Optical Engineering

OpticalEngineering.SPIEDigitalLibrary.org

Pose estimation method for planar mirror based on one-dimensional target

Wenlei Liu
Sentang Wu
Xiaolong Wu

SPIE.

Wenlei Liu, Sentang Wu, Xiaolong Wu, "Pose estimation method for planar mirror based on one-dimensional target," *Opt. Eng.* **57**(7), 073101 (2018), doi: 10.1117/1.OE.57.7.073101.

Pose estimation method for planar mirror based on one-dimensional target

Wenlei Liu,[a,*] Sentang Wu,[a] and Xiaolong Wu[b]
[a]Beihang University, School of Automation Science and Electrical Engineering, Beijing, China
[b]Navigation and Control Technology Institute of NORINCO Group, Beijing, China

Abstract. This paper presents a pose estimation method for planar mirror based on a one-dimensional (1-D) target. First, a 1-D target is maneuvered in front of a camera so as to allow the camera to capture several images of both the original and virtual feature points on the target. Next, the normal vector of the mirror and the distance between the mirror and the camera are calculated by a closed-form method. Last, a nonlinear optimization is used to refine the closed-form solution by minimizing the reprojection error of reflection projection. Synthetic data and real experiments show that the proposed method is accurate and reliable. The proposed method is simple and flexible. No prior information about the target motion and camera pose is assumed. A minimum of two nonparallel target positions are required. It is suitable for the pose estimation of planar mirrors, especially when they are in fixed positions. We can find the posture of the fixed camera according to the normal vector of the plane mirror. [DOI: 10.1117/1.OE.57.7.073101]

Keywords: reflection; calibration; mirrors; estimation; camera.

Paper 180459 received Mar. 28, 2018; accepted for publication Jun. 12, 2018; published online Jul. 5, 2018.

1 Introduction

Vision sensors have advantages of flexibility and noncontact measurement. Planar mirrors are utilized in a wide variety of applications, such as fringe-reflection measurement,[1] planar catadioptric vision sensors,[2] stereovision by a single camera, and calibration of multiple cameras with nonoverlapping fields of view (FOVs).[3,4] A common major challenge of these applications is the pose estimation of the planar mirror. It is important and necessary to develop accurate and flexible pose estimation methods for planar mirrors.

There has been some research on the pose estimation of planar mirrors. Lebraly et al.[5] used a planar mirror to create an overlap between cameras by attaching markers on the mirror's surface. However, pasting markers on the mirror was inconvenient and troublesome. Sturm et al.,[6] Rodrigues et al.,[7] Hesch et al.,[8] and Kosuke Takahashi et al.[9] proposed different pose estimation methods based on the fact that successive reflections in two-mirror planes were a fixed-axis rotation around the line of intersection of two mirror planes. Reference 10 proposed a practical visual inertial calibration method, which is based on visual observations taken in front of a planar mirror. Reference 11 presented a design and construction of a new system for stereovision with a single CCD camera. Reference 12 proposed a method for the calibration, which is used the orthogonal phase-shifted sinusoidal patterns during the calibration process.

Xiao et al.[1] calculated the transformation between the camera and the LCD screen by a markerless flat mirror. Li et al.[13] measured the pose of a flat mirror by capturing the reflected structured light pattern. The translation vector was calculated by the mirror motion without any tilt. However, pure translations can hardly be achieved without external equipment. Ren et al.[14] calibrated a stereodeflectometry system, where two cameras captured the fringe patterns produced by an LCD screen reflected by a flat mirror. The above methods require that the mirror is placed at several positions. However, these methods fail when the camera and the mirror are in fixed positions.

In Ref. 15, an algorithm is based on an orthogonality constraint, which should be satisfied by all families of reflections of a single reference object. It can return a unique solution of three P3P problems from three mirrored images. Reference 16 presents a method that utilizes reasonable assumptions about geometric features such as a plane or an intersecting line of two planes. It obtains an initial solution via singular value decomposition (SVD), which is then refined via nonlinear optimization, which does not require any assumption about structures but estimates the parameters of geometric features. In Ref. 17, an accurate closed-form solution with explicit geometric meanings is proposed. It minimizes an error metric based on all reflections of rotation, and then derives the optimal solution that gives the real camera rotation directly by SVD of the sum of mirrored camera rotations. Then, it can minimize object-space collinearity error.

This paper is also inspired by Shah[18] and Li et al.[17] Shah found the best rotation matrix with determinant 1. In contrast, we seek the best orthogonal matrix with determinant -1, and the best reflection matrix can be constructed by a matrix with determinant -1. Li et al.[17] calculated the camera rotation by minimizing the error of all rotation reflections. In this paper, we initialize the normal vector by minimizing the error of reflections of the unit direction vectors.

In this paper, we focus on the pose estimation of the planar mirror based on a one-dimensional (1-D) target. 1-D target has advantages of simple structure and flexibility. First, instead of moving the mirror, we move the target in

*Address all correspondence to: Wenlei Liu, E-mail: liuwenlei@buaa.edu.cn

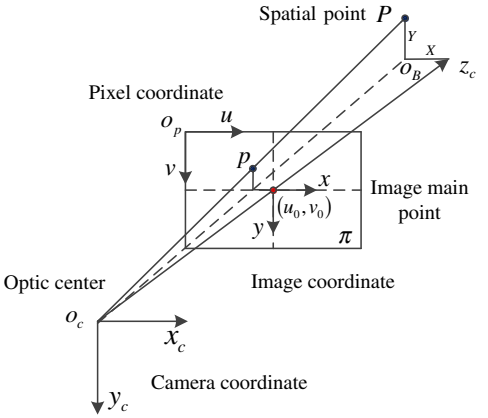

Fig. 1 The perspective projection model and coordinate relation.

front of the camera to provide multiple (at least two) views of both the original and virtual targets. The camera and the mirror are fixed during the process. Next, the normal vector of the planar mirror is estimated based on the direction vectors of the original and virtual targets. The distance between the camera and the mirror is calculated based on reflection equations. Last, the closed-form solution thus obtained is refined using nonlinear optimization by minimizing the reprojection error of reflection projection. Compared with conventional methods, the proposed method is suitable for the pose estimation of planar mirrors, especially when camera and the mirror are in fixed positions. After obtaining the position information in the plane mirror, the camera can be observed the carrier virtual image through plane mirror, then the rotation matrix, and translation vector of the image machine coordinate system $\{\mathbf{C}\}$ and the carrier coordinate system $\{\mathbf{B}\}$ can be calculated. The calibration of the installation parameters of the image machine is completed.

The rest of this paper is organized as follows: preliminary work is introduced in Sec. 2. The principle of the proposed method is described in Sec. 3. Synthetic data and real-data experiments are performed in Sec. 4. The conclusions are given in Sec. 5.

2 Preliminaries

2.1 *Camera Model*

Perspective projection is used as the camera model. The parameters and coordinates of the perspective projection model are shown in Fig. 1. The relation between a spatial point $\mathbf{P} = [X, Y, Z]^T$ in $\{\mathbf{C}\}$ and the corresponding image point $\mathbf{p} = [u, v]^T$ can be described as follows:

$$s\begin{bmatrix}\mathbf{p}\\1\end{bmatrix} = \mathbf{KP}, \quad \mathbf{K} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{1}$$

where $\mathbf{s}$ is a scale factor, $\mathbf{K}$ is the intrinsic matrix in camera. $\mathbf{f_x}$ and $\mathbf{f_y}$ are the equivalent focal lengths in horizontal and vertical directions, respectively. (u_0, v_0) denotes the principal point.

