www.ietdl.org

Published in IET Computer Vision
Received on 31st January 2014
Revised on 11th July 2014
Accepted on 1st September 2014
doi: 10.1049/iet-cvi.2014.0181

ISSN 1751-9632

Calibration of multiple fish-eye cameras using a wand

Qiang Fu[1], *Quan Quan*[1,2], *Kai-Yuan Cai*[1]

[1]*School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, People's Republic of China*
[2]*State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, People's Republic of China*
E-mail: fq@buaa.edu.cn

Abstract: Fish-eye cameras are becoming increasingly popular in computer vision, but their use for three-dimensional measurement is limited partly because of the lack of an accurate, efficient and user-friendly calibration procedure. For such a purpose, the authors propose a method to calibrate the intrinsic and extrinsic parameters (including radial distortion parameters) of two/multiple fish-eye cameras simultaneously by using a wand under general motions. Thanks to the generic camera model used, the proposed calibration method is also suitable for two/multiple conventional cameras and mixed cameras (e.g. two conventional cameras and a fish-eye camera). Simulation and real experiments demonstrate the effectiveness of the proposed method. Moreover, the authors develop the camera calibration toolbox, which is available online.

1 Introduction

Camera calibration is very important in computer vision, and numerous studies have been carried out on it. Most of these studies are based on conventional cameras, which obey the pinhole projection model and provide a limited overlap region of the field of view (FOV). The overlap region can be expanded greatly by using fish-eye cameras [1], because fish-eye cameras can provide images with a very large FOV (about 180°) without requiring external mirrors or rotating devices [2]. Fish-eye cameras have been used in many applications, such as robot navigation [3], three-dimensional (3D) measurement [4] and city modelling [5]. The drawbacks of fish-eye cameras are low resolution and significant distortion. Their use for 3D measurement is limited partly because of the lack of an accurate, efficient and user-friendly calibration procedure.

So far, many methods of calibrating conventional cameras [6, 7] have been proposed, but they are inapplicable to fish-eye camera calibration directly because the pinhole camera model no longer holds for cameras with a very large FOV. Existing methods of calibrating fish-eye cameras are roughly classified into three categories: (i) methods based on 3D calibration patterns [8, 9], (ii) methods based on 2D calibration patterns [10–12] and (iii) self-calibration methods [13, 14]. The most widely-used methods are based on 2D calibration patterns, which are often applied to a single camera. To calibrate the geometry relation among multiple cameras, it is required that all cameras observe a sufficient number of feature points simultaneously [6]. It is difficult to achieve by 3D/2D calibration patterns if two of the cameras faces each other. On the other hand, many wand-based calibration methods [6, 15, 16] were proposed for motion capture systems consisting of multiple cameras, such as the well-known Vicon system [17]. However, most of them were dedicated to dealing with conventional cameras. Calibration methods for fish-eye cameras with a 1D wand have not been discussed in the literature as far as we know.

For such a purpose, we propose a new method to calibrate the intrinsic and extrinsic parameters (including radial distortion parameters) of two/multiple fish-eye cameras simultaneously with a freely-moving wand. Thanks to the generic camera model used, the proposed calibration method is also suitable for two/multiple conventional cameras and mixed cameras (e.g. two conventional cameras and a fish-eye camera). The calibration procedure of two cameras is summarised as follows. First, the intrinsic and extrinsic parameters are initialised and optimised by using some prior information such as the real wand lengths and the nominal focal length provided by the camera manufacturer. Then, the bundle adjustment [18] is adopted to refine all unknowns, which consist of the intrinsic parameters (including radial distortion parameters), extrinsic parameters and coordinates of 3D points. With the help of vision graphs in [19], the proposed method is further extended to the case of multiple cameras, which does not require all the cameras to have a common FOV. The calibration procedure of multiple cameras is summarised as follows. First, the intrinsic and extrinsic parameters of each camera are initialised by involving pairwise calibration results. Then, the bundle adjustment is used to refine all unknowns, which consist of the intrinsic and extrinsic parameters (including radial distortion parameters) of each camera, and coordinates of 3D points.

The main contributions of this paper are: (i) a method to calibrate two/multiple cameras equipped with generic lens using a wand under general motions [As far as we know, it is the first time.]; (ii) better calibration results obtained by using the proposed method than 2D pattern-based methods

in case of two fish-eye cameras or two mixed cameras (a conventional camera and a fish-eye camera); (iii) a camera calibration toolbox freely available online [Available at http://quanquan.buaa.edu.cn/CalibrationToolbox.html.]. This paper is organised as follows. Some preliminaries are introduced in Section 2. In Section 3, the calibration algorithm for two cameras and multiple cameras is presented. Then the experimental results are reported in Section 4, followed by the conclusions in Section 5.

2 Preliminaries

2.1 Generic camera model

The perspective projection is described by the following equation [10]

$$r_1(f, \theta) = f \tan\theta \quad \text{(perspective projection)} \tag{1}$$

where θ is the angle between the optical axis and the incoming ray, the focal length f is fixed for a given camera, and $r_1(f, \theta)$ is the distance between the image point and the principal point. By contrast, fish-eye lenses are usually designed to obey one of the following projections

$$r_2(f, \theta) = f\theta \quad \text{(equidistance projection)} \tag{2}$$

$$r_3(f, \theta) = f \sin\theta \quad \text{(orthogonal projection)} \tag{3}$$

$$r_4(f, \theta) = 2f \tan(\theta/2) \quad \text{(stereographic projection)} \tag{4}$$

$$r_5(f, \theta) = 2f \sin(\theta/2) \quad \text{(equisolid angle projection)} \tag{5}$$

In practice, the real lenses do not satisfy the designed projection model exactly. A generic camera model for fish-eye lenses is proposed as follows [10]

$$r(\theta) = k_1\theta + k_2\theta^3 + k_3\theta^5 + k_4\theta^7 + k_5\theta^9 + \cdots \tag{6}$$

It is found that the first five terms can approximate different projection curves well. Therefore, in this paper we choose the model that contains only the five parameters k_1, k_2, k_3, k_4 and k_5.

As shown in Fig. 1, a 3D point $\boldsymbol{P}$ is imaged at $\boldsymbol{p}$ by a fish-eye camera, while it would be $\boldsymbol{p}'$ by a pinhole camera. Let $O_c - X_cY_cZ_c$ denotes the camera coordinate system and $o - xy$ is the image coordinate system (unit mm). We can obtain the image coordinates of $\boldsymbol{p}$ in $o - xy$ by

$$\begin{pmatrix} x \\ y \end{pmatrix} = r(\theta)\begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix} \tag{7}$$

where $r(\theta)$ is defined in (6) and φ is the angle between the radial direction and the x-axis. Then we can obtain the pixel coordinates (u, v) from

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{bmatrix} m_u & 0 \\ 0 & m_v \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} u_0 \\ v_0 \end{pmatrix} \tag{8}$$

where (u_0, v_0) is the principal point and m_u, m_v are the number of pixels per unit distance in horizontal and vertical directions, respectively. Thus, for each fish-eye camera, the intrinsic parameters are (k_1, k_2, m_u, m_v, u_0, v_0, k_3, k_4 and k_5).

Note that in this paper we do not choose the equivalent sphere model in [20]. If this generic model is used, the following calibration process will not be changed except for some intrinsic parameters. The tangential distortion is not considered here for simplicity. As pointed out in [21], the lens manufacturing technology now is of sufficiently high levels so that the tangential distortion can be ignored. Otherwise, the tangential distortion terms need to be taken into account in (6). With them, the following calibration process will not be changed except for some additional unknown parameters.

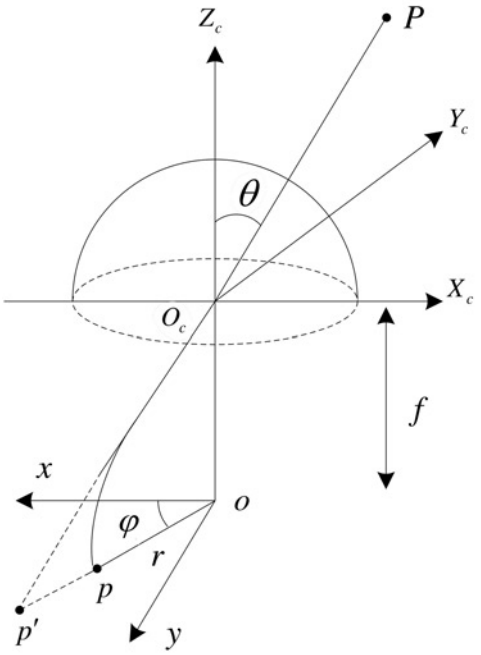


Fig. 1 *Fish-eye camera model [10]*

3D point $\boldsymbol{P}$ is imaged at $\boldsymbol{p}$ by a fish-eye camera, while it would be $\boldsymbol{p}'$ by a pinhole camera

2.2 Essential matrix

As shown in Fig. 2, the 1D wand has three collinear feature points $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ ($\boldsymbol{A}_j$, $\boldsymbol{B}_j$, $\boldsymbol{C}_j$ denote their locations for the jth image pair), which satisfy

$$\|\boldsymbol{A} - \boldsymbol{B}\| = L_1, \quad \|\boldsymbol{B} - \boldsymbol{C}\| = L_2, \quad \|\boldsymbol{A} - \boldsymbol{C}\| = L$$

where $\|\cdot\|$ denotes the Euclidean vector norm. Let $O_0 - X_0Y_0Z_0$ and $O_1 - X_1Y_1Z_1$ denote the camera coordinate systems of the left and the right cameras, respectively. The

