Minimal Solutions to Uncalibrated Two-view Geometry with Known Epipoles

Gaku Nakano
NEC Corporation
1753 Shimonumabe, Kawasaki, Japan
g-nakano@nec.com

Abstract

This paper proposes minimal solutions to uncalibrated two-view geometry with known epipoles. Exploiting the epipoles, we can reduce the number of point correspondences needed to find the fundamental matrix together with the intrinsic parameters: the focal length and the radial lens distortion. We define four cases by the number of available epipoles and unknown intrinsic parameters, then derive a closed-form solution for each case formulated as a higher-order polynomial in a single variable. The proposed solvers are more numerically stable and faster by orders of magnitude than the conventional 6- or 7-point algorithms. Moreover, we demonstrate by experiments on the human pose dataset that the proposed method can solve two-view geometry even with 2D human pose, of which point localization is noisier than general feature point detectors.

1. Introduction

Solving the two-view geometry of uncalibrated cameras is one of the basic tasks for many computer vision applications such as the structure-from-motion [1, 31], Visual-SLAM [6, 26], and novel view synthesis [4, 25]. The goal of this problem is to find the fundamental matrix and the intrinsic parameters of the two cameras from point correspondences. Since point correspondences are generally contaminated by outliers, many efforts have been devoted to develop efficient and numerically stable solvers that can be incorporated into RANSAC [9] to extract a set of inlier point pairs.

It is known that the fundamental matrix with an unknown focal length and known principal points can be determined by six point correspondences. Stevénius et al. [34] first developed a minimal solution to this problem by manually deriving Gröbner basis. Then, other methods were proposed based on various approaches, e.g., the hidden variable technique [23], the polynomial eigenvalue problem [19], the automatic generator [21]. Following the success of the 6-point algorithms, research interests have been expanded to deal with the radial lens distortion. Jiang et al. [15] and Kuang et al. [18] showed that the fundamental matrix with both unknown focal length and unknown radial distortion can be solved by seven point correspondences. However, with more intrinsic parameters to be estimated, new issues arise in terms of numerical stability and computational efficiency. The root cause is that those methods assume general conditions with no restrictions on the camera motion and thus cannot reduce the number of point correspondences.

To address the above issue, we assume that two cameras are mutually observed, as shown in Fig. 1. This is not a rare and limited situation but occurs in real applications. For example, multiple cameras are installed to capture human behaviors from 360° views in dataset creation (Figs. 2a and 2b). In dynamic scenes such as multi-agent Visual-SLAM, robots collaboratively create a single map of a wide area (Fig. 2c). Visual-SLAM can also be applied to calibrate camera networks consisting of fixed and moving cameras (Fig. 2d). In those scenarios, epipoles or 2D positions of the camera center are directly obtainable from images. Epipoles constrain a fundamental matrix to be rank-deficient, thus reducing the number of point correspondences. Nistér [27] showed a theory for finding a relative rotation of calibrated cameras with a single epipole. Ito et al. [14] utilized epipole(s) for fundamental matrix estimation, then Sato [30] generalized it to three- and four-view geometry. However, both methods are based on the DLT algorithm [11], thus not minimal solutions. They also do not address how to find the intrinsic parameters.

In this paper, we propose minimal solutions utilizing epipole(s) to determine a fundamental matrix with unknown intrinsic camera parameters: focal length and radial lens distortion. The proposed method has four variations depending on the number of available epipoles and the number of unknown parameters to be estimated. Moreover, we introduce a simple regularization term to fix the epipoles during the non-linear optimization, which is conducted on inliers after RANSAC. Through exhaustive experiments, we show that the proposed algorithms provide more reliable estimations than the conventional methods while achieving a significant acceleration by orders of magnitude.
where $\otimes$ denotes the Kronecker product and $f$ is a 9-dim vector representation of $F$.

Epipoles, $e = [x, y, 1]^T$ and $e' = [x', y', 1]^T$, are the 2D projections of the image center of one camera onto the image plane of the other camera, which satisfy

$$Fe = F^Te' = 0.$$  \hfill (2)

Equation (2) leads to one of the constraints on $F$, i.e., $\det(F) = 0$. Thus, $F$ is of rank two.

