

Telecentric stereo micro-vision system: Calibration method and experiments



Zhong Chen^{a,b,*}, Huiyang Liao^{a,b}, Xianmin Zhang^{a,b}

^a School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510640, China

^b Guangdong Provincial Key Laboratory of Precision Equipment and Manufacturing Technology, South China University of Technology, Guangzhou 510640, China

ARTICLE INFO

Article history:

Received 7 September 2013

Received in revised form

11 January 2014

Accepted 20 January 2014

Available online 4 February 2014

Keywords:

Stereo micro-vision

Calibration

Machine vision

Telecentric camera

ABSTRACT

When a telecentric stereo micro-vision system is used for 3D measurement, the fundamental and key issue is the calibration of the system. However, a telecentric lens generally possesses a small field of view (FOV), which renders the calibration complicated and difficult. In our case, the FOV is $2.4 \text{ mm} \times 3.2 \text{ mm}$. From the existing literature published to date, little attention has been paid to the calibration of visual systems with limited FOVs; therefore, a highly effective calibration method is required. In this paper, a new and accurate calibration method is proposed. First, we present a geometric model of the telecentric camera used in this study, considering lens distortion. Afterward, a method for single-camera calibration is described in detail. With respect to recovering the camera rotation matrix, the problem of sign ambiguity induced by the planar-object-based calibration technique is successfully solved. Based on this technique, a method for calibrating a telecentric stereo micro-vision system is presented. The results of experiments conducted in this study show that the method is quite accurate and reliable.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Stereo micro-vision systems have been widely used in the non-contact measurement of 3D micro-scale objects. One of the broadest applications of such systems is in stereo light microscopes [1–3], which consist of a combination of conventional lenses. Compared to these lenses, telecentric lenses possess numerous unparalleled advantages, such as high resolution, nearly zero distortion and constant magnification, although they are more expensive and have comparatively much smaller fields of view (FOVs). Due to these features, a stereo micro-vision system composed of two telecentric lenses can also be utilized to measure micro-scale objects in 3D space. Because telecentric lenses can eliminate optical parallax, determine the precise size of an object and produce high measurement accuracy, such a system can be widely applied for defect detection of printed circuit boards (PCBs), 3D measurement and reconstruction of micro-scale objects and digital-image-correlation (DIC) measurement of shape, motion and deformation, etc. [4–7].

Telecentric stereo micro-vision systems must be calibrated before being used for measurement. To date, many calibration methods have been proposed. Among them, self-calibration methods [8–10] apply to calibrating moving cameras, whereas conventional methods [11–13]

are used for fixed-camera calibration. Most of these methods generally aim at the calibration of pinhole cameras with conventional lenses instead of telecentric lens. Unlike the perspective projection of pinhole cameras, telecentric cameras perform orthographic projection, which gives rise to a geometric model of the projection that is different from that of the pinhole camera. Thus, many of the existing calibration methods are not applicable to telecentric cameras. In our telecentric stereo micro-vision system, both telecentric lenses only produce a $2.4 \text{ mm} \times 3.2 \text{ mm}$ FOV. To date, little attention has been paid to the calibration of such systems with limited FOV. Hence, a highly effective and easy-to-implement calibration method is required.

In general, the existing calibration methods that are suitable for telecentric cameras can be classified into three categories:

- 1) **Self-calibration.** This category of methods is appropriate for calibrating moving cameras. The camera moves to capture a sequence of calibration images of a scene at several positions. Horaud [14] proposed a method for reconstructing the Euclidean structure of an object with an uncalibrated affine camera. The camera, mounted on a robot hand that could provide controlled motion, captured images at various positions. Horaud then developed an algebraic decomposition method for determining the camera transformation matrix. This method was observed to be applicable only for calibrating cameras with controlled motions. Based on the invariance of a camera's intrinsic parameters, Quan [15] presented an approach for computing the

* Corresponding author at: School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510640, China.
Tel.: +86 15918687239.

E-mail address: mezchen@scut.edu.cn (Z. Chen).

shape and motion from a sequence of images without knowing the camera's parameters. The limitation of this method was that it is not suitable for the calibration of fixed cameras. Tomasic [16] developed a singular value decomposition method to factorize the measurement matrix into two matrices representing the object shape and camera rotation. The shape and motion were recovered under orthography from an image stream. Similarly, Tomasic's method should be classified as self-calibration. In fact, many other affine motion recovery methods [17–19] all fall into this category.

- 2) **3D-object-based calibration.** This category of methods normally uses a 3D calibration cube whose geometry in 3D space is known with very good precision. The cube usually comprises two or three planes perpendicular to each other. Methods in this category apply to the calibration of fixed cameras. Basically, a direct linear transformation (DLT) algorithm is used to estimate camera parameters. Although 3D-object-based calibration methods are simple, direct and easy to understand, the apparent limitation is that it is very expensive to fabricate an accurate 3D micro-scale calibration cube that can fit into the commonly small FOV of telecentric lenses. Gorpas [20] adopted this method to calibrate a telecentric stereovision system used for the detection of skin cancer tumors.
- 3) **Planar-object-based calibration.** This category of methods uses a planar calibration pattern that provides mark points with good precision. The camera is fixed but takes pictures of the pattern placed at several different orientations. This is virtually the reverse scenario of the self-calibration method. At each position, a homographic matrix H maps 2D mark points of the pattern to 2D image points. Based on the invariance of the intrinsic parameters of the camera, a factorization method can be developed to decompose the constraints provided by a number of homographic matrices. For telecentric cameras, this category of methods is the best option due to its ease of use, cost-effectiveness and good accuracy [11]. Haskamp [21] directly used a nonlinear optimization approach instead of the factorization method to estimate camera parameters. However, due to the randomly chosen initial guesses of all of the parameters, the iteration process might require a very long time to run, and the global minimum might not be achieved. Li [22] proposed a method for calibrating a camera with a telecentric lens. However, Li's method presented some problems in recovering the extrinsic parameters of the camera. Lanman [23] discussed a factorization method that successfully calibrated a telecentric camera that was used for 3D shape reconstruction. However, in Lanman's case, the adopted telecentric lens had a comparatively large FOV, which allowed the camera to be calibrated much more easily.

The stereo micro-vision system, in our case, is composed of two fixed telecentric cameras whose FOV is very small. Hence,

we discarded the former two categories of methods and chose the planar-object-based calibration method. During the calibration process, the calibration pattern was mounted on a precision micro-positioning stage, which could provide fine adjustment for the pattern and move it into the view of the two small FOV lenses.

In this paper, first, we discuss the geometric model of the telecentric camera. The mathematical description [19] we adopt is the orthographic projection model. Lens distortions are also discussed even though they are small. However, only radial distortion, the major source of distortions of telecentric lenses, is taken into consideration.

Next, we describe the procedure for calibrating a single telecentric camera. A factorization method for estimating the camera parameters is introduced in detail. Inherent to the planar-object-based calibration method, a sign ambiguity problem appears in the process of recovering the first two elements of the third column of camera rotation matrix. If their signs cannot be disambiguated, then the full camera transformation matrix cannot be possibly recovered. This problem is perfectly solved by a technique introduced in Section 3.3. Afterward, all of the estimated camera parameters undergo refinement through non-linear optimization.

Finally, we discuss the principle of stereovision. After the calibration of a single telecentric camera was successfully achieved, we calibrated the telecentric stereo micro-vision system. The strategy implemented for the system calibration is described in Section 4.2. To validate the method, we conducted an experiment and created three cases of reconstruction using the calibration pattern. The experimental results are shown in Section 5.

2. Geometric model of telecentric camera

2.1. Telecentric lens and telecentric perspective

For object-side telecentric lens, the aperture stop is placed at the image-side focal point [24]. Thus, only the light rays parallel to the optical axis from the object side are allowed to pass through. In the optics, the entrance pupil is the virtual image of the aperture stop, as observed through the front of the lens. Now that the aperture stop is placed at the image-side focal point f' , the entrance pupil, which acts as the projection center, is therefore located at infinity on the object side. For this reason, the projection center lies at infinity on the object side as well, and the principal ray is parallel to the optical axis [25]. See Fig. 1(a). This property of the object-side telecentric lens produces an orthographic projection of an object in the world space [26]. As a consequence, all equally sized objects in the world space, however close to or far from the lens, would have the same sized projection in the image. This explains why telecentric lens produces constant magnification [27]. In conclusion, an object-side telecentric lens produces orthographic projection.

