Efficient DLT-Based Method for Solving PnP, PnPf, and PnPfr Problems

Gaku NAKANO*1, Member

SUMMARY This paper presents an efficient method for solving PnP, PnPf, and PnPfr problems, which are the problems of determining camera parameters from 2D–3D point correspondences. The proposed method is derived based on a simple usage of linear algebra, similarly to the classical DLT methods. Therefore, the new method is easier to understand, easier to implement, and several times faster than the state-of-the-art methods using Gröbner basis. Contrary to the existing Gröbner basis methods, the proposed method consists of three algorithms depending on the number of the points and the 3D point configuration. Experimental results show that the proposed method is as accurate as the state-of-the-art methods even in near-planar scenes while achieving up to three times faster.

Keyword: perspective-n-point (PnP) problem, absolute camera pose estimation, direct linear transform (DLT), focal length, radial lens distortion

1. Introduction

Estimating camera parameters from a single image has been a fundamental problem in computer vision applications, such as incremental structure from motion [1], geolocalization [2], camera tracking [3], etc. The basic problem, called Perspective-n-Point (PnP) problem [4], [5], is to find extrinsic parameters of a calibrated camera from \( n \geq 3 \) pairs of a 2D–3D point correspondence, i.e. rotation and translation.

Extensive studies on PnP problem [6]–[11] have revealed that methods minimizing algebraic error can efficiently achieve comparable accuracy to those minimizing reprojection error. Moreover, Gröbner basis approaches [12]–[14] are useful to derive globally optimal methods that are generalized to handle both planar and non-planar 3D points without explicitly differentiating the point configuration. As a result of the success of the PnP solvers, research interests have been extended to more complicated and practical cases where some intrinsic parameters are unknown: PnPf problem for unknown focal length, PnPfr problem for both unknown focal length and unknown lens distortion.

Minimal solvers, which find both extrinsic and intrinsic parameters from the minimum number of the points, were proposed for P4P problem [15]–[18] and P4Pfr or P5Pfr problems [19]–[22]. These methods are mainly aimed to be used for outlier rejection by incorporating into a RANSAC scheme [23]. For least squares case, reprojection error minimization is typically applied to refine the accuracy of a solution obtained by a minimal solver in spite of a disadvantage on computational cost.

To obtain a least squares solution in an efficient way, similarly to the PnP solvers, some methods for PnPfr [24]–[27] and PnPfr [28] problems have been developed that minimize algebraic error. Among them, Nakano [28] proposed the first least squares method for PnPfr problem with \( n \geq 5 \) points, called VPnPfr. One of the contributions of the VPnPfr method is that it derives a common way to solve the aforementioned three problems, i.e. PnP, PnPf, and PnPfr problems.

The VPnPfr method divides an algebraic cost function into two approximate subproblems: the first subproblem is solved by a Gröbner basis approach which is common for the three problems; the second subproblem is solved by a linear method which slightly differs depending on each problem. The VPnPfr method can handle both planar and non-planar scenes by a single solver due to the Gröbner basis approach; however, the Gröbner basis approach is not computationally efficient for calculating a Gauss-Jordan elimination on a 97 \( \times \) 117 matrix and an eigenvalue decomposition on a 20 \( \times \) 20 matrix.

This paper proposes a new linear-based method to improve the efficiency of the VPnPfr method. Given \( n \geq 6 \) points, we show that the first subproblem of VPnPfr method is decomposed into three linear equations that the direct linear transform (DLT) [29] can be applied. Since the whole procedure of the VPnPfr method can be solved by simple linear computations without a Gröbner basis approach, the proposed method is easier to understand, easier to implement, and more efficient than the VPnPfr method. Similarly to the VPnPfr method, the proposed approach is able to apply PnP and PnPf problems as well. We demonstrate by experiments that our method, named EDPnPfr (Efficient DLT-based method for PnPfr problem), is faster than the VPnPfr method by up to three times while performing comparable accuracy and robustness with the state-of-the-art methods.

This paper is organized as follows. In Sect. 2, we give a mathematical formulation of PnPfr problem and briefly revisit the VPnPfr method. Section 3 describes the details of the proposed method. We report experimental results on both synthetic and real data in Sect. 4. According to the experimental results, future work is discussed in Sect. 5. Finally, we will give concluding remarks in Sect. 7.
2. Preliminaries

This section first describes mathematical formulations of PnPfr problem, then, reviews the state-of-the-art algorithms, called the VPnPfr method [28]. Since PnP and PnPf problems are subcategories of PnPfr problem, which can be derived by setting some particular parameters known, we will mention only PnPfr problem in this section.