When the two cameras can be assumed to have the same focal length $f$, another constraint is given by

$$2FQF^TQF - tr(FQF^TQ)FQ = 0,$$  \hfill (3)

where $Q = \text{diag}(f^2, f^2, 1)$. Equation (3) is the necessary and sufficient conditions to recover $R$ and $t$ by decomposing $F$. Eliminating $f$, we can rewrite Eq. (3) in a closed-form w.r.t. $F$ [20] as follows:

$$h(F) = F_{11}F_{33}^2F_{31} + F_{13}^2F_{21}F_{32} + F_{11}F_{13}F_{23}^2F_{31}
+ F_{21}F_{33}^2F_{31} - F_{11}F_{13}^2F_{23} - F_{21}F_{33}^2F_{31} + F_{11}F_{13}F_{23}^2F_{31}
+ F_{21}F_{33}^2F_{31} - F_{11}F_{13}^2F_{23} + F_{21}F_{33}^2F_{31} - F_{11}F_{13}F_{23}^2F_{31}
- F_{11}F_{13}F_{23}^2F_{31} - F_{21}F_{33}^2F_{31} - F_{11}F_{13}F_{23}^2F_{31}
- F_{11}F_{13}F_{23}^2F_{31} - F_{21}F_{33}^2F_{31} - F_{11}F_{13}F_{23}^2F_{31}
- F_{11}F_{13}F_{23}^2F_{31} - F_{21}F_{33}^2F_{31} - F_{11}F_{13}F_{23}^2F_{31}
- F_{11}F_{13}F_{23}^2F_{31} - F_{21}F_{33}^2F_{31} - F_{11}F_{13}F_{23}^2F_{31}.$$  \hfill (4)

where $F_{ij}$ denotes the $(i, j)$ element of $F$.

Once we obtain an $F$ such that $\det(F) = 0$ and Eq. (3) (or Eq. (4)), the focal length $f$ can be computed in a closed-form [20]:

$$f^2 = \frac{F_{12}F_{13}F_{33}^2 + F_{22}F_{23}F_{33}^2}{F_{11}F_{13}F_{23}^2 + 2F_{21}F_{23}F_{33} + F_{12}F_{13}F_{23}^2} - \frac{F_{13}^2F_{23}F_{33} - F_{23}^2F_{33}F_{23}}{F_{11}F_{13}F_{23}^2 + 2F_{21}F_{23}F_{33} + F_{12}F_{13}F_{23}^2}.$$  \hfill (5)

It is well known that the relative motion, $R$ and $t$, can be obtained by using a singular value decomposition [11] or a closed-form method [13].

If the cameras have a wide angle lens with a radial distortion parameter $k$, a 2D point $m_i$ is observed as a radially distorted point $\hat{m}_i$. The relation can be written by using the division model [10]:

$$m = \begin{pmatrix} \hat{u} \\ \hat{v} \end{pmatrix} \propto \begin{pmatrix} \hat{u} \\ \hat{v} + kr \end{pmatrix},$$  \hfill (6)

where $r = \hat{u}^2 + \hat{v}^2$. Hereafter, we use the hat symbol $\hat{\cdot}$ to express radially distorted values.
3. Finding F-matrix with known epipoles

In this section, we propose four different algorithms depending on the number of unknown intrinsic parameters and the number of known epipoles. The basic strategy commonly used in all methods is to parameterize a fundamental matrix $F$ as a linear combination of the null space of a coefficient matrix computed from epipoles and point correspondences. We reduce the number of unknown variables to one and employ Eq. (4) to lead to a higher-order polynomial equation of the single variable. The fundamental matrix $F$ is obtained by substituting the real roots of the polynomial equation into the linear combination of the null space. The null space is derived from the epipoles, and Eq. (4) is used for the root finding, therefore, both the intrinsic and extrinsic parameters can be recovered from the computed $F$.

