RESEARCH **Open Access**

CrossMark

Calibration and rectification of vertically aligned binocular omnistereo vision systems

Chengtao Cai, Xiangyu Weng[*], Bing Fan and Qidan Zhu

Abstract

Omnidirectional stereo vision systems have been widely used as primary vision sensors in intelligent robot 3D measurement tasks, which require stereo calibration and rectification. Current stereo calibration and rectification methods suffer from complex calculations or a lack of accuracy. This paper establishes a simple and effective equivalency between an omnidirectional stereo vision system and a perspective vision system by studying stereo calibration and rectification methods. First, we improved the stereo calibration method. By applying the essential matrix, the complicated calibration process of the original method is simplified. By using a manual extraction method to extract corner points, noise error is eliminated and high precision is ensured. Second, we propose a new rectification method. By using the proposed simple rectification model and calibration data, the baseline length and an accurate column-aligned image pair are easily obtained, which reduces the computation time. The proposed stereo calibration and rectification method can simply and effectively obtain two key parameters of the triangulation formula for 3D measurement tasks: baseline length and parallax. Using real data captured by equipment, we performed experiments covering all the necessary stages to obtain a high-performance omnidirectional stereo vision system. Statistical analyses of the experimental results demonstrate the effectiveness of the proposed method.

Keywords: Robot, Stereo calibration, Rectification, Stereo vision system

1 Introduction

Omnidirectional stereo (omnistereo) vision systems composed of omnidirectional cameras offer the possibility of providing 3D measurement information for a 360° field of view. Several interesting configurations of omnistereo systems, such as binocular omnistereo [1], N-ocular omnistereo [2], circular projection omnistereo [3], and dynamic omnistereo [4], have been designed to achieve different mission requirements. Vertically aligned binocular (V-binocular) omnistereo vision systems, composed of two vertical coaxial catadioptric omnidirectional cameras, provide certain advantages over other types of omnistereo vision systems. (a) These systems possess a simple epipolar geometry correspondence. (b) The depth accuracy of the V-binocular omnistereo vision system is isotropic, and there are no occlusions of the image pair due to the coaxial installation. Due to the above advantages, V-binocular omnistereo vision systems have been widely used in many intelligent robot tasks [5–9]. In our research, to obtain a high-performance V-binocular omnistereo vision system, we focused on stereo calibration and rectification.

For a stereo system, calibration is the process of calibrating the camera intrinsic parameters and the camera-camera extrinsic relationship. There are two categories of current stereo calibration methods for omnistereo vision systems. One category is the calibration of the relative parameters between the camera and calibration boards. These parameters are then transformed into the camera-camera relationship [10]. Such methods provide high precision but typically require multiple calibration stereo pairs, and the Levenberg-Marquardt [11] iterative algorithm is required to reduce errors. Thus, significantly more work is required to configure the control points and measurement process. The other category consists of methods that calibrate the absolute parameters in the world coordinates [12] based on epipolar geometry [13]. The method uses only self-point correspondences in one

* Correspondence: wengxiangyu@hrbeu.edu.cn
College of Automation, Harbin Engineering University, Harbin 150001, China

Cai *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:46

Page 2 of 14

image pair without requiring prior knowledge about the scene. However, accuracy suffers, making self-point correspondences in one image pair unsuitable for 3D information measurement tasks [14].

Stereo rectification aligns the corresponding points on the same column [15]. Current omnistereo rectification models also suffer from various defects. Some are limited to articular mirrors and produce heavily distorted images. Other models are not scan-line methods, and thereby lose the important advantage of simplified stereo matching. Y. Wang et al. proposed an omnistereo rectification method [14], which is a scan-line method and avoids heavy distortion; however, the rectification model is complicated and not suitable for real-time 3D-information measurement tasks.

To overcome these difficulties and obtain a high-performance omnistereo system, first, we improved the stereo calibration method [10] based on an epipolar geometry [12], which requires only a few matching points manually extracted from one image pair to reduce the complexity of the calibration and ensure accuracy. Second, we propose a simple rectification method. The calculation of baseline length and accurate column-aligned image pairs are easily achieved by using the proposed simple rectification model and calibration data. The proposed rectification method can reduce effort while ensuring a real-time calculation. After the proposed procedure, two key parameters of the triangulation formula for 3D measurement tasks, baseline length and parallax, are easily obtained.

Using real data captured by our system, we performed experiments with the proposed stereo calibration and rectification method and compared the data with those from some existing methods. We also performed other necessary experiments to verify the high performance of the proposed method, including stereo matching, 3D reconstruction and depth estimation. Statistical analyses of the experimental results demonstrate the effectiveness of the system.

2 Stereo calibration method

2.1 Single-viewpoint system calibration

Nayar and Baker use a mathematical formula to prove that a single-viewpoint catadioptric omnidirectional system mirror section must be a quadratic curve [16]. Geyer and Daniilidis [17] also demonstrate that a central catadioptric system is coincident with the unified sphere-imaging model. The unified sphere-imaging model projection process, as shown in Fig. 1, isolates the non-linear transformation from the projection, substantially simplifying the subsequent analysis and calculation.

Due to the lens distortion of the perspective camera, the resulting omnidirectional camera calibration errors must be considered. For the calibration of internal

Fig. 1 Unified sphere-imaging model for a single-view-point, catadioptric, omnidirectional system
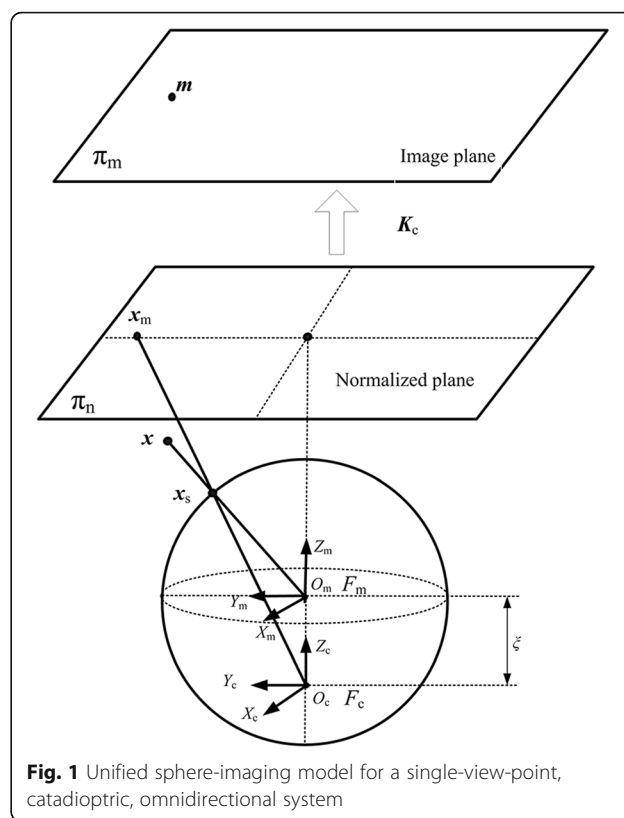
parameters, we referred to the single-view, omnidirectional camera calibration algorithm proposed by Mei and Rives [10] in which camera lens distortion is introduced into the projection image process and the installation error of the structure is compensated. In addition, this method provides the best results with respect to catadioptric omnidirectional systems among existing methods [18]. The first step is the single-camera calibration using the OpenSource toolbox [10] based on the unified sphere-imaging model [19]. Figure 2 shows the transformation of the omnidirectional camera calibration model's coordinate system. The omnidirectional camera coordinate system is the same as the mirror coordinate system, and the origin of the coordinate system is the internal focus of the mirror.

The calibration model projection process includes unknown parameters, as described below.