2.1 PnPfr Problem

We assume that intrinsic parameters of a pinhole camera are known except for focal length \( f \) and radial lens distortion \( k \). The radial distortion is modeled by the three-parameter division model [30]. Let \( R \) and \( t = [r_x, r_y, r_z]^T \) be a \( 3 \times 3 \) rotation matrix and a \( 3 \times 1 \) translation vector, respectively. We can formulate a projective equation between an \( i \)-th 3D point \( p_i = [x_i, y_i, z_i]^T \) and its 2D projection \( m_i = [u_i, v_i, w_i]^T \) in the homogeneous coordinate by

\[
m_i \propto K (R p_i + t),
\]

where \( \propto \) represents equality up to scale and \( K = \text{diag}(1, 1, 1^T) \). The homogeneous term \( w_i \) including radial distortion can be written as

\[
w_i = 1 + k_i^d d_i,
\]

where \( d_i = [u_i^2 + v_i^2, (u_i^2 + v_i^2)^2, (u_i^2 + v_i^2)^3]^T \).

The cost function of Eq. (3) is called algebraic error. Although algebraic error minimization is not geometrically meaningful, it results in comparable accuracy with reprojection error minimization for PnP and PnPf problems [9], [27].

2.2 Existing Solvers for PnPfr Problem

Given \( n = 4 \) or \( n = 5 \) points, Eq. (3) becomes P4Pfr or P5Pfr problem, respectively, which are special and the minimal cases of PnPfr problem. Josephson and Byröd [19] proposed the first solution to P4Pfr problem that is a general single solver with one radial distortion parameter for handling both planar and non-planar scenes. Since their solver is not efficient enough for real-time processing due to the generalization by a Gröbner basis method, Bujnak et al. [20] provided a faster P4Pfr solver by developing two algorithms for planar and non-planar scenes separately. By introducing new constraint for the projection matrix, Larsson et al. [22] extended Bujnak et al.’s method to a single generalized solver. To deal with up to the three radial distortion parameters, Kukelova et al. [21] proposed a fast and general P5Pfr solver based on a simple usage of linear algebra.

For the least squares case, \( n \geq 5 \) points, Nakano [28] proposed the first solution to PnPfr problem, called VPnPfr, by extending the Kukelova et al.’s solver. The VPnPfr method consists of a three-step approach, which enables to solve PnP and PnPf problems as well as PnPfr problem in the same manner. Since the proposed method in this paper is an alternative for solving the first step among the three steps, we will briefly review the VPnPfr method in the next section. Table 1 summarizes a comparison of the existing and the proposed PnPfr solvers.

2.3 VPnPfr Method

The VPnPfr method gives a solution through three steps. The first and the second steps sequentially solve two subproblems, which are approximations of Eq. (3). Finally, an approximate solution from the second step is refined by solving the original problem. The main contribution of VPnPfr is to derive the two subproblems: the first subproblem expressed by a part of the extrinsic parameters, which are common unknowns for PnP, PnPf, and PnPfr problems; the
second subproblem composed of remaining parameters depend on each problem.

Let us move on the derivation of VPnPfr. Define \(a_i^T\), \(b_i^T\), and \(c_i^T\) by the first, the second, and the third row of \([m_i]_x\) in Eq. (4), respectively. Thus, PnPfr problem, Eq. (3), can be equivalently rewritten by

\[
\min_{\mathbf{r}_i, \mathbf{f}_i, k} \sum_{i=1}^{n} (a_i^T \mathbf{q}_i)^2 + (b_i^T \mathbf{q}_i)^2 + (c_i^T \mathbf{q}_i)^2 \tag{5}
\]

\[
\text{s.t. } \mathbf{R}^T \mathbf{R} = \mathbf{I}, \text{ } \det(\mathbf{R}) = 1,
\]

where

\[
\mathbf{R} = \begin{bmatrix} r_1 & r_2 & r_3 \end{bmatrix}^T,
\]

\[
\mathbf{q}_i = \mathbf{K} (\mathbf{R} \mathbf{p}_i + \mathbf{t}),
\]

\[
a_i^T \mathbf{q}_i = -w_i (\mathbf{r}_i \mathbf{p}_i + \mathbf{t}_i) + v_i f_i^{-1} (\mathbf{r}_i \mathbf{p}_i + \mathbf{t}_i),
\]

\[
b_i^T \mathbf{q}_i = w_i (\mathbf{r}_i \mathbf{p}_i + \mathbf{t}_i) - u_i f_i^{-1} (\mathbf{r}_i \mathbf{p}_i + \mathbf{t}_i),
\]

\[
c_i^T \mathbf{q}_i = -w_i (\mathbf{r}_i \mathbf{p}_i + \mathbf{t}_i) + u_i (\mathbf{r}_i \mathbf{p}_i + \mathbf{t}_i),
\]

Considering that Eq. (10) contains only \(r_1, r_2, r_3\), and \(t_y\), we can derive the first step by

\[
\min_{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{t}_y} \sum_{i=1}^{n} (c_i^T \mathbf{q}_i)^2 \tag{11}
\]

\[
\text{s.t. } ||\mathbf{r}_1||^2 = 1, \mathbf{r}_1^T \mathbf{r}_2 = 0, ||\mathbf{r}_1||^2 - ||\mathbf{r}_2||^2 = 0.
\]

