Int J Comput Vis (2008) 79: 209–223
DOI 10.1007/s11263-007-0114-4

# Detecting and Handling Unreliable Points for Camera Parameter Estimation

**Yihong Wu · Youfu Li · Zhanyi Hu**

Received: 19 December 2006 / Accepted: 21 November 2007 / Published online: 8 December 2007


**Abstract** The popularly used DLT method sometimes fails to give reliable camera parameter estimation. It is therefore important to detect the unreliability and provide the corresponding solutions. Based on a complete framework of invariance for six points, we construct two evaluation functions to detect the unreliability. The two evaluation functions do not involve any computations for the camera projective matrix or optical center and thus are efficient to perform the detection. Then, the guidelines corresponding to the different detection results are presented. In particular, a filtering RANSAC method to remove the detected unreliable points is provided. The filtering RANSAC proves to be successful in removing the unreliable points even if these points are of a large proportion.

**Keywords** Bracket algebra · Invariant · Camera calibration

## 1 Introduction

Camera parameter estimation is a key problem in many fields such as 3D reconstruction, augmented reality, visual surveillance, and industrial photogrammetry. A very popular way to achieve the estimation is by the Direct-Linear-Transformation (DLT) method which uses at least six pairs of space points and their image points (Abdel-Aziz and Karara 1971; Slama 1980; Sutherland 1963; Hartley and Zisserman 2000). The advantage of this method is that the camera parameters can be determined linearly, making the result accurate in general.

However, sometimes the expected accuracy of the estimated camera parameters cannot be obtained by using either an optimized DLT method or a non-optimized DLT method. The accuracy of the camera parameters will severely affect the implementation result of a subsequent vision task. For example, the reconstructed shape of an object may be deformed, the virtual object and the real object cannot be coincident, or the measurement will have low accuracy. Therefore, detecting the unreliability of space points and their image points for camera parameter estimation and giving the corresponding solutions are extremely important.

The cause for the unreliability can be traced to the following two cases:

(i) Camera and space points lie on a degenerate configuration.

If a spatial configuration is degenerate mathematically but the noise from the measured image makes it non-degenerate, any estimation under such configuration would be useless (Weng et al. 1989). Notice that by using the DLT method, many degenerate configurations may occur. There are systematic analyses for these degenerate configurations in Chap. 21 of (Hartley and Zisserman 2000) as two cases: incidence case and non-incidence case. In the incidence case, some of the space points are collinear or coplanar, or they are collinear or coplanar with the camera optical center, which can be detected by checking the linearly dependent relations among the space points or the image points. In the non-incidence case, the space points and the camera optical

Y. Wu (✉) · Y. Li
Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong
e-mail: yhwu@nlpr.ia.ac.cn

Y. Wu · Z. Hu
National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, P.O. Box 2728, Beijing 100080, China

center lie on a proper twisted cubic, on which no three points are collinear and no four points are coplanar. It is difficult to detect this non-incidence case. The method of (Wu and Hu 2003) can detect this degenerate configuration. However, it requires estimation of the camera optical center which is sensitive to noise. In (Wu and Hu 2006), a robust method is presented but needs to fix some point order such that using this method is not convenient. Determining the rank of the coefficient matrix of the linear equations on the camera projective matrix may also be a way to detect this non-incidence degenerate configuration. Unfortunately, as pointed out in the discussion section of (Kahl and Henrion 2007), this is a difficult task to preset a threshold to discriminate the degeneracy from the non-degeneracy.

(ii) The space points and their image points are inconsistent with a camera projective matrix due to large noise in some of the image or space points, or due to mismatching between space and image points.

Mismatching and overlarge noise among the used space points and their image points can occur in practice. If a vision task has been performed and is found to be unsatisfactory, then people need to intervene to check every step of their algorithm. This will involve checking whether there is overlarge noise or mismatching of the points. This checking is tedious. Automatically detecting and removing these points can make a system more usable and efficient. In (Förstner 1987), the author studied the influence of overlarge noise of the input data on the reliability of a linear system, where the analysis is dependent on the coordinate system under which the input data are taken. RANSAC idea may remove the mismatching and overlarge noise of the points. However, whether the initial samples are consistent with a camera projective matrix or not, RANSAC would always choose them to start the estimation. This will lead to unreliable estimation for camera parameters, in particular when the inconsistent points are of a large proportion.

The above two reasons are the unreliability of using DLT method. In this work, we develop the detection method for the unreliability by two constructed evaluation functions. Then, the corresponding solutions for the different detection results are provided. In particular, a filtering RANSAC method is proposed to remove the points with overlarge noise or mismatched. This filtering RANSAC method can also prevent the remaining points from lying on a degenerate configuration. The advantages of our methods include: (i) no estimation of the camera projective matrix or optical center in the detection is needed; (ii) the error distribution of measurements on the space points or image points does not need to be known; (iii) the detection is independent of the coordinate system, which is simple and compact, requiring only comparison of the values of the evaluation functions with given thresholds; (iv) the initial samples with poor performance are detected and filtered out in the proposed filtering RANSAC before a general RANSAC method is employed. Experiments performed show that the detection method is efficient and the filtering RANSAC has a high success rate in removing the unreliable points. Comparison between the filtering RANSAC and the general RANSAC is also made, showing the importance of the filtering.

These advantages are attributed to the two constructed evaluation functions based on the invariant relationship of six space points and their image points. Thanks to the unified and complete framework of invariance on six points (Wu and Hu 2005), the two evaluation functions are established for identifying whether or not the camera and space points lie on a twisted cubic degenerate configuration and whether or not there is overlarge noise or mismatching among the used points.

The organization of the paper is as follows. Some preliminaries are listed in Sect. 2. The two evaluation functions for six points are constructed in Sect. 3. Section 4 gives the detection method of the unreliability for six or more than six points and provides the corresponding solutions to the different detection results. In Sect. 5, we report the experiments, followed are conclusions in Sect. 6.

## 2 Preliminaries

In this paper, a bold capital letter denotes either a homogeneous 4-vector or a matrix, a bold small letter denotes a homogeneous 3-vector, a bracket "[ ]" denotes the determinant of vectors in it.

There exist relations among determinants. One kind of these relations are the following Grassmann-Plücker relations:

$$[\mathbf{a}_1\mathbf{a}_2\mathbf{a}_5][\mathbf{a}_3\mathbf{a}_4\mathbf{a}_5] - [\mathbf{a}_1\mathbf{a}_3\mathbf{a}_5][\mathbf{a}_2\mathbf{a}_4\mathbf{a}_5] + [\mathbf{a}_1\mathbf{a}_4\mathbf{a}_5][\mathbf{a}_2\mathbf{a}_3\mathbf{a}_5] = 0, \tag{1}$$

$$[\mathbf{A}_1\mathbf{A}_2\mathbf{A}_5\mathbf{A}_6][\mathbf{A}_3\mathbf{A}_4\mathbf{A}_5\mathbf{A}_6] - [\mathbf{A}_1\mathbf{A}_3\mathbf{A}_5\mathbf{A}_6][\mathbf{A}_2\mathbf{A}_4\mathbf{A}_5\mathbf{A}_6] + [\mathbf{A}_1\mathbf{A}_4\mathbf{A}_5\mathbf{A}_6][\mathbf{A}_2\mathbf{A}_3\mathbf{A}_5\mathbf{A}_6] = 0 \tag{2}$$

where (1) is with respect to homogeneous 3-vectors $\mathbf{a}_i, i = 1..5$ and (2) is with respect to homogeneous 4-vectors $\mathbf{A}_i, i = 1..6$. They are often used to simplify bracket computations.

Under a pinhole camera, a space point $\mathbf{M}_i$ is projected to an image point $\mathbf{m}_i$ by:

$$x_i\mathbf{m}_i = \mathbf{K}(\mathbf{R}, \mathbf{t})\mathbf{M}_i, \tag{3}$$

where $\mathbf{K}$ is the $3 \times 3$ matrix of camera intrinsic parameters, $\mathbf{R}$ a $3 \times 3$ rotation matrix, $\mathbf{t}$ a translation vector with 3 elements, and $x_i$ a nonzero scalar. If $x_i$ were zero, $\mathbf{M}_i$ could not be projected to the image plane. If the optical center, denoted as $\mathbf{O}$, is not at infinity, its non-homogeneous coordinates are $\hat{\mathbf{O}} = -\mathbf{R}^T\mathbf{t}$. We assume that the camera optical center $\mathbf{O}$ and the space points $\mathbf{M}_i$ are not at infinity throughout this paper. The DLT method solves the camera parameters $\mathbf{K}, \mathbf{R}, \mathbf{t}$ by solving for the matrix $\mathbf{K}(\mathbf{R}, \mathbf{t})$ from the pairs $\mathbf{M}_i \leftrightarrow \mathbf{m}_i$ with $i \geq 6$.

Let the non-homogeneous coordinate of $\mathbf{M}_i$ be $\hat{\mathbf{M}}_i$, so $\mathbf{M}_i = (\hat{\mathbf{M}}_i^T, 1)^T$. By (3) or (Carlsson 1995) we have:

$$
\begin{aligned}
&x_i x_j x_k [\mathbf{m}_i, \mathbf{m}_j, \mathbf{m}_k] \\
&\quad = [x_i \mathbf{m}_i, x_j \mathbf{m}_j, x_k \mathbf{m}_k] \\
&\quad = \det(\mathbf{K})[\mathbf{R}\hat{\mathbf{M}}_i + \mathbf{t}, \mathbf{R}\hat{\mathbf{M}}_j + \mathbf{t}, \mathbf{R}\hat{\mathbf{M}}_k + \mathbf{t}] \\
&\quad = \det(\mathbf{K})[\hat{\mathbf{M}}_i + \mathbf{R}^\tau\mathbf{t}, \hat{\mathbf{M}}_j + \mathbf{R}^\tau\mathbf{t}, \hat{\mathbf{M}}_k + \mathbf{R}^\tau\mathbf{t}] \\
&\quad = \det(\mathbf{K})[\hat{\mathbf{M}}_i - \hat{\mathbf{O}}, \hat{\mathbf{M}}_j - \hat{\mathbf{O}}, \hat{\mathbf{M}}_k - \hat{\mathbf{O}}] \\
&\quad = \det(\mathbf{K}) \begin{bmatrix} \hat{\mathbf{M}}_i, & \hat{\mathbf{M}}_j, & \hat{\mathbf{M}}_k, & \hat{\mathbf{O}} \\ 1, & 1, & 1, & 1 \end{bmatrix} \\
&\quad = \det(\mathbf{K})[\mathbf{M}_i, \mathbf{M}_j, \mathbf{M}_k, \mathbf{O}]. \qquad (4)
\end{aligned}
$$

Then, $[\mathbf{m}_i, \mathbf{m}_j, \mathbf{m}_k] = 0$ if and only if $[\mathbf{M}_i, \mathbf{M}_j, \mathbf{M}_k, \mathbf{O}] = 0$. Namely, $\mathbf{m}_i, \mathbf{m}_j, \mathbf{m}_k$ are collinear if and only if $\mathbf{M}_i, \mathbf{M}_j, \mathbf{M}_k, \mathbf{O}$ are coplanar.

