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.

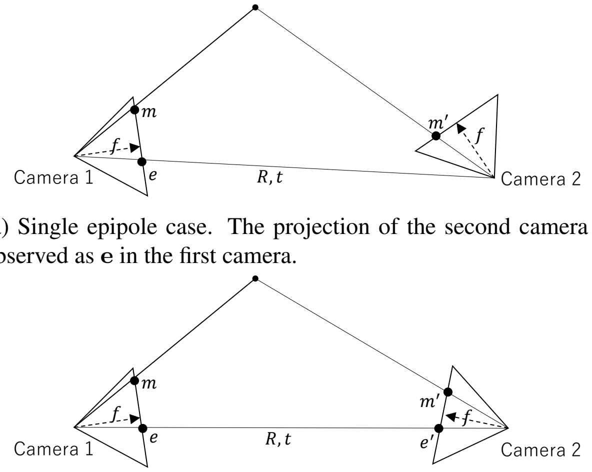
1. Introduction

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. Stewénius *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.

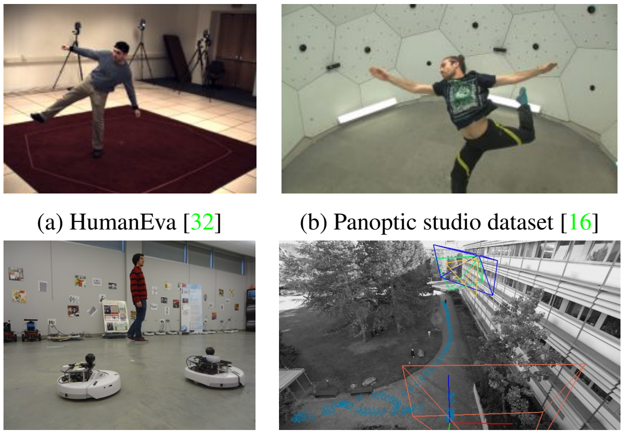
In this paper, we propose minimal solutions utilizing epipole(s) to determine a fundamental matrix with unknown intrinsic camera parameters: focal length and radial lens distortion. The proposed method has four variations depending on the number of available epipoles and the number of unknown parameters 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.



(a) Single epipole case. The projection of the second camera is observed as $\mathbf{e}$ in the first camera.

(b) Two epipoles case. The two cameras are mutually observed as $\mathbf{e}$ and $\mathbf{e}'$ in each image.

Figure 1. Two-view geometry with known epipoles.



(a) HumanEva [32]

(b) Panoptic studio dataset [16]

(c) Collaborative Visual-SLAM [28, 36]

(d) SLAM-based calibration for surveillance and moving cameras [29]

Figure 2. Mutual camera projections in applications. (a)–(b): Human pose dataset. (c)–(d): Multi-agent Visual-SLAM.

2. Uncalibrated two-view geometry

This section describes the theoretical background of two-view geometry for uncalibrated cameras. We represent 2D points as their homogeneous coordinates, *i.e.* $\mathbf{m} = [u, v, 1]^\mathsf{T}$, and assume a pinhole camera model whose intrinsic parameters are partially known: the principal point is approximated by the image center; the skewness is zero. Solving uncalibrated two-view geometry is to find the focal length f and the radial lens distortion k as well as the relative rotation $\mathtt{R}$ and translation $\mathbf{t}$ between the two cameras.

Given n pairs of a 2D point correspondence $\{\mathbf{m}_i \leftrightarrow \mathbf{m}'_i\}$ between the two cameras, each point pair is constrained by a 3×3 fundamental matrix $\mathtt{F}$:

$$\mathbf{m}_i'^\mathsf{T} \mathbf{F} \mathbf{m}_i = (\mathbf{m}'_i \otimes \mathbf{m}_i)^\mathsf{T} \mathbf{f} = 0, \tag{1}$$

where $\otimes$ denotes the Kronecker product and $\mathbf{f}$ is a 9-dim vector representation of $\mathtt{F}$.

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

$$\mathtt{F}\mathbf{e} = \mathtt{F}^\mathsf{T}\mathbf{e}' = \mathbf{0}. \tag{2}$$

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

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

$$2\mathtt{FQF}^\mathsf{T}\mathtt{QFQ} - \operatorname{tr}(\mathtt{FQF}^\mathsf{T}\mathtt{Q})\mathtt{FQ} = \mathbf{0}, \tag{3}$$

