Optics and Lasers in Engineering 48 (2010) 1252–1261

Contents lists available at ScienceDirect

Optics and Lasers in Engineering

journal homepage: www.elsevier.com/locate/optlaseng

An accurate stereo vision system using cross-shaped target self-calibration method based on photogrammetry

Zhenzhong Xiao [a,*], Jin Liang [a], Dehong Yu [a], Zhengzong Tang [a], Anand Asundi [b]

[a] School of Mechanical Engineering, Xi'an Jiaotong University, No.28, Xianning West Road, Xi'an 710049, Shaanxi, PR China
[b] School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798, Singapore

ARTICLE INFO

Article history:
Received 20 January 2010
Received in revised form
4 June 2010
Accepted 12 June 2010
Available online 3 July 2010

Keywords:
Stereo vision system
Self-calibration
Photogrammetry
Cross target

ABSTRACT

This paper presents an accurate stereo vision system for industrial inspection, which uses a self-calibration method based on photogrammetry. A cross-shaped calibration pattern which is portable and easy to be manufactured is designed. The cross target can be used to calibrate stereo vision systems and obtain higher measurement precision conveniently. The mathematical model of the stereo vision system with 10 distortion parameters for each camera is proposed. The feature point detection method with sub-pixel accuracy is explored. The calibration initial values are computed using the relative orientation method and the direct linear transform (DLT) method of photogrammetry. The bundle adjustment algorithm is used to optimize the calibration parameters as well as the 3D coordinates of the feature points. Experiment results show that the RMS error of the reprojection in our method is less than 0.05 pixels and the distance measurement error is 0.031 mm with a high precision scale bar which length is 221.001 ± 0.003 mm.

1. Introduction

Vision inspection has been widely used in industrial measurement and scientific research, owing to its noncontact manner. Three-dimensional measurement for industrial products inspection requires high accuracy and flexibility. Camera calibration is a key issue that affects the accuracy of stereo vision system.

Camera calibration is the process of determining the internal and external parameters of camera via capturing external reference objects. The single camera calibration has been studied extensively [1–3,13,14]. A popular and practical algorithm is the method developed by Tsai [1] using radial alignment constraint (RAC), but in this method, initial camera parameters are required and only lens radial distortion is considered. Weng et al. [2] proposed a camera calibration with distortion models. Zhang [3] proposed a flexible technique for camera calibration by viewing a plane from different unknown orientations. This algorithm assumes that the calibration pattern is an ideal plane, and ignores the actual errors in manufacturing. However, all these approaches assume that the world coordinates of the feature points of the calibration pattern are accurately known. Therefore they did not consider the inevitable manufacturing errors and measurement errors of these points. These errors could affect the accuracy of the calibration, especially in the stereo measurement system with volatile measurement volume.

The calibration of camera systems without relying on external 3D measurements was also investigated by a number of researchers [4–9,15–17]. Self-calibration only requires the corresponding points of images, and thus is more flexible in practical applications. In [4], Heyden and Astrom proposed a self-calibration algorithm that uses explicit constraints from the assumption of the intrinsic parameters of the camera. They proved that self-calibration is possible under different kind of cameras with the assumptions that the aspect ratio is known and there is no skew. They solved the problem using the bundle adjustment method [6] that requires simultaneous minimization on all reconstructed points and cameras. However, the initialization problem was not properly presented in it. Bougnoux [5] proposed a practical self-calibration algorithm that used the constraints derived from [4]. A linear initialization step is used in the nonlinear minimization, and the bundle adjustment is used in the projective reconstruction step. But none of them is suitable for practical on-site calibration in industrial inspection because of the limit of precision and field of view.

In this paper, an accurate stereo vision system for industrial inspection is proposed. Our system uses a cross target with ring coded points as calibration object. The world coordinates of these ring coded points are not required. The proposed procedure does not require any expensive equipment. Moreover, it is very fast and completely automatic, as the user is only requested to capture a few images of the calibration cross target. The calibration initial value is computed by using the relative orientation method and the direct

* Corresponding author. Tel./fax: 86 29 82669103.
E-mail address: xiaoxice@hotmail.com (Z. Xiao).