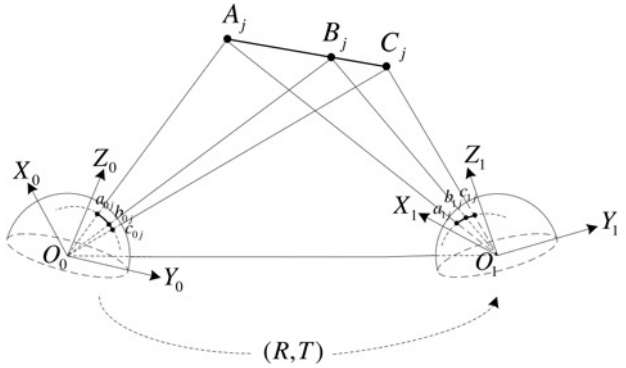


Fig. 2 *Illustration of 1D calibration wand*

3D points $\boldsymbol{A}_j$, $\boldsymbol{B}_j$, $\boldsymbol{C}_j$ denote their locations for the jth image pair

3D points A_j, B_j, C_j are projected to a_{0j}, b_{0j}, c_{0j} on the unit hemisphere centred at O_0 and a_{1j}, b_{1j}, c_{1j} on the unit hemisphere centred at O_1. The extrinsic parameters are the rotation matrix $\boldsymbol{R} \in \mathbb{R}^{3\times3}$ and translation vector $\boldsymbol{T} = (t_x, t_y, t_z)^{\mathrm{T}} \in \mathbb{R}^3$ from the left camera to the right camera.

Suppose that a 3D point $\boldsymbol{M_j} \in \mathbb{R}^3$ is projected to

$$\boldsymbol{m_{0j}} = \begin{pmatrix} \sin\theta_{0j}\cos\varphi_{0j} \\ \sin\theta_{0j}\sin\varphi_{0j} \\ \cos\theta_{0j} \end{pmatrix}, \quad \boldsymbol{m_{1j}} = \begin{pmatrix} \sin\theta_{1j}\cos\varphi_{1j} \\ \sin\theta_{1j}\sin\varphi_{1j} \\ \cos\theta_{1j} \end{pmatrix}$$

on the unit hemisphere centred at O_0 and the unit hemisphere centred at O_1, respectively. Since $\boldsymbol{m_{0j}}$, $\boldsymbol{m_{1j}}$, $\boldsymbol{T}$ are all coplanar, we have [22]

$$\boldsymbol{m_{1j}^T}[\boldsymbol{T}]_\times \boldsymbol{R}\boldsymbol{m_{0j}} = 0 \tag{9}$$

where

$$[\boldsymbol{T}]_\times = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \tag{10}$$

Furthermore, (9) is rewritten in the form as

$$\boldsymbol{m_{1j}^T}\boldsymbol{E}\boldsymbol{m_{0j}} = 0 \tag{11}$$

where $\boldsymbol{E} = [\boldsymbol{T}]_\times \boldsymbol{R}$ is known as the 'essential matrix'.

2.3 Reconstruction algorithm

In this section, a linear reconstruction algorithm for spherical cameras is proposed, which is the direct analogue of the linear triangulation method for perspective cameras [18]. Suppose that the homogeneous coordinates of a 3D point $\boldsymbol{M} \in \mathbb{R}^3$ are

$$\boldsymbol{M_0} = \begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \\ 1 \end{pmatrix}, \quad \boldsymbol{M_1} = \begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \\ 1 \end{pmatrix} = [\boldsymbol{R}, \boldsymbol{T}]\boldsymbol{M_0}$$

in $O_0 - X_0Y_0Z_0$ and $O_1 - X_1Y_1Z_1$, respectively. The 3D point M is projected to

$$\boldsymbol{m_0} = \begin{pmatrix} \sin\theta_0\cos\varphi_0 \\ \sin\theta_0\sin\varphi_0 \\ \cos\theta_0 \end{pmatrix}, \quad \boldsymbol{m_1} = \begin{pmatrix} \sin\theta_1\cos\varphi_1 \\ \sin\theta_1\sin\varphi_1 \\ \cos\theta_1 \end{pmatrix}$$

on the unit hemisphere centred at O_0 and the unit hemisphere centred at O_1, respectively. Then we have

$$\begin{cases} s_0\boldsymbol{m_0} = \boldsymbol{Q_0}\boldsymbol{M_0} \\ s_1\boldsymbol{m_1} = \boldsymbol{Q_1}\boldsymbol{M_1} \end{cases} \tag{12}$$

where s_0, s_1 are scale factors and $\boldsymbol{Q_0} = [\boldsymbol{I_3}, \boldsymbol{0_{3\times1}}] \in \mathbb{R}^{3\times4}$, $\boldsymbol{Q_1} = [\boldsymbol{R}, \boldsymbol{T}] \in \mathbb{R}^{3\times4}$. For each image point on the unit hemisphere, the scale factor can be eliminated by a cross product to give three equations, two of which are linearly independent. So the four independent equations are written in the form as follows

$$\boldsymbol{A}\boldsymbol{M_0} = \boldsymbol{0} \tag{13}$$

with

$$\boldsymbol{A} = \begin{bmatrix} \sin\theta_0\cos\varphi_0\boldsymbol{Q_{0,3}} - \cos\theta_0\boldsymbol{Q_{0,1}} \\ \sin\theta_0\sin\varphi_0\boldsymbol{Q_{0,3}} - \cos\theta_0\boldsymbol{Q_{0,2}} \\ \sin\theta_1\cos\varphi_1\boldsymbol{Q_{1,3}} - \cos\theta_1\boldsymbol{Q_{1,1}} \\ \sin\theta_1\sin\varphi_1\boldsymbol{Q_{1,3}} - \cos\theta_1\boldsymbol{Q_{1,2}} \end{bmatrix} \tag{14}$$

where $\boldsymbol{Q}_{0,i}$ and $\boldsymbol{Q_{1,i}}$ are the ith row of $\boldsymbol{Q_0}$ and $\boldsymbol{Q_1}$, respectively. Based on (13), $\boldsymbol{M_0}$ is the singular vector corresponding to the smallest singular value of $\boldsymbol{A}$. So far, given $\boldsymbol{m_0}$, $\boldsymbol{m_1}$, $\boldsymbol{R}$, $\boldsymbol{T}$, the homogeneous coordinates of $\boldsymbol{M} \in \mathbb{R}^3$ in $O_0 - X_0Y_0Z_0$, namely $\boldsymbol{M_0}$, is reconstructed. This is called 'the linear reconstruction algorithm' here.

Note that (13) provides only a linear solution, which is not very accurate in presence of noises. It could be refined by minimising reprojection errors or Sampson errors [18]. However, since the reconstruction algorithm is carried out at each optimisation iteration, it is more efficient to choose the linear reconstruction algorithm mentioned above. Furthermore, the linear reconstruction algorithm can be extended easily to the case of n-view ($n > 2$) triangulation for calibration of multiple cameras (Section 3.2) [18].

3 Calibration algorithm

3.1 Calibration of two cameras

Based on the preliminaries mentioned in Section 2, we next present a generic method to simultaneously calibrate the intrinsic and extrinsic parameters (including radial distortion parameters) of two cameras with a freely-moving 1D wand, which contains three points in known positions, as shown in Fig. 7*a*. This method is simple, user-friendly and can be used to calibrate two fish-eye cameras. Let the intrinsic parameters of the ith camera be $(k_1^i, k_2^i, m_u^i, m_v^i, u_0^i, v_0^i, k_3^i, k_4^i, k_5^i)$. Without loss of generality, we take the 0th camera and the 1st camera as an example in this subsection. The first three steps of the calibration procedure involve only 12 intrinsic parameters $(k_1^0, k_2^0, m_u^0, m_v^0, u_0^0, v_0^0, k_1^1, k_2^1, m_u^1, m_v^1, u_0^1, v_0^1)$, leaving the other parameters dealt with only in the final step.