As shown in Fig. 2(a), both the original and the virtual targets are in the camera's FOV. $O_c x_c y_c z_c$ is the coordinate frame of the camera, denoted as $\{\mathbf{C}\}$.

2.2 *Pose Estimation of One-Dimensional Target*

As shown in Fig. 2, $\mathbf{P}_{i,j}$ and $\tilde{\mathbf{P}}_{i,j}$ denote the spatial coordinates of the j'th original and the j'th virtual feature points at the i'th position in $\{\mathbf{C}\}$, respectively. $\mathbf{p}_{i,j}$ and $\tilde{\mathbf{p}}_{i,j}$ are the corresponding image coordinates, $i = 1, \ldots, M$, $j = 1, \ldots, N$. M is the number of target positions and N is the number of feature points on the target.

The original target and the virtual target at i'th position are uniquely characterized by $\mathbf{P_{i,1}}$, $\mathbf{d_i}$ and $\tilde{\mathbf{P}}_{i,1}, \tilde{\mathbf{d}}_i$, respectively. $\mathbf{d_i}$ and $\tilde{\mathbf{d}}_i$ are the unit direction vectors of the original and virtual targets at ith position, respectively.

For the ith position, $\mathbf{P_{i,1}}$ and $\mathbf{d_i}$ can be obtained by the method proposed by Liu et al.[19,20] It requires that at least three collinear feature points on the target. Suppose $\mathbf{P}_{i,1}(x_1, y_1, z_1)$, $\mathbf{P}_{i,2}(x_2, y_2, z_2)$, and $\mathbf{P}_{i,3}(x_3, y_3, z_3)$, then

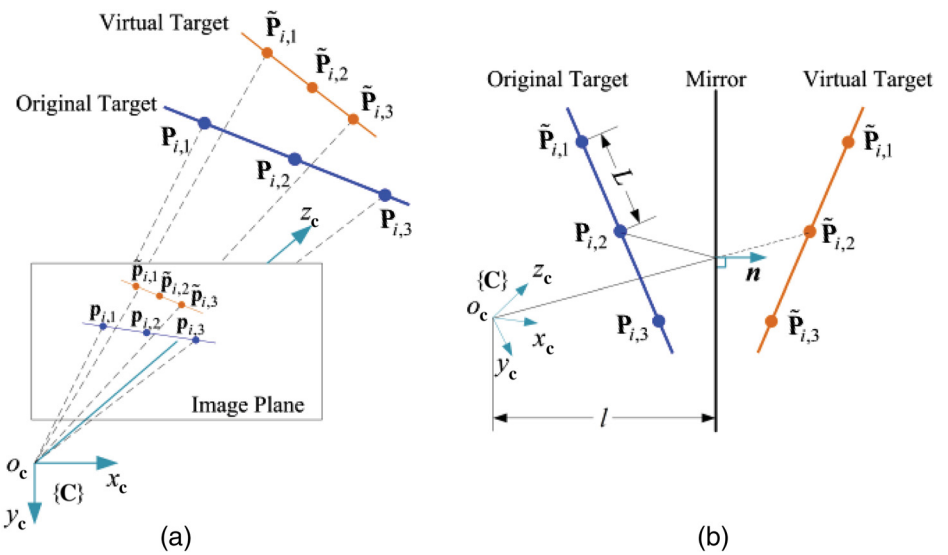

Fig. 2 The 1-D target is placed in front of the camera. (a) Projections of the original and virtual targets at the i'th position and (b) reflection through the planar mirror.

$$\frac{x_3 - x_1}{x_2 - x_1} = \frac{y_3 - y_1}{y_2 - y_1} = \frac{z_3 - z_1}{z_2 - z_1} = \rho, \tag{2}$$

where $\rho = \|\mathbf{P}_{i3} - \mathbf{P}_{i1}\| / \|\mathbf{P}_{i2} - \mathbf{P}_{i1}\|$ can be computed from the distances between the points $\mathbf{P}_{i,1}$, $\mathbf{P}_{i,2}$, $\mathbf{P}_{i,3}$.

Suppose that in the image coordinate system, $\mathbf{p}_{i,1}(u_1, v_1)$, $\mathbf{p}_{i,2}(u_2, v_2)$, and $\mathbf{p}_{i,3}(u_3, v_3)$. Combining Eqs. (1) and (2), we can obtain

$$\begin{cases} (u_1 - u_0)z_1 = x_1 f_x \\ (v_1 - v_0)z_1 = y_1 f_y \\ (u_2 - u_0)z_2 = x_2 f_x \\ (v_2 - v_0)z_2 = y_2 f_y \\ (u_3 - u_0)z_3 = x_3 f_x \\ (v_3 - v_0)z_3 = y_3 f_y \\ (\rho - 1)x_1 = \rho x_2 - x_3 \\ (\rho - 1)y_1 = \rho y_2 - y_3 \\ (\rho - 1)z_1 = \rho z_2 - z_3 \end{cases}. \tag{3}$$

By solving Eq. (3), the coordinates of $\mathbf{p}_{i,1}(x_1, y_1, z_1)$, $\mathbf{p}_{i,2}(x_2, y_2, z_2)$, and $\mathbf{p}_{i,3}(x_3, y_3, z_3)$ can be retrieved and the initial value of the normal vector $\mathbf{d}_i$ can be obtained. Due to the image noise, if there are more than three feature points, results of different point triplets, e.g., $(\mathbf{P}_{i,1}, \mathbf{P}_{i,2}, \mathbf{P}_{i,3})$ and $(\mathbf{P}_{i,1}, \mathbf{P}_{i,2}, \mathbf{P}_{i,4})$, are not equivalent.

Let $\hat{\mathbf{p}}_{i,j}$ be the image coordinates of the reprojection feature point

$$s\begin{bmatrix} \hat{\mathbf{p}}_{i,j} \\ 1 \end{bmatrix} = \mathbf{K}[\mathbf{P}_{i,1} + (j-1)\mathrm{L}\mathbf{d}_i], \tag{4}$$

where $\mathbf{s}$ is a scale factor, $\mathbf{L}$ is the known interval of two adjacent feature points on the target, and $\mathbf{K}$ denotes the intrinsic matrix in camera.

To utilize the redundant information of feature points, $\mathbf{P}_{i,1}$ and $\mathbf{d}_i$ are refined by minimizing the following reprojection error using Levenberg–Marquardt algorithm:[21]

$$f(\Omega) = \sum_{j=1}^{N} d^2(\mathbf{p}_{i,j}, \hat{\mathbf{p}}_{i,j}), \tag{5}$$

where $\Omega = (\mathbf{P}_{i,1}, \mathbf{d}_i)$. $\mathbf{d}_i = [\cos\theta_i \cos\varphi_i, \sin\varphi_i, \cos\theta_i \sin\varphi_i]^T$. θ_i and φ_i are the pitch angle and yaw angle of the i'th unit direction vector with respect to $\{\mathbf{C}\}$, respectively. $d(\cdot)$ is the distance function between two points. The initial value of $(\mathbf{P}_{i,1}, \mathbf{d}_i)$ can be obtained by Liu's method.[19]

The pose of the i'th virtual target $(\tilde{\mathbf{P}}_{i,1}, \tilde{\mathbf{d}}_i)$ can also be obtained similarly.