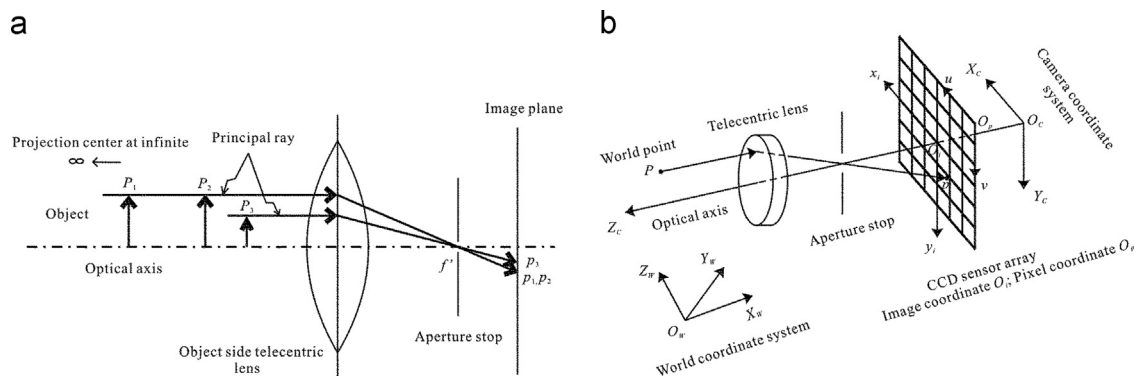


Fig. 1. Schematic of (a) an object-side telecentric lens and (b) the object-side telecentric projection model.

2.2. Telecentric camera model

Fig. 1(b) illustrates how a telecentric camera projects a 3D world point P to a 2D point p in the image. $O_w(X_w, Y_w, Z_w, 1)$, $O_c(X_c, Y_c, Z_c, 1)$, $O_i(x_i, y_i, 1)$ and $O_p(u, v, 1)$, respectively, represent the world, the camera, the image and the pixel coordinate systems in homogeneous form.

As mentioned in the previous section, the projection center of an object-side telecentric camera lies at infinity, which means that the left-hand 3×3 block of the camera matrix M is singular [26]. Thus, the last row of the camera matrix M has the form $(0, 0, 0, 1)$. Consequently, the projection model can be geometrically described in homogeneous form as follows:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \underbrace{\begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ 0 & 0 & 0 & 1 \end{pmatrix}}_M \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}, \quad (1)$$

Denoted by a rotation R in the form of a 3×3 orthogonal matrix and a translation vector $T = (t_x, t_y, t_z)^T$, a rigid transformation can convert a world point $(X_w, Y_w, Z_w, 1)$ to a point in the camera coordinate system $O_c(X_c, Y_c, Z_c, 1)$. As for telecentric camera, t_z is meaningless because of the nature of the orthographic projection, which can be presented by a matrix **Orth** of the form

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

This matrix maps a camera coordinate point $(X_c, Y_c, Z_c, 1)^T$ to an image coordinate point $(x_i, y_i, 1)^T$. Similarly, when the projection point is transformed from the image coordinate system $O_i(x_i, y_i, 1)$ to the pixel coordinate system $O_p(u, v, 1)$, there is also a 2D rigid transformation. Hence, Eq. (1) can be decomposed into the following form [26]:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha & \gamma & 0 & 0 \\ 0 & \beta & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_C \underbrace{\begin{pmatrix} R_{3 \times 3} & T_{3 \times 1} \\ 0 & 1 \end{pmatrix}}_K \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}, \quad (2)$$

where α and β are the axis factors of the pixel coordinates and γ is a factor related to the skew angle of the CCD pixel array. These factors are the inherent elements that constitute the camera's intrinsic matrix C . The camera's extrinsic matrix is denoted as K .

The camera's intrinsic matrix C can be further extended as follows:

$$C = \begin{pmatrix} \frac{\varepsilon}{s_u} & -\frac{\varepsilon \cot \theta}{s_u} & 0 & 0 \\ 0 & \frac{\varepsilon}{s_v \sin \theta} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (3)$$

where ε is the magnification of the lens; s_u and s_v are the scaling factors (expressed in unit metric length/pixels) [28]; and θ is the skew angle of the CCD pixel array. Because an object-side telecentric camera has no projection center (lies at infinite), the first two elements in the last column of C have no other values than zero. If any, they should be the pixel coordinates of the image's center, which indicate the center of the lens' radial distortion.

For convenience, Eq. (1) can be rewritten in inhomogeneous form as

$$\begin{pmatrix} u \\ v \end{pmatrix} = \tilde{M}_{2 \times 3} \begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix} + \mathbf{n}_{2 \times 1}, \quad (4)$$

where $\tilde{M}_{2 \times 3} = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{pmatrix}$ and $\mathbf{n}_{2 \times 1} = (m_{14} \ m_{24})^T$.

2.3. Lens distortion

Fundamentally, every camera lens produces distortions, which include radial, tangential and thin prism distortions [29]. Telecentric lenses are no exception, but their distortions are small compared to those of other lenses. Among all types of distortions, radial distortion, by and large, dominates telecentric lens distortions.

Radial distortion is caused by imperfect lens shape. The distortion is symmetric and distorts the image points along the radial direction from the distortion center. In our case, tangential and thin prism distortions are not taken into account because they usually tend to be much smaller than radial distortion [30]. Hence, the distortion equations are expressed as follows:

$$\begin{aligned} \delta_x &= x_i(k_1 r^2 + k_2 r^4 + k_3 r^6 + \dots) \\ \delta_y &= y_i(k_1 r^2 + k_2 r^4 + k_3 r^6 + \dots), \end{aligned} \quad (5)$$

where radial distance $r = \sqrt{x_i^2 + y_i^2}$; (x_i, y_i) are the ideal undistorted image coordinates; k_1 and k_2 are the coefficients of radial distortion.

With respect to radial distortion, the real distorted image coordinates can be given by

$$\begin{aligned} \tilde{x}_i &= x_i + \delta_x \\ \tilde{y}_i &= y_i + \delta_y. \end{aligned} \quad (6)$$

Using only the first two radial distortion terms in Eq. (5), we obtain [11]

$$\begin{aligned} \tilde{u} &= u + (u - u_0)[k_1(x_i^2 + y_i^2) + k_2(x_i^2 + y_i^2)^2] \\ \tilde{v} &= v + (v - v_0)[k_1(x_i^2 + y_i^2) + k_2(x_i^2 + y_i^2)^2], \end{aligned} \quad (7)$$

where (u, v) are the pixel coordinates of the feature points estimated by the camera model; (\tilde{u}, \tilde{v}) are the pixel coordinates of the feature points extracted by using image processing techniques. In the case of a telecentric camera, (u_0, v_0) have no other meaning than simply indicating the center of the lens's radial distortion. Thus, whether they are included or excluded in Eq. (7) will produce no different distortion coefficients.

3. Single camera calibration

The calibration pattern we used during the process was a planar grid point board. Multiple views of the calibration pattern along different orientations were captured by a fixed camera. By convention, for each view, we assume that the calibration pattern lies on $Z=0$ of the world coordinate system. Therefore, we can quickly write Eq. (2) as

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha & \gamma & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & t_x \\ r_{21} & r_{22} & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ 1 \end{pmatrix} = \mathbf{A} \mathbf{K}_s \mathbf{P} = \mathbf{H} \mathbf{P}, \quad (8)$$

where \mathbf{A} and \mathbf{K}_s are, respectively, the truncated intrinsic and extrinsic matrices of the camera. Unlike the pinhole camera model, there is no arbitrary scale factor on the left-hand side of Eq. (8) due to the orthographic projection of the object-side telecentric lens. Here, it can be observed that the equation is linear. Inspired by Zhang [11], we use a factorized method to decompose the homographic matrix \mathbf{H} to obtain both the intrinsic and extrinsic parameters of the camera.