Since Eq. (11) does not have any constraints about the translation, we can equivalently rewrite Eq. (11) as a constrained optimization of the rotation matrix where \(r_x\) and \(r_y\) are represented by a function of the rotation:

\[
\min_{\mathbf{r}_1, \mathbf{r}_2} \left[ \begin{array}{c} r_1^T \mathbf{r}_2 \\ r_1^T \mathbf{r}_2 \end{array} \right] \mathbf{M} \left[ \begin{array}{c} r_1 \\ r_2 \end{array} \right] \tag{12}
\]

\[
\text{s.t. } ||\mathbf{r}_1||^2 = 1, ||\mathbf{r}_1||^2 - ||\mathbf{r}_2||^2 = 0, \mathbf{r}_1^T \mathbf{r}_2 = 0,
\]

where

\[
\mathbf{M} = \mathbf{A}^T \mathbf{A} - \mathbf{A}^T \mathbf{B} (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{A},
\]

\[
\mathbf{A} = \begin{bmatrix} -v_1 \mathbf{P}_1^T & u_1 \mathbf{P}_1^T \\ \vdots & \vdots \\ -v_n \mathbf{P}_n^T & u_n \mathbf{P}_n^T \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} -v_1 & u_1 \\ \vdots & \vdots \\ -v_n & u_n \end{bmatrix}.
\]

The translation components are given by

\[
\begin{bmatrix} \mathbf{t}_x \\ \mathbf{t}_y \end{bmatrix} = -(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{A} \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{bmatrix}.
\]

There are no intrinsic parameters in Eq. (12). Hence, the first subproblem can be a common step to be solved for PnP, PnPf, and PnPfr problems regardless of the differences of unknown intrinsic parameters. Nakano [28] revealed that Eq. (12) has up to 20 real stationary points without sign ambiguity and developed a Gröbner basis method to retrieve the global optimum among all stationary points.

To obtain remaining parameters, \(r_x\), \(f\), and \(k\), the second subproblem is constructed by using Eqs. (8) and (9):

\[
\min_{r_x, f, k} \sum_{i=1}^{n} (a_i^T q_i)^2 + (b_i^T q_i)^2 = ||\mathbf{x} + \mathbf{g}\|^2, \tag{16}
\]

where

\[
\mathbf{L} = \begin{bmatrix} v_1 & v_1 \mathbf{z}_1^T & -y_1 \mathbf{d}_1^T \\ -u_1 & -u_1 \mathbf{z}_1^T & x_1^T \mathbf{d}_1^T \\ \vdots & \vdots & \vdots \\ v_n & v_n \mathbf{z}_n^T & -y_n \mathbf{d}_n^T \\ -u_n & -u_n \mathbf{z}_n^T & x_n^T \mathbf{d}_n^T \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} f^{-1} t_x \\ \vdots \\ k \end{bmatrix}, \quad \mathbf{g} = \begin{bmatrix} -y_i^T \mathbf{x}_i^T \\ \vdots \\ -y_n^T \mathbf{x}_n^T \end{bmatrix}, \quad \mathbf{x}_i^T = r_i^T \mathbf{p}_i + t_x, \quad y_i^T = r_i^T \mathbf{p}_i + t_y, \quad \mathbf{z}_i^T = r_i^T \mathbf{p}_i.
\]

As Eqs. (16) and (17) show, the second subproblem is a linear form of the unknowns, which can be simply solved by computing \(\mathbf{x} = -(\mathbf{L}^T \mathbf{L})^{-1} \mathbf{L}^T \mathbf{g}\).

Since Eqs. (12) and (16) are approximations of Eq. (5), the solution obtained by the two subproblems is obviously not a local minimum of the original PnPfr problem. To refine the accuracy of the parameters, a simple Gauss-Newton method is carried out on Eq. (5) at the third step with the parameters estimated at the previous steps as an initial guess.

The three-step approach does not guarantee to converge to the global optimum. However, it has been experimentally validated that the above procedure gives comparable performance to the state-of-the-art methods that have the global optimality [28].

3. Proposed Method

This section describes an efficient linear method for solving the first step of VPnPfr in the case of \(n \geq 6\) points. The difference against VPnPfr is that the proposed method consists of three algorithms: two for non-planar scenes and one for planar scenes. The 3D point configuration can be categorized by checking the rank of a covariance matrix of the 3D points [31]. The second and the third steps are exactly the same as in VPnPfr. Therefore, the proposed method can also be applicable to PnP and PnPf problems. Algorithm 1 shows a pseudo code of the proposed method with MATLAB like expressions for matrix indexing and operators.

3.1 Overview

The basic strategy of the proposed method is to find a solution of Eq. (10) based on a DLT-like method depending on the 3D point configuration. DLT-like means, as the name implies, that it first solves an optimization problem by a linear least squares without considering non-linear constraints, next, corrects the solution to satisfy the constraints.