2.3 *Planar Mirror Reflection*

As shown in Fig. 2(b), camera $\{\mathbf{C}\}$ is in front of a planar mirror Π. The mirror is uniquely determined by its unit normal vector $\mathbf{n}$ and the distance $\mathbf{l}$ with respect to $\{\mathbf{C}\}$.[2,17] A spatial point $\mathbf{P}$ lies on the mirror plane if and only if it satisfies

$$\mathbf{n}^T\mathbf{P} = l. \tag{6}$$

The Householder matrix $\mathbf{H}$[22,23] is widely used in the mirror reflection model, and its equation is as follows:

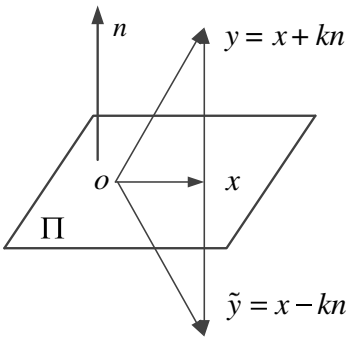

Fig. 3 The schematic diagram of a mirror surface emission model.

$$\mathbf{H} = I - 2\mathbf{n}\mathbf{n}^T. \tag{7}$$

The Householder transformation is a linear transformation that transforms a vector into a mirror image reflected by a hyperplane. The Householder transformation can zero some elements of a vector and keep the norm of the vector constant.

The schematic diagram of a mirror surface emission model is shown in Fig. 3. The equation of the plane Π is as follows:

$$\mathbf{n}^T x = 0 \quad \forall\, x \in \Pi. \tag{8}$$

If $x \in \Pi$, then

$$\mathbf{H}x = (I - 2\mathbf{n}\mathbf{n}^T)x = x. \tag{9}$$

If $y \notin \Pi$, then

$$\mathbf{H}y = (I - 2\mathbf{n}\mathbf{n}^T)(x + k\mathbf{n}) = x + k\mathbf{n} - 2\, k\mathbf{n}\mathbf{n}^T\mathbf{n} = x - k\mathbf{n} = \tilde{y}. \tag{10}$$

The reflection in Π can be represented by the following matrix:

$$\mathbf{S} = \begin{bmatrix} \mathbf{H} & 2\ln \\ 0 & 1 \end{bmatrix}. \tag{11}$$

Through the conversion of $\mathbf{S}$ can obtain the position of the virtual mirror $\tilde{\mathbf{P}}$ corresponding to the space point $\mathbf{P}$ in the camera coordinate system. The space point $\mathbf{P}$ and the corresponding virtual point are related as follows:

$$\begin{bmatrix} \tilde{\mathbf{P}} \\ 1 \end{bmatrix} = \mathbf{S}\begin{bmatrix} \mathbf{P} \\ 1 \end{bmatrix}. \tag{12}$$

3 Principle of the Mirror Pose Estimation

3.1 *Analysis of the Unique Solution*

The calculation of the normal vector $\mathbf{n}$ is equivalent to solving the Householder matrix $\mathbf{H}$. According to Eqs. (11) and (12), $\mathbf{d}_i$ and $\tilde{\mathbf{d}}_i$ are related by the Householder matrix as follows:

$$\tilde{\mathbf{d}}_i = \mathbf{H}\mathbf{d}_i. \tag{13}$$

From Eq. (13), one position is not enough to obtain a unique solution. For two positions, we have

$$\tilde{\mathbf{d}}_1 = \mathbf{H}\mathbf{d}_1, \tilde{\mathbf{d}}_2 = \mathbf{H}\mathbf{d}_2. \tag{14}$$

According to the principle of mirror image, assume that target real image is positioned for the right-handed coordinate system, and in the target virtual image, the left and right, upside down, and up and down directions remain the same with respect to the left-hand target image coordinate system.[4] Therefore, the relationship between real image $\mathbf{d}_1 \times \mathbf{d}_2$ and virtual image $\tilde{\mathbf{d}}_1 \times \tilde{\mathbf{d}}_2$ is as follows:

$$-\tilde{\mathbf{d}}_1 \times \tilde{\mathbf{d}}_2 = \mathbf{H}(\mathbf{d}_1 \times \mathbf{d}_2). \tag{15}$$

Combining Eqs. (14) and (15), we have

$$[\,\tilde{\mathbf{d}}_1 \quad \tilde{\mathbf{d}}_2 \quad -\tilde{\mathbf{d}} \times \tilde{\mathbf{d}}_2\,] = \mathbf{H}[\,\mathbf{d}_1 \quad \mathbf{d}_2 \quad \mathbf{d}_1 \times \mathbf{d}_2\,]. \tag{16}$$

If $\mathbf{d}_1$ is not parallel to $\mathbf{d}_2$, matrix $[\,\mathbf{d}_1 \quad \mathbf{d}_2 \quad \mathbf{d}_1 \times \mathbf{d}_2\,]$ is invertible. Thus, $\mathbf{H}$ can be uniquely determined as follows:

$$\mathbf{H} = [\,\tilde{\mathbf{d}}_1 \quad \tilde{\mathbf{d}}_2 \quad -\tilde{\mathbf{d}} \times \tilde{\mathbf{d}}_2\,][\,\mathbf{d}_1 \quad \mathbf{d}_2 \quad \mathbf{d}_1 \times \mathbf{d}_2\,]^{-1}. \tag{17}$$

If $\mathbf{d}_1$ is parallel to $\mathbf{d}_2$, the target motion is a pure translation, $\det[\,\mathbf{d}_1 \quad \mathbf{d}_2 \quad \mathbf{d}_1 \times \mathbf{d}_2\,] = 0$, thus a unique solution cannot be found.

3.2 *Computing the Normal Vector n*

If the target is placed more than two times, the best Householder matrix can be computed from the solution to the orthogonal Procrustes problem.[21] According to Eq. (13), the sum of the normal vector error is

$$E = \sum_{i=1}^{M} e_i = \sum_{i=1}^{M} \|\tilde{d}_i - Hd_i\|_F^2, \tag{18}$$

where $\|\cdot\|_{\mathbf{F}}$ denotes the Frobenius norm, $\|\mathbf{x}\|_{\mathrm{F}}^2 = \mathrm{tr}(\mathbf{x}^{\mathrm{T}}\mathbf{x}) = \mathrm{tr}(\mathbf{x}\mathbf{x}^{\mathrm{T}})$. If measurements are noiseless, $\mathbf{E} = 0$.

The best approximation of matrix $\mathbf{H}$ is equivalent to the matrix that solves the minimization problem of Eq. (18). According to the properties of the Householder matrix, Eq. (18) is expanded as

$$E = \sum_{i=1}^{M} \mathrm{tr}[\left(\tilde{\mathbf{d}}_i^T - \mathbf{d}_i^T\mathbf{H}^T\right)\left(\tilde{\mathbf{d}}_i - \mathbf{H}\mathbf{d}_i\right)] = 2M - 2\mathrm{tr}(\mathbf{H}\boldsymbol{\Delta}), \tag{19}$$

where $\boldsymbol{\Delta} = \sum_{i=1}^{M} \mathbf{d}_i\tilde{\mathbf{d}}_i^T$.

