Relative Pose Estimation of Calibrated Cameras with Known SE(3) Invariants

Bo Li[1][0000−0002−9336−1862], Evgeniy Martyushev[2][0000−0002−6892−079X],
Gim Hee Lee[1][0000−0002−1583−0475]

[1] National University of Singapore, Singapore
prclibo@gmail.com, gimhee.lee@nus.edu.sg
[2] South Ural State University, 454080 Chelyabinsk, Russia
martiushevev@susu.ru

Abstract. The SE(3) invariants of a pose include its rotation angle and screw translation. In this paper, we present a complete comprehensive study of the relative pose estimation problem for a calibrated camera constrained by known SE(3) invariant, which involves 5 minimal problems in total. These problems reduces the minimal number of point pairs for relative pose estimation and improves the estimation efficiency and robustness. The SE(3) invariant constraints can come from extra sensor measurements or motion assumption. Unlike conventional relative pose estimation with extra constraints, no extrinsic calibration is required to transform the constraints to the camera frame. This advantage comes from the invariance of SE(3) invariants cross different coordinate systems on a rigid body and makes the solvers more convenient and flexible in practical applications. In addition to the concept of relative pose estimation constrained by SE(3) invariants, we also present a comprehensive study of existing polynomial formulations for relative pose estimation and discover their relationship. Different formulations are carefully chosen for each proposed problems to achieve best efficiency. Experiments on synthetic and real data shows performance improvement compared to conventional relative pose estimation methods. Our source code is available at: http://github.com/prclibo/relative_pose.

1 Introduction

Minimal relative pose solver of a camera is a fundamental component in modern 3D vision applications including robot localization and mapping, augmented reality, autonomous driving, 3D modeling, etc. Well-known solvers include the 7-point algorithm [11] and the 5-point algorithm [30,39]. It is generally admitted that an n-point solver with smaller n performs more robustly, has less degenerate configurations and requires less iterations when integrated in a RANSAC framework.

As the first contribution of this paper, we show that two measurements – rotation angle and screw translation – can be respectively integrated into relative pose solvers to reduce the number n of minimal points. Typical scenarios for these

measurements include a robot equipped with a camera and an IMU, and a robot with planar motion. These measurements are referred as SE(3) invariants as they stay invariant cross different coordinate systems on a rigid body. Consequently, the proposed methods do not require known extrinsic pose of the camera with respect to the IMU or the motion plane, which is an important advantage over previous relative pose estimation methods [8,13,23,35,4,36,20]. This advantage make the proposed methods more flexible and convenient. For example, when estimating visual odometry to hand-eye calibrate the camera-IMU extrinsics, the proposed methods improve trajectory estimation even though the extrinsics are unavailable. For robot systems subjected to long term operation, the proposed methods avoid re-calibration. All the 5 minimal problems introduced by different combination of SE(3) invariants as constraints are comprehensively studied in this paper.

The second contribution is a comprehensive study of all existing polynomial formulations for relative pose solvers. We show pros and cons of each formulation under different minimal problem settings and reveal connections between the formulations. For each proposed relative pose problem with SE(3) invariants, we evaluate these formulations and propose solvers with the best efficiency.

2 Related Works

A fundamental matrix for a pair of pinhole cameras has 7 DoF and can be estimated minimally from 7 and linearly from 8 point correspondences [10,11,25]. The estimation can be reduced to the 6-point algorithm [12,16,40] when all camera intrinsics except a common focal length are calibrated. In the case where the focal length is also calibrated, the 5-point algorithm [30,39] is naturally introduced.

Beyond 5-point solution, extra constraints can be exploited. With known gravity direction measured by IMU or by the knowledge of motion plane, [8,13,35,4] obtain two rotation angles and hence reduce the minimal number of point pairs to 3. Camera extrinsics are required to be calibrated for these methods to transform the two angles to the camera frame. In [36], it is assumed that the camera follows the Ackermann motion. In this case only one point is needed for pose estimation. This method requires the camera to be specifically mounted. [23] proposes the 4-point algorithm given a known rotation angle measurement from other sensors. This is the first work on integrating SE(3) invariants in relative pose estimation. The known rotation angle can also be used for the camera self-calibration as demonstrated in [27]. As mentioned in the previous section, extrinsic calibration is not required to use SE(3) invariants.

Minimal problems are usually formulated in terms of multivariate polynomial systems and a plenty of methods have been proposed to solve these systems. Some of these methods make use of the Gröbner basis computation [16,15,27,28,39]. The roots are then derived from the eigenvectors of the so-called action matrix constructed from the Gröbner basis [5]. Apart from the action matrix, alternative matrix decomposition methods are also proposed, e.g. PolyEig [17] and

QuEst [33]. To avoid significant computational cost of matrix decomposition, the hidden variable approach has been used in several solvers [12,30]. This approach reduces the problem to finding real roots of a univariate polynomial.

3   Preliminaries

3.1   Notation

We preferably use α, β, ... for scalars, a, b, ... for column 3-vectors, and A, B, ... for matrices. For a matrix A, the transpose is $A^\top$, the determinant is $\det A$, and the trace is $\operatorname{tr} A$. For two 3-vectors a and b the cross product is $a \times b$. For a vector a, the entries are a_i, the notation $[a]_\times$ stands for the skew-symmetric matrix such that $[a]_\times b = a \times b$ for any vector b. We use I for the identity matrix and $\| \cdot \|$ for the Frobenius norm. A notation f_*^* is used to refer a polynomial.

A rotation matrix R can be represented by a unit quaternion $[\sigma \; u^\top]$ as follows

$$R = 2(uu^\top - \sigma[u]_\times) + (\sigma^2 - \|u\|^2)I, \tag{1}$$

$$\mathsf{f}^\sigma := \|u\|^2 + \sigma^2 - 1 = 0. \tag{2}$$

With θ as the rotation angle, we have

$$\operatorname{tr} R = 4\sigma^2 - 1 = 2\cos\theta + 1, \tag{3}$$

Another way to represent a rotation matrix R comes from the Cayley transform if and only if it is not a rotation through an angle $\pi + 2\pi k$.

$$R = (I - [v]_\times)(I + [v]_\times)^{-1}, \tag{4}$$

where $v = u/\sigma$ is a 3-vector.