3.1.1 Step 1: Initialisation of intrinsic parameters:

For the ith camera, the principal point (u_0^i, v_0^i) is initialised by the coordinates of the image centre, and the pixel sizes m_u^i and m_v^i are given by the camera manufacturer. If the ith camera is a conventional or fish-eye camera, then the initial values of $\boldsymbol{k^i} = (k_1^i, k_2^i)$ are obtained by fitting the model (6) to the projections (1)–(5). Concretely, let the interval $[0, \theta_{\max}^i]$ be equally divided into many pieces $(\theta_1^i, \theta_2^i, \ldots, \theta_p^i) \in \mathbb{R}^p$. Then we have

$$\begin{bmatrix} \theta_1^i & \theta_1^{i3} \\ \theta_2^i & \theta_2^{i3} \\ \vdots & \vdots \\ \theta_p^i & \theta_p^{i3} \end{bmatrix} \begin{pmatrix} k_{1,s}^i \\ k_{2,s}^i \end{pmatrix} = \begin{pmatrix} r_s(f^i, \theta_1^i) \\ r_s(f^i, \theta_2^i) \\ \vdots \\ r_s(f^i, \theta_p^i) \end{pmatrix}, \quad s = 1, 2, \ldots, 5 \tag{15}$$

where the nominal focal length of the ith camera is f^i and the maximum viewing angle is θ_{max}^i provided by the camera manufacturer. Based on (15), for the ith camera, $\boldsymbol{k^i}$ is determined by

doi: 10.1049/iet-cvi.2014.0181

$$\left(\boldsymbol{k}^{\boldsymbol{i}}, s^*\right) = \arg \min_{k^i_{1,s}, k^i_{2,s}, s \in \{1,2,\ldots,5\}} \sum_{j=1}^{p} \left(r_s(f^i, \theta^i_j) - k^i_{1,s}\theta^i_j - k^i_{2,s}\theta^{i3}_j\right)^2 \quad (16)$$

So far, we obtain the initialisation of intrinsic parameters $(k^i_1, k^i_2, m^i_u, m^i_v, u^i_0, v^i_0)$, $i = 0, 1$. Note that it is required to specify the projection type of cameras in advance in [10]. Otherwise, it is possible to obtain inaccurate calibration results. However, this is not a problem in this paper because we obtain the best initialisation of $\boldsymbol{k}^{\boldsymbol{i}}$ automatically. Besides this, the initialisation of the principle point is reasonable, because the principal point of modern digital cameras lies close to the centre of the image [18].

3.1.2 Step 2: Initialisation of extrinsic parameters:

With the intrinsic parameters $(k^i_1, k^i_2, m^i_u, m^i_v, u^i_0, v^i_0)$, $i = 0, 1$ and the pixel coordinates of image points for the jth image pair, we can compute θ_{0j}, φ_{0j}, θ_{1j} and φ_{1j} by (6)–(8). Therefore, according to (11), the essential matrix $\boldsymbol{E}_{\boldsymbol{01}}$ is obtained by using the 5-point random sample consensus (RANSAC) algorithm [23] if five or more corresponding points are given.

If the essential matrix $\boldsymbol{E}_{\boldsymbol{01}}$ is known, then the initial values for the extrinsic parameters $\boldsymbol{R}_{\boldsymbol{01}}$ and $\bar{\boldsymbol{T}}_{\boldsymbol{01}}$ are obtained by the singular value decomposition of $\boldsymbol{E}_{\boldsymbol{01}}$ [18]. Note that $\|\bar{\boldsymbol{T}}_{\boldsymbol{01}}\| = 1$, so the obtained translation vector $\bar{\boldsymbol{T}}_{\boldsymbol{01}}$ differs from the real translation vector $\boldsymbol{T}_{\boldsymbol{01}}$ by a scale factor. Let $\boldsymbol{A}^{\boldsymbol{r}}_{\boldsymbol{j}}, \boldsymbol{C}^{\boldsymbol{r}}_{\boldsymbol{j}}$ denote the reconstructed points of $\boldsymbol{A}$, $\boldsymbol{C}$ for the jth image pair, which are given by the linear reconstruction algorithm based on (13) with the intrinsic and extrinsic parameters obtained above. To minimise errors, the scale factor λ is

$$\lambda = \frac{1}{N}\sum_{j=1}^{N} \frac{L}{\|\boldsymbol{A}^{\boldsymbol{r}}_{\boldsymbol{j}} - \boldsymbol{C}^{\boldsymbol{r}}_{\boldsymbol{j}}\|} \quad (17)$$

where N is the number of image pairs. Finally, the initial value for the translation vector is

$$\boldsymbol{T}_{\boldsymbol{01}} = \left(t_x, t_y, t_z\right)^{\mathrm{T}} = \lambda\, \bar{\boldsymbol{T}}_{\boldsymbol{01}} \in \mathbb{R}^3 \quad (18)$$

Thus, we obtain the initialisation of extrinsic parameters $\boldsymbol{R}_{\boldsymbol{01}}$ and $\boldsymbol{T}_{\boldsymbol{01}}$.

3.1.3 Step 3: Non-linear optimisation of intrinsic and extrinsic parameters:

Denote the reconstructed points of $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ for the jth image pair by $\boldsymbol{A}^{\boldsymbol{r}}_{\boldsymbol{j}}, \boldsymbol{B}^{\boldsymbol{r}}_{\boldsymbol{j}}, \boldsymbol{C}^{\boldsymbol{r}}_{\boldsymbol{j}}$, respectively, which are given by the linear reconstruction algorithm based on (13) with the intrinsic and extrinsic parameters obtained above. Because of noises, there exist distance errors as follows

$$g_{1,j}(x) = L_1 - \left\|\boldsymbol{A}^{\boldsymbol{r}}_{\boldsymbol{j}} - \boldsymbol{B}^{\boldsymbol{r}}_{\boldsymbol{j}}\right\| \quad (19)$$

$$g_{2,j}(x) = L_2 - \left\|\boldsymbol{B}^{\boldsymbol{r}}_{\boldsymbol{j}} - \boldsymbol{C}^{\boldsymbol{r}}_{\boldsymbol{j}}\right\| \quad (20)$$

$$g_{3,j}(x) = L - \left\|\boldsymbol{A}^{\boldsymbol{r}}_{\boldsymbol{j}} - \boldsymbol{C}^{\boldsymbol{r}}_{\boldsymbol{j}}\right\| \quad (21)$$

where $\boldsymbol{x} = (k^0_1, k^0_2, m^0_u, m^0_v, u^0_0, v^0_0, k^1_1, k^1_2, m^1_u, m^1_v, u^1_0, v^1_0, r_1, r_2, r_3, t_x, t_y, t_z) \in \mathbb{R}^{18}$. In particular, $\boldsymbol{r}_{\boldsymbol{01}} = (r_1, r_2, r_3)^{\mathrm{T}} \in \mathbb{R}^3$ and the rotation matrix $\boldsymbol{R}_{\boldsymbol{01}}$ are related by the Rodrigues formula, namely $\boldsymbol{R}_{\boldsymbol{01}} = e^{[\boldsymbol{r}_{\boldsymbol{01}}]_\times}$ [18, p. 585]. Therefore, according to (19)–(21), the objective function for optimisation is

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} \sum_{j=1}^{N} \left(g^2_{1,j}(\boldsymbol{x}) + g^2_{2,j}(\boldsymbol{x}) + g^2_{3,j}(\boldsymbol{x})\right) \quad (22)$$

which is solved by using the Levenberg–Marquardt method [18].

3.1.4 Step 4: Bundle adjustment:

The solution above can be refined through the bundle adjustment [18], which involves both the camera parameters and 3D space points. For the jth image pair, we can compute $\boldsymbol{A}^{\boldsymbol{r}}_{\boldsymbol{j}}, \boldsymbol{B}^{\boldsymbol{r}}_{\boldsymbol{j}}, \boldsymbol{C}^{\boldsymbol{r}}_{\boldsymbol{j}}$ by the linear reconstruction algorithm based on (13) with the camera parameters $\boldsymbol{x}^*$ obtained in *Step 3*. If

$$\left|\frac{L - \left\|\boldsymbol{A}^{\boldsymbol{r}}_{\boldsymbol{j}} - \boldsymbol{C}^{\boldsymbol{r}}_{\boldsymbol{j}}\right\|}{L}\right| > 1\%$$

then the jth image pair is removed from the observations. After this, the number of image pairs reduces from N to N_1. Without loss of generality, the image pairs from $(N_1 + 1)$th to Nth are removed. Since the 3D space points $\boldsymbol{A}_{\boldsymbol{j}}$, $\boldsymbol{B}_{\boldsymbol{j}}$ and $\boldsymbol{C}_{\boldsymbol{j}}$ are collinear, they have the relation as follows

$$\begin{cases} \boldsymbol{B}_{\boldsymbol{j}} = f_B(\boldsymbol{A}_{\boldsymbol{j}}, \phi_j, \theta_j) = \boldsymbol{A}_{\boldsymbol{j}} + L_1 \cdot \boldsymbol{n}_{\boldsymbol{j}} \\ \boldsymbol{C}_{\boldsymbol{j}} = f_C(\boldsymbol{A}_{\boldsymbol{j}}, \phi_j, \theta_j) = \boldsymbol{A}_{\boldsymbol{j}} + L \cdot \boldsymbol{n}_{\boldsymbol{j}} \end{cases} \quad (23)$$

where ϕ_j, θ_j are spherical coordinates centred at $\boldsymbol{A}_{\boldsymbol{j}}$ and $\boldsymbol{n}_{\boldsymbol{j}} = (\sin\phi_j\cos\theta_j,\ \sin\phi_j\sin\theta_j,\ \cos\phi_j)^{\mathrm{T}}$ denotes the orientation of the 1D wand.