0143-8166/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.optlaseng.2010.06.006

linear transform (DLT) method of photogrammetry. The bundle adjustment algorithm is used to optimize the calibration parameters including the 3D coordinates of the ring coded points. It also considers the errors of feature point identification, the radial distortion, the tangential distortion and thin prism distortion for the lens distortion model. A high precision scale bar is used to evaluate the measurement accuracy in the experiments.

2. System model and principle

2.1. Mathematical model of the stereo vision system

In the field of computer vision, the camera model is used to solve the correspondence problem between the 3D object points and 2D image points, for which the perspective projection model is widely used [6]. The perspective projection model is shown in Fig. 1.

In Fig. 1, P is the world coordinate of the object point, S is the coordinate of the projection center, (x,y) is the coordinate of the image point, (x_0,y_0) is the coordinate of the image point O, and f is the focal length. After considering the actual lens distortion, the relationship among the 3D point, the projection center, and the image point is as follows:

$$x-x_0+dx=-f\frac{a_1(X-X_s)+b_1(Y-Y_s)+c_1(Z-Z_s)}{a_3(X-X_s)+b_3(Y-Y_s)+c_3(Z-Z_s)}$$
$$y-y_0+dy=-f\frac{a_2(X-X_s)+b_2(Y-Y_s)+c_2(Z-Z_s)}{a_3(X-X_s)+b_3(Y-Y_s)+c_3(Z-Z_s)} \qquad (1)$$

where (X, Y, Z) is the world coordinate of the object point, (X_s, Y_s, Z_s) is the coordinate of the projection center, and dx, dy are

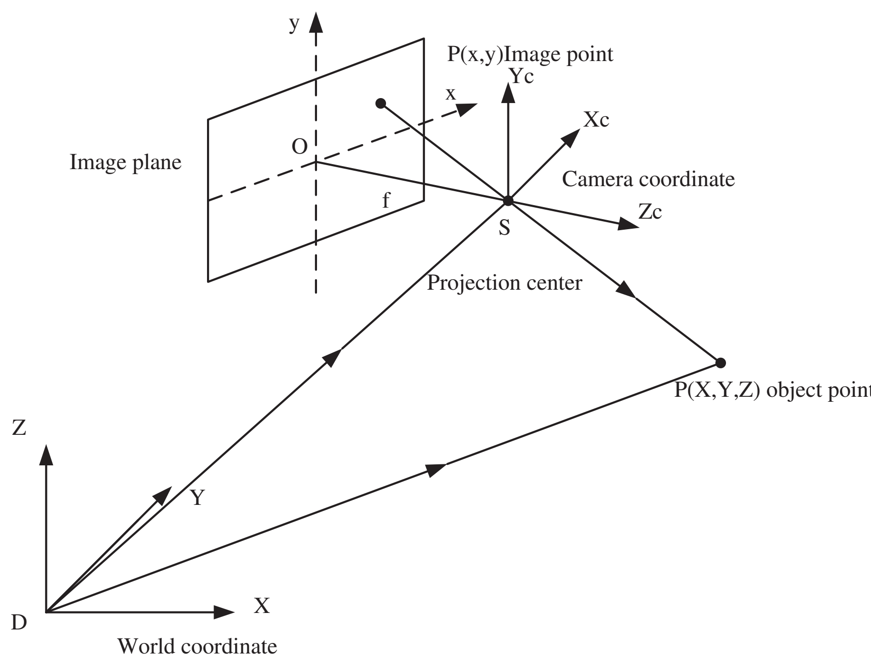

Fig. 1. Perspective projection model of camera.

