometry: The first 30 years and fundamentals. IEEE Robotics & Automation Magazine **18**(4), 80–92 (2011)
43. Schönberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 4104–4113 (2016)
44. Stewénius, H., Nistér, D., Kahl, F., Schaffalitzky, F.: A minimal solution for relative pose with unknown focal length. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 789–794 (2005)
45. Stewénius, H., Oskarsson, M., Aström, K., Nistér, D.: Solutions to minimal generalized relative pose problems. In: Workshop on Omnidirectional Vision in conjunction with ICCV. pp. 1–8 (2005)
46. Sturm, P., Ramalingam, S.: A generic concept for camera calibration. In: European Conference on Computer Vision. vol. 3022, pp. 1–13. Springer-Verlag (2004)
47. Sweeney, C., Flynn, J., Turk, M.: Solving for relative pose with a partially known rotation is a quadratic eigenvalue problem. In: International Conference on 3D Vision. pp. 483–490 (2014)
48. Ventura, J., Arth, C., Lepetit, V.: An efficient minimal solution for multi-camera motion. In: IEEE International Conference on Computer Vision. pp. 747–755 (2015)
49. Zhao, J., Kneip, L., He, Y., Ma, J.: Minimal case relative pose computation using ray-point-ray features. IEEE Transactions on Pattern Analysis and Machine Intelligence **42**(5), 1176–1190 (2020)
50. Zhao, J., Xu, W., Kneip, L.: A certifiably globally optimal solution to generalized essential matrix estimation. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 12034–12043 (2020)
51. Zheng, E., Wu, C.: Structure from motion using structure-less resection. In: IEEE International Conference on Computer Vision. pp. 2075–2083 (2015)