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

$$\begin{aligned} h(\mathtt{F}) = {} & F_{11}F_{13}^3F_{31} + F_{13}^2F_{21}F_{23}F_{31} + F_{11}F_{13}F_{23}^2F_{31} \\ & + F_{21}F_{23}^3F_{31} - F_{11}F_{13}F_{31}^3 - F_{21}F_{23}F_{31}^3 + F_{12}F_{13}^3F_{32} \\ & + F_{13}^2F_{22}F_{23}F_{32} + F_{12}F_{13}F_{23}^2F_{32} + F_{22}F_{23}^3F_{32} \\ & - F_{12}F_{13}F_{31}^2F_{32} - F_{22}F_{23}F_{31}^2F_{32} - F_{11}F_{13}F_{31}F_{32}^2 \\ & - F_{21}F_{23}F_{31}F_{32}^2 - F_{12}F_{13}F_{32}^3 - F_{22}F_{23}F_{32}^3 \\ & - F_{11}^2F_{13}^2F_{33} - F_{12}^2F_{13}^2F_{33} - 2F_{11}F_{13}F_{21}F_{23}F_{33} \\ & - 2F_{12}F_{13}F_{22}F_{23}F_{33} - F_{21}^2F_{23}^2F_{33} - F_{22}^2F_{23}^2F_{33} \\ & + F_{11}^2F_{31}^2F_{33} + F_{21}^2F_{31}^2F_{33} + 2F_{11}F_{12}F_{31}F_{32}F_{33} \\ & + 2F_{21}F_{22}F_{31}F_{32}F_{33} + F_{12}^2F_{32}^2F_{33} + F_{22}^2F_{32}^2F_{33} = 0, \end{aligned} \tag{4}$$

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

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

$$f^2 = \frac{\begin{gathered} F_{12}F_{13}F_{33}^2 + F_{22}F_{23}F_{33}^2 \\ - F_{13}^2F_{32}F_{33} - F_{23}^2F_{32}F_{33} \end{gathered}}{\begin{gathered} F_{11}F_{13}F_{31}F_{32} + F_{21}F_{23}F_{31}F_{32} + F_{12}F_{13}F_{32}^2 \\ + F_{22}F_{23}F_{32}^2 - F_{12}^2F_{32}F_{33} - F_{22}^2F_{32}F_{33} \\ - F_{11}F_{12}F_{31}F_{33} - F_{21}F_{22}F_{31}F_{33} \end{gathered}}. \tag{5}$$

It is well known that the relative motion, $\mathtt{R}$ and $\mathbf{t}$, 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 $\mathbf{m}$ is observed as a radially distorted point $\hat{\mathbf{m}}$. The relation can be written by using the division model [10]:

$$\mathbf{m} = \frac{1}{1 + kr}\hat{\mathbf{m}} \propto \begin{bmatrix} \hat{u} \\ \hat{v} \\ 1 + kr \end{bmatrix}, \tag{6}$$

where $r = \hat{u}^2 + \hat{v}^2$. Hereafter, we use the hat symbol $\hat{}$ 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 $\mathtt{F}$ 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 $\mathtt{F}$ 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 $\mathtt{F}$.

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

We start with the simplest case where two epipoles $\mathbf{e}$ and $\mathbf{e}'$ are observed by each camera with an unknown focal length f but no lens distortion. If we define the 3×3 identity matrix by $\mathtt{I}$ and a 6×9 matrix $\mathtt{N}_1$ by

$$\mathtt{N}_1 = \begin{bmatrix} \mathbf{e}^\mathsf{T} & & \\ & \mathbf{e}^\mathsf{T} & \\ & & \mathbf{e}^\mathsf{T} \\ x'\mathtt{I} & y'\mathtt{I} & \mathtt{I} \end{bmatrix}, \tag{7}$$

Eq. (2) can be written in the form

$$\mathtt{N}_1 \mathbf{f} = \mathbf{0}. \tag{8}$$

Here, $\mathtt{N}_1$ is of size 6×9 but its row rank is only five.

Utilizing two point correspondences, a 8×9 coefficient matrix $\mathtt{M}_1$ is given by

$$\mathtt{M}_1 = \begin{bmatrix} \mathtt{N}_1 \\ (\mathbf{m}'_1 \otimes \mathbf{m}_1)^\mathsf{T} \\ (\mathbf{m}'_2 \otimes \mathbf{m}_2)^\mathsf{T} \end{bmatrix}. \tag{9}$$

Then, we have

$$\mathtt{M}_1 \mathbf{f} = \mathbf{0}. \tag{10}$$

Since the row rank of $\mathtt{M}_1$ is seven, $\mathbf{f}$ can be parameterized by a linear combination of two null vectors $\mathbf{v}_1$, $\mathbf{v}_2$ of $\mathtt{M}_1$:

$$\mathbf{f} = \alpha \mathbf{v}_2 + \mathbf{v}_1. \tag{11}$$

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

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

Once we found the roots of Eq. (12), the fundamental matrix $\mathtt{F}$ 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 $\mathtt{F} + f$ problem by adding more two point correspondences. Using four point correspondences together with the single epipole $\mathbf{e}$, we can construct a 7×9 coefficient matrix $\mathtt{M}_2$ by

$$\mathtt{M}_2 = \begin{bmatrix} \mathbf{e}^\mathsf{T} & & \\ & \mathbf{e}^\mathsf{T} & \\ & & \mathbf{e}^\mathsf{T} \\ & (\mathbf{m}'_1 \otimes \mathbf{m}_1)^\mathsf{T} & \\ & \vdots & \\ & (\mathbf{m}'_4 \otimes \mathbf{m}_4)^\mathsf{T} & \end{bmatrix}, \tag{13}$$