If we assume the unit-vector constraint instead of the rotation matrix constraints, Eq. (12) becomes a simpler form

\[
\text{rank}(\sum_{i=1}^{n} (\mathbf{p}_i - \overline{\mathbf{p}}) (\mathbf{p}_i - \overline{\mathbf{p}})^T) = 3: \text{ non-plane, } 3 \text{ with a large condition number: near-plane, } 2: \text{ plane, where } \overline{\mathbf{p}} = 1/n \sum_{i=1}^{n} \mathbf{p}_i \text{ is the center of mass of the 3D points.}
\]
obtained by the singular value decomposition (SVD) of 1470
also the solution because the unit-vector constraint has a
not true rotation matrices. To obtain the nearest rotation ma-
U
end
else
12: else % planar scenes (z = 0)
13: \( \hat{M} = M([1 : 2, 4 : 5], [1 : 2, 4 : 5]) \)
14: \([\sim, \sim, V] = \text{svd}(\hat{M}) \)
15: \([\hat{F}_1; \hat{F}_2] = V(:, 4) \)
16: \( r^\alpha_{13} = \text{solveEq22or23}(\hat{v}_1, \hat{v}_2) \)
17: \( r_{13} = \pm \sqrt{r^\alpha_{13}^2 + r_{23}^2} \)
18: \( [r_1; r_2] = [\hat{r}_1; \hat{r}_2; r_{23}] \)
end
20: \( R = \text{recoverRotationEqs19and20}(r_1, r_2) \)
21: \([t_1; t_3] = -B[A [r_1; r_2] \]
22: \([L, g] = \text{calcEq17}(Y, R, t_1, t_3) \)
23: \( x = -L \backslash g \)
24: \( f = 1 / x(2), \quad k = x(3 : 5) \)
25: \([R, t, f, k] = \text{GaussNewtonEq5}(X, Y, R, t, f, k) \)
26: end
end
\end{algorithm}

\begin{align}
\min_{r_1, r_2} \begin{bmatrix} r_1^T, r_2^T \end{bmatrix} M \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \\
\text{s.t.} \quad ||r_1||^2 + ||r_2||^2 = 1.
\end{align}

Once we found a solution of Eq. (18), \( -r_1 \) and \( -r_2 \) are also the solution because the unit-vector constraint has a sign ambiguity. Thus, we can recover two rotation matrices by

\begin{align}
R = \begin{bmatrix} r_1^T \\ r_2^T \\ (r_1 \times r_2)^T \end{bmatrix}, \\
\quad R = \begin{bmatrix} -r_1^T \\ -r_2^T \\ -(r_1 \times r_2)^T \end{bmatrix}.
\end{align}

We ignored the rotation matrix constraints when solving Eq. (18), therefore, the above two matrices are actually not true rotation matrices. To obtain the nearest rotation matrix in the sense of Frobenius norm, a well known constraint enforcement [32] can be applied by

\begin{align}
R = U \diag(1, 1, \text{det}(UV^T)) V,
\end{align}

where \( U \) and \( V \) are the left and the right orthonormal matrices obtained by the singular value decomposition (SVD) of \( R \) in Eq. (19), respectively.

In the following section, we will provide the details of the proposed method for solving Eq. (18) depending on the 3D point configuration.

3.2 Non-Planar Scenes

3.2.1 \( n = 6 \) Points

The \( 6 \times 6 \) matrix \( M \) is of rank four in the six non-planar point case for both noiseless and noisy data. Because of the rank deficiency, the unknown vector \( [r_1^T, r_2^T] \) can be expressed as a linear combination of two nullspace vectors \( v_1 \) and \( v_2 \) of \( M \), i.e.

\begin{align}
[r_1] = \alpha v_1 + v_1, \\
[r_2] = \alpha v_2 + v_2,
\end{align}

where \( \alpha \) is a new unknown. In this case, the cost function of Eq. (18) is always zero regardless of the value of \( \alpha \).

To determine \( \alpha \), we can use one of the orthonormal constrains that \( r_1 \) and \( r_2 \) are perpendicular and have the same norm. Those constraints are represented by

\begin{align}
(\alpha v_1 + v_1)^T(\alpha v_2 + v_2) = 0, \\
||\alpha v_1 + v_1||^2 + ||\alpha v_2 + v_2||^2 = 0.
\end{align}

Both two equations result in a quadratic polynomial equation in \( \alpha \), which can be solved in a closed form. For example, Eq. (22) leads to

\begin{align}
(v_1^T v_1)\alpha^2 + 2(v_1^T v_1 + v_1^T v_2)\alpha + v_1^T v_2 = 0.
\end{align}

Thus, we can find up to two real solutions of \( \alpha \) by solving either of Eq. (22) or Eq. (23).

3.2.2 \( n \geq 7 \) Points

When we have more than or equal to seven non-planar points, \( M \) is of rank five and six for noiseless and noisy data, respectively. In this case, as is well known, the unique solution of Eq. (18) is given by the right nullspace vector corresponding to the smallest singular value. This is the same procedure typically used by DLT methods in computer vision [29].