The special Euclidean group SE(3) consists of all orientation-preserving rigid motions of 3-dimensional Euclidean space. Any element $H \in \mathrm{SE}(3)$ can be represented by a 4×4 matrix of the form

$$H = \begin{bmatrix} R & t \\ 0^\top & 1 \end{bmatrix}, \tag{5}$$

where $R \in \mathrm{SO}(3)$ and $t \in \mathbb{R}^3$ are the rotational and translational parts of H respectively. In the sequel, saying about elements of group SE(3) we always imply 4×4 matrices of type (5).

3.2   Epipolar Constraint

Let $P' = [R' \; t']$ and $P'' = [R'' \; t'']$, where $R', R'' \in \mathrm{SO}(3)$ and $t', t'' \in \mathbb{R}^3$, be calibrated camera matrices. Let q_i' and q_i'' be the corresponding images of a 3D point Q_i. Then the epipolar constraint reads

$$\mathsf{f}_i := q_i''^\top (R[t']_\times - [t'']_\times R)q_i' = 0, \tag{6}$$

where i counts the point pairs and $R = R''R'^{\top}$ is called the relative rotation matrix. We notice that Eq. (6) can be rewritten in form

$$q_i''^{\top}[t]_{\times}Rq_i' = 0, \tag{7}$$

where $t = Rt' - t''$ is called the relative translation. Matrix $E = [t]_{\times}R$ is well known in the computer vision community as an essential matrix.

3.3 SE(3) invariants

Given an element $H \in \mathrm{SE}(3)$, with its rotational part R represented by (1). We denote the unit rotation axis of R by $r = \frac{u}{\|u\|}$, then the value

$$\delta = r^{\top}t \tag{8}$$

is the *screw translation* of H. In this paper, we are specifically interested in the case of $\delta = 0$, which is also equivalent to

$$\mathsf{f}^0 := u^{\top}t = 0. \tag{9}$$

Consider a robot with planar motion. Its rotation axis r must be the normal vector of the motion plane, and its translation vector t must lie on the motion plane. Therefore, it is obvious that the condition of zero screw translation ($\delta = 0$) holds for any planar motion regardless of the camera orientation with respect to the ground plane direction.

Figure 1 illustrates the definition of θ and δ. We refer them as the SE(3) invariants, i.e. scalar values invariant under the conjugation by an SE(3) element. In robotics, this conjugation is known as the hand eye transformation. The difference between SE(3) invariants and an easily mixed-up concept bi-invariant metrics can be found from [3].

Theorem 1 (SE(3) **Invariants**). *For a transform $H \in \mathrm{SE}(3)$, its rotation angle θ and screw translation δ are invariant under the hand eye transformation $H' = X^{-1}HX$ with $X \in \mathrm{SE}(3)$.*

Proof. Let the rotational and translational parts of H, H', X be R, R', R_X and t, t', t_X respectively. Then we have

$$H' = \begin{bmatrix} R_X^{\top} & -R_X^{\top}t_X \\ 0^{\top} & 1 \end{bmatrix} \begin{bmatrix} R & t \\ 0^{\top} & 1 \end{bmatrix} \begin{bmatrix} R_X & t_X \\ 0^{\top} & 1 \end{bmatrix} = \begin{bmatrix} R_X^{\top}RR_X & R_X^{\top}(Rt_X - t_X + t) \\ 0^{\top} & 1 \end{bmatrix}, \tag{10}$$

that is $R' = R_X^{\top}RR_X$ and $t' = R_X^{\top}(Rt_X - t_X + t)$. The invariance of θ follows from Eq. (3), since $\mathrm{tr}(R') = \mathrm{tr}(R_X^{\top}RR_X) = \mathrm{tr}(R)$.

Furthermore, let r and r' be the unit rotation axes of R and R', respectively. It is clear that $r = Rr$ and $r' = R'r' = R_X^{\top}RR_Xr'$ (Lemma II [43]). Hence, axes r and r' are related by $r' = R_X^{\top}r$. Substituting this into the definition of δ' yields

$$\delta' = r'^{\top}t' = r^{\top}R_XR_X^{\top}(Rt_X - t_X + t) = r^{\top}Rt_X - r^{\top}t_X + r^{\top}t = r^{\top}t = \delta. \tag{11}$$

$\square$

Theorem 1 is a well-known result in robotics [2,3]. Different proof for the rotation part can also be found from [38,2,3].
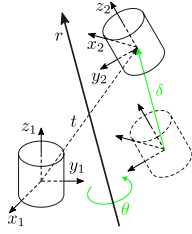
Fig. 1. A rigid motion in SE(3) can always be decomposed as a rotation around an axis r with angle θ, and a screw translation δ along r. θ and δ remain consistent for different parts of the rigid body, regardless of the part offset and the local coordinate system.

4   Minimal Problem Formulations

The relative pose estimation problem aims to solve for the relative rotation R and relative translation t given several image point pairs. It is well known that if no additional constraints are used, the relative pose can be estimated minimally from 5 point pairs [30,39,12]. With known SE(3) invariants (rotation angle θ and screw translation δ), the number of point pairs required for a minimal solution is reduced. Table 1 summarizes the minimal relative pose estimation problems that can be formulated for different combinations of image data and SE(3) invariants.