which satisfies

$$\mathtt{M}_2 \mathbf{f} = \mathbf{0}. \tag{14}$$

The row rank of $\mathtt{M}_2$ is seven, which is identical to $\mathtt{M}_1$. Consequently, we can find $\mathtt{F}$ 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, $\hat{\mathbf{e}}$ and $\hat{\mathbf{e}}'$, are observed in the distorted image coordinates, as with the point correspondences $\hat{\mathbf{m}}$. Using Eq. (6), we can represent the undistorted epipoles, $\mathbf{e}$ and $\mathbf{e}'$, as

$$\begin{aligned} \mathbf{e} &= \frac{1}{1+kz}\hat{\mathbf{e}} \propto \begin{bmatrix} \hat{x} \\ \hat{y} \\ 1+kz \end{bmatrix}, \\ \mathbf{e}' &= \frac{1}{1+kz'}\hat{\mathbf{e}}' \propto \begin{bmatrix} \hat{x}' \\ \hat{y}' \\ 1+kz' \end{bmatrix}, \end{aligned} \tag{15}$$

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

$$\mathtt{N}_2 \mathbf{f} = \mathbf{0}, \tag{16}$$

where

$$\mathtt{N}_2 = \begin{bmatrix} \mathbf{e}^\mathsf{T} & & \\ & \mathbf{e}^\mathsf{T} & \\ & & \mathbf{e}^\mathsf{T} \\ \hat{x}'\mathtt{I} & \hat{y}'\mathtt{I} & (1+kz')\mathtt{I} \end{bmatrix}. \tag{17}$$

The null space of $\mathtt{N}_2$ can be symbolically given by

$$\mathtt{V}_1 = \begin{bmatrix} \hat{y}\hat{y}' & \hat{y}'(1+kz) & \hat{y}(1+kz') & (1+kz)(1+kz') \\ -\hat{x}\hat{y}' & 0 & -\hat{x}(1+kz') & 0 \\ 0 & -\hat{x}\hat{y}' & 0 & -\hat{x}(1+kz') \\ -\hat{x}'\hat{y} & -\hat{x}'(1+kz) & 0 & 0 \\ \hat{x}\hat{x}' & 0 & 0 & 0 \\ 0 & \hat{x}\hat{x}' & 0 & 0 \\ 0 & 0 & -\hat{x}'\hat{y} & -\hat{x}'(1+kz) \\ 0 & 0 & \hat{x}\hat{x}' & 0 \\ 0 & 0 & 0 & \hat{x}\hat{x}' \end{bmatrix}. \quad (18)$$

Thus, the fundamental matrix can be represented as a linear combination of the column vectors of $\mathtt{V}_1$, *i.e*.

$$\mathbf{f} = \mathtt{V}_1\boldsymbol{\alpha}_1, \quad (19)$$

where $\boldsymbol{\alpha}_1$ is a 4-dim vector.

We now have three point correspondences, which satisfy $(\mathbf{m}'_i \otimes \mathbf{m}_i)^\mathsf{T}\mathbf{f} = 0$. If we define a 3×4 matrix $\mathtt{G}_1$ by

$$\mathtt{G}_1 = \begin{bmatrix} (\mathbf{m}'_1 \otimes \mathbf{m}_1)^\mathsf{T} \\ (\mathbf{m}'_2 \otimes \mathbf{m}_2)^\mathsf{T} \\ (\mathbf{m}'_3 \otimes \mathbf{m}_3)^\mathsf{T} \end{bmatrix} \mathtt{V}_1, \quad (20)$$

we obtain

$$\begin{bmatrix} (\mathbf{m}'_1 \otimes \mathbf{m}_1)^\mathsf{T} \\ (\mathbf{m}'_2 \otimes \mathbf{m}_2)^\mathsf{T} \\ (\mathbf{m}'_3 \otimes \mathbf{m}_3)^\mathsf{T} \end{bmatrix} \mathbf{f} = \mathtt{G}_1\boldsymbol{\alpha}_1 = \mathbf{0}. \quad (21)$$

The unknown vector $\boldsymbol{\alpha}_1$ can be determined as the kernel of $\mathtt{G}_1$ up to scale. Since $\mathtt{G}_1$ is of size 3×4, the kernel can be symbolically computed as a 4-dim vector:

$$\boldsymbol{\alpha}_1 \propto \mathrm{Ker}(\mathtt{G}_1) = \mathtt{Q}_1\boldsymbol{\beta}_1, \quad (22)$$

where $\mathtt{Q}_1$ is a 4×5 matrix and $\boldsymbol{\beta}_1 = [k^4, k^3, k^2, k, 1]^\mathsf{T}$. We can replace $\boldsymbol{\alpha}_1$ with $\mathtt{Q}_1\boldsymbol{\beta}_1$, then Eq. (19) can be rewritten as

$$\mathbf{f} = \mathtt{V}_1\boldsymbol{\alpha}_1 = \mathtt{V}_1\mathtt{Q}_1\boldsymbol{\beta}_1. \quad (23)$$