It can be seen that minimization of $\mathbf{E}$ is equivalent to maximize $\mathrm{tr}(\mathbf{H}\boldsymbol{\Delta})$. Therefore, the matrix $\mathbf{H}$ that solves the minimization problem of Eq. (18) is equivalent to $\mathbf{H}$ that solves

$$\max_{\mathbf{H}} \mathrm{tr}(\mathbf{H}\boldsymbol{\Delta}). \tag{20}$$

There are some researches[18,22] on finding the best rotation matrix with determinant $+1$. However, the Householder matrix is not a rotation but actually a reflection with determinant -1. To construct the best reflection, a lemma is introduced.

Lemma 1 For a given 3×3 matrix $\boldsymbol{\Delta}$ and a Householder matrix $\mathbf{H}$

$$\mathrm{tr}(\mathbf{H}\boldsymbol{\Delta}) \le \mathrm{tr}(D\Sigma), \tag{21}$$

where $U\Sigma V^T$ is the full SVD of $\boldsymbol{\Delta}$, the singular values are in decreasing order, and

$$\mathbf{D} = \mathrm{diag}\left[1, 1, -\det(\mathbf{V}\mathbf{U}^T)\right]. \tag{22}$$

Proof Using $\mathrm{tr}(\mathbf{A}\mathbf{B}) = \mathrm{tr}(\mathbf{B}\mathbf{A})$ and the SVD result of matrix $\boldsymbol{\Delta}$, we have

$$\mathrm{tr}(\mathbf{H}\Delta) = \mathrm{tr}(\mathbf{H}\mathbf{U}\Sigma\mathbf{V}^T) = (\mathbf{V}^T\mathbf{H}\mathbf{U}\mathbf{D}\mathbf{D}\Sigma), \tag{23}$$

where $\mathbf{D}^2 = I$ is the identity matrix. Let $\hat{\mathbf{H}} = \mathbf{V}^T\mathbf{H}\mathbf{U}\mathbf{D}$.

$$\hat{\mathbf{H}}\hat{\mathbf{H}}^T = (\mathbf{V}^T\mathbf{H}\mathbf{U}\mathbf{D})(\mathbf{V}^T\mathbf{H}\mathbf{U}\mathbf{D})^T = I, \tag{24}$$

$$\begin{aligned}\det(\hat{\mathbf{H}}) &= \det(\mathbf{V}^T\mathbf{H}\mathbf{U}\mathbf{D}) = \det(\mathbf{V}^T)\det(\mathbf{H})\det(\mathbf{U})\det(\mathbf{D}) \\ &= -[\det(\mathbf{V}^T)^2][\det(\mathbf{U})]^2\det(\mathbf{H}) = 1.\end{aligned} \tag{25}$$

According to Eqs. (24) and (25), we can come to the conclusion that $\hat{\mathbf{H}}$ is an orthogonal matrix with determinant 1 and hence a rotation matrix.

Suppose that $\hat{\mathbf{H}} = \begin{bmatrix} h_{11} & \cdots & h_{1j} \\ \vdots & \ddots & \vdots \\ h_{i1} & \cdots & h_{ij} \end{bmatrix}$, $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \sigma_3)$, we have

$$\mathrm{tr}(\mathbf{H}\boldsymbol{\Delta}) = \mathrm{tr}(\hat{\mathbf{H}}\mathbf{D}\Sigma) = \sigma_1 h_{11} + \sigma_2 h_{22} - \sigma_3 h_{33}\ \det(\mathbf{V}\mathbf{U}^T). \tag{26}$$

Therefore, the trace is maximized if $h_{ii} = 1$. As $\hat{\mathbf{H}}$ is an orthogonal matrix, this means that $\hat{\mathbf{H}}$ would have to be the identity matrix. Then

$$\mathrm{tr}(\mathbf{H}\boldsymbol{\Delta}) \le \sigma_1 + \sigma_2 - \sigma_3\det(\mathbf{V}U^T) = \mathrm{tr}(D\Sigma). \tag{27}$$

According to two sets of the original and virtual direction vectors, the best reflection matrix can be obtained by Theorem 1.

Theorem 1. The solution to the maximization problem of Eq. (20) is as follows:

$$\mathbf{H} = \mathbf{V}\mathbf{D}\mathbf{U}^T, \tag{28}$$

where $\mathbf{U}\Sigma\mathbf{V}^T$ is the full SVD of $\boldsymbol{\Delta}$ with decreasing singular values and $\mathbf{D}$ is subject to Eq. (22).

Proof. From Lemma 1, the maximization problem of Eq. (20) is achieved if a Householder matrix $\mathbf{H}$ can be constructed such that

$$\mathrm{tr}(\mathbf{H}\boldsymbol{\Delta}) = \mathrm{tr}(\mathbf{D}\Sigma). \tag{29}$$

Let $\mathbf{H} = \mathbf{V}\mathbf{D}\mathbf{U}^T$, then $\mathrm{tr}(\mathbf{H}\boldsymbol{\Delta}) = \mathrm{tr}(\mathbf{V}\mathbf{D}\mathbf{U}^T U\Sigma\mathbf{V}^T) = \mathrm{tr}(\mathbf{D}\Sigma)$. When the Householder matrix $\mathbf{H}$ is determined, the unit normal vector $\mathbf{n}$ is obtained.

3.3 *Computing the Distance l*

From Eq. (12), we have

$$2\ \mathrm{ln} = \tilde{\mathbf{P}}_{i,j} - \mathbf{H}\mathbf{P}_{i,j}. \tag{30}$$

Therefore, the distance $\mathbf{l}$ can be computed as

$$l = \frac{1}{2MN}\sum_{i=1}^{M}\sum_{j=1}^{N}[\mathbf{n}^T(\tilde{\mathbf{P}}_{i,j} - \mathbf{H}\mathbf{P}_{i,j})], \tag{31}$$

where $\mathbf{P}_{i,j}$ and $\tilde{\mathbf{P}}_{i,j}$ can be calculated according to the Eq. (3). $\mathbf{H}$ can be calculated according to Eq. (17).

3.4 *Nonlinear Optimization*

The results obtained in Sec. 3.2 and Sec. 3.3, which were called as the closed-form solution. The closed-form calculation is fast, but it is sensitive to the noise. In this subsection, the closed-form solution is refined with nonlinear optimization by minimizing the reprojection error of reflection. According to the camera model, we have

$$s\begin{bmatrix}\widehat{\mathbf{p}}_{i,j}\\ 1\end{bmatrix} = \mathbf{K}[\mathbf{I}\quad 0]\begin{bmatrix}\mathbf{I}-2\mathbf{n}\mathbf{n}^T & 2\ \mathrm{ln}\\ 0 & 1\end{bmatrix}\begin{bmatrix}\mathbf{p}_{i,j}\\ 1\end{bmatrix}, \tag{32}$$

where s is a scale factor, $\mathbf{K}$ denotes the intrinsic matrix in camera, $\mathbf{I}$ is a 3×3 identity matrix, and 0 is a 1×3 zero vector. $\widehat{\mathbf{p}}_{i,j}$ is the reprojected image coordinates.

