Equivalent Constraints for Two-View Geometry: Pose Solution/Pure Rotation Identification and 3D Reconstruction

Qi Cai · Yuanxin Wu · Lilian Zhang · Peike Zhang

Received: 3 May 2017 / Accepted: 20 November 2018 / Published online: 30 November 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract
Two-view relative pose estimation and structure reconstruction is a classical problem in computer vision. The typical methods usually employ the singular value decomposition of the essential matrix to get multiple solutions of the relative pose, from which the right solution is picked out by reconstructing the three-dimension (3D) feature points and imposing the constraint of positive depth. This paper revisits the two-view geometry problem and discovers that the two-view imaging geometry is equivalently governed by a Pair of new Pose-Only (PPO) constraints: the same-side constraint and the intersection constraint. From the perspective of solving equation, the complete pose solutions of the essential matrix are explicitly derived and we rigorously prove that the orientation part of the pose can still be recovered in the case of pure rotation. The PPO constraints are simplified and formulated in the form of inequalities to directly identify the right pose solution with no need of 3D reconstruction and the 3D reconstruction can be analytically achieved from the identified right pose. Furthermore, the intersection inequality also enables a robust criterion for pure rotation identification. Experiment results validate the correctness of analyses and the robustness of the derived pose solution/pure rotation identification and analytical 3D reconstruction.

Keywords Relative pose · Coplanar relationship · Same-side constraint · Intersection constraint · Pure rotation

1 Introduction
It is well known that the relative pose and 3D points can be generally recovered up to a scale from two views. The relationship of the image point pairs in two views is well described by two-view imaging geometry in Fig. 1. For the uncalibrated camera, it is specified as the fundamental matrix whose properties have been studied by Beardsley and Zisserman (1995) and Vieville and Lingraud (1996). Hartley (1992) and Hartley et al. (1992) proposes algorithms for uncalibrated camera pose estimation. For the calibrated camera, the essential matrix was first introduced to the computer vision field by Longuet-Higgens (1981), which provides a way to get the solution to the relative pose. Faugeras and Maybank (1990), Huang and Faugeras (1989), Maybank (1993) and Hartley (1995) have extensively studied the properties of the essential matrix.

The essential matrix is subject to the constraint $EE^T E - Tr(E E^T) E / 2 = 0$ (Faugeras and Maybank 1990), which can be used in degenerate cases with less than eight point pairs. The degenerate cases are solved by Huang and Shim (1988), Maybank (1993), Nister (2004) and Stewénius et al. (2006). Usually the essential matrix linearly estimated using the normalized points would not be the optimum, and there exist nonlinear methods to refine the pose to fulfill the above constraint, as described in Hartley and Zisserman (2003), Horn (1990) and Zhang (1998). When a proper essential matrix is achieved, it can be used to get multiple solutions
to the camera relative pose. The popular method is based on the singular value decomposition (SVD) of the essential matrix, which originated from the proof in Huang and Faugeras (1989). The SVD method was further developed and summarized in Hartley (1995), Hartley and Zisserman (2003) and Wang and Tsui (2000). The solutions are used to reconstruct the 3D points by triangulation and it is only the right solution that yields positive depth (Longuethiggins 1981; Ling et al. 2011). It has been believed that the SVD method requires non-zero translation (Ma et al. 2004; Bazin et al. 2010). Recently, Kneip et al. (2012) puts forward a new nonlinear rotation constraint independent of translation and uses the Grobner basis method to solve the rotation for the 5-point case. That work also raises a new inequality to disambiguate the rotation matrix.

This paper is motivated by answering the following three questions:

1. How to explain the multiple pose solutions to the essential matrix equation from the aspect of solving equation?
2. How to explain the experiment phenomenon that the rotation part can still be recovered from the essential matrix equation in the pure rotation case?
3. Is it possible to directly identify the right pose solution without 3D reconstruction?

The main contribution of the paper is multiple-fold: (1) the two-view imaging geometry is found to be equivalently governed by a Pair of Pose-Only (PPO) constraints: the same-side constraint and the intersection constraint; (2) the basic two-view imaging equation is for the first time formulated as a function of pure pose, independent of 3D feature point coordinates. The relative depth information is totally parameterized by relative pose; (3) the complete pose solutions to the essential matrix equation are explicitly derived, which explains that the orientation can still be recovered in the pure rotation case; (4) aided by two inequalities derived from the same-side and intersection constraints, the right pose solution can be provably identified without resorting to 3D reconstruction. In other words, the pose identification is decoupled from 3D reconstruction, while 3D reconstruction can be analytically achieved from the identified right pose; (5) a robust criterion of pure rotation identification is derived from the intersection inequality.

The paper is organized as follows. Section 2 raises the PPO (same-side and intersection) constraints and proves their equivalency to the basic two-view imaging geometry. Section 3 derives the pose solutions to the essential equation. Section 4 simplifies the PPO constraints yielding the same-side inequality and the intersection inequality to help identify the right pose solution without 3D reconstruction of feature points. A pure rotation identification method is also derived from the intersection inequality. Section 5 reports the experiment results for pose solution/pure rotation identification and analytical reconstruction. Section 6 concludes the paper.

A number of frequently-used equalities of matrices, vectors and products are presented in “Appendix” for easy reference.

2 Two-View Imaging Geometry and Equivalent Constraints

2.1 Two-View Geometry

The two-view imaging geometry is illustrated in Fig. 1. Suppose the point set \( P = \{X_i = (x_i, y_i, 1)^T | i = 1, \ldots, m\} \) and the corresponding point set \( P' = \{X'_i = (x'_i, y'_i, 1)^T | i = 1, \ldots, m\} \) are the sets of normalized image point pairs, while \( P'_w = (X'_w = (x'_w, y'_w, z'_w)^T \) or \( X'_w = (x'_w, y'_w, z'_w)^T | i = 1, \ldots, m\) are the projected 3D world feature points. Throughout the paper, we use the normalized image coordinates, unless explicitly stated otherwise.

The world point coordinates in the two camera frames, centered at \( C \) and \( C' \) respectively, are characterized by

\[
X'_{wi} = RX_{wi} + t
\]  

where \( R \) and \( t \) are respectively the true rotation matrix and translation vector between the two camera frames. It is readily apparent that the two normalized image point sets are related by

\[
z'_w X'_i = z_w R X_i + t
\]  

which can be rewritten as the two-view imaging equation

\[
X'_i = \lambda_i (RX_i + s_i t)
\]  

where the depth-related factors \( \lambda_i \triangleq z_w / z'_w \in \mathbb{R}^+ \) and \( s_i \triangleq 1 / z'_w \in \mathbb{R}^+ \). The equation is important to almost all geometric computer vision problems.

Fig. 1 Sketch of two-view imaging geometry
Left multiplying (3) by \( X_i^T \) removes the depth-related factors (Ma et al. 2004)

\[
0 = X_i^T [t] \cdot X_i' = \lambda_i X_i^T [t] X_i
\]

(4)

which is right the well-known essential matrix equation (Hartley and Zisserman 2003)

\[
X_i^T E X_i = 0
\]

(5)

where \( E = [t] \times \) and \( [t] = \) the skew-symmetric matrix formed by \( t \). Note that the development from (3) to (5) is not reversible, that is to say, some information has been lost as will be shown below. If the translation vector \( t = 0 \), the essential matrix \( E \) reduces to zero or arguably is not well defined, but from the viewpoint of equation, (5) still exists with non-zero solutions, as discussed in next section.

Referring to Fig. 1, (5) depicts the co-planar relationship of the projection rays of the two views for the \( i \)-th point pair \( (X_i, RX_i) \) and the translation vector \( (t) \). Note that it embeds the imaging geometry of two views, but loses the important imaging relationship among the translation vector and two projection rays. For example, the two projection rays always point to the same side of the translation vector (Kneip et al. 2012), as shown in Fig. 1. However, the obtained benefit of the coplanar relationship (5) is that the depth information \( (\lambda_i, s_i) \) is isolated from pose and it is linear in \( E \) which can be exploited to get pose solutions by linear methods like SVD.