3.1. 2-pt algorithm for F+f with two epipoles

We start with the simplest case where two epipoles $e$ and $e'$ are observed by each camera with an unknown focal length $f$ but no lens distortion. If we define the $3 \times 3$ identity matrix by $I$ and a $6 \times 9$ matrix $N_1$ by

$$N_1 = \begin{bmatrix} e^T & e'^T & x'I & y'I & I \end{bmatrix}, \tag{7}$$

Eq. (2) can be written in the form

$$N_1 f = 0. \tag{8}$$

Here, $N_1$ is of size $6 \times 9$ but its row rank is only five.

Utilizing two point correspondences, a $8 \times 9$ coefficient matrix $M_1$ is given by

$$M_1 = \begin{bmatrix} \left[ N_1 \right] \left[ m'_1 \otimes m_1 \right]^T \left[ m'_2 \otimes m_2 \right]^T \end{bmatrix}. \tag{9}$$

Then, we have

$$M_1 f = 0. \tag{10}$$

Since the row rank of $M_1$ is seven, $f$ can be parameterized by a linear combination of two null vectors $v_1, v_2$ of $M_1$:

$$f = \alpha v_2 + v_1. \tag{11}$$

Substituting Eq. (11) into Eq. (4), we obtain a fifth-order equation for the single variable $\alpha$, which is represented by

$$h(F(\alpha)) = c_5 \alpha^5 + c_4 \alpha^4 + c_3 \alpha^3 + c_2 \alpha^2 + c_1 \alpha + c_0 = 0. \tag{12}$$

Once we found the roots of Eq. (12), the fundamental matrix $F$ is then obtained from Eq. (11). This algorithm gives at most five solutions.

3.2. 4-pt algorithm for F+f with single epipole

When one of the two epipoles is only available, we can solve the $F + f$ problem by adding more two point correspondences. Using four point correspondences together with the single epipole $e$, we can construct a $7 \times 9$ coefficient matrix $M_2$ by

$$M_2 = \begin{bmatrix} e^T & (m'_4 \otimes m_4)^T \end{bmatrix}, \tag{13}$$

which satisfies

$$M_2 f = 0. \tag{14}$$

The row rank of $M_2$ is seven, which is identical to $M_1$. Consequently, we can find $F$ using exactly the same approach as in Eqs. (11) and (12). The number of the real solutions of this algorithm is at most five.

3.3. 3-pt algorithm for F+f+k with two epipoles

We address a more complicated problem where a radial distortion parameter $k$ is additionally unknown. Since we have two epipoles in this case, this problem can be solved using three point correspondences.

In this problem, the epipoles, $\hat{e}$ and $\hat{e}'$, are observed in the distorted image coordinates, as with the point correspondences $\hat{m}$. Using Eq. (6), we can represent the undistorted epipoles, $e$ and $e'$, as

$$e = \frac{1}{1 + k z} \hat{e} \propto \begin{bmatrix} \hat{x} \\ \hat{y} \\ 1 + k z \end{bmatrix}, \tag{15}$$

$$e' = \frac{1}{1 + k' z'} \hat{e}' \propto \begin{bmatrix} \hat{x}' \\ \hat{y}' \\ 1 + k' z' \end{bmatrix},$$

where $z = \hat{x}^2 + \hat{y}^2$ and $z' = \hat{x}'^2 + \hat{y}'^2$. Note that the radial distortion $k$ is an unknown parameter.

Similar to Eqs. (7) and (8), we can rewrite Eq. (2) by

$$N_2 f = 0, \tag{16}$$

where

$$N_2 = \begin{bmatrix} e^T & e'^T & (m'_1 \otimes m_1)^T \end{bmatrix}. \tag{17}$$
The null space of \( N_2 \) can be symbolically given by

\[
V_1 = \begin{bmatrix}
\hat{y}'(1 + kz) & \hat{y}(1 + k'z') & (1 + kz)(1 + k'z') \\
-x'y' & 0 & -\hat{x}(1 + k'z') \\
0 & -x'y' & 0 \\
-x' \hat{y} & -x'(1 + kz) & 0 \\
\hat{x} \hat{x}' & 0 & 0 \\
0 & \hat{x} \hat{x}' & 0 \\
0 & 0 & -\hat{x}'(1 + kz) \\
0 & 0 & \hat{x}' \\
0 & 0 & 0
\end{bmatrix}
\]

Thus, the fundamental matrix can be represented as a linear combination of the column vectors of \( V_1 \), i.e.

\[
f = V_1 \alpha_1,
\]

where \( \alpha_1 \) is a 4-dim vector.

