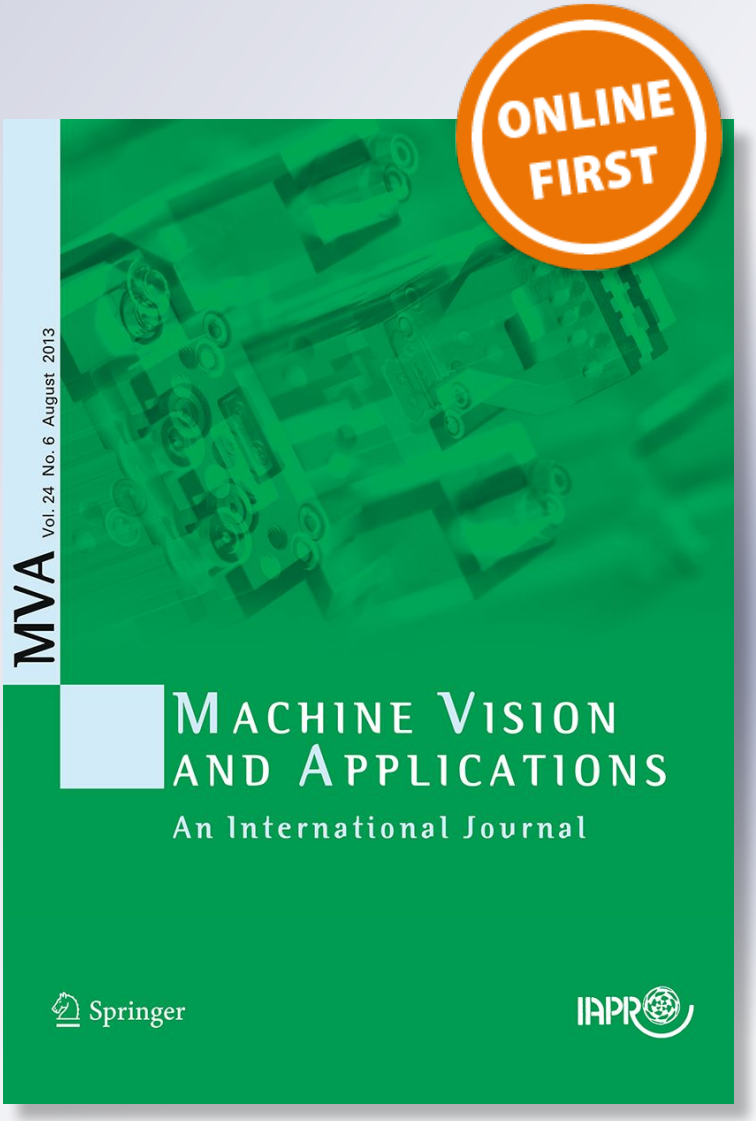
A closed-form single-pose calibration method for the camera–projector system

Zexi Feng, Zhiquan Cheng & Zhan Song

Machine Vision and Applications

ISSN 0932-8092

Machine Vision and Applications
DOI 10.1007/s00138-019-01002-2



Springer

Machine Vision and Applications
https://doi.org/10.1007/s00138-019-01002-2

ORIGINAL PAPER

A closed-form single-pose calibration method for the camera–projector system

Zexi Feng[1,2] · Zhiquan Cheng[1,3] · Zhan Song[1]

Received: 4 March 2018 / Revised: 8 October 2018 / Accepted: 2 January 2019

Abstract
To improve the calibration efficiency of the structured light stereo vision system, a closed-form single-pose calibration method is proposed. Previous works calibrate the structured light stereo vision system via the planar-based method since the world coordinates of the projected pixels are easy to find. But the disadvantage of these methods is that multiple poses are required, so that the mechanical movement is necessary in these calibration methods. In this paper, the inherent geometry between the spheres and their image is exploited deeply. Thus, a sphere-based calibration method is proposed. The proposed method first calibrates the camera from the image of the absolute conic through an improved algorithm. Then, the three-dimensional coordinates of the projected pixels are found via the inherent geometrical relation. Next, the projector is calibrated using the direct linear method. Finally, the distortion of the projector is found through the nonlinear optimization method with an improved updating direction. Experiments demonstrate that not only the calibration efficiency is improved, but also the calibration accuracy is improved.

Keywords Calibration · IAC · Optimization

1 Introduction

Structured light is a kind of image registration technique [1–3] in the stereo vision field. A typical structured light system consists of a camera and a projector. To get the depth information via the stereo vision technique, the first step is to find a set of matrixes for the camera–projector system that model the perspective transformation [4–7]. Then, the projector projects a sequence of specially designed patterns onto the target scene. Meanwhile, the camera captures these encoded images in the same order. Next, a set of strategies are employed to find out the correspondences between the image of the camera and the image of the projector. Finally, a linear equation [8] is employed to determine the depth information for each pair of the corresponding pixels.

To determine the depth information from the corresponding pixels, the perspective projection matrices for the camera and the projector must be determined beforehand. The method, which is employed to determine the perspective projection matrices, is called calibration. The traditional direct linear transformation (DLT) like method [4] employs a cube as the calibration artifact, in which, marker points are placed so that the world coordinate system can be set up. The disadvantage of the DLT like method is that the calibration artifact is not easy to produce.

Planar-based calibration method [9] uses chessboards as the calibration artifact. By taking the corners on the chessboard as the marker points, a three-dimensional world coordinate system is set up and every marker point is assigned a unique world coordinate. Then, the matrix, which is employed to model the perspective transformation of the camera, is determined by the linear mapping from the world coordinates to the image coordinates.

As the homography transformation proposed in [9] is suitable to determine the three-dimensional coordinate of the

✉ Zexi Feng
fengzexi2002@aliyun.com

Zhiquan Cheng
cheng.zhiquan@gmail.com

Zhan Song
zhan.song@siat.ac.cn

1 ShenZhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

2 University of Chinese Academy of Sciences, No. 19(A) Yuquan Road, Shijingshan District, Beijing 100049, China

3 Avatar Science Company, Changsha City, Hunan Province 410073, China

Published online: 18 January 2019

projected features, the checker board is employed to calibrate the camera–projector system in [10–13]. After the camera is calibrated, the three-dimensional coordinates corresponding to the projected pixels are determined. Finally, the projector, which is modeled as a camera but cannot capture image directly, is calibrated as a pseudo camera in the same way.

In addition to the chessboard, the sphere is employed as the calibration artifact [14–17]. If there are more than two spheres in an image, then the intrinsic parameters of the camera are determined linearly by the elliptical outlines. The advantages of employing sphere as the calibration artifact is that a single pose is enough to complete the calibration. But the sphere-based calibration method cannot be applied to the projector directly. This is because the outline of the sphere is illegible in the projector's view due to the angle of reflection. Besides, part of the outline is blocked due to the parallax between the camera and the projector, which destroys the elliptical outlines of the spheres.

Recent methods [18, 19] applied the sphere-based calibration method to the structured light system. By employing the sphere as the calibration artifact, the operations in the calibration process are simplified. However, these methods either employ additional constrains or are numerically unstable.

In this paper, a closed-form method is proposed to improve the calibration efficiency of the structured light stereo vision system. The proposed method first calibrates the camera through an improved IAC estimation method and then employs the gray code method [20] to get the pixel correspondences between the camera and the projector. After that, the projected features, which are located on the spheres, are translated into three-dimensional points via the inherent geometry. Finally, the projector is calibrated by the direct linear transform method and the trust region reflective method [21]. Because the proposed method calibrates the camera–projector system in a single pose, the calibration efficiency is improved. Experiments demonstrate that the propose method is accurate and robust.

This paper is organized as follows: Sect. 2 introduces the current theories for the camera–projector calibration problem. Section 3 presents a robust algorithm to determine the outline and the interior points from the image of the spheres. Section 4 presents the proposed calibration methods in detail. Section 5 shows the experimental result on real data. Section 6 discusses the result and compares the result with other method. Section 7 gives the conclusions.

2 Previous work

In the DLT like method, the camera is calibrated from the geometry between the world coordinates and the image points. Let $\boldsymbol{P}$ be the projection matrix of the camera, and let $\boldsymbol{p}_i$ be the ith row of $\boldsymbol{P}$. If the world coordinate system is set up and every marker point is assigned with a three-dimensional coordinate, then $\boldsymbol{P}$ is determined by the linear mapping from the world coordinate to the image coordinate. Let $\boldsymbol{G}_i = [a_i b_i c_i]$ be the world coordinate, and let $\boldsymbol{g}_i = [u_i v_i]$ be the corresponding image coordinate, then the mapping from $\boldsymbol{G}_i$ to $\boldsymbol{g}_i$ is $u_i = \left(\boldsymbol{p}_1 \cdot \bar{\boldsymbol{G}}_i^T\right)/\left(\boldsymbol{p}_3 \cdot \bar{\boldsymbol{G}}_i^T\right)$ and $v_i = \left(\boldsymbol{p}_2 \cdot \bar{\boldsymbol{G}}_i^T\right)/\left(\boldsymbol{p}_3 \cdot \bar{\boldsymbol{G}}_i^T\right)$. Here, $\bar{\boldsymbol{G}}_i = [\boldsymbol{G}_i 1]$ is the augmented vector of $\boldsymbol{G}_i$. After $\boldsymbol{g}_i$ is detected in the feature point detection procedure, $\boldsymbol{P}$ is found in (1).

$$\begin{bmatrix} \cdots & \cdots & \cdots \\ \overline{\mathbf{G}_i} & \mathbf{0} & -u_i \cdot \overline{\mathbf{G}_i} \\ \mathbf{0} & \overline{\mathbf{G}_i} & -v_i \cdot \overline{\mathbf{G}_i} \\ \cdots & \cdots & \cdots \end{bmatrix} \begin{bmatrix} \mathbf{p}_1^T \\ \mathbf{p}_2^T \\ \mathbf{p}_3^T \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \tag{1}$$

Equation (1) is solved by taking the singular vector that corresponds to its smallest singular value as the solution. If the number of points in (1) is large, then the singular value decomposition algorithm may exceed the memory limit. In this situation, (1) is solved by setting the first element of $\boldsymbol{P}$ to 1 and then employs the least square method to find the solution. Let $||\boldsymbol{p}_3||$ be the 2-norm of $\boldsymbol{p}_3$. Because $\boldsymbol{p}_3$ is the rotate axis of the coordinate system, so that $\boldsymbol{P}$ is scaled by $1/||\boldsymbol{p}_3||$ after the solution is found.