Let multiplying (3) by \( t^T \) leads to

\[
0 = \lambda_i t^T X_i' \cdot (RX_i + s_i t) = \lambda_i t^T X_i' \
\Rightarrow t^T (X_i' \times RX_i) = 0
\]

(6)

It means all vectors \( \{X_i' \times RX_i | i = 1, \ldots, m\} \) lie on a plane with the normal vector \( t \). Denote \( B = (X_1' \times RX_1, \ldots, X_m' \times RX_m) \). (6) is equivalent to rank\( (B) \leq 2 \). This is exactly the rotation constraint independent of translation proposed by Kneip et al. (2012).

### 2.2 Equivalent PPO Constraints

Next we will present the PPO constraints governing the physical formation of two overlapped views, namely, the same-side constraint and the intersection constraint (see Fig. 2). Define \( \theta_1, \theta_2 \) and \( \theta_3 \) to denote the angles among the translation vector and the projection rays, and \( \theta_i \in [0, \pi], i = 1, 2, 3 \). Hereafter in this section, the translation vector is divided into the unit-direction part and the magnitude part by \( t = e_i \|t\| \), where the unit vector \( e_i \) is the direction vector of \( t \). Note that for zero translation, \( \|t\| = 0 \) and \( e_i \) is an arbitrary unit vector.

**Same-Side Constraint** The projection rays should point to the same side of the translation vector, that is to say, the vector products of \( t \times RX_i \) and \( t \times X_i' \) are in the same direction.

The same-side constraint can be mathematically represented as

\[
\frac{t \times RX_i}{\|t \times RX_i\|} = \frac{t \times X_i'}{\|t \times X_i'\|} \Rightarrow \frac{e_i \times RX_i}{\|RX_i\|} = \frac{e_i \times X_i'}{\|X_i'\|}
\]

(7)

The first equality is not well-defined for the zero-translation case. For zero translation, \( \theta_3 = 0 \) and \( \theta_3 = 1 \), and as a result (7) reduces to

\[
\frac{e_i \times RX_i}{\|RX_i\|} = \frac{e_i \times X_i'}{\|X_i'\|} \Rightarrow e_i \times \left( \frac{RX_i}{\|RX_i\|} - \frac{X_i'}{\|X_i'\|} \right) = 0
\]

(8)

which means \( RX_i / \|RX_i\| = X_i' / \|X_i'\| \), as \( e_i \) is an arbitrary unit vector. In other words, the projection rays \( X_i' \) and \( RX_i \) are collinear in the same direction, so the second equality of (7) is a constraint for the same direction as well. That is to say, all translation cases are well defined by the second equality of (7).

**Intersection Constraint** The projection rays intersect with each other only if \( \theta_2 = \theta_1 + \theta_3 \) is satisfied.

The three angles in Fig. 2 are obviously related to the vector inner products by

\[
\cos \theta_1 = \frac{X_i^T t}{\|X_i\| \|t\|}, \cos \theta_2 = \frac{(RX_i)^T t}{\|RX_i\| \|t\|}, \cos \theta_3 = \frac{X_i'^T RX_i}{\|X_i'\| \|RX_i\|}
\]

(9)

Note the arccosine function’s value field is \([0, \pi]\). Taking cosine on both sides, the geometric constraint \( \theta_2 = \theta_1 + \theta_3 \) is equivalent to
The same-side constraint can be rewritten as
\[
\frac{(RX_i)^T t}{\|X_i\| \|t\|} = \frac{X_i^T t}{\|X_i\| \|t\|} X_i^T RX_i - \frac{X_i^T \times t}{\|X_i\| \|t\|} \times X_i^T RX_i \geq 0
\]
(11)
The inequality constraint of (10) is equivalent to \(\cos(\pi - \theta_3)\), and we have
\[
\frac{X_i^T t}{\|X_i\| \|t\|} + \frac{X_i^T RX_i}{\|X_i\| \|t\|} \geq 0
\]
(12)
Collectively, the intersection constraint is rewritten as
\[
\begin{align*}
\left( \frac{(RX_i)^T t}{\|X_i\| \|t\|} \right) &= \frac{X_i^T t}{\|X_i\| \|t\|} X_i^T RX_i - \frac{X_i^T \times t}{\|X_i\| \|t\|} \times X_i^T RX_i \\
&\geq 0
\end{align*}
\]
(13)
or alternatively,
\[
\begin{align*}
\left\{ \begin{array}{l}
\|X_i^T e_i\| \|X_i^T RX_i\| = X_i^T RX_i X_i^T e_i - \|X_i^T X_i^T R^T e_i\| \\
\|X_i\| X_i^T e_i + X_i^T RX_i \geq 0
\end{array} \right.
\]
(14)
Similarly for the zero-translation case, (13) is not well-defined. In such a case, however, that the inequality in (14) is always true for any \(e_i\) means \(X_i^T RX_i \geq \max_{t}(-\|X_i\| X_i^T e_i) = \|X_i\| X_i^T \). As \(X_i^T RX_i \leq \|X_i\| \|X_i^T\|\), we can derive that \(X_i^T RX_i = \|X_i\| \|X_i^T\|\). This happens only when the projection rays \(X_i^T\) and \(RX_i\) are collinear in the same direction, i.e., \(\theta_3 = 0\). It shows that the inequality in (14) also gives a constraint for the same direction. We can easily verify that the equality in (14) becomes identity when motion is pure rotation. It is exactly what the intersection constraint means for the zero-translation case. That is to say, all translation cases are well defined by (14).

**Proposition 1** The two-view imaging equation is equivalent to the combination of the same-side and intersection constraints, namely, (7)+(14) \(\Leftrightarrow\) (3).

**Proof of Sufficiency** The same-side constraint can be rewritten as
\[
h \triangleq e_i \times w RX_i = e_i \times X_i^T
\]
(15)
where \(w = \|e_i \times X_i^T\| / \|e_i \times RX_i\|\). Define \(q = h \times e_i\), \(X_i^T\) can be expressed as a linear combination of \(t\) and \(q\)
\[
X_i^T = ae_i + bq
\]
(16)

With (15),
\[
q = h \times e_i = (e_i \times w RX_i) \times e_i = w RX_i - w \left( X_i^T R^T e_i \right) e_i
\]
(17)
From the second line to the third line, the vector product equality for any triple vectors is used (see “Appendix”). Substituting (16) into (15) gives
\[
h = e_i \times (ae_i + bq) = be_i \times q = bh \Rightarrow b = 1
\]
(18)
Considering (16)–(18), we have
\[
X_i^T = ae_i + q
\]
(19)
Contrasting (3) and (19), we will prove that \(w = \lambda_i\) and \((a - w X_i^T R^T e_i) = \lambda_i s_i /\|t\|\).

As for \(w\),
\[
w = \frac{\|e_i \times X_i^T\|}{\|e_i \times RX_i\|} = \frac{z_{wi}}{z_{wi}'} \|e_i \times X_i^T\| = \frac{z_{wi}}{z_{wi}'} = \lambda_i
\]
(20)
where \(\|e_i \times X_i^T\|\) and \(\|e_i \times RX_i\|\) both describe the distance from the 3D point to the translation baseline.