We now have three point correspondences, which satisfy \((m_i' \otimes m_i)^T f = 0\). If we define a \( 3 \times 4 \) matrix \( G_1 \) by

\[
G_1 = \begin{bmatrix}
(m_1' \otimes m_1)^T \\
(m_2' \otimes m_2)^T \\
(m_3' \otimes m_3)^T
\end{bmatrix} V_1,
\]

we obtain

\[
\begin{bmatrix}
(m_1' \otimes m_1)^T \\
(m_2' \otimes m_2)^T \\
(m_3' \otimes m_3)^T
\end{bmatrix} f = G_1 \alpha_1 = 0.
\]

The unknown vector \( \alpha_1 \) can be determined as the kernel of \( G_1 \) up to scale. Since \( G_1 \) is of size \( 3 \times 4 \), the kernel can be symbolically computed as a 4-dim vector:

\[
\alpha_1 \propto \text{Ker}(G_1) = Q_1 \beta_1,
\]

where \( Q_1 \) is a \( 4 \times 5 \) matrix and \( \beta_1 = [k^4, k^3, k^2, k, 1]^T \). We can replace \( \alpha_1 \) with \( Q_1 \beta_1 \), then Eq. \((19)\) can be rewritten as

\[
f = V_1 \alpha_1 = V_1 Q_1 \beta_1.
\]

Plugging Eq. \((23)\) into Eq. \((4)\), we finally obtain a 16th-degree polynomial equation in \( k \):

\[
h(F(k)) = c_{16} k^{16} + \cdots + c_1 k + c_0 = 0.
\]

We can recover \( f \) by substituting the real roots of Eq. \((24)\) into Eq. \((23)\). This algorithm provides at most 16 real solutions.

### 3.4. 5-pt algorithm for \( F+f+k \) with single epipole

Given only a single epipole \( e \), we need two additional point pairs to solve the \( F+f+k \) problem. We omit the details of the derivation and describe the differences only because the procedure of this algorithm is basically similar to Sec. 3.3.

The null space of \[
\begin{bmatrix}
e^T \\
e^T \\
e^T
\end{bmatrix}
\]

can be symbolically given by

\[
V_2 = \begin{bmatrix}
-\hat{y} & 1 + kz & 0 & 0 & 0 & 0 \\
\hat{x} & 0 & 0 & 0 & 0 & 0 \\
0 & -\hat{x} & 0 & 0 & 0 & 0 \\
0 & 0 & -\hat{y} & 1 + kz & 0 & 0 \\
0 & 0 & \hat{x} & 0 & 0 & 0 \\
0 & 0 & 0 & -\hat{x} & 0 & 0 \\
0 & 0 & 0 & 0 & -\hat{y} & 1 + kz \\
0 & 0 & 0 & 0 & \hat{x} & 0 \\
0 & 0 & 0 & 0 & 0 & -\hat{x}
\end{bmatrix}.
\]

Since \( V_2 \) is of size \( 9 \times 6 \), the unknown vector \( \alpha_2 \) for this problem becomes a 6-dim vector. Using the five point correspondences, \( Q_2 \) is represented as a \( 5 \times 6 \) matrix, as shown in Eq. \((20)\). Then, we can symbolically obtain the kernel of \( Q_2 \), which is given as the matrix-vector product of a \( 6 \times 6 \) matrix \( Q_2 \) and \( \beta_2 = [k^5, k^4, k^3, k^2, k, 1]^T \). Now we can express the fundamental matrix by

\[
f = V_2 Q_2 \beta_2.
\]

Substituting Eq. \((26)\) into Eq. \((4)\), we finally obtain a 21st-order polynomial equation in \( k \), i.e.

\[
h(F(k)) = c_{21} k^{21} + \cdots + c_1 k + c_0 = 0.
\]

Thus, we can obtain at most 21 real solutions.