The mirror normal vector $\mathbf{n}$ and the distance $\mathbf{l}$ are refined by minimizing the following function using the Levenberg–Marquardt algorithm[21]

$$f(\Omega) = \sum_{i=1}^{M}\sum_{j=1}^{N}\omega^2(\tilde{\mathbf{p}}_{i,j}, \widehat{\mathbf{p}}_{i,j}), \tag{33}$$

where $\Omega = (\mathbf{n}, \mathbf{l})$, $\mathbf{n} = [\cos\theta\cos\varphi, \sin\theta, \cos\theta\sin\varphi]^T$. θ and φ are the pitch angle and yaw angle of the normal vector $\mathbf{n}$ with respect to $\{\mathbf{C}\}$, respectively. $\omega(\cdot)$ is the distance between image points and $\tilde{\mathbf{p}}_{ij}$ is the image coordinate corresponding to the imaginary point $\widehat{\mathbf{p}}_{i,j}$. After optimization, the optimal value of the normal vector $\mathbf{n}$ and the distance $\mathbf{l}$ is obtained.

3.5 *Calculating the Relative Position of the Image Machine $\{C\}$ and the Carrier $\{B\}$*

Figure 4 shows the relationship diagram of the vector coordinate system $\{\mathbf{B}\}$, the vector image frame $\{\mathbf{B}'\}$, and the machine coordinate system $\{\mathbf{C}\}$, where the $\mathbf{R_{BC}}$ and $\mathbf{R_{B'C}}$ stand for the rotation vector, the $\mathbf{t_{BC}}$ and $\mathbf{t_{B'C}}$ represent the translation vector, and the $\mathbf{S}$ mean mirror reflection matrix.

The image of the mounting carrier in the mirror is observed by the plane mirror, so as to obtain the transformation relationship between the plane frame $\{\mathbf{C}\}$ and the vector virtual frame $\{\mathbf{B}'\}$. According to the above results, the transformation relationship between the image coordinate system $\{\mathbf{C}\}$ and the carrier image coordinates $\{\mathbf{B}\}$ can be obtained directly. Thus, the calibration of the installation parameters of the imager can be completed.

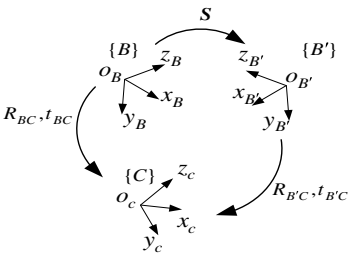

Fig.4 The relationship diagram of the vector coordinate system $\{\mathbf{B}\}$, the vector image frame $\{\mathbf{B}'\}$, and the machine coordinate system $\{\mathbf{C}\}$.

$$\begin{bmatrix}\mathbf{R}_{BC} & \mathbf{t}_{BC}\\ 0 & 1\end{bmatrix} = \begin{bmatrix}\mathbf{R}_{B'C} & \mathbf{t}_{B'C}\\ 0 & 1\end{bmatrix}\begin{bmatrix}\mathbf{I}-2\mathbf{n}\mathbf{n}^T & 2\ \mathrm{ln}\\ \mathbf{0} & 1\end{bmatrix}, \tag{34}$$

where $\mathbf{R}$ is the rotation matrix, $\mathbf{t}$ is the translation vector, subscripts $B'C$ and BC, respectively, represent the conversion relationship between the camera coordinate system $\{\mathbf{C}\}$ and the virtual carrier coordinate $\{\mathbf{B}'\}$, and the transformation relationship between the camera coordinate system $\{\mathbf{C}\}$ and the carrier image coordinates $\{\mathbf{B}\}$.

3.6 *Method Summary*

The calibration method of the installation parameters of the image machine based on the plane mirror is proposed in this section, which keeps the relative position of the image machine and the plane mirror unchanged. The main processes of the proposed method are as follows:

Step 1: The intrinsic parameters of the camera are calibrated using the Bouguet Camera Calibration Toolbox based on Zhang's method.[24,25] The internal parameters of the camera mainly include the following: f_x, f_y, u_0, v_0.

Step 2: The 1-D target with N feature points is placed in front of the camera for M times. For each position, one image is taken by the camera. Both the original and virtual targets are in the camera's FOV. All the images are undistorted according to the intrinsic calibration results. Thus, image coordinates are distortionless.

Step 3: The optimal estimation value of matrix $\mathbf{H}$ can be obtained by Orthogonal Procrustes problem. For the ith position, compute the value of $\mathbf{P}_{i,1}, \mathbf{d}_i$ and $\tilde{\mathbf{P}}_{i,1}, \tilde{\mathbf{d}}_i$, respectively. Repeat this step for $i = 1, \ldots, M$.

Step 4: By obtaining the minimum value of the reprojection error function, the optimal value of the mirror normal vector $\mathbf{n}$ and the image machine-mirror distance $\mathbf{l}$ are obtained.

Step 5: The rotation matrix $\mathbf{R_{BC}}$ and the translation vector $\mathbf{t_{BC}}$ between the image machine coordinate system and the carrier coordinate system are calculated, and the calibration of the installation parameters of the image machine is completed.

4 Experimental Results

Through the previous derivation, the mirror position (the mirror vector $\mathbf{n}$ and the image machine-mirror distance $\mathbf{l}$) are the key parameters in the calibration process of the

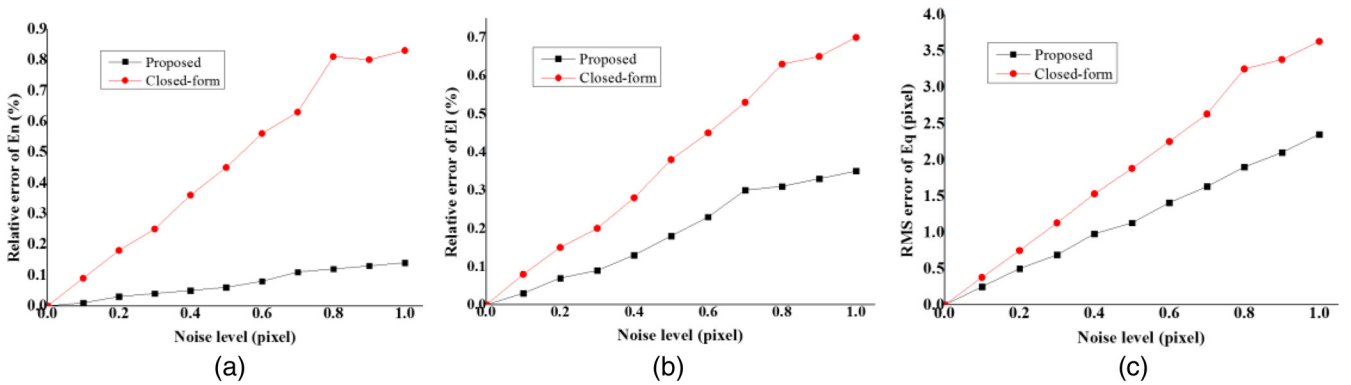

Fig. 5 Error of pose estimation versus the noise level: (a) the relative error of En versus noise level, (b) the relative error of El versus noise level, and (c) the RMS error of Eq versus noise level.

installation parameters of the image machine. Therefore, this section takes the pose error of plane mirror as the accuracy evaluation index, analyzes the related factors that affect the accuracy through simulation data and physical experiments, and compares it with other calibration methods.