3.1. Estimation of camera parameters

Homographic matrix \mathbf{H} can be easily computed by using a DLT algorithm [26]. The only constraint on \mathbf{H} , for a telecentric camera,

is that the last row of \mathbf{H} be $(0, 0, 1)$. Because $\mathbf{H} = \mathbf{A}\mathbf{K}_s$, we can easily obtain

$$(\mathbf{A}^{-1}\mathbf{H})^T\mathbf{A}^{-1}\mathbf{H} = \begin{pmatrix} \mathbf{R}_s^T\mathbf{R}_s & \mathbf{R}_s^T\mathbf{T}_s \\ \mathbf{T}_s^T\mathbf{R}_s & \mathbf{T}_s^T\mathbf{T}_s + 1 \end{pmatrix}, \quad (9)$$

where $\mathbf{R}_s = \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix}$ and $\mathbf{T}_s = (t_x \ t_y)^T$. Because \mathbf{R}_s is defined as the truncated rotation matrix, we can write the rotation matrix \mathbf{R} as

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_s & \mathbf{B} \\ \mathbf{C}^T & d \end{pmatrix}, \quad (10)$$

where $\mathbf{C}^T = (r_{31} \ r_{32})$; $\mathbf{B} = (r_{13} \ r_{23})^T$; $d = r_{33}$. Thus, we obtain

$$\mathbf{R}^T\mathbf{R} = \begin{pmatrix} \mathbf{R}_s^T\mathbf{R}_s + \mathbf{C}\mathbf{C}^T & \mathbf{R}_s^T\mathbf{B} + d\mathbf{C} \\ \mathbf{B}^T\mathbf{R}_s + d\mathbf{C}^T & \mathbf{B}^T\mathbf{B} + d^2 \end{pmatrix}. \quad (11)$$

It is noteworthy that the rotation matrix \mathbf{R} is orthogonal, so $\mathbf{R}^T\mathbf{R} = \mathbf{I}_{3 \times 3}$, where $\mathbf{I}_{3 \times 3}$ is the 3×3 identity matrix. As a result, the upper-left 2×2 submatrix on the right of the Eq. (11) can be written as

$$\mathbf{R}_s^T\mathbf{R}_s + \mathbf{C}\mathbf{C}^T = \mathbf{I}_{2 \times 2}. \quad (12)$$

Because $\det(\mathbf{C}\mathbf{C}^T) = 0$, we can quickly obtain

$$\det(\mathbf{R}_s^T\mathbf{R}_s - \mathbf{I}_{2 \times 2}) = 0. \quad (13)$$

Eqs. (9) and (13) can be combined to obtain

$$l_1 - l_2(h_{21}^2 + h_{22}^2) - l_3(h_{11}^2 + h_{12}^2) + 2l_4(h_{11}h_{21} + h_{12}h_{22}) = -(h_{11}h_{22} - h_{12}h_{21})^2, \quad (14)$$

where $l_1 = \alpha^2\beta^2$, $l_2 = \alpha^2 + \gamma^2$, $l_3 = \beta^2$ and $l_4 = \beta\gamma$; h_{ij} is the i th row and j th column element of homographic matrix \mathbf{H} .

Given n images, we can obtain n different homographic matrices. Without loss of generality, the element in the i th row and j th column of the n th homographic matrix $\mathbf{H}^{(n)}$ is denoted as $h_{ij}^{(n)}$. For the n th image, let us write $G_1^{(n)} = -(h_{21}^{(n)2} + h_{22}^{(n)2})$, $G_2^{(n)} = -(h_{11}^{(n)2} + h_{12}^{(n)2})$, $G_3^{(n)} = 2(h_{11}^{(n)}h_{21}^{(n)} + h_{12}^{(n)}h_{22}^{(n)})$, and $w^{(n)} = -(h_{11}^{(n)}h_{22}^{(n)} - h_{12}^{(n)}h_{21}^{(n)})^2$. Thus, Eq. (14) can be stacked to obtain n equations in total, which can be written as a matrix equation as follows

$$\underbrace{\begin{pmatrix} 1 & G_1^{(1)} & G_2^{(1)} & G_3^{(1)} \\ 1 & G_1^{(2)} & G_2^{(2)} & G_3^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & G_1^{(n)} & G_2^{(n)} & G_3^{(n)} \end{pmatrix}}_{\mathbf{G}} \underbrace{\begin{pmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \end{pmatrix}}_{\mathbf{L}} = \underbrace{\begin{pmatrix} w^{(1)} \\ w^{(2)} \\ \vdots \\ w^{(n)} \end{pmatrix}}_{\mathbf{w}}. \quad (15)$$

From Eq. (15), we note that at least four images are required to solve vector \mathbf{L} . The linear least-square solution is thus given by

$$\mathbf{L} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{w}. \quad (16)$$

We can then immediately recover

$$\alpha = \sqrt{\frac{l_2 l_3 - l_4^2}{l_3}}, \quad \beta = \sqrt{l_3}, \quad \gamma = \sqrt{l_2 - \frac{l_1}{l_3}}. \quad (17)$$

For the i th image, the truncated extrinsic matrix of the camera $\mathbf{K}_s^{(i)}$ can be derived from

$$\mathbf{K}_s^{(i)} = \mathbf{A}^{-1}\mathbf{H}^{(i)}, \quad i = 1, 2, 3, \dots \quad (18)$$

3.2. Parameter refinement by non-linear optimization

All of the intrinsic and extrinsic parameters of the camera require further refinement after they have been estimated by

using the method described above. Therefore, we use the maximum likelihood estimation to minimize the distances between the extracted and predicted pixel coordinates of the pattern's mark points. This is a nonlinear optimization problem given by

$$\sum_{i=1}^n \sum_{j=1}^m \|p_{ij} - \tilde{p}(\mathbf{A}, \mathbf{K}_s^{(i)}, k_1, k_2, \mathbf{P}_j)\|^2, \quad (19)$$

where p_{ij} are the extracted pixel coordinates of the j th mark point in the i th image and $\tilde{p}(\mathbf{A}, \mathbf{K}_s^{(i)}, k_1, k_2, \mathbf{P}_j)$ are the corresponding predicted pixel coordinates. This nonlinear problem can be solved by using the Levenberg–Marquardt algorithm [31]. The initial guesses of all of the camera parameters can be obtained by using the method described in the previous section.

3.3. Recover rotation matrix $\mathbf{R}^{(i)}$

For each image, the camera's extrinsic matrix varies. From Eq. (8), we note that, for the i th image, only the elements in the upper-left 2×2 submatrix of the rotation matrix $\mathbf{R}^{(i)}$ can be directly recovered from $\mathbf{K}_s^{(i)}$. Providing $\mathbf{R}^{(i)}$ is unitary and orthogonal, the remaining matrix elements can be obtained from

$$r_{13} = \pm \sqrt{1 - r_{11}^2 - r_{12}^2}, \quad (20)$$

$$r_{23} = \pm \sqrt{1 - r_{21}^2 - r_{22}^2}, \quad (21)$$

$$\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2, \quad (22)$$

where \mathbf{r}_1 , \mathbf{r}_2 and \mathbf{r}_3 are the row vectors of the matrix $\mathbf{R}^{(i)}$. From Eqs. (20) and (21), it is obvious that there exists uncertainty in the signs of r_{13} and r_{23} , whereas the orthogonormality of $\mathbf{R}^{(i)}$ can only provide one constraint by

$$-(r_{11}r_{21} + r_{12}r_{22}) = r_{13}r_{23}. \quad (23)$$

There is still one degree of freedom left. To solve this problem of sign ambiguity, we require an extra constraint. From Eq. (8), it is comprehensible that the elimination of r_{13} and r_{23} from $\mathbf{R}^{(i)}$ is due to the definition that the calibration pattern lies on $Z=0$. Notice that

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha & \gamma & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}. \quad (24)$$

To obtain r_{13} and r_{23} , we need to acquire the information of Z_w . However, the calibration pattern is planar and cannot provide mark points that contain Z -axis information. Under these circumstances, we successfully solved the problem with the assistance of a micro-positioning stage; after the i th image was taken, we drove the stage to provide a translational displacement Z_d , along the Z axis of the calibration pattern, and then, we captured an additional image that together with the i th image grabbed before the displacement could be used to recover the signs of r_{13} and r_{23} quickly. Here, we call these two images an image pair. From Eq. (24), we obtain

$$\begin{pmatrix} r_{13} \\ r_{23} \\ 0 \end{pmatrix} = \mathbf{A}^{-1} \begin{pmatrix} \frac{u_d}{Z_d} \\ \frac{v_d}{Z_d} \\ \frac{1}{Z_d} \end{pmatrix} - \mathbf{K}_s^{(i)} \begin{pmatrix} \frac{X_w}{Z_d} \\ \frac{Y_w}{Z_d} \\ \frac{1}{Z_d} \end{pmatrix}, \quad (25)$$

where (u_d, v_d) are the pixel coordinates of the world point (X_w, Y_w, Z_d) .