The six additional camera parameters $\left(k^0_3, k^0_4, k^0_5, k^1_3, k^1_4, k^1_5\right)$ for the two cameras are initialised to zero first, which together with $\boldsymbol{x}^*$ constitute

$$\boldsymbol{y} = \left(k^0_1, k^0_2, m^0_u, m^0_v, u^0_0, v^0_0, k^0_3, k^0_4, k^0_5, k^1_1, k^1_2, m^1_u, m^1_v, u^1_0, v^1_0, k^1_3, k^1_4, k^1_5, r_1, r_2, r_3, t_x, t_y, t_z\right) \in \mathbb{R}^{24}$$

Let functions $P_i(\boldsymbol{y}, \boldsymbol{M})$ denote the projection of a 3D point $\boldsymbol{M}$ onto the ith camera image plane under the parameter $\boldsymbol{y}$, $i = 0, 1$. Bundle adjustment minimises the following reprojection error (see (24) at the bottom of the next page)

where a_{ij}, b_{ij}, c_{ij} are the image points of 3D points $\boldsymbol{A}_{\boldsymbol{j}}$, $\boldsymbol{B}_{\boldsymbol{j}}$, $\boldsymbol{C}_{\boldsymbol{j}}$ in the ith camera, respectively. Since $\boldsymbol{A}^{\boldsymbol{r}}_{\boldsymbol{j}}, \boldsymbol{B}^{\boldsymbol{r}}_{\boldsymbol{j}}, \boldsymbol{C}^{\boldsymbol{r}}_{\boldsymbol{j}}$ are known, we could obtain ϕ^r_j, θ^r_j from (23). Then, $\boldsymbol{A}_{\boldsymbol{j}}$, ϕ_j, θ_j are initialised by $\boldsymbol{A}^{\boldsymbol{r}}_{\boldsymbol{j}}$, ϕ^r_j, θ^r_j, respectively. After all the optimisation variables are initialised, the non-linear minimisation is done by using the sparse Levenberg–Marquardt algorithm [24].

Note that the main difference here from existing work is to take the extra parameters of the radial distortion in the set of unknowns into bundle adjustment.

3.2 Calibration of multiple cameras

3.2.1 Step 1: Initialisation of intrinsic and extrinsic parameters:

The multiple camera system could be represented by a weighted undirected graph as in [19]. For example, the vision graph of a system consisting of five cameras is shown in Fig. 3. Each vertex represents an

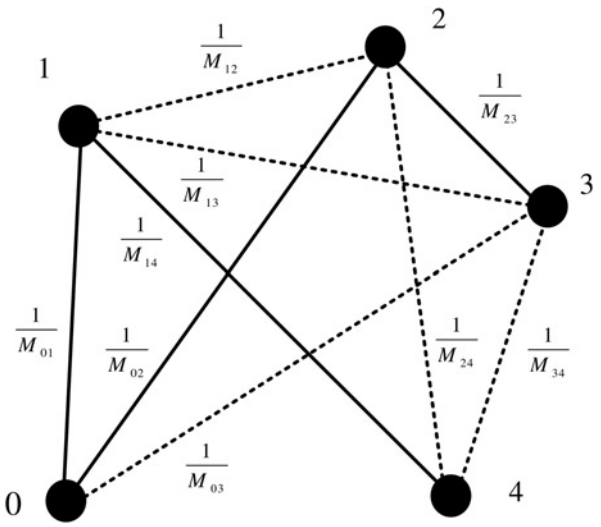


Fig. 3 *Vision graph and the optimal path from reference camera 0 to the other four cameras in solid lines*

M_{ij} is the number of common points between cameras and $\frac{1}{M_{ij}}$ is the corresponding weight
Vertices 0 and 4 are not connected because $M_{04} = 0$

individual camera and the weights w_{ij} are given as $\frac{1}{M_{ij}}$ where M_{ij} is the number of points in the common FOV of the two cameras. If $M_{ij} = 0$, then the vertices corresponding to the two cameras are not connected. Next, we use the Dijkstra's shortest path algorithm [25] to find the optimal path from a reference camera to other cameras. With the shortest paths from the reference camera to other cameras and corresponding pairwise calibration results, we could get the rotation matrices and translation vectors that represent the transformation from the reference camera to other cameras. For example, if the transformations from the ith camera to the jth camera and from the jth camera to the kth camera are ($\boldsymbol{R_{ij}}$, $\boldsymbol{T_{ij}}$) and ($\boldsymbol{R_{jk}}$, $\boldsymbol{T_{jk}}$), respectively, then the transformation from the ith camera to the kth camera is obtained as follows

$$\begin{cases} \boldsymbol{R_{ik}} = \boldsymbol{R_{jk}}\boldsymbol{R_{ij}} \\ \boldsymbol{T_{ik}} = \boldsymbol{R_{jk}}\boldsymbol{T_{ij}} + \boldsymbol{T_{jk}} \end{cases} \tag{25}$$

If the length of a path from the reference camera is longer than two, we could apply (25) sequentially to cover the entire path. Besides, the initial value of each camera's intrinsic parameter is determined from the corresponding pairwise calibration results when the most points exist in the common FOV of two cameras.

Note that only the pairwise calibration involved in the optimal path is performed by using the calibration algorithm of two cameras mentioned before. However, if all the camera pairs are calibrated as in [19], then it will be very time-consuming, especially when the number of cameras is large.

3.2.2 Step 2: Bundle adjustment: As in the calibration algorithm of two cameras, $\boldsymbol{A_j^r}$, $\boldsymbol{B_j^r}$, $\boldsymbol{C_j^r}$ are computed by n-view ($n \geq 2$) triangulation method in Section 2.3 and a distance error threshold can be set to remove outliers. The intrinsic and extrinsic parameters of $m+1$ cameras (except the extrinsic parameter of the reference camera – the 0th camera, as it is constantly $\boldsymbol{I_{3\times 3}}$ and $\boldsymbol{0_{3\times 1}}$) constitute $\boldsymbol{y} \in \mathbb{R}^{15m+9}$. Let functions $P_i(\boldsymbol{y}, \boldsymbol{M})$ ($i = 0, 1, \ldots, m$) define projection of a 3D point $\boldsymbol{M}$ onto the ith camera image plane, then bundle adjustment minimises the following reprojection error (see (26))

where a_{ij}, b_{ij}, c_{ij} are the image points of 3D points $\boldsymbol{A_j}$, $\boldsymbol{B_j}$, $\boldsymbol{C_j}$ in the ith camera and N_i is the number of times $\boldsymbol{A_j}$, $\boldsymbol{B_j}$, $\boldsymbol{C_j}$ are viewed in the ith camera. After all the optimisation variables are initialised, the non-linear minimisation is done by using the sparse Levenberg–Marquardt algorithm [24].

4 Experimental results

4.1 Simulation experiments

4.1.1 Simulation setting: In the simulation experiments, the 0th, 1st and 2nd fish-eye cameras all have image resolutions of 640 pixels × 480 pixels with pixel sizes of 5.6 μm × 5.6 μm and FOVs of 185°. As for the 1D calibration wand, the feature points $\boldsymbol{A}$ and $\boldsymbol{B}$, $\boldsymbol{C}$ satisfy

$$\begin{aligned} L_1 &= \|\boldsymbol{A} - \boldsymbol{B}\| = 400\,\text{mm} \\ L_2 &= \|\boldsymbol{B} - \boldsymbol{C}\| = 200\,\text{mm} \\ L &= \|\boldsymbol{A} - \boldsymbol{C}\| = 600\,\text{mm} \end{aligned}$$

Suppose that the 1D calibration wand undertake 300 times with general motions inside the volume of $[-0.35, 0.35]\,\text{m} \times [-0.35, 0.35]\,\text{m} \times [0.7, 1]\,\text{m}$. The rotation matrices from the 0th to the 1st and 2nd cameras are $(28.65, 28.65, 28.65)^{\text{T}}$, $(57.3, 57.3, 57.3)^{\text{T}}$ (in the form of Euler angles, unit: degree), respectively. The translation vectors from the 0th to the 1st and 2nd cameras are $(-700, 100, 200)^{\text{T}}$, $(-1200, -200, 700)^{\text{T}}$, respectively. The calibration error of rotation is measured by the absolute error in degrees between the true rotation matrix $\boldsymbol{R}_{\text{true}}$ and the estimated rotation matrix $\boldsymbol{R}$ defined as [26]

$$E_r = \max_{k=1}^{3} \left\| \arccos\langle \boldsymbol{r}_{\text{true}}^{\boldsymbol{k}}, \boldsymbol{r}^{\boldsymbol{k}} \rangle \right\| \times 180/\pi \tag{27}$$

where $\boldsymbol{r}_{\text{true}}^{\boldsymbol{k}}$ and r^k are the kth column of $\boldsymbol{R}_{\text{true}}$ and $\boldsymbol{R}$, respectively. The calibration error of translation is measured by

$$E_t = \frac{\left\| \boldsymbol{T}_{\text{true}} - \boldsymbol{T} \right\|}{\left\| \boldsymbol{T}_{\text{true}} \right\|} \tag{28}$$

where the true translation vector is $\boldsymbol{T}_{\text{true}}$ and the estimated translation vector is $\boldsymbol{T}$. If there are n 3D points viewed by a

$$\min_{\boldsymbol{y}, \boldsymbol{A_j}, \phi_j, \theta_j} \sum_{i=0}^{1} \sum_{j=1}^{N_1} \left(\|a_{ij} - P_i(\boldsymbol{y}, \boldsymbol{A_j})\|^2 + \|b_{ij} - P_i(\boldsymbol{y}, f_B(\boldsymbol{A_j}, \phi_j, \theta_j))\|^2 + \|c_{ij} - P_i(\boldsymbol{y}, f_C(\boldsymbol{A_j}, \phi_j, \theta_j))\|^2 \right) \tag{24}$$