In order to prove \((a - w X_i^T R^T e_i) = \lambda_i s_i /\|t\|\), we must take into account of the intersection constraint. Define the left side of the equality in (14) as \(w_2 \triangleq \|e_i \times X_i^T\| \|X_i^T \times RX_i\|\), Substituting (16) and (18) into the equality in (14), we have
\[
w_2 = ae_i^T RX_i q^T e_i + aq^T RX_i + q^T RX_i q^T e_i - 2ae_i^T q X_i^T R^T e_i - q^T q X_i^T R^T e_i
\]
(21)
Considering (17), we have
\[
q^T e_i = 0
\]
\[
q^T RX_i = w X_i^T X_i - w \left( X_i^T R^T e_i \right)^2
\]
\[
q^T q = w^2 X_i^T X_i - w^2 \left( X_i^T R^T e_i \right)^2
\]
(22)
So, (21) reduces to
\[
w_2 = \left( a - w X_i^T R^T e_i \right) w \left( X_i^T X_i - \left( X_i^T R^T e_i \right)^2 \right)
\]
(23)
Define \(w_3 \triangleq w \left( X_i^T X_i - \left( X_i^T R^T e_i \right)^2 \right)\), then we have
\[
\left( a - w X_i^T R^T e_i \right) = \frac{w_2}{w_3}
\]
(24)
Substituting into (19), we get

\[ X'_i = w \left( RX_i + \frac{w_2}{w_3} e_t \right) \]

(25)
in which

\[
\begin{align*}
\frac{w_2}{w_3} & = \frac{1}{w^2} \left( \| e_t \times X'_i \| \left\| X'_i \times RX_i \right\| \right) \\
& = \frac{1}{w^2} \left( \| X'_i \|^2 \| X_i \| \sin \theta_1 \sin \theta_3 \right)
\end{align*}
\]

(26)

Considering the relationship between normalized image points and the world coordinates of 3D points defined in (1) and (2), the above equation becomes

\[
\begin{align*}
\frac{w_2}{w_3} & = \frac{1}{w^2} \left( \frac{z_{wi}}{z_{wi}^2} \left\| X'_wi \right\|^2 \sin \theta_1 \sin \theta_3 \right) \\
& = \frac{1}{w^2} \left( \frac{z_{wi}}{z_{wi}^2} \| t \| \left\| X'_wi \right\| \sin \theta_3 \right) \\
& = \frac{1}{w^2} \left( \frac{z_{wi}}{z_{wi}^2} \| t \| \left\| e_t \times RX_{wi} \right\| \sin \theta_3 \right)
\end{align*}
\]

(27)

As \( \| X'_wi \| \sin \theta_3 \) and \( \| t \| \sin \theta_3 \) both describe the distance from \( C' \) to the projection ray \( X_i \) and by using (20), it can be further reduced to

\[
\frac{w_2}{w_3} = \left( \frac{z_{wi}'}{z_{wi}} \right)^2 \left( \frac{z_{wi}'}{z_{wi}^2} \right) \| t \| = \| t \| s_j
\]

(28)

Combing (20) and (26), (25) yields

\[ X'_i = \lambda_i(RX_i + s_i \| e_t \| e_t) = \lambda_i(RX_i + s_i t) \]

(29)

**Proof of Necessity** Substitute (3) into the second equality of (7),

\[
\frac{e_t \times X'_i}{\| e_t \times X'_i \|} = \frac{e_t \times \lambda_i(RX_i + s_i t)}{\| e_t \times \lambda_i(RX_i + s_i t) \|} = \frac{e_t \times RX_i}{\| e_t \times RX_i \|}
\]

(30)

Similarly substitute (3) into the equality in (14),

\[
X'_i ^T RX_i X'_i ^T e_t - \left| X'_i \right|^2 e_t = \lambda_i(RX_i + s_i t)^T RX_i \lambda_i(RX_i + s_i t)^T e_t - \| \lambda_i(RX_i + s_i t) \|^2 e_t
\]

(31)

Expanding the above equation yields

\[
\begin{align*}
\lambda_i^2 \left( s_i X'_i ^T X_i t^T e_t - s_i X'_i ^T t X_i ^T e_t \right) \\
& = \lambda_i^2 \| t \| \left( \| X_i \|^2 - \left( X'_i ^T e_t \right)^2 \right) \\
& = \lambda_i^2 \| t \| \left( X'_i \times e_t \right) \| X'_i \times RX_i \| \times RX_i \\
& = \| X'_i \times e_t \| \left( \| X_i \| + e_t ^T RX_i \right) \| X_i \| \times e_t ^T RX_i \}
\end{align*}
\]

(32)

Substitute (3) into the inequality in (14),

\[
\| X_i \| \times X_i ^T e_t + X_i ^T RX_i \\
= \lambda_i(RX_i + s_i t)^T (\| X_i \| e_t + RX_i) \\
= \lambda_i \left[ \| X_i \| \left( \| X_i \| + e_t ^T RX_i \right) + s_i \| t \| \left( \| X_i \| + e_t ^T RX_i \right) \right] \geq 0
\]

(33)

\[ \square \]

**Proposition 2** The same-side constraint implies the co-planar relationship.

**Proof** Left multiply \( \| t \| X'_i ^T \) on both sides of the second equality of (7)

\[
\| t \| X'_i ^T \left( e_t \times RX_i \right) \| e_t \times RX_i \| = \| t \| X'_i ^T \left( e_t \times X'_i \right) \| e_t \times X'_i \|
\]

\[
\Rightarrow X'_i ^T e_t + RX_i = \| t \| \left( \| e_t \times RX_i \| \left( e_t \times X'_i \right) \right) = 0
\]

(34)

\[ \square \]

**Proposition 3** The two-view imaging Eq. (3) can be rewritten as

\[ X'_i = \frac{\| e_t \times X'_i \| \times RX_i}{\| e_t \times RX_i \|} + \frac{\| X'_i \times RX_i \|}{\| e_t \times RX_i \|} e_t
\]

(35)

and the relative ratios of depth

\[
\frac{z_{wi}}{\| t \|} = \frac{\| e_t \times X'_i \|}{\| X'_i \times RX_i \|} \text{ and } \frac{z_{wi}'}{\| t \|} = \frac{\| e_t \times RX_i \|}{\| X'_i \times RX_i \|}
\]

(36)

**Proof** From (20),

\[
\lambda_i = \frac{\| e_t \times X'_i \|}{\| e_t \times RX_i \|}
\]

(37)
According to (20) and (26),

\[
\begin{align*}
    s_i &= \frac{w_2}{\|t\| w_3} = \frac{1}{\|t\|^2} \left\| I_{3 \times 3} t - (X_i^T R_i e_i)^T \right\| \\
    &= \frac{1}{\|t\|^2} \left( \left\| e_i \times X_i^\prime \right\| X_i^\prime \times RX_i \right) \\
    &= \frac{1}{\|t\|^2} \left( \left\| e_i \times X_i^\prime \right\| \right)(RX_i + \frac{1}{\|t\|^2} \left\| e_i \times X_i^\prime \right\| t) \\
    \Rightarrow \|t\| \left\| e_i \times RX_i \right\| X_i^\prime &= \|t\| \left\| e_i \times X_i^\prime \right\| RX_i + t \\
\end{align*}
\]

Comparing the above equation with (2) by coefficients, we have

\[
\begin{align*}
    z_{wi} &= \|t\| \left\| e_i \times X_i^\prime \right\| \\
    z_{wi}^\prime &= \|t\| \left\| e_i \times RX_i \right\| \\
    \end{align*}
\]

\[
\begin{align*}
    z_{wi} &= \|t\| \left\| e_i \times X_i^\prime \right\| \\
    z_{wi}^\prime &= \|t\| \left\| e_i \times RX_i \right\| \\
\end{align*}
\]  

Proposition 4 For all 3D points other than those on the baseline, \( \|X_i^\prime \times RX_i\| = 0 \) if and only if \( z_{wi}/\|t\| = +\infty \) and \( z_{wi}^\prime/\|t\| = +\infty \).

Proof of Sufficiency For the ith 3D point,

\[
\begin{align*}
    \frac{z_{wi}}{\|t\|} &= +\infty \quad \text{and} \quad \frac{z_{wi}^\prime}{\|t\|} = +\infty \\
\end{align*}
\]

Proposition 5 Pure rotation leads to \( \|X_i^\prime \times RX_i\| = 0 \) for all point pairs, and an infinite 3D point or a point on the translation baseline only leads to \( \|X_i^\prime \times RX_i\| = 0 \) for the corresponding point pair.

Proof From Proposition 4, for the case of pure rotation motion, \( \|t\| = 0 \) makes every point pair satisfy \( \|X_i^\prime \times RX_i\| = 0 \). Similarly, for an infinite 3D point, \( z_{wi}/\|t\| = +\infty \) and \( z_{wi}^\prime/\|t\| = +\infty \), the corresponding point pair satisfies \( \|X_i^\prime \times RX_i\| = 0 \). And for a projected point on the baseline, \( X_i^\prime \) and \( RX_i \) are collinear so that \( \|X_i^\prime \times RX_i\| = 0 \).

Proposition 1 shows that the two-view imaging Eq. (3) is equivalent to the PPO constraints. In Sect. 4, we will see that the pair of constraints can be used to pick the right pose solution out of the multiple solutions obtained from (5). In Sect. 5, Proposition 3 will be used to analytically reconstruct the 3D point, and Proposition 4 will be used to discriminate infinite points or pure rotation motion.