The r_{13} and r_{23} recovered from Eq. (25) may not precisely satisfy Eqs. (20) and (21), but their signs are all perfectly disambiguated, which results in the complete recovery of $\mathbf{R}^{(i)}$ and solves the key problem of this calibration method.

4. Calibration of the telecentric stereo micro-vision system

From the foregoing discussion, there is no difficulty in understanding that the 3D space information of an object cannot be uniquely recovered by only using one fixed camera. Indeed, two cameras are required to view the object simultaneously from different angles. These two cameras together achieve binocular stereovision.

4.1. Principle of telecentric stereovision

An object-side telecentric camera produces an orthographic projection, which is one of the typical forms of projection of an affine camera. Therefore, two of these cameras combined produce affine stereovision, whose schematic is illustrated in Fig. 2. O_{LC} and O_{RC} are, respectively, the origins of the left and right camera coordinates. However, in fact, for telecentric cameras, the origins lie at infinity; therefore, the two principal rays, PO_{LC} and PO_{RC} , are actually parallel to the optical axes, $O_{LC}Z_L$ and $O_{RC}Z_R$, and vanish at infinity. With respect to a single telecentric camera, it is clear that any world point P on the same principal ray produces the projection at an identical point in the image. As a result, two cameras are required to determine the real world point P . If we know that p and p' are the corresponding projections of P in the left and right views, respectively, then P can be uniquely determined.

From Eq. (4), if $p = (u, v)^T \leftrightarrow p' = (u', v')^T$ is an image point correspondence, and both cameras are calibrated, then the world point $P = (X, Y, Z)^T$ can be recovered by

$$\begin{pmatrix} \tilde{M} \\ \tilde{M}' \end{pmatrix} P = \begin{pmatrix} p - n \\ p' - n' \end{pmatrix}, \quad (26)$$

where \tilde{M} and n belong to the left camera and \tilde{M}' and n' to the right camera.

Eq. (26) is an over-determined set of equations; thus, the world point P can be obtained by using linear least-square method. To

obtain P , the stereovision system needs to be calibrated so that the camera matrix M and a common world coordinate system O_w for both the left and right cameras can be obtained in advance.

4.2. Calibrating the telecentric stereo micro-vision system

When applied to 3D object measurement, the telecentric stereo micro-vision system should define a common world coordinate system for both cameras. The world coordinate system is chosen according to various measurement requirements. In our case, the planar calibration pattern was mounted on the PI micro-positioning stage; thus, for simplicity, the world coordinate system was defined by placing the pattern on an orientation in which the world coordinate system coincided with the PI micro-positioning stage's coordinates of motion. In this orientation, both cameras simultaneously captured pictures of the pattern. To fully recover the extrinsic matrix, the PI stage should drive the pattern by a certain Z-axis displacement where an additional pair of images is taken. This common world coordinate system for both cameras is called the reference world coordinate system on which all subsequent measurements are based.

Afterward, more calibration pictures were taken by both cameras independently but not simultaneously. The cameras were independently calibrated by using the method described in the previous section. The reasons for this are obvious; first, due to the limited FOV of the telecentric lenses, it is quite difficult for the two cameras, on the one hand, to capture simultaneous images of the calibration pattern under each orientation and, on the other hand, to avoid degeneracy [25]; second, in practice, the subsequent measurement only requires one common world coordinate frame, and hence, there is no need for multiple different simultaneous views of the pattern by both cameras.

Once the stereo vision system is successfully calibrated, Eq. (26) can be used to measure a 3D object in the reference world coordinate frame.

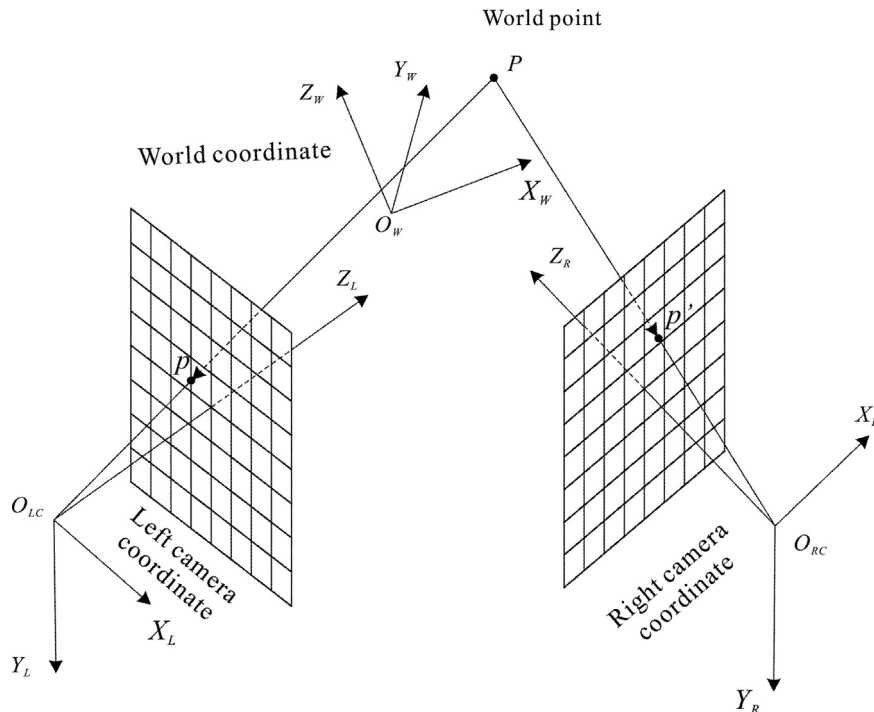


Fig. 2. The imaging model of a stereovision system.

5. Experiments and discussion

Experiments were conducted to verify the calibration method described above. The telecentric stereo micro-vision system adopted in the experiment was composed of two AVT Guppy Pro F125B cameras (Resolution: 1292×964 ; sensor size: $1/3''$; cell size: $3.75 \mu\text{m}$) mounted with two Moritex MML2-ST110D telecentric lenses (Working distance: 110 mm; FOV: $2.4 \text{ mm} \times 3.2 \text{ mm}$; DOF: 0.66 mm; magnification: $2 \times$; resolution: $11 \mu\text{m}$). The planar calibration pattern used in the experiment was a standard $1 \text{ mm} \times 1 \text{ mm}$ Halcon black-dot grid plate with a distance of $125 \mu\text{m}$ between every two neighboring dots. The grid plate was mounted on a PI M-126DG micro-positioning stage. The lighting source was an LED illuminator.

5.1. Minimal configuration of the calibration pictures

From Eq. (15), we know that at least four images are required to recover the intrinsic and extrinsic matrices of the camera. Moreover, the four calibration pictures should avoid causing similar degeneracy to the camera calibration. In other words, if the pattern is not rotated by all of its axes, then the depth change of the calibration pattern in each image will not be sharp, causing the homographic matrices to be similar and the coefficient matrix G in Eq. (15) to be singular, which gives rise to the degeneracy problem. For instance, if the four calibration pictures are taken by simply displacing the pattern along its Z axis, then the patterns will be

parallel to each other, and it will be impossible to calibrate the camera with these pictures.

Therefore, the minimal configuration for calibrating a camera is four pictures of the pattern rotated fully by all of its axes. If there are more than four pictures, the problem may not be so fatal because Eq. (15) becomes a linear least-square problem in this scenario. In our case, the FOV of the lenses is very small; thus, it is difficult to ensure a non-degenerate minimal configuration. For this reason, the more pictures that are captured, the more accurate the calibrating results becomes.

5.2. Image acquisition and processing

Due to the small FOV of the lenses, the micro-positioning stage is needed not only to drive the calibration pattern into the working distance of the lenses but also to provide micro-adjustments to make the pattern's imaging clear and sharp. Both cameras captured 24 calibration pictures. See the first 24 pictures in Fig. 3 (a) and (b). Among them, only the 1st image was taken in an orientation in which the calibration pattern could be simultaneously viewed by both cameras. The 25th pictures in Fig. 3(a) and (b), which were obtained after displacing the pattern in the 1st picture by $125 \mu\text{m}$ along its Z axis, are used to solve the rotation matrix R of each camera. All of the other remaining pictures were captured independently by the two cameras. The pixel coordinates of the mark point's center were extracted by using an ellipse fitting algorithm in Matlab.