In the following, for the notational convenience, if no ambiguity can be aroused, $\mathbf{M}_i, i = 1..6$ will be simply denoted as $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}$, and the commas in the brackets will be omitted.

There is a unique proper quadric cone with $\mathbf{1}$ as the vertex and passing through $\mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}$ with no three collinear and no four coplanar. Any point $\mathbf{X}$ is on this quadric cone if and only if (Wu and Hu 2003):

$$
\frac{[\mathbf{1246}][\mathbf{1356}]}{[\mathbf{1236}][\mathbf{1456}]} = \frac{[\mathbf{124X}][\mathbf{135X}]}{[\mathbf{123X}][\mathbf{145X}]}. \qquad (5)
$$

This representation is not unique because a permutation of $\mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}$ also results in a representation of the same quadric cone.

There is a unique proper twisted cubic passing through $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}$ with no three collinear and no four coplanar. Any point $\mathbf{X}$ is on this twisted cubic if and only if (Wu and Hu 2003):

$$
\begin{cases}
\dfrac{[\mathbf{1246}][\mathbf{1356}]}{[\mathbf{1236}][\mathbf{1456}]} = \dfrac{[\mathbf{124X}][\mathbf{135X}]}{[\mathbf{123X}][\mathbf{145X}]}, \\
\dfrac{[\mathbf{1246}][\mathbf{2356}]}{[\mathbf{1236}][\mathbf{2456}]} = \dfrac{[\mathbf{124X}][\mathbf{235X}]}{[\mathbf{123X}][\mathbf{245X}]}, \\
\quad \text{and } \mathbf{X} \text{ is not on line } \mathbf{12} \text{ except } \mathbf{1}, \mathbf{2}.
\end{cases} \qquad (6)
$$

This representation is not unique because a permutation of $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}$ also results in a representation of the same twisted cubic.

We can see that each bracket in the first equation of (6) has the point $\mathbf{1}$. So according to (5), its geometric meaning is that $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}, \mathbf{X}$ lie on a quadric cone with $\mathbf{1}$ as the vertex (Wu and Hu 2003). Similarly since each bracket of the second equation of (6) has the point $\mathbf{2}$, this equation means that $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}, \mathbf{X}$ lie on a quadric cone with $\mathbf{2}$ as the vertex. These meanings are consistent with the theorem in (Semple and Kneebone 1952) that a twisted cubic can be the intersection of two quadrics.

## 3 Construction of Evaluation Functions for Six Points from a Complete Framework of Invariance

In this section, for six space points we assume that no five of them are coplanar and no three of them are collinear. Also assume that the camera optical center is non-incident with the space points. These assumptions are denoted by *NC*.

### 3.1 Invariance between Space Points and their Image Points

Between six space points $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}$ and their image points $\mathbf{m}_i, i = 1..6$, there exists an invariant relationship as (Wu and Hu 2005; Carlsson 1995; Bayro-Corrochano and Banarer 2002; Carlsson 1998; Quan 1995):

$$
\begin{aligned}
&f_{1234;56} \\
&\quad = [\mathbf{m_3 m_4 m_5}][\mathbf{m_1 m_2 m_6}][\mathbf{1235}][\mathbf{1245}][\mathbf{1346}][\mathbf{2346}] \\
&\qquad + [\mathbf{m_3 m_4 m_6}][\mathbf{m_1 m_2 m_5}][\mathbf{1236}][\mathbf{1246}][\mathbf{1345}][\mathbf{2345}] \\
&\qquad + [\mathbf{m_2 m_3 m_5}][\mathbf{m_1 m_4 m_6}][\mathbf{1245}][\mathbf{1345}][\mathbf{1236}][\mathbf{2346}] \\
&\qquad + [\mathbf{m_2 m_3 m_6}][\mathbf{m_1 m_4 m_5}][\mathbf{1246}][\mathbf{1346}][\mathbf{1235}][\mathbf{2345}] \\
&\qquad - [\mathbf{m_2 m_4 m_5}][\mathbf{m_1 m_3 m_6}][\mathbf{1235}][\mathbf{1345}][\mathbf{1246}][\mathbf{2346}] \\
&\qquad - [\mathbf{m_2 m_4 m_6}][\mathbf{m_1 m_3 m_5}][\mathbf{1236}][\mathbf{1346}][\mathbf{1245}][\mathbf{2345}] \\
&\quad = 0. \qquad (7)
\end{aligned}
$$

This expression is symmetric with respect to $(\mathbf{5}, \mathbf{6})$ and symmetric with respect to $(\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4})$. Thus it is denoted as $f_{1234;56}$. After a permutation of $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}$ and their corresponding image points in (7), the obtained equation called a permuted equation is still an invariant relationship but not independent of (7). Under the assumption *NC*, there is at least one term in $f_{1234;56}$ being nonzero. Dividing $f_{1234;56}$ by this nonzero term gives an expression on cross ratios of the image points and cross ratios of the space points. Thus, the relationship is an invariant relationship, also called as

invariance. For example, if the first term is nonzero, (7) becomes:

$$\frac{[\mathbf{m}_1\mathbf{m}_2\mathbf{m}_5][\mathbf{m}_3\mathbf{m}_4\mathbf{m}_6]}{[\mathbf{m}_3\mathbf{m}_4\mathbf{m}_5][\mathbf{m}_1\mathbf{m}_2\mathbf{m}_6]}I_1 - \frac{[\mathbf{m}_1\mathbf{m}_3\mathbf{m}_5][\mathbf{m}_2\mathbf{m}_4\mathbf{m}_6]}{[\mathbf{m}_3\mathbf{m}_4\mathbf{m}_5][\mathbf{m}_1\mathbf{m}_2\mathbf{m}_6]}I_2 + \frac{[\mathbf{m}_1\mathbf{m}_4\mathbf{m}_5][\mathbf{m}_2\mathbf{m}_3\mathbf{m}_6]}{[\mathbf{m}_3\mathbf{m}_4\mathbf{m}_5][\mathbf{m}_1\mathbf{m}_2\mathbf{m}_6]}I_3 + \frac{[\mathbf{m}_2\mathbf{m}_3\mathbf{m}_5][\mathbf{m}_1\mathbf{m}_4\mathbf{m}_6]}{[\mathbf{m}_3\mathbf{m}_4\mathbf{m}_5][\mathbf{m}_1\mathbf{m}_2\mathbf{m}_6]}I_4 - \frac{[\mathbf{m}_2\mathbf{m}_4\mathbf{m}_5][\mathbf{m}_1\mathbf{m}_3\mathbf{m}_6]}{[\mathbf{m}_3\mathbf{m}_4\mathbf{m}_5][\mathbf{m}_1\mathbf{m}_2\mathbf{m}_6]}I_5 + 1 = 0$$

where

$$I_1 = \frac{[\mathbf{1345}][\mathbf{2345}][\mathbf{1236}][\mathbf{1246}]}{[\mathbf{1235}][\mathbf{1245}][\mathbf{1346}][\mathbf{2346}]}, \qquad I_2 = \frac{[\mathbf{2345}][\mathbf{1236}]}{[\mathbf{1235}][\mathbf{2346}]},$$

$$I_3 = \frac{[\mathbf{2345}][\mathbf{1246}]}{[\mathbf{1245}][\mathbf{2346}]}, \qquad I_4 = \frac{[\mathbf{1345}][\mathbf{1236}]}{[\mathbf{1235}][\mathbf{1346}]},$$

$$I_5 = \frac{[\mathbf{1345}][\mathbf{1246}]}{[\mathbf{1245}][\mathbf{1346}]}$$

are on cross ratios of space points, and their coefficients are on cross ratios of the image points (Carlsson 1995, 1998; Wu 2001).

When the space points and the camera optical center lie on some specific configuration, these space points and their image points still satisfy this relationship $f_{1234;56}$. However, the relationship cannot be a sufficient condition. The necessary and sufficient conditions of two specific configurations are given as follows.

**Proposition 1** *Let the quadric cone with* $\mathbf{1}$ *as the vertex and through* $\mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}$ *be QC. Then, the camera optical center* $\mathbf{O}$ *lies on QC if and only if*:

$$g_{1,(25;34)} = [\mathbf{m}_1\mathbf{m}_2\mathbf{m}_3][\mathbf{m}_1\mathbf{m}_4\mathbf{m}_5][\mathbf{1246}][\mathbf{1356}] - [\mathbf{m}_1\mathbf{m}_2\mathbf{m}_4][\mathbf{m}_1\mathbf{m}_3\mathbf{m}_5][\mathbf{1236}][\mathbf{1456}] = 0, \quad (8)$$

*where the expression is denoted as* $g_{1,(25;34)}$ *because each bracket contains* $\mathbf{1}$ *or* $\mathbf{m}_1$ *and because the expression is symmetric with respect to* $(\mathbf{2}, \mathbf{5})$ *or* $(\mathbf{3}, \mathbf{4})$. *After a permutation of* $\mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}$ *and their corresponding image points in* (8), *the obtained equation is still an invariant relationship of* $\mathbf{O}$ *lying on the same quadric cone and still dependent on the one before permutation.*

*Proof* If no three points are collinear and no four points are coplanar for $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}, \mathbf{O}$, then according to (5), $\mathbf{O}$ lies on *QC* if and only if

$$\frac{[\mathbf{1246}][\mathbf{1356}]}{[\mathbf{1236}][\mathbf{1456}]} = \frac{[\mathbf{124O}][\mathbf{135O}]}{[\mathbf{123O}][\mathbf{145O}]}. \quad (9)$$

Moreover, by (4) we have

$$\frac{[\mathbf{124O}][\mathbf{135O}]}{[\mathbf{123O}][\mathbf{145O}]} = \frac{[\mathbf{m}_1\mathbf{m}_2\mathbf{m}_4][\mathbf{m}_1\mathbf{m}_3\mathbf{m}_5]}{[\mathbf{m}_1\mathbf{m}_2\mathbf{m}_3][\mathbf{m}_1\mathbf{m}_4\mathbf{m}_5]}.$$

It follows that we get

$$\frac{[\mathbf{1246}][\mathbf{1356}]}{[\mathbf{1236}][\mathbf{1456}]} = \frac{[\mathbf{m}_1\mathbf{m}_2\mathbf{m}_4][\mathbf{m}_1\mathbf{m}_3\mathbf{m}_5]}{[\mathbf{m}_1\mathbf{m}_2\mathbf{m}_3][\mathbf{m}_1\mathbf{m}_4\mathbf{m}_5]}. \quad (10)$$

Expanding (10) gives $g_{1,(25;34)} = 0$ of (8).