After $\boldsymbol{P}$ is determined, the internal matrix $\boldsymbol{K}$ and the relative position, which are $\boldsymbol{R}$ and $\boldsymbol{T}$, is determined by the QR decomposition algorithm. Due to the distortion, the decomposed $\boldsymbol{R}$ is not a strictly Rodrigues rotation matrix [22], so that $\boldsymbol{R}$ is refined by a singular value-based approximating processes in [9]. In [9], when the initial value of $\boldsymbol{K}$, $\boldsymbol{R}$ and $\boldsymbol{T}$ is determined, the distortion parameters are solved by the Levenberg–Marquardt algorithm [23].

As the three-dimensional artifact, which is employed to setup the world coordinate system, is not easy to manufacture, the planar-based method [9] is developed to calibrate the camera. The planar-based method employs checker board as the calibration artifact. Because the world coordinates corresponding to the checker board are two dimensional, the third column of $\boldsymbol{P}$ cannot be determined directly. Hence, multiple poses are required to determine the parameters of the camera.

Because the projector cannot be calibrated by itself, previous methods calibrate the projector with the help of the camera. As the features are easy to track when they are projected onto a plane, the planar-based method is employed to calibrate the projector in [10–13, 19]. Because the projector is treated as a pseudo camera, the projector is calibrated after the camera. Once the camera is calibrated, the points in the captured patterns are translated into world coordinates via the homography matrix $\boldsymbol{H}$. Let $\boldsymbol{P}_{\text{cam}}$ be the projection matrix of the camera, then $\boldsymbol{H}^{-1}$ is determined by the first column,

the second column and the forth column of $\boldsymbol{P}_{\text{cam}}$. Let $\boldsymbol{G}_i = [a_i b_i 0]$ be the world coordinates of the projected features, and let $\bar{\boldsymbol{G}}_i = [a_i b_i 1]$ be the augmented vector of $\boldsymbol{G}_i$. Let $\bar{\boldsymbol{g}}_i = [u_i v_i 1]$ be the homogeneous vector of the captured pixel, and let $\boldsymbol{h}_i$ be the ith row of $\boldsymbol{H}^{-1}$, then the mapping from $\bar{\boldsymbol{g}}_i$ to $\bar{\boldsymbol{G}}_i$ is determined as $\bar{\boldsymbol{G}}_i = \lambda \boldsymbol{H}^{-1} \bar{\boldsymbol{g}}_i^{\mathrm{T}}$. Here, $\lambda = 1/\left(\boldsymbol{h}_3 \cdot \bar{\boldsymbol{g}}_i^{\mathrm{T}}\right)$ is the scale factor. In [10–12], after the world coordinates of the projected pattern are determined, the internal matrix and the relative position of the projector are determined via the planar-based method.

As mentioned in [14–17], the internal matrix of the camera could also be determined by the image of spheres. By decomposing the matrixes of ellipses, the image of the absolute conic is found. In [16], the absolute conic is found by (2) where $\boldsymbol{\omega} = \boldsymbol{K}^{-\mathrm{T}} \boldsymbol{K}^{-1}$ is the absolute conic, $\boldsymbol{l}$ is the vanishing line, and $\boldsymbol{v}$ is the vanishing point. Let $\bar{\boldsymbol{g}}_i^{\mathrm{T}} \boldsymbol{C}_1 \bar{\boldsymbol{g}}_i = 0$ and $\bar{\boldsymbol{g}}_j^{\mathrm{T}} \boldsymbol{C}_2 \bar{\boldsymbol{g}}_j = 0$ be the equation of 2 ellipses, then $\boldsymbol{l}$ and $\boldsymbol{v}$ are found from $\boldsymbol{C}_1 \boldsymbol{C}_2^{-1}$ as the eigenvector and the by-product of the eigenvectors in [16]. If there are three ellipses, then there are three pairs of $\boldsymbol{l}$ and $\boldsymbol{v}$. Hence, the image of the absolute conic is linearly solved. Because the eigenvalues of $\boldsymbol{\omega}$ are not 1, the 2-norm of $\boldsymbol{l}$ and $\boldsymbol{v}$ is different. In [16], the 2-norm of $\boldsymbol{l}$ and $\boldsymbol{v}$ is scaled by the eigenvalues of $\boldsymbol{C}_1 \boldsymbol{C}_2^{-1}$. The advantage of employing spheres as the calibration artifact is that sphere is easy to obtain and a single pose is enough to complete the calibration.

$$\mathbf{l} = \boldsymbol{\omega} \mathbf{v} \tag{2}$$

In [19], the sphere-based calibration method is applied to the structured light system. Their method first calibrates the camera from the image of spheres via a nonlinear approach and then calibrates the projector via the planar-based method. Although the camera is calibrated in a single pose, their calibration method requires multiple poses to find a closed-form solution.

With unmarked spheres, the internal parameters of the camera are determined, but the projection matrix of the projector is not determined. Although the exact shape of the calibration artifact is known, the three-dimensional coordinates of the projected pixels are undetermined. If the three-dimensional coordinates of the projected pixels can be found, then the projector can be calibrated using (1). In this paper, an improved algorithm, which is employed to find the image of the absolute conic, is proposed. With the proposed algorithm, the accuracy of the absolute conic is improved. Besides, a closed-form method is proposed to find the three-dimensional coordinates on the scanned spheres. The proposed method determines the corresponding three-dimensional coordinates via the inherent geometrical relation. The detail of the proposed method is described in the following sections.

Table 1 Symbols and properties in Sect. 3

Symbol	Property
$\boldsymbol{A}_1$	A vector, which is the major axis of the three-dimensional ring
$\boldsymbol{A}_2$	A vector, which is the minor axis of the three-dimensional ring
$\boldsymbol{A}_4$	A vector, which is the center of the three-dimensional ring
$\boldsymbol{Q}_i$	A point on the ring, which satisfies $\boldsymbol{Q}_i = \cos\theta \boldsymbol{A}_1 + \sin\theta \boldsymbol{A}_2 + \boldsymbol{A}_4$
$\boldsymbol{q}_i$	A point on the image of the ring
$\bar{\boldsymbol{q}}_i$	The augmented coordinate of $\boldsymbol{q}_i$, which satisfies $\boldsymbol{q}_i = \lambda_i \boldsymbol{K} \boldsymbol{Q}_i$
$\{\boldsymbol{m}_i\}$	A collection of pixels, which is on the image of spheres
$\{\boldsymbol{n}_i\}$	A collection of pixels, which is on the background
L_j	The possible radius of a circular areas
M_k	The number of pixels in a connected area
th_j	An local threshold employed to segment the local image

3 Extracting the contours and the outlines of the spheres

Section 3 is organized as follows: Sect 3.1 introduces the inherent relation between the ellipse and the three-dimensional ring. Section 3.2 introduces the detail of the proposed extracting strategy. The symbols employed in Sect. 3 are shown in Table 1.