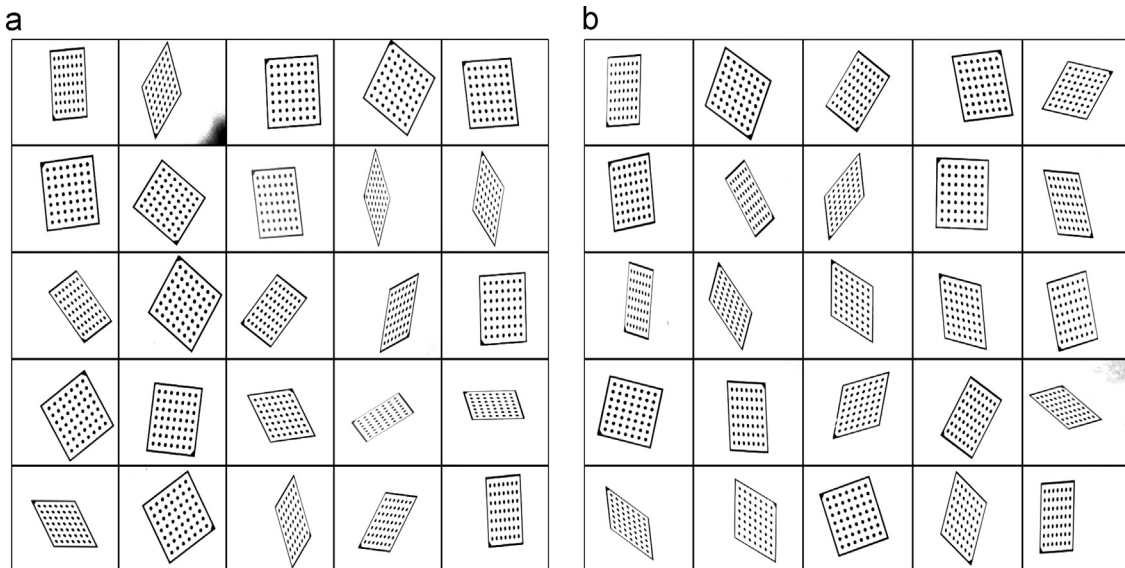


Fig. 3. Twenty-five calibration pictures of (a) the left camera and (b) the right camera.

Table 1
Calibration results for both cameras.

Item	Left camera	Right camera
Camera intrinsic matrix C	$\begin{pmatrix} 522.53 & -0.010 & 0 & 0 \\ 0 & 522.50 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 525.73 & 0.0651 & 0 & 0 \\ 0 & 525.29 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$
Rotation matrix R	$\begin{pmatrix} 0.6508 & -0.0559 & 0.7571 \\ -0.0499 & -0.9978 & -0.0198 \\ 0.7579 & -0.0094 & -0.6522 \end{pmatrix}$	$\begin{pmatrix} 0.6138 & 0.0265 & -0.7890 \\ -0.0374 & -0.9966 & -0.0734 \\ -0.7883 & 0.0746 & -0.6107 \end{pmatrix}$
Translation (t_x, t_y)	(0.8260, 1.0140)	(0.4643, 1.0990)
Distortion coefficients (k_1, k_2)	$(-8.972 \times 10^{-5}, 9.145 \times 10^{-6})$	$(-3.304 \times 10^{-5}, 3.014 \times 10^{-6})$

5.3. Experiment results

The estimated camera parameters are listed in Table 1. All of the parameters were refined through nonlinear optimization.

From Eq. (3), we know that the skew angle of the camera's CCD sensor array can be calculated by using $\theta = \arccot(-\gamma/\alpha)$. Referring to the recovered camera's intrinsic matrix \mathbf{C} , we can quickly calculate the skew angles for the left and right cameras to be 89.9989° and 90.0071° , respectively. Both angles are quite close to 90° , as expected with almost all modern CCD cameras.

For both cameras, α and β are nearly the same. Given that the scaling factors $s_u = s_v = 0.00375$, the magnification of both lenses can be computed by using $\varepsilon = (s_x\alpha + s_y\beta\sin\theta)/2$. Thus, the calculated

magnification for the left camera is 1.96, and that of the right camera is 1.97. Both values approach 2.0, the given nominal magnification of the lenses.

Moreover, we notice that the computed radial distortion coefficients k_1 , k_2 are very small, which in fact is consistent with the low-distortion property of telecentric lenses. Normally, other types of distortions such as tangential and prism distortions are much smaller than the radial distortion; therefore, it is reliable that we only take into account the radial distortion in the model.

After both cameras were successfully calibrated, we performed the re-projection using the model. Then, we compared the predicted pixel coordinates of the pattern's mark points with the corresponding extracted ones. Fig. 4 shows the re-projection errors of both cameras. The errors δ can be used to evaluate the calibration method we proposed. For each camera, there are a total of 1176 mark points in 24 images.

Table 2 shows the mean values and standard deviations of the re-projection errors for both the left and right cameras. The results indicate that for the two cameras, the absolute mean errors in the u and v axes are both less than one pixel. The standard deviations are all near 0.1, which means that the re-projection errors, for 1176 points of each camera, are intensively small, which proves that our method is accurate and reliable.

5.4. Degenerate scenario

To illustrate how important image acquisition is during the process, we created an example of an ill calibration. Fig. 5(a) shows six images taken by the right camera. These six patterns do not have sharp depth changes because the pattern was not been fully rotated around all of its three axes. We used them to calibrate the right camera. The re-projection errors are shown in Fig. 5(b). The mean errors in the u and v axes are, respectively, -9.3082 and -6.4525 . The errors are too large, and therefore, the results of the calibration should not be trusted.

If the six patterns are parallel to each other, the problem will become much more ill-conditioned. It is usually difficult to take calibration pictures that can satisfy the minimal configuration under a small FOV. Consequently, the best solution to avoid degeneracy is to take as many as possible pictures of the pattern in different orientations.

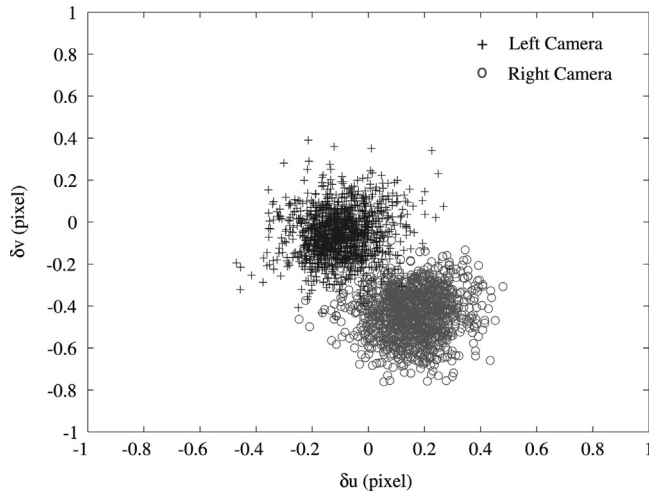


Fig. 4. Re-projection errors of the left and right cameras.

Table 2

The mean values and standard deviations of the re-projection errors.