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

$$h(\mathtt{F}(k)) = c_{16}k^{16} + \cdots + c_1k + c_0 = 0. \quad (24)$$

We can recover $\mathbf{f}$ 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 $\mathbf{e}$, we need two additional point pairs to solve the $\mathtt{F} + f + k$ 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 $\begin{bmatrix} \mathbf{e}^\mathsf{T} & & \\ & \mathbf{e}^\mathsf{T} & \\ & & \mathbf{e}^\mathsf{T} \end{bmatrix}$ can be symbolically given by

$$\mathtt{V}_2 = \begin{bmatrix} -\hat{y} & 1+kz & 0 & 0 & 0 & 0 \\ \hat{x} & 0 & 0 & 0 & 0 & 0 \\ 0 & -\hat{x} & 0 & 0 & 0 & 0 \\ 0 & 0 & -\hat{y} & 1+kz & 0 & 0 \\ 0 & 0 & \hat{x} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\hat{x} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\hat{y} & 1+kz \\ 0 & 0 & 0 & 0 & \hat{x} & 0 \\ 0 & 0 & 0 & 0 & 0 & -\hat{x} \end{bmatrix}. \quad (25)$$

Since $\mathtt{V}_2$ is of size 9×6, the unknown vector $\boldsymbol{\alpha}_2$ for this problem becomes a 6-dim vector. Using the five point correspondences, $\mathtt{G}_2$ is represented as a 5×6 matrix, as shown in Eq. (20). Then, we can symbolically obtain the kernel of $\mathtt{G}_2$, which is given as the matrix-vector product of a 6×6 matrix $\mathtt{Q}_2$ and $\boldsymbol{\beta}_2 = [k^5, k^4, k^3, k^2, k, 1]^\mathsf{T}$. Now we can express the fundamental matrix by

$$\mathbf{f} = \mathtt{V}_2\mathtt{Q}_2\boldsymbol{\beta}_2. \quad (26)$$

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

$$h(\mathtt{F}(k)) = c_{21}k^{21} + \cdots + c_1k + c_0 = 0. \quad (27)$$

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*.

$$L_{\mathrm{sym}} = \sum_{i=1}^{n} \frac{(\mathbf{m}'^{\mathsf{T}}_i\mathtt{F}\mathbf{m}_i)^2}{[\mathtt{F}\mathbf{m}_i]_1^2 + [\mathtt{F}\mathbf{m}_i]_2^2} + \frac{(\mathbf{m}'^{\mathsf{T}}_i\mathtt{F}\mathbf{m}_i)^2}{[\mathtt{F}^\mathsf{T}\mathbf{m}'_i]_1^2 + [\mathtt{F}^\mathsf{T}\mathbf{m}'_i]_2^2}, \quad (28)$$

where $[\mathtt{F}\mathbf{m}_i]_j^2$ denotes the square of the j-th element of the vector $\mathtt{F}\mathbf{m}_i$.

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 $\mathtt{F} = \mathrm{diag}(1, 1, f_2)\ \mathtt{E}\ \mathrm{diag}(1, 1, f_1)$, where $\mathtt{E}$ represents an essential matrix. Thus, we can use the Helmke's method [12] to update $\mathtt{E}$ 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:

$$L_{\text{epi}} = \|\mathtt{F}\mathbf{e}\|^2 + \left\|\mathtt{F}^\mathsf{T}\mathbf{e}'\right\|^2. \tag{29}$$

Note that the epipoles, $\mathbf{e}$ and $\mathbf{e}'$, are undistorted using Eq. (15) with k_1 and k_2. In the case of the single epipole, the second term $\left\|\mathtt{F}^\mathsf{T}\mathbf{e}'\right\|^2$ is simply dropped.

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

$$L = L_{\text{sym}} + wL_{\text{epi}}. \tag{30}$$

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 $\mathtt{F} + f$ 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 $\mathtt{F} + f$ problem cannot be solved is that two epipoles and two point correspondences are collinear under forward motion. For example, the row rank of $\mathtt{M}_1$ is less than seven in Eq. (9) when $\mathbf{e} = \mathbf{e}' = [0, 0, 1]^\mathsf{T}$, $\mathbf{m}_i = [u_i, 0, 1]^\mathsf{T}$, $\mathbf{m}'_i = [-u_i, 0, 1]^\mathsf{T}, i \in \{1, 2\}$. 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 $\mathtt{N}_1$ and $\mathtt{M}_1$ ($\mathtt{N}_2$ and $\mathtt{M}_2$ 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 $\mathtt{F}+f$ problem (L6f) [21], Larsson's 7-point solver for $\mathtt{F} + f + k$ problem (L7fk) [21], and Martyushev's 7-point solver for $\mathtt{F} + f + k$ 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:

- $\epsilon_\mathtt{F}$ Symmetric epipolar distance in pixels by applying $\mathtt{F}_{\text{est}}$ onto the ground-truth 2D points.
- ϵ_f Relative error by $|f_{\text{est}} - f_{\text{gt}}|/f_{\text{gt}} \times 100$ [%].
- ϵ_k Normalized pixel distance [%] between undistorted coordinates using k_{est} and k_{gt}. 100 points were uniformly sampled, and the mean distance is normalized by the diagonal length of the image.
- $\epsilon_\mathtt{R}$ Angle-axis error by $\cos^{-1}\left(\frac{\mathrm{tr}(\mathtt{R}_{\text{gt}}^\mathsf{T}\mathtt{R}_{\text{est}})-1}{2}\right)$ [degrees].
- $\epsilon_\mathbf{t}$ Cosine similarity by $\cos^{-1}\left(\frac{\mathbf{t}_{\text{gt}}^\mathsf{T}\mathbf{t}_{\text{est}}}{\|\mathbf{t}_{\text{gt}}\|\|\mathbf{t}_{\text{est}}\|}\right)$ [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 × 5.0 m and 2.0 m high to project each other, as shown in Fig. 1b. Small perturbations within 5 degrees and ± 0.2 m were added to the rotation and position of the cameras, respectively. The image resolution was 640 × 480, the focal length was set to

Method	Reference	Problem	Decomposability Eqs. (3) and (4)	# of points	# of epipoles	# of solutions	Approach
L6f	[21]	$\mathtt{F}+f$	✓	6	0	15	Gröbner basis
L7fk	[21]	$\mathtt{F}+f+k$	✓	7	0	68	
M7fk	[24]	$\mathtt{F}+f+k$	✓	7	0	68	
I5	[14]	$\mathtt{F}$		5	1	1	DLT
I3	[14]	$\mathtt{F}$		3	2	1	
P4f	this paper	$\mathtt{F}+f$	✓	4	1	5	Closed-form
P2f		$\mathtt{F}+f$	✓	2	2	5	
P5fk		$\mathtt{F}+f+k$	✓	5	1	16	
P3fk		$\mathtt{F}+f+k$	✓	3	2	21	

Table 1. Methods compared in the experiments.

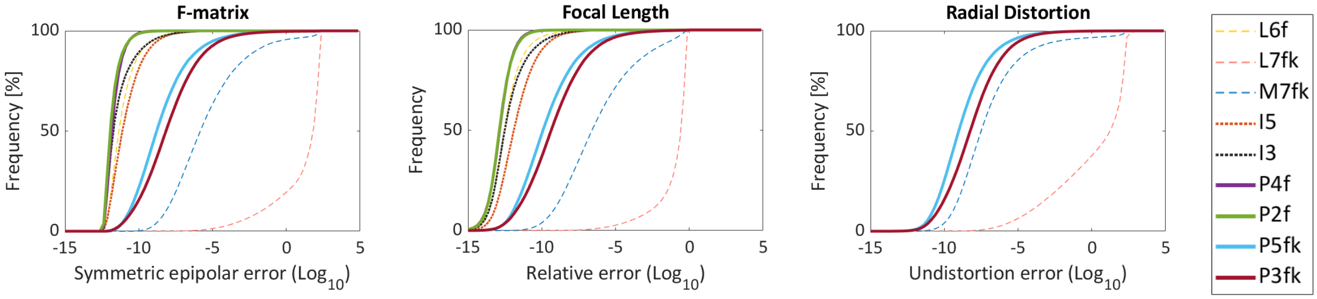


Figure 3. The cumulative histogram of numerical accuracy for 10^6 trials of randomly generated data without noise.

$f = 381.3$, which is approximately 80° HFOV. The radial lens distortion was set to $k = -0.2/f^2$ for the $\mathtt{F}+f+k$ problem. We uniformly generated 100 3D points in the common viewing frustum of the two cameras and projected the 3D points onto the image plane. Then, we randomly picked the minimal number of the points needed for each method.

5.3.2 Numerical stability

In the first test, we investigated the numerical stability of the methods in noise free data. Figure 3 summarizes a cumulative histogram of $\epsilon_{\mathtt{F}}$, ϵ_f, and ϵ_k over 10^6 independent trials. The figure indicates that the proposed methods are numerically stable. Comparing the methods in each categories, we see that P2f and P4f are better than I3, I5, and L6f in the $\mathtt{F}+f$ problem, and P5fk and P3fk are better than L7fk and M7fk in the $\mathtt{F}+f+k$ problem. Here, it should be noted that L7fk is so unstable as to be impractical. We further investigate this issue in the following experiments.

5.3.3 Robustness w.r.t. noise on points

To evaluate the robustness against the image noise, we measured the estimation errors by adding the zero-mean Gaussian noise onto point correspondences with the standard deviation $0 \leq \sigma \leq 2$ pixels. In this experiment, the image noise was not added to the epipoles, that is, the ground-truth values are given for the epipoles. Figure 4 shows the mean error over 10^5 independent trials for each noise level. From the figure, we see that the proposed methods are the most robust against noise. Since the two DLT methods, I3 and I5, do not consider the decomposability conditions described in Eqs. (3) and (4), their focal length estimation are sensitive to image noise, which is even worse than the general methods, L6f and M7fk. The proposed methods outperform M7fk in estimation of the fundamental matrix and lens distortion while showing comparable accuracy in the focal length estimation.