3.1 The inherent relation between the ellipse and the ring

Consider that a sphere is observed by a camera. The contour of the sphere is a three-dimensional ring, which is formulated in (3). Here, $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are two orthogonal vectors, $\boldsymbol{A}_4$ is the center of the ring, and θ_i is the parameter.

$$\mathbf{Q}_i = \cos\theta_i \cdot \mathbf{A}_1 + \sin\theta_i \cdot \mathbf{A}_2 + \mathbf{A}_4 \tag{3}$$

Let $\boldsymbol{K}$ be the intrinsic matrix of the camera, then the image points of the ring is $\bar{\boldsymbol{q}}_i = \lambda_i \cdot \boldsymbol{K} \boldsymbol{Q}_i$. Here, $\boldsymbol{q}_i$ is the image point, and $\bar{\boldsymbol{q}}_i = \left[\boldsymbol{q}_i; 1\right]$ is the augmented vector of $\boldsymbol{q}_i$, and λ_i is a scale factor. If the sphere is located at the center of the camera, then $\boldsymbol{A}_1$ is $[l; 0; 0]$, $\boldsymbol{A}_2$ is $[0; l; 0]$, and $\boldsymbol{A}_4$ is $[0; 0; d]$. Here, l is the side length of the ring, and d is the distance from the camera to the center of the ring. In this case, $\boldsymbol{q}_i$ satisfies (4).

$$\mathbf{q}_i = \cos\theta_i \cdot \begin{bmatrix} f_u \cdot l/d \\ 0 \end{bmatrix} + \sin\theta_i \cdot \begin{bmatrix} s \cdot f_v \cdot l/d \\ f_v \cdot l/d \end{bmatrix} + \begin{bmatrix} c_u \\ c_v \end{bmatrix} \tag{4}$$

Equation (4) suggests that $\boldsymbol{q}_i$ satisfies $\boldsymbol{q}_i = \cos\theta_i \cdot \boldsymbol{n} + \cos\theta_i \cdot \boldsymbol{m} + \boldsymbol{c}$. Although $\boldsymbol{n}$ and $\boldsymbol{m}$ are not orthogonal to each other, $\boldsymbol{q}_i$ is still a two-dimensional ellipse. A simple proof is given below: Suppose $\boldsymbol{n} = [1, 0]$, $\boldsymbol{m} = [o_x, o_y]$ and $\boldsymbol{c} = [0, 0]$, then any general cases can be considered as a linear transformation of the given case with a scale factor. So that any point $\boldsymbol{q}_i = [\delta_i, \varepsilon_i]$, which is located on the non-standard ellipse, is determined as $\delta_i = \cos\theta_i + \sin\theta_i \cdot o_x$ and $\varepsilon_i = \sin\theta_i \cdot o_y$. It is easy to verify that there is a group of value $b = -2o_x/o_y$, $f = (1 + o_x^2)/o_y^2$, which satisfies an elliptic function of $\delta_i^2 + b \cdot \delta_i\varepsilon_i + f \cdot \varepsilon_i^2 - 1 = 0$. Hence, (4) is an ellipse.

Let the equation of the two-dimensional ellipse be $a \cdot x^2 + b \cdot xy + c \cdot y^2 + \mathrm{e} \cdot x + g \cdot y + h = 0$. Rewritten the function of the ellipse as a symmetric matrix, then $\bar{\boldsymbol{q}}_i$ satisfies (5).

$$\bar{\mathbf{q}}_i^{\mathrm{T}} \begin{bmatrix} a & b/2 & e/2 \\ b/2 & c & g/2 \\ e/2 & g/2 & h \end{bmatrix} \bar{\mathbf{q}}_i = 0 \tag{5}$$

If the sphere is not located at the center of the camera, then there is a rotation matrix $\boldsymbol{R}$ that could transform the sphere back to the center of the camera. Let $\boldsymbol{O}_i$ be the ring that is not located at the center of the camera, and let $\boldsymbol{Q}_i = \boldsymbol{R}\boldsymbol{O}_i$ be the transformed ring that is located at the center of the camera. Let $\boldsymbol{o}_i$ be the image points of $\boldsymbol{Q}_i$, then $\boldsymbol{o}_i$ satisfies (6).

$$\bar{\mathbf{o}}_i^{\mathrm{T}} \mathbf{R}^{\mathrm{T}} \begin{bmatrix} a & b/2 & e/2 \\ b/2 & c & g/2 \\ e/2 & g/2 & h \end{bmatrix} \mathbf{R}\bar{\mathbf{o}}_i = 0 \tag{6}$$

Equation (6) suggests that no matter where the sphere is, the image of its contour is an ellipse.

3.2 Extraction ellipses from the image

Because the minimal number of spheres, which is required in the calibrating process, is 3, there are several elliptical regions in a binary image. To separate and extract these elliptical regions reliably, an extracting strategy is proposed in this paper.

Suppose there are 2 types of pixels in the binary map. Let the first type be the image of spheres, and let the second type be the image of the background. Let $\{\boldsymbol{m}_i\}$ be the coordinates of the pixels belong to the first type, and let $\{\boldsymbol{n}_j\}$ be the coordinate of pixels belong to the second type, so that $\{\boldsymbol{m}_i\}$ and $\{\boldsymbol{n}_j\}$ can be separated by setting a global threshold value th_1 to the gray image. After $\{\boldsymbol{m}_i\}$ is found, the connected area is determined by [24]. For each of the connected regions in $\{\boldsymbol{m}_i\}$, a possible radius is determined by (7). Here, M_k is the number of pixels in the connected area. After the possible radius is found, the center of the circular area is determined as the geometrical center of the connected area. Let $\{L_j\}$ be the possible radius of the connected regions. If there are k spheres, then the elliptical regions of the spheres are determined as the connected regions corresponding to the k largest values in $\{L_j\}$. The elliptical extracting process is shown in Fig. 1. Here, the gray levels in Fig. 1 is linearly related to the possible radius of the connected regions.

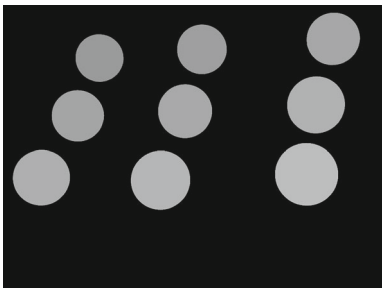


Fig. 1 Find the elliptical areas

$$L = \sqrt{M_k/\pi} \tag{7}$$

Let $\{\boldsymbol{c}_j\}$ be the center of the jth sphere. After the elliptical regions are found, the edges and the weights of the elliptical regions are determined in the following three steps. Firstly, a local threshold value for the jth sphere is determined by (8). Here, w_k is the image response in the gray image, S is the pixels in $\{\boldsymbol{n}_j\}$ whose distance to $\{\boldsymbol{c}_j\}$ is within $(\beta_1 \cdot L_j, \beta_2 \cdot L_j)$, while α is the parameter. Equation (8) means that the average gray level of the surrounding pixels is taken as the basis of the local threshold. Secondly, the region of the interested image, for which the distance to $\{\boldsymbol{c}_j\}$ is less than $\beta_2 \cdot L_j$, is segmented by th_j. Thirdly, the image points, whose Euclidean distance to the background is 1, are taken as the edge points. Besides, the gray levels corresponding to the edge points are taken as the weights.

$$\mathrm{th}_j = \alpha \cdot \sum_{k \subset S} w_k / \mathrm{count}(S) \tag{8}$$

After the edge points and the weights are found, a weighted vision of Fitzgibbon's ellipse fitting method [25] is employed to find the function of the ellipse. In [25], the initial center is determined by the mean value of the edge points, and the polynomial terms are separated to the quadratic terms and the linear terms. In the weighted vision, the initial center is determined by the weighted mean value of the edge points, while the quadratic terms and the linear terms are multiplied by the weights. With the weights of the edge points, the images of the rings are tracked more accurately.

Table 2 Additional symbols and properties in Sect. 4

Symbol	Property
$\boldsymbol{C}_i$	The matrix representation of an ellipse, the corresponding function of the ellipse is $ax^2 + bxy + cy^2 + dx + ey + f = 0$
$\boldsymbol{\omega}$	A matrix, which is the image of the absolute conic
$\boldsymbol{l}$	A vanishing line found by a combination of $\boldsymbol{C}_i\boldsymbol{C}_j^{-1}$
$\boldsymbol{v}$	A vanishing line found by a combination of $\boldsymbol{C}_i\boldsymbol{C}_j^{-1}$
ω_{ij}^*	The element of $\boldsymbol{\omega}^{-1}$
$\boldsymbol{skc}$	The distortion of the camera
$\boldsymbol{skp}$	The distortion of the projector
$\{\boldsymbol{E}_j\}_i$	A collection of the edge points on the image of the ith sphere
$\{\boldsymbol{E}'_j\}_i$	The corresponding edge points on the normalized image
$\boldsymbol{q}_j$	A corresponding pixel in the image of the camera, which is obtained from the gray code method
$\boldsymbol{p}_j$	A corresponding pixel in the image of the projector, which is obtained from the gray code method
$\boldsymbol{Q}_j$	The corresponding three-dimensional points of $\boldsymbol{q}_j$ and $\boldsymbol{p}_j$
$\boldsymbol{A}_i$	The axis and the center of a sphere, which satisfies $\boldsymbol{Q}_i = \sin\alpha \cdot \cos\beta \boldsymbol{A}_1 + \sin\alpha \cdot \sin\beta \boldsymbol{A}_2 + \cos\alpha \boldsymbol{A}_3 + \boldsymbol{A}_4$
e_x	The horizontal re-projection error of the image points
e_y	The vertical re-projection error of the image points

4 The proposed calibration method

In this paper, the camera is calibrated from the outline of the spheres. This idea is borrowed from Zhang et al. [16]. To suit for more than three spheres and make the result numerically stable, an improved version of Zhang's algorithm is proposed in the first part of this section.

In the second part of this section, a closed-form method, which is employed to determine the spatial position of the projected points, is proposed. When the spatial positions of the projected points are determined, the projector is calibrated by (1). Additionally, the partial derivative, which is employed for the re-projection error, is researched and explained clearly. The additional symbols employed in Sect. 4 are shown in Table 2.