Item	Left camera	Right camera
Mean value ($\bar{\delta}_u$, $\bar{\delta}_v$)	(-0.1010 , -0.0579)	(0.1591 , -0.4354)
Standard deviation (σ_u , σ_v)	(0.1016 , 0.1149)	(0.1055 , 0.1139)

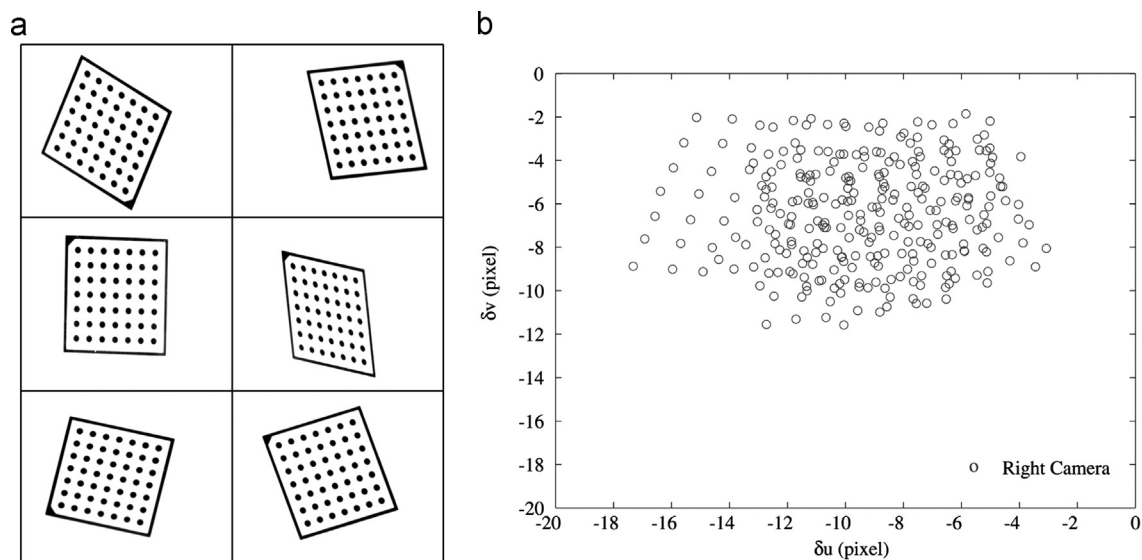


Fig. 5. Illustration of degeneracy. (a) Six calibration pictures of the right camera. The pattern was not fully rotated and translated by all of its three axes. (b) The corresponding re-projection errors.

5.5. Simple validation of the telecentric stereo micro-vision system

From Eq. (26), once the telecentric stereo micro-vision system is calibrated, we can use it for 3D measurement and reconstruction. To verify the calibration of the system, we carried out two types of reconstructions.

In the first case, we made use of the 1st and 25th calibration pictures in both Fig. 3(a) and (b). Fig. 6(a) shows the four pictures. Because the 25th picture was obtained after displacing the pattern in the 1st picture by 125 μm along its Z axis, we could soon verify the effectiveness of the system calibration by comparing the real displacement values with the results estimated by the reconstruction illustrated in Fig. 6(b). For the 49 mark points on the pattern, their

corresponding estimated displacements given by the reconstruction are all very close to 125 μm . Fig. 6(c) depicts the estimation errors. The average estimation error is 1.1 μm , with a standard deviation of 0.11.

Because the 1st pattern in the first case defined the reference world coordinate system, we performed another reconstruction to further illustrate the generality of the calibration effectiveness. In this case, the pattern was arbitrarily placed. Afterward, we drove the stage to displace the pattern by $-250 \mu\text{m}$ along the Z axis of the reference world coordinate system. Fig. 7(a) and (b) shows the pictures and the result of the reconstruction, respectively. For the 49 mark points on the pattern, the displacement values estimated by the reconstruction, compared to the real one, have an average error of 2.9 μm , and the standard deviation of the errors is 0.20. In fact, the

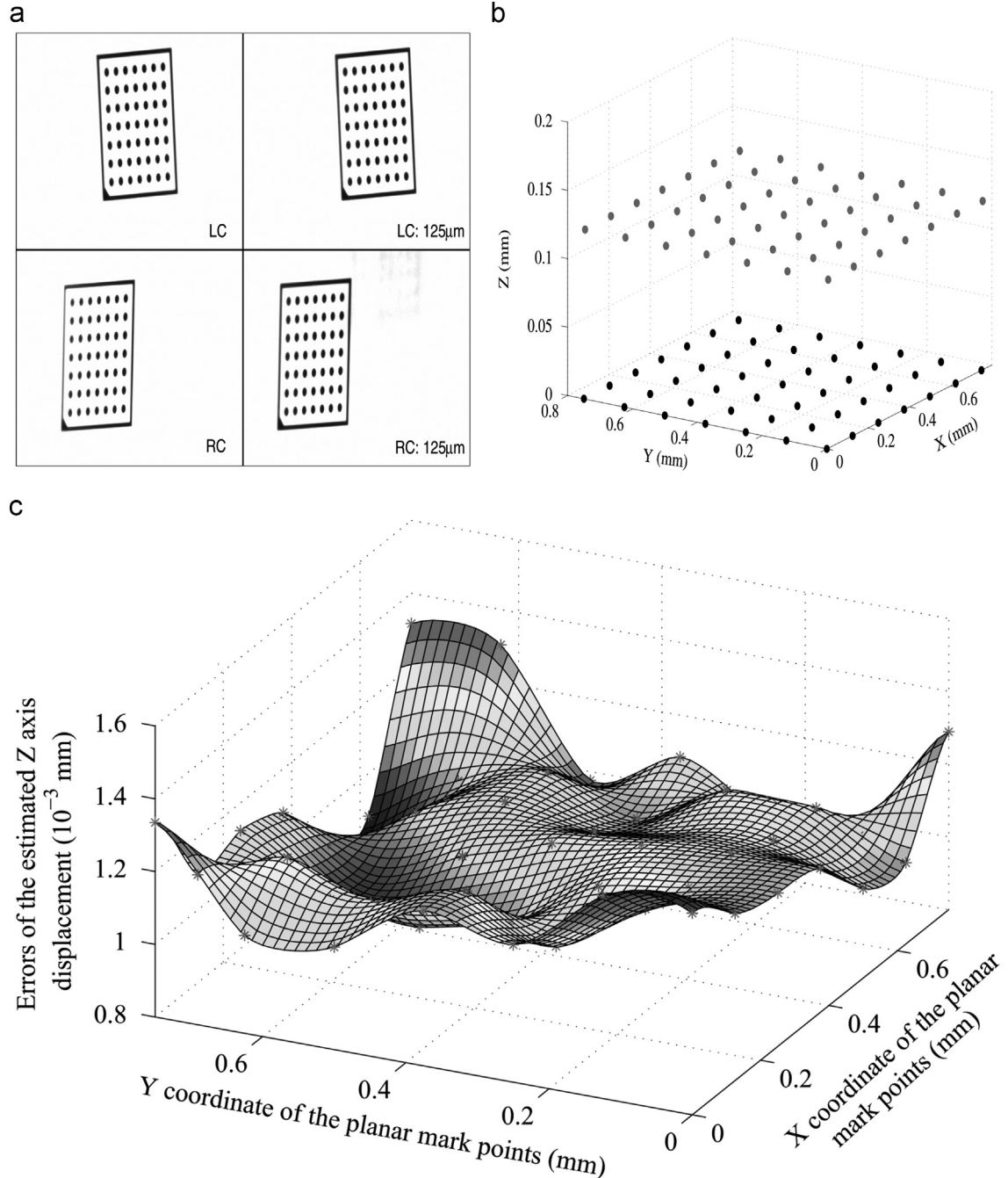


Fig. 6. The reconstruction and measurement of the pattern at $Z=0$ plane of the world coordinate system after a 125- μm displacement along its Z axis. (a) Pictures taken by the left and right cameras before and after the 125- μm displacement. (b) The bottom and upper grid plates are, respectively, the reconstructions of the pattern before and after the displacement. (c) Errors of the displacement estimated by the stereo micro-vision system.

error is nearly 90 times smaller than the real 250- μm displacement. Fig. 7(c) depicts the reconstruction errors.

The two cases discussed above represent a simple validation of the calibrated telecentric stereo micro-vision system. The results prove that the system calibration is effective and that the reconstruction generated by the system is reliable.

5.6. Reconstruction of arbitrarily-placed patterns

Because we have validated the system, we use it to reconstruct arbitrarily placed patterns. To this end, we choose seven pairs of pictures that were taken by both cameras simultaneously. The pictures are shown in Fig. 8(a) and (b). The corresponding reconstructions, illustrated in Fig. 8(c), recover the scene of how

the seven patterns were placed in the 3D space. Fig. 8(d) shows the Y–Z view of the reconstructions, which clearly demonstrates that pattern No. 2 is displaced from pattern No. 1 by 125 μm along the Z axis, which is exactly the first case described in Section 5.5.