1. Extrinsic parameters: The relationship between the plane calibration plate coordinate system and the panoramic-camera coordinate system can be expressed by the formula $x = PX$, where $P = [R_w, T_w]$ is a 3×4 matrix, as shown in Fig. 2. The rotation matrix R_w is a quaternion notation, wherein W represents the projection process, and $V_1 = [q_o, q_1, q_2, q_3, t_1, t_2, t_3]$ represents the unknown variables in the projection matrix P.
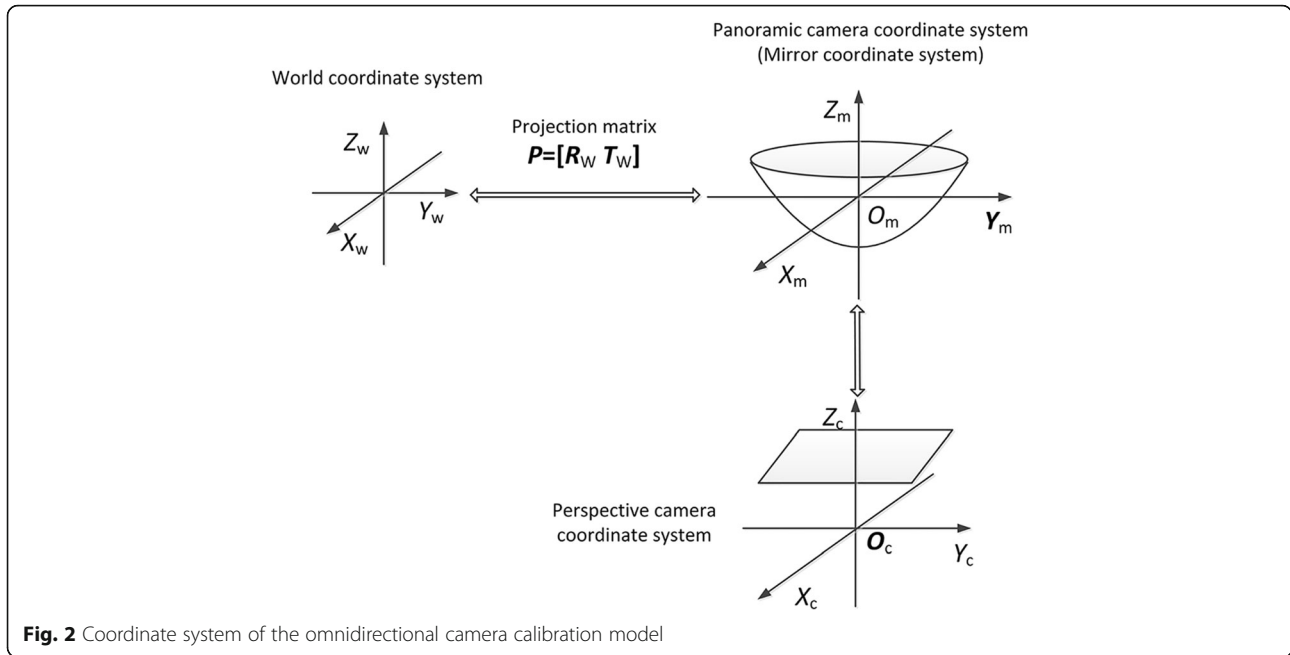
Fig. 2 Coordinate system of the omnidirectional camera calibration model

2. Nonlinear projection transformation: Assuming that the projection coordinates under the mirror coordinate system are known, the projection point coordinates on the metric plane can be calculated. H represents the nonlinear projection equation, and $V_2 = [\xi]$ represents the unknown variable of the mirror.

3. Distortion: The model introduces two primary distortions: radial distortion, which is caused by changes in radial curvature, and eccentric distortion, which is caused by the incomplete co-linearity of the axes of the optical lens. These distortions occur in optical systems as a result of assembly errors.

   Distortions can be expressed by five parameters, in which three are radial distortion δ_r factors:

   $$\delta_r = 1 + k_1\rho^2 + k_2\rho^4 + k_5\rho^6 \qquad (1)$$

   where $\rho = \sqrt{x^2 + y^2}$. The remaining two parameters are eccentric distortion δ_d factors:

   $$\delta_d = \begin{bmatrix} 2k_3xy + k_4(\rho^2 + 2x^2) \\ k_3(\rho^2 + 2y^2) + 2k_4xy \end{bmatrix} \qquad (2)$$

   D represents the distortion equation, and $V_3 = [k_1, k_2, k_3, k_4, k_5,]$ represents the distortion variable.

4. Perspective camera model: the projection process from the normalized plane to the image plane can be expressed using the generalized camera projection matrix K_c:

$$K_c = \begin{bmatrix} \gamma_1 & \alpha & u_0 \\ 0 & \gamma_2 & v_0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (3)$$

where $\gamma_i = f_i \cdot \eta$, f is the camera focal length, η is the mirror parameter, and $V_4 = [\alpha, \gamma_1, \gamma_2, u_0, v_0]$ represents the unknown variable.

5. Final projection equation: G represents the all-projection equation, and V includes the 18 unknowns:

$$G = K_c \times D \times H \times W, = V = [V_1, V_2, V_3, V_4] \qquad (4)$$

Assuming that there are n images in the calibration board, each image has m corner points, and all the unknown parameters' maximum likelihood solutions can be obtained by calculating the minimum value of the formula

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\left\|G(V_{1i}, V_2, V_3, V_4, X_{ij}) - m_{ij}\right\|^2 \qquad (5)$$

where $G(V_{1i}, V_2, V_3, V_4, X_{ij})$ is the projection of the calibration board's corner points, and X_{ij} and m_{ij} contain the corresponding image coordinates. Equation 5 is a nonlinear optimization equation that can be solved using the Levenberg-Marquardt optimization algorithm. The initial-value selection problem has been well analyzed in Mei and Rives [19] and will not be discussed in this study.

2.2 A modification theory

After omnidirectional camera calibration, all variables are known except for the extrinsic parameters. The toolbox proposed by Mei and Rives does not include the function to calculate extrinsic parameters between the board and camera. Thus, we developed this function independently from the toolbox. The theory is as follows: P' is the projection matrix from the plane calibration board's coordinate system to the omnidirectional camera coordinate system, and $V_1' = [q_0, q_1, q_2, q_3, t_1, t_2, t_3]$ contains the unknown variables. Assuming that the calibration block has m corner points, the homogeneous coordinates are X_i', and the corresponding image coordinates of X_i' are m_i'. Then, the maximum likelihood solutions of the external parameters can be obtained by calculating the minimum value of the formula

$$\sum_{i=1}^{m} \left\| G\left(V_1', V_2, V_3, V_4, X_i'\right) - m_i' \right\|^2 \tag{6}$$

where V_2, V_3 and V_4 are internal omnidirectional camera parameters that are calculated after calibration and are used as known values to obtain V_1' based on the L-M nonlinear optimization algorithm. The calibration block corners' coordinates in the omnidirectional coordinate system are calculated by

$$x = P'X_i' \tag{7}$$

2.3 Cylindrical expansion model

A cylindrically expanded, panoramic image is based on the unified sphere-imaging model. By cutting the cylindrical surface radially and tiling it, a 2D, rectangular, cylindrical, panoramic image can be obtained. This procedure eliminates the scene distortion in the restored image. Experiments were conducted with stereo vision systems and image processing software that was independently developed by our laboratory, as shown in Fig. 3.

As shown in Fig. 3, the omnidirectional system's effective viewpoint is considered to be the origin of the mirror coordinate system $O_m X_m Y_m Z_m$, and the virtual imaging plane is considered to be a cylindrical surface with a coaxial omnidirectional system whose radius is f. Assuming that the cylindrically expanded image resolution is $W \times H$ and that the pitch angles of the cylinder image's upper and lower edges are α_1 and α_2, the cylindrical image height is $H = f \tan \alpha_1 + f \tan \alpha_2$, as shown in Fig. 4. $m' = [i, j]^T$ is a point on the cylindrically expanded image, and the 3D coordinate x of m' in the mirror coordinate system can be expressed as follows:

$$\boldsymbol{x} = [f \cos\theta, \ f \sin\theta, \ f \tan\alpha_1 - j]^T \tag{8}$$
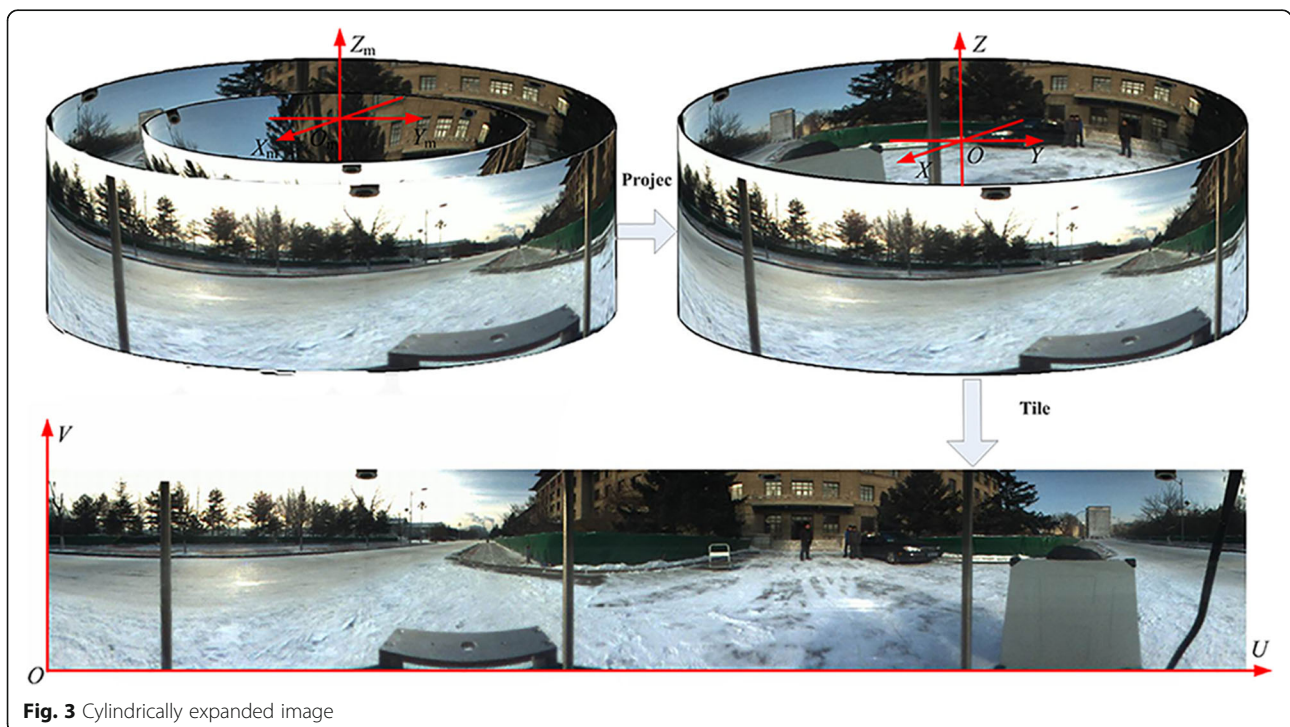
where $\theta = 2\pi i / W$.