4.1 Calibrate the camera

4.1.1 The improved calibration algorithm

Let $\{\boldsymbol{C}_i\}$ be the elliptical functions of the spheres found in the previous section. In [16], it is proven that the image of the absolute conic is determined by its polar and pole, while the polar and pole are determined by $\boldsymbol{C}_i\boldsymbol{C}_j^{-1}$. But, the problem is that the eigenvalue of $\boldsymbol{\omega} = \boldsymbol{K}^{-\mathrm{T}}\boldsymbol{K}^{-1}$, which is mentioned in (2), is not 1. Hence, it will change the size of $\boldsymbol{v}$. In [16], $\boldsymbol{l}$ and $\boldsymbol{v}$ are scaled by the eigenvalues of $\boldsymbol{C}_i\boldsymbol{C}_j^{-1}$. But, the problem is that the eigenvalues of $\boldsymbol{C}_i\boldsymbol{C}_j^{-1}$ is different from that of $\boldsymbol{\omega}$. This is because $\boldsymbol{C}_i\boldsymbol{C}_j^{-1} = \boldsymbol{K}^{-\mathrm{T}}\mathrm{diag}\{1, 1, \gamma_2^2/\gamma_1^2\}\boldsymbol{K}^{\mathrm{T}}$ and γ_k is different for each of the ellipses. Hence, γ_2^2/γ_1^2 is not 1. For this reason, the eigenvalues of $\boldsymbol{\omega}$ cannot be determined from $\boldsymbol{C}_i\boldsymbol{C}_j^{-1}$. If $\boldsymbol{\omega}$ is found by scaling $\boldsymbol{l}$ and $\boldsymbol{v}$ according to the eigenvalues of $\boldsymbol{C}_i\boldsymbol{C}_j^{-1}$, then the solution is not accurate. In this paper, an improved algorithm is proposed. Experiments demonstrate that the accurate solution is found by the improved algorithm.

Because $\boldsymbol{l}$ and $\boldsymbol{v}$ must be scaled, (2) is changed to (9).

$$\boldsymbol{\omega}^{-1}\mathbf{l} = \lambda \cdot \mathbf{v} \tag{9}$$

Because $\boldsymbol{\omega}^{-1}$ (also referred to as $\boldsymbol{\omega}^*$) is a symmetric matrix, so that the number of unknowns in (9) is $6+k$, where k is the number of spheres. If there are at least 3 spheres, then $\boldsymbol{\omega}^{-1}$ is determined. Change (9) to a matrix format as shown in (10), then $\boldsymbol{\omega}^{-1}$ is found. Here l_{ij} is the element of the line $\boldsymbol{l}_{\mathrm{i}}$ in the ith combination of $\boldsymbol{C}_i$ and $\boldsymbol{C}_j^{-1}$, $\boldsymbol{v}_{\mathrm{i}}$ is determined by the eigenvectors of $\boldsymbol{C}_i\boldsymbol{C}_j^{-1}$, while v_{ij} is the element of $\boldsymbol{v}_{\mathrm{i}}$.

$$\begin{bmatrix} v_{11} & v_{12} & v_{13} & 0 & 0 & 0 & l_{11} & 0 & \cdots \\ 0 & v_{11} & 0 & v_{12} & v_{13} & 0 & l_{12} & 0 & \cdots \\ 0 & 0 & v_{11} & 0 & v_{12} & v_{13} & l_{13} & 0 & \cdots \\ v_{21} & v_{22} & v_{23} & 0 & 0 & 0 & 0 & l_{21} & \cdots \\ 0 & v_{21} & 0 & v_{22} & v_{23} & 0 & 0 & l_{22} & \cdots \\ 0 & 0 & v_{21} & 0 & v_{22} & v_{23} & 0 & l_{23} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \begin{bmatrix} \omega_{11}^* \\ \omega_{12}^* \\ \omega_{13}^* \\ \omega_{23}^* \\ \omega_{24}^* \\ \omega_{33}^* \\ \lambda_1 \\ \lambda_2 \\ \vdots \end{bmatrix} = \mathbf{0} \tag{10}$$

To avoid a trivial solution, (10) is solved by taking the singular vector corresponding to its smallest singular value as the solution.

After $\boldsymbol{\omega}^{-1}$ is determined, then the intrinsic parameters of the camera are determined by applying the Cholesky factorization to $\boldsymbol{\omega}^{-1}$.

For multiple spheres, the center of each sphere is determined by the lines, which cross the projecting center of the spheres. As shown in Fig. 2, the lines, which cross the projecting center of $\boldsymbol{C}_i$, are found by $\boldsymbol{C}_i\boldsymbol{C}_j^{-1}$. After the lines, which cross each pair of ellipses $\boldsymbol{C}_i$ and $\boldsymbol{C}_j$, are found, the projection center of $\boldsymbol{C}_i$ is determined as the intersection point of the lines. In this paper, the intersection point is found using the least square method.

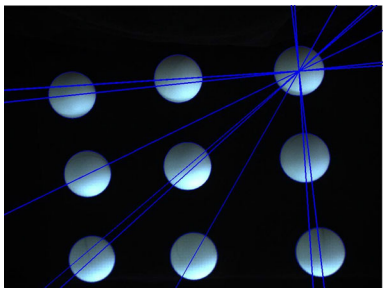


Fig. 2 The center of the ellipse is determined by the decomposed lines

4.1.2 Find the distortion of the camera

After the initial value of the intrinsic matrix is solved, the distortion of the camera is found along with the optimized value of the intrinsic matrix.

Let $\boldsymbol{K}_{\text{cam}}$ be the intrinsic parameters of the camera determined from the above process. Because the distortion is not considered in the above process, so that the trust region reflective method [21] is employed to find the distortion, along with the optimized value of the internal parameters.

In this paper, the first, the second and the third orders of the radial distortion coefficients are considered. Besides, the vertical and the horizontal distortion coefficients are considered. Let $\boldsymbol{skc} = [k_1, k_2, k_3, k_4, k_5]$ be the distortion coefficients of the camera, then the undistorted points are found in (11).

$$\begin{cases} x = (q_x - cc_x)/fc_x \\ y = (q_y - cc_y)/fc_y \\ d = x^2 + y^2 \\ x' = 2k_4xy + k_5(d + 2x^2) \\ y' = 2k_4(d + 2y^2) + k_5xy \\ x'' = x(1 + k_1d + k_2d^2 + k_3d^3) + x' \\ y'' = y(1 + k_1d + k_2d^2 + k_3d^3) + y' \end{cases} \tag{11}$$

If the perfect distortion coefficient is found, then the undistorted edge should be a perfect ellipse. Hence, the distance to the fitted ellipse is considered as the fitting error. Let $\{\boldsymbol{E}_j\}_i$ be the edge points in the ith ring, let $\left\{\boldsymbol{E}'_j\right\}_i$ be the undistorted points, and let $\boldsymbol{C}'_i$ be the fitted ellipse of the ith ring (normalized). After $\left\{\boldsymbol{E}'_j\right\}_i$ is found by (11), $\boldsymbol{C}'_i$ is determined by the weighted vision of Fitzgibbon's method. Hence, the distance from $\left\{\boldsymbol{E}'_j\right\}_i$ to $\boldsymbol{C}'_i$ is determined by (12). Here, $\boldsymbol{l}_A$ is the major axis of $\boldsymbol{C}'_i$, $\boldsymbol{l}_I$ is the minor axis of $\boldsymbol{C}'_i$, $[c_x, c_y]^{\mathrm{T}}$ is the center of $\boldsymbol{C}'_i$, while θ is the parameter. Equation (12) is solved by changing $\sin\theta$ to $\pm\sqrt{1-\cos^2\theta}$ and then taking the solution, which is closer to $[x'', y'']^{\mathrm{T}}$, as the solution.

$$\begin{aligned} &\arg\min(L(q_x, q_y)) \\ &\text{s.t.} \begin{cases} \begin{bmatrix} q_x \\ q_y \end{bmatrix} = \cos\theta \cdot \vec{\boldsymbol{l}}_A + \sin\theta \cdot \vec{\boldsymbol{l}}_I + \begin{bmatrix} c_x \\ c_y \end{bmatrix} \\ L(x, y) : (y - c_y)(x'' - c_x) = (x - c_x)(y'' - c_y) \end{cases} \end{aligned} \tag{12}$$

After (12) is solved, the deviation between $[x'', y'']^{\mathrm{T}}$ and $[q_x, q_y]^{\mathrm{T}}$ is found. Hence, the distortion coefficient is found using the trust region reflective method. Let $\rho_x = p_x - cc_x$ and $\rho_y = p_y - cc_y$, then the partial derivative is found by (13) and (14).