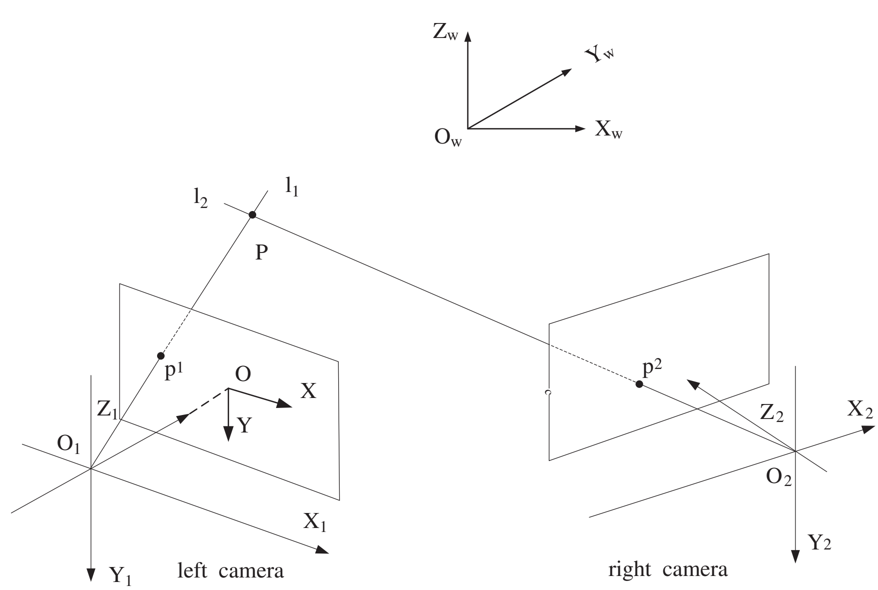

Fig. 2. Mathematical model of stereo vision system.

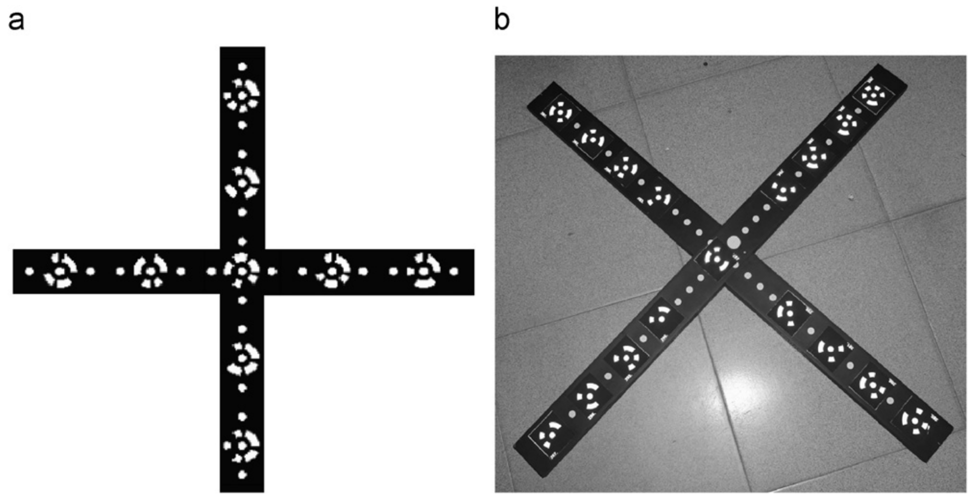

Fig. 3. Calibration cross: (a) design sketch and (b) in-kind photo.

Fig. 4. Two kinds of feature point.

the magnitudes of the lens distortion. The matrix

$$\begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}$$

is the transform matrix from the world coordinate to camera coordinate.

Zhang's calibration algorithm in [3] only considers the two coefficients of the radial symmetrical distortion of the lens. In this paper, we also consider the tangential distortion and thin prism distortion. The new distortion model used in our method is

$$\begin{aligned} dx &= A_1x(r^2-r_0^2)+A_2x(r^4-r_0^4)+A_3x(r^6-r_0^6)+B_1(r^2+2x^2) \\ &\quad +2xyB_2+C_1x+C_2y \\ dy &= A_1y(r^2-r_0^2)+A_2y(r^4-r_0^4)+A_3y(r^6-r_0^6)+B_2(r^2+2y^2)+2xyB_1 \end{aligned} \tag{2}$$

where A_1, A_2, A_3 are radial distortion parameters, B_1 and B_2 are tangential distortion parameters, and C_1 and C_2 are thin prism distortion parameters, r is the image radius and r_0 is the second zero crossing of the distortion curve. This model is slightly different from Bouguet's model [10] since r_0 is employed for balance parameters. It can reduce the RMS error of the reprojection.