$$\min_{\boldsymbol{y}, \boldsymbol{A_j}, \phi_j, \theta_j} \sum_{i=0}^{m} \sum_{j=1}^{N_i} \left(\|a_{ij} - P_i(\boldsymbol{y}, \boldsymbol{A_j})^2 + \|b_{ij} - P_i(\boldsymbol{y}, f_B(\boldsymbol{A_j}, \phi_j, \theta_j))^2 + \|c_{ij} - P_i(\boldsymbol{y}, f_C(\boldsymbol{A_j}, \phi_j, \theta_j))\|^2 \right) \tag{26}$$

doi: 10.1049/iet-cvi.2014.0181

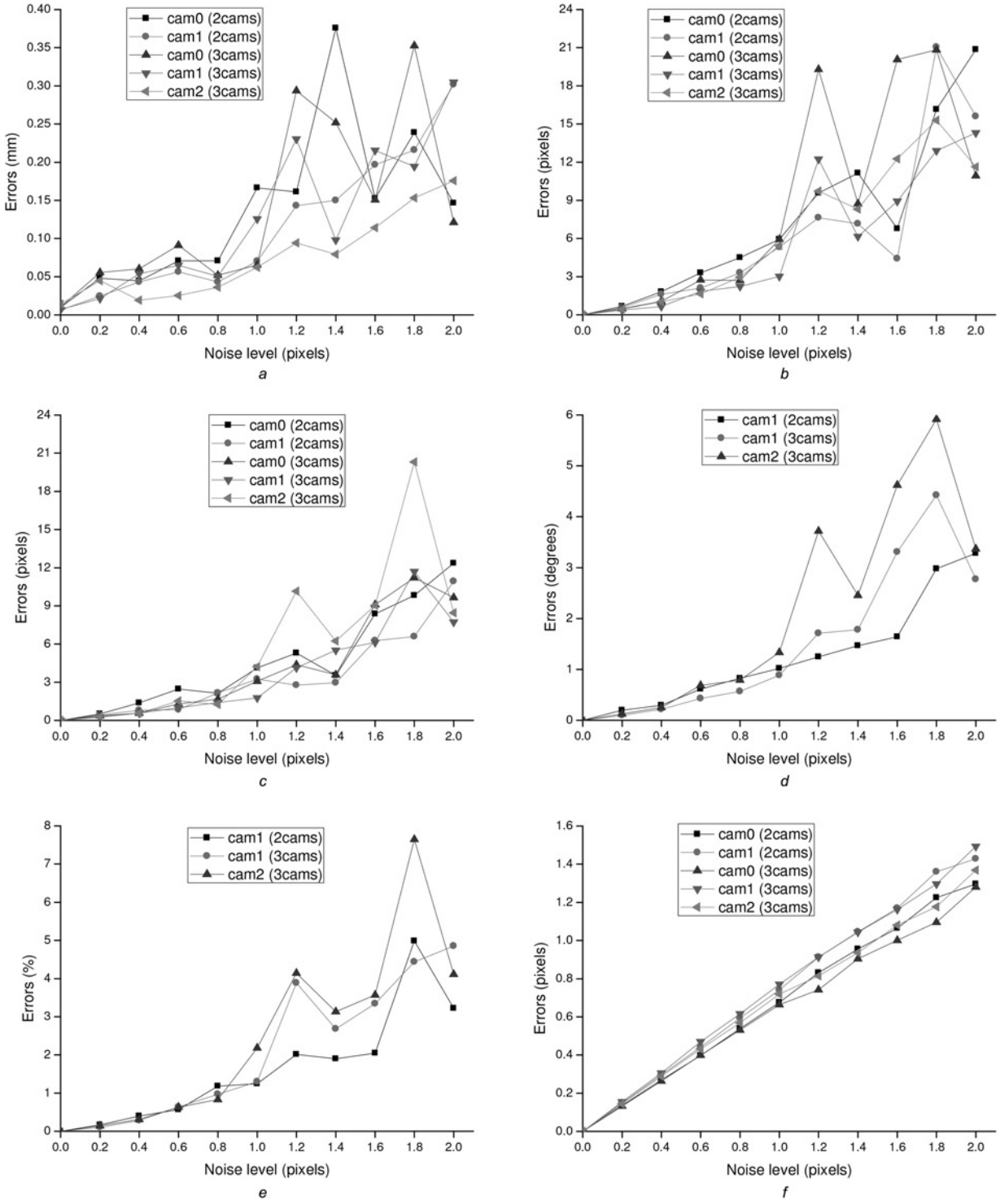


Fig. 4 *Calibration errors of the intrinsic and extrinsic parameters and reprojection errors for different noise levels*

'2cams' means calibration of the 0th and 1st fish-eye cameras and '3cams' means calibration of the 0th, 1st and 2nd fish-eye cameras
a Focal lengths of two and multiple cameras
b Principle points (u_0) of two and multiple cameras
c Principle points (v_0) of two and multiple cameras
d Calibration error of rotation E_r of two and multiple cameras
e Calibration error of translation E_t of two and multiple cameras
f RMS reprojection error E_{RMS} of two and multiple cameras

camera, the global calibration accuracy of this camera is evaluated by the root-mean-squared (RMS) reprojection error

$$E_{\mathrm{RMS}} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(\boldsymbol{u}_j - \hat{\boldsymbol{u}}_j\right)^2} \tag{29}$$

where $\boldsymbol{u}_j$ denotes the image point of the jth 3D point and $\hat{\boldsymbol{u}}_j$ is the corresponding reprojection point obtained by using calibration results. Next, we perform simulation for both two cameras (the 0th and 1st fish-eye cameras) and multiple cameras (the 0th, 1st and 2nd fish-eye cameras).

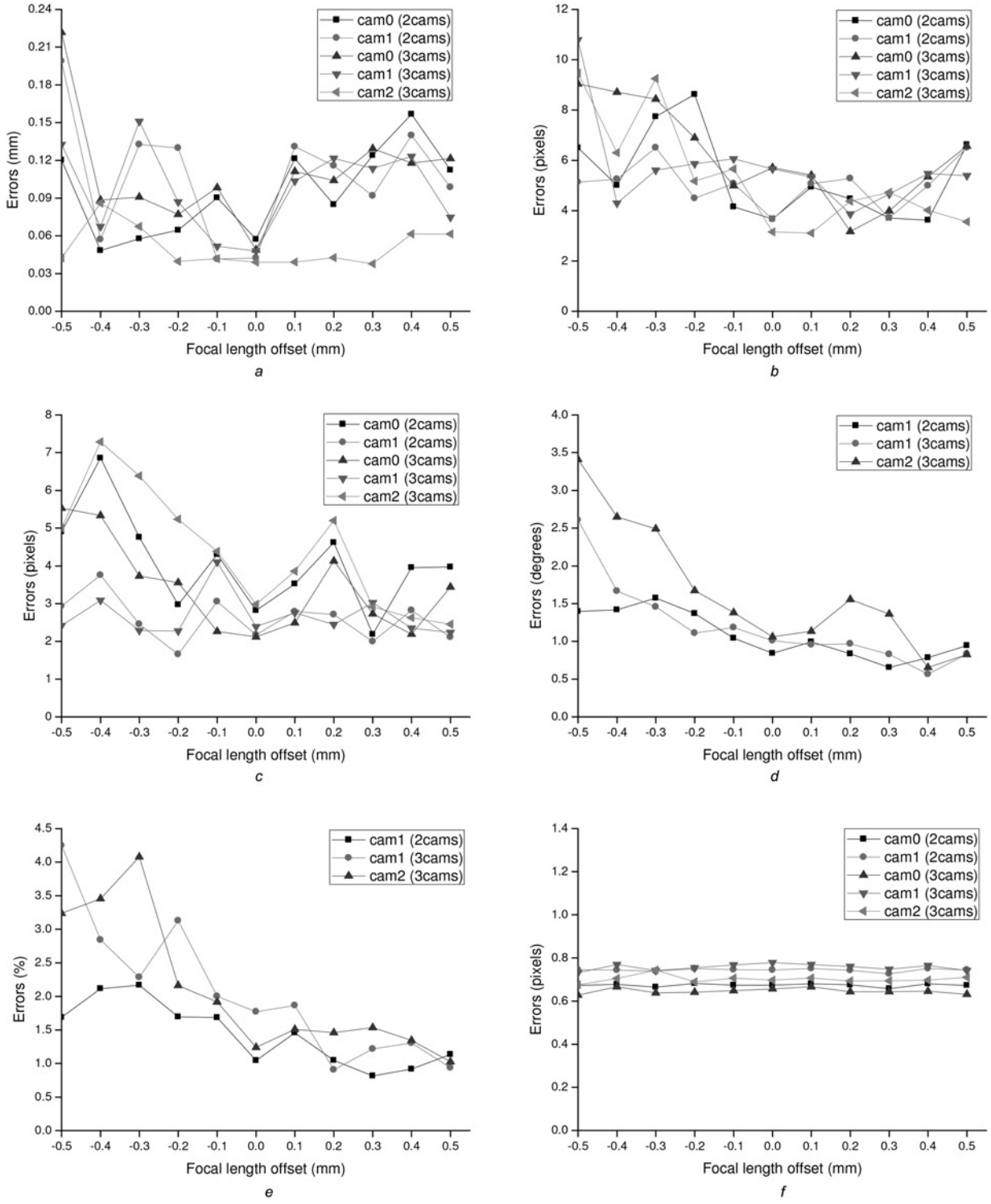


Fig. 5 *Calibration errors of the intrinsic and extrinsic parameters and reprojection errors for different focal length offsets*