Now that we have successfully recovered the camera's rotation matrix \mathbf{R} for both cameras, we add the left and right camera coordinate frames in the reconstruction figure to demonstrate the relation between the world coordinate system, the two camera coordinate systems and the arbitrarily placed patterns. Due to the very small size of the pattern, to better demonstrate the relation, we adjusted the translational values of both cameras but did not change their rotation matrices. How the patterns were placed in the world space can be clearly observed in Fig. 8(e), and the relation can be expressed as follows:

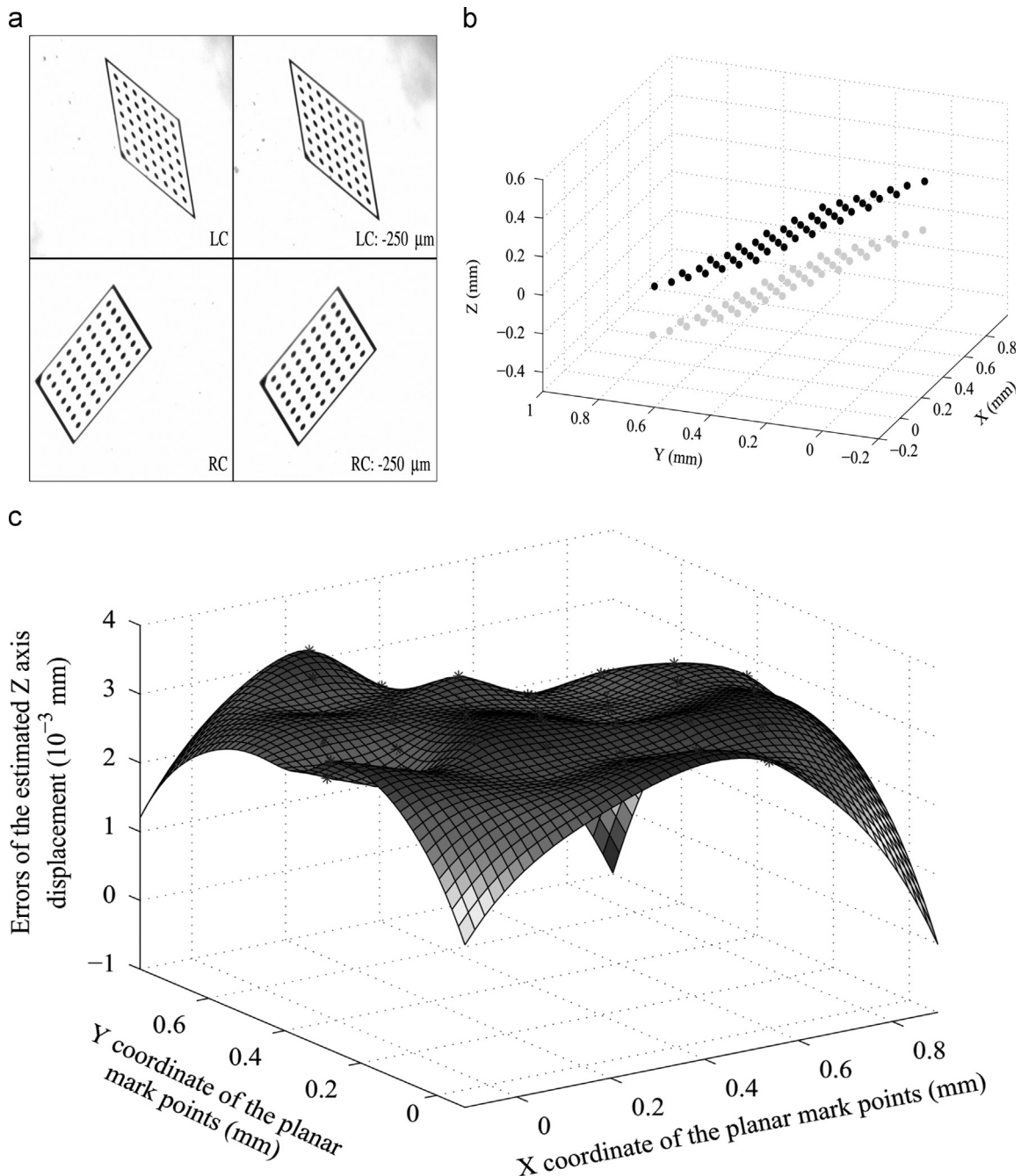


Fig. 7. The reconstruction and measurement of the pattern in an arbitrarily placed orientation after a $-250\text{-}\mu\text{m}$ displacement along the Z axis of the world coordinate system. (a) Pictures taken by the left and right cameras before and after the $-250\text{-}\mu\text{m}$ displacement. (b) The upper and bottom grid plates are, respectively, the reconstructions of the pattern before and after the displacement. (c) Errors of the displacement estimated by the stereo micro-vision system.

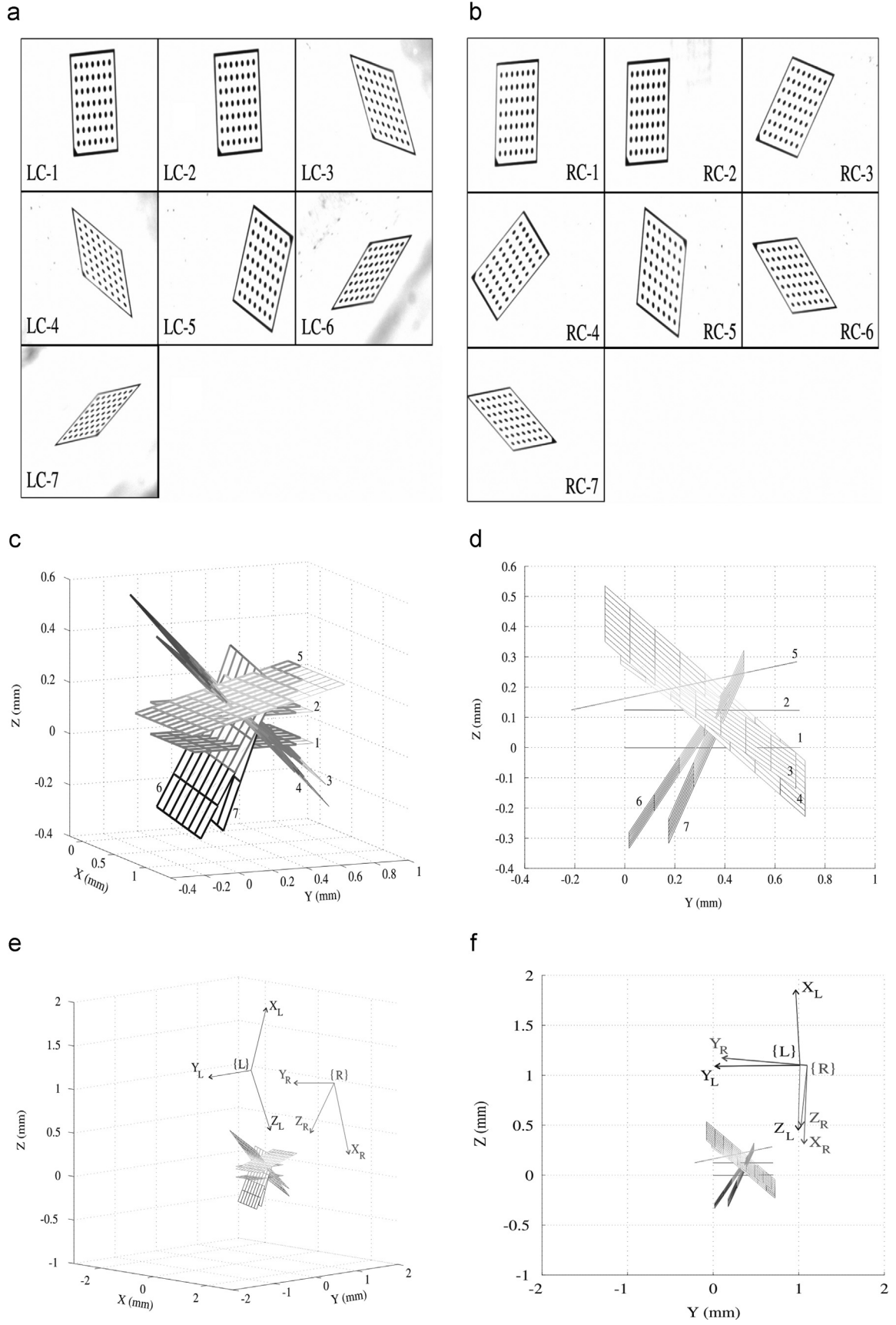


Fig. 8. Reconstructions of seven arbitrarily-placed patterns. (a) Seven pictures of the arbitrarily placed pattern taken by the left camera. (b) The corresponding seven pictures taken by the right camera. (c) Reconstructions of the seven patterns. (d) Y-Z view of the reconstructions. (e) The reconstructions with the left camera and right camera coordinate systems on. (f) Y-Z view of the reconstructions with the left camera and right camera coordinate systems on.