The stereo vision system includes two cameras and a digital light procession (DLP) projector. Fig. 2 shows the 3D measurement model of the binocular system.

The stereo vision system should be calibrated by calculating two cameras' internal and external parameters, and the relative position of the two cameras is obtained. The rotation matrix R and

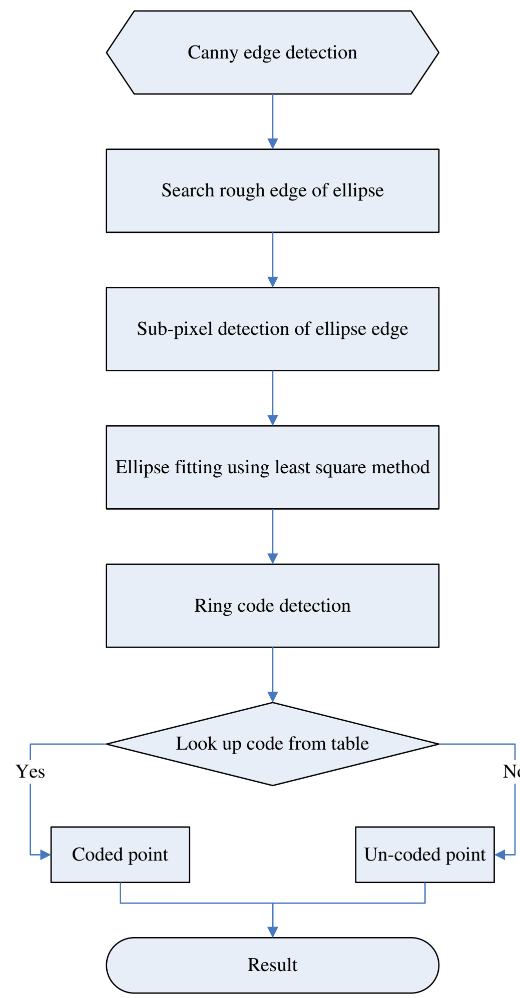

Fig. 5. Flowchart of feature point detection.

translation matrix T are:

$$R=\begin{pmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{pmatrix}, \quad T=\begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}$$

The point coordinates of the object is then calculated as

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = (A^T A)^{-1} A^T B \tag{3}$$

where

$$A = \begin{bmatrix} 1 & 0 & -x_1 \\ 0 & 1 & -y_1 \\ r_1 - r_7 x_2 & r_2 - r_8 x_2 & r_3 - r_9 x_2 \\ r_4 - r_7 y_2 & r_5 - r_8 y_2 & r_6 - r_9 y_2 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ x_2 t_z - t_x \\ y_2 t_z - t_y \end{bmatrix}$$

(x_1, y_1) is the image point coordinate of the left camera and (x_2, y_2) is the image point coordinate of the right camera.

To sum up, all parameters that we need to compute are: left camera: focal length f, principle point (x_0,y_0), distortion parameters A_1, A_2, A_3, B_1, B_2, C_1, C_2; right camera: focal length f', principle point (x'_0,y'_0), distortion parameters $A'_1,A'_2,A'_3,B'_1,B'_2,C'_1,C'_2$; the rotation matrix R and the translation matrix T between the left camera and right camera.

2.2. Calibration cross design and feature point detection

A cross-shaped calibration target is designed in this paper. Since it is portable and easy to be manufactured, it is more convenient than the planar pattern used in conventional methods, especially for large industrial field calibration. The structure does not need the feature points to be in the same plane, and it does not need to know the actual 3D coordinates of the feature points. The only necessary information is the distance between two points as a scale. The calibration process can be completed automatically (Fig. 3).