3.3 Planar Scenes

Without loss of generality, we can assume that all 3D points lie on a plane \( z = 0 \) for planar scenes. Since the third and the sixth columns of \( A \) are zeros, we let \( \hat{M} \) be a \( 4 \times 4 \) matrix excluding the third and the sixth columns and rows of \( M \), respectively. Thus, we can rewrite Eq. (18) for planar scenes by

\begin{align}
\min_{r_1, r_2} \begin{bmatrix} r_1^T, r_2^T \end{bmatrix} M \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \\
\text{s.t.} \quad ||r_1||^2 + ||r_2||^2 = 1.
\end{align}

\begin{align}
\min_{\hat{r}_1, \hat{r}_2} \begin{bmatrix} \hat{r}_1^T, \hat{r}_2^T \end{bmatrix} M \begin{bmatrix} \hat{r}_1 \\ \hat{r}_2 \end{bmatrix} \\
\text{s.t.} \quad ||\hat{r}_1||^2 + ||\hat{r}_2||^2 = 1.
\end{align}

Here, \( \hat{r}_1 = [r_{11}, r_{12}]^T \), \( \hat{r}_2 = [r_{21}, r_{22}]^T \), and \( r_{ij} \) denotes the \((i, j)\) element of \( R \). Similarly to the non-planar case with \( n \geq 7 \) points, Eq. (25) is a linear least squares problem.
NAKANO: DLT-BASED METHOD FOR PNP, PNPF, AND PNPFR PROBLEMS

Therefore, \( \hat{r}_1 \) and \( \hat{r}_2 \) can be obtained by SVD of \( \hat{M} \).

After having obtained \( \hat{r}_1 \) and \( \hat{r}_2 \), we use the orthonormal constraints between \( r_1 \) and \( r_2 \) to find the solution of two scalar unknowns \( r_{13} \) and \( r_{23} \). These constraints lead to two equations,

\[
\hat{r}_1^T \hat{r}_2 + r_{13} r_{23} = 0, \tag{26}
\]

\[
\hat{r}_1^T \hat{r}_1 + r_{13}^2 - (\hat{r}_1^T \hat{r}_2 + r_{23}^2) = 0. \tag{27}
\]

Eliminating one of the unknowns from the above equations, we obtain a fourth degree polynomial equation in the other unknown. For example, if we select \( r_{23} \) to be eliminated, the quartic equation has the form

\[
r_{13}^4 + (\hat{r}_1^T \hat{r}_1 - \hat{r}_2^T \hat{r}_2) r_{13}^2 - \hat{r}_1^T \hat{r}_2 = 0. \tag{28}
\]

Regarding Eq. (28) as a quadratic polynomial in \( r_{13}^2 \), we can obtain up to four positive real solutions. Substituting \( \pm \sqrt{r_{13}^2} \) into Eq. (26) gives the corresponding \( r_{23} \) by

\[
r_{23} = -\frac{\hat{r}_1^T \hat{r}_2}{r_{13}}. \tag{29}
\]

3.4 Degeneracy

Degenerate configurations for PnP and PNPF problems have been widely studied in the literature [4], [16], e.g. the danger cylinder\(^{\dagger} \) in P3P problem, front-parallel planar scenes\(^{\ddagger} \) in PNPF problem. In this section, we will investigate degenerate configurations that are inherent to the proposed method.

There exists critical point configurations that lead to a degeneracy in the proposed method: \( \text{rank}(M) < 5 \) for the first subproblem and \( \text{rank}(L^T L) < 5 \) for the second subproblem.

A simple situation where \( \text{rank}(M) < 5 \) occurs is that 3D points are collinear, i.e. points on a single line. In this case, we cannot calculate \( M \) because \( A^T A \) and \( B^T B \) are not invertible. A 3D line gives two constraints, therefore, we need to observe more than three distinct lines to obtain a valid \( M \). Note that \( B^T B \) is rank-deficient in PNPF problem if non-collinear 3D points are projected as 2D points on a line.

\(^{\dagger}\) A camera is positioned on a circular cylinder passing through the three points.

\(^{\ddagger}\) The z-axis of a camera and the normal of a 3D plane are parallel to each other.
due to lens distortion. A special case of lines is a degenerate configuration for \( \text{rank}(L^TL) < 5 \). If 2D points or 3D points are on a line \( v_i = -u_i \) or \( y_i = -x_i \), any additional points do not constraint \( L^TL \) to be of full rank. The collinearity can be detected by validating the rank of a covariance matrix.

Another condition is that 3D points can be expressed by a linear combination of \( r_1 \) and \( r_2 \). In this case, \( \text{rank}(L^TL) = 4 \) because the second column of \( L \) becomes a zero vector. Since we have a rotation matrix before the second subproblem, this condition is also avoidable by removing such 3D points.

4. Experiments

We evaluated the performance of the proposed method on both synthetic and real data by comparing to the existing methods. On the synthetic data evaluation, we investigated parameter estimation accuracy in PnP, PnPf, and PnPfr problems with varying the number of the points and Gaussian image noise. Moreover, we measured computational time of all methods on synthetic data. For the real image test, we visually checked the quality of undistorted images because the groundtruth data is not measurable in practical situations.