3 Complete Pose Solutions to Essential Matrix Equation

The most well-known pose decomposition from the essential matrix is the SVD methods by Hartley (1992) and others. But these methods are actually based on the implicit assumption of \( t \neq 0 \), namely, the essential matrix \( E = \|t\| R \) is well defined, so there exists an incorrect believing that in pure rotation case the essential matrix equation cannot be used to compute the rotation (Ma et al. 2004; Bazin et al. 2010).

This section is devoted to deriving all pose solutions to the essential equation (5) and fundamentally explaining the phenomenon that the orientation can be recovered from the case of pure rotation, which is incorrectly interpreted by Ma et al. (2004) as the result of spurious translation caused by image coordinate noise. As a simplification of the two-view imaging equation, (5) might produce solutions that do not satisfy (3).

We first discuss the essential matrix solutions to the essential matrix equation (5) and then the pose solutions decomposed from the essential matrix solutions.

3.1 Essential Matrix Equation Solutions

In order to reduce the confusion, (5) is re-written as

\[
X_i^T Q X_i = 0 \iff \left( X_i^T \otimes X_i^T \right) \text{vec}(Q) = 0 
\]

where \( \otimes \) denotes the Kronecker product and \( \text{vec}(\cdot) \) is the vectorization of a matrix (see “Appendix”). The popular method to solve \( \text{vec}(Q) \) is the linear methods by restricting \( \text{vec}(Q) \) to have unity length (Hartley and Zisserman 2003), in which the case of pure rotation has not been well considered.

Considering (3), we have

\[
\left( X_i^T \otimes [X_i^T s_i] \right) \left( I_3 \otimes (R t)^T \right) \text{vec}(Q) = 0 \]

in which \( \lambda_i \) is omitted as \( \lambda_i > 0 \). Then for the whole point set, we have
\[
\begin{pmatrix}
X_1^T \otimes [X_1^T s_1] \\
\vdots \\
X_m^T \otimes [X_m^T s_m]
\end{pmatrix}
( I_3 \otimes (R t)^T) \vec{c}(Q) = 0
\] (44)

Let
\[
L \triangleq \begin{pmatrix}
X_1^T \otimes [X_1^T s_1] \\
\vdots \\
X_m^T \otimes [X_m^T s_m]
\end{pmatrix},
y \triangleq ( I_3 \otimes (R t)^T) \vec{c}(Q)
\] (45)

Expanding the matrix \( L \)
\[
L = \begin{pmatrix}
x_1^2 & x_1 y_1 & x_1 s_1 & y_1 s_1 & y_1 & y_1 s_1 & x_1 & 1 & s_1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
x_m^2 & x_m y_m & x_m s_m & y_m s_m & y_m & y_m s_m & x_m & 1 & s_m
\end{pmatrix}
\] (46)

and (44) is equivalent to
\[
Ly = 0
\] (47)

As columns 2, 3 and 7 are respectively equal to 5, 9 and 10, we have \( \text{rank}(L) \leq 9 \). Let \( A = L ( I_3 \otimes (R t)^T) \), we have \( \text{rank}(A) \leq 9 \) and (44) can be written as \( A \vec{c}(Q) = 0 \). Therefore, when \( m \geq 9 \) the homogeneous equation (47) has three linearly independent special solutions
\[
\xi_1 = (0 \ 1 \ 0 \ 0 \ -1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0) ^T, \\
\xi_2 = (0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ -1 \ 0 \ 0) ^T, \\
\xi_3 = (0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ -1 \ 0) ^T
\] (48)

So the solution space of \( y \) is given by
\[
y = ( I_3 \otimes (R t)^T) \vec{c}(Q) = (a_1 \xi_1 + a_2 \xi_2 + a_3 \xi_3)
\] (49)

where \( a_1, a_2, a_3 \) are real numbers. By using the Kronecker product equality (Golub and Van Loan 1996)
\[
( I_3 \otimes (R t)^T) \vec{c}(Q) = \vec{c}((R t)^T Q)
\] (50)

(49) can be described in the form of matrix as
\[
(R t)^T Q = \begin{pmatrix}
0 & -a_1 & -a_2 \\
a_1 & 0 & -a_3 \\
a_2 & a_3 & 0
\end{pmatrix}
\] (51)

Then we get two equations for \( Q \)
\[
Q = R \begin{pmatrix}
0 & -a_1 & -a_2 \\
a_1 & 0 & -a_3 \\
a_2 & a_3 & 0
\end{pmatrix} = R[a]_x
\] (52)

\( t^T Q = 0 \) (53)

where \( a \triangleq (a_3, -a_2, a_1)^T \) and \( a \) can be determined without translation information from (52). Equation (52) indicates that the solutions to (42) are definitely essential matrices.

For \( t = 0 \), it is obvious that \( Q \) is still related to \( R \), but its connection to \( t \) would be gone. It can be derived from (52) that \( ( I_3 \otimes R^T) \vec{c}(Q) = \vec{c}([a]_x) \) which means
\[
||a|| = \sqrt{a_1^2 + a_2^2 + a_3^2} = \frac{\sqrt{2}}{2} ||\vec{c}(Q)||
\] (54)

as the matrix \( I_3 \otimes R^T \) is orthogonal.

### 3.2 Complete Pose Solutions

According to (52), \( Q \) and \([a]_x\) have the same singular values. By SVD, the skew-symmetric matrix can be decomposed as \([a]_x = U_x \Sigma V_x^T\), where \( U_x, V_x \) are orthogonal matrices and \( \Sigma = \text{diag}(||a||, ||a||, 0) \) (Golub and Van Loan 1996). Then \( Q \) can be written as
\[
Q = R[a]_x = RU_x \Sigma V_x^T \triangleq U_Q \Sigma V_Q^T
\] (55)

where \( U_Q = RU_x \) and \( V_Q = V_x \). Note that the SVD decomposition is not unique. Actually, to get \( R \) and \( t \), there is no need to get the exact form of \( V_x \). As \([a]_x\) is a skew-symmetric matrix, one of the singular values is equal to zero and the corresponding singular value vector is \( a \) (Golub and Van Loan 1996). Let \( U_x = (u_1, u_2, u_3) \) and \( V_x = (v_1, v_2, v_3) \), where \( v_3 = \pm a/||a|| \) and \( v_1, v_2 \) are the singular value vectors for the singular value \( ||a|| \). It can be derived from \([a]_x = U_x \Sigma V_x^T\) that
\[
(u_1, u_2) = [a]_x (v_1, v_2) \begin{pmatrix}
||a|| & 0 \\
0 & ||a||
\end{pmatrix}^{-1}
= \frac{1}{||a||} ([a]_x v_1, [a]_x v_2)
= \begin{pmatrix}
a/||a|| \\
||a|| \times v_1, ||a|| \times v_2
\end{pmatrix}
\] (56)

The explicit relationship between \( U_x \) and \( V_x \) is complicated, due to different combinations of the determinant of \( U_x \) and \( V_x \). We next examine it case by case.

If \( \text{det}(V_x) = 1 \),

Case (1). \( v_3 = a/||a|| \) & \( u_3 = v_3 \rightarrow U_x = V_x W_1 \)

Case (2). \( v_3 = -a/||a|| \) & \( u_3 = v_3 \rightarrow U_x = V_x W_2 \)

Case (3). \( v_3 = a/||a|| \) & \( u_3 = -v_3 \rightarrow U_x = V_x W_3 \)

Case (4). \( v_3 = -a/||a|| \) & \( u_3 = -v_3 \rightarrow U_x = V_x W_4 \)

If \( \text{det}(V_x) = -1 \),
Taking (59) into consideration, the translation vector can be known as
\[
\det(U Q V_Q) = \det(U Q W_T^T V_Q^T) = \det(R) = 1
\]
be known as \(R = U Q W_T^T V_Q^T\) and
\[
\det(U Q V_Q) = \det(-R) = -1
\]
which is equivalent to \(E = 0\) up to scale; when \(t = 0\), \(E = 0\) and \(Q\) has nothing to do with \(E\).