'2cams' means calibration of the 0th and 1st fish-eye cameras and '3cams' means calibration of the 0th, 1st and 2nd fish-eye cameras
a Focal lengths of two and multiple cameras
b Principle points (u_0) of two and multiple cameras
c Principle points (v_0) of two and multiple cameras
d Calibration error of rotation E_r of two and multiple cameras
e Calibration error of translation E_t of two and multiple cameras
f RMS reprojection error E_{RMS} of two and multiple cameras

4.1.2 Noise simulations: The truth values of the three cameras' focal lengths and principal points are 2 mm and (310, 250), while initial values are 1.8 mm and (320, 240), respectively. Gaussian noises with the mean value $\mu = 0$ and the standard deviation σ varying from 0 to 2 pixels are added to the image points. Simulations are performed ten times for each noise level and the average of estimated parameters is taken as the result. Figs. 4*a*–*e* show the calibration errors of the intrinsic and extrinsic parameters, whereas Fig. 4*f* gives the RMS reprojection errors of the

doi: 10.1049/iet-cvi.2014.0181

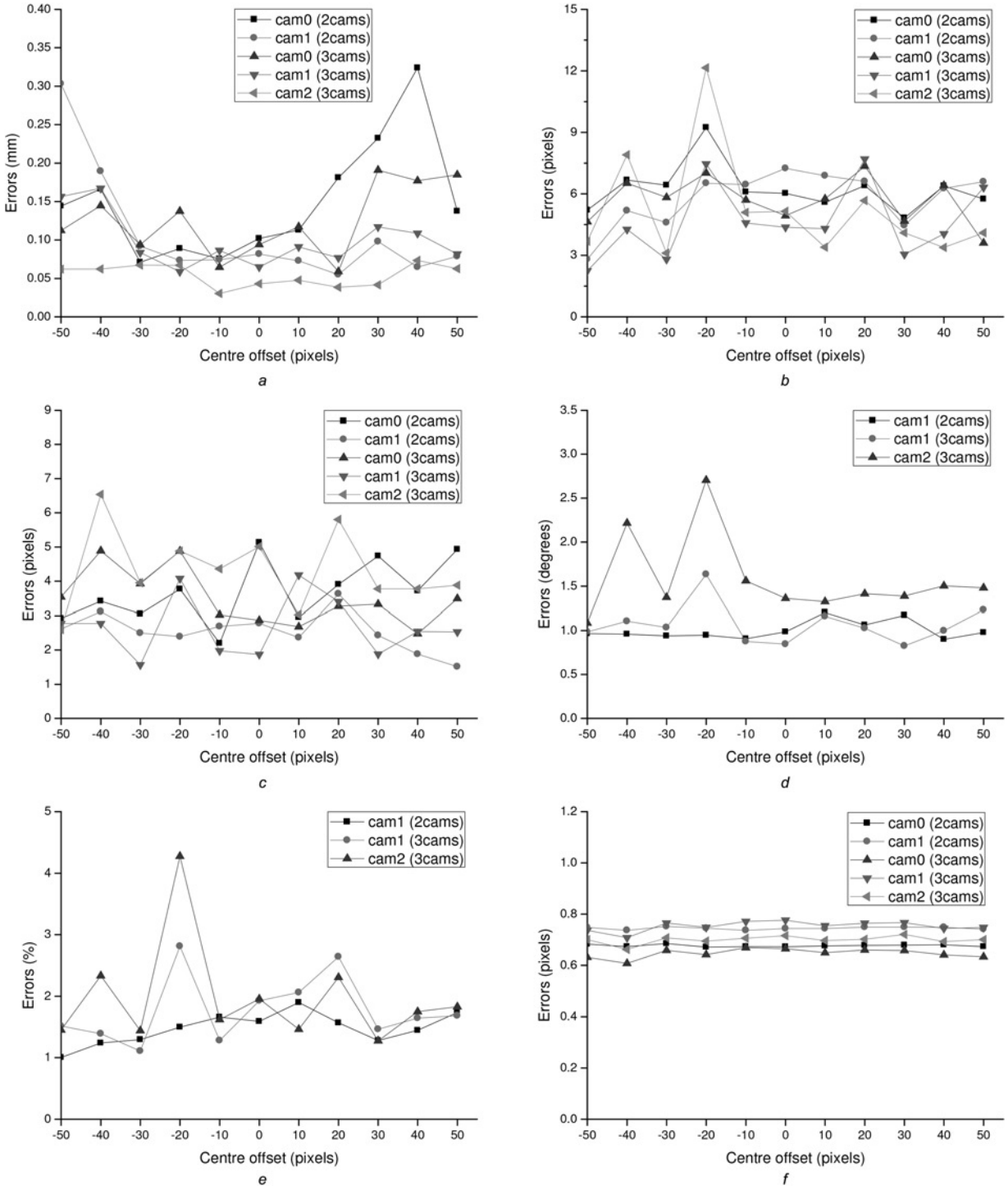


Fig. 6 *Calibration errors of the intrinsic and extrinsic parameters and reprojection errors for different centre offsets of the principle points*

'2cams' means calibration of the 0th and 1st fish-eye cameras and '3cams' means calibration of the 0th, 1st and 2nd fish-eye cameras
a Focal lengths of two and multiple cameras
b Principle points (u_0) of two and multiple cameras
c Principle points (v_0) of two and multiple cameras
d Calibration error of rotation E_r of two and multiple cameras
e Calibration error of translation E_t of two and multiple cameras
f RMS reprojection error E_{RMS} of two and multiple cameras

cameras. In Fig. 4, '2cams' means calibration of the 0th and the 1st fish-eye cameras (two cameras) and '3cams' means calibration of the 0th, 1st and 2nd fish-eye cameras (multiple cameras).

As shown in Fig. 4, the calibration errors of the intrinsic and extrinsic parameters do not change drastically with the noise level. Moreover, the RMS reprojection errors of the cameras increase almost linearly with the noise level. All these errors are small even when $\sigma = 2$ pixels. This shows that the calibration algorithm in this paper performs well and achieves high stability for the cases of both two cameras and multiple cameras.

4.1.3 Initial value simulations: First, we conduct experiments with initial focal lengths at different 'distances' from the truth values to analyse the convergence and the quality of the solution. The truth values of the three cameras' principal points are (310, 250), while initial values are (320, 240). Gaussian noise with the mean value $\mu = 0$ and the standard deviation $\sigma = 1$ pixel are added to the image points. The truth values of the three cameras' focal lengths vary from 1.5 to 2.5 mm, while the initial values are fixed to 2 mm. Simulations are performed ten times for each focal length and the average of involving parameters is taken as the result. Figs. 5*a–e* show the calibration errors of the intrinsic, whereas Fig. 5*f* gives the extrinsic parameters and the RMS reprojection errors of the cameras.

Next, we conduct experiments with initial principle points at different 'distances' from the truth values to analyse the convergence and the quality of the solution. The true values of the three cameras' focal lengths are 2 mm, while the initial values are 1.8 mm. Gaussian noise with the mean value $\mu = 0$ and the standard deviation $\sigma = 1$ pixel is added to the image points. The truth values of the three cameras' principle points vary from (270, 190) to (370, 290) along the diagonal line $u = v$, while the initial values are fixed to (320, 240). Experiments are performed ten times for each principle point and the average of estimated parameters is taken as the result. Figs. 6*a–e* show the calibration errors of the intrinsic and extrinsic parameters, whereas Fig. 6*f* gives the RMS reprojection errors of the cameras.

As shown in Figs. 5 and 6, the calibration errors of the intrinsic and extrinsic parameters change to a small extent with the focal length offset or the centre offset of the principle point. Moreover, the RMS reprojection errors of the cameras remain almost constant. In summary, the optimisation always converges to a good solution even when initial solutions largely differ from the true solution.

4.2 Real experiments

In the real experiments, we use Basler scA640–120 gm/gc cameras with the image resolution of 658 pixels × 492 pixels, equipped with conventional lenses (Pentax C60402KP) having a FOV of 86.77° or fish-eye lenses (Fujinon FE185C057HA-1) having a FOV of 185°. The nominal focal lengths of conventional lenses and fish-eye lenses are 4.2 and 1.8 mm, respectively. The 1D calibration wand is a hollow wand with three collinear LEDs on it (see Fig. 7*a*) and the distances between the LEDs therein are

$$L_1 = \|\boldsymbol{A} - \boldsymbol{B}\| = 400\,\text{mm}$$
$$L_2 = \|\boldsymbol{B} - \boldsymbol{C}\| = 200\,\text{mm}$$
$$L = \|\boldsymbol{A} - \boldsymbol{C}\| = 600\,\text{mm}$$

The wand with LEDs is easy to manufacture, and the manufacture accuracy of the 1D calibration wand used in this paper is not high. In order to clearly observe the LEDs, the outside light could be minimised by setting the exposure time of each camera to a small value. Let the wand undertake general rigid motion for many times so that image points fill the image plane as far as possible. Meanwhile, the pixel coordinates of corresponding image points are obtained by using the geometry of the three collinear LEDs.