There are two kinds of feature points on the calibration cross as shown in Fig. 4. The left one called un-coded point is a white point on black background. The right one called coded point has a center point and surrounding discrete ring which denote its exclusive identification. It is used for correlating the coded feature points in different images.

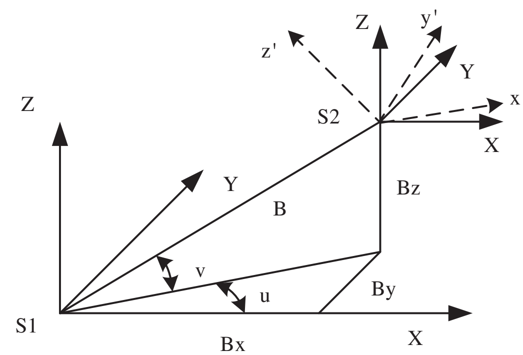

Fig. 7. Relative orientation.

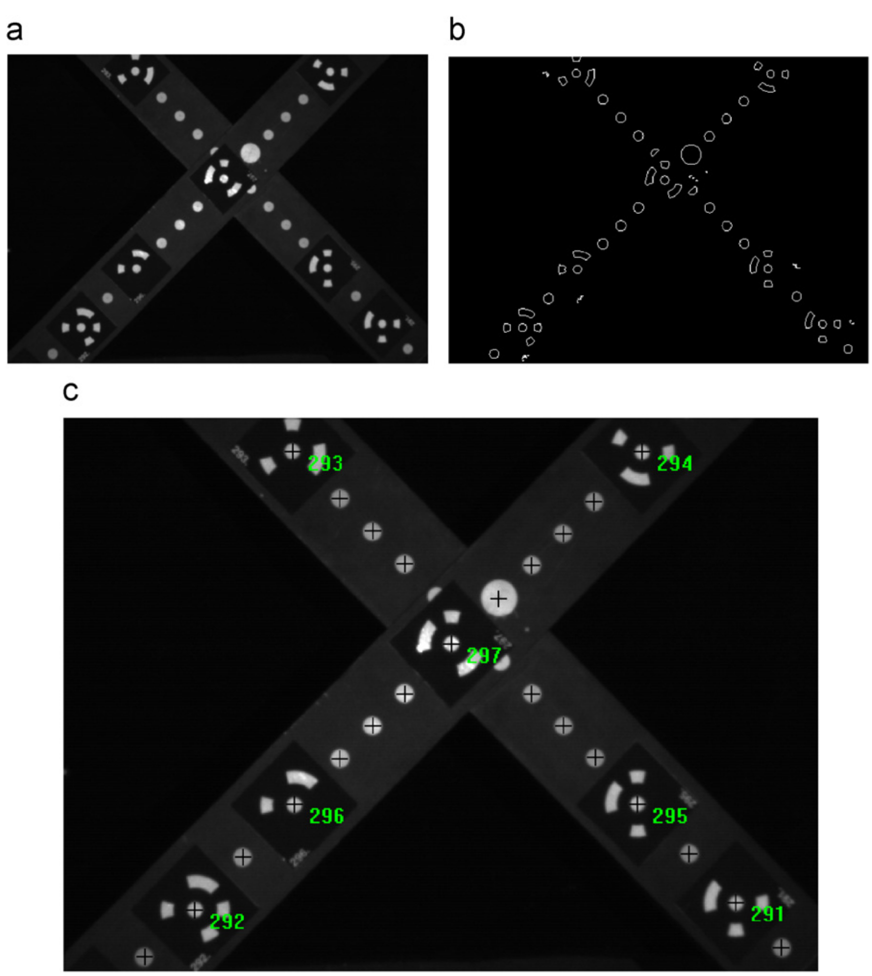

Fig. 6. Feature point detection: (a) calibration image; (b) rough edge; and (c) detection result.