**Proof** From (52) and (61), if \(t \neq 0\)
\[
Q = R[a]_x = \pm R \frac{a}{\|a\|} R^T t \quad \text{or} \quad Q = R \frac{a}{\|a\|} R^T t = \pm \|a\| R R^T t = \pm \frac{a}{\|a\|} E
\]

**Proposition 7** For the case \(t = 0\), the right rotation matrix \(R\) can be obtained as one of the solutions; the translation estimate is always wrong.

**Proof** For the case of \(t = 0\), the right rotation matrix can be obtained from (57) because (52) is still valid. However, the translation vector estimate satisfies \(\hat{t} = \pm u_{q3} = \pm Ra / \|a\|\) according to (59). It has nothing to do with the true translation, namely, zero translation. As we know, the norm of \(\hat{t}\) is always equal to 1.

### 4 Identify Right Solution Without 3D Reconstruction

This section will simplify the PPO equality constraints into linear inequality constraints and then use them to identify the right solution from the multiple solutions to the essential equation. Section 3 explores the theoretical pose solutions to the essential matrix equation (5). In applications, an estimate of the essential matrix \(\hat{Q}\) is obtained as the solution of (42). Hereafter we use the hatted symbols to distinguish the estimate from the theoretical solution above. Then by the popular SVD method, four pose solutions

| \(\det(U Q V_Q) = 1\) | \(R = U Q W_T^T V_Q^T\) → \(\det(R) = 1\) | Valid |
| --- | --- | --- |
| \(\det(U Q V_Q) = -1\) | \(R = U Q W_T^T V_Q^T\) → \(\det(R) = -1\) | Invalid |

**Table 1** Validity of rotation matrix

In summary, the number of complete pose solutions to the essential matrix equation (42) is eight, which are now concluded as \(\{R|t|\}\), where
\[
\begin{align*}
R &= R_i |t| = U Q W_T^T V_Q^T, \quad i = 1, 2, 3, 4, \\
\hat{t} &= t_1 \text{ or } t_2 |t_1| = \pm u_{q3} t_2 = \pm u_{q4}.
\end{align*}
\]

They are consistent with those in the previous work of Wang and Tsui (2000). As \(U Q\) and \(V_Q\) are usually obtained by the general SVD, \(\det(U Q V_Q)\) may be \(1\) or \(-1\). As shown in Table 1, if \(\det(U Q V_Q) = 1\), \(R_1\) and \(R_2\) are valid; otherwise, \(R_3\) and \(R_4\) are valid. So in practical code implementation, we can reduce the number of potential pose solutions from eight to four by simply restricting \(\det(U Q V_Q) = 1\), namely, \([R_1 |t_1|, R_2 |t_1|, R_1 |t_2|, R_2 |t_2|\). Hartley and Zisserman (2003) proposed these four solutions by imposing the constraint \(\det(U Q) = \det(V_Q) = 1\) that falls in the scope of \(\det(U Q V_Q) = 1\). Therefore, the essential matrix equation (5) has no less than four pose solutions.

**Proposition 6** For the case \(t \neq 0\), \(Q = \pm E \|a\| / \|t\|\) (or \(Q\) is equivalent to \(E\) up to scale); when \(t = 0\), \(E = 0\) and \(Q\) has nothing to do with \(E\).

**Proof** From (52) and (61), if \(t \neq 0\)
\[
Q = R[a]_x = \pm R \left[ \frac{a}{\|a\|} R^T t \right]_x = \pm \frac{a}{\|a\|} R R^T t = \pm \frac{a}{\|a\|} E
\]
can be obtained by imposing the constraint \( \det \left( U_Q V_Q \right) = 1 \), namely, \( \begin{bmatrix} \hat{R}_1 & \hat{R}_1 \end{bmatrix} \) and \( \begin{bmatrix} \hat{R}_2 & \hat{R}_2 \end{bmatrix} \), where \( \hat{R}_i = U_Q W_i^T V_Q^T \) for \( i = 1, 2 \). \( \hat{Q}_1 = u_{q3} \) and \( \hat{Q}_2 = -u_{q3} \), as in Hartley and Zisserman (2003) and Wang and Tsui (2000).

## 4.1 Same-Side Inequality Constraint

From the same-side constraint (7),

\[
\left( \frac{e_t \times RX_i}{\|e_t \times RX_i\|} \right)^T \frac{e_t \times X'_i}{\|e_t \times X'_i\|} = 1
\]

(64)

It can be reduced to

\[
(e_t \times RX_i)^T (e_t \times X'_i) = \|e_t \times RX_i\| \|e_t \times X'_i\| = X'_i R^T [e_t \times e_t] X'_i > 0
\]

(65)

This is in spirit similar to the inequality to disambiguate the rotation matrix by Kneip et al. (2012), which is nonlinear in the rotation matrix. Note that this inequality has a quite wide threshold, namely \( \|e_t \times RX_i\| \|e_t \times X'_i\| \), which is very helpful in robustly identifying the right solution, as shown in next section.

Considering (52) and (61), \( [e_t \times e_t] = Q Q^T / \|a\|^2 \), so the constraint for a single point is formulated as

\[
m_1(R) \triangleq X'_i R Q Q^T RX_i > 0
\]

(66)

where \( m_1(R) \) is a function of \( R \) when others are known.

It can be checked that the above inequality is able to discriminate the right orientation. Given \( \hat{Q} \) and \( \hat{R} = \hat{R}_1 \) or \( \hat{R}_2 \),

\[
X'_i \hat{Q} \hat{R} X_i = X'_i Q \hat{Q} T \sigma Q \Sigma Q^T W_i^T V_Q T X_i \text{ or } X'_i Q \hat{Q} T \sigma Q \Sigma Q^T W_2^T V_Q T X_i
\]

As \( \Sigma Q \Sigma Q^T W_i^T = -\Sigma Q \Sigma Q^T W_2^T \), we have

\[
m_1(\hat{R}_1) = -m_1(\hat{R}_2)
\]

(68)

i.e.

\[
m_1(\hat{R}_1) m_1(\hat{R}_2) < 0
\]

(69)

The equation shows that the two orientation solutions yield \( m_1(\hat{R}) \) with opposite sign, so it is feasible to get the right orientation by using \( m_1(\hat{R}) > 0 \).

For all point pairs, the same-side linear inequality constraint can be collectively written as

\[
M_1(\hat{R}) \triangleq \begin{pmatrix} X'_1 \otimes X'^T_1 \\ \vdots \\ X'_m \otimes X'^T_m \end{pmatrix} \text{vec} \left( \hat{Q} \hat{Q}^T \hat{R} \right) > 0
\]

(70)

which can be used to select the right rotation matrix.

### 4.2 Intersection Inequality Constraint

Since \( \theta_2 = \theta_1 + \theta_3 \) and \( \theta_i \in [0, \pi] \), we have \( \pi \geq \theta_2 > \theta_1 \geq 0 \). It means

\[
\cos^{-1} \left( \frac{RX_i}{\|RX_i\|} \|t\| \right) > \cos^{-1} \left( \frac{X'_i}{\|X'_i\|} \|t\| \right)
\]

i.e.

\[
\frac{X'_i^T e_t}{\|X'_i\|} > \frac{X'^T_i R^T e_t}{\|X'_i\|}
\]

(72)

It can be written as

\[
m_2(R, t) \triangleq \left( \frac{\|X'_i\|}{\|X'^T_i\|} \frac{\|X_i\|}{\|X'^T_i\|} \right) \left( \begin{array}{cc} -R^T & \hat{t} \end{array} \right) > 0
\]

(73)

For the translation solutions in Sect. 3, we can see that \( m_2(R, t) \) is effective as \( m_2(\hat{R}, \hat{t}_1) m_2(\hat{R}, \hat{t}_2) < 0 \). Actually, the sign of \( m_2(R, t) \) is independent of the right or wrong \( R \). Let \( \hat{t}_1 = u_{q3} \), for \( \hat{R}_1 \) and \( \hat{R}_2 \), we have

\[
\left( -\hat{R}_1^T I_3 \right) \hat{t}_1 = \left( -U_Q W_1^T V_Q^T \right) u_{q3} = \left( -V_Q W_1 U_Q^T u_{q3} \right)
\]

\[= \left( -V_Q W_1 \begin{pmatrix} 0 \\ 0 \end{pmatrix} u_{q3} \right)
\]

(74)

and

\[
\left( -\hat{R}_2^T I_3 \right) \hat{t}_1 = \left( -V_Q W_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} u_{q3} \right)
\]

(75)

So \( m_2(\hat{R}_1, \hat{t}_1) = m_2(\hat{R}_2, \hat{t}_1) \) when \( \hat{t}_1 = u_{q3} \) (or \( \hat{t}_2 = -u_{q3} \)). In conclusion, the intersection linear inequality constraint \( m_2(R, t) \) can be described as \( m_2(t) > 0 \).