Fig. 3 Cylindrically expanded image

Cai *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:46
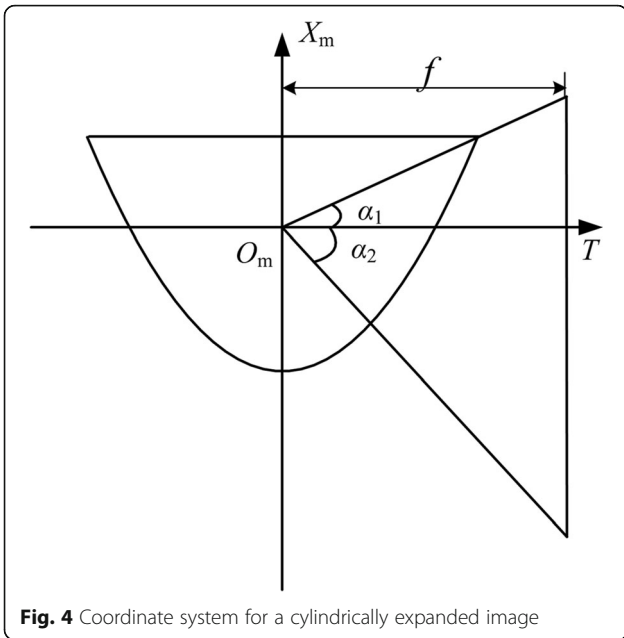
Page 5 of 14

Fig. 4 Coordinate system for a cylindrically expanded image

2.4 Epipolar geometry and essential matrix

The epipolar geometry describes a geometric relationship between the positions of the corresponding points in the two images acquired by central cameras [13]. Because epipolar geometry is a property of central projection perspective cameras, it also exists for central catadioptric cameras and is represented by a matrix as the essential matrix. Figure 5 shows the epipolar geometry

model of a general stereo vision system based on the unified sphere-imaging model. This model is a general model and also applies to the vertically arranged stereo omnidirectional system. Without loss of generality, we used this model to introduce the epipolar geometry. The epipolar constraint of the stereo image pair simplifies the search procedure from a 2D search into a 1D search.

F_1 and F_2 represent two single-viewpoint vision-system coordinate systems whose origins are O_{m1} and O_{m2}. The two corresponding image planes are π_{p1} and π_{p2}, and $x = [x, y, z]^T$ is a point in the 3D space. Without loss of generality, it is assumed that the world coordinate system coincides with the coordinate system F_1; l_1 and l_2 are a polar pair correspondence; the image points $\boldsymbol{m}_1 = [u_1, v_1]$ and $\boldsymbol{m}_2 = [u_2, v_2]$ are two image points of x on the image plane, where $\boldsymbol{m}_1 = \in l_1$, $\boldsymbol{m}_2 = \in l_2$; and polar points are denoted as $e_{ij}(i = 1, 2, j = 1, 2)$. Projecting x onto the two-sphere surface, we obtain $\boldsymbol{x}_{s1} = [x_{s1}, y_{s1}, z_{s1}]^T$ and $\boldsymbol{x}_{s2} = [x_{s2}, y_{s2}, z_{s2}]^T$. The points O_{m1}, O_{m2}, and x constitute the polar plane, and the line connecting O_{m1} and O_{m2} is the baseline. $\boldsymbol{R}$ and $\boldsymbol{T}$ are the rotation matrix and the translation vector between the two omnidirectional coordinate systems, and vector $\boldsymbol{T}$ also represents the coordinate of O_{m2} in the coordinate system F_1. Because $\boldsymbol{T}$ and $\boldsymbol{x}_{s1}$ are on the polar plane, the normal equation of the polar plane can be expressed as $N_{F1} = \boldsymbol{T} \times \boldsymbol{x}_{s1}$, where x represents the outer product. The normal equation of the polar plane under the F_2 coordinate system can be expressed as $N_{F2} = \boldsymbol{R}N_{F1} = \boldsymbol{R}(\boldsymbol{T} \times \boldsymbol{x}_{s1})$. Because the point x_{s2} is also on
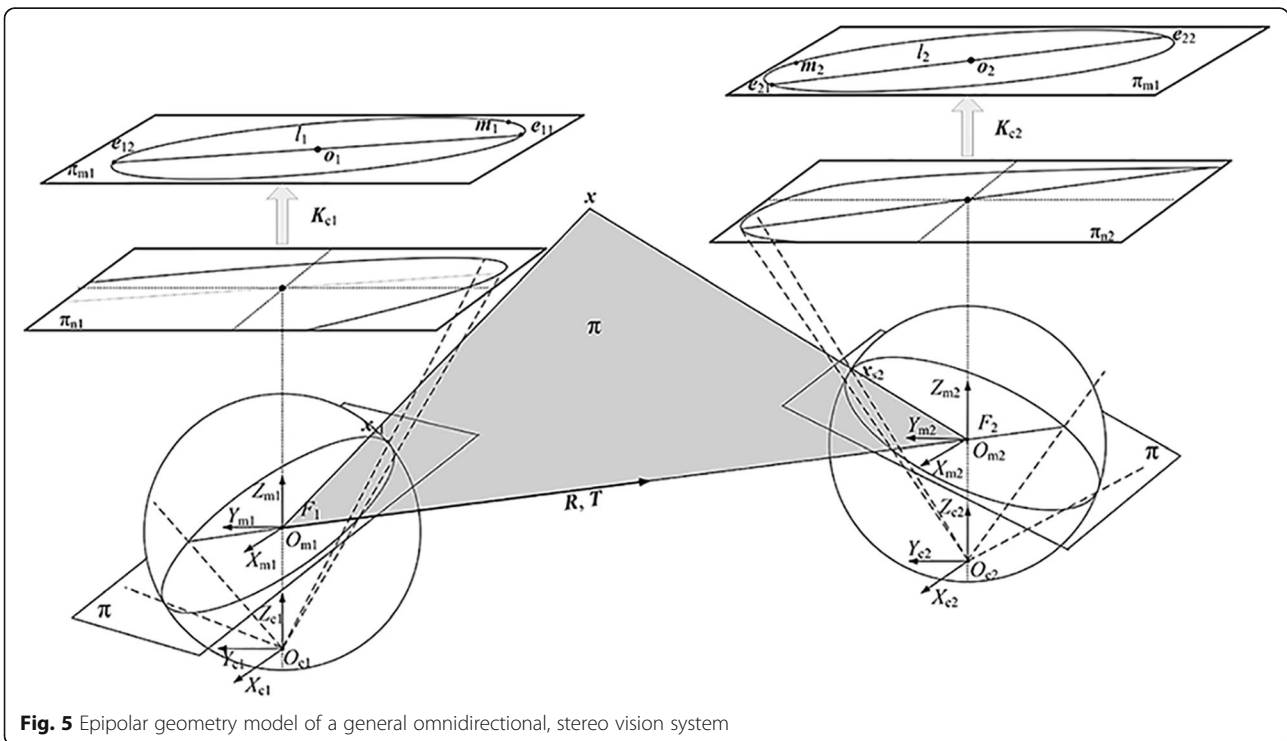
Fig. 5 Epipolar geometry model of a general omnidirectional, stereo vision system

Cai *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:46

Page 6 of 14

the intersection line of the polar plane and the unit sphere, $x_{s2}^T \cdot N_{F2} = 0$ and

$$x_{s2}^T \cdot R(T \times x_{s1}) = 0 \qquad (9)$$

Assuming that $T = [t_x, \ t_y, \ t_z]^T$, we obtain

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

where $[T]_\times$ is the anti-symmetric matrix of the vector T. Therefore, Eq. 9 can be expressed as follows:

$$x_{s2}^T \cdot R(T \times x_{s1}) = x_{s2}^T E x_{s1} = 0 \qquad (10)$$

where $E = R[T]_\times$ is an essential 3×3 matrix with a rank of 2. Notably, the essential matrix has only 5° of freedom. Based on Eq. 10, each pair of points can provide two linear constraint equations of the essential matrix; thus, calculating the essential matrix requires a minimum of eight pairs of points. Because of the homogeneity of Eq. 10, the essential matrix E can be obtained only with the difference of a non-zero factor, indicating that the equation $E = [R, \overline{T}]_\times$ with the motion parameters $[R, T]$ can only be obtained as $[R, \overline{T}]$, that is, with the difference of a non-zero factor, where $T = \lambda \overline{T}$. Here, the physical meaning of the non-zero factor λ is the baseline length.