Problem	SE(3) Inv	n	DoF	Constraints
5P [12,30,39]	-	5	5	f_1, f_2, f_3, f_4, f_5
4P-RA [23,28]	θ	4	5	$f_1, f_2, f_3, f_4, f^\sigma$
4P-ST0	$\delta = 0$	4	5	f_1, f_2, f_3, f_4, f^0
3P-RA-ST0	$\theta, \delta = 0$	3	5	$f_1, f_2, f_3, f^\sigma, f^0$
5P-ST1	$\delta \neq 0$	5	6	f_1, f_2, f_3, f_4, f_5
4P-RA-ST1	$\theta, \delta \neq 0$	4	6	$f_1, f_2, f_3, f_4, f^\sigma$

Table 1. Relative pose problems with SE(3) invariants.   RA: Relative angle (θ)   ST0: Zero screw translation ($\delta = 0$)   ST1: Non-zero screw translation ($\delta \neq 0$) n: Number of points for minimal cases

Remark 1. In 5P, 4P-RA, 4P-ST0 and 3P-RA-ST0, each of SE(3) invariants f^σ and f^0 can replace one of point correspondences f_i as constraints and the relative pose can be estimated up to an ambiguous scale (5 DoF). In 5P-ST1 and 4P-RA-ST1, condition (8) with $\delta \neq 0$ can not be used to replace point correspondences as the essential matrix does not change regardless of the value of a non-zero δ. Instead, δ can be used to determine the length of translation, and hence the overall scale is observable (6 DoF). 5P-ST1 and 4P-RA-ST1 are therefore equivalent to problems 5P and 4P-RA, respectively.

Remark 2 (Beyond Pinhole Cameras). In this paper, we only focus on relative pose estimation for a pinhole camera. However, it is worth mentioning the difference between pinhole camera and generalized camera models under known SE(3) invariants. For a generalized camera model [32,41], the relative translation length is observable. The vanilla version of relative pose estimation problem requires 6 points to fully recover the 6 DoF relative pose. When the screw translation is known, regardless of being zero or non-zero, 5 points are required to fully recover the 6 DoF relative pose. The case of known relative rotation angle for generalized cameras was covered in [28].

5 Solution Formulations for Relative Pose Estimation

Various solutions have been studied in the past decades on relative estimation problems, i.e. to solve (6) or (7) under different constraints. We provide a comprehensive summary for these formulations in this section and also discover how SE(3) invariants can be denoted for each formulation. All mentioned previous formulations are also listed in Table 2.

Problem	Form	#R	#t	Templ	#S
PC 5P [15]	SIR3	9 (R)	3	66×197	20
PC 5P [6]	SIR6	4 (u,σ)	3	40×56	35
PC 5P [14]	Direct	4 (u,σ)	3	NR	80
PC+θ [23]	SIR3	3 (u)	3	270×290	20
PC+θ [28]	SIR2	3 (u)	2	16×36	20
PC+g (IMU) [13]	SIR3	1 (yaw)	3	CF	4
PC+g (pl) [4]	Direct	1 (yaw)	1	CF	4
PC+Ack [37]	Direct	1 (yaw)	0	CF	1
GC 6P [39]	SIR2	3 (v)	2	60×120[1]	64
GC+θ [28]	SIR2	3 (u)	2	37×81	44
GC+g (IMU) [20]	SIR3	1 (yaw)	3	CF	8
GC+g (pl) [22]	SIR3	1 (yaw)	2	CF	6
GC+Ack [21]	Direct	1 (yaw)	1	CF	2

Problem	Form	Templ	#S
PC 5P [30,39]	NullE	10×20	10
PC−fc [16,40]	NullE	31×46	15
PC+θ−fc−pp [27]	NullE	19×32[2]	6
PC+g (IMU) [8]	NullE	6×10	4

* #R, #t: Number of parameters
* #S: Number of solutions
* +/−: With a constraint or an unknown
* PC/GC: Pinhole camera / generalized camera
* NR: Not reported
* CF: Closed form solution w/o template matrix
* g: Vertical direction
* Ack: Ackermann motion model
* fc/pp: Focal length / principle point

[1] Insufficient for full Gröbner basis generation

[2] The largest of cascaded templates reported

Table 2. Representative works of polynomial formulations for relative pose estimation.

5.1 Solutions by Decomposing E

Directly Solving R and t. As an intuitive start, it is possible to directly solve (6), (7) by considering R and t as polynomial unknowns. In [23], R is parameterized by angle-axis to integrate rotation angle and t is constrained by $\|t\|^2 = 1$ to remove scale ambiguity. In theory R can be also parameterized by a quaternion with constraint of f^σ or by 9 matrix elements with constraint of $R^\top R = I$. However, these formulations involve $6 \sim 12$ unknowns (including 3 for translation) and require quite complicated polynomial elimination, which makes high computational burden in real-time applications. A simpler specialization is when the vertical direction is known from IMU [13] or known ground plane

[4], R can be parameterized by a yaw angle rotation. With Ackermann motion assumed, the parameters can even be further reduced [21,37].

Since each f_i is linear in t, cf. Eq. (7), several formulations have been proposed in previous works to eliminate translation variables and solve rotation parameters only for simplicity.

SIR3: Solving Isolated Rotation by a 3×3 Determinant. We can isolate unknown translation by rewriting the epipolar constraints (7) for n point pairs in the form

$$G\ t = 0, \tag{12}$$

where G is a matrix of size $n \times 3$. Elements of the i-th row of matrix G are polynomials in unknown rotation parameters and known q_i', q_i''. It follows from Eq. (12) that all 3×3 minors of G must vanish for a valid translation. Thus we obtain new polynomial constraints of the total degree 6 on the rotation parameters only. This solution formulation will be further referred to as *SIR3*. If the rotation matrix is represented by (1) (resp. by (4)), then the formulation is denoted by SIR3+u (resp. SIR3+v). When rotation is constrained to have only one unknown, SIR3 generates a closed form univariate polynomial [13,20]. However, with more unknowns in rotation, it is still not satisfactory as leads to large matrix templates for the Gröbner basis computation. For example, in [15] SIR3 is used to solve 5P by a reduction on a 66×197 matrix. In [23], the SIR3+u formulation of 4P-RA involves even larger template matrix and has to be solved by numerical search.