If there are three points collinear, this is contrary to the assumption *NC* of this section. If there are four points coplanar, this is also contrary to *NC*. Because at this time, *QC* is degenerate that consists of two planes or one plane. That *QC* is a plane is contrary to *NC* that no five points are coplanar. That *QC* consists of two planes is contrary to *NC* that the optical center is non-incident with the space points. □

Similarly, by (4) and (6), we have the following proposition.

**Proposition 2** *The camera optical center* $\mathbf{O}$ *lies on the proper twisted cubic determined by* $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}$ *if and only if*

$$\begin{cases} g_{1,(25;34)} = [\mathbf{m}_1\mathbf{m}_2\mathbf{m}_3][\mathbf{m}_1\mathbf{m}_4\mathbf{m}_5][\mathbf{1246}][\mathbf{1356}] \\ \qquad\qquad - [\mathbf{m}_1\mathbf{m}_2\mathbf{m}_4][\mathbf{m}_1\mathbf{m}_3\mathbf{m}_5][\mathbf{1236}][\mathbf{1456}] = 0, \\ g_{2,(15;34)} = [\mathbf{m}_1\mathbf{m}_2\mathbf{m}_3][\mathbf{m}_2\mathbf{m}_4\mathbf{m}_5][\mathbf{1246}][\mathbf{2356}] \\ \qquad\qquad - [\mathbf{m}_1\mathbf{m}_2\mathbf{m}_4][\mathbf{m}_2\mathbf{m}_3\mathbf{m}_5][\mathbf{1236}][\mathbf{2456}] = 0. \end{cases} \quad (11)$$

*After a permutation of* $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}$ *and their corresponding image points in* (11), *the obtained equation system is still the invariant relationships of* $\mathbf{O}$ *lying on the same twisted cubic and still dependent on the one before permutation.*

Notice that in (6), there is another condition: $\mathbf{O}$ is not on the line through $\mathbf{1}$ and $\mathbf{2}$. Here this additional condition in (11) is unnecessary because we have assumed that $\mathbf{O}$ is non-incident with the space points.

### 3.2 A Unified and Complete Framework of Invariance for Six Points

We have a unified and complete framework of invariance for six space points (Wu and Hu 2005). For self-containing, the main outline is given as follows.

For the optical center $\mathbf{O}$ and five spatial points $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}$, there is a Grassmann-Plücker relation of (2) as:

$$[\mathbf{1245}][\mathbf{3O45}] - [\mathbf{1345}][\mathbf{2O45}] + [\mathbf{2345}][\mathbf{1O45}] = 0.$$

From this equation and from (4), we have:

$$h_{123,45} = x_3[\mathbf{1245}][\mathbf{m}_3\mathbf{m}_4\mathbf{m}_5] - x_2[\mathbf{1345}][\mathbf{m}_2\mathbf{m}_4\mathbf{m}_5] + x_1[\mathbf{2345}][\mathbf{m}_1\mathbf{m}_4\mathbf{m}_5] = 0.$$

Similarly from other five-group points of **1**, **2**, **3**, **4**, **5**, **6**, we obtain the corresponding equations:

$$h_{123,46}=0, \qquad h_{124,35}=0, \qquad h_{124,36}=0,$$
$$h_{125,36}=0, \qquad h_{126,35}=0.$$

These equations are all linear with $x_i, i = 1..6$. We do not consider other equations $h_{ijk,mn} = 0$ because there are at most such six equations independent: for six points there are in total 18 equations of (3), from which after eliminating 12 parameters of $\mathbf{K}(\mathbf{R}, \mathbf{t})$, only 6 equations on $x_i, \mathbf{m}_i, \mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}$ are left to be independent. $h_{ijk,mn} = 0$ are the equations on $x_i, \mathbf{m}_i, \mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}$ and so at most six of them are independent.

Since the invariants are not related to the depths, we eliminate $x_i, i = 1..6$ from the above six equations $h_{ijk,mn} = 0$. During the eliminations, we obtain two equations on $x_1, x_2$:

$$\begin{aligned} &x_1([\mathbf{m}_1\mathbf{m}_4\mathbf{m}_5][\mathbf{m}_3\mathbf{m}_4\mathbf{m}_6][\mathbf{2345}][\mathbf{1246}] \\ &\quad -[\mathbf{m}_3\mathbf{m}_4\mathbf{m}_5][\mathbf{m}_1\mathbf{m}_4\mathbf{m}_6][\mathbf{1245}][\mathbf{2346}]) \\ &= x_2([\mathbf{m}_2\mathbf{m}_4\mathbf{m}_5][\mathbf{m}_3\mathbf{m}_4\mathbf{m}_6][\mathbf{1345}][\mathbf{1246}] \\ &\quad -[\mathbf{m}_3\mathbf{m}_4\mathbf{m}_5][\mathbf{m}_2\mathbf{m}_4\mathbf{m}_6][\mathbf{1245}][\mathbf{1346}]), \end{aligned} \tag{12}$$

$$\begin{aligned} &x_1([\mathbf{m}_1\mathbf{m}_3\mathbf{m}_5][\mathbf{m}_3\mathbf{m}_4\mathbf{m}_6][\mathbf{2345}][\mathbf{1236}] \\ &\quad -[\mathbf{m}_3\mathbf{m}_4\mathbf{m}_5][\mathbf{m}_1\mathbf{m}_3\mathbf{m}_6][\mathbf{1235}][\mathbf{2346}]) \\ &= x_2([\mathbf{m}_2\mathbf{m}_3\mathbf{m}_5][\mathbf{m}_3\mathbf{m}_4\mathbf{m}_6][\mathbf{1345}][\mathbf{1236}] \\ &\quad -[\mathbf{m}_3\mathbf{m}_4\mathbf{m}_5][\mathbf{m}_2\mathbf{m}_3\mathbf{m}_6][\mathbf{1235}][\mathbf{1346}]). \end{aligned} \tag{13}$$

Let the coefficients of $x_1, x_2$ in (12) be $c_1, c_2$, and the ones in (13) be $e_1, e_2$. Because the depths $x_1, x_2$ cannot be zero, we have $c_1 = 0$ if and only if $c_2 = 0$, also $e_1 = 0$ if and only if $e_2 = 0$. Thus, there are only three cases: (i) $c_1 \neq 0$ and $e_1 \neq 0$; (ii) $c_1 = 0$ but $e_1 \neq 0$, or $c_1 \neq 0$ but $e_1 = 0$; (iii) $c_1 = 0$ and $e_1 = 0$. For each case, continuously eliminate $x_1, x_2$, the corresponding invariants can be established as:

(i) $c_1 \neq 0$ and $e_1 \neq 0$. At the time, also $c_2 \neq 0$ and $e_2 \neq 0$. From (12) and (13), there is $\frac{x_1}{x_2} = \frac{c_2}{c_1} = \frac{e_2}{e_1}$. It follows that $c_2e_1 - c_1e_2 = 0$. Simplifying this equation by (1) and (2), we obtain the invariant relation (7). This is the general case.

(ii) $c_1 = 0$ but $e_1 \neq 0$, or $c_1 \neq 0$ but $e_1 = 0$. In this case, the invariant is just $c_1 = 0$ or $e_1 = 0$. By Proposition 1, we know the camera optical center and the space points lie on a quadric cone configuration. This is called quadric cone case.

(iii) $c_1 = 0$ and $e_1 = 0$. In this case, the invariants are just $c_1 = 0$ and $e_1 = 0$. By Proposition 2, we know the camera and the space points lie on a twisted cubic configuration. This is called twisted cubic case.

Due to the above complete classification and due to considering all the independent equations $h_{ijk,mn} = 0$ of eliminating $\mathbf{K}(\mathbf{R}, \mathbf{t})$ from (3), the invariants in these three cases constitute a complete framework of invariance for six points.

### 3.3 Evaluation Function Construction

Each invariant in above complete framework just can be used to identify type of the geometric information between the camera and the scene. However, due to the non-uniform noise of points, stability of a single invariant is much affected by the point orders. It follows that we need to construct evaluation functions that are independent of the point orders. These constructions for the complete three cases are respectively as follows.

*Case I. The General Case*

In this case, the camera optical center and the space points are in the general position and the camera projective matrix can be determined uniquely. The space points and the image points satisfy (7) and all its permuted equations $f_{ijkl;pq} = 0$, but do not satisfy any permuted equations $g_{k,(ij;pq)} = 0$ of (8). The evaluation function is defined as:

$$I_{\text{general}} = \frac{1}{15} \sum_{ijkl;pq \in S} \frac{1}{W^2_{ijkl;pq}} f^2_{ijkl;pq},$$

where $S$ is a set consisting of all the different combinations of $(ijkl; pq)$ from $1, 2, 3, 4, 5, 6$ with 15 elements in total. $W_{ijkl;pq}$ is a weight to $f_{ijkl;pq}$ given as follows. Let $w_{n1}$ be the absolute value of the product of space-point brackets in the $n$-th term of $f_{ijkl;pq}$, and let $w_{n2}$ be that of image-point brackets in the $n$-th term of $f_{ijkl;pq}$. Then sort in ascending order all $w_{n1}$ with varying $n$ and let the result be $B_1$. Sort in ascending order all $w_{n2}$ with varying $n$ and let the result be $B_2$. The weight $W_{ijkl;pq}$ is given as the product of the fourth element of $B_1$ and the fourth element of $B_2$.