2.5 Relative positional calculation of the omnidirectional, stereo vision system

Based on the epipolar geometry of the omnidirectional stereo vision system, the relative position of the two omnidirectional cameras is equal to the essential matrix. Using the improved eight-corner-point algorithm [20], the eigenvalue of the decomposed essential matrix is $E = UDV^T$, where $D = diag(\sigma, \ \sigma, \ 0)$; thus, the solution of R and $\overline{T}$ can be expressed as follows:

$$[t]_\times \approx VZV^T, R = UGV^T \text{or} R = UG^T V^T \qquad (11)$$

where

$$G = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, Z = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The symbol $\approx$ indicates that there is a difference in the proportional constant factor λ, which can be calculated by the real translation vector $T = \lambda \overline{T}$, where $\overline{T}$ is the unitized vector of the translation vector solution in Eq. 11. In this case, the physical meaning of λ is the baseline length. The calculation method of λ and the real translation vector T will be described after the stereo rectification method is described.

3 Stereo rectification method

3.1 A new rectification model

In the standard V-binocular omnistereo vision system shown in Fig. 6, the triangulation formula can be simplified as a standard, binocular, visual triangulation formula for perspective stereo systems as follows:

$$r = f \frac{d}{v_2 - v_1} = f \frac{d}{v} \qquad (12)$$

where $v = v_2 - v_1$ describes the pixel disparity of the vertical axis, f is the camera focal length, and d is the length of the vertical baseline. Equation 12 shows that the positioning accuracy is isotropic in a vertically aligned stereo vision system and is not affected by the field of view. The epipolar line is the vertical axis of the cylindrically expanded image, which makes the corresponding polar line simple to determine.

An ideal, vertical-baseline, omnidirectional, stereo pair has a linear, epipolar, geometric relationship in the radial direction; however, this assumption is ideal in that inevitable misalignment errors exist between the two optical axes when applied in practice. Our rectification is a procedure used to obtain a column alignment image pair in our vertically arranged systems. A standard, stereo, cylindrical image pair can be generated using the rotation matrix R, and the unit transition vector $\overline{T}$ can be
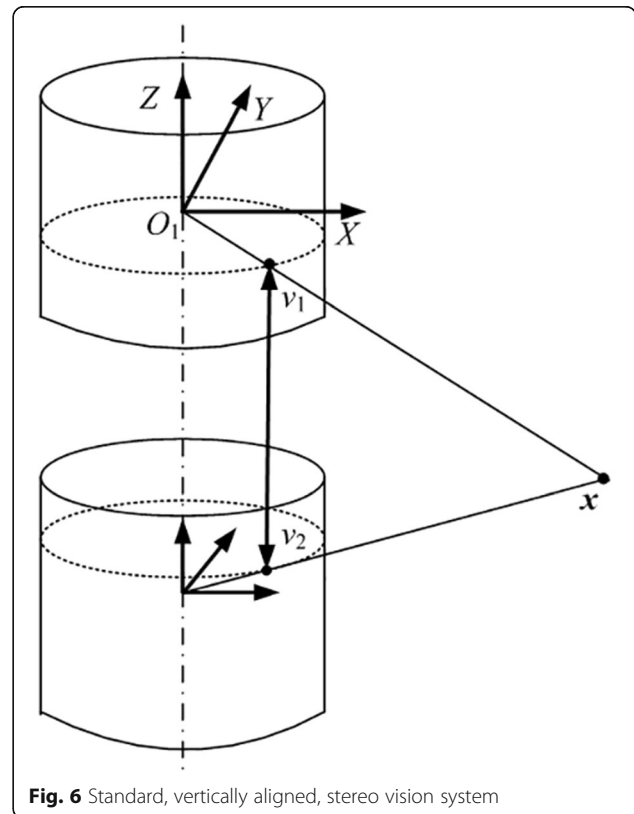

Fig. 6 Standard, vertically aligned, stereo vision system

Cai *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:46

Page 7 of 14

calculated using the essential matrix. This significantly improves the speed of the algorithm making it suitable for real-time applications. After applying our proposed procedure, two important parameters, namely, (a) baseline length (obtained via stereo calibration) and (b) pixel disparity of the vertical axis (easily obtained from a paralleled cylindrical image pair obtained by rectification), can be used in Eq. 12 to calculate metric scene measurements for robot tasks.

Figure 7 shows the cylindrical image, rectification model of the stereo vision system with a deviation in the optical axis. The epipolar line of the upper and lower cylindrical image is not the vertical line of the expanded cylindrical image when the expansion model that we

Fig. 7 Cylindrical-image, rectification model

mentioned is applied directly. However, based on the rectification model shown in Fig. 7, the vertical baseline (the connection of the two origins) of the panoramic, stereo, expanded, cylindrical image pair described in Fig. 6 can be obtained directly. The corresponding polar line and triangulation formulas are thus simplified.

By using the effective viewpoints of the two omnidirectional vision systems as origins to establish the omnidirectional vision system's coordinates, F_u and F_d and O_{m1} and O_{m2} are the origins of the systems, indicating O_{m1} as the origin of the world coordinate system. R and $\overline{T}$ can be calculated, and the physical meaning of the unit translation vector $\overline{T}$ is the unit vector of O_{m2} in F_u. Point x_u is a point of the upper cylindrical expansion image in F_u, and x_d is a point of the lower cylinder expansion image in F_d. Thus, the following holds:

$$x_u = Rx_d + \lambda\overline{T} \qquad (13)$$

The line connecting the two origins is the axis of the new annular, cylindrical expansion image. In Fig. 7, the normal vector of the annular, cylindrical top plane π can be expressed as $n = [n_x, n_y, n_z] = -\overline{T}$ in F_u. F_u is transformed to make the plane π perpendicular to the Z'_{m1} axis in the new coordinate system F'_u via a rotation transformation, which can be expressed as

$$x'_u = R_u x_u \qquad (14)$$

where

$$R_u = \begin{bmatrix} \dfrac{\overline{(m \times n) \times n}}{\overline{(m \times n)}} \\ n \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{n_x n_z}{\sqrt{1-n_z^2}} & \dfrac{n_y n_z}{\sqrt{1-n_z^2}} & -\sqrt{1-n_z^2} \\ \dfrac{-n_x}{\sqrt{1-n_z^2}} & \dfrac{-n_y}{\sqrt{1-n_z^2}} & 0 \\ n_x & n_y & n_z \end{bmatrix}$$

$$(15)$$

and $m = [0, 0, 1]^T$. F'_u is the expanded cylindrical coordinate system shown in Fig. 4, and x_u is described by Eq. 8. Similarly, transforming the expanded cylindrical image in the coordinate system F_d, where $n = -R\overline{T}$, the formula for R_d can be described by Eq. 15.