### 3.5. Non-linear refinement

After running a RANSAC scheme, we generally obtain many inlier points and perform a non-linear refinement to polish the accuracy of the parameters. For example, minimizing the symmetric epipolar distance is one of the major cost functions in the literature \[11\], i.e.

\[
L_{sym} = \sum_{i=1}^n \frac{(m_i^T F m_i)^2}{|F m_i|^2 + |F|^2 |m_i|^2} + \frac{(m_i^T F m_i)^2}{|F|^2 |m_i|^2}.
\]

where \( |F m_i|^2 \) denotes the square of the j-th element of the vector \( F m_i \).

Since we have an initial guess of the focal length of the two images as \( f_1 = f_2 = f \), we can parameterize the fundamental matrix by \( F = \text{diag}(1,1,f) \).
where E represents an essential matrix. Thus, we can use the Helmke’s method [12] to update E with preserving 5 DoFs. Also, we initialize the lens distortion of the two images by \(k_1 = k_2 = k\).

However, those typical approaches of the non-linear refinement do not assume that epipoles are known, which in turn may degrade the fundamental matrix obtained by the proposed method. To avoid this issue, we introduce a regularization term of Eq. (2) as follows:

\[
L_{\text{epi}} = \|Fe\|^2 + \|FT e'\|^2.
\]  

(29)

Note that the epipoles, e and e’, are undistorted using Eq. (15) with \(k_1\) and \(k_2\). In the case of the single epipole, the second term \(\|FT e'\|^2\) is simply dropped.

The total cost function of the proposed scheme is given by jointly optimizing \(L_{\text{sym}}\) and \(L_{\text{epi}}\) with a weight \(w\):

\[
L = L_{\text{sym}} + wL_{\text{epi}}.
\]  

(30)

4. Degenerate configurations

Analysis of the degenerate configurations is important to avoid ambiguous 3D shape reconstructions and use the proposed method in real applications. Due to limitations of space, we give several simple examples for \(F + f\) problem with two known epipoles in this section.

Degenerate configurations of two-view geometry in general cases have been extensively studied for the past decades [5, 17, 33, 35]. One typical situation that could often occur in practice is forward motion, where both focal length and radial distortion cannot be uniquely determined regardless of the point distributions. Forward motion alone is not the case for the proposed solvers, but forward motion with special point distributions is. One configuration where the proposed 2-point solver for \(F + f\) problem cannot be solved is that two epipoles and two point correspondences are collinear under forward motion. For example, the row rank of \(M_1\) is less than seven in Eq. (9) when \(e = e' = [0, 0, 1]^T\), \(m_i = [u_i, 0, 1]^T\), \(m_i' = [-u_i, 0, 1]^T\), \(i \in \{1, 2\}\). Another well-known degenerate configuration for the general solvers is that all 3D points are on a plane. However, in our case, unique solutions can be obtained in planar scenes when two epipoles are known [30].

The rigorous conditions can be given by analyzing the rank of submatrices of \(N_1\) and \(N_1 (N_2 + M_2)\) for the other problems as well). In other words, we first find epipoles and image points that cause rank deficiency, then link them to actual camera motions and 3D point distributions. However, as reported in the previous studies, it is not easy to reveal all possible ambiguities. We would leave the in-depth discussion to future research.

5. Experiments

This section reports experimental results on synthetic and real data. We conducted all experiments on a PC with Core i7-13700k.

5.1. Implementation

We have implemented five conventional methods and the four proposed methods in MATLAB, as shown in Tab. 1. We chose three Gröbner basis solvers that are the state-of-the-art based on the efficient parameterization (Eq. (4)): Larsson’s 6-point solver for \(F + f\) problem (L6f) [21], Larsson’s 7-point solver for \(F + f + k\) problem (L7fk) [21], and Martyushev’s 7-point solver for \(F + f + k\) problem (M7fk) [24]. Those are designed for general scenes, therefore, epipoles are not assumed to be available. Moreover, we used two DLT-based approaches by Ito et al. [14]: the 3-point solver with two epipoles (I3), and the 5-point solver with a single epipole (I5). Although I3 and I5 do not guarantee to satisfy Eq. (3) or Eq. (4), we calculated focal length by simply applying Eq. (5) to a fundamental matrix given by I3 or I5. We used the root function for solving the polynomials of the proposed solvers, i.e. Eqs. (12), (24) and (27).