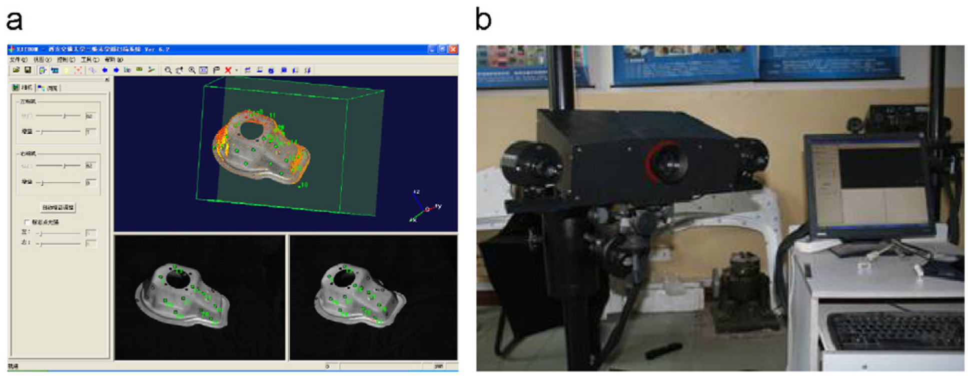

Fig. 8. The stereo vision system: (a) software and (b) hardware.