In Eq. 15, the calculation of the rotation matrix R_u does not rely on the real solution of the translation vector T; indeed, the rotation matrix R_u can be calculated using the unit translation vector $\overline{T}$, which is directly decomposed by the essential matrix E. Using Eqs. 8 and 13–15 to calculate the point's 3D coordinates of the expanded cylindrical image in the mirror coordinate system, the unified sphere-imaging model's projection
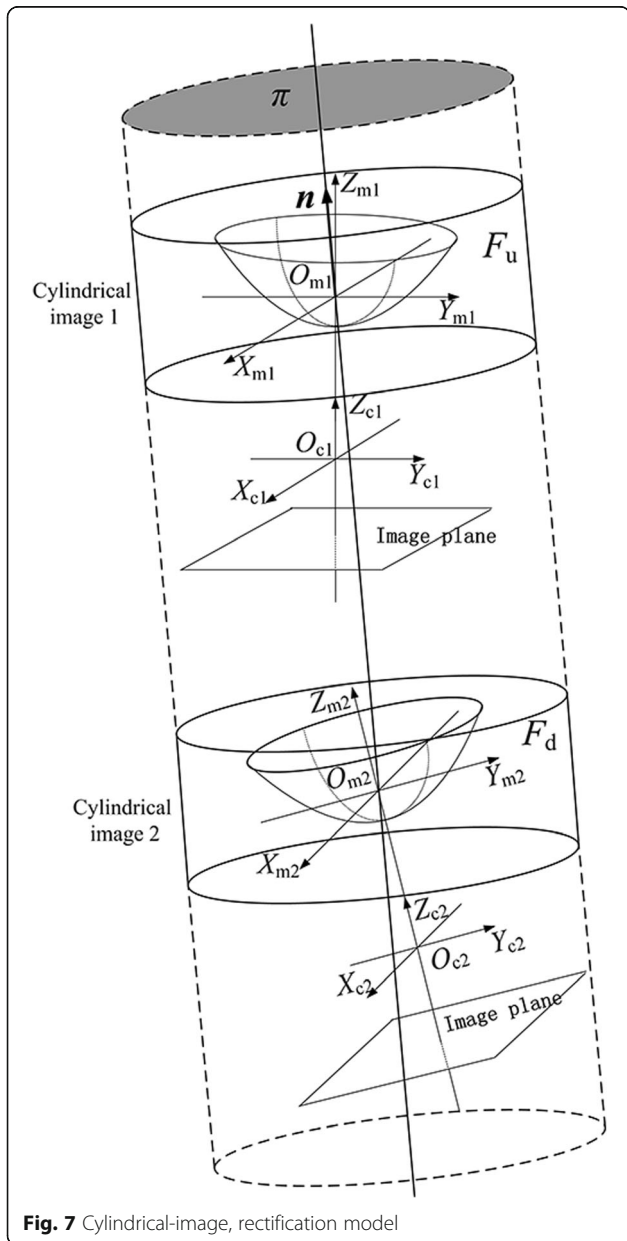
Fig. 8 V-binocular omnistereo vision system equipped on a crawler-type intelligent mobile robot
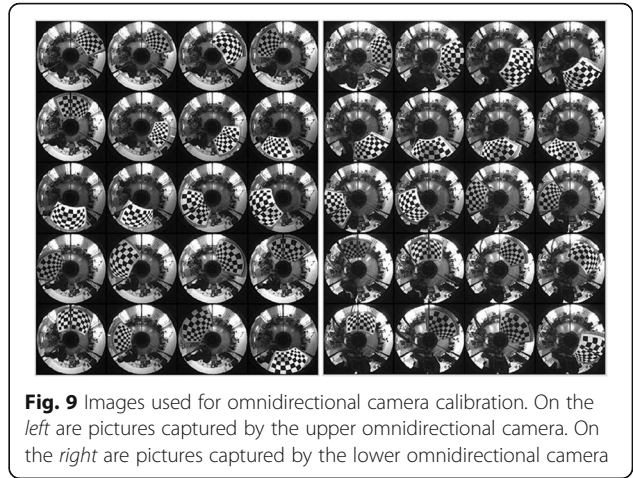

Fig. 9 Images used for omnidirectional camera calibration. On the *left* are pictures captured by the upper omnidirectional camera. On the *right* are pictures captured by the lower omnidirectional camera

point coordinates in the cylindrical image are equal to $\boldsymbol{m} = [u, v]^{\mathrm{T}}$. The corresponding point's incident light vector can be determined using Eq. 8, where $\boldsymbol{x} = [x, y, z]^{\mathrm{T}}$; thus, the corresponding point's angle of incident light can be expressed as follows:

$$\theta = 2\pi u/W, \alpha = \arctan\left(\frac{z}{\sqrt{x^2 + y^2}}\right) \tag{16}$$

This angle's corresponding parallax value is assumed to be v', and as a result, based on Eq. 12, the corner space coordinates can be expressed as

$$\begin{cases} X = r\cos\theta \\ Y = r\sin\theta , \\ Z = r\tan\alpha \end{cases} \qquad r = f\frac{d}{v'} = f\frac{\lambda}{v'} \tag{17}$$

The length between the two corner points of the calibration board is assumed to be equal to L; then, the coordinates in the omnidirectional coordinate system can be expressed as $X_1 = [X_1, Y_1, Z_1]^{\mathrm{T}}$ and $X_2 = [X_2, Y_2, Z_2]^{\mathrm{T}}$, and the parallax values are v'_1 and v'_2, respectively. Thus, the following holds:

formula can be applied to facilitate the cylindrical expansion of the upper and lower panoramic stereo image pair to obtain the standard, cylindrical, stereo image pair. Then, the standard, vertical-baseline, omnidirectional, stereo-vision-positioning model in Fig. 6 is used to calculate the 3D coordinates of the spatial points according to Eq. 12.

3.2 Baseline length calculation
The 3D coordinates of one of the calibration board's corner points are $\boldsymbol{X} = [X, Y, Z]^{\mathrm{T}}$, and the corresponding

Table 1 Mirror parameters and camera parameters given by the manufacturer

Mirror parameters		Camera parameters	
a (major axis)	31.2888 mm	Maximum resolution	2448 × 2048 pixels
b (minor axis)	51.1958 mm	Effective resolution	1360 × 1360 pixels
ξ (mirror parameter)	0.82	Frame rate	10 frames/s
Unilateral vertical viewing angle	120°	Interface	1394b

Table 2 Intrinsic calibration parameters of the omnidirectional cameras

	The upper camera	The lower camera
Main point position $[u_0, v_0]$/(pixels)	[673.59, 683.82]	[669.98, 676.12]
Equivalent focal length $[\gamma_1, \gamma_2]$	[410.44, 411.28]	[412.37, 414.78]
Mirror parameter ξ	0.83176	0.84279
Radial distortion $[k_1, k_2]$	[-0.08661, 0.00732]	[-0.09310, 0.00961]
Tangential distortion $[k_3, k_4]$	[-0.00131, 0.00279]	[0.00445, 0.00387]
Re-projection error	[0.63084, 0.59738]	[0.99979, 0.95924]

Cai et al. EURASIP Journal on Image and Video Processing (2017) 2017:46

Page 9 of 14

$$L = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2} = \sqrt{\left(f\frac{\lambda}{v_1}\cos\theta_1 - f\frac{\lambda}{v_2}\cos\theta_2\right)^2 + \left(f\frac{\lambda}{v_1}\sin\theta_1 - f\frac{\lambda}{v_2}\sin\theta_2\right)^2 + \left(f\frac{\lambda}{v_1}\tan\alpha_1 - f\frac{\lambda}{v_2}\tan\alpha_2\right)^2}$$

$$= \lambda\sqrt{\left(\frac{f}{v_1}\cos\theta_1 - \frac{f}{v_2}\cos\theta_2\right)^2 + \left(\frac{f}{v_1}\sin\theta_1 - \frac{f}{v_2}\sin\theta_2\right)^2 + \left(\frac{f}{v_1}\tan\alpha_1 - \frac{f}{v_2}\tan\alpha_2\right)^2}$$

where the proportional coefficient λ is described by

$$\lambda = L \Bigg/ \sqrt{\begin{array}{l}\left(\dfrac{f}{v'_1}\cos\theta_1 - \dfrac{f}{v'_2}\cos\theta_2\right)^2 + \left(\dfrac{f}{v'_1}\sin\theta_1 - \dfrac{f}{v'_2}\sin\theta_2\right)^2 \\ + \left(\dfrac{f}{v'_1}\tan\alpha_1 - \dfrac{f}{v'_2}\tan\alpha_2\right)^2\end{array}}$$

(18)

Therefore, an accurate value of the proportional constant factor λ can be calculated from two corner points that are accurately located on the calibration board; the real value of the translation vector $\boldsymbol{T}$ can thus be obtained.

4 Results and discussion
4.1 V-binocular omnistereo vision system
Figure 8 presents our experimental equipment. A V-binocular omnistereo system was used as a vision sensor in an REVV-B32 crawler-type intelligent mobile robot. The robot was used for the tracking and localization of moving targets in our lab. The two single-viewpoint omnidirectional cameras consisted of high-accuracy hyperbolic mirrors and GREY POINT 1394b cameras, and the parameters of which are shown in Table 1, which were given by the manufacturer.

The camera base was equipped with a 3°-of-freedom adjustment device. By using a single-view-point constraint determination method [20] and by adjusting the device, the single-view-point constraint was considered to be satisfied during camera-mirror assembly, and the installation accuracy of the mechanical structures was

guaranteed. There were no changes in the extrinsic parameters of our hardware configuration. The baseline was defined as the distance from the focus of the two mirrors under the unified sphere-imaging model, which was accurately measured directly by a long Vernier caliper. The installation spacing of the vertical baseline was 332 mm, which was used as a ground truth to validate the accuracy of the stereo calibration method.

4.2 Single camera calibration experiment
The OpenSource calibration toolbox [10] was directly used to independently calibrate the intrinsic parameters of the upper and lower omnidirectional cameras. A total of 20 images were obtained by each camera. The calibration images are shown in Fig. 9.