We call the proposed three Efficient DLT-based solvers as EDPnP, EDPnPf, and EDPnPfr for PnP, PnPf, and PnPfr problems, respectively. In each problem, we compared the following methods:

**PnP problem (six methods)**
- DLT [29], EPnP+GN [33], OPnP [9], UPnP [10], VPnP [28], EDPnP

**PnPf problem (five methods)**
- DLT [29], GPnPf+GN [25], EPnPfR [26], VPnPf [28], EDPnPf

**PnPfr problem (three methods)**
- P5Pfr+NL [21] (P5Pfr with RANSAC and reprojection error minimization), VPnPfr [28], EDPnPfr

We implemented the proposed solvers on MATLAB. For the existing methods, we used the original MATLAB or C++ code publicly available on the web except for P5Pfr and the VPnP variants, which were implemented by the author of this paper. All evaluations were conducted on Core i7-6700 and MATLAB.

### 4.1 Synthetic Data

3D points were randomly generated in the x-, y-, and z-range of \([-2, 2] \times [-2, 2] \times [4, 8] \) in the camera coordinate...
Fig. 3 Median error of the extrinsic parameters w.r.t. varying Gaussian image noise (1 ≤ σ ≤ 5) with fixed number of points (n = 20). The x-axis is the image noise in pixels, and the y-axis shows the absolute error in degrees for rotation and the relative error in percentage for the others. 500 independent trials were conducted for each σ.

for non-planar scenes, [−2, 2] × [1, 2] × [4, 8] in the camera coordinate for near-planar scenes, and [−2, 2] × [−2, 2] × [0, 0] in the world coordinate for planar scenes. For planar scenes, the depth of 3D points were configured within the range of [4, 8] in the camera coordinates similarly to the non-planar and the near-planar scenes. These 3D points were projected onto a camera having focal length f = 500, zero skew and the principal point at the image center. Radial distortion was set to be zero $k = [0, 0, 0]^T$ in PnP and PnPfr problems and $k = [-3/f^2, -0.5/f^4, -0.05/f^6]^T$ in PnPfr problem. The rotation and translation of the camera were randomly determined on each trial. Parameter accuracy against the groundtruth was measured by the absolute error in degrees for the rotation and the relative error in percentage for the others.

### 4.1.1 Accuracy w.r.t. Varying Number of Points

In the first experiment, we have studied the accuracy with respect to varying number of the points under a fixed image noise. We varied the number of the points from 6 to 100 (6 ≤ n ≤ 100) and fixed Gaussian image noise to be zero mean and standard deviation $σ = 2$ pixels. Then, we conducted 500 independent trials for each n in three configurations of the 3D points, i.e. non-planar, near-planar, and planar scenes, on three pose estimation problems. Figures 1 and 2 show the median error$^\dagger$ of parameter estimation accuracy for the extrinsic parameters and the intrinsic parameters, respectively.

We can observe that the proposed solvers achieve the comparable accuracy to the state-of-the-art methods and work well even in near-planar scenes. The VPnPfr paper [28] has experimentally showed that the three-step approach gives globally optimal solutions, which do not explicitly differentiate the 3D point configuration between planar and non-planar scenes. The results in this section verify that the proposed approach, switching several algorithms depending on the scene, is as effective as VPnPfr in spite of its simpleness.

In PnPfr problem, EDPnPfr and VPnPfr have similar performance to P5Pfr+NL for rotation but not for the other

$^\dagger$Since mean error curves by some methods are not smooth but zig-zag along the x-axis due to unstable estimations, we used the median error for readability.
parameters. The lens distortion errors seem to be significantly larger than the estimation error of the other parameters; however, this is because the absolute value of the groundtruth is very small, e.g. $|k_2| < 10^{-11}$. A notable result here is a performance gap between EDPnPfr, VPnPfr, and P5Pfr+NL. While the estimation error against $\sigma=2$ image noise by EDPnPfr and VPnPfr reaches the lower bound at $n=100$ points, P5Pfr+NL still shows the possibility of reducing the estimation error by using $n \geq 100$ points. This is caused by the difference of the cost function to be minimized: algebraic error in EDPnPfr and VPnPfr, reprojection error in P5Pfr+NL. These observations may suggest the limitation of methods minimizing algebraic error in PnPfr problem.

4.1.2 Accuracy w.r.t. Varying Image Noise

In the next experiment, we have studied the accuracy with respect to varying image noise $1 \leq \sigma \leq 5$ under fixed $n=20$ points. We measured the median error of 500 independent trials for each $\sigma$. The results of the extrinsic and the intrinsic parameters are shown in Figs. 3 and 4, respectively.

In PnP and PnPfr problems, similarly to the previous experiment, the proposed EDPnP and EDPnPfr have compara-
Table 2  Results on real image data. The original image has 1024 × 1024 resolution and was captured by a camera with approximately 118° HFOV. The parameters were estimated by detecting chessboard corners. All methods estimate almost identical values and successfully provide undistorted images where curved lines are straightened.