4.1 *Simulation Data*

To evaluate the accuracy, the estimation errors are defined as follows:

(a) Error of the mirror normal vector $\mathbf{n}$ against the truth value $\mathbf{n_g}$ is defined as follows:

$$E_n = \arccos(\mathbf{n}^T \mathbf{n}_g). \tag{35}$$

(b) Error of the distance $\mathbf{l}$ against the truth value $\mathbf{l_g}$ is defined as follows:

$$E_l = l - l_g. \tag{36}$$

(c) Reprojection error is defined as follows:

$$E_q = \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N} \frac{\|\tilde{\mathbf{p}}_{i,j} - \hat{\mathbf{p}}_{i,j}\|^2}{2MN}}. \tag{37}$$

In this experiment, one simulated camera and a planar mirror are used to construct the system. The intrinsic parameters of the camera are $f_x = f_y = 995.556$, $u_0 = 512$, $v_0 = 384$. The image resolution is 1024 pixel × 768 pixel. The mirror normal vector $\mathbf{n} = [-0.5, 0, 0.866]^{\mathrm{T}}$, i.e., $\theta = 0$, $\varphi = 120$ deg. The distance $\mathbf{l} = 400$ mm, the interval $\mathbf{L} = 25$ mm.

The positions of target are randomly generated. Only the original and virtual targets in the camera's FOV are used. Gaussian noise with zero mean and different standard deviation σ is added to the image coordinates of feature points. The effects of several factors on the accuracy of the pose estimation are evaluated quantitatively by synthetic data as follows.

4.1.1 *Influence of the characteristic point noise on the accuracy*

In this experiment, $M = 10$, $N = 9$. Gaussian noise is added to the image coordinates of the feature points. The noise level varies from 0 to 1.0 pixel. For each level, 100 independent trials are performed to compute the relative errors of $\mathbf{E_n}$, $\mathbf{E_l}$ and RMS errors of $\mathbf{E_q}$.

From Fig. 5, the error increases linearly with the noise level. It indicates that the closed-form solution can be effectively adjusted by the proposed method. If the noise level of a system is <0.5 pixel, relative errors of the normal vector $\mathbf{n}$ and the distance $\mathbf{l}$ are <0.06% and 0.5%, respectively, which are applicable for common applications.

4.1.2 *Influence of the number of target placement on the accuracy*

In this experiment, Gaussian noise with $\sigma = 0.2$ pixel is added to the image coordinates of the feature points. M varies from 2 to 12, N is set to 9. For each level, 100 independent trials are performed and are used to compute relative errors.

From Fig. 6, the error decreases with the increasing number of target positions. When the target is placed only twice, the pose estimation error is rather high. The estimation error decreases rapidly as the number of positions increases. When there are enough positions, the contribution of additional positions decreases.

4.1.3 *Influence of the number of target feature points on the accuracy*

In this experiment, Gaussian noise with $\sigma = 0.2$ pixel is added to the image coordinates of the feature points. M is set to 10, N varies from 3 to 16. For each level, 100 independent trials are performed and are used to evaluate relative errors.

From Fig. 7, increasing number of feature points benefits the estimation accuracy. When there are only three feature points on the target, the estimation error reaches maximum. The accuracy improves sharply when N increases from 3 to 8. When $N > 10$, the accuracy stables at the certain level.

4.2 *Physical Experiments*

As shown in Fig. 8(a), a Canon 60D digital camera and a planar mirror are in fixed positions. The image resolution

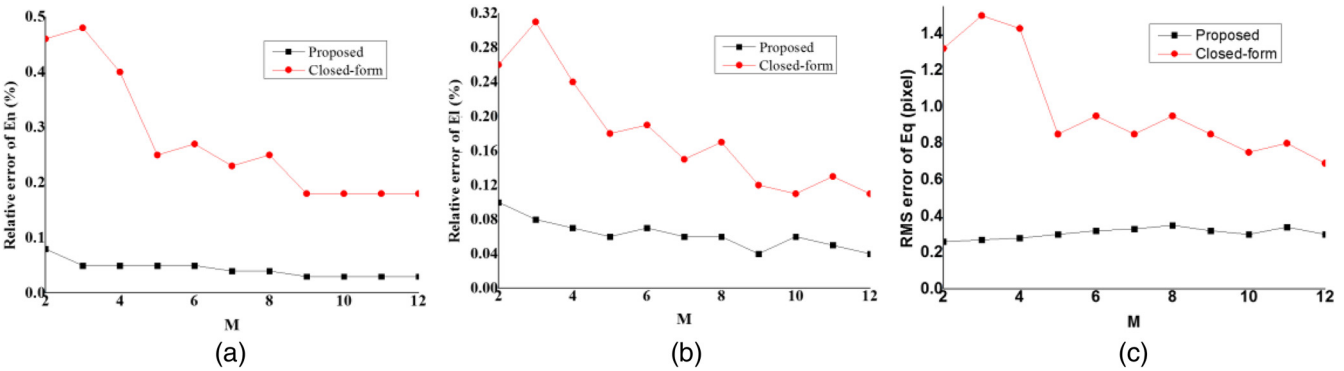

Fig. 6 Error of pose estimation vs. the number of target positions M (a) the relative error of En versus the number of target positions, (b) the relative error of El versus the number of target positions, and (c) the RMS error of Eq versus the number of target positions.

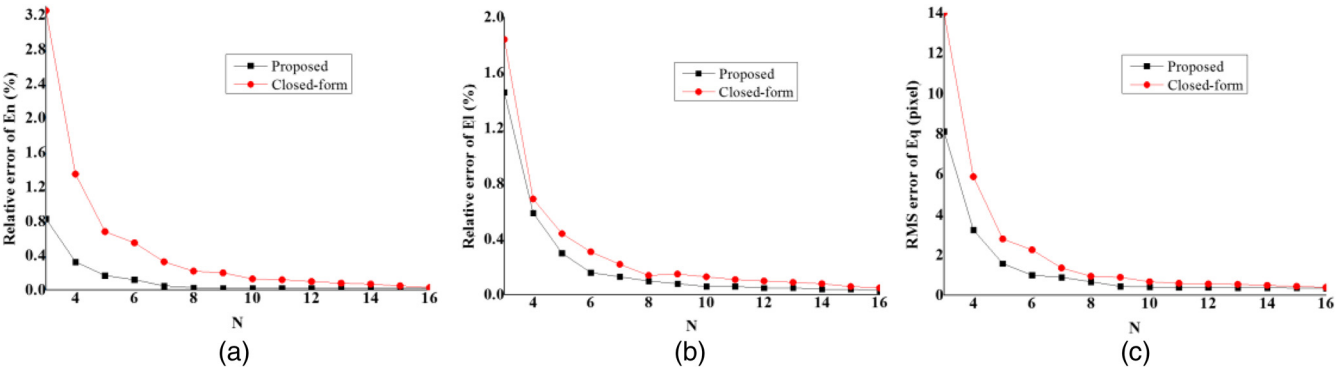

Fig. 7 Error of pose estimation versus the number of feature points N (a) the relative error of En versus the number of feature points, (b) the relative error of El versus the number of feature points, and (c) the RMS error of Eq versus the number of feature points.

Fig. 8 (a) An image captured by the camera. Both the original and virtual 1-D target are in the camera's FOV and (b) a checkerboard is used to evaluate the accuracy.

of the camera is 1920 pixel × 1280 pixel. The intrinsic parameters of the vision sensor are calibrated using the Bouguet's calibration toolbox,[15] as shown in Table 1. We use the intersection of two rows of checkerboards equivalently to the 1-D target. The distance between neighboring feature points in the checkerboard is 20 mm, i.e., $\mathbf{L} = 20$ mm. The two-row checkerboard is used to define 1-D target for feature extraction.