The calibration results are shown in Table 2. The accuracy of the toolbox is demonstrated in [18]. From Table 2, the calibrated parameter of the upper mirror parameter was $\xi = 0.83176$, and the lower mirror parameter was $\xi = 0.84279$. Compared with the designed parameter $\xi = 0.82$ from Table 1, the calibration results indicate that the upper and lower omnidirectional cameras were both correctly installed.

The results of cylindrical expansion using the intrinsic calibration results, the expansion model, and a stereo pair (shown in Fig. 10) are shown in Fig. 11a left. Details are provided in Fig. 11a right, which shows that the spatial imaging pixels are out of alignment due to co-axial installation errors. Therefore, stereo calibration and rectification are required to obtain a pixel-aligned image pair.
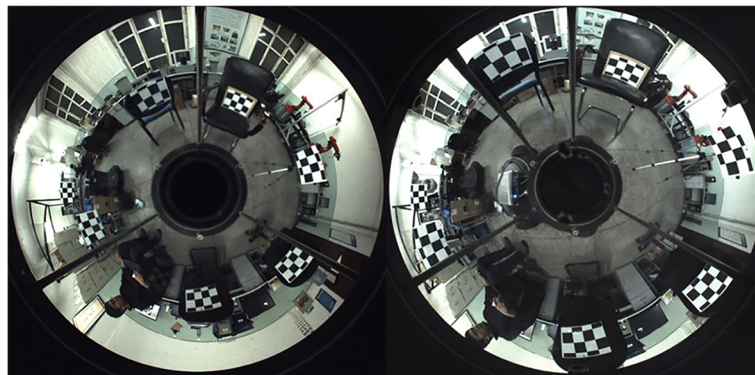

Fig. 10 Panoramic stereo pair with seven calibration boards

Cai *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:46
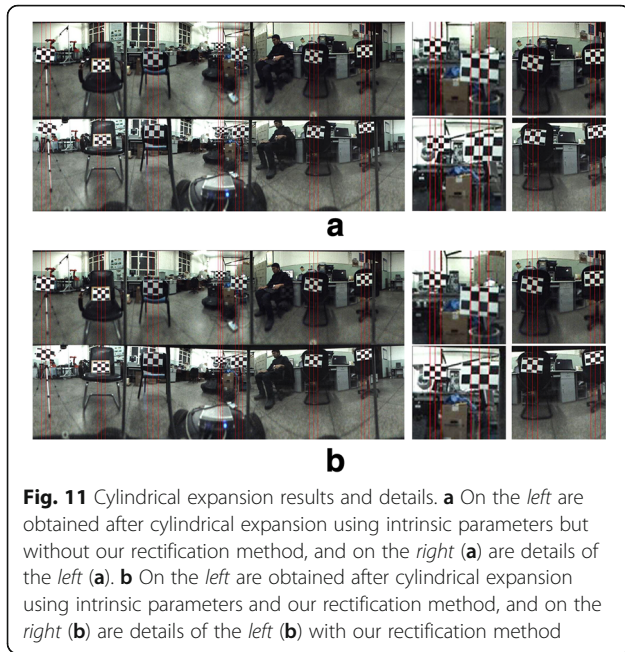
Page 10 of 14

Fig. 11 Cylindrical expansion results and details. **a** On the *left* are obtained after cylindrical expansion using intrinsic parameters but without our rectification method, and on the *right* (**a**) are details of the *left* (**a**). **b** On the *left* are obtained after cylindrical expansion using intrinsic parameters and our rectification method, and on the *right* (**b**) are details of the *left* (**b**) with our rectification method

4.3 Stereo calibration

Seven calibration boards are presented in Fig. 10, and we manually measured the distance to the corner points of the seven calibration boards with respect to the center of the sensor when taking these photos. Three corner matching points from each calibration board are extracted from the image pair. We manually extracted the initial matching points and then used an extraction algorithm with subpixel accuracy in a 9-pixel neighborhood of the initial points to extract matching points. This can significantly reduce matching errors, thereby improving the calculation accuracy and eliminating noise interference. Using the improved method based on the essential matrix, the calculated results for the camera pose using 21 pairs of matching points are shown in Table 3 after the incorrect results are rejected.

The original method [5] was also used for comparative experiment results. We used the same image (Fig. 10) in the contrast stereo calibration method. The following experiment was performed. First, we calculated the extrinsic parameter matrices $[R_1, T_1]$ and $[R_2, T_2]$. The coordinates of a calibration point in the world coordinate system and in the two omnidirectional coordinate systems were $X = [X, Y, Z]^\mathrm{T}$, $x_1 = [x_1, y_1, z_1]^\mathrm{T}$, and $x_2 = [x_2, y_2, z_2]^\mathrm{T}$.

Then, the transformation between X, x_1, and x_2 is achieved as follows:

$$\begin{cases} x_1 = R_1 X + T_1 \\ x_2 = R_2 X + T_2 \end{cases} \tag{19}$$

After eliminating X, the following is obtained:

$$x_2 = R_2 R_1^{-1} x_1 + T_2 - R_2 R_1^{-1} T_1 = R x_1 + T \tag{20}$$

where $R = R_2 R_1^{-1}$ and $T = T_2 - R_2 R_1^{-1} T_1$ are the rotation matrix and translation vector of the two cameras.

The comparison results of our stereo calibration method with the contrast method are shown in Table 4.

The proportional constant factor λ was calculated using Eq. 18. The result of the improved method was $\lambda = 330.1302$ and $T = \lambda \overline{T}$; therefore, the final translation vector obtained using our calibration method was $T = [0.3301, 1.3205, -330.1302]$. The modulus of vector T is the calibrated baseline length. The baseline length of our system was 332 mm. The deviation between our calibration and true value was 1.8698 mm with an error of 0.56%. The translation vector calculated by the original method in Table 4 was $T = [-7.9351, 4.4249, -322.8688]$, whose model is 322.9966. The deviation from the true value of 332 mm was −9.0034 mm, giving an error of 2.71%. Thus, the improved method has better performance.

When stereo calibration is performed using the contrast method, the calibrated results for each stereo pair will be slightly different. This is because the method [10] does not allow for the manual extraction of grid points. Mei and Rives only considered images wherein the grid points were successfully extracted, which increases noise and rounding errors. The results obtained using Eq. 20 can only be used as an initial approximation of the real results. The Levenberg-Marquardt iterative algorithm is then used to perform the calculation, which minimizes the projection error. Such methods typically require multiple calibration stereo pairs, and thus, significantly more work is required to configure the control points and measurement process, which requires rigorous and complex calculations.

In contrast, in our method, we only need to calculate one stereo pair to obtain the rotation matrix and translation vector via our manual extraction method, therein allowing the methods to eliminate noise error to obtain

Table 3 Rotation matrix and unit translation vector results of the improved method

Essential matrix	Rotation matrix R	Unit translation vector $\overline{T}$
$E = \begin{bmatrix} 0.0257 & 0.9993 & 0.0040 \\ 0.9996 & -0.0255 & -0.0011 \\ 0.0084 & 0.0283 & 0.0001 \end{bmatrix}$	$R = \begin{bmatrix} 0.9993 & 0.0259 & -0.0269 \\ -0.0255 & 0.9996 & 0.0131 \\ 0.0273 & -0.0124 & 0.9996 \end{bmatrix}$	$\overline{T} = \begin{bmatrix} 0.0010 \\ 0.0040 \\ -1.000 \end{bmatrix}$

Cai *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:46

Page 11 of 14

Table 4 Comparison results of our stereo calibration method with the original method

	Rotation matrix **R**	Translation vector **T**
Stereo calibration results of the improved method	$\boldsymbol{R} = \begin{bmatrix} 0.9993 & 0.0259 & -0.0269 \\ -0.0255 & 0.9996 & 0.0131 \\ 0.0273 & -0.0124 & 0.9996 \end{bmatrix}$	$\boldsymbol{T} = \begin{bmatrix} 0.3301 \\ 1.3205 \\ -330.1302 \end{bmatrix}$
Stereo calibration results of the original method	$\boldsymbol{R} = \begin{bmatrix} 0.9997 & 0.0217 & 0.0115 \\ -0.0215 & 0.9995 & -0.0223 \\ -0.0120 & 0.0221 & 0.9997 \end{bmatrix}$	$\boldsymbol{T} = \begin{bmatrix} -7.9351 \\ 4.4249 \\ -322.8688 \end{bmatrix}$

the maximum amount of available data. Our calibration method is easier to implement.

4.4 Stereo rectification

A rectification experiment of the upper and lower images in Fig. 10 was performed using the calibrated intrinsic parameters, the calculated rotation matrix, the unit translation vector, and our proposed stereo rectification method. We saved the rectification transforms as lookup tables. The results are shown in Fig. 11b left. Figure 11b right presents the details from the rectified image pair. We can observe that the stereo correspondences fall on the same line and that the pixels are aligned. Comparing the details before rectification with those after rectification, we can conclude that our rectification method is effective.