| Parameter               | Spec sheet | P5Pfr+NL [21] | VPnPfr [28] | EDPnPfr (proposed) |
|-------------------------|------------|---------------|-------------|--------------------|
| HFOV [degrees]          | 118        | 111.58        | 111.60      | 111.60             |
| Focal length $f$ [pixels]| 307.64     | 348.11        | 347.90      | 347.90             |
| Radial distortion $k_1$ | n.a.       | $-3.43 \times 10^{-6}$ | $-3.41 \times 10^{-6}$ | $-3.41 \times 10^{-6}$ |
| Radial distortion $k_2$ | n.a.       | $-2.05 \times 10^{-12}$ | $-2.57 \times 10^{-12}$ | $-2.57 \times 10^{-12}$ |
| Radial distortion $k_3$ | n.a.       | $-2.57 \times 10^{-18}$ | $5.01 \times 10^{-19}$ | $5.01 \times 10^{-19}$ |

Table 2 shows estimated values of the intrinsic parameters and qualitative results, i.e., the original and undistorted images rectified by using the estimated distortion coefficients. The estimated values are almost identical each other and close to those calculated by the specifications. Although $k_3$ varies a lot in each method, the differences do not affect the visual quality where curved lines are successfully corrected to straight lines. From these observations, we can consider that these results are reasonable and proper.

5. Discussion

The experimental results in Sects. 4.1 and 4.2 show that the proposed method is more efficient than P5Pfr+NL but less accurate for PnPfr problem. Considering the trade-off, we can conclude that the proposed method is suitable for real-time applications, e.g., visual SLAM and augmented reality, which do not require rigorous accuracy at every input image. In many of those applications, 2D points are localized within a few pixel errors, which corresponds to $\sigma < 2$, and reprojection error minimization can be conducted at regular intervals to correct drift errors. Therefore, the proposed method can be used as a good estimator for the initial guess to obtain the global optimum. Moreover, since EDPnPfr is the only fast method for PnPfr problem, it can be used as a local optimizer for LO-RANSAC [34], which performs a least-squares estimation on tentative inlier points to achieve fast and stable convergence.
6. Future Work

One way to improve the estimation accuracy of algebraic methods against image noise is to apply non-minimization approaches [35] such as the Taubin method and hyper least squares. Those methods are derived by a statistical error analysis of algebraic equations in high order terms and result in a closed-form solution, e.g. the generalized eigenvalue problem. By introducing this technique, we can possibly obtain more accurate methods for finding rotation (Eqs. (18) and (25)) and the other parameters (Eq. (17)) in the first and the second subproblems, respectively. Particularly, a non-minimization method is expected to work for removing a statistical bias of radial lens distortion \( k = [k_1, k_2, k_3]^T \), of which coefficient \( d_i = [u_i^2 + v_i^2, (u_i^2 + v_i^2)^2, (u_i^2 + v_i^2)^3]^T \) has a large error up to six degrees. Unbiased solutions to a camera pose estimation problem with lens distortion have not been reported yet, thus, this is an interesting topic for future work.

7. Conclusions

In this paper, we have proposed a new efficient method for solving PnP, PnPf, and PnPfr problems. The proposed method is similar to the classical DLT method, therefore, the derivation and implementation are simpler and easier than the existing methods employing a Gröbner basis approach. We also discussed degenerate point configurations of the proposed method as well as workaround to avoid them. We have demonstrated by both synthetic and real data experiments that the proposed method is faster by more than two times while maintaining comparable accuracy to the state-of-the-art methods. In future work, we will work for a potential limitation of the algebraic error minimization, as shown in the synthetic data evaluation, that the intrinsic parameter estimation is sensitive to a large image noise in PnPfr problem.