As shown in Fig. 8(b), a checkerboard is placed between the camera and the mirror. Both the original and virtual

Table 1 Intrinsic parameters of the camera.

	f_x	f_y	u_0	v_0
Camera	1548	1548	960	640
Bouguet	1561.293	1560.153	972.497	597.061
Error (%)	0.86	0.79	1.3	6.7

checkerboards are in the camera's FOV. The spatial coordinates of the corner points of the checkerboards can be obtained by the Bouguet's toolbox. In this experiment, reflective error $\mathbf{E_s}$ and the reprojection error $\mathbf{E_q}$ are used to evaluate the accuracy

$$E_s = \frac{1}{(n_x+1)(n_y+1)} \sum_{i=1}^{n_x+1} \sum_{j=1}^{n_y+1} \|\tilde{\mathbf{Q}}_{i,j} - (\mathbf{H}\mathbf{Q}_{i,j} + 2\ \ln)\|, \tag{38}$$

where $\mathbf{Q}_{i,j}$ and $\tilde{\mathbf{Q}}_{i,j}$ are the spatial coordinates of the original and virtual corner points of the checkerboards, with respect to $\{\mathbf{C}\}$. $\mathbf{n_x}$ and $\mathbf{n_y}$ are the number of squares along two orthogonal directions, respectively.

To verify the effect of several factors shown in the synthetic experiment, pose estimations of different position numbers and different feature points are implemented.

4.2.1 *Influence of the number of target placement on the accuracy*

In this experiment, the number of target positions M is changed from 2 to 12, the number of feature points N is set to 9. From Fig. 9, the result is similar to the result of the synthetic experiment in Sec. 4.1.2. An increasing number of target positions benefits the accuracy, especially when the number of target positions is relatively small.

4.2.2 *Influence of the number of target feature points on the accuracy*

In this experiment, the number of target positions M is set to 10, the number of feature points N is changed from 3 to 12. From Fig. 10, it also indicates a similar result of the synthetic experiment in Sec. 4.1.3. The estimation error peaks at $N = 3$, which corresponds to the minimum quantity of feature points. Then, the estimation error decreases sharply with the increasing number of feature points. When $N > 8$, the estimation error flattens out gradually.

4.2.3 *Discussions*

Evidently, there exists a trade-off between the accuracy and the workload. According to the effect, analysis of the number of target positions M and the number of feature points N, $M = 10$ and $N = 9$ is a balanced choice. When the error of an image point is 0.5 pixel, the synthetic experiment shows that relative errors of the normal vector $\mathbf{n}$ and the distance $\mathbf{l}$ are 0.06% and 0.45%, respectively. It is acceptable for common applications such as the camera-to-base calibration.[8]

In addition, there is a gap between the results of synthetic data and real-data experiments. There may be some reasons for this. First, the feature point extraction method is a commonly used method with good generality and acceptable accuracy. Extraction algorithms with higher accuracy improve the accuracy, which will be further investigated. Second, the mirror used in this paper is a domestic appliance,

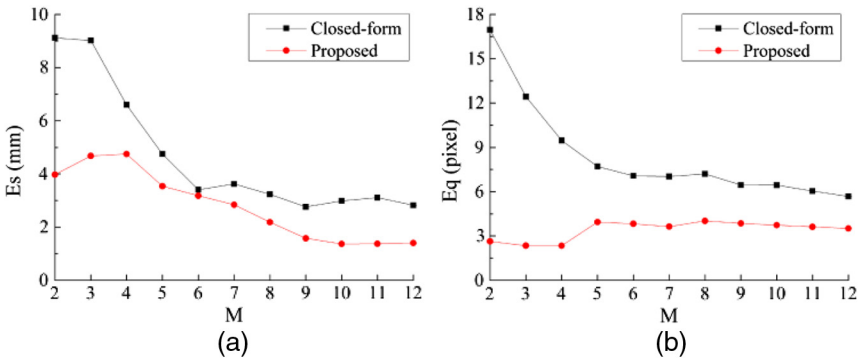

Fig. 9 Error of pose estimation versus the number of target positions M (a) the reflective error Es versus target position number and (b) the reprojection error Eq versus target position number.

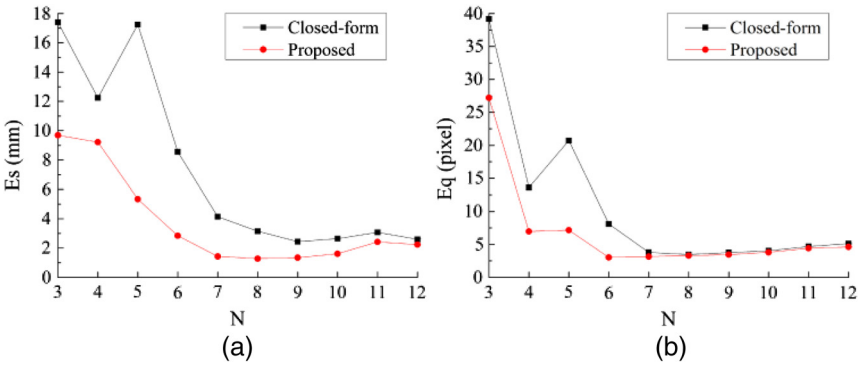

Fig. 10 Error of pose estimation versus the number of feature points N (a) the reflective error ES versus the number of feature points and (b) the reprojection error Eq versus the number of feature points.

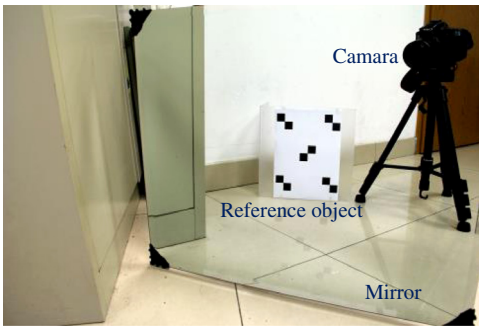

Fig. 11 A virtual image of the reference object.

not a precision optical device. Precision mirrors can be used to improve the accuracy.

4.3 *Comparisons*

There are many methods for the pose estimation. Here, three typical methods are described as follows:

Method 1: This method is similar to the proposed method. As shown in Fig. 8(b), a checkerboard is used as the calibration target, instead of the 1-D target. The corner points are extracted and calibrated by the Bouguet's toolbox. The pose of the mirror is also refined using nonlinear optimization. We use method 1 as the criterion for evaluating the merits of the proposed method.

Method 2: As shown in Fig. 11, a reference object is placed on the ground and remains fixed. Then, the pose of the mirror can be calculated by Li's method.[17]

For each method, 10 images are captured to estimate the mirror pose. The reflective error Es is used to evaluate the accuracy. The results of different methods are shown in Table 2.

From Table 2, the comparisons show that the proposed method achieves a good accuracy. Compared with the closed-form solution, the closed solution is simple and fast, but it is more sensitive to noise and larger in error. The same line determined by the selection of different points is different. Therefore, the normal vectors determined by these lines will also have certain differences, which will produce certain errors. Compared with the checkerboard of method 1, the errors of the two methods are relatively small. However, the number of feature points of 1-D target is reduced a lot, under the condition that the computation accuracy is not decreased. Therefore, according to the 1-D target of the known size, the calibration of the imaging machine can be achieved without reducing the precision, and the limitation of the complex calibration tool can be removed, so that the calibration is more flexible and convenient. The accuracy of method 3 is roughly the same as the closed-form solution.