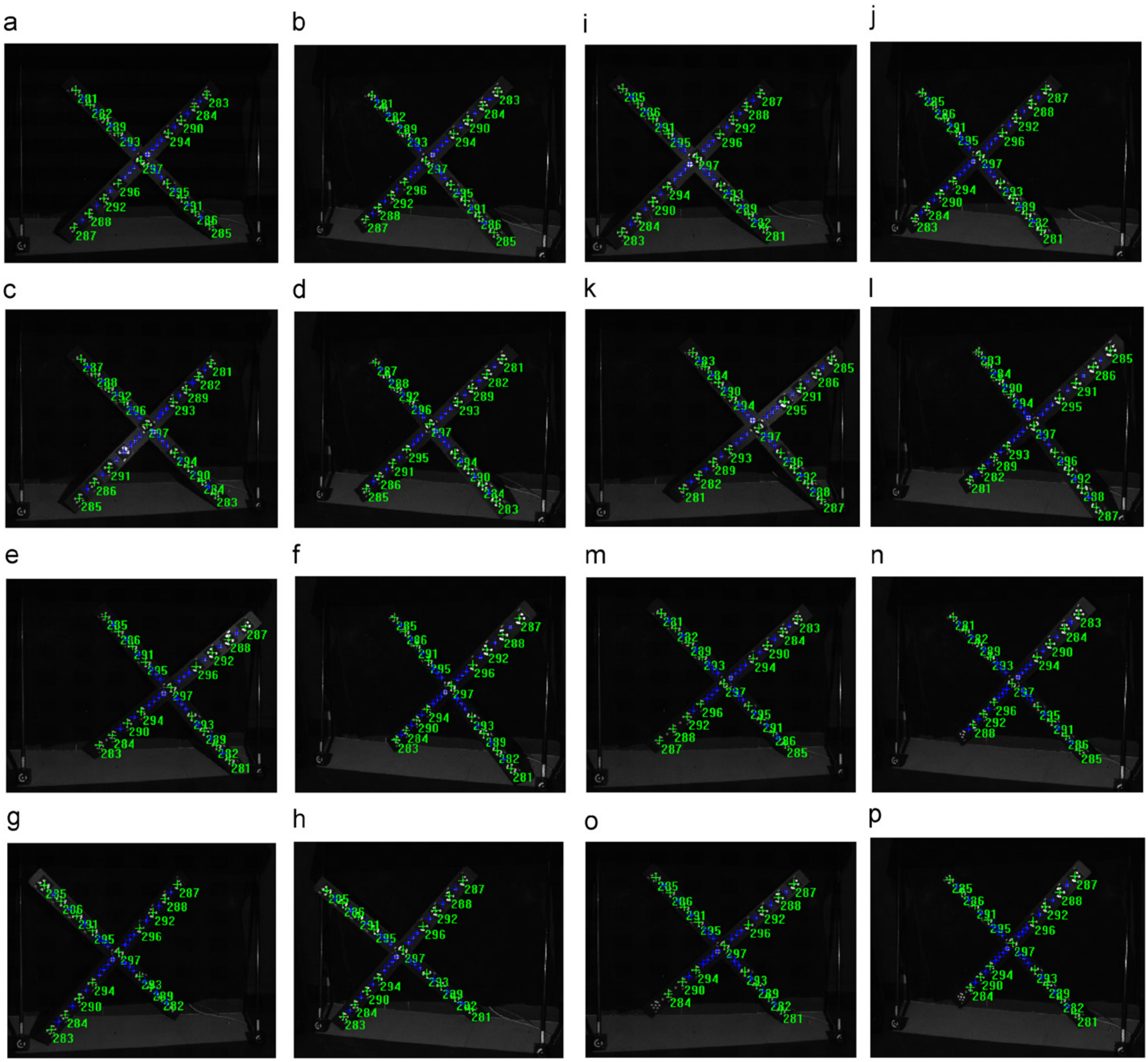

Fig. 9. Images captured for calibration with feature point detection result: (a) left 1; (b) right 1; (c) left 2; (d) right 2; (e) left 3; (f) right 3; (g) left 4; (h) right 4; (i) left 5; (j) right 5; (k) left 6; (l) right 6; (m) left 7; (n) right 7; (o) left 8, and (p) right 8.

Table 1
Internal parameters of left camera and right camera.

Parameter/camera	Left camera		Right camera	
	Initial	Final	Initial	Final
f (pixels)	3076.92	3094.906	3076.92	3088.445
(x_0,y_0)	(640, 512)	(642.768, 538.679)	(640, 512)	(642.272, 546.812)
A_1	0	−2.322028e-008	0	−1.791926e-008
A_2	0	4.366657e-014	0	2.096009e-014
A_3	0	−4.013798e-020	0	−1.458508e-020
B_1	0	−2.980062e-008	0	2.461222e-007
B_2	0	−1.616889e-007	0	−5.099662e-008
C_1	0	−2.990895e-004	0	1.475318e-004
C_2	0	−2.104225e-005	0	1.383162e-004

A circular point becomes an ellipse on the image because of projectivity. The procedure to extract ring feature points from image is illustrated in Fig. 5.

Among the variety of types of edge detecting algorithms, Canny edge detector [11] is widely used in varying applications for its efficiency and continuity. However, the outputs of Canny edge detector are step-like discontinuities expressed in pixel level accuracy as shown in Fig. 6.

For the sake of sub-pixel definition requirement for calibration processing, an improved edge localization algorithm is explored after applying Canny edge detector in image, this method using gradient of adjacent pixels around edge pixel to adjust the edge position, result in a sub-pixel edge. To locate the center of the point, a least-square ellipse fitting algorithm [12] is employed on the sub-pixel edges. The general equation of an ellipse is described as

$$x^2+2Bxy+Cy^2+2Dx+2Ey+F=0 \tag{4}$$

where ellipse parameters B, C, D, E, and F are solved by the least-square ellipse fitting method. The center coordinates of the ellipse are computed by

$$\left.\begin{aligned} x_0 &= \frac{BE-CD}{C-B^2} \\ y_0 &= \frac{BD-E}{C-B^2} \end{aligned}\right\} \tag{5}$$