In the following, we investigate the performance of the proposed method [The camera calibration toolbox that implements the proposed method is available at http://quanquan.buaa.edu.cn/CalibrationToolbox.html.] and compare it with the state-of-the-art checkboad-based methods proposed by Bouguet [27] and Kannala and Brandt [10]. In this paper, we use a 7 × 10 checkboard (see Fig. 7*b*) pattern and the corner points are detected automatically by using the method in [28]. Compared to conventional checkboard-based methods, the proposed method is more convenient and more efficient especially when there are many cameras to calibrate. The deficiency of the proposed method is less accurate than conventional checkboard-based methods, because the feature extraction is less accurate.

First, we perform experiments on two cameras, including two conventional cameras, two fish-eye cameras and two mixed cameras (camera 0 is a fish-eye camera and camera 1 is a conventional camera). With some prior knowledge given by the camera manufacturer and wand constraints, the intrinsic and extrinsic parameters (including radial distortion parameters) of the two cameras can be calibrated simultaneously by the proposed algorithm in Section 3.1. The calibration results are shown in Tables 1–3, from which we find that the three methods give similar calibration results.

We also perform experiments on multiple cameras, including two conventional cameras (camera 0 and camera 2) and a fish-eye camera (camera 1). The two conventional cameras have a small common FOV, so it is impractical to use checkboard-based methods to calibrate the intrinsic and extrinsic parameters of these three cameras simultaneously. However, it is easy to finish this task by using the proposed algorithm in Section 3.2. Fig. 8 shows the vision graph generated from the calibration with camera 0 chosen as the reference camera. Owing to a small overlap between camera 0 and camera 2, the optimal transformation path for this camera is 0–1-2, rather than 0–2. The calibration results of intrinsic parameters are shown in Table 4, from which it is found that the three methods give similar results again.

After calibration, we perform 3D reconstruction with the calibration results for all the camera setups above. Put the 1D calibration wand randomly at 20 different places in a measurement volume of 3 m × 3 m × 3 m. Thus, for each camera 20 images are taken, samples of which are shown in Fig. 9. The corresponding pixel coordinates of 3D points $\boldsymbol{A}$, $\boldsymbol{C}$ are extracted manually. Tables 1–4 also give the

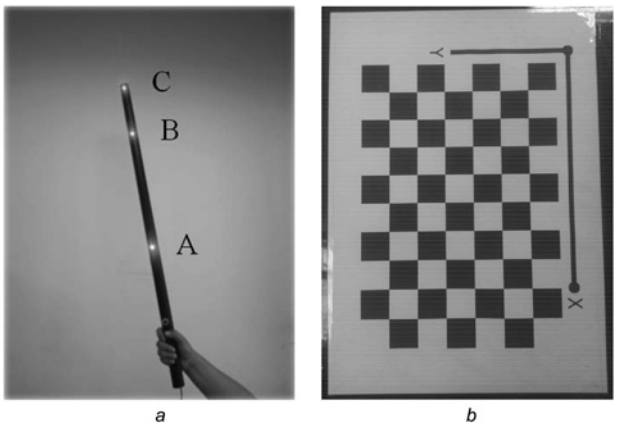


Fig. 7 *Calibration objects*

a 1D calibration wand used in the proposed method
b Checkboard used in the comparison of the proposed method

doi: 10.1049/iet-cvi.2014.0181

Table 1 Calibration results of the intrinsic and extrinsic parameters and reconstruction results of two conventional cameras

Method	Proposed		Bouguet [27]		Kannala and Brandt [10]	
camera	cam 0	cam 1	cam 0	cam 1	cam 0	cam 1
k_1	4.5932	3.3307	N/A	N/A	4.3547	4.0564
k_2	−0.6424	0.0200	N/A	N/A	−0.5023	0.2402
u_0, pixel	355.4040	376.8750	354.8109	370.3515	343.2948	361.6145
v_0, pixel	236.2023	271.7577	230.5293	268.0877	223.6586	293.3133
$\boldsymbol{R}$	$(0.0729, 0.6276, 0.2053)^{\mathrm{T}}$		$(0.0774, 0.6365, 0.1997)^{\mathrm{T}}$		N/A	
$\boldsymbol{T}$, mm	$(-1373.47, -180.03, 504.96)^{\mathrm{T}}$		$(-1371.46, -174.70, 529.24)^{\mathrm{T}}$		N/A	
E_{RMS}, pixel	0.5817	0.5421	N/A	N/A	N/A	N/A
D_{RMS}, mm	6.3160		5.9858		N/A	

Table 2 Calibration results of the intrinsic and extrinsic parameters and reconstruction results of two fish-eye cameras.

Method	Proposed		Bouguet [27]		Kannala and Brandt [10]	
Camera	cam 0	cam 1	cam 0	cam 1	cam 0	cam 1
k_1	1.8449	1.7273	N/A	N/A	1.7558	1.7083
k_2	−0.0033	0.0753	N/A	N/A	0.0706	0.1061
u_0, pixel	350.2229	352.2591	355.6940	357.9337	344.8255	356.7091
v_0, pixel	238.8122	256.1896	236.9247	257.9265	237.3625	248.5101
$\boldsymbol{R}$	$(0.0679, 0.7277, 0.2314)^{\mathrm{T}}$		$(0.0749, 0.7283, 0.2272)^{\mathrm{T}}$		N/A	
$\boldsymbol{T}$, mm	$(-1277.65, -149.39, 475.19)^{\mathrm{T}}$		$(-1276.43, -149.40, 452.62)^{\mathrm{T}}$		N/A	
E_{RMS}, pixel	0.3948	0.3575	N/A	N/A	N/A	N/A
D_{RMS}, mm	5.0890		16.1052		N/A	

Table 3 Calibration results of the intrinsic and extrinsic parameters and reconstruction results of two mixed cameras

Method	Proposed		Bouguet [27]		Kannala and Brandt [10]	
Camera	cam 0	cam 1	cam 0	cam 1	cam 0	cam 1
k_1	1.8192	4.1128	N/A	N/A	1.7448	4.0191
k_2	−0.1012	0.5818	N/A	N/A	0.0235	−0.5901
u_0, pixel	357.1181	381.2913	348.0850	362.4492	359.5363	360.5308
v_0, pixel	237.8219	258.8621	243.3855	271.0786	240.3054	248.9361
$\boldsymbol{R}$	$(0.0649, 0.7165, 0.2285)^{\mathrm{T}}$		$(0.0743, 0.7095, 0.2360)^{\mathrm{T}}$		N/A	
$\boldsymbol{T}$, mm	$(-1297.01, -149.46, 450.24)^{\mathrm{T}}$		$(-1270.74, -150.47, 485.44)^{\mathrm{T}}$		N/A	
E_{RMS}, pixel	0.3526	0.6980	N/A	N/A	N/A	N/A
D_{RMS}, mm	8.3876		12.3939		N/A	

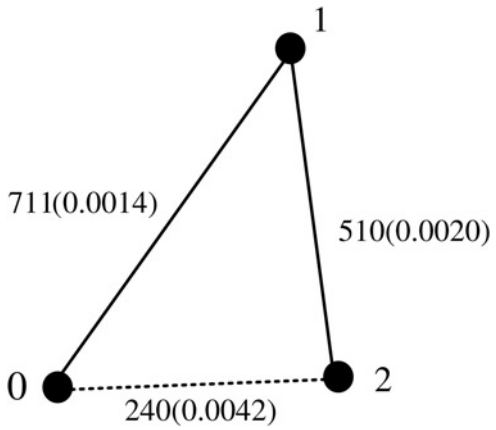


Fig. 8 *Vision graph and the optimal path from reference camera 0 to the other two cameras in solid lines*

Numbers indicate the number of common points between cameras and their corresponding weights

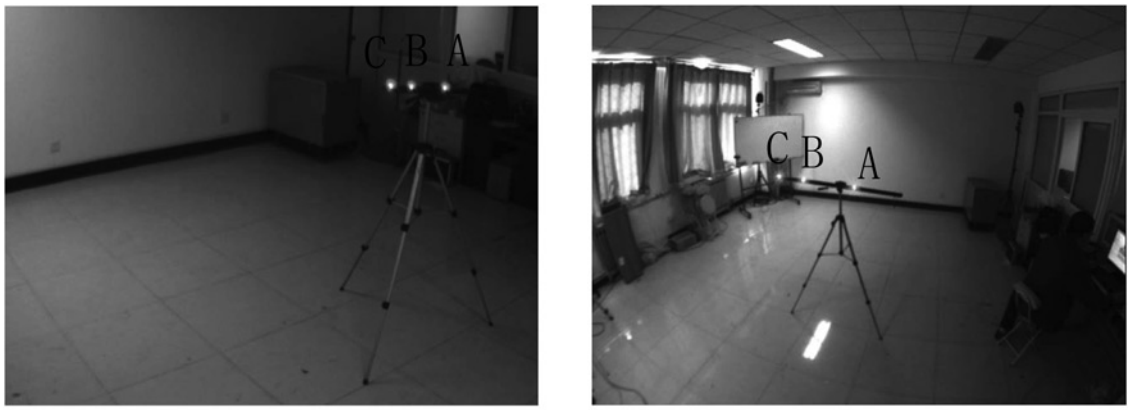
reconstruction results, where

$$D_{\mathrm{RMS}} = \sqrt{\frac{1}{20}\sum_{j=1}^{20}\left(L - \left\|\boldsymbol{A}_j^r - \boldsymbol{C}_j^r\right\|\right)^2} \tag{30}$$

with $\boldsymbol{A}_j^r$, $\boldsymbol{C}_j^r$ being the reconstructed points of $\boldsymbol{A}$, $\boldsymbol{C}$ for the jth image pair or image triple. We know from these tables that the proposed method and [27] have similar measurement accuracy in the case of two conventional cameras. However, if there are two fish-eye cameras or two mixed cameras, then our method gives better results than [27]. This is probably because: (i) the 1D calibration wand is freely placed in the scene volume, and this can increase the calibration accuracy; and (ii) our 2D pattern is a printed paper on a board, thus it is not accurate enough. From Tables 1–4, it is concluded that the measurement error of the proposed method is about 1% for all camera setups.