$$\begin{cases} \partial e_x/\partial f_x = -\rho_x/fc_x^2 \cdot x'' + \rho_x/fc_x \cdot (-2k_1\rho_x^2/fc_x^3 - \\ 4k_2(\rho_x^2/fc_x^2 + \rho_y^2/fc_y^2)\rho_x^2/fc_x^3 - 6k_3(\rho_x^2/fc_x^2 + \\ \rho_y^2/fc_y^2)^2\rho_x^2/fc_x^3 - 2k_4\rho_x\rho_y/fc_x^2/fc_y - 6k_5\rho_x^2/fc_x^3) \\ \partial e_x/\partial f_y = \rho_x/fc_x \cdot (-2k_1\rho_y^2/fc_y^3 - 4k_2(\rho_x^2/fc_x^2 + \\ \rho_y^2/fc_y^2)\rho_y^2/fc_y^3 - 6k_3(\rho_x^2/fc_x^2 + \rho_y^2/fc_y^2)^2\rho_y^2/fc_y^3 - \\ 2k_4\rho_x\rho_y/fc_y^2/fc_x - 2k_5\rho_y^2/fc_y^3) \\ \partial e_x/\partial c_x = -x''/fc_x + \rho_x/fc_x \cdot (-2k_1\rho_x/fc_x^2 - \\ 4k_2(\rho_x^2/fc_x^2 + \rho_y^2/fc_y^2)\rho_x/fc_x^2 - 6k_3(\rho_x^2/fc_x^2 + \\ \rho_y^2/fc_y^2)^2\rho_x/fc_x^2 - 2k_4\rho_y/fc_x/fc_y - 6k_5\rho_x/fc_x^2) \\ \partial e_x/\partial c_y = \rho_x/fc_x \cdot (-2k_1\rho_y/fc_y^2 - 4k_2(\rho_x^2/fc_x^2 + \\ \rho_y^2/fc_y^2)\rho_y/fc_y^2 - 6k_3(\rho_x^2/fc_x^2 + \rho_y^2/fc_y^2)^2\rho_y/fc_y^2 - \\ 2k_4\rho_x/fc_x/fc_y - 2k_5\rho_y/fc_y^2) \\ \partial e_y/\partial f_x = \rho_y/fc_y \cdot (-2k_1\rho_x^2/fc_x^3 - 4k_2(\rho_x^2/fc_x^2 + \\ \rho_y^2/fc_y^2)\rho_x^2/fc_x^3 - 6k_3(\rho_x^2/fc_x^2 + \rho_y^2/fc_y^2)^2\rho_x^2/fc_x^3 - \\ 2k_5\rho_x\rho_y/fc_x^2/fc_y - 2k_4\rho_x^2/fc_x^3) \\ \partial e_y/\partial f_y = -\rho_y/fc_y^2 \cdot y'' + \rho_y/fc_y \cdot (-2k_1\rho_y^2/fc_y^3 - \\ 4k_2(\rho_x^2/fc_x^2 + \rho_y^2/fc_y^2)\rho_y^2/fc_y^3 - 6k_3(\rho_x^2/fc_x^2 + \\ \rho_y^2/fc_y^2)^2\rho_y^2/fc_y^3 - 2k_5\rho_x\rho_y/fc_y^2/fc_x - 6k_4\rho_y^2/fc_y^3) \\ \partial e_y/\partial c_x = -\rho_y/fc_y \cdot (-2k_1\rho_x/fc_x^2 - 4k_2(\rho_x^2/fc_x^2 + \\ \rho_y^2/fc_y^2)\rho_x/fc_x^2 - 6k_3(\rho_x^2/fc_x^2 + \rho_y^2/fc_y^2)^2\rho_x/fc_x^2 - \\ 2k_5\rho_y/fc_x/fc_y - 2k_4\rho_x/fc_x^2) \\ \partial e_y/\partial c_y = -y''/fc_y + \rho_y/fc_y \cdot (-2k_1\rho_y/fc_y^2 - \\ 4k_2(\rho_x^2/fc_x^2 + \rho_y^2/fc_y^2)\rho_y/fc_y^2 - 6k_3(\rho_x^2/fc_x^2 + \\ \rho_y^2/fc_y^2)^2\rho_y/fc_y^2 - 2k_5\rho_x/fc_x/fc_y - 6k_4\rho_y/fc_y^2) \end{cases} \tag{13}$$

$$\begin{cases} \partial e_x/\partial k_1 = \rho_x/fc_x \cdot (\rho_x^2/fc_x^2 + \rho_y^2/fc_y^2) \\ \partial e_x/\partial k_2 = \rho_x/fc_x \cdot (\rho_x^2/fc_x^2 + \rho_y^2/fc_y^2)^2 \\ \partial e_x/\partial k_3 = \rho_x/fc_x \cdot (\rho_x^2/fc_x^2 + \rho_y^2/fc_y^2)^3 \\ \partial e_x/\partial k_4 = 2\rho_x\rho_y/fc_x/fc_y \\ \partial e_x/\partial k_5 = 3\rho_x^2/fc_x^2 + \rho_y^2/fc_y^2 \\ \partial e_y/\partial k_1 = \rho_y/fc_y \cdot (\rho_x^2/fc_x^2 + \rho_y^2/fc_y^2) \\ \partial e_y/\partial k_2 = \rho_y/fc_y \cdot (\rho_x^2/fc_x^2 + \rho_y^2/fc_y^2)^2 \\ \partial e_y/\partial k_3 = \rho_y/fc_y \cdot (\rho_x^2/fc_x^2 + \rho_y^2/fc_y^2)^3 \\ \partial e_y/\partial k_4 = 2\rho_x\rho_y/fc_x/fc_y \\ \partial e_y/\partial k_5 = 3\rho_y^2/fc_y^2 + \rho_x^2/fc_x^2 \end{cases} \tag{14}$$

4.2 Calibrate the projector

In the previous part of this section, the camera is calibrated. When the feature points are projected onto the spheres, the corresponding points are found. Although the exact shape of the calibration artifact is known, the three-dimensional coordinates of the projected features are unknown. If the three-dimensional points of the projected feature are found, then the projector is calibrated by (1). In this part of this section, the proposed method translates the projected features, which are located on the spheres, into three-dimensional points via the inherent geometry. Hence, the projector is fully calibrated.

4.2.1 Translate the image points into three-dimensional points

Let $\{\boldsymbol{q}_j\}_i$ be the set of image points located within ellipse $\boldsymbol{C}_i$, and let $\{\boldsymbol{Q}_j\}_i$ be the corresponding three-dimensional points. Let $\boldsymbol{A}_1$, $\boldsymbol{A}_2$ and $\boldsymbol{A}_3$ be three unit vectors perpendicular to each other (the axes). Let $\boldsymbol{A}_4$ be the center of the sphere, and let I be the radius of the sphere. The parametric equation of the sphere, which is show in (15), is employed to find $\boldsymbol{Q}_j$.

$$\mathbf{X}_j = I(\mathbf{A}_1 \sin\alpha_j \cos\beta_j + \mathbf{A}_2 \sin\alpha_j \sin\beta_j + \mathbf{A}_3 \cos\alpha_j) + \mathbf{A}_4 \tag{15}$$

After the image points are corrected by (11), the edge points should be an ellipse and the projected features should be within that ellipse. Let $\boldsymbol{C}_i$ be the ellipse, and let $\boldsymbol{c}_i$ be the center of $\boldsymbol{C}_i$. If $\boldsymbol{C}_i$ is located at the distortion center of the image, then $\boldsymbol{A}_1$ should be $[1, 0, 0]$, $\boldsymbol{A}_2$ should be $[0, 1, 0]$, $\boldsymbol{A}_3$ should be $[0, 0, 1]$, and $\boldsymbol{A}_4$ should be $[0, 0, t]$. In this situation, the sphere is located at the center axes of the camera, and the image of the sphere is a circle in the normalized image. This means that α_j and β_j can be extracted from the image, while t_i can be determined by the geometry.

Let $\boldsymbol{B} = [B_0, B_1, B_2]$ be a point on the sphere and let $\boldsymbol{b} = [b_0, b_1]$ be the image of $\boldsymbol{B}$ in the normalized image. If the sphere is located at the center axes of the camera, then $\boldsymbol{b}$ is simplified to (16).

$$\begin{cases} b_0 = (I \sin\alpha_j \cos\beta_j)/(I \cos\alpha_j + t) \\ b_1 = (I \sin\alpha_j \sin\beta_j)/(I \cos\alpha_j + t) \end{cases} \tag{16}$$

As shown in (16), the parameter β_j is linearly related to $\boldsymbol{b}$ as $\beta_j = \arctan(b_1/b_0)$.

Let the augmented vector of $\boldsymbol{b}$ be $\bar{\boldsymbol{b}} = [b_0, b_1, 1]$, if the radius of the circle is known, then the parameter α is linearly related to $\boldsymbol{b}$. As shown in Fig. 3, let μ be the radius of the circle, which is in the normalized image, and let γ be the tangent angle between the camera and the sphere. Hence, γ is determined as $\gamma = \arctan(\mu)$ and t is found as $t = I/\sin(\gamma)$.

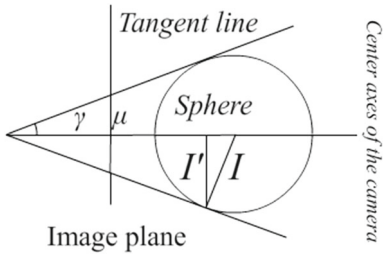


Fig. 3 The radial of the circle and the size of the sphere are related by the tangent angle

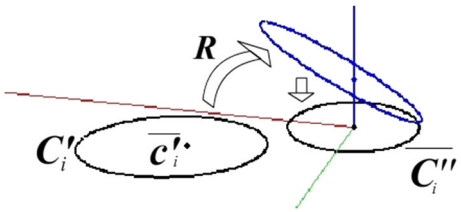


Fig. 4 Transform the image points to the center of the normalized image

Consider the distance from $\boldsymbol{b}$ to the distortion center, α_j satisfies (17).

$$b_0^2 + b_1^2 = \frac{\sin^2\alpha_j}{(\cos\alpha_j + 1/\sin\gamma)^2} \tag{17}$$

Replace the $\sin^2\alpha_j$ in (17) with $1 - \cos^2\alpha_j$, then (17) is changed to a one variable quadratic equation. Solve (17) with the quadratic formula, then the parameter α_j is found. Here, α_j is the angle belongs to $(\pi/2, \pi]$. After α_j, β_j and t are found, $\boldsymbol{Q}_j$ is determined.