*Error Analysis for the Added Weights* Why do we take the fourth elements of $B_1$ and $B_2$ when assigning the weight $W_{ijk;opq}$ to $f_{ijkl;pq}$? The reason is given as follows. $f_{ijkl;pq}$ has six terms. Let $F = \sum_{i=1}^{6} v_i$ be a general function containing six terms $v_i, i = 1..6$. The error for $v_i$ is denoted as $e_i$. In practice the obtained function value is the value of $F^e = \sum_{i=1}^{6}(v_i + e_i)$. Without loss of generality, we assume $|v_i + e_i| \leq |v_4 + e_4| \leq |v_5 + e_5| \leq |v_6 + e_6|, i = 1..3$, namely the $|v_4 + e_4|$ term is the fourth element of the list $|v_i + e_i|$ after being ordered ascendingly. Since for our case of $f_{ijkl;pq}$, the values of $v_i$ are much bigger than the errors $e_i$, from $|v_i + e_i| \leq |v_4 + e_4| \leq |v_5 + e_5| \leq |v_6 + e_6|$ there is still $|v_i| \leq |v_4| \leq |v_5| \leq |v_6|$. We assign the weight $v_4 + e_4$ to $F^e$ and then have the weight function $\frac{F^e}{v_4+e_4}$. The error for

the weight function is the value of $|\frac{F^e}{v_4+e_4} - \frac{F}{v_4}|$, denoted as *ER*. We expand *ER* in a Taylor series to order 1 and obtain:

$$ER \approx \left| \frac{e_1+e_2+e_3+e_5+e_6}{v_4} - \frac{(v_1+v_2+v_3+v_5+v_6)e_4}{v_4^2} \right|. \quad (14)$$

For some parameters, if the value of $F$ is zero, then we have $v_4 = -(v_1+v_2+v_3+v_5+v_6)$. Substitute $v_4 = -(v_1+v_2+v_3+v_5+v_6)$ into (14), we obtain: $ER \approx \frac{|e_1+e_2+e_3+e_4+e_5+e_6|}{|v_4|}$. Since $|v_4|$ is much bigger than the errors $e_i$, *ER* is small and thus for the zero $F$, the weight function $\frac{F^e}{v_4+e_4}$ is stable to noise $e_i$. Clearly, the bigger the denominator, the smaller the error *ER*. However, for some parameters, the values of $F$ should not be zero, the bigger weight is not the better choice. Assume we assigned the term $|v_6+e_6|$ of having the maximal absolute value in the list $|v_i+e_i|, i = 1..6$ to $F^e$. Then, $|\frac{F^e}{v_6+e_6}| \le |\frac{F^e}{v_4+e_4}|$. Therefore, the weight function by $v_6+e_6$ is more close to zero than the weight function by $v_4+e_4$. This means that the ability to discriminate between the zero values of $F$ and the nonzero values of $F$ by $|\frac{F^e}{v_6+e_6}|$ is poor. Thus, taking the fourth element as the weight is a sensible tradeoff between the stability to noise and the discriminability from zero to nonzero.

*Case II. The Quadric Cone Case*

In this case, the camera optical center and the space points lie on a quadric cone and the camera projective matrix can be determined uniquely. Let the vertex of the quadric cone be **1**, then the space points and their image points satisfy (8) and all its permuted equations $g_{1,(ij;pq)} = 0$. The evaluation function is defined as:

$$I_{1,\text{cone}} = \frac{1}{15} \sum_{(ij;pq)\in S} \frac{1}{W_{1,(ij;pq)}^2} g_{1,(ij;pq)}^2,$$

where $S = \{(23;45), (24;35), (25;34), (23;46), (24;36), (26;34), (23;56), (25;36), (26;35), (24;56), (25;46), (26;45), (34;56), (35;46), (36;45)\}$ has 15 elements, and $W_{1,(ij;pq)}$ is a weight to $g_{1,(ij;pq)}$ given as the mean of the absolute values of the two terms in $g_{1,(ij;pq)}$.

*Error Analysis for the Added Weights* $W_{1,(ij;pq)}$ is taken as the mean of the absolute values of the two terms in $g_{1,(ij;pq)}$. The reason is as follows. Let $F = v_1 + v_2$ be a general function containing two terms $v_1, v_2$. The error for $v_i$ is denoted as $e_i$. In practice the obtained function value is the value of $F^e = (v_1+e_1)+(v_2+e_2)$. We assign the weight $\frac{|v_1+e_1|+|v_2+e_2|}{2}$ to $F^e$ and then have $\frac{F^e}{(|v_1+e_1|+|v_2+e_2|)/2}$. If for some parameters $v_1 + v_2 = 0$, then the error of the weight function is $ER = \frac{|e_1+e_2|}{(|v_1+e_1|+|v_2+e_2|)/2}$. For our case of $g_{1,(ij;pq)}$, the values of $v_i$ is much bigger than the errors $e_i$. Thus, $ER \approx 0$. If for some parameters the value of $F$ is nonzero, the values of the weight function are within the scope:

$$\left[ \frac{F^e}{\max(|v_1+e_1|,|v_2+e_2|)}, \frac{F^e}{\min(|v_1+e_1|,|v_2+e_2|)} \right] \approx \left[ \frac{|v_1+v_2|}{\max(|v_1|,|v_2|)}, \frac{|v_1+v_2|}{\min(|v_1|,|v_2|)} \right].$$

*Case III. The Twisted Cubic Case*

In this case, the camera optical center and the space points lie on a twisted cubic and the camera projective matrix cannot be determined uniquely. Equation (11) and all their permuted equations $g_{k,(ij;pq)} = 0$ are satisfied by the space points and their image points. The evaluation function is defined as:

$$I_{\text{tc}} = \frac{1}{6}(I_{1,\text{cone}} + I_{2,\text{cone}} + I_{3,\text{cone}} + I_{4,\text{cone}} + I_{5,\text{cone}} + I_{6,\text{cone}}),$$

where $I_{i,\text{cone}}$ is the evaluation function of the quadric cone case with **i** being the cone vertex.

In each above case, the constructed weights are always nonzero under *NC*. As we consider all the different $f_{ijkl;pq} = 0, g_{k,(ij;pq)} = 0$ from all different point orders, the evaluation functions are independent of the point orders, which makes them robust to noise. Taking square of $f_{ijkl;pq}$ and $g_{1,(ij;pq)}$ instead of their absolute values also makes the evaluation functions with higher discriminative ability. Coble studied the invariants of points between spaces of different dimensions (Coble 1922). Then, Buchanan interpreted the invariants of six points from 2D space and 3D space in an easy readable way (Buchanan 1992). The invariants are obtained by using six basic Joubert invariants, of which each Joubert invariant has five terms. Here for easy implementation, we use the invariant (7).

## 4 Unreliable Point Detection and Solutions

In practice, more than six space points and their image points are often used to estimate camera parameters by DLT. Based on the evaluation functions in Sect. 3, this section gives algorithms to detect the unreliability of $N \ge 6$ points and then provides the corresponding solutions.

The unreliability of $N \ge 6$ space points and their image points for camera parameter estimation is caused by the two cases: (u-i) degenerate configuration; (u-ii) mismatching or overlarge noise, as given in the introduction section. For these two cases, neither an optimized DLT method

nor a non-optimized DLT method can give satisfactory results. The constructed $I_{\text{tc}}$ can be used to detect case (u-i) and the constructed $I_{\text{general}}$ can be used to detect case (u-ii). The function $I_{i,\text{cone}}$ serves as the basis for establishing $I_{\text{tc}}$. The quadric cone configuration is not degenerate for computing the camera parameters. Moreover, the space and image points of this configuration also satisfy the function $I_{\text{general}} = 0$. Thus, in the following we do not care about the quadric cone case, and just use $I_{\text{tc}}$ and $I_{\text{general}}$ to detect the two cases of (u-i) and (u-ii).

The detection on the unreliability of six points is given by the following Algorithm 1. Then, the algorithm is extended to more than six points by Algorithm 2. Following Algorithm 2 are the corresponding solution methods, of which one is the filtering RANSAC Algorithm 3.

Let $\varepsilon_1, \varepsilon_2$ be two given thresholds.

**Algorithm 1**

Input: the coordinates of six pairs of space and image points, where the points satisfy *NC*.

Output: whether or not the input data are reliable for camera parameter estimation.

Step 1. Determine whether $I_{\text{tc}} < \varepsilon_1$ on the input data. If yes, then output: the space points and the camera optical center lie on a degenerate configuration, and the input data are unreliable; Otherwise, do Step 2.

Step 2. Determine whether $I_{\text{general}} < \varepsilon_2$ on the input data. If yes, then output: the input data are reliable; Otherwise, output: some space points or some image points have large noise or are mismatched, with the result that they are unreliable.

*Remark 1* If the used six space points and their image points do not satisfy *NC*, the degeneracy is in the incidence case as mentioned in the introduction section. This degeneracy can be detected easily by detecting the linearly dependent relations among these points.

A six-point group means a set consisting of six space points and their image points.

**Algorithm 2**

Input: the coordinates of $N$ ($\geq 6$) space points and their image points.

Output: one of the following for camera parameter estimation

(i) all the space points and the camera lie on a degenerate configuration and the input data are unreliable;
(ii) all the input data are in the general position but unreliable;
(iii) all the input data are reliable;
(iv) some of the input data are reliable and some are not.

Step 1. Construct a set $G$ consisting of six-point groups: From all pairs of the space points and their image points, choose five pairs that satisfy *NC*. Denote these five space points as $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}$ and combine them with the remaining space points $\mathbf{M}_i$. We put the combinations satisfying *NC* into a set $G = \{\{\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{M}_i\} \mid i = 1..N_1\}$. If all the space points appear in $G$, do the next step. Otherwise, choose another five space points and repeat the process. The obtained new set of six-point groups and the original $G$ is united into another new set. The result is still denoted as $G$. Until all the space points are considered, do the next step.