References

[1] N. Snavely, S.M. Seitz, and R. Szeliski, “Photo tourism: Exploring photo collections in 3D,” ACM Trans. Graph. (TOG), vol.25, no.3, pp.835–846, 2006.
[2] Y. Li, N. Snavely, D.P. Huttenlocher, and P. Fua, “Worldwide pose estimation using 3D point clouds,” in Large-Scale Visual Geo-Localization, pp.147–163, Springer, 2016.
[3] G. Klein and D. Murray, “Parallel tracking and mapping for small AR workspaces,” 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR 2007, pp.225–234, 2007.
[4] B.M. Haralick, C.-N. Lee, K. Ottenberg, and M. Nölle, “Review and analysis of solutions of the three point perspective pose estimation problem,” International Journal of Computer Vision, vol.13, no.3, pp.331–356, 1994.
[5] Y. Wu and Z. Hu, “PnP problem revisited,” Journal of Mathematical Imaging and Vision, vol.24, no.1, pp.131–141, 2006.
[6] G. Schweighofer and A. Pinz, “Globally optimal O(n) solution to the PnP problem for general camera models,” BMVC, pp.1–10, 2008.
[7] H. Hmam and J. Kim, “Optimal non-iterative pose estimation via convex relaxation,” Image and Vision Computing, vol.28, no.11, pp.1515–1523, 2010.
[8] Y. Zheng, S. Sugimoto, and M. Okutomi, “ASPnP: An accurate and scalable solution to the perspective-n-point problem,” IEICE Trans. Inf. Syst., vol.E96-D, no.7, pp.1525–1535, July 2013.
[9] Y. Zheng, Y. Kuang, S. Sugimoto, K. Åström, and M. Okutomi, “Revisiting the PnP problem: A fast, general and optimal solution,” 2013 IEEE International Conference on Computer Vision (ICCV), pp.2344–2351, 2013.
[10] L. Kneip, H. Li, and Y. See, “UPnP: An optimal O(n) solution to the absolute pose problem with universal applicability,” in Computer Vision, ECCV 2014, pp.127–142, Springer, 2014.
[11] G. Nakano, “Globally optimal DLS method for PnP problem with Cayley parameterization,” Proc. British Machine Vision Conference (BMVC), pp.78.1–78.11, BMVA Press, Sept. 2015.
[12] D.A. Cox, J. Little, and D. O’shea, Using algebraic geometry, Graduate Texts in Mathematics, Springer, 2005.
[13] Z. Kukelova, M. Bujnak, and T. Pajdla, “Automatic generator of minimal problem solvers,” in Computer Vision, ACCV 2010, pp.302–315, Springer, 2008.
[14] V. Larsson, K. Åström, and M. Oskarsson, “Efficient solvers for minimal problems by syzygy-based reduction,” Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.820–829, 2017.
[15] M.A. Abidi and T. Chandra, “A new efficient and direct solution for pose estimation using quadrangular targets: Algorithm and evaluation,” IEEE Trans. Pattern Anal. Mach. Intell., vol.17, no.5, pp.534–538, 1995.
[16] B. Triggs, “Camera pose and calibration from 4 or 5 known 3D points,” Proc. Seventh IEEE International Conference on Computer Vision, pp.278–284, 1999.
[17] C. Wu, “P3P: Pose estimation with unknown focal length,” Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.2440–2448, 2015.
[18] M. Bujnak, Z. Kukelova, and T. Pajdla, “A general solution to the P4P problem for camera with unknown focal length,” IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, pp.1–8, 2008.
[19] K. Josephson and M. Byröd, “Pose estimation with radial distortion and unknown focal length,” IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp.2419–2426, 2009.
[20] M. Bujnak, Z. Kukelova, and T. Pajdla, “New efficient solution to the absolute pose problem for camera with unknown focal length and radial distortion,” in Computer Vision, ACCV 2010, pp.11–24, Springer, 2011.
[21] Z. Kukelova, M. Bujnak, and T. Pajdla, “Real-time solution to the absolute pose problem with unknown angular distortion and focal length,” Proc. IEEE International Conference on Computer Vision, pp.2816–2823, 2013.
[22] V. Larsson, Z. Kukelova, and Y. Zheng, “Making minimal solvers for absolute pose estimation compact and robust,” Proc. IEEE International Conference on Computer Vision, pp.2335–2343, 2017.
[23] M.A. Fischler and R.C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” Commun. ACM, vol.24, no.6, pp.381–395, 1981.
[24] A. Penate-Sanchez, J. Andrade-Cetto, and F. Moreno-Noguer, “Exhaustive linearization for robust camera pose and focal length estimation,” IEEE Trans. Pattern Anal. Mach. Intell., vol.35, no.10, pp.2387–2400, 2013.
[25] Y. Zheng, S. Sugimoto, L. Sato, and M. Okutomi, “A general and simple method for camera pose and focal length determination,” Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.430–437, 2014.
[26] E. Kanaeva, L. Gurevich, and A. Vakhitov, “Camera pose and focal length estimation using regularized distance constraints,” Proc. British Machine Vision Conference (BMVC), pp.162.1–162.12, BMVA Press, Sept. 2015.
[27] Y. Zheng and L. Kneip, “A direct least-squares solution to the PnP
problem with unknown focal length,” Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.1790–1798, 2016.

[28] G. Nakano, “A versatile approach for solving PnP, PnPf, and PnPfr problems,” European Conference on Computer Vision, pp.338–352, Springer International Publishing, 2016.

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

[30] A.W. Fitzgibbon, “Simultaneous linear estimation of multiple view geometry and lens distortion,” Proc. 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, pp.125–132, 2001.

[31] S. Li, C. Xu, and M. Xie, “A robust $O(n)$ solution to the perspective-n-point problem,” IEEE Trans. Pattern Anal. Mach. Intell., vol.34, no.7, pp.1444–1450, 2012.

[32] S. Umeyama, “Least-squares estimation of transformation parameters between two point patterns,” IEEE Trans. Pattern Anal. Mach. Intell., vol.13, no.4, pp.376–380, 1991.

[33] V. Lepetit, F. Moreno-Noguer, and P. Fua, “EPnP: An accurate $O(n)$ solution to the PnP problem,” International Journal of Computer Vision, vol.81, no.2, pp.155–166, 2009.

[34] O. Chum, J. Matas, and J. Kittler, “Locally optimized RANSAC,” Joint Pattern Recognition Symposium, pp.236–243, Springer, 2003.

[35] K. Kanatani, Y. Sugaya, and Y. Kanazawa, Guide to 3D Vision Computation, Springer, 2016.

Gaku Nakano received the M.S. and Ph.D. degrees in Engineering from the University of Tsukuba in 2008 and 2021, respectively. He has been a computer vision researcher at the Central Research Laboratories of NEC Corporation since 2008. He is a member of IEICE and IPSJ.