The known rotation angle can be easily integrated into SIR3+u. If the motion is planar, i.e. the screw translation is zero, polynomial f^0 from (9) can be written as a new row $u^\top$ of matrix G.

SIR6: Solving Isolated Rotation by a 6×6 Determinant. Let Q_i be the i-th 3D point so that

$$\lambda_i q_i' = P'Q_i, \quad \mu_i q_i'' = P''Q_i, \tag{13}$$

where λ_i and μ_i are some scalars. The relative pose R and t satisfies $\lambda_i q_i'' = \mu_i Rq_i' + t$. Consider this equation for the i, j, k-th points and subtract the i-th equation over the j-th and k-th respectively to eliminate t. We can obtain two 3D linear equations, forming a 6×6 matrix M_{ijk}.

$$M_{ijk}[\lambda_i\ \mu_i\ \lambda_j\ \mu_j\ \lambda_k\ \mu_k]^\top = 0. \tag{14}$$

Similar to SIR3, the determinant of M_{ijk} must vanish, resulting in polynomials on the rotation parameters only. SIR6 is proposed by [6] as an alternative solution to 5P. Symbolic computation reveals the relationship between SIR6 and SIR3:

Theorem 2. *Consider a 3×3 submatrix G_{ijk} of matrix G whose three rows correspond to the i-th, j-th, and k-th point pairs. We have $\det G_{ijk} = \det M_{ijk}$, up to a sign.*

SIR2: Solving Isolated Rotation by a 2×2 Determinant. Using the rigid motion ambiguity of the world coordinate frame, we set $Q_i = [0\ 0\ 0\ 1]^\top$ in (13) for a certain i. This yields

$$t' = \lambda_i q'_i,\ t'' = \mu_i q''_i. \tag{15}$$

Substituting t' and t'' into Eq. (6) for a j-th pair with $j \neq i$, we convert Eq. (6) into $F_{ij} \begin{bmatrix} \lambda_i\ \mu_i \end{bmatrix}^\top = 0$. Construct F_{ik} from a k-th point pair and stack with F_{ij} as F_{ijk}, we have

$$F_{ijk} \begin{bmatrix} \lambda_i\ \mu_i \end{bmatrix}^\top = 0. \tag{16}$$

where matrix F_{ijk} is of size 2×2. F_{ijk} must have zero determinant. This leads to degree 4 polynomial equations in the rotation parameters. The proposed solution formulation will be further referred to as *SIR2*. The two versions of SIR2 corresponding to the quaternion (1) and Cayley (4) parametrizations are denoted by SIR2+u and SIR2+v respectively. The SIR2+v form was earlier used in [41] for solving the relative pose problem for generalized cameras. In [28], SIR2+u was used to solve 4P-RA. Symbolic computation also reveals the connection between SIR2 and SIR3 and explains why SIR2 is a simpler formulation compared to SIR3/SIR6.

Theorem 3. *Consider G_{ijk} in Theorem 2. Under SIR2/SIR3+u, we have:*

$$\det G_{ijk} = (\|u\|^2 + \sigma^2) \cdot \det F_{ijk}, \tag{17}$$

up to a sign. This equation also holds if replacing $(\|u\|^2 + \sigma^2)$ with $(\|v\|^2 + 1)$ under the SIR2/SIR3+v.

Remark 3. According to Eq. (17), since equation $\|v\|^2 + 1 = 0$ has infinite number of complex solutions, 5P in SIR3+v is not zero-dimensional over $\mathbb{C}$.

The known rotation angle can be easily integrated into SIR2+u. If the motion is planar, i.e. $\delta = 0$, then polynomial f^0 can add a row $[-u^\top q'_i\ u^\top q''_i]$ to F_{ijk}.

5.2 Solutions by Constraining E

NullE: Solving Essential Matrix Represented by NullSpace Bases. Instead of direct solving for R and t, a more classical approach to the relative pose problem is solving first for the essential matrix E which is a mixed form of rotation and translation parameters. Unknown E is parameterized by $\sum_{i=1}^{9-n} \gamma_i E^{(i)}$, where matrices $E^{(i)}$ form the nullspace basis of the underdetermined linear system $\{f_i \mid i = 1, \dots, n\}$, γ_i are new unknowns which are usually scaled so that $\gamma_1 = 1$. Traditional 5P solvers [30,39,12] use the following constraints to form a polynomial system on γ_i:

$$\det E = 0, \tag{18}$$

$$2EE^\top E - \mathrm{tr}(EE^\top)E = 0. \tag{19}$$

In addition, [8] found that known vertical can be denoted as constraints on E. Known SE(3) invariants also can be formulated as constraints on E as follows.

Theorem 4 ([27]). *Let $E = [t]_\times R$ be an essential matrix and* $\operatorname{tr} R = \tau$. *Then E fulfills the following equation*

$$\frac{1}{2}\left(\tau^2 - 1\right)\operatorname{tr}(EE^\top) + (\tau + 1)\operatorname{tr}(E^2) - \tau \operatorname{tr}^2 E = 0. \tag{20}$$

Theorem 5. *Let $E = [t]_\times R$ be an essential matrix and R be a rotation through an angle θ around a vector $r \neq 0$. If $\delta = r^\top t = 0$, then*

$$\operatorname{tr} E = 0. \tag{21}$$

Conversely, if $\operatorname{tr} E = 0$, *then either* $\delta = 0$ *or* $\theta = \pi k$ *for a certain integer* k.

Proof. We utilize Proposition 2.20 from [29] that $\operatorname{tr} E = -2\sin\theta(r^\top t)$ and the statement follows. □

Using Eqs. (18) – (21), problems 4P-RA, 4P-ST0, and 3P-RA-ST0 can be formulated in terms of matrix E. Table 3 summarizes these formulations with reference template matrix size generated by automatic polynomial solver generators [16] and [24].