Step 2. For each six-point group in $G$, compute its values of $I_{\text{tc}}$ and $I_{\text{general}}$. If for all groups, $I_{\text{tc}} < \varepsilon_1$, then output (i); Otherwise if for all groups, $I_{\text{general}} > \varepsilon_2$, then output (ii); Otherwise if for all groups, $I_{\text{general}} < \varepsilon_2$, then output (iii); Otherwise, output (iv).

*Remark 2* $I_{\text{tc}} = 0$ is the necessary and sufficient condition that six space points and camera optical center lie on a twisted cubic degenerate configuration. $I_{\text{general}} = 0$ is a necessary condition that six space points and their image points are consistent. Therefore, the outputs (i), (ii), (iv) of Algorithm 2 can be fully trusted. For the output (iii) of Algorithm 2, it can be further enhanced by a general RANSAC method as pointed out below.

According to the different outputs from Algorithm 2, we take different actions. If the output is (i), we need to adjust the camera position; if the output is (ii), we re-match the space points and the image points or re-extract the image points or the space points; if the output is (iii), we estimate the camera parameters by the following general RANSAC from Algorithm 3; if the output is (iv), we estimate the camera parameters by the following filtering RANSAC from Algorithm 3.

Let $\varepsilon_3$ be another given threshold.

**Algorithm 3**

Input: the coordinates of $N$ ($\geq 6$) space points and their image points.

Output: the camera parameters.

Step 1. Construct a set $G$ consisting of six-point groups as Step 1 of Algorithm 2.

Step 2. Remove the unreliable six-point groups in $G$ by Algorithm 1.

Step 3. Score each pair of space and image points $(\mathbf{M}_k, \mathbf{m}_k)$ in $G$ by considering each six-point group $spg_i$ in $G$ respectively:

(i) For each $spg_i$, compute a camera projective matrix $\mathbf{P}_i$ and find its corresponding inlier set $is_i$. A pair $(\mathbf{M}_k, \mathbf{m}_k)$ is an inlier if

$d(\mathbf{m}_k, \mathbf{P}_i\mathbf{M}_k) < \varepsilon_3$. Then, compute another camera projective matrix $\mathbf{P}'_i$ by all pairs in $spg_i \cup is_i$.

(ii) Put all $spg_i \cup is_i$ from different $i$ together and count the times that each $(\mathbf{M}_k, \mathbf{m}_k)$ appears. The result is the score of $(\mathbf{M}_k, \mathbf{m}_k)$. Let $st_k$ be the set consisting of the subscripts for different $spg_i \cup is_i$ when $(\mathbf{M}_k, \mathbf{m}_k)$ appears. If two pairs $(\mathbf{M}_{k_j}, \mathbf{m}_{k_j})$, $j = 1, 2$ have the same score, order them by using $S_{k_j} = \{d(\mathbf{m}_{k_j}, \mathbf{P}'_{i_0}\mathbf{M}_{k_j}) \mid i_0 \in st_{k_j}\}$. The sum of mean and standard deviation of $S_{k_j}$ is denoted as $ms_{k_j}$. The pair with smaller $ms_{k_j}$ has a higher score.

Step 4. Take the six pairs of space and image points with the highest six scores not on a degenerate configuration, where $I_{\text{tc}}$ is again used to detect the degeneracy. Then from these six pairs and their inliers, compute the camera parameters and output the results.

Algorithm 3 is a RANSAC to estimate camera parameters which can deal with the image points and space points containing overlarge noise, mismatched pairs, or in a degenerate configuration. Step 2 filters out the unreliable six-point groups by using $I_{\text{tc}}, I_{\text{general}}$ through Algorithm 1. Step 4 chooses the six-point group with the highest score which is not on a degenerate configuration detected by $I_{\text{tc}}$. We call this algorithm a **filtering RANSAC method (FRM)**. The algorithm without the filtering in Step 2 and without the detection by $I_{\text{tc}}$ in Step 4 is called a **general RANSAC method (GRM)**. As previously pointed out, if the output of Algorithm 2 is (iii), we use GRM and if the output of Algorithm 2 is (iv), we use FRM. We have compared FRM and GRM in experiments showing that FRM has much higher success rate than GRM if image points and space points containing overlarge noise, mismatched pairs, or in a degenerate configuration. Therefore, it is necessary and important to filter these points out before applying the general RANSAC. It should be noticed that the computational cost by FRM is higher.

*Remark 3* Another general way of RANSAC is to score each six-point group by the number of its inliers, then to use the group with the highest score and its inliers together for computing the camera projective matrix. However we find that this method almost fails in removing the mismatched pair or the pair with overlarge noise.

*Learning the Thresholds* How to choose the values of the thresholds $\varepsilon_1, \varepsilon_2, \varepsilon_3$? We learn them from extensive data. All the $\varepsilon_1, \varepsilon_2, \varepsilon_3$ are the thresholds to discriminate two groups of values of a function: the group denoted as $GR_1$ whose values should be zero and the group denoted as $GR_2$ whose values should be nonzero. Due to noise, the values in $GR_1$ become nonzero but are close to zero. Let the functions for the three thresholds be unified as $FUN$. Then, from the values of $FUN$ on many data, we make clustering by $k$-means method (Duda et al. 2000). $k$ is not set as 2 since the nonzero values of $FUN$ are looser. While, we round all the values of $FUN$ and the number of these different results is taken as $k$. These different results by rounding the values of $FUN$ are just as the initial cluster centers. After obtaining $k$ clusters, we are to find $GR_1$ and $GR_2$. For $\varepsilon_1, \varepsilon_2$, the obtained cluster with the smallest values among all the $k$ clusters is taken as $GR_1$ and the remaining clusters are combined into one as $GR_2$. For $\varepsilon_3$, because the corresponding $FUN$ is sensitive to noise, the half clusters of the $k$ clusters with smaller values are combined into $GR_1$ and another half with larger values are combined into $GR_2$. Lastly, take the maximum value of $GR_1$ and the minimum value of $GR_2$. Then, the threshold is set as a value between them.

The flow chart of the proposed methods in this section can be summarized as Fig. 1.

## 5 Experiments

The thresholds $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ are learned as $(1.1, 1, 3.8)$ by the $k$-means. We used them below unless otherwise stated. The notations $\mathbf{K}, \mathbf{R}, \mathbf{t}$ are as in (3).

### 5.1 Simulations

The stabilities of Algorithms 1, 2, 3 were tested. The details are reported in this subsection. Since the space points and their image points are in a relative position, the noise in the space points can be regarded as having been transferred to the image points. Thus, we only consider noise on the image points in the following simulations.

Testing the stability of Algorithm 1 is equivalent to testing the stability of the evaluation functions $I_{\text{tc}}$ and $I_{\text{general}}$. Extensive experiments performed show that the evaluation functions $I_{\text{tc}}$ and $I_{\text{general}}$ are very stable. Six samples are shown below, whose data are denoted as $D_i, i = 1..6$ with image sizes not greater than $1000 \times 1000$ pixels.

The world coordinate system was taken as the camera coordinate system and the simulated camera intrinsic parameter matrix was taken as:

$$\mathbf{K} = \begin{pmatrix} 1000 & 0 & 512 \\ 0 & 900 & 384 \\ 0 & 0 & 1 \end{pmatrix}.$$

Different seven-point groups of space points $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}, \mathbf{7}$ were generated, where $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}, \mathbf{O}$ do not lie on a twisted cubic, but $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{7}, \mathbf{O}$ do lie. By $\mathbf{K}$ and these space points, different views were produced.

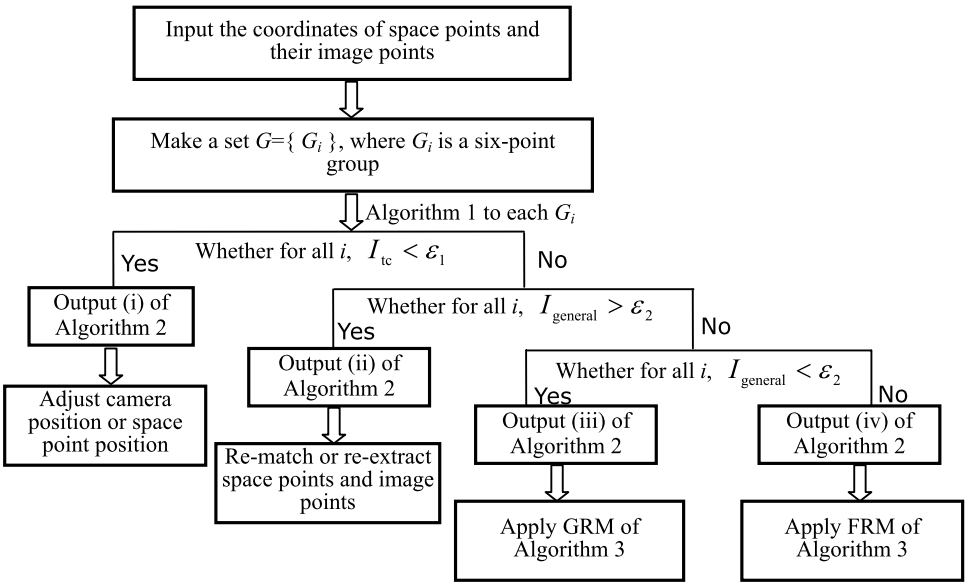



**Fig. 1** Flow chart of the detections and the corresponding solutions

Gaussian noise with 0 mean and standard deviation ranging from 0 to 2 pixels was directly added to each image point. Then in each view, the following values were computed:

$\text{inv}_1$ : the value of $I_{\text{tc}}$ on $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{7}, \mathbf{m}_i, i = 1..5, \mathbf{m}_7$;

$\text{inv}_2$ : the value of $I_{\text{tc}}$ on $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}, \mathbf{m}_i, i = 1..6$;

$\text{inv}_3$ : the value of $I_{\text{general}}$ on $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}, \mathbf{m}_i, i = 1..6$;

$\text{inv}_4$ : the value of $I_{\text{general}}$ on $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}, \mathbf{m}_i, i = 1..5$, $\mathbf{m}_6 + (30, 20, 0)^T$;

$\text{inv}_5$ : the value of $I_{\text{general}}$ on $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}, \mathbf{m}_i, i = 1..5$, $\mathbf{m}_6 + (70, 80, 0)^T$.