5.3.4 Robustness w.r.t. noise on epipoles

In this experiment, we studied the noise robustness of the six methods using epipoles: two Ito *et al.*'s and four proposed methods. We added the Gaussian noise on epipoles varying $0 \leq \sigma \leq 2$ while fixing the image noise as $\sigma = 0.5$. Figure 5 reports the mean error over 10^5 independent trials for each noise level. The results have similar trends to the previous experiments. Among the proposed methods, P4f and P5fk are more robust to image noise than P2f and P4fk because P4f and P5fk utilizes a single epipole.

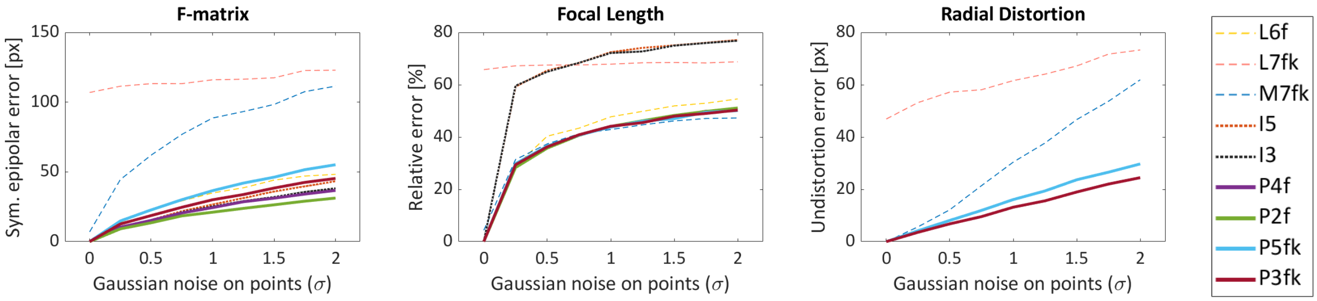


Figure 4. Mean error w.r.t. the image noise on point correspondences over 10^5 trials for each noise level.

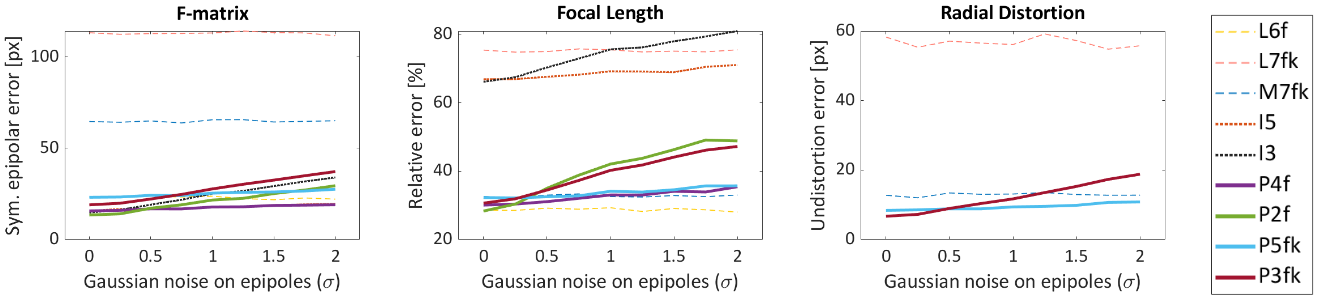


Figure 5. Mean error w.r.t. the image noise on epipoles over 10^5 trials for each noise level.

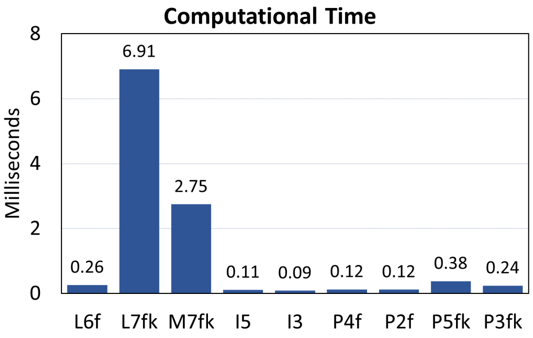


Figure 6. Runtime comparison of the minimal solvers.

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 × 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.

[1] http://humaneva.is.tue.mpg.de/

[2] http://domedb.perception.cs.cmu.edu/

Method	$\epsilon_{\mathtt{R}}$	$\epsilon_{\mathtt{t}}$	ϵ_f	ϵ_k	iter.	time
L6f	28.0	14.3	48.8	0.78	1241	0.25
M7fk	29.8	15.1	56.1	1.46	2819	7.69
I3	32.8	16.1	70.0	**0.26**	92	0.04
P2f	**22.7**	**11.3**	48.0	**0.26**	**39**	**0.03**
P3fk	24.7	12.3	**46.9**	1.11	146	0.04

(a) HumanEva: `S2` (2 epipoles, 2 sequences)

Method	$\epsilon_{\mathtt{R}}$	$\epsilon_{\mathtt{t}}$	ϵ_f	ϵ_k	iter.	time
L6f	17.3	8.5	18.6	2.70	10000	2.68
M7fk	23.8	12.2	19.8	2.61	10000	33.43
I5	15.5	12.4	30.7	2.69	10000	**0.83**
P4f	11.3	6.8	**18.0**	2.68	10000	1.22
P5fk	**11.2**	**6.5**	18.2	**2.09**	10000	3.64