If $\boldsymbol{c}_i$ is not located at the distortion center of the image, then $\boldsymbol{C}_i$ should be transformed to the distortion center before $\boldsymbol{Q}_j$ can be determined. Let $\bar{\boldsymbol{c}}_i = [\boldsymbol{c}_i 1]$ be the augmented vector of $\boldsymbol{c}_i$, then there is a rotation matrix $\boldsymbol{R}$ that could transform $\boldsymbol{c}_i$ to $[0, 0, ||\boldsymbol{c}_i||]$. In this paper, the direction of the rotation vector $\boldsymbol{r}$, which corresponds to $\boldsymbol{R}$, is determined as the cross product of $\bar{\boldsymbol{c}}_i$ and $[0, 0, 1]$. And, the size of $\boldsymbol{r}$ is determined by the angle between $\bar{\boldsymbol{c}}_i$ and $[0, 0, 1]$.

Let $\bar{\boldsymbol{q}}_j$ be the augmented vector of $\boldsymbol{q}_j$. After the rotation matrix $\boldsymbol{R}$ is found, $\bar{\boldsymbol{q}}_j$ is transformed to $\bar{\boldsymbol{q}}'_j = \boldsymbol{R}\bar{\boldsymbol{q}}_j$. Let $\bar{\boldsymbol{q}}'_j = \left[q'_{xj}, q'_{yj}, q'_{zj}\right]$ be the rotated coordinate, then the transformed coordinate in the normalized image is $\boldsymbol{q}'_j = \left[q'_{xj}/q'_{zj}, q'_{yj}/q'_{zj}\right]$. Similarly, the function of the ellipse is transformed to the distortion center. Arrange the function of ellipse as a symmetry matrix, then the transformation function is $\boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}_i\boldsymbol{R}$. The above transformation process is shown in Fig. 4.

After the image points and the ellipse are transformed to the distortion center, α_j, β_j and t are found by (16). Hence, $\boldsymbol{Q}_j$ is determined. Multiply $\boldsymbol{Q}_j$ with $\boldsymbol{R}^{\mathrm{T}}$, then the three-dimensional points are found as $\boldsymbol{R}^{\mathrm{T}}\boldsymbol{Q}_j$.

For every image point in $\left\{\{\boldsymbol{q}_j\}_i\right\}$, the above process finds its three-dimensional coordinates $\left\{\{\boldsymbol{Q}_j\}_i\right\}$. Along with the three-dimensional points, the centers of the spheres are found as $\{\boldsymbol{A}_4\}_i$.

4.2.2 Calibrate and optimize the projector

Let $\left\{\{\boldsymbol{q}_j\}_i\right\}$ and $\left\{\{\boldsymbol{p}_j\}_i\right\}$ be the pair of image points obtained from the gray code method [20], and let $\left\{\{\boldsymbol{Q}_j\}_i\right\}$ be the three-dimensional coordinates correspond to $\left\{\{\boldsymbol{q}_j\}_i\right\}$. Here, $\{\boldsymbol{q}_j\}$ are the image points inside the image of the ith sphere, while $\{\boldsymbol{p}_j\}$ are the corresponding points in the image of the projector. The proposed method finds $\left\{\{Q_{j_j}\}_i\right\}$ in previous part of this section. After $\left\{\{\boldsymbol{Q}_j\}_i\right\}$ is found, the projector is calibrated by (1).

After the projection matrix is determined, the internal matrix and the relative position are separated by the QR decomposition. Let $\boldsymbol{K}_{\text{prj}}$ be the intrinsic parameters, and let $\boldsymbol{Rt}_{\text{prj}} = \left[\boldsymbol{R}_{\text{prj}}, \boldsymbol{T}_{\text{prj}}\right]$ be the relative position of the projector. As the radial distortion of the projector is not considered in the above process, so that the trust region reflective method is employed to find the distortion of the projector.

For the updating directions, the partial derivative is found in the following way. Consider the situation that a three-dimensional point $\boldsymbol{Q}_j$ is transformed from the world coordinate to the image coordinate. Let $\boldsymbol{Q}_j = [X_1, X_2, X_3]$, then the translation from the world coordinate to the image coordinate is shown in (18).

$$\begin{cases} x = X_1/X_3 \\ y = X_2/X_3 \\ d = x^2 + y^2 \\ x' = 2k_4xy + k_5(d + 2x^2) \\ y' = 2k_4(d + 2y^2) + k_5xy \\ x'' = x(1 + k_1d + k_2d^2 + k_3d^3) + x' \\ y'' = y(1 + k_1d + k_2d^2 + k_3d^3) + y' \\ p_x = f_x \cdot x'' + c_x \\ p_y = f_y \cdot y'' + c_y \end{cases} \tag{18}$$

According to (18), the partial derivatives of the internal parameters are found by (19) and (20).

$$\begin{cases} \partial(e_x)/\partial(fp_x) = x'' \\ \partial(e_x)/\partial(fp_y) = 0 \\ \partial(e_x)/\partial(cp_x) = 1 \\ \partial(e_x)/\partial(cp_y) = 0 \\ \partial(e_x)/\partial(k_1) = fp_x(x^3 + xy^2) \\ \partial(e_x)/\partial(k_2) = fp_x \cdot x(x^2 + y^2)^2 \\ \partial(e_x)/\partial(k_3) = fp_x \cdot x(x^2 + y^2)^4 \\ \partial(e_x)/\partial(k_4) = 2fp_x \cdot xy \\ \partial(e_x)/\partial(k_5) = fp_x \cdot (3x^2 + y^2) \end{cases} \tag{19}$$

$$\begin{cases} \partial(e_y)/\partial(fp_x) = 0 \\ \partial(e_y)/\partial(fp_y) = y'' \\ \partial(e_y)/\partial(cp_x) = 0 \\ \partial(e_y)/\partial(cp_y) = 1 \\ \partial(e_y)/\partial(k_1) = fp_y(x^3 + xy^2) \\ \partial(e_y)/\partial(k_2) = fp_y \cdot y(x^2 + y^2)^2 \\ \partial(e_y)/\partial(k_3) = fp_y \cdot y(x^2 + y^2)^4 \\ \partial(e_y)/\partial(k_4) = fp_y \cdot (x^2 + 3y^2) \\ \partial(e_y)/\partial(k_5) = 2fp_y \cdot xy \end{cases} \tag{20}$$

Equation (18) is suit for the three-dimensional points that lay in the coordinate system of the projector. However, the coordinates lay in the coordinate system of the camera. Hence, the rotation and the translation must be applied to $\boldsymbol{Q}_j$ before the projection process. After applying $\boldsymbol{Rt}_{\text{prj}}$ to $\boldsymbol{Q}_j$, $\boldsymbol{Q}_j$ is transformed to $\boldsymbol{Q}'_j = \boldsymbol{R}\boldsymbol{Q}_j + \boldsymbol{T}$. After applying $\boldsymbol{Rt}_{\text{prj}}$, $\boldsymbol{Q}'_j$ is in the coordinate system of the projector. Hence, the partial derivatives of the internal parameters are found.

For the relative positions, the partial derivatives are found in the following way. According to [22], let $\boldsymbol{r} = [r_1, r_2, r_3]$ be the rotation vector, then the rotation matrix is determined by (21).

$$\begin{cases} \theta = ||\mathbf{r}|| \\ \boldsymbol{\eta} = \mathbf{r}/\theta \\ R_{11} = \cos\theta + (1 - \cos\theta)\eta_1^2 \\ R_{12} = (1 - \cos\theta)\eta_1\eta_2 - \sin\theta\eta_3 \\ R_{13} = (1 - \cos\theta)\eta_1\eta_3 - \sin\theta\eta_2 \\ R_{21} = (1 - \cos\theta)\eta_1\eta_2 + \sin\theta\eta_1 \\ R_{22} = \cos\theta + (1 - \cos\theta)\eta_2^2 \\ R_{23} = (1 - \cos\theta)\eta_2\eta_3 - \sin\theta\eta_1 \\ R_{31} = (1 - \cos\theta)\eta_1\eta_3 - \sin\theta\eta_2 \\ R_{32} = (1 - \cos\theta)\eta_2\eta_3 + \sin\theta\eta_3 \\ R_{33} = \cos\theta + (1 - \cos\theta)\eta_3^2 \end{cases} \tag{21}$$

Let $\boldsymbol{T} = [T_1, T_2, T_3]$ be the translation vector, then x and y in (18) is changed to (22).