5.2. Evaluation metrics

We measured the following metrics to quantitatively evaluate the estimation accuracy:

\(\epsilon_F\) Symmetric epipolar distance in pixels by applying \(F_{\text{est}}\) onto the ground-truth 2D points.

\(\epsilon_f\) Relative error by \(|f_{\text{est}} - f_{\text{gt}}| / f_{\text{gt}} \times 100\%|\).

\(\epsilon_k\) Normalized pixel distance [%] between undistorted coordinates using \(k_{\text{est}}\) and \(k_{\text{gt}}\). 100 points were uniformly sampled, and the mean distance is normalized by the diagonal length of the image.

\(\epsilon_R\) Angle-axis error by \(\cos^{-1}(\frac{\text{tr}(R_{\text{gt}} R_{\text{est}}^{-1}) - 1}{2})\) [degrees].

\(\epsilon_t\) Cosine similarity by \(\cos^{-1}(\frac{t_{\text{gt}}^T t_{\text{est}}}{\|t_{\text{gt}}\| \|t_{\text{est}}\|})\) [degrees].

5.3. Synthetic data

We investigate the performance of the minimal solvers using synthetic data in this section.

5.3.1 Scene configuration

We created synthetic scenes as follows. Two cameras were randomly located in the area of 5.0 m \(\times\) 5.0 m and 2.0 m high to project each other, as shown in Fig. 1b. Small perturbations within 5 degrees and \(\pm 0.2\) m were added to the rotation and position of the cameras, respectively. The image resolution was \(640 \times 480\), the focal length was set to...
Method Reference Problem Decomposability Eqs. (3) and (4) # of points # of epipoles # of solutions Approach
L6f [21] F + f ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓
5.3.5 Computational time

Figure 6 shows the mean runtime of all methods over $10^6$ independent trials. The computational time of the proposed methods is less than 1 msec, which is not as fast as the Ito et al.’s methods, but is still reasonably fast enough for real-time applications.

5.4. Real data

In this experiments, we incorporated each method into RANSAC together with the non-linear refinement to investigate the performance in real data. We excluded L7fk due to its instability and high computational cost shown in the synthetic data experiments in Sec. 5.3.

5.4.1 Dataset

We used two publicly available dataset where multiple cameras are mutually projected:

**HumanEva** [32]¹ Four RGB cameras with the $656 \times 490$ resolution are installed in the corners of a room. The intrinsic parameters are calibrated by Zhang’s method, then the extrinsics are calculated using Vicon sensors. A single person is asked to do simple actions such as walking, jogging, etc. There are two combinations of mutual camera projections with two epipoles.

**Panoptic studio** [16]² 30 HD cameras, 480 VGA cameras, and 5 Kinect2 are installed on the walls of a dome-shaped studio. Both the intrinsic and extrinsic parameters of all cameras are simultaneously calibrated based on the structure-from-motion. All sensors are used for recording multi-person interactions. There are 20 combinations of a single epipole and 80 combinations of two epipoles in the HD cameras.

We chose a video sequence from each dataset in which a single person moves around the scene: S2 sequence of HumanEva (Fig. 2a) and dance1 of Panoptic studio (Fig. 2b). Then, 2D human joints on the video frames were detected using OpenPose [7]. The neck and hip joints were selected as the point correspondences because they are less likely to be occluded and are easily observed from any camera. We manually labeled epipole locations in each image.

---

¹http://humaneva.is.tue.mpg.de/
²http://domedb.perception.cs.cmu.edu/
Table 2. Quantitative results of real image dataset (best in bold and second-best underlined). iter.: the number of RANSAC iterations, time: the total processing time of RANSAC and the non-linear refinement in seconds.

5.4.2 Quantitative results

For testing the methods, we configured RANSAC to have the threshold by 3 pixels, the confidence by 0.995, and the maximum iteration number by 10000. The random seed was fixed for all methods during a single trial so that the same sample points were drawn. After RANSAC, the non-linear refinement described in Sec. 3.5 was performed on inliers to polish the camera parameters to be more accurate.