(b) Panoptic studio: `dance1` (single epipole, 20 sequences)

Method	$\epsilon_{\mathtt{R}}$	$\epsilon_{\mathtt{t}}$	ϵ_f	ϵ_k	iter.	time
L6f	10.3	4.9	19.8	2.67	10000	1.80
M7fk	12.0	5.9	20.0	2.85	10000	24.88
I3	13.1	8.0	33.7	2.66	2001	0.15
P2f	**6.1**	**2.8**	**17.7**	2.66	**318**	**0.04**
P3fk	6.6	3.2	17.9	**1.72**	2438	0.40

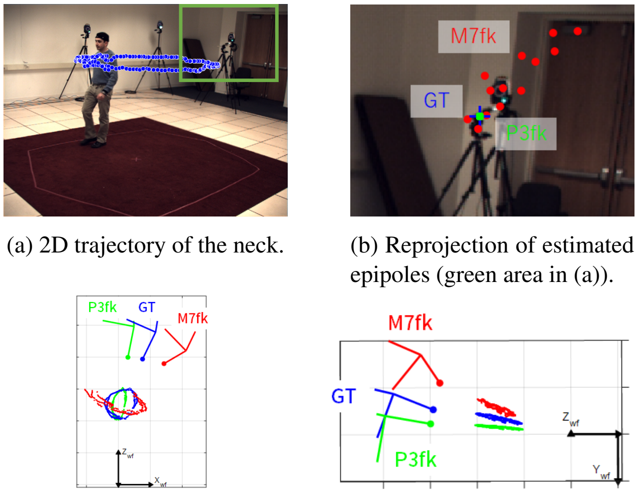
(c) Panoptic studio: `dance1` (2 epipoles, 80 sequences)

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. The two general solvers, L6f and M7fk, optimized only the symmetric epipolar distance L_{sym} in Eq. (29). The rest of the solvers utilizing epipoles employed the total cost function L in Eq. (30) with a weight $w = 100 \times n$ to balance L_{sym} and L_{epi} depending on the number of the points n.

We conducted 100 and 10 trials for each camera pair of `S2` and `dance1`, respectively. Table 2 reports the average values of ($\epsilon_{\mathtt{R}}$, $\epsilon_{\mathtt{t}}$, ϵ_f, ϵ_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.



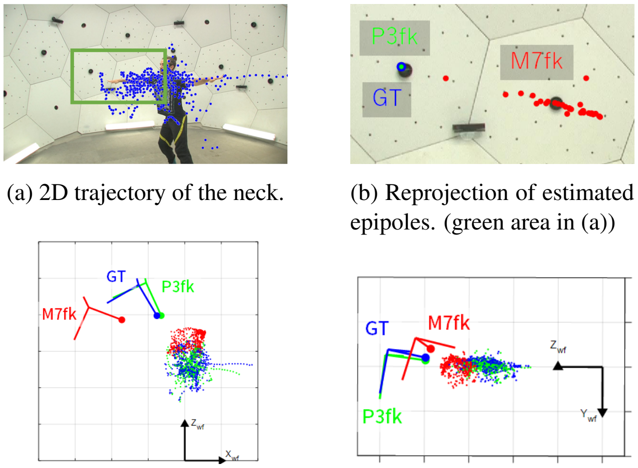
(a) 2D trajectory of the neck.

(b) Reprojection of estimated epipoles (green area in (a)).

(c) Top view.

(d) Side view.

Figure 7. 3D shape reconstruction of HumanEva `S2`.



(a) 2D trajectory of the neck.

(b) Reprojection of estimated epipoles. (green area in (a))

(c) Top view.

(d) Side view.

Figure 8. 3D shape reconstruction of Panoptic studio `dance1`.

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 $\mathtt{F}$ and $\mathbf{f}$. 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 S2 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

[1] Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M Seitz, and Richard Szeliski. Building rome in a day. *Communications of the ACM*, 54(10):105–112, 2011. 1

[2] Farhoosh Alghabi and Mohsen Soryani. Direct computation of epipoles using two pairs of point-line correspondences. In *2008 Second International Symposium on Intelligent Information Technology Application*, volume 3, pages 486–489, 2008. 9

[3] Safar M Asaad and Halgurd S Maghdid. A comprehensive review of indoor/outdoor localization solutions in iot era: Research challenges and future perspectives. *Computer Networks*, page 109041, 2022. 9

[4] Shai Avidan and Amnon Shashua. Novel view synthesis in tensor space. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1034–1040. IEEE, 1997. 1

[5] GN Newsam DQ Huynh MJ Brooks and HP Pan. Recovering unknown focal lengths in self-calibration: An essentially linear algorithm and degenerate configurations. In *Proc. ISPRS-Congress*, volume 31, pages 575–580. Citeseer, 1996. 5

[6] Cesar Cadena, Luca Carlone, Henry Carrillo, Yasir Latif, Davide Scaramuzza, José Neira, Ian Reid, and John J Leonard. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on robotics*, 32(6):1309–1332, 2016. 1