Under each noise level, we performed 100 runs, and the averaged results are shown in Table 1. We can see that each $\text{inv}_i$ is stable. Also $\text{inv}_1$ is indeed quite small, which is consistent with that $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{7}, \mathbf{O}$ lie on a twisted cubic. While $\text{inv}_2$ is not, which is consistent with that $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}, \mathbf{O}$ do not lie on a twisted cubic. The threshold $\varepsilon_1$ is 1.1 and there are always $\text{inv}_1 < 1.1, \text{inv}_2 > 1.1$. Moreover, we see that $\text{inv}_3$ is small, whereas $\text{inv}_4, \text{inv}_5$ are not. $\varepsilon_2$ is 1, there are always $\text{inv}_3 < 1, \text{inv}_4 > 1, \text{inv}_5 > 1$ except only $\text{inv}_4$ in $D_1$. If we take $\varepsilon_2$ as 0.38, the discriminations are all successful. The variances of the computations were also calculated showing that all of them are smaller than 0.1 and only one is 0.12 for $\text{inv}_1, \text{inv}_2, \text{inv}_3$, the variances for $\text{inv}_4$ are all smaller than 3.1, and the variances for $\text{inv}_5$ are all smaller than 5.1. In all of the extensive experiments, we find using $\varepsilon_2 = 1$ failed at only one time as shown in $\text{inv}_4$ of $D_1$.

Algorithms 2, 3 have been tested too. Repeated experiments show similar results. One experimental sample is presented below.

Ten space points $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}, \mathbf{7}, \mathbf{8}, \mathbf{9}, \mathbf{10}$, of which $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}$ and the camera optical center $\mathbf{O}$ lie on a twisted cubic but other space points do not, were generated and then projected onto the simulated image plane by the camera parameters:

$$\mathbf{K} = \begin{pmatrix} 1000 & 0.8 & 512 \\ 0 & 900 & 384 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{R} = \begin{pmatrix} 0.9330 & -0.3536 & 0.0670 \\ 0.3536 & 0.8660 & -0.3536 \\ 0.0670 & 0.3536 & 0.9330 \end{pmatrix},$$

$$\mathbf{t} = \begin{pmatrix} -0.9330 \\ -0.3536 \\ -20.0670 \end{pmatrix}.$$

The image size is $500 \times 500$ pixels. Gaussian noise with 0 mean and standard deviation ranging from 0 to 2 pixels was directly added to each image point. Then the following disturbances were added:

(i) $\mathbf{m}_{10}$ is disturbed by (30, 20);
(ii) $\mathbf{m}_{10}$ is disturbed by (70, 80);
(iii) $\mathbf{m}_9$ is disturbed by (30, 20) and $\mathbf{m}_{10}$ is disturbed by (40, 30);

**Table 1** The $\text{inv}_i$ values under different noise levels (pixel)

| Noise level | | 0 | 0.5 | 1.0 | 1.5 | 2.0 |
|---|---|---|---|---|---|---|
| $D_1$ | $\text{inv}_1$ | 0.0000 | 0.0003 | 0.0011 | 0.0026 | 0.0047 |
| | $\text{inv}_2$ | 2.4098 | 2.4102 | 2.4102 | 2.4098 | 2.4108 |
| | $\text{inv}_3$ | 0.0000 | 0.0136 | 0.0437 | 0.1057 | 0.1592 |
| | $\text{inv}_4$ | 0.3927 | 0.4091 | 0.4343 | 0.4893 | 0.5642 |
| | $\text{inv}_5$ | 13.1137 | 13.1235 | 13.0734 | 12.8677 | 12.9511 |
| $D_2$ | $\text{inv}_1$ | 0.0000 | 0.0005 | 0.0018 | 0.0044 | 0.0065 |
| | $\text{inv}_2$ | 2.7224 | 2.7231 | 2.7196 | 2.7176 | 2.7129 |
| | $\text{inv}_3$ | 0.0000 | 0.0120 | 0.0411 | 0.0826 | 0.1688 |
| | $\text{inv}_4$ | 7.4103 | 7.4677 | 7.4015 | 7.1962 | 7.6928 |
| | $\text{inv}_5$ | 36.8287 | 36.9302 | 36.7053 | 36.1357 | 37.1052 |
| $D_3$ | $\text{inv}_1$ | 0.0000 | 0.0021 | 0.0073 | 0.0200 | 0.0582 |
| | $\text{inv}_2$ | 2.9356 | 2.9359 | 2.9351 | 2.9351 | 2.9384 |
| | $\text{inv}_3$ | 0.0000 | 0.0054 | 0.0172 | 0.0580 | 0.0996 |
| | $\text{inv}_4$ | 9.6281 | 9.5428 | 9.6578 | 9.8161 | 9.3806 |
| | $\text{inv}_5$ | 90.6676 | 90.5046 | 90.7381 | 90.6296 | 90.4451 |
| $D_4$ | $\text{inv}_1$ | 0.0000 | 0.0001 | 0.0005 | 0.0014 | 0.0018 |
| | $\text{inv}_2$ | 2.6865 | 2.6867 | 2.6865 | 2.6865 | 2.6881 |
| | $\text{inv}_3$ | 0.0000 | 0.0221 | 0.0973 | 0.1663 | 0.3795 |
| | inv4 | 5.2267 | 5.2556 | 5.4291 | 5.4783 | 6.2405 |
| | $\text{inv}_5$ | 6.3549 | 6.4018 | 6.6079 | 6.7081 | 7.6497 |
| $D_5$ | $\text{inv}_1$ | 0.0000 | 0.0003 | 0.0009 | 0.0019 | 0.0045 |
| | $\text{inv}_2$ | 1.6501 | 1.6504 | 1.6517 | 1.6488 | 1.6523 |
| | $\text{inv}_3$ | 0.0000 | 0.0014 | 0.0068 | 0.0124 | 0.0304 |
| | $\text{inv}_4$ | 1.1369 | 1.1285 | 1.1380 | 1.1032 | 1.2051 |
| | $\text{inv}_5$ | 11.2435 | 11.2339 | 11.3106 | 11.1331 | 11.4879 |
| $D_6$ | $\text{inv}_1$ | 0.0000 | 0.0004 | 0.0014 | 0.0037 | 0.0056 |
| | $\text{inv}_2$ | 2.3361 | 2.3362 | 2.3356 | 2.3363 | 2.3350 |
| | $\text{inv}_3$ | 0.0000 | 0.0242 | 0.0977 | 0.2313 | 0.3575 |
| | $\text{inv}_4$ | 2.8085 | 2.7908 | 2.8073 | 3.1459 | 2.8030 |
| | $\text{inv}_5$ | 8.9641 | 8.8869 | 8.9528 | 9.1748 | 8.9284 |

(iv) $\mathbf{m}_9$ is disturbed by (30, 20) and $\mathbf{m}_{10}$ is disturbed by (70, 80);
(v) $\mathbf{m}_9$ is disturbed by (60, 50) and $\mathbf{m}_{10}$ is disturbed by (70, 80).

Under each noise level and each disturbance, we performed 100 runs of Algorithms 2, 3, and calculated the number of the runs that failed to remove the above disturbed pairs or that made the remaining space points and $\mathbf{O}$ lie on a degenerate configuration. The results are shown in Table 2. Table 2 demonstrates that the FRM is much more stable than the GRM and the success rate of the FRM is higher than that of the GRM.

## 5.2 Experiments with Real Data

### *5.2.1 Experiment on a Grid Object*

The used images were taken by a NIKON COOLPIX990 camera. A sample image of a calibration object with size $1024 \times 768$ pixels is shown in Fig. 2. Canny edge detector was used to extract the lines of the image and then their intersection points were calculated to get the image points. We

set up the world coordinate system in the grid and matched the space points with the image points. We obtained 108 matching pairs.

Applying Algorithm 2 to these 108 pairs, we obtained the conclusion that the input data are reliable for camera parameter estimation. Then we estimated the camera parameters from these 108 pairs, with the results given below:

**Fig. 2** An image of a calibration grid

$$\mathbf{K} = \begin{pmatrix} 2049.8128 & -2.7983 & 523.9202 \\ 0 & 2050.5605 & 394.1385 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{R} = \begin{pmatrix} 0.7784 & -0.6272 & 0.0270 \\ -0.2648 & -0.3671 & -0.8917 \\ 0.5692 & 0.6870 & -0.4518 \end{pmatrix},$$

$$\mathbf{t} = \begin{pmatrix} -0.7503 \\ 4.8624 \\ 30.7296 \end{pmatrix}.$$

Furthermore, from the 108 pairs we combined 125 six-point groups and calculated the corresponding values of $I_{\text{general}}$ and $I_{\text{tc}}$. The values of $I_{\text{general}}$ are shown in Fig. 3a demonstrating that each group is consistent with a projective matrix. The values of $I_{\text{tc}}$ are shown in Fig. 3b demonstrating that there exist groups whose space points and the camera lie on a degenerate configuration.