$$\begin{cases} x = \frac{R_{11}X_1 + R_{12}X_2 + R_{13}X_3 + T_1}{R_{31}X_1 + R_{32}X_2 + R_{33}X_3 + T_3} \\ y = \frac{R_{21}X_1 + R_{22}X_2 + R_{23}X_3 + T_2}{R_{31}X_1 + R_{32}X_2 + R_{33}X_3 + T_3} \end{cases} \tag{22}$$

$$\begin{cases} \frac{\partial e_x}{\partial r_i} = f_x \cdot \frac{\partial e_x}{\partial x''}\left[\frac{\partial x''}{\partial x}\sum_j \frac{\partial x}{\partial \eta_j}\frac{\partial \eta_j}{\partial r_i} + \frac{\partial x''}{\partial y}\sum_j \frac{\partial y}{\partial \eta_j}\frac{\partial \eta_j}{\partial r_i}\right] \\ \frac{\partial e_x}{\partial T_1} = f_x \cdot \frac{\partial e_x}{\partial x''}\left[\frac{\partial x''}{\partial x}\frac{\partial x}{\partial T_1} + \frac{\partial x''}{\partial y}\frac{\partial y}{\partial T_1}\right] \\ \frac{\partial q_x}{\partial T_2} = 0 \\ \frac{\partial e_x}{\partial T_3} = f_x \cdot \frac{\partial e_x}{\partial x''}\left[\frac{\partial x''}{\partial x}\frac{\partial x}{\partial T_3} + \frac{\partial x''}{\partial y}\frac{\partial y}{\partial T_3}\right] \end{cases} \tag{23}$$

$$\begin{cases} \frac{\partial e_y}{\partial r_i} = f_y \cdot \frac{\partial e_y}{\partial y''}\left[\frac{\partial y''}{\partial x}\sum_j \frac{\partial x}{\partial \eta_j}\frac{\partial \eta_j}{\partial r_i} + \frac{\partial y''}{\partial y}\sum_j \frac{\partial y}{\partial \eta_j}\frac{\partial \eta_j}{\partial r_i}\right] \\ \frac{\partial q_y}{\partial T_1} = 0 \\ \frac{\partial e_y}{\partial T_2} = f_y \cdot \frac{\partial e_y}{\partial y''}\left[\frac{\partial y''}{\partial x}\frac{\partial x}{\partial T_1} + \frac{\partial y''}{\partial y}\frac{\partial y}{\partial T_1}\right] \\ \frac{\partial e_y}{\partial T_3} = f_y \cdot \frac{\partial e_y}{\partial y''}\left[\frac{\partial y''}{\partial x}\frac{\partial x}{\partial T_3} + \frac{\partial y''}{\partial y}\frac{\partial y}{\partial T_3}\right] \end{cases} \tag{24}$$

According to (18) and (22), the partial derivative of $\boldsymbol{r}$ and $\boldsymbol{T}$ are determined by (23) and (24). With the partial derivative and the re-projection error, the updating direction is found. Hence, the distortion coefficients of the projector, along with the optimized value of the projection matrix, are found by the trust region reflective method. In this paper, the initial guess for all the distortion coefficients is 0.

5 Experiment

The resolution of the projector in the experiment is 1024×768, and the resolution of the camera in the experiment is 2048×1552. The ideal spheres employed in the experiment are the standard white ping-pong with a diameter of 40 mm.

5.1 Execute the calibration

Experiments start from pasting the spheres on a black board, which is shown in Fig. 5, and then fasten the homemade calibration tool in front of the camera–projector stereo vision system. Then, the gain and the shutter are adjusted for the environment. Next, the camera is calibrated by the image of the spheres via the proposed method. After that, the gain and shutter are changed to suit for the projector. Next, the gray code method [20], in which the reverse pattern is employed too, is employed to scan the spheres in column and row sequences so that the correspondences between $\left\{\{\boldsymbol{q}_j\}_i\right\}$ and $\left\{\{\boldsymbol{p}_j\}_i\right\}$ are established. Finally, the projector is calibrated via the proposed method.

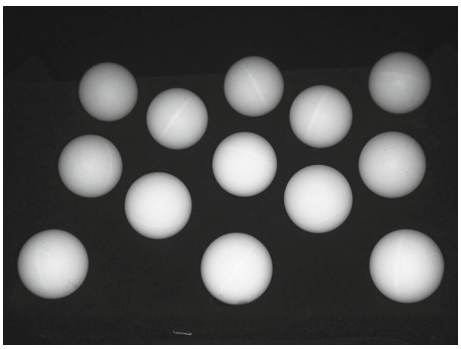


Fig. 5 The homemade calibration tool employed in the experiment

Table 3 A typical calibration result (1 pose)

$\boldsymbol{fc}$	[3803.1, 3831.0]
$\boldsymbol{cc}$	[1133.9, 767.6]
$\boldsymbol{fp}$	[1317.6, 1317.7]
$\boldsymbol{cp}$	[498.2, 782.5]
$\boldsymbol{r}$	[0.2588, − 0.2670, 0.0321]
$\boldsymbol{T}$	[217.30, 18.37, 42.02]
$\boldsymbol{skc}$	[− 0.0013, − 0.0127, − 0091, 0.0046, 0.0000]
$\boldsymbol{skp}$	[0.0152, 0.0239, 0.0192, 0.0134, − 0.0107]

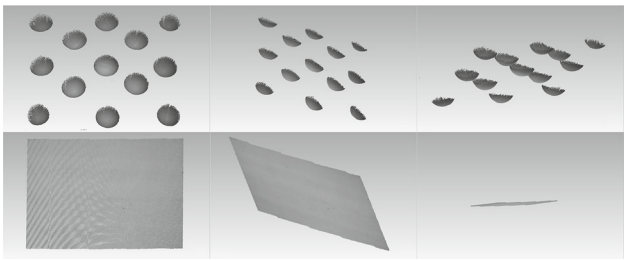


Fig. 6 The reconstructed result of the spheres and the plane. The left is the front view, while the center and the right are the side views

5.2 Result

In the experiment, the parameters for the image segmentation process, which is mentioned in Sect. 3.1, is set as $\text{th}_1 = 15$, $\beta_1 = 1.15$, $\beta_2 = 1.25$, $\alpha = 4$. A typical calibration result is shown in Table 3, where $\boldsymbol{fc}$ is the focal length of the camera, $\boldsymbol{cc}$ is the distortion center of the camera, $\boldsymbol{fp}$ is the focal length of the projector, $\boldsymbol{cp}$ is the distortion center of the projector, $\boldsymbol{r}$ is the relative rotation between the camera and the projector, $\boldsymbol{T}$ is the relative translation between the camera and the projector, $\boldsymbol{skc}$ is the distortion coefficients of the camera, and $\boldsymbol{skp}$ is the distortion coefficients of the projector. Some reconstructed result is shown in Fig. 6.

The calibration result is evaluated in two ways. Firstly, the calibration result is evaluated by comparing the size of the reconstructed objects. Secondly, the calibration result is evaluated by comparing the flatness. In this paper, the size of the reconstructed object is evaluated by comparing the average size of the fitted sphere with the ground truth.

In the experiments, the average reconstructed radius of the 13 spheres (20 mm radius in average, see Fig. 6) is

Table 4 The minimal number of poses and the average RMS

	Poses	Mean RMS (mm)
The proposed method	1	0.11
Planar-based method [10–13, 26]	3	0.08–0.11
Xu's method [18]	3	0.4

20.0083 mm with the standard deviation of 0.1054 mm. The standard deviation of the fitted plane (a 130 mm × 100 mm plane, see Fig. 6) is 0.3442 mm. We have tested that if the number of the spheres is 9–13, then the average reconstructed radius is 19.97–20.01 mm and the RMS of the sphere is 0.1040–0.1058 mm.

6 Comparison and discussion

A comparison between the proposed method and other calibration methods is shown in Table 4. Although the standard deviation of the fitted plane is associated with the hardware (the brightness of the projector and the reflectance of the plane), which is hard to compare, the average size of the fitted spheres is equal or better than other calibration method.

For the single-pose calibration problem, it is tested that the proposed method is robust and accurate. With the proposed method, one pose is enough to complete the calibration. It is noticed that if the homemade tool is obliquely placed (about 60° to the ground), then the calibration result of the projector is more accurate. This is because the range of the scanned scene has increased when the homemade tool is obliquely placed, so that the solution found by (1) is more reliable. Besides, it is also noticed that if the ring light is employed in the camera calibration process, the calibration result is more accurate. This is because the ambient light is usually directional, so that the accuracy the outline is affected in the shaded area.

Compared with the planar-based calibration method [10–13, 26], the proposed method has the advantage that it requires less mechanical operation to the calibration artifact (less calibration pose is required). In our experience, the mechanical movement is the most time consuming and energy consumption activity during the calibration. In other words, the calibration efficiency has improved.