Remark 4. As is well known, the 5P problem in the NullE formulation has 10 solutions. Each essential matrix corresponds to a twisted pair of rotations [11] and each rotation, being represented by a unit quaternion, doubles due to the sign ambiguity. Therefore, as shown in Table 3, the 5P problem in SIR2/SIR3+u has 40 solutions while in SIR2/SIR3+v has only 20 solutions. The 4P-RA problem in the NullE formulation has 20 solutions, corresponding to 40 rotations. For each pair of rotations, there is a unique one whose rotation angle equals known θ. Similarly for 4P-ST0 in NullE, each pair of rotations corresponding to an essential matrix contains a unique valid rotation.

Remark 5. The 3P-RA-ST0 problem has 12 solutions in SIR2+u. However, in NullE the system consisting of Eqs. (18) – (21) has 20 solutions. The obtained contradiction indicates that there must exist additional polynomial constraints on essential matrix E. Using the implicitization algorithm [5], we found that the entries of E additionally satisfy 7 cubic equations. We provide them in the supplementary material. The above polynomial system complemented with the new 7 cubics has 12 solutions (NullEx in Table 3). However, due to the lack of geometric interpretability for the additional cubics, in this paper we use a hand-crafted solver with slightly larger template matrix.

6 Minimal Relative Pose Solvers with SE(3) Constraints

The goodness of different solver formulations is reflected by the size of the matrix template for Gröbner basis computation since it directly affects both the speed and numerical accuracy of a minimal solver. Different formulated solvers are reported in Table 3. We compared our proposed formulation with [16], the most widely used generator the past years, and [24], a wrapper of a newer generator [19].

Problem	Form	#V	D	#S	AG	GAPS	Proposed
5P	SIR3+u	4	6	40	146×186	116×200[1]	-
5P	SIR3+v	3	6	∞	-	-	-
5P	SIR2+u	4	4	40	90×130	60×80[1]	-
5P	SIR2+v	3	4	20	31×51	36×56	-
5P	NullE	3	3	10	10×20	10×20	-
4P-RA	SIR2+u	3	4	20	26×46	36×56	16×36 [28]
4P-RA	NullE	4	3	20	34×54[2]	50×70	-
4P-ST0	SIR2+u	3	4	20	62×82[2]	38×65[1]	-
4P-ST0	SIR2+v	3	4	10	25×35	27×35	-
4P-ST0	NullE	3	3	10	10×20	10×20	10×20
3P-RA-ST0	SIR2+u	3	4	12	23×35[2]	28×35	13×25
3P-RA-ST0	NullE	5	3	20	34×54	50×70	-
3P-RA-ST0	NullEx	5	3	12	22×35[2]	53×65	-

Table 3. Comparison of different solution formulations for the minimal relative pose problems with SE(3) invariants.   #V: Number of variables   D: Highest degree #S: Number of solutions   AG: Generator from [16]   GAPS: Generator from [24] 1: Mirrored roots u merged by [18]   2: The largest of cascaded templates reported

6.1   5P, 4P-RA, 5P-ST1 and 4P-RA-ST1

NullE is the most widely used polynomial formulation for 5P, with a template matrix of 10×20. Template matrix of 4P-RA was recently reduced from 270×290 to 16×36 using SIR2-u [28]. 5P-ST1 can be solved by a 5P solver and multiply the unit translation solution t by $\frac{\delta}{r^\top t}$. 4P-RA-ST1 can be solved by a 4P-RA solver in the same way.

6.2   4P-ST0

The NullE formulation for 4P-ST0 produces the smallest template of size 10×20. Note that in NullE of 4P-ST0, Eq. (21) replaces an epipolar constraint of 5P and they are both linear on E. Therefore, 4P-ST0 can be simply solved by a NullE 5P solver by replacing the coefficients of one epipolar constraint.

6.3   3P-RA-ST0

For problem 3P-RA-ST0, the SIR2+u formulation is preferable as it leads to the smallest 13×25 matrix template. The algorithm is summarized as follows.

Three image point pairs are first used to form a 2×2 matrix F_{123}, see Subsect. 5.1. We set $u^\top = \begin{bmatrix} \alpha \ \beta \ \gamma \end{bmatrix}$. Then our system consists of the following polynomial equations:

- 10 equations of $m \cdot f^\sigma = 0$ for m being every monomial with degree up to 2;
- 1 equation $\det F_{123} = 0$;

– 12 equations of $m \cdot \det F'_{ij} = 0$, with $i \neq j$, $m \in \{\alpha, \beta, \gamma, 1\}$ and

$$F'_{ij} = \begin{bmatrix} F_{ij} \\ -u^\top q'_i \, u^\top q''_i \end{bmatrix}.$$

In matrix form the system can be written as $Ax = 0_{23 \times 1}$, where A is the 23×35 coefficient matrix whose i-th row consists of coefficients of the i-th polynomial, x is a monomial vector. Matrix A is exactly the template produced by the Automatic Generator, see Table 3. However, the template's size can be further reduced if we take into account the special structure of matrix A. Namely, if the first 10 monomials in x are

$$\alpha^4, \ \alpha^3\beta, \ \alpha^2\beta^2, \ \alpha^3\gamma, \ \alpha^2\beta\gamma, \ \alpha^2\gamma^2, \ \alpha^3, \ \alpha^2\beta, \ \alpha^2\gamma, \ \alpha^2, \tag{22}$$

then matrix A has the following block form $A = \begin{bmatrix} U & V \\ W & X \end{bmatrix}$, where U is an upper-triangular 10×10 matrix with 1's on its main diagonal. We conclude that matrix A is equivalent to $\begin{bmatrix} U & V \\ 0_{13 \times 10} & B \end{bmatrix}$, where matrix $B = X - WU^{-1}V$ is our final template of size 13×25. Matrix B contains all necessary data for deriving solutions either by constructing an action matrix or by forming the 12-th degree univariate polynomial in accordance with the hidden variable method. We provided more details on the 3P-RA-ST0 solver in the supplementary material. Readers can refer to [7,28] for more usage of the above simplification.