Pattern No. 1 defines a common world coordinate frame for both cameras. The {L} and {R} coordinate frames indicate the left and right camera coordinate systems, respectively. The roll-pitch-yaw angles of the left camera coordinates relative to the world coordinates are 176.26° , 49.22° and 4.91° ; the right camera coordinates are 173.24° , -52.08° and -2.47° . As shown in Fig. 8 (f), the two cameras are not strictly perpendicular to each other.

From the reconstructions, we can immediately obtain the orientations of all of the other six patterns in the reference world coordinate system defined by the first pattern. The reason we only reconstructed a planar object is that the calibration pattern can directly provide point correspondences, which makes the reconstructions much easier. Because there is a triangle mark used for axis identification and mark point matching in one corner of the calibration pattern's rectangle frame, the image point correspondences can be easily detected. In fact, the system can reconstruct 3D micro-objects if a sufficient number of image point correspondences are detected. In addition, the reference world coordinate is defined to be Euclidean, so the reconstruction is the object's Euclidean structure.

6. Conclusion

A complete and accurate procedure for calibrating a telecentric stereo micro-vision system is proposed in this paper. The explicit and detailed procedure of the calibration is also presented. It is concluded that our method applies to any type of telecentric camera and can be extended to the calibration of affine cameras that perform orthographic projection.

The experimental results verify the effectiveness of our method. In sum, the method has three major advantages. First, in the case of limited FOV, compared to the 3D-object-based calibration method, our approach is more cost-effective because it avoids the utilization of an expensive micro-scale 3D calibration cube. Second, in the case of fixed camera calibration, the method is easy to implement and directly provides accurate measures of the camera's intrinsic and extrinsic parameters, which is quite beneficial and helpful in some occasions in which the camera's extrinsic parameters are required to be known. Third, because the 2D calibration pattern defines a Euclidean world coordinate space for both cameras, the structure reconstructed by using this method is Euclidean as well. Therefore, we do not have to recover the real world structure from its affine reconstruction, which is just what the self-calibration method normally needs to deal with after the camera is calibrated.

Furthermore, even in some cases in which the camera moves, this method is still applicable. Therefore, our method possesses good generality of application.

Acknowledgment

This work was supported by the Major Program of Natural Science Foundation of China (No. 91223201) and also by the Science and Technology Planning Project of Guangdong Province, China (Nos. 2012B011300010 and 2012B091100141).

References

- [1] Schreier HW, Garcia D, Sutton MA. Advances in light microscope stereo vision. *Exp Mech* 2004;44:278–88.
- [2] Sutton MA, Ke X, Lessner SM, Goldbach M, Yost M, Zhao F, et al. Strain field measurements on mouse carotid arteries using microscopic three-dimensional digital image correlation. *J Biomed Mater Res Part A* 2008;86A:178–90.
- [3] Hu Z-X, Luo H-Y, Du Y-J, Lu H-B. Fluorescent stereo microscopy for 3D surface profilometry and deformation mapping. *Opt Express* 2013;21:11808–18.
- [4] Arai Y, Shimamura R, Yokozeki S. Dynamic out-of-plane deformation measurement using virtual speckle patterns. *Opt Lasers Eng* 2009;47:563–9.
- [5] Li EB, Tieu AK, Yuen WYD. Application of digital image correlation technique to dynamic measurement of the velocity field in the deformation zone in cold rolling. *Opt Lasers Eng* 2003;39:479–88.
- [6] Orteu JJ. 3-D computer vision in experimental mechanics. *Opt Lasers Eng* 2009;47:282–91.
- [7] Luo P-F, Huang F-C. Application of stereo vision to the study of mixed-mode crack-tip deformations. *Opt Lasers Eng* 2000;33:349–68.
- [8] Teng C-H, Chen Y-S, Hsu W-H. Camera self-calibration method suitable for variant camera constraints. *Appl Opt* 2006;45:688–96.
- [9] Faugeras OD, Luong QT, Maybank SJ. Camera self-calibration: theory and experiments. In: Proceedings of the 2nd European conference on computer vision. Santa Margherita Ligure; 1992. p. 321–34.
- [10] Maybank SJ, Faugeras OD. A theory of self-calibration of a moving camera. *Int J Comput Vis* 1992;8:123–51.
- [11] Zhang Z-Y. A flexible new technique for camera calibration. *IEEE Trans Pattern Anal Mach Intell* 2000;22:1330–4.
- [12] Heikkilä J, Silven O. A four-step camera calibration procedure with implicit image correction. In: Proceedings of the IEEE conference on computer vision and pattern recognition. San Juan; 1997. p. 1106–12.
- [13] Tsai R. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE J Robot Automat* 1987;3:323–44.
- [14] Horaud R, Christy S, Mohr R. Euclidean reconstruction and affine camera calibration using controlled robot motions. In: Proceedings of IEEE international conference on intelligent robots and systems. Grenoble; 1997. p. 1575–82.
- [15] Quan L. Self-calibration of an affine camera from multiple views. *Int J Comput Vis* 1996;19:93–105.
- [16] Tomasi C, Kande T. Shape and motion from image streams under orthography: a factorization method. *Int J Comput Vis* 1992;9:137–54.
- [17] Koenderink JJ, van Doorn AJ. Affine structure from motion. *J Opt Soc Am A* 1992;18:377–85.
- [18] Kahl F, Heyden A. Affine structure and motion from points, lines, and conics. *Int J Comput Vis* 1999;33:163–80.
- [19] Shaprio LS, Zisserman A, Brandy M. 3D motion recovery via affine epipolar geometry. *Int J Comput Vis* 1995;16:147–82.
- [20] Gorpas DS, Politopoulos K, Yova D. Development of a computer vision binocular system for non-contact small animal model skin cancer tumour imaging. In: Proceedings of the SPIE diffuse optical imaging of tissue. Munich; 2007.
- [21] Haskamp K, Kästner M, Reithmeier E. Accurate calibration of a fringe projection system by considering telecentricity. In: Proceedings of the SPIE optical measurement systems for industrial inspection VII. Munich; 2011.
- [22] Dong L, Tian J-d. An accurate calibration method for a camera with telecentric lenses. *Opt Lasers Eng* 2013;51:538–41.
- [23] Lanman D, Hauagge DC, Taubin G. Shape from depth discontinuities under orthographic projection. In: Proceedings of the 2009 IEEE 12th international conference on computer vision workshops. Kyoto; 2009. p. 1550–7.
- [24] Pan B, Yu L-P, Wu D-F. High-accuracy 2D digital image correlation measurements with bilateral telecentric lens: error analysis and experimental verification. *Exp Mech* 2013;53:1719–33.
- [25] Steger C, Ulrich M, Wiedemann C. Machine vision algorithms and applications. New Jersey: Wiley; 2008([chapter 3]).
- [26] Hartley R, Zisserman A. Multiple view geometry in computer vision. 2nd ed.. UK: Cambridge University Press; 2004; 87–109.
- [27] Watanabe M, Nayar S. Telecentric optics for computational vision. In: Proceedings of the 4th European conference on computer vision. Cambridge, UK; 1996. p. 439–51.
- [28] Sutton MA, Orteu J-J, Schreier HW. Image correlation for shape, motion and deformation measurements. New York: Springer-Verlag; 2009; 27–30.
- [29] Weng J, Cohen P, Herniou M. Camera calibration with distortion models and accuracy evaluation. *IEEE Trans Pattern Anal Mach Intell* 1992;14:965–80.
- [30] Pan B, Yu L-P, Wu D-F, Tang L-Q. Systematic errors in two-dimensional digital image correlation due to lens distortion. *Opt Lasers Eng* 2013;51:140–7.
- [31] Moré JJ. The Levenberg–Marquardt algorithm: implementation and theory. *Numer Anal* 1987;630:105–16.