We chose the group with the minimal value of $I_{\text{tc}}$ and the group with the maximal value of $I_{\text{tc}}$, and denoted them as $G_{\min}, G_{\max}$ respectively. The values of $I_{\text{tc}}$ from $G_{\min}$ and

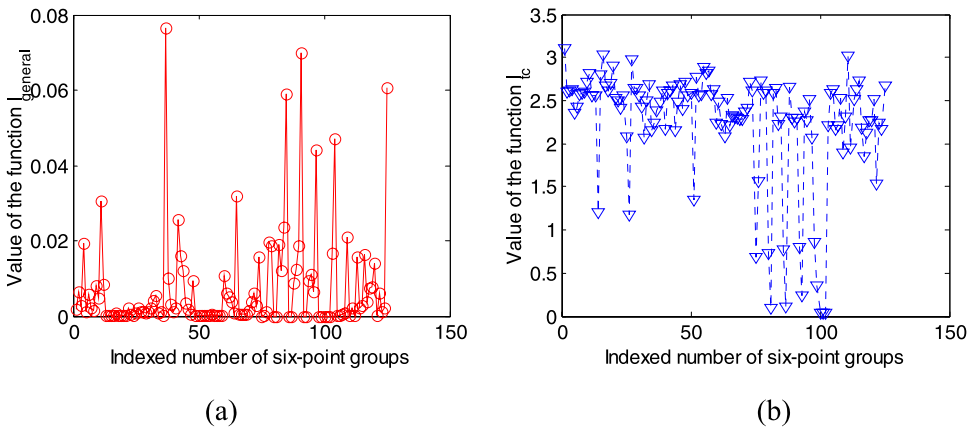
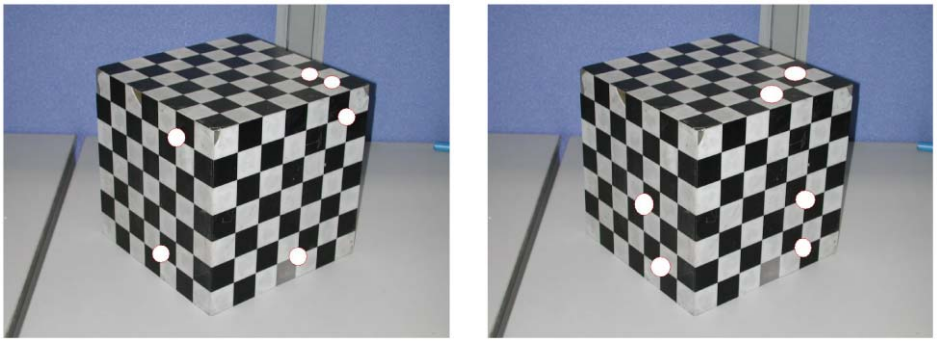



**Fig. 3** The values of the evaluation function **a** $I_{\text{general}}$; **b** $I_{\text{tc}}$

**Table 2** The failure rate

| Noise level(pixel) | | 0 | 0.5 | 1.0 | 1.5 | 2.0 |
|---|---|---|---|---|---|---|
| Disturbance (i) | FRM | 0 | 0 | 0 | 0 | 2 |
| | GRM | 0 | 3 | 4 | 2 | 4 |
| Disturbance (ii) | FRM | 0 | 0 | 0 | 0 | 0 |
| | GRM | 0 | 5 | 3 | 3 | 3 |
| Disturbance (iii) | FRM | 0 | 0 | 0 | 2 | 0 |
| | GRM | 0 | 3 | 9 | 12 | 7 |
| Disturbance (iv) | FRM | 0 | 0 | 0 | 0 | 2 |
| | GRM | 0 | 5 | 15 | 7 | 16 |
| Disturbance (v) | FRM | 0 | 0 | 0 | 0 | 3 |
| | GRM | 0 | 10 | 8 | 17 | 12 |



(a) (b)

**Fig. 4** Six image points, denoted in white color, **a** from the group with the minimal value of $I_{tc}$; **b** from the group with the maximal value of $I_{tc}$

$G_{max}$ are 0.0380 and 3.0997 respectively. We show the image points from $G_{min}$ in Fig. 4a and the image points from $G_{max}$ in Fig. 4b. For both of these two groups, no four space points are coplanar and no three image points are collinear.

We calibrated the camera parameters from $G_{min}$, with the results:

$$\mathbf{K}_1 = \begin{pmatrix} 980.4078 & 26.8782 & 430.9372 \\ 0 & 870.5114 & 541.8497 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{R}_1 = \begin{pmatrix} -0.6666 & 0.7454 & 0.0062 \\ -0.0205 & -0.0266 & 0.9994 \\ 0.7451 & 0.6661 & 0.0330 \end{pmatrix},$$

$$\mathbf{t}_1 = \begin{pmatrix} -1.0457 \\ -1.5375 \\ -18.0317 \end{pmatrix}.$$

Also, we calibrated the camera parameters from $G_{max}$, with the results:

$$\mathbf{K}_2 = \begin{pmatrix} 2140.9987 & -2.1069 & 570.3262 \\ 0 & 2138.6413 & 452.6338 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{R}_2 = \begin{pmatrix} 0.7668 & -0.6409 & 0.0358 \\ -0.2772 & -0.3808 & -0.8822 \\ 0.5790 & 0.6665 & -0.4696 \end{pmatrix},$$

$$\mathbf{t}_2 = \begin{pmatrix} -1.4424 \\ 3.9693 \\ 32.1664 \end{pmatrix}.$$

These estimations are evaluated by comparing them with the above $\mathbf{K}$, $\mathbf{R}$, $\mathbf{t}$:

$$\mathbf{K}_1 - \mathbf{K} = \begin{pmatrix} -1069.4050 & 29.6765 & -92.9830 \\ 0 & -1180.0491 & 147.7112 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{K}_2 - \mathbf{K} = \begin{pmatrix} 91.1859 & 0.6914 & 46.4060 \\ 0 & 88.0809 & 58.4953 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{R}_1 - \mathbf{R} = \begin{pmatrix} -1.4450 & 1.3725 & -0.0208 \\ 0.2443 & 0.3405 & 1.8911 \\ 0.1759 & -0.0208 & 0.4848 \end{pmatrix},$$

$$\mathbf{R}_2 - \mathbf{R} = \begin{pmatrix} -0.0116 & -0.0137 & 0.0087 \\ -0.0123 & -0.0137 & 0.0095 \\ 0.0098 & -0.0205 & -0.0178 \end{pmatrix},$$

$$\mathbf{T}_1 - \mathbf{T} = \begin{pmatrix} -0.2954 \\ -6.3999 \\ -48.7613 \end{pmatrix}, \qquad \mathbf{T}_2 - \mathbf{T} = \begin{pmatrix} -0.6922 \\ -0.8931 \\ 1.4368 \end{pmatrix}.$$

We observe that the estimation from $G_{max}$ is much better than the estimation from $G_{min}$.

The test results for Algorithms 2, 3 are reported below.

Among the 108 pairs of the space points and their image points, we deliberately mismatched 12, 22, 36, 54 pairs by matching the space points with their projected next image points. Then for each of these cases, Algorithm 2 was applied and we obtained the conclusion (iv) of the output. So FRM of Algorithm 3 was applied following. For all the cases with 12, 22, 36, 54 mismatched pairs, all of the mismatched pairs were removed successfully by FRM and none of the true matched pairs was removed. The calibration results after removing the mismatched pairs for these cases are respectively:

$$\mathbf{K} = \begin{pmatrix} 2051.7 & -2.7 & 521.0 \\ 0 & 2052.6 & 394.5 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{R} = \begin{pmatrix} 0.7793 & -0.6261 & 0.0263 \\ -0.2648 & -0.3671 & -0.8917 \\ 0.5680 & 0.6879 & -0.4519 \end{pmatrix},$$

$$\mathbf{t} = \begin{pmatrix} -0.7066 \\ 4.8568 \\ 30.7588 \end{pmatrix},$$

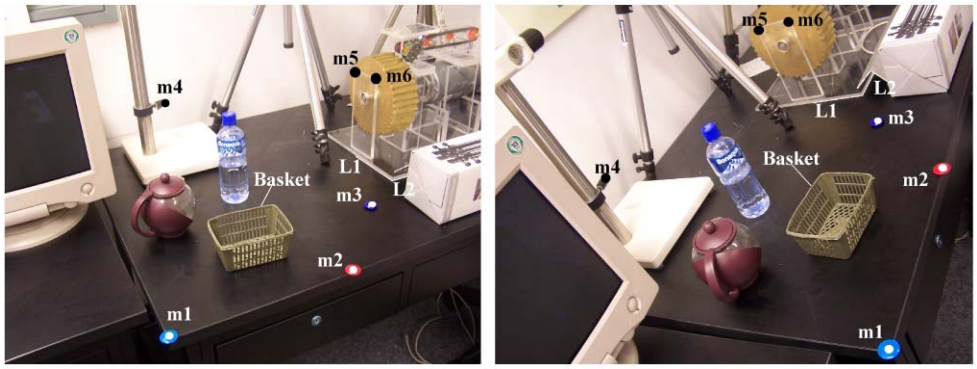



**Fig. 5** Two images of a real scene

$$\mathbf{K} = \begin{pmatrix} 2052.4 & -2.3 & 522.6 \\ 0 & 2053.7 & 395.8 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{R} = \begin{pmatrix} 0.7788 & -0.6267 & 0.0269 \\ -0.2650 & -0.3676 & -0.8914 \\ 0.5686 & 0.6871 & -0.4524 \end{pmatrix},$$

$$\mathbf{t} = \begin{pmatrix} -0.7322 \\ 4.8382 \\ 30.7746 \end{pmatrix},$$

$$\mathbf{K} = \begin{pmatrix} 2049.4 & -2.7 & 524.9 \\ 0 & 2051.2 & 387.6 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{R} = \begin{pmatrix} 0.7782 & -0.6275 & 0.0271 \\ -0.2635 & -0.3654 & -0.8928 \\ 0.5701 & 0.6876 & -0.4497 \end{pmatrix},$$

$$\mathbf{t} = \begin{pmatrix} -0.7650 \\ 4.9617 \\ 30.7124 \end{pmatrix},$$

$$\mathbf{K} = \begin{pmatrix} 2057.3 & -2.4 & 523.4 \\ 0 & 2059.5 & 391.4 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{R} = \begin{pmatrix} 0.7786 & -0.6270 & 0.0270 \\ -0.2645 & -0.3668 & -0.8919 \\ 0.5691 & 0.6872 & -0.4515 \end{pmatrix},$$

$$\mathbf{t} = \begin{pmatrix} -0.7421 \\ 4.9037 \\ 30.8510 \end{pmatrix}.$$

On the other hand, by GRM of Algorithm 3, only for the cases with 12, 22 mismatched pairs, all of the mismatched pairs were removed. At the time, we obtained the same calibration results from the FRM and GRM. For the case with 36 mismatched pairs by GRM, 99 pairs were removed and solving for the camera projective matrix from the remaining pairs gave zero solution of the camera projective matrix. For the case with 54 mismatched pairs, 8 pairs were mistakenly removed by GRM and from the remaining data the calibration result is not better than that by FRM.