Compared with Douet's method [19], the proposed method is robust and accurate. Douet's method employs elliptical function to model the images of the cutting sections, but the images of the three-dimensional rings are ellipses only if they are in proper perspective, otherwise the fitted ellipses are not accurate because the images just approach to an ellipse. Besides, there is another problem which prevents building up the correspondence between the ellipse and the three-dimensional ring. Let the function of a three-dimensional ring be $\boldsymbol{A}_1\cos\theta + \boldsymbol{A}_2\sin\theta + \boldsymbol{A}_4$, where $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are two orthogonal axes, while $\boldsymbol{A}_4$ is the center of the ring. And let $\boldsymbol{o}_1\cos\theta + \boldsymbol{o}_2\sin\theta + \boldsymbol{o}_4$ be the fitted ellipse, where $\boldsymbol{o}_1$ and $\boldsymbol{o}_2$ are the major axes and the mineral axes of the ellipse while $\boldsymbol{o}_4$ is the center of the ellipse. Because the projection of $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ may not be orthogonal to each other in the perspective projection transform, so that $\boldsymbol{o}_1$ and $\boldsymbol{o}_2$ are not the projection of $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$. Let $\boldsymbol{o}_1^{'}$ and $\boldsymbol{o}_2^{'}$ be the image of $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$. As mentioned in Sect. 3.1 that although $\boldsymbol{o}_1^{'}$ and $\boldsymbol{o}_2^{'}$ are not orthogonal to each other, the function $\boldsymbol{o}_1^{'}\cos\theta' + \boldsymbol{o}_2^{'}\sin\theta' + \boldsymbol{o}_4$ is still an ellipse. This means that there are infinite pairs of $\boldsymbol{o}_1^{'}$ and $\boldsymbol{o}_2^{'}$ which satisfy the fitted ellipse, but there is only one pair of values corresponding to $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$. Although the elliptical functions are equal, the correct direction of $\boldsymbol{o}_1^{'}$ and $\boldsymbol{o}_2^{'}$ must be determined, nor the incorrect mapping will cause the intrinsic matrixes un-solvable. Unluckily, there is no direct clue where $\boldsymbol{o}_1^{'}$ and $\boldsymbol{o}_2^{'}$ should be. Besides, the projection of $\boldsymbol{A}_4$ is not $\boldsymbol{o}_4$, which makes the homography transformation unreliable. Only if the camera is calibrated and the centers of the spheres are determined, then the mapping between the image and the cutting sections can be reliably established.

Compared with Xu's method [18], the proposed method requires fewer pose and is easy to deploy. This is because their method employs the homography transformation to determine the three-dimensional coordinates of the projected points, so that the calibration process requires at least three poses to calibrate the projector. If the spheres are pasted on a plane, the super glue may change the height of the ideal plane. If the ideal plane is not flat due to the super glue, the accuracy of Xu's method may be affected. The proposed method determines the three-dimensional coordinates of the projected points based on the inherent geometrical relation, so that the spheres don't have to be accurately placed on an ideal plane. This means that if the spheres are visible, then the calibration artifact is more flexible.

Compared with Resch's method [27], the proposed method calibrates the structured light system without additional assumption on the intrinsic and extrinsic parameters of the projector-camera system, while their method assumes that the initial calibration parameters are roughly known in some other way. Besides, the resolution of the projector and the accurate size of the calibration artifact will only affect the projector's calibration process in the proposed method while if the size of the calibration artifact is not accurate, both the calibration result of the camera and the calibration result of the projector will be affected in their method.

7 Conclusion and future work

In this paper, a closed-form single-pose calibration method for the structured light stereo vision system is proposed.

The proposed method improves the efficiency of the calibration process for the structured light system by reducing the mechanical operations required in the calibration process. Experiments demonstrate that the proposed method is accurate and robust.

The proposed method can be used to calibrate multiple structured light scanners to improve the calibration efficiency. This is because the entire surface of the sphere is shared by all of the scanners in different perspective. When each of the structured light scanners is calibrated, the fitted centers of the spheres are also found. Hence, the relative positions between the scanners are determined by the fitted centers. This means that when each of the scanners is calibrated, at the same time, the relative positions between the scanners are found. Hence, the calibration efficiency is improved because no additional operation is required when determining the relative positions.

References

1. Posdamer, J.L., Altschuler, M.D.: Surface measurement by space-encoded projected beam systems. Comput. Graph. Image Process. **18**(1), 1–17 (1982)
2. Inokuchi, S., Sato, K., Matsuda, F.: Range imaging system for 3-D object recognition. In: Proc. ICPR, pp. 806–808 (1984)
3. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. Proc. CVPR **1**, 195–202 (2003)
4. Abdel-Aziz, Y., Karara, H.: Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. In: Symposium on Close Range Photogrammetry, pp. 1–18 (1971)
5. Faig, W.: Calibration of close-range photogrammetry systems: mathematical formulation. Photogramm. Eng. Remote Sens. **41**(12), 1479–1486 (1975)
6. Weng, J.Y., Cohen, P., Herniou, M.: Camera calibration with distortion models and accuracy evaluation. IEEE Trans. Pattern Anal. Mach. Intell. **14**(10), 965–980 (1992)
7. Faugeras, O.: Modeling and calibrating cameras. In: Three-Dimensional Computer Vision: A Geometric Viewpoint, pp. 33–66. MIT, London (1993)
8. Hartley, R., Zisserman, A.: Camera geometry and single view geometry. In: Hartley, R., Zisserman, A. (eds.) Multiple View Geometry in Computer Vision, 2nd edn, pp. 151–176. Cambridge University, Cambridge (2003)
9. Zhang, Z.Y.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. **22**(11), 1330–1334 (2000)
10. Audet, S., Okutomi, M.: A user-friendly method to geometrically calibrate projector-camera systems. In: CVPR Workshpos, vol 1, 2, pp. 586–593 (2009)
11. Song, W., Xie, X., Li, G.L., Wang, Z.H.: Flexible method to calibrate projector–camera systems with high accuracy. Electron. Lett. **50**(23), 1685–1686 (2014)
12. Beiwen, L., Song, Z.: Structured light system calibration method with optimal fringe angle. Appl. Opt. **53**(33), 7942–7950 (2014)
13. Li, T.T., Hu, F., Geng, Z.: Geometric calibration of a camera–projector 3D imaging system. In: Proc. VRCAI, pp. 187–194 (2011)
14. Lu, Y., Payandeh, S.: On the sensitivity analysis of camera calibration from images of spheres. Comput. Vis. Image Underst. **114**(1), 8–20 (2010)
15. Ying, X.H., Hu, Z.Y.: Catadioptric camera calibration using geometric invariants. Trans. Pattern Anal. Mach. Intell. **26**(10), 1260–1271 (2004)
16. Zhang, H., Wong, K.Y.K., Zhang, G.Q.: Camera calibration from images of spheres. Trans. Pattern Anal. Mach. Intell. **29**(3), 499-U1 (2007)
17. Ying, X.H., Zha, H.B.: A novel linear approach to camera calibration from sphere images. In: Proc. ICPR, vol 1, pp. 535–538, (2006)
18. Xu, J., Douel, J., Zhao, J., Chen, K.: A simple calibration method for structured light-based 3D profile measurement. Opt. Laser Technol. **48**(2), 187–193 (2013)
19. Douet, J., Xu, J., Chen, K.: Sphere-based geometrical calibration method for structured light surface measurement systems. Adv. Mater. Res. **571**, 433–438 (2012)
20. Valkenburg, R.J., McIvor, A.M.: Accurate 3d measurement using a structured light system. Image Vis. Comput. **16**(2), 99–110 (1998)
21. Coleman, T.F., Li, Y.Y.: An interior trust region approach for nonlinear minimization subject to bounds. SIAM J. Optim. **6**(2), 418–445 (1996)
22. Faugeras, O.: Three-Dimensional Computer Vision: A Geometric Viewpoint. MIT Press, Cambridge (1993)
23. Moré, J.J.: The Levenberg-Marquardt algorithm: implementation and theory. In: Watson, G.A. (ed.) Numerical Analysis, vol. 630, pp. 105–116. Springer, Berlin (1978)
24. Haralick, R.M., Shapiro, L.G.: Computer and Robot Vision, vol. I, pp. 28–48. Addison-Wesley, Boston (1992)
25. Fitzgibbon, A.W., Pilu, M., Fisher, R.B.: Direct least squares fitting of ellipses. Trans. Pattern Anal. Mach. Intell. **21**(5), 476–480 (1999)
26. Chen, R., Xu, J., Chen, H.P., Su, J.H., Zhang, Z.H., Chen, K.: Accurate calibration method for camera and projector in fringe patterns measurement system. Appl. Opt. **55**(16), 4293–4300 (2016)
27. Resch, C., Naik, H., Keitler, P., Benkhardt, S., Klinker, G.: On-site semi-automatic calibration and registration of a projector-camera system using arbitrary object with known geometry. Trans. Vis. Comput. Graph. **21**(11), 1211–1220 (2015)

Zexi Feng received BSc and MSc degrees from the School of Computer Science, SiChuan University, in 2008 and 2011, respectively. He was an employee of Huawei Company from 2011 to 2013. He became a PhD student of ShenZhen Institutes of Advanced Technology, Chinese Academy of Sciences, in 2014. His research interests include computer graphics, computer networks and software infrastructure.

Zhiquan Cheng received BSc, MSc, and PhD degrees from Computer School at National University of Defense Technology in 2000, 2002, and 2008, respectively. He was a lecturer at the PDL Laboratory, Computer School, National University of Defense Technology (NUDT). He is now a post-doctoral of ShenZhen Institutes of Advanced Technology, Chinese Academy of Sciences. His research interests include computer graphics and digital geometry processing.

Zhan Song is a professor in the Shenzhen Institutes of Advanced Technology (SIAT) at the Chinese Academy of Sciences. His current research interests include structured light-based sensing, image processing, 3D face recognition, and human–computer interaction. Song has a PhD in mechanical and automation engineering from the Chinese University of Hong Kong.