Handling Degeneracy

Condition (8) becomes degenerate when the rotation matrix is close to I. In this case the rotation axis r is ill-posed and vector u becomes arbitrarily small. Enforcing condition (9) in this case might lead to a large deviation in the direction of translation. Nevertheless, this degenerate case can be easily covered by fitting relative pose to a translation-only motion (2P-TO). The skew-symmetric essential matrix $[t]_\times$ can be easily estimated from two image feature pairs. In this paper, we estimate 4P-ST0 and 3P-RA-ST0 together with 2P-TO and accept the results with more inliers.

7   Experiments

7.1   Implementation Details

All algorithms compared in experiments are implemented by C++. The hidden variable method is used to derive solutions of polynomial systems. Roots of univariate polynomials are found using Sturm sequences. We implement 4P-RA [28], 4P-ST0, and 3P-RA-ST0. The C++ 5P solver from [12] is used, which is regarded as the state-of-the-art fast implementation. Runtime statistics on an i5-4288U is listed in Table 4.
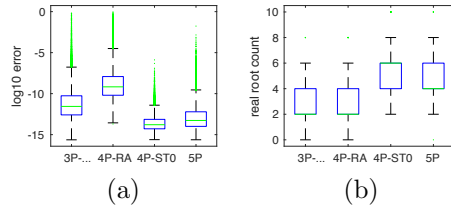
Fig. 2. (a) Numerical accuracy comparison of the solvers; (b) Statistics on the number of real roots for each solver

Minimal Solver	3P-RA-ST0	4P-RA	4P-ST0	5P
Average Time	28 μs	34 μs	26 μs	25 μs

**Table 4.** Average runtime comparison of the solvers

7.2 Synthetic Data

Synthetic data are generated to illustrate the algorithm performance. Synthetic image features are generated from a 60° field of view with focal length in 500px. We test algorithm performance under Gaussian image noise whose std ranges in 0–1px. Synthetic data is generated for forward motion and sideway motion. Rotation angle of a pose pair is randomly generated from Gaussian with std of 5°. The rotation angle measurement is disturbed by Gaussian noise (derived from the widely used Brownian process model for IMU noise) with std ranging in 0–1°. To test the performance of 4P-ST0 and 3P-RA-ST0 under non-perfectly planar motion, we first produces unit translation with zero component on rotation axis. Then the translation is disturbed along the rotation axis with Gaussian noise whose std ranges in 0–5%.

The numerical accuracy of each algorithm is compared and listed in Fig. 2(a). The numerical error is measured by the value $\min_i \|R_i - \bar{R}\|$, where i counts all real solutions and $\bar{R}$ is the ground truth relative rotation matrix. The number of real roots is also counted for each algorithm and listed in Fig. 2(b). We observe that 4P-ST0 in NullE formulation generally has more real roots compared to 5P. The number of real roots also affects the computational efficiency in some RANSAC frameworks like OpenCV where each real solution must be verified by computing the reprojection error over all image feature pairs.

In the experiments on both synthetic data and real data, the error of rotation is measured by the rotation angle between the estimated and groundtruth rotation. The error of translation is measure by the angle between the unit groundtruth translation and the estimated translation. Forward motion and sideway motion are experimented separately. The mean estimation error of 4P-ST0 and 5P against image ray disturb is shown in Fig. 3. The estimation is executed on 100 image feature pairs with 30% outliers under RANSAC. Green curves from bottom to top represent 4P-ST0 estimation with different screw translation disturbance along rotation axis {0%, 1.66%, 3.33%, 5%}. As it is mentioned in many previous works, the rotation error is generally small for different solvers. Regarding translation error, the advantage of 4P-ST0 over 5P is more significant for forward motion, which is considered as the more common and difficult case than sideway motion.
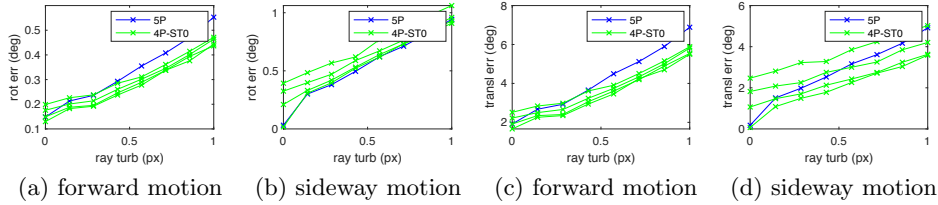
Fig. 3. Estimation error plot of 4P-ST0 and 5P on synthetic data: (a, b) rotation errors; (c, d) translation errors
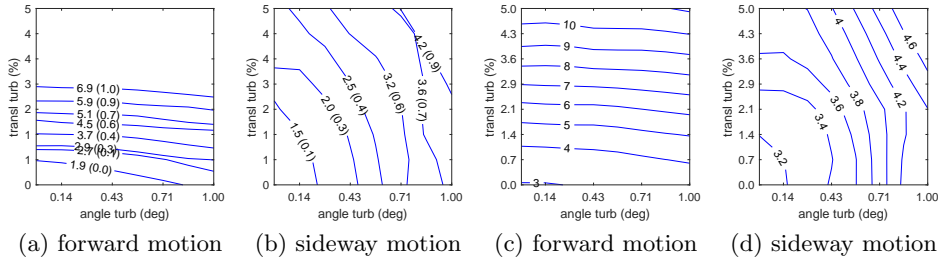
Fig. 4. Estimation error comparison of 3P-RA-ST0 and 5P: (a, b) contour curves $\xi(\epsilon)$ denote that under image feature noise std ϵ px, when the error of rotation angle and screw translation is at the bottom left of the curve, 3P-RA-ST0 outperforms 5P with translation error no more than ξ; (c, d) translation error contour curve over different rotation angle and screw translation disturbance, with image feature noise std fixed as 1px