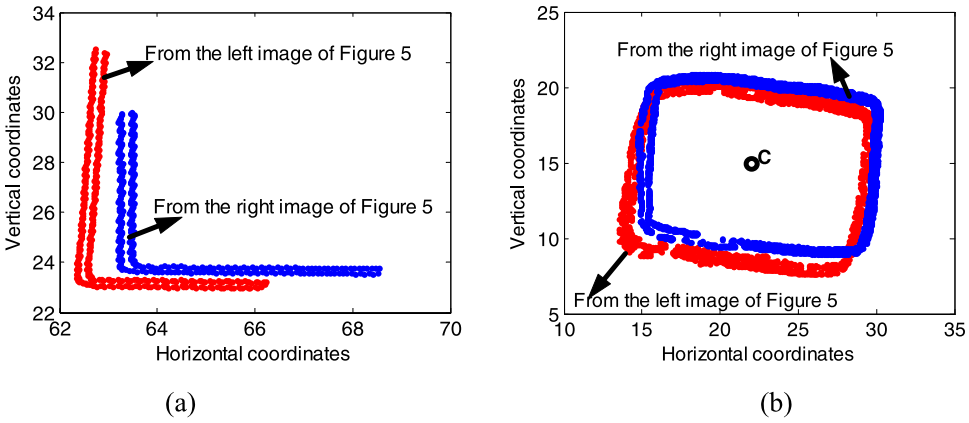
We did not mismatch but disturbed the above 12, 22, 36, 54 image points by noise $(8, 6)$. For the cases with 12, 22 disturbed image points, we had the same results as above. For the case with 36 disturbed image points, still 99 pairs were removed and the obtained solution of the camera projective matrix was still zero by GRM; by FRM only 14 pairs were mis-removed. For the case with 54 disturbed image points, just the 54 disturbed image points were removed by FRM, whereas 72 image points were removed by GRM. If we disturbed image points by larger noise than $(8, 6)$, all of the pairs can be successfully removed by both FRM and GRM.

These results show the efficiency of Algorithms 2, 3. They also demonstrate the importance and necessity of the developed filtering method by the evaluation functions in Algorithm 3.

### 5.2.2 *Experiment on Real Scenes*

We also performed experiment from images of a real scene. The used two images were captured by a KODA 6490 camera and are shown in Fig. 5, where $\mathbf{m}_i, i = 1..6$ in each are from the same six space points. The sizes of the images are $2304 \times 1728$ pixels. The world coordinate system was set up as: the line through the space points of $\mathbf{m}_1, \mathbf{m}_2$ is the $X$-axis, the line of the edge of the tabletop perpendicular to the $X$-axis is the $Y$-axis, and the line perpendicular to the tabletop is the $Z$-axis.

For each view, from $\mathbf{m}_i, i = 1..6$ and their space points, we computed the value of $I_{\text{tc}}$. The result for the left image is 0.4103, while that for the right image is 1.6308. The value 0.4103 is smaller than the threshold 1.1. So, the output of Algorithm 1 for the left image is that the space points and their image points are unreliable. To the right image, Step 2 of Algorithm 1 was continually applied and the obtained value of $I_{\text{general}}$ is 0.0084 lower than the threshold 1. Therefore lastly, the output of Algorithm 1 for the right image is that




(a) (b)

**Fig. 6** Reconstruction results, where the points in red are from the left image of Fig. 5 and the points in blue are from the right image of Fig. 5: **a** Reconstructions of two vertical lines; **b** Reconstructions of the upper edge of a basket

the space points and their image points are reliable for the camera parameter estimation.

To verify whether the detections are valid or not, we estimated the camera parameters and used them to perform reconstruction. The camera parameters estimated from the left image are:

$$\mathbf{K}_1 = \begin{pmatrix} 3609.4854 & -553.1948 & 826.7216 \\ 0 & 2993.1786 & -1000.7742 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{R}_1 = \begin{pmatrix} 0.9253 & -0.3632 & -0.1093 \\ -0.0857 & 0.0807 & -0.9931 \\ 0.3695 & 0.9282 & 0.0436 \end{pmatrix},$$

$$\mathbf{T}_1 = \begin{pmatrix} 15.1853 \\ 106.8644 \\ 123.0356 \end{pmatrix},$$

$$\mathbf{P}_1 = \begin{pmatrix} 3599.3259 & -573.2380 & 185.8930 & 94948.6774 \\ -610.3517 & -670.0271 & -2939.7460 & 191761.5251 \\ 0.3601 & 0.9048 & 0.0425 & 119.9262 \end{pmatrix}.$$

The estimations for camera parameters from the right image are:

$$\mathbf{K}_2 = \begin{pmatrix} 1926.4461 & -111.5432 & 1576.3127 \\ 0 & 1719.4163 & 1076.0902 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{R}_2 = \begin{pmatrix} 0.2894 & -0.8771 & -0.3834 \\ -0.8304 & -0.0309 & -0.5562 \\ 0.4760 & 0.4794 & -0.7373 \end{pmatrix},$$

$$\mathbf{T}_2 = \begin{pmatrix} 9.9875 \\ 18.1978 \\ 53.1970 \end{pmatrix},$$

$$\mathbf{P}_2 = \begin{pmatrix} 1380.0008 & -916.9166 & -1811.6972 & 99581.3631 \\ -902.1840 & 455.8733 & -1724.1366 & 87234.2112 \\ 0.4690 & 0.4723 & -0.7265 & 52.4157 \end{pmatrix}.$$

From $\mathbf{K}_1$, we see that the principal point is at (826.7216, −1000.7742), which is outside the image region and thus is unreasonable. Two vertical lines and the upper edge of a basket from each view were also reconstructed by the corresponding camera parameters estimated. The images of the two vertical lines are shown as $L_1, L_2$ and the basket images are also denoted in Fig. 5. Back projecting the points on $L_1, L_2$ in each view to the world $X$–$Y$ plane by the corresponding camera projective matrix estimated above and back projecting the image points on the upper edge of the basket image in each view to the world plane $Z = 6.8$ by the corresponding camera projective matrix estimated above, we obtained their reconstructions, where 6.8 is the prior known height of the basket under the same unit as the space points of $\mathbf{m}_i, i = 1..6$. The reconstruction results are demonstrated in Fig. 6, where the points in red are the reconstructions from the left image of Fig. 5 by $\mathbf{P}_1$, and the points in blue are the reconstructions from the right image of Fig. 5 by $\mathbf{P}_2$. In Fig. 6a, we can see that the two lines reconstructed from the left image of Fig. 5 are clearly not vertical. In Fig. 6b, we plot the ground truth of the center of the basket upper edge as point **C**. We observe that **C** is very close to the center of the edge reconstructed from the right image of Fig. 5 but far away from that reconstructed from the left image of Fig. 5.

This experiment validates Algorithm 1. The detected unreliable space points and image points indeed fail to give good estimation for the camera parameters, which adversely affect the subsequent shape reconstruction. Algorithm 1 is the basis of all algorithms and its validity also ensures the validities of Algorithms 2 and 3.

## 6 Conclusions

That DLT method may give unreliable estimations for camera parameters confuses people long time. We give the detection method for the unreliability and provide the solutions of the detection results. The detection method is based on

two evaluation functions constructed in a complete framework of invariance for six points. Not involving camera optical center or projective matrix and independent of point orders, these two evaluation functions make the detection method stable and flexible. In the solutions for the detection results, a filtering RANAC is proposed to filter the detected unreliable points before a general RANSAC is applied. The filtering is important and necessary for a successful RANSAC.

**Acknowledgements** This work was supported by a grant from the Research Grants Council of Hong Kong [Project No. CityU117106] and supported by the National Key Basic Research and Development Program (973) under Grant No. 2004CB318107 and the National Natural Science Foundation of China under Grant No. 60633070.

## References

Abdel-Aziz, Y. I., & Karara, H. M. (1971). Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. In *Proceedings of the ASP/UI symposium on close range photogrammetry* (pp. 1–18).

Bayro-Corrochano, E., & Banarer, V. (2002). A geometric approach for the theory and applications of 3D projective invariants. *Journal of Mathematical Imaging and Vision*, *16*(2), 131–154.

Buchanan, T. (1992). *The projection of six points from 2-space and 3-space*. Preprint, Eberstadt, Troyesstr. 78, D-6100 Darmstadt, Germany. http://www.iks.hs-merseburg.de/~buchanan/pubs/Coble.pdf.

Carlsson, S. (1995). *View variation and linear invariants in 2-D and 3-D* (Tech. Rep. ISRN KTH/NA/P–95/22–SE), December 1995.

Carlsson, S. (1998). Symmetry in perspective. In *ECCV* (pp. 249–263).

Coble, A. B. (1922). Associated sets of points. *Transactions of the American Mathematical Society*, *24*(1), 1–20.

Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern classification*. (2nd ed.). New York: Wiley-Interscience.

Förstner, W. (1987). Reliability analysis of parameter estimation in linear models with applications to mensuration problems in computer vision. *Computer Vision, Graphics, and Image Processing*, *40*(3), 273–310.

Hartley, R., & Zisserman, A. (2000). *Multiple view geometry in computer vision*. Cambridge: Cambridge University Press.

Kahl, F., & Henrion, D. (2007). Globally optimal estimates for geometric reconstruction problems. *International Journal of Computer Vision*, *74*(1), 3–15.

Quan, L. (1995). Invariants of six points and projective reconstruction from three uncalibrated images. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, *17*(1), 34–46.

Semple, J., & Kneebone, G. (1952). *Algebraic projective geometry*. Oxford: Oxford University Press.

Slama, C. (1980). *Manual of photogrammetry* (4th ed.). Falls Church: American Society of Photogrammetry.

Sutherland, I. E. (1963). *Sketchpad: a man-machine graphical communications system* (Technical report 296). MIT Lincoln Laboratory. Also published by Garland publishing, New York (1980).

Weng, J., Huang, T. S., & Ahuja, N. (1989). Motion and structure from two perspective views: algorithms, error analysis, and error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*(5), 451–476.

Wu, Y. H. (2001). *Bracket algebra, affine bracket algebra, and automated geometric theorem proving*. PhD dissertation, Institute of Systems Science, Chinese Academy of Sciences, Beijing.

Wu, Y. H., & Hu, Z. Y. (2003). The invariant representations of a quadric cone and a twisted cubic. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, *25*(10), 1329–1332.

Wu, Y. H., & Hu, Z. Y. (2005). A unified and complete framework of invariance for six points. In H. Li, P. J. Olver, & G. Sommer (Eds.), *Lecture notes in computer science: Vol. 3519. Computer algebra and geometric algebra with applications* (pp. 403–417). Berlin: Springer.

Wu, Y. H., & Hu, Z. Y. (2006). A robust method to recognize critical configuration for camera calibration. *Image and Vision Computing*, *24*(12), 1313–1318.