For all point pairs, the intersection inequality constraint can be collectively written as

\[
M_2(\hat{t}) \triangleq \begin{pmatrix} \|X'_1\| \|X'^T_1\| \|X_1\| \|X'^T_1\| \\ \vdots \\ \|X'_m\| \|X'^T_m\| \|X_m\| \|X'^T_m\| \end{pmatrix} \left( \begin{array}{cc} -\hat{t}^T & \hat{t} \end{array} \right) > 0
\]

(76)

which can be used to select the right translation vector.

In contrast to the robustness in identifying the rotation matrix by \( M_1(\hat{R}) > 0 \), finding the right translation vector by \( M_2(\hat{t}) > 0 \) in the case of large
depth/translation ratio should be cautious. In these cases, according to Proposition 4, \( \|X'_i \times RX_i\| \approx 0 \) and \( X'_i \approx RX_i \).\( e_t \times RX_i \| = \lambda_i RX_i \) from (35). \( m_2(t) \) might be very near to zero because
\[
m_2(t) = \left( -\|X'_i\| \|X'_i\|^T + \|X_i\|X_i^T \right) e_t \\
\approx \left( -\lambda_i RX_i \|X'_i\|^T + \|X_i\| (\lambda_i RX_i)^T \right) e_t \\
= 0
\]

(77)

This might cause some problem in translation identification in the case of nearly pure rotation. However, translation estimation under this case is of little significance. In this aspect, it would be very helpful for judging the significance of the obtained translation vector if the pure rotation case could be successfully identified.

4.3 Robust Relative Pose Algorithm Without 3D Reconstruction

Table 2 summarizes the algorithm of linear solutions of the essential matrix and how to robustly identify the right solution using the proposed two inequality constraints.

Due to the effect of point matching error, there might exist some point pairs that violate the constraints. To improve the robustness against such errors, we could choose the solution as the right one with the maximum number of point pairs that satisfy \( M_1(\hat{R}) > 0 \) and \( M_2(\hat{i}) > 0 \).

4.4 Relative Rotation Identification

Propositions 4 and 5 prove the property of pure rotation and its difference from infinite points and baseline points (i.e., points on the translation baseline). In this section, \( R \) denotes the right rotation matrix identified by \( M_1(\hat{R}) > 0 \). To identify pure rotation, we might manually set a threshold on \( X'_i \times RX_i \) for all point pairs. In other words, we can identify the pure rotation motion by
\[
m_3(R) \triangleq \frac{\|X'_i \times RX_i\|}{\|X'_i\| \|X_i\|}, \quad M_3(R) \triangleq \text{avg} \left( \frac{\|X'_i \times RX_i\|}{\|X'_i\| \|X_i\|} \right) < \delta_0
\]

where \( \delta_0 \) is a threshold and \( \text{avg}(\cdot) \) is the average function. We identify the camera motion as pure rotation if \( M_3(R) \) is greater than some-prescribed confidence level, e.g., 0.95. There exists a problem that \( \delta_0 \) needs to be manually tuned under different cases.

Proposition 8 For all 3D points other than those on the baseline, \( \|X'_i \times RX_i\| = 0 \) if and only if \( m_2(t) = 0 \).

Proof of Necessity If \( \|X'_i \times RX_i\| = 0 \), according to Proposition 4 and referring to the development in (77), \( X'_i = RX_i \).\( e_t \times RX_i \| = \lambda_i RX_i \) from (35). Then we have \( m_2(t) = 0 \).

Proof of Sufficiency If \( m_2(t) = 0 \), referring to Fig. 2, we have
\[
m_2(t) = -\|X'_i\| \|X'_i\|^T e_t + \|X_i\|X_i^T e_t = 0 \\
\Rightarrow \|X_i\| \|X'_i\| \cos \theta_1 - \|X_i\| \|X'_i\| \cos \theta_2 = 0 \\
\Rightarrow \cos \theta_1 = \cos \theta_2
\]

(79)

which means \( \theta_1 = \theta_2 \) as \( \theta_1, \theta_2 \in [0, \pi] \) and the cosine function is monotonic. From the intersection constraint, the angle \( \theta_3 \) must be zero, namely, \( X'_i \times RX_i = 0 \).

Note that the normal scenes will not contain infinite points and translation baseline points only. According to Propositions 4, 5 and 8 actually tells that \( m_2(t) = 0 \) could be alternatively used as a criterion for identifying pure rotation.

5 Test Results

In our simulation tests, the 3D feature points are centered at known coordinates and subjected to uniform distributions, which are mapped to two views to generate image pair coordinates in pixels. For example, we use \( U((0, 0, d), (30, 30, d)) \) to denote the uniform distribution in a cube that is centered at \( [0 0 d]^T \) and of length 30, 30 and \( d \) (in meters) along X, Y and Z axes, respectively. The extrinsic parameter matrix of the left view is constantly set to \( (I_3 | 0_{3 \times 1}) \) and the intrinsic parameter matrix of two cameras is
\[
\begin{pmatrix}
0 & 0 & 512 \\
0 & 800 & 512 \\
0 & 0 & 1
\end{pmatrix}
\]

Springer
We generate \( N_{pts} \) 3D points in \( U((0, 0, d), (30, 30, d)) \). To get \( N_s \) poses of the right view, we generate the rotation matrices of which the corresponding Euler angles obey \( N(\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T, diag[5^2, 5^2, 20^2]) \) in degrees, and the unit direction vector of translation in the OXY plane. Each translation length is given by \( \|t\| = \alpha \|t\|_{\text{max}} \), where the parallax factor \( \alpha \in [10^{-3}, 1] \) and \( \|t\|_{\text{max}} = 4.2 \) denotes the maximum length of translation. Figure 3 demonstrates 3D points and the poses of right views for \( d = 30 \), \( N_{pts} = 50 \) and \( N_s = 10 \).

Image points are obtained by projecting 3D points onto the pair of views and the Gaussian noise is added to the generated image pair coordinates, with the standard deviation varying between 0.1 and 5 pixels for \( N_{std} \) times. \( N_{mc} \) Monte Carlo runs are carried out for each case. In the following tests, we set \( (N_s, N_{pts}, N_{std}, N_{mc}) = (200, 50, 200, 30) \) and \( d = 30 \) m unless explicitly stated otherwise.

The experimental phenomenon that the rotation part can still be recovered from the essential matrix equation in the pure rotation case has been reported in very few literature, e.g., Ma et al. (2004). Table 3 lists the pose decomposition results in a case of pure rotation. As predicted in Sect. 4, we see that the rotation matrix can be still recovered, but the translation vector is wrong.

### 5.1 Pose Solution Identification Result

We define the discrepancy \( \varepsilon_{i,j}^R \) between the identified rotation matrix and the true rotation matrix as

\[
\varepsilon_{i,j}^R = \frac{1}{3} \|dcm2angle(R_i^T \hat{R}_{i,j})\|_{L1}
\]

(80)

where \( R_i \) is the true rotation of the \( i \)-th two-view and \( \hat{R}_{i,j} \) is the identified rotation matrix in the \( j \)-th Monte Carlo run.

\( dcm2angle() \) means the Euler angle vector derived from the rotation matrix, and \( \|\cdot\|_{L1} \) is the 1-norm of a vector.