Table 2 The estimation errors of different methods.

	Proposed method ($M = 10$, $N = 9$)	Closed-form solution ($M = 10$, $N = 9$)	Method 1	Method 2
E_s/mm	1.41	6.38	1.37	5.71

However, method 3 fails when the camera and the mirror cannot be moved.

5 Conclusion

In this paper, a pose estimation method for planar mirrors is presented. The method only needs a simple 1-D target and a digital camera. First, the camera captures the target in different positions. The pose of the target in each position is obtained. Then, the normal vector and the distance between the camera and the mirror are calculated by a closed-form method, followed by a nonlinear optimization.

Our method is simple and flexible. A minimum of two nonparallel target positions and three collinear feature points on the target are needed. No prior information about the target motion and the camera's pose is required. The proposed method is accurate and robust to noises. Both synthetic data and real experiments demonstrate the effectiveness of our method. In addition, both the camera and the mirror do not need to move during the process, so it is suitable for application where the camera and the mirror need to remain fixed.

In some applications, the transformation between the body frame and the camera frame need to be accurately calibrated. However, cameras mounted on the body usually cannot observe the body directly. We plan to apply our method for the camera-to-body calibration in the future.

Acknowledgments

The authors acknowledge the support by the Industrial Technology Development Program under Grant No. B1120131046.

References

1. Y. Xiao, X. Su, and W. Chen, "Flexible geometrical calibration for fringe-reflection 3D measurement," *Opt. Lett.* **37**(4), 620–622 (2012).
2. G. L. Mariottini et al., "Planar mirrors for image-based robot localization and 3-D reconstruction," *Mechatronics* **22**, 398–409 (2012).
3. X. Feng and D. Pan, "Research on the application of single camera stereo vision sensor in three-dimensional point measurement," *J. Mod. Opt.* **62**(15), 1204–1210 (2015).
4. R. K. Kumar et al., "Simple calibration of non-overlapping cameras with a mirror," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, Anchorage, pp. 2602–2608 (2008).
5. P. Lebraly et al., "Flexible extrinsic calibration of non-overlapping cameras using a planar mirror: Application to vision-based robotics," in *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, (IROS, 2010)* pp. 5640–5647 (2010).
6. P. Sturm and T. Bonfort, "How to compute the pose of an object without a direct view?" *Lect. Notes Comput. Sci.* **3852**, 21–31 (2006).
7. R. Rodrigues, J. P. Barreto, and U. Nunes, "Camera pose estimation using images of planar mirror reflections," *Lect. Notes Comput. Sci.* **6314**, 382–395 (2010).
8. J. A. Hesch, A. I. Mourikis, and S. I. Roumeliotis, "Mirror-based extrinsic camera calibration," in *Algorithmic Foundations of Robotics VIII*, Springer Tracts in Advanced Robotics, G. S. Chirikjian et al., Eds. Vol. **57**, Springer, Berlin, Heidelberg (2009).
9. K. Takahashi, S. Nobuhara, and T. Matsuyama, "Mirror-based camera pose estimation using an orthogonality constraint," *IPSJ Trans. Comput. Vision Appl.* **8**, 11–19 (2016).
10. G. Panahandeh, M. Jansson, and P. Händel, "Calibration of an IMU-camera cluster using planar mirror reflection and its observability analysis," *IEEE Trans. Instrum. Meas.* **64**(1), 75–88 (2015).
11. T. P. Pachidis and J. N. Lygouras, "Pseudostereo-vision system: a monocular stereo-vision system as a sensor for real-time robot applications," *IEEE Trans. Instrum. Meas.* **56**(6), 2547–2560, (2007).
12. Y. Xu et al., "A calibration method for non-overlapping cameras based on mirrored absolute phase target," *Int. J. Adv. Manuf. Technol.* (2018).
13. L. Li et al., "Flat mirror tilt and piston measurement based on structured light reflection," *Opt. Express* **22**(22), 27707–27716 (2014).
14. H. Ren, F. Gao, and X. Jiang, "Iterative optimization calibration method for stereo deflectometry," *Opt. Express* **23**(17), 22060–22068 (2015).
15. J. Y. Bouguet, "Camera calibration toolbox for Matlab," http://www.vision.caltech.edu/bouguetj (2013).

16. T. Matsuyama, S. Nobuhara, and K. Takahashi, "A new mirror-based extrinsic camera calibration using an orthogonality constraint" in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1051–1058 (2012).
17. X. Li et al., "Accurate mirror-based camera pose estimation with explicit geometric meanings," *Sci. China-Technol. Sci.* **57**(12), 2504–2513 (2014).
18. M. Shah, "Comparing two sets of corresponding six degree of freedom data," *Comput. Vis. Image Understanding* **115**(10), 1355–1362 (2011).
19. Z. Liu et al., "Novel calibration method for non-overlapping multiple vision sensors based on 1D target," *Opt. Lasers Eng.* **49**(4), 570–577 (2011).
20. Z. Liu, F. J. Li, and G. J. Zhang, "An external parameter calibration method for multiple cameras based on laser rangefinder," *Measurement* **47**, 954–962 (2014).
21. J. J. Moré, "The Levenberg–Marquardt algorithm: implementation and theory," in *Numerical Analysis*, pp. 105–116, Springer (1978).
22. F. Zhang, "Matrix analysis and applications," *Int. J. Inf. Syst. Sci.* **309**(1), i (2004).
23. K. Kanatani, "Analysis of 3-D rotation fitting," *IEEE Trans. Pattern Anal. Mach. Intell.* **16**(5), 543–549 (1994).
24. Z. Y. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(11), 1330–1334 (2000).
25. Y. Bok et al., "Extrinsic calibration of non-overlapping camera-laser system using structured environment" in *IEEE/RSJ Int. Conf. on Intelligent Robots & Systems*, pp. 436–443 (2014).

Wenlei Liu received his BS degree at Shandong University of Science and Technology, Qingdao, in 2012 and his MS degree at Beihang University, Beijing, in 2016. Currently, he is pursuing his PhD at Beihang University, Beijing, China. He mainly studies the high-precision visual relative navigation, and also studies the data fusion of multi data sensors to achieve the purpose of cooperative detection and guidance.

Sentang Wu received his PhD in dynamics, ballistics, and aircraft motion control systems from National Aviation University, Ukraine, in 1992. Currently, he is a professor of automation science and electrical engineering, and a PhD tutor at Beihang University, Beijing, China. He is also the navy missile expert in the National Defense Basic Research Institute and a member of the academic committee. His research interests include the theory and application of nonlinear stochastic systems, computer information processing and control, aircraft cooperative control, precision, and guidance.

Xiaolong Wu received his BS degree in Sichuan University, Chengdu, in 2007 and his PhD in Beihang University, Beijing, in 2017. He mainly studies the information acquisition system of the missile's autonomous formation, including the relative navigation method of the missile's autonomous formation, the calibration of the vision sensor's installation parameters and the optimization of the field of view, and the cooperative detection and guidance of the missile's autonomous formation.