To accurately evaluate the precision of the stereo rectification process, nine corner coordinates were extracted from each calibration board in Fig. 11b left. We manually extracted the initial matching points and used an extraction algorithm in the 9-pixel neighborhood of the initial points with subpixel accuracy to extract the matching points. The abscissa parallax values of the corresponding corner points were also calculated. Selected

results are shown in Table 5. The rectification accuracy can be determined from the row coordinate deviation value of corner points between the upper and lower images. Table 5 shows that all the row coordinate deviations are at the subpixel level. The noise error can be eliminated via manual extraction of the corner points. The mean value of the abscissa parallax of 63 corresponding corner points was 0.5875 pixels.

With the same picture taken during our experiment and the rectification method proposed by Wang et al. [14], a contrast experiment can be performed under the same conditions. We cannot visually see the difference between the algorithm proposed in this paper and the contrast algorithm visually from the image pair, so we compared the quantized data. The mean value of the abscissa parallax of 63 corresponding corner points was 0.9225 pixels. This finding indicates that our rectification method provides a higher pixel-alignment accuracy. We used 500 pictures to record the computation time of our proposed algorithm and contrast algorithm. The average time of our algorithm was 97 ms per frame, while the average calculation time of the contrast algorithm was 151 ms per frame. Thus, our algorithm requires less computational time than the comparison algorithm.

Table 5 Stereo rectification accuracy

Upper image (pixel)		Lower image (pixel)		Row coordinate parallax (pixel)	Column coordinate parallax (pixel)
x coordinates	y coordinates	x coordinates	y coordinates		
52.13	113.16	52.38	64.42	−0.26	48.74
63.35	113.08	63.26	64.47	0.09	48.61
73.72	113.10	74.20	64.55	−0.48	48.55
363.22	177.98	362.57	73.40	0.65	104.58
389.30	177.51	388.68	72.71	0.61	104.79
414.88	176.49	414.38	73.03	0.49	103.46
887.12	159.46	887.14	55.00	−0.02	104.47
907.97	157.41	907.42	51.14	0.54	106.27
928.93	155.11	928.52	47.51	0.41	107.60
1144.03	136.51	1144.16	40.67	−0.13	95.84
1163.51	136.89	1163.31	41.79	0.20	95.10
1181.96	137.40	1182.18	43.42	−0.22	93.98

Cai et al. EURASIP Journal on Image and Video Processing (2017) 2017:46

Page 12 of 14

Fig. 12 Disparity image obtained using a matching method and the cylindrical expansion image pair from Fig. 11b *left*

There is no ground truth data for the rotation matrix. However, because the rectification method uses R and $\overline{T}$, the rectification accuracy can also laterally show the accuracy of the calculated rotation matrix R.

4.5 Off-line experiments in practice: stereo matching, 3D reconstruction, and depth estimation

We also conducted stereo matching, 3D reconstruction and off-line depth estimation experiments to show the accuracy of our proposed stereo calibration and rectification method.

4.5.1 Stereo matching

Stereo matching was performed on the rectified cylinder stereo image pairs using the semi-global matching algorithm [21], and the matching results are shown in Fig. 12. The gray scale of the pixel points in the disparity map in Fig. 12 is in proportion to the distance between the points and the omnidirectional system (white is close; black is further away). Based on the experimental results, the disparity map represents the distance between most objects in the area of high texture (e.g., the area near the system where the depth information is valid after recovery). By using a subpixel accuracy algorithm to extract the parallax values of the corner points from the calibration board, the disparity between the determined values and the real values was less than 1 pixel. The deviation between the measurement points was less than 60 mm following the image-sampling measurement of seven calibration boards in the range of approximately 1–2 m.
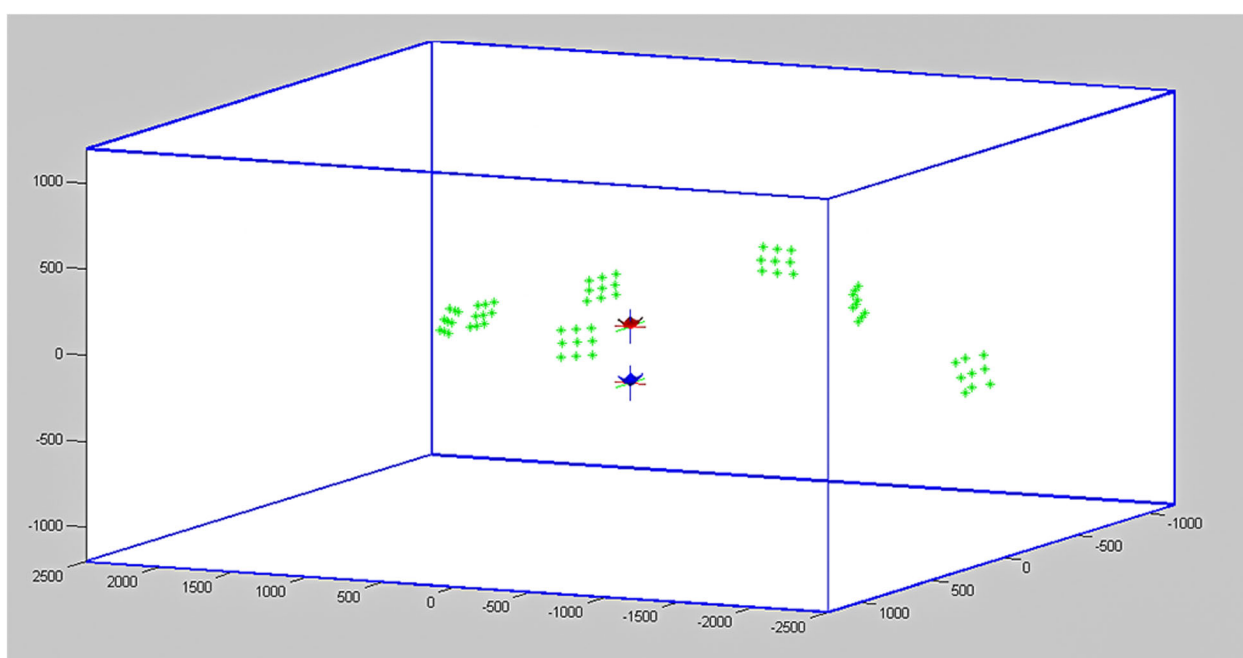

Fig. 13 The 3D positive view of the 3D reconstruction of the seven calibration boards in Fig. 11b *left*. The origin is the middle of the sensor (the upper camera is shown in *red*, and the lower camera is shown in *blue*). The coordinates of all matching points can be extracted from the generated picture
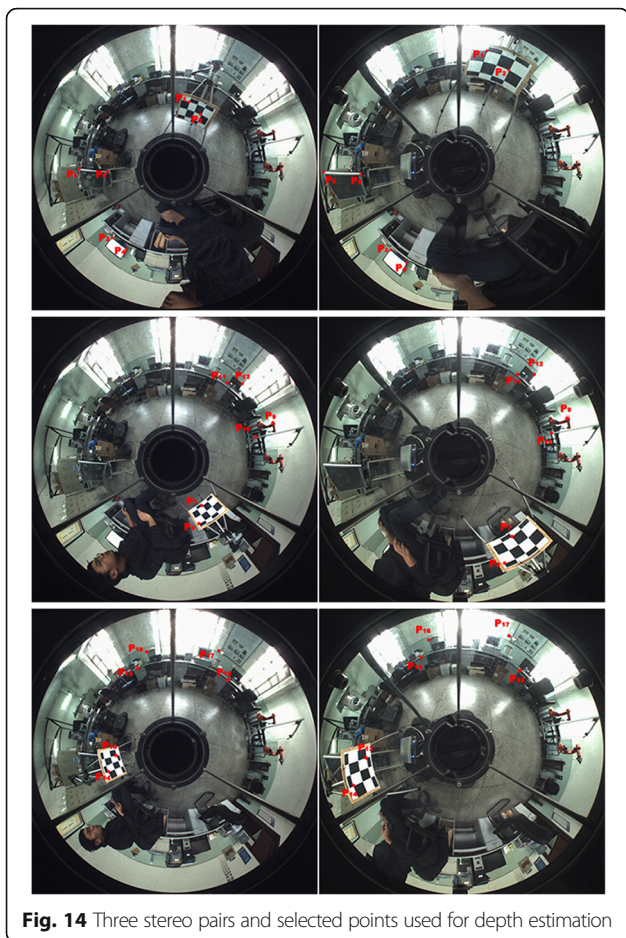
Cai et al. EURASIP Journal on Image and Video Processing (2017) 2017:46

Page 13 of 14

Fig. 14 Three stereo pairs and selected points used for depth estimation

4.5.2 3D reconstruction

Following stereo matching, 3D reconstruction of the seven calibration boards in Fig. 11b left using the unified sphere-imaging model was performed. We used the stereo pair in Fig. 11b left as the input images. A total of nine matching points from each of the seven calibration boards were extracted via the manual extraction method. The 3D reconstruction results are shown in Fig. 13.