The performance of 3P-RA-ST0 is affected by rotation angle error, screw translation error and also image feature error. To simplify our visualization, in Fig. 4(a) and Fig. 4(b), we fix the image feature error at different level and consider the translation error surface against the rotation angle and screw translation errors. The intersection contour of this surface with the error surface of 5P is plotted. The bottom-left area of each curve denotes the error level under which 3P-RA-ST0 can outperform 5P. We only compare the translation error here as the rotation error is generally similar for different methods. Compared to 4P-ST0, we find that 3P-RA-ST0 is more sensitive to screw translation error in forward motion. For example, with perfect rotation angle and image feature noise std as 1px in forward motion, 3P-RA-ST0 outperforms 5P only when screw translation error is less than 3% (Fig. 4(a)), while 4P-ST0 outperforms 5P even when screw translation error is 5% (Fig. 3(c)).

7.3 Real-World Data

We compare our approaches on multiple datasets collected on indoor mobile robots or outdoor autonomous vehicles, which are two popular modern robot applications with planar motion. Experimented datasets include:

- RawSeeds-Bicocca [1]: Indoor mobile robot data, with IMU and odometer available for rotation angle. Front camera (FC) images are used.

- TUM-RGBD-SLAM [42], Robot@Home [34]: Indoor mobile robot data, without available angle measurement. Images from the front RGBD camera (FC) are used. The left RGBD camera (LC) of Robot@Home is also experimented.

- KITTI [9], UMich [31], RobotCar [26]: Autonomous vehicle data, with fused GPS/INS data available. Front camera (FC) images are used. The left camera (LC) of UMich is also used. Rotation angle from the fused GPS/INS pose is used.

Consecutive image pairs with translational movement larger than 0.1m in indoor data and 1m in outdoor data are used in experiments. Performance comparison on indoor data is shown in Appendix Table 1. It is seen that 4P-ST0 outperforms 5P in almost all cases, which is consistent with the synthetic data results. With IMU data on RawSeeds, 3P-RA-ST0 further improves the estimation. Results with odometry angle have no improvement, implying the accuracy of odometry angle might be low in this dataset.

However, the performance varies on autonomous driving data in Appendix Table 2. We observe that 4P-ST0 outperforms 5P for broad road or highway environments. For urban narrow road environments, 5P has the better accuracy. The difference in the environments corresponds to different screw translation disturbance on a planar motion assumption as roads are less planar and vehicles might tilt more on urban road. With further analysis, we found that a portion of relative poses have screw translation of more than 20% ($r^\top t/\|t\| > 0.2$) in urban autonomous driving scenarios, which explains the poor performance of 4P-ST0.

We also note that 3P-RA-ST0 performs better on UMich left camera than 4P-ST0, corresponding to the observation in Fig. 4 that 3P-RA-ST0 is less sensitive to screw translation error for sideway motion.

8   Conclusions

In this paper, we show that known SE(3) invariants can be used to constrain the minimal relative pose estimation problem. Compared to existing relative pose problems with contraints, the proposed methods are more flexible and convenient since extrinsics are not required to transform SE(3) invariant to the camera frame. We also comprehensively revise and relate to each other existing formulations of the relative pose problem. The discovered relationship provides a deeper understanding to these previous methods. This knowledge help formulate the most efficient solvers for the proposed relative pose problem with SE(3) constraints. A series of experiments on synthetic and real datasets show practicality of the proposed solvers in robotic perception especially for indoor robots.

References

1. Bonarini, A., Burgard, W., Fontana, G., Matteucci, M., Sorrenti, D.G., Tardos, J.D.: Rawseeds: Robotics advancement through web-publishing of sensorial and elaborated extensive data sets. In: In proceedings of IROS. vol. 6 (2006)
2. Chen, H.: A screw motion approach to uniqueness analysis of head-eye geometry. In: Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 145–151 (1991). https://doi.org/10.1109/CVPR.1991.139677
3. Chirikjian, G.S.: Partial Bi-Invariance of SE(3) Metrics 1. Journal of Computing and Information Science in Engineering **15**(1), 011008 (2014). https://doi.org/10.1115/1.4028941
4. Choi, S., Kim, J.H.: Fast and reliable minimal relative pose estimation under planar motion. Image and Vision Computing **69**, 103–112 (2018)
5. Cox, D., Little, J., O'Shea, D.: Ideals, Varieties, and Algorithms, vol. 3. Springer (2007)
6. Fathian, K., Jin, J., Wee, S.G., Lee, D.H., Kim, Y.G., Gans, N.R.: Camera relative pose estimation for visual servoing using quaternions. Robotics and Autonomous Systems **107**, 45–62 (2018). https://doi.org/10.1016/j.robot.2018.05.014
7. Faugère, J.C., Lachartre, S.: Parallel Gaussian Elimination for Gröbner bases computations in finite fields **68**(2), 5 (2007). https://doi.org/10.1145/1837210.1837225
8. Fraundorfer, F., Tanskanen, P., Pollefeys, M.: A minimal case solution to the calibrated relative pose problem for the case of two known orientation angles. In: European Conference on Computer Vision. pp. 269–282. Springer (2010)
9. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2012)
10. Hartley, R.: In defence of the 8-point algorithm. In: Computer Vision, 1995. Proceedings., Fifth International Conference on. pp. 1064–1070. IEEE (1995)
11. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2003)
12. Hartley, R.I., Li, H.: An Efficient Hidden Variable Approach to Minimal-Case Camera Motion Estimation. IEEE Trans. Pattern Anal. Mach. Intell. **34**(12), 2303–2314 (2012)
13. Kalantari, M., Hashemi, A., Jung, F., Guédon, J.P.: A new solution to the relative orientation problem using only 3 points and the vertical direction. Journal of Mathematical Imaging and Vision **39**(3), 259–268 (2011)
14. Kalantari, M., Jung, F., Guedon, J.P., Paparoditis, N.: The five points pose problem: A new and accurate solution adapted to any geometric configuration. In: Pacific-Rim Symposium on Image and Video Technology. pp. 215–226. Springer (2009)
15. Kneip, L., Siegwart, R., Pollefeys, M.: Finding the exact rotation between two images independently of the translation. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **7577 LNCS**(PART 6), 696–709 (2012)
16. Kukelova, Z., Bujnak, M., Pajdla, T.: Automatic generator of minimal problem solvers. In: European Conference on Computer Vision. pp. 302–315. Springer (2008)
17. Kukelova, Z., Bujnak, M., Pajdla, T.: Polynomial eigenvalue solutions to the 5-pt and 6-pt relative pose problems. In: British Machine Vision Conference. vol. 2 (2008)