We conducted 100 and 10 trials for each camera pair of HumanEva S2 and dance1, respectively. Table 2 reports the average values of ($\epsilon_R$, $\epsilon_t$, $\epsilon_f$, $\epsilon_k$), the number of RANSAC iterations, and the total processing time. Comparing Tab. 2a with Tabs. 2b and 2c, we can see that all methods show better scores in dance1 even though two epipoles are available in HumanEva. The person in dance1 is captured in the higher-resolution image than in HumanEva, resulting in higher localization accuracy of the 2D joints.

Table 2 indicates that the proposed solvers outperform the conventional methods in either case of one or two epipoles. The proposed solvers show the best or the second-best scores in most evaluation criteria. For example, P2f and P4f are more than 100 times faster than M7fk while achieving lower estimation errors. Although P2f and P4f does not estimate the radial distortion $k$ in RANSAC, the non-linear refinement successfully converges to an accurate solution of $k$ due to a good initial guess of $F$ and $f$. On the other hand, in spite of the use of epipoles, I3 and I5 are inferior to the general solvers L6f and M7fk. As reported in Sec. 5.3, two Ito’s methods are sensitive to image noise. According to the above observations, we can conclude that satisfying the decomposability conditions, Eqs. (3) and (4), significantly improves the estimation accuracy even with known epipoles.
5.4.3 Qualitative results

Figures 7 and 8 visualize a 3D shape reconstruction of $s_2$ and dance1, respectively. For ease of viewing, only neck joints are shown in both 2D and 3D images. We plotted the position of estimated epipoles over 30 independent trials in Figs. 7b and 8b. The P3fk’s epipole and the ground-truth overlap so closely that the difference is not visible. On the other hand, the camera motion given by M7fk has large errors and the epipole is estimated on a different place in each trial. This is consistent with the result in Tab. 2 where the proposed methods have the smallest motion errors. Also, we can visually validate that P3fk reconstructs a 3D shape more accurately than M7fk in Figs. 7c, 7d, 8c and 8d.

6. Limitations and future work

One of the limitations is that epipole positions are manually annotated. This is not a critical issue in data set creation (Figs. 2a and 2b) because device installations and data capturing spend time dominantly. However, a fast-automatic epipole detector is required to use the proposed method in real-time applications such as multi-agent Visual-SLAM [28, 29, 36] shown in Figs. 2c and 2d. Although direct epipole estimation has been studied for several decades [2, 8, 22], unfortunately it cannot be said as active nor extensive. We think that the manual labeling is merely a current limitation, which can be resolved by future effort in the computer vision community. For instance, a quick workaround is to place a camera at the center of a colored sphere or circle and detect the camera’s center in images by ellipse fitting. Alternatively, there are more sophisticated ways that use wifi/radio waves or IMUs to find the relative locations between devices in 3D space [3].

Another future work is to extend the proposed approach to multi-view geometry, just as Sato [30] extended Ito et al.’s DLT method [14] for obtaining a trifocal or quadrifocal tensor. Combining the multi-view extension with an automatic epipole detector, we can expect that the proposed approach stabilizes multi-agent Visual-SLAM with wide baselines for large-scale scene.

7. Conclusion

We have presented four minimal solutions to the uncalibrated two-view geometry where one or two epipoles are visible in the image plane. All solvers are manually formulated as higher-order polynomials in a single variable without an automatic Gröbner basis generator. We briefly discussed the degenerate conditions and showed several examples where the proposed solvers are feasible but not the conventional methods. In the experiments, we first compared the proposed solvers with the conventional methods in the numerical stability and robustness against image noise using synthetic data. Then, we demonstrated that the proposed solvers successfully stabilize the structure-from-motion using human joints as point correspondences, which are noisy and less accurate than image feature points. Through these experiments, the proposed solvers outperform the conventional methods in estimation accuracy with maintaining computational efficiency. Finally, we discussed the current limitations of the proposed approach and future work to be applicable in real-time applications. We hope this paper will motivate the computer vision community to develop automatic epipole detectors more extensively.

References