The distance between the calibration boards and the panorama system was less than 1–2 m. The origin was the center of the sensor; coordinates of all matching points were extracted from the generated picture. The corner points of the calibration boards are co-planar following reconstruction. By comparing the real distance of these corner points that we manually measured when taking these photos with the coordinates extracted from the generated picture, the average distance error of the 63 matching points was 1.16%. The 3D calculation results demonstrate the precision of our stereo calibration and rectification method.

4.5.3 Depth estimation

We also conducted a depth estimation experiment using the experimental depth estimation procedures in reference [22]. The system configuration remains the same as that introduced at the beginning of the experimental section. In our work, depth was defined as the distance from the middle of the cylindrical shell of the stereo vision system to the object point plus the radius of the shell. Figure 14 shows 18 ground truth points in three unexpanded stereo pairs. We manually measured the distance of each chosen point when taking these photos and considered them to be ground truth data. Following stereo calibration and rectification, we manually selected the target point and extracted the column coordinate deviation using an extraction algorithm subpixel accuracy to elimination of noise error and matching error. Using triangulation Eq. 12, we calculated the estimated values. Table 6 compares the estimated depth with the ground truth depth for the selected points.

According to Table 6, we calculated the average depth estimation error ratio with respect to the ground truth data to be 3.37%. This indicates that depth information can be effectively obtained following stereo calibration and rectification using our method.

5 Conclusions

In this paper, we have proposed a general, comprehensive stereo calibration and rectification method suitable for any V-binocular stereo vision system. We have provided the key techniques required to establish a simple and effective equivalency between an omnidirectional stereo vision system and a perspective vision system,

Table 6 Depth estimation results compared with ground truth depth data

Selected points	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9
Ground truth, m	1.69	1.67	4.01	3.95	2.18	2.24	1.83	1.84	8.02
Estimated depth, m	1.64	1.63	4.16	4.07	2.13	2.18	1.79	1.81	8.57
Depth error ratio, %	2.96	2.40	3.74	3.04	2.29	2.68	2.19	1.63	6.86
Selected points	P_{10}	P_{11}	P_{12}	P_{13}	P_{14}	P_{15}	P_{16}	P_{17}	P_{18}
Ground truth, m	7.87	5.39	5.67	1.89	1.88	4.26	4.62	5.16	4.79
Estimated depth, m	8.36	5.16	5.39	1.92	1.93	4.37	4.76	5.35	4.98
Depth error ratio, %	6.22	4.27	4.94	1.59	2.66	2.58	3.03	3.68	3.97

Cai *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:46

Page 14 of 14

including stereo calibration and rectification. The stereo calibration method was improved. The stereo calibration procedure was simplified based on epipolar geometry. The rounding error was reduced, and the accuracy was ensured by using a manual extraction method. The experimental results verified that the improved stereo calibration method is more accurate than the original method and reduces the complexity of the algorithm. We proposed a simple rectification model. The experimental results verified that the computation time of the proposed rectification method is shorter, and the accuracy is higher than the existing method, which makes it more suitable for real-time vision tasks. Other experiments, such as stereo matching, 3D reconstruction, and depth estimation were also conducted. Experimental results and analyses also support the effectiveness of our methods. In conclusion, our methods can effectively meet the requirements of high-precision vision sensors for robot tasks.

Funding

Authors' contributions

The work presented in this paper was carried out in collaboration between all authors. CC, XW, and QZ conceived the research theme, designed and implemented the feature optimization procedure, and prepared the manuscript. XW and BF wrote the code. XW performed the experiments and analyzed the data. XW reviewed and edited the manuscript. All authors discussed the results and implications, commented on the manuscript at all stages, and approved the final version.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Y Tang, Q Wang, M Zong, J Jiang, Y Zhu, Design of vertically aligned binocular omnistereo vision sensor. EURASIP J Image Video Process **2010**, 1–24 (2010). doi:10.1155/2010/624271
2. S Gulati, T George, *Multi parallax exploitation for Omni-directional imaging electronic eye. United States Patent US 8*, 2012, p. 326
3. G Chen, TD Bui, S Krishnan, S Dai, Circular projection for pattern recognition, in *Advances in Neural Networks: 10th International Symposium on Neural Networks, Dalian, China, July 4-6, 2013. Proceedings, part I*, ed. by C Guo, ZG Hou, Z Zeng (Springer, Berlin, 2013), pp. 429–436. doi:10.1007/978-3-642-39065-4_52
4. Z Zhu, G Wolberg, JR Layne, *Dynamic pushbroom stereo vision. 3D imaging for safety and security* (Springer Verlag, Amsterdam, 2007), pp. 173–199
5. J Cacace, A Finzi, V Lippiello, G Loianno, D Sanzone, Aerial service vehicles for industrial inspection: task decomposition and plan execution. Appl Intell **42**, 49–62 (2015). doi:10.1007/s10489-014-0542-0
6. Q Zhu, X Liu, C Cai, Feature optimization for long-range visual homing in changing environments. Sensors **14**, 3342–3361 (2014). doi:10.3390/s140203342
7. ZE Kadmiri, OE Kadmiri, SE Joumani, Z Kaddouri, Color based omnidirectional target tracking. Int J Imaging Robotics **16**, 1 (2016)
8. M Mendonça, LVR de Arruda, F Neves Jr, Autonomous navigation system using event driven-fuzzy cognitive maps. Appl Intell **37**, 175–188 (2012). doi:10.1007/s10489-011-0320-1
9. H Korrapati, Y Mezouar, Multi-resolution map building and loop closure with omnidirectional images. Auton Robot **41**, 967–987 (2017). doi:10.1007/s10514-016-9560-6
10. C Mei, P Rives, Single view point omnidirectional camera calibration from planar grids, in *Proceedings of the 2007 IEEE International Conference on Robotics and Automation*, 2007, pp. 3945–3950. doi:10.1109/ROBOT.2007.364084
11. JJ Moré, The Levenberg-Marquardt algorithm: implementation and theory. Lecture Notes Math **630**, 105–116 (1978). doi:10.1007/BFb0067700
12. B Micusík, T Pajdla, Structure from motion with wide circular field of view cameras. IEEE Transactions Pattern Anal Mach Intell **28**, 1135–1149 (2006). doi:10.1109/TPAMI.2006.151
13. T Svoboda, T Pajdla, Epipolar geometry for central catadioptric cameras. Int J Comput Vis **49**, 23–37 (2002). doi:10.1023/A:1019869530073
14. Y Wang, X Gong, Y Lin, J Liu, Stereo calibration and rectification for omnidirectional multi-camera systems. Int J Adv Robotic Syst **9**, 1–12 (2012). doi:10.5772/50541
15. YP Tang, CJ Pang, ZS Zhou, YY Chen, Binocular omni-directional vision sensor and epipolar rectification in its omni-directional images. J Zhejiang Univ Technol **1**, 20 (2011)
16. S Baker, SK Nayar, *A theory of catadioptric image formation, in IEEE International Conference on Computer Vision, Mumbai, India*, 1998, pp. 35–42. doi:10.1109/ICCV.1998.710698
17. C Geyer, K Daniilidis, *Conformal rectification of omnidirectional stereo pairs, in Conference on Computer Vision and Pattern Recognition Workshop* (Madison Book Company, Madison, 2003)
18. L Puig, J Bermúdez, P Sturm, JJ Guerrero, Calibration of omnidirectional cameras in practice: a comparison of methods. Comput Vis Image Underst **116**, 120–137 (2012). doi:10.1016/j.cviu.2011.08.003
19. C Geyer, K Daniilidis, A unifying theory for central panoramic systems and practical implications, in *Computer Visition ECCV 2000*, ed. by D Vernon (Berlin, Springer, 2000), pp. 445–461. doi:10.1007/3-540-45053-X_29
20. Q Zhu, F Zhang, K Li, L Jing, On a new calibration method for single viewpoint constraint for catadioptric omnidirectional vision. J Hua Zhong Univ Sci Tech **38**, 115–118 (2010)
21. Hirschmuller H, Accurate and efficient stereo processing by semi-global matching and mutual information, in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (IEEE, 2005), p. 807-814. doi: 10.1109/CVPR.2005.56.
22. Z-H Xiong, I Cheng, W Chen, A Basu, M-J Zhang, Depth space partitioning for omni-stereo object tracking. IET Comput Vis **6**, 153–163 (2012). doi:10.1049/iet-cvi.2010.0115