[7] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 7

[8] Huiqin Chen, Emanuel Aldea, and Sylvie Le Hégarat-Mascle. Determining epipole location integrity by multimodal sampling. In *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–8. IEEE, 2019. 9

[9] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. 1

[10] Andrew W Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I. IEEE, 2001. 2

[11] Richard I. Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004. 1, 2, 4

[12] Uwe Helmke, Knut Hüper, Pei Yean Lee, and John Moore. Essential matrix estimation using gauss-newton iterations on a manifold. *International Journal of Computer Vision*, 74(2):117–136, 2007. 5

[13] Berthold KP Horn. Recovering baseline and orientation from essential matrix. *J. Opt. Soc. Am*, 110, 1990. 2

[14] M. Ito, T. Sugimura, and J. Sato. Recovering structures and motions from mutual projection of cameras. In *2002 International Conference on Pattern Recognition*, volume 3, pages 676–679 vol.3, 2002. 1, 5, 6, 9

[15] Fangyuan Jiang, Yubin Kuang, Jan Erik Solem, and Kalle Åström. A minimal solution to relative pose with unknown focal length and radial distortion. In *Asian Conference on Computer Vision*, pages 443–456. Springer, 2014. 1

[16] Hanbyul Joo, Hao Liu, Lei Tan, Lin Gui, Bart Nabbe, Iain Matthews, Takeo Kanade, Shohei Nobuhara, and Yaser Sheikh. Panoptic studio: A massively multiview system for social motion capture. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3334–3342, 2015. 2, 7

[17] Kenichi Kanatani and Chikara Matsunaga. Closed-form expression for focal lengths from the fundamental matrix. In *Proc. 4th Asian Conf. Comput. Vision*, volume 1, pages 128–133. Citeseer, 2000. 5

[18] Yubin Kuang, Jan E Solem, Fredrik Kahl, and Kalle Astrom. Minimal solvers for relative pose with a single unknown radial distortion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 33–40, 2014. 1

[19] Zuzana Kukelova, Martin Bujnak, and Tomas Pajdla. Polynomial eigenvalue solutions to the 5-pt and 6-pt relative pose problems. In *Proceedings of the British Machine Vision Conference*, pages 56.1–56.10. BMVA Press, 2008. doi:10.5244/C.22.56. 1

[20] Zuzana Kukelova, Joe Kileel, Bernd Sturmfels, and Tomas Pajdla. A clever elimination strategy for efficient minimal solvers. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4912–4921, 2017. 2

[21] Viktor Larsson, Kalle Astrom, and Magnus Oskarsson. Efficient solvers for minimal problems by syzygy-based reduction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 820–829, 2017. 1, 5, 6

[22] Jonathan Lawn and Roberto Cipolla. Epipole estimation using affine motion parallax. In *Proceedings of the British Machine Vision Conference*, pages 38.1–38.10. BMVA Press, 1993. doi:10.5244/C.7.38. 9

[23] Hongdong Li. A simple solution to the six-point two-view focal-length problem. In *European Conference on Computer Vision*, pages 200–213. Springer, 2006. 1

[24] Evgeniy Martyushev, Jana Vrablikova, and Tomas Pajdla. Optimizing elimination templates by greedy parameter search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15754–15764, 2022. 5, 6

[25] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European conference on computer vision*, pages 405–421. Springer, 2020. 1

[26] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. Orb-slam: a versatile and accurate monocular slam system. *IEEE transactions on robotics*, 31(5):1147–1163, 2015. 1

[27] David Nistér and Frederik Schaffalitzky. Four points in two or three calibrated views: Theory and practice. *International Journal of Computer Vision*, 67:211–231, 2006. 1

[28] Jacob M Perron, Rui Huang, Jack Thomas, Lingkang Zhang, Ping Tan, and Richard T Vaughan. Orbiting a moving target with multi-robot collaborative visual slam. In *Proceedings of the Workshop on Multi-View Geometry in Robotics (MVIGRO), Rome, Italy*, pages 1339–1344, 2015. 2, 9

[29] Thomas Pollok and Eduardo Monari. A visual slam-based approach for calibration of distributed camera networks. In *2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 429–437, 2016. 2, 9

[30] Jun Sato. Recovering multiple view geometry from mutual projections of multiple cameras. *International Journal of Computer Vision*, 66(2):123–140, 2006. 1, 5, 9

[31] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 1

[32] Leonid Sigal, Alexandru O Balan, and Michael J Black. Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *International journal of computer vision*, 87(1):4–27, 2010. 2, 7

[33] Carsten Steger. Estimating the fundamental matrix under pure translation and radial distortion. *ISPRS Journal of Photogrammetry and Remote Sensing*, 74:202–217, 2012. 5

[34] Henrik Stewénius, David Nistér, Fredrik Kahl, and Frederik Schaffalitzky. A minimal solution for relative pose with unknown focal length. *Image and Vision Computing*, 26(7):871–877, 2008. 1

[35] Changchang Wu. Critical configurations for radial distortion self-calibration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 25–32, 2014. 5

[36] Danping Zou and Ping Tan. Coslam: Collaborative visual slam in dynamic environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):354–366, 2013. 2, 9