18. Larsson, V., Åström, K.: Uncovering symmetries in polynomial systems. In: European Conference on Computer Vision. pp. 252–267. Springer (2016)
19. Larsson, V., Astrom, K., Oskarsson, M.: Efficient solvers for minimal problems by syzygy-based reduction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 820–829 (2017)
20. Lee, G., Pollefeys, M., Fraundorfer, F.: Relative pose estimation for a multi-camera system with known vertical direction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 540–547. IEEE (2014)
21. Lee, G.H., Faundorfer, F., Pollefeys, M.: Motion estimation for self-driving cars with a generalized camera. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2746–2753 (2013)
22. Lee, G.H., Fraundorfer, F., Pollefeys, M.: Structureless pose-graph loop-closure with a multi-camera system on a self-driving car. In: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 564–571. IEEE (2013)
23. Li, B., Heng, L., Lee, G., Pollefeys, M.: A 4-point algorithm for relative pose estimation of a calibrated camera with a known relative rotation angle. In: IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 1595–1601. IEEE (2013)
24. Li, B., Larsson, V.: Gaps: Generator for automatic polynomial solvers. arXiv preprint arXiv:2004.11765 (2020)
25. Longuet-Higgins, H.C.: A computer algorithm for reconstructing a scene from two projections. Nature **293**(5828), 133 (1981)
26. Maddern, W., Pascoe, G., Linegar, C., Newman, P.: 1 Year, 1000km: The Oxford RobotCar Dataset. The International Journal of Robotics Research (IJRR) **36**(1), 3–15 (2017). https://doi.org/10.1177/0278364916679498
27. Martyushev, E.: Self-calibration of cameras with Euclidean image plane in case of two views and known relative rotation angle. In: European Conference on Computer Vision. pp. 435–449. Springer (2018)
28. Martyushev, E., Li, B.: Efficient relative pose estimation for cameras and generalized cameras in case of known relative rotation angle (2020). https://doi.org/10.1007/s10851-020-00958-5
29. Maybank, S.: Theory of reconstruction from image motion, vol. 28. Springer Science & Business Media (2012)
30. Nistér, D.: An efficient solution to the five-point relative pose problem. IEEE Transactions on Pattern Analysis and Machine Intelligence **26**(6), 756–770 (2004)
31. Pandey, G., McBride, J.R., Eustice, R.M.: Ford campus vision and lidar data set. The International Journal of Robotics Research **30**(13), 1543–1552 (2011)
32. Pless, R.: Using many cameras as one. In: CVPR (2). pp. 587–593. IEEE Computer Society (2003)
33. Ramirez-paredes, J.P., Doucette, E.A., Curtis, J.W., Gans, N.R.: QuEst : A Quaternion-Based Approach for Camera. IEEE Robotics and Automation Letters **3**(2), 857–864 (2018)
34. Ruiz-Sarmiento, J.R., Galindo, C., González-Jiménez, J.: Robot@home, a robotic dataset for semantic mapping of home environments. International Journal of Robotics Research (2017)
35. Saurer, O., Vasseur, P., Boutteau, R., Demonceaux, C., Pollefeys, M., Fraundorfer, F.: Homography Based Egomotion Estimation with a Common Direction. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(2), 327–341 (2017). https://doi.org/10.1109/TPAMI.2016.2545663

36. Scaramuzza, D.: 1-point-ransac structure from motion for vehicle-mounted cameras by exploiting non-holonomic constraints. International journal of computer vision **95**(1), 74–85 (2011)
37. Scaramuzza, D., Fraundorfer, F.: Tutorial: Visual odometry. IEEE Robot Autom. Mag. **18**, 80–92 (2011)
38. Shiu, Y.C., Ahmad, S.: Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form ax= xb. IEEE Transactions on robotics and automation **5**(1), 16–29 (1989)
39. Stewénius, H., Engels, C., Nistér, D.: Recent developments on direct relative orientation. ISPRS Journal of Photogrammetry and Remote Sensing **60**(4), 284–294 (2006)
40. Stewénius, H., Nistér, D., Kahl, F., Schaffalitzky, F.: A minimal solution for relative pose with unknown focal length. Image and Vision Computing **26**(7), 871–877 (2008)
41. Stewénius, H., Nistér, D., Oskarsson, M., Åström, K.: Solutions to minimal generalized relative pose problems. In: In Workshop on Omnidirectional Vision (ICCV) (2005)
42. Sturm, J., Engelhard, N., Endres, F., Burgard, W., Cremers, D.: A benchmark for the evaluation of rgb-d slam systems. In: Proc. of the International Conference on Intelligent Robot Systems (IROS) (2012)
43. Tsai, R.Y., Lenz, R.K.: Real time versatile robotics hand/eye calibration using 3d machine vision. In: Proceedings. 1988 IEEE International Conference on Robotics and Automation. pp. 554–561. IEEE (1988)