Table 4 Calibration results of the intrinsic parameters and reconstruction results of multiple cameras

Method	Proposed			Bouguet [27]			Kannala and Brandt [10]		
Camera	cam 0	cam 1	cam 2	cam 0	cam 1	cam 2	cam 0	cam 1	cam 2
k_1	3.2958	1.7573	4.1353	N/A	N/A	N/A	4.1519	1.6884	4.1626
k_2	−1.0775	−0.0872	−1.6900	N/A	N/A	N/A	−0.1672	−0.0177	−0.2902
u_0, pixel	338.1441	346.9810	356.9225	324.4709	354.5205	347.8188	324.8254	355.0662	347.5127
v_0, pixel	231.6928	251.2369	269.0179	228.0510	260.3225	267.9654	228.9251	259.4588	260.4136
E_{RMS}, pixel	0.5364	0.3148	0.5763	N/A	N/A	N/A	N/A	N/A	N/A
D_{RMS}, mm		3.9091			N/A			N/A	



a *b*

Fig. 9 *Sample images of the 1D object captured by the cameras for reconstruction*

a Conventional camera
b Fish-eye camera

Note that the methods in [3, 4] require parallel stereo vision to perform 3D measurement. However, this is not a requirement for the proposed method in this paper. In the experiments above, the differences between the measuring distances and the ground truth may come from several sources, such as the inaccurate extraction of image points and the manufacture errors of the 1D calibration wand. Although there are so many error sources, the calibration accuracy obtained by using the proposed method is satisfying. Also, these experiments demonstrate the practicability of the proposed calibration method.

5 Conclusions

A calibration method with a 1D object under general motions is proposed to calibrate multiple fish-eye cameras in this paper. Simulations and real experiments have demonstrated that the calibration method is accurate, efficient and user-friendly. The proposed method is generic and also suitable for two/multiple conventional cameras and mixed cameras. When there are two conventional cameras, the proposed method and 2D pattern-based methods have similar calibration accuracy. However, the proposed method gives more accurate calibration results in the case of two fish-eye cameras or two mixed cameras (a conventional camera and a fish-eye camera). The achieved level of accuracy for two fish-eye cameras is promising, especially considering their use for 3D measurement purposes. In addition, 2D/3D pattern-based methods are inapplicable when multiple cameras have little or no common FOV, whereas it is not a problem for the proposed method.

6 Acknowledgments

The authors would like to thank the anonymous reviewers for their helpful comments. This work was supported by the National Natural Science Foundation of China (61473012) and "Young Elite" of High Schools in Beijing City (YETP1071).

7 References

1 Li, S.: 'Binocular spherical stereo', *IEEE Trans. Intell. Transp. Syst.*, 2008, **9**, (4), pp. 589–600
2 Abraham, S., Förstner, W.: 'Fish-eye-stereo calibration and epipolar rectification', *ISPRS J. Photogramm. Remote Sens.*, 2005, **59**, (5), pp. 278–288
3 Shah, S., Aggarwal, J.K.: 'Mobile robot navigation and scene modeling using stereo fish-eye lens system', *Mach. Vis. Appl.*, 1997, **10**, (4), pp. 159–173
4 Yamaguchi, J.: 'Three dimensional measurement using fisheye stereo vision', in Bhatti, A. (Ed.): 'Advances in theory and applications of stereo vision' (InTech, 2011), pp. 151–164
5 Havlena, M., Pajdla, T., Cornelis, K.: 'Structure from omnidirectional stereo rig motion for city modeling'. Proc. of the Int. Conf. on Computer Vision Theory and Applications, Funchal, Portugal, 2008, pp. 407–414
6 Zhang, Z.: 'Camera calibration with one-dimensional objects', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2004, **26**, (7), pp. 892–899
7 Heikkilä, J.: 'Geometric camera calibration using circular control points', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2000, **22**, (10), pp. 1066–1077
8 Puig, L., Bastanlar, Y., Sturm, P., Guerrero, J.J., Barreto, J.: 'Calibration of central catadioptric cameras using a DLT-like approach', *Int. J. Comput. Vis.*, 2011, **93**, (1), pp. 101–114
9 Du, B., Zhu, H.: 'Estimating fisheye camera parameters using one single image of 3D pattern'. Proc. of the Int. Conf. on Electric Information and Control Engineering, Wuhan, China, 2011, pp. 367–370

doi: 10.1049/iet-cvi.2014.0181

10 Kannala, J., Brandt, S.: 'A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2006, **28**, (8), pp. 1335–1340
11 Mei, C., Rives, P.: 'Single view point omnidirectional camera calibration from planar grids'. Proc. of the IEEE Int. Conf. on Robotics and Automation, Roma, Italy, 2007, pp. 3945–3950
12 Feng, W., Röning, J., Kannala, J., Zong, X., Zhang, B.: 'A general model and calibration method for spherical stereoscopic vision'. Proc. of the SPIE 8301, Intelligent Robots and Computer Vision XXIX: Algorithms and Techniques, Burlingame, USA, 2012, p. 830107
13 Micusik, B., Pajdla, T.: 'Structure from motion with wide circular field of view cameras', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2006, **28**, (7), pp. 1135–1149
14 Espuny, F., Burgos Gil, J.: 'Generic self-calibration of central cameras from two rotational flows', *Int. J. Comput. Vis.*, 2011, **91**, (2), pp. 131–145
15 De França, J.A., Stemmer, M.R., França, M.B.d.M., Piai, J.C.: 'A new robust algorithmic for multi-camera calibration with a 1D object under general motions without prior knowledge of any camera intrinsic parameter', *Pattern Recogn.*, 2012, **45**, (10), pp. 3636–3647
16 Pribanic, T., Sturm, P., Peharec, S.: 'Wand-based calibration of 3D kinematic system', *IET Comput. Vis.*, 2009, **3**, (3), pp. 124–129
17 'Vicon Real-time motion capture system', http://www.vicon.com/System/Calibration, accessed June 2014
18 Hartley, R., Zisserman, A.: 'Multiple view geometry in computer vision' (Cambridge University Press, 2004, 2nd edn.)
19 Kurillo, G., Li, Z., Bajcsy, R.: 'Wide-area external multi-camera calibration using vision graphs and virtual calibration object'. Proc. of Second ACM/IEEE Int. Conf. on Distributed Smart Cameras, Stanford, USA, 2008, pp. 1–9
20 Geyer, C., Daniilidis, K.: 'A unifying theory for central panoramic systems and practical implications'. Proc. of the European Conf. on Computer Vision, Dublin, Ireland, 2000, pp. 445–461
21 Kanatani, K.: 'Calibration of ultrawide fisheye lens cameras by eigenvalue minimization', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2013, **35**, (4), pp. 813–822
22 Svoboda, T., Pajdla, T., Hlaváč, V.: 'Motion estimation using central panoramic cameras'. Proc. of the IEEE Conf. on Intelligent Vehicles, Stuttgart, Germany, 1998, pp. 335–340
23 Nistér, D.: 'An efficient solution to the five-point relative pose problem', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2004, **26**, (6), pp. 756–770
24 Lourakis, M.: 'Sparse non-linear least squares optimization for geometric vision'. Proc. of the European Conf. on Computer Vision, Crete, Greece, 2010, pp. 43–56
25 Chen, J.: 'Dijkstra's shortest path algorithm', *Formalized Math.*, 2003, **11**, (3), pp. 237–247
26 Zheng, Y., Kuang, Y., Sugimoto, S., Åström, K., Okutomi, M.: 'Revisiting the pnp problem: a fast, general and optimal solution'. Proc. of the IEEE Int. Conf. on Computer Vision, Sydney, Australia, 2013, pp. 2344–2351
27 Bouguet, J-Y.: 'Camera calibration toolbox for matlab', http://www.vision.caltech.edu/bouguetj/calib_doc/, accessed June 2014
28 Geiger, A., Moosmann, F., Car, Ö., Schuster, B.: 'Automatic camera and range sensor calibration using a single shot'. Proc. of the Int. Conf. on Robotics and Automation, Saint Paul, USA, 2012, pp. 3936–3943