Similarly, define the discrepancy \( \varepsilon_{i,j}^t \) between the identified translation vector and the true translation by

\[
\varepsilon_{i,j}^t = \arccos\left( \frac{\hat{t}_{i,j}^T t_i}{\|\hat{t}_{i,j}\| \|t_i\|} \right)
\]

(81)

Figure 4 plots the rotation and translation discrepancies (RMSE), averaged across Monte Carlo runs, as the function of noise standard deviation and the parallax factor \( \alpha \). The result in Fig. 4a accords with Proposition 7 that we can still get good rotation estimates even when \( \alpha \) approaches zero. Actually, the rotation identification is always right. However, as shown in Fig. 4b, the estimated translation vector is susceptible to the noise standard deviation or the parallax factor. The translation discrepancy even reaches 90° at the right-bottom triangular area where the noise standard deviation is large or the parallax factor is small. This observation shows a fact that the translation estimate obtained from the essential equation is unsatisfactory in accuracy.

In order to investigate the robustness of the pose identification method proposed in Sect. 4, we compare it with that of the traditional method by

\[
\varepsilon_{diff}^t = \varepsilon_{rad}^t - \varepsilon_{new}^t
\]

(82)

where \( \varepsilon_{rad}^t \) and \( \varepsilon_{new}^t \) are translation discrepancies for the traditional method (Hartley and Zisserman 2003) and the new method, respectively, and \( \varepsilon_{diff}^t \) is the difference of the two translation discrepancies, as shown in Fig. 5. In order to get a more intuitive result of relative advantage, we argue that the average results of all parallax factors for a definite noise standard deviation and vice versa, as shown in Fig. 6a, b for \( d = 30 \) m, 50 m and 80 m. If we regard Fig. 5 as a matrix, they respectively correspond to column averaging and row averaging. In Fig. 6, the translation discrepancy differences are almost all above zero, which shows that the new identification method performs more robustly. As shown in Fig. 6a, the larger noise standard deviation becomes, the better robustness.
ness is. In Fig. 6b, the trend is similar for small parallax factors, and along with further increased parallax factor, the translation discrepancy difference turns to decrease and stops just above zero (see $d=30$ m for example).

### 5.2 Analytical Reconstruction Result

According to Proposition 3, the relative depth has been expressed in terms of the pose. It actually gives a new method to analytically reconstruct 3D points. There are two depths existing in (36), one for the left view and the other for the right view. The difference between the two depths is negligible and they are averaged as the depth estimate of the new method.

We compare the new method with the traditional 3D point reconstruction method by the DLT-based linear triangulation (Hartley and Zisserman 2003). The depth of both methods is computed using the identified solution in Sect. 5.1.

The reconstruction error ratio is defined as

$$\text{ratio}^{3D} = \frac{\varepsilon_{3D}^{\text{anal}}}{\varepsilon_{3D}^{\text{trad}}}$$

where $\varepsilon_{3D}^{\text{trad}}$ and $\varepsilon_{3D}^{\text{anal}}$ are respectively the 3D point average reconstruction error (RMSE across Monte Carlo runs) of the analytical method and the traditional method. Figure 7 presents the reconstruction error ratio as a function of noise standard deviation and parallax scale. For normal cases with large parallax, the two reconstruction methods perform almost identically, but for small parallax and large noise standard deviation, $\text{ratio}^{3D}$ tends to be much less than 1, which means that the analytical method is much better in reconstruction accuracy.

Real tests are carried out to confirm the above reconstruction advantage. Two 2D barcode papers are placed onto mutually-perpendicular faces of a table. Pictures are taken with seven selected poses so that the two barcode pictures are well observed (see Fig. 8). Figure 8a gives an example picture of the table and barcode papers, in which 62 feature points are observed, and Fig. 8b depicts the seven camera poses relative to the barcode papers.

We use the PnP method (Garro et al. 2012; Ferraz et al. 2014) to compute the true camera poses and the true 3D points. Specifically, a separate spatial frame is attached to each barcode paper and the PnP method is used to compute the camera poses relative to each barcode paper frame. The ground-truth camera poses and 3D barcode feature point coordinates are obtained by unifying the two spatial frames. The reconstruction error ratios for all $C_2^7 = 21$ combinations (blue circles) are plotted in Fig. 9. Except the 3–4 pair of
Fig. 6 a Translation discrepancy difference averaged across all parallax factors. b Translation discrepancy difference averaged across all noise standard deviations.

Fig. 7 Reconstruction error ratio of analytical method relative to traditional method, as a function of parallax factor and noise standard deviation.

Fig. 8 a Test table and barcode papers with feature points. b Ground-truth camera poses and 3D structure in real tests.

Fig. 9 Reconstruction error ratios for 21 two-view combinations in real tests.

views, the analytical reconstruction method is overwhelmingly better in reconstruction accuracy.
Considering that the initial pose solution may likely be too coarse to suitable for reconstruction, the bundle adjustment (BA) (Hartley and Zisserman 2003) is used to optimize 3D points and the pose. The BA is respectively initialized by the analytical reconstruction result and the traditional reconstruction result. The reconstruction error ratios after the BA are also presented in red circles in Fig. 9. It shows that the BA does not significantly change the reconstruction error ratio, in other words, the analytical reconstruction method is still much preferred.

Two demonstrating reconstructed barcode papers together with the truth are given in Fig. 10, one is the 3–4 pair of views with the largest parallax, the other one 4–6 pair of views with the second largest parallax. We can see that the two methods perform similarly for the largest parallax case in Fig. 10a, but the analytical reconstruction method is much better for the second largest parallax case in Fig. 10b. For 3–4 and 4–6 view pairs, the analytically-initialized BA and traditionally-initialized BA are almost identically good (see Fig. 9), so only the analytically-initialized BA is plotted in Fig. 10. In addition, Fig. 10c, d give the reconstructed barcode papers
5.3 Pure Rotation Identification Result

Propositions 4, 5 and 8 actually present two alternative methods to identify pure rotation or infinite 3D points. In Sect. 4.4, we have compared them for identifying the pure rotation and recommended the method derived from Proposition 8.

Define a pure rotation indicator (PRI) as the average of $M_2(\hat{t})$ in (76) across all feature points

$$PRI(\hat{t}) = \frac{1}{|\mathcal{F}|} \sum_{f \in \mathcal{F}} \frac{M_2(\hat{t})}{m_2(\hat{t})}$$

and identify the motion as a pure rotation if

$$PRI(\hat{t}) < \delta_{pri}$$

where $\delta_{pri}$ is a threshold. Actually, it is ready to check $|m_2(\hat{t}_1)| = |m_2(\hat{t}_2)|$, so we simply denote it as PRI.

Referring to Fig. 4b, for the case of pure rotation ($\alpha = 0$) and zero noise, the translation discrepancy is roughly $90^\circ$. For a fixed nonzero parallax factor, e.g., $\alpha = 0.2$, as the noise standard deviation increases the translation discrepancy also

Fig. 11 a PRI as a function of parallax factor and noise standard deviation. b $M_3(R)$ as a function of parallax factor and noise standard deviation

Fig. 12 a Reconstruction errors by analytical method and analytically-initialized BA for 21 two-view combinations in ascending order of PRI. b Reconstruction errors by analytical method and analytically-initialized BA for 21 two-view combinations in ascending order of $M_3(R)$ for the 1–4 pair of views. The analytical reconstruction initialization leads to a well-behaved BA in Fig. 10c, while the traditional reconstruction initialization leads to a diverging BA in Fig. 10d. This can be confirmed by the reconstruction error ratio in Fig. 9, almost zero for the 1–4 view pair.
gradually approaches $90^\circ$. The cases of about $90^\circ$ translation discrepancy roughly forms a triangle at the lower-right corner. Translation estimate with so large discrepancy is usually of little use, so it is preferable to identify all of them by a chosen threshold $\delta_{pri}$.

Figure 11 plots the PRI and $M_3(R)$ with the same simulation settings as Fig. 4. Referring to Fig. 11a, for the case of pure translation and zero noise, PRI is approximately zero as predicted by Proposition 8. For a fixed nonzero parallax factor, e.g., $\alpha = 0.2$, increasing noise standard deviation tends to reduce the PRI to zero. Notably, the color pattern of PRI much resembles that of the translation discrepancy in Fig. 4b. This enables us to easily find a threshold to identify the cases with large translation discrepancy. For instance, $\delta_{pri} = 0.015$ corresponding to the magenta color is a good choice to identify the cases with $90^\circ$ translation discrepancy. In contrast, the $M_3(R)$ in Fig. 11b has no such a nice property, namely, we cannot find a proper threshold to do the identification. We have tested 3D points with different depths ($d = 30$ m, 50 m and 80 m) and found that $\delta_{pri} = 0.015$ is still a good threshold for all of these scenarios.

Figure 12a, b plot the reconstruction errors by the analytical method and the analytically-initialized BA for 21 two-view combinations, respectively in the ascending order of PRI and $M_3(R)$. The PRI has a much sharper rising than $M_3(R)$. Larger PRI predicts those view pairs (1–4, 4–6, 4–7, 3–4) that have smaller initial reconstruction errors and converged BA results. In comparison, $M_3(R)$ cannot effectively pick the 1–4 view pair out using a threshold, because those view pairs (2–4, 2–3, 4–5, 3–6, 3–7) to the right have diverged BA results.

5.4 Computation Cost

Time costs as a function of the number of feature points are shown in Fig. 13. The elapsed time are averaged across 20 Monte Carlo runs. ‘Method I’ means the two inequalities are employed to identify the right pose solution, and ‘Method II’ additionally uses the right pose from ‘Method I’ to analytically reconstruct 3D points. In contrast, the traditional method identifies the right solution by the DLT-based 3D reconstruction and subsequent positive depth check for each of four solutions (Hartley and Zisserman 2003).

Note that the time cost of computing the feature pairs is not taken into account. The matrix $\left( X_1 \otimes X'_1, \ldots, X_m \otimes X'_m \right)^T$ in $M_1(R)$ uses the same one as in solving the matrix $Q$ in (42). Figure 13 clearly shows that Methods I and II have almost the same time cost as the analytical reconstruction is very cost-efficient. As compared with the traditional method, they reduce the time cost by about 7–9 times.

6 Conclusions

The well-known essential matrix equation only represents the coplanar relationship in the two-view imaging geometry and loses the important connection among the translation vector and two projection rays. The paper comes up with the PPO (same-side and intersection) constraints for the two-view geometry problem that are proven to be equivalent to the two-view imaging geometry. The complete pose solutions to the essential matrix equation are explicitly derived from the perspective of equation solving. It is shown that the
orientation can still be recovered in the pure rotation case. Two inequalities are formulated by simplifying the two new constraints, so as to help directly identify the right solution. It does not need 3D points reconstruction and depth check that are required in traditional methods. The intersection inequality constraint lends itself to a criterion for identifying the pure rotation motion of the camera.

The relationship among the problems discussed in this paper is summarized in Fig. 14 for easy reference. Test results demonstrate the usefulness of the PPO constraints in robustly identifying the right pose solution and pure rotation, as well as in 3D point reconstruction.

Acknowledgements Thanks to anonymous reviewers for their constructive comments and Dr. Danping Zou for group talks. The work is funded by National Natural Science Foundation of China (61422311, 61673263, 61503403) and Hunan Provincial Natural Science Foundation of China (2015JJ1021).

Appendix

Cross Product The cross product of two vectors $a$ and $b$ is defined only in three-dimensional space and is denoted by $a \times b$. The vector triple product is the cross product of a vector with the result of another cross product, and is related to the dot product by the following formula

$$ (a \times b) \times c = (a^T c) b - (b^T c) a $$

and

$$ a \times (b \times c) = (a^T c) b - (a^T b) c $$

Kronecker Product If $A$ is an $m \times n$ matrix and $B$ is a $p \times q$ matrix, then the Kronecker product $A \otimes B$ is the $mp \times nq$ block matrix:

$$ A \otimes B = \begin{bmatrix} a_{11} B & \cdots & a_{1n} B \\ \vdots & \ddots & \vdots \\ a_{m1} B & \cdots & a_{mn} B \end{bmatrix} $$

Vectorization the vectorization of an $m \times n$ matrix $A$, denoted $vec(A)$, is the $mn \times 1$ column vector obtained by stacking the columns of the matrix $A$ on top of one another:

$$ vec(A) = [a_1, \ldots, a_{m1}, a_{12}, \ldots, a_{m2}, \ldots, a_{1n}, \ldots, a_{mn}]^T $$

Here, $a_{ij}$ represents the element of row-i and column-j. The vectorization is frequently used together with the Kronecker product to express matrix multiplication as a linear transformation on matrices. In particular,

$$ vec(ABC) = \left(c^T \otimes A\right) vec(B) $$

for matrices $A$, $B$, and $C$ of dimensions $k \times l$, $l \times m$, and $m \times n$.

References

Bazin, J. C., Demonceaux, C., Vasseur, P., & Kweon, I. S. (2010). Motion estimation by decoupling rotation and translation in catadioptric vision. Computer Vision and Image Understanding, 114, 254–273.

Beardsley, P. A., & Zisserman, A. (1995). Affine calibration of mobile vehicles. In Europe-China Workshop on Geometrical Modelling and Invariants for Computer Vision, pp. 214–221.

Faugeras, O. D., & Maybank, S. (1990). Motion from point matches: Multiplicity of solutions. International Journal of Computer Vision, 4, 225–246.

Ferraz, L., Binefa, X., & Moreno-Noguer, F. (2014). Very fast solution to the PnP problem with algebraic outlier rejection. In Computer Vision and Pattern Recognition, pp. 501–508.

Garro, V., Crosilla, F., & Fusiello, A. (2012). Solving the PnP problem with anisotropic orthogonal proructs analysis. In Second International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, pp. 262–269.

Golub, G. H., & Van Loan, C. F. (1996) Matrix computations (3rd ed.). Johns Hopkins University Press.

Hartley, R. I. (1992). Estimation of relative camera positions for uncalibrated cameras. Presented at the ECCV '92 Proceedings of the Second European Conference on Computer Vision.

Hartley, R. I. (1995). An investigation of the essential matrix. Report Ge.

Hartley, R., Gupta, R., & Chang, T. (1992). Stereo from uncalibrated cameras. In Proceedings CVPR '92. 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 761–764.

Hartley, R., & Zisserman, A. (2003). Multiple view geometry in computer vision. Cambridge: Cambridge University Press.

Horn, B. K. P. (1990). Relative orientation. International Journal of Computer Vision, 4, 59–78.
Huang, T. S., & Faugeras, O. D. (1989). Some properties of the E matrix in two-view motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 11*, 1310–1312.

Huang, T. S. & Shim, Y. S. (1988). Linear algorithm for motion estimation: How to handle degenerate cases. In *International Conference on Pattern Recognition*, pp. 439–447.

Kneip, L., Siegwart, R., & Pollefeys, M. (2012). Finding the exact rotation between two images independently of the translation. In *European Conference on Computer Vision*, pp. 696–709.

Ling, L., Cheng, E., & Burnett, I. S. (2011). Eight solutions of the essential matrix for continuous camera motion tracking in video augmented reality. In *Proceedings of the 2011 IEEE International Conference on Multimedia and Expo (ICME 2011)*, Barcelona, Spain, pp. 1–6.

Longuet-Higgins, H. C. (1981). A computer algorithm for reconstructing a scene from two projections. *Nature, 293*, 133–135.

Ma, Y., Soatto, S., Kosecka, J., & Sastry, S. S. (2004). *An invitation to 3-D vision*. Berlin: Springer.

Maybank, S. (1993). *Theory of reconstruction from image motion* (Vol. 28). Berlin: Springer.

Nister, D. (2004). An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 26*, 756–777.

Stewénius, H., Engels, C., & Nistér, D. (2006). Recent developments on direct relative orientation. *ISPRS Journal of Photogrammetry & Remote Sensing, 60*, 284–294.

Vieville, T., & Lingrand, D. (1996). Using singular displacements for uncalibrated monocular visual systems. In *European Conference on Computer Vision*, pp. 207–216.

Wang, W., & Tsui, H. T. (2000). An SVD decomposition of essential matrix with eight solutions for the relative positions of two perspective cameras. In *International Conference on Pattern Recognition*, Vol. 1, pp. 362–365.

Zhang, Z. (1998). Determining the epipolar geometry and its uncertainty: A review. *International Journal of Computer Vision, 27*, 161–195.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.