2.3. Stereo vision geometry and orientation methods

The relative orientation is a method to determine the relationship between two images. There are six exterior orientation elements to determine the position of an image. Therefore, to determine orientation of two images we need the 12 exterior orientation elements:

$$\text{Image 1}: X_{s1},Y_{s1},Z_{s1},\varphi_1,\omega_1,\kappa_1$$

$$\text{Image 2}: X_{s2},Y_{s2},Z_{s2},\varphi_2,\omega_2,\kappa_2$$

As shown in Fig. 7, the exterior orientation elements of the image 2 minus the exterior orientation elements of the image 1:

$$\Delta X_s = X_{s2}-X_{s1}$$

$$\Delta Y_s = Y_{s2}-Y_{s1}$$

$$\Delta Z_s = Z_{s2}-Z_{s1}$$

$$\Delta\varphi = \varphi_2-\varphi_1$$

$$\Delta\omega = \omega_2-\omega_1$$

$$\Delta\kappa = \kappa_2-\kappa_1$$

where ΔX_s, ΔY_s, ΔZ_s are the projection of the photography baseline (two cameras projection center connection) in the three coordinate axis of the world coordinate system, marked B_x, B_y, B_z. where $B=\sqrt{B_x{}^2+B_y{}^2+B_z{}^2}$, $\tan(\mu)=B_y/B_x$, $\sin(\nu)=B_z/B$, then B_x, B_y, B_z can be instead of the elements of B, μ, ν.

And then, we can see that the length of the baseline B affects only the stereo pair of the scale, does not affect their relative position. As a result, the relative orientation elements of the stereo pair of images need only five parameters, namely, μ, ν, $\Delta\varphi$, $\Delta\omega$, $\Delta\kappa$.

Suppose the coordinates of a point P in image 1 is $(x_1, y_1, -f)$ in S_1-xyz coordinate system, the coordinates of the point P in the image 2 is $(x_2', y_2', -f)$ in $S_2-x'y'z'$ coordinate system, the coordinates of the point P is (x_2, y_2, z_2) in S_2-xyz coordinate system, and the coordinates of the point S_2 is (B_x, B_y, B_z) in S_1-xyz coordinate system. Assuming the coordinate rotation matrix R is formed by $S_2-x'y'z'$ and S_2-xyz, the relationship between (x_2, y_2, z_2) and $(x_2', y_2', -f)$ is described as

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = R \begin{bmatrix} x_2' \\ y_2' \\ -f \end{bmatrix} \tag{6}$$

$$\begin{vmatrix} B_x & B_y & B_z \\ x_1 & y_1 & -f \\ x_2 & y_2 & z_2 \end{vmatrix} = 0 \tag{7}$$

where $B_y = B_x\tan(u) \approx B_x u$, and $B_z = B_x\tan(v) \approx B_x v$

Eq. (7) can be simplified as

$$\begin{vmatrix} 1 & u & v \\ x_1 & y_1 & -f \\ x_2 & y_2 & z_2 \end{vmatrix} = 0 \tag{8}$$

Therefore, as long as two images have more than 5 common points, their relative positions μ, ν, $\Delta\varphi$, $\Delta\omega$, $\Delta\kappa$ can be solved using Eqs. (6) and (8).

2.4. System self-calibration based on photogrammetry

Our calibration procedure consists of the following six steps:

(1) place the calibration target 1 m from the stereo vision system, and the two cameras capture eight images in different locations and orientations simultaneously by moving the cross-shaped target;
(2) identify feature points in the eight groups of images, and get the image coordinates of circular feature points' centers and the codes of ring coded feature points;

(3) calculate the relative orientation for the first two images, reconstruct the 3D coordinates of the ring coded feature points;
(4) compute the orientation of other images by using the direct linear transform (DLT) method;
(5) use the bundle adjustment method to iteratively adjust two cameras' internal and external parameters and the world coordinates of the feature points on the calibration cross; and
(6) the final calibration parameters of the stereo vision system are obtained after the bundle adjustment iteration.

The method of bundle adjustment [6] in photogrammetry is based on the collinear equation, which estimates 3D object coordinates, image orientation parameters, and any additional model parameters together with related statistical information about accuracy and reliability. Since all observed (measured) values and all unknown parameters of a photogrammetric project are taken into account within one simultaneous calculation, the bundle adjustment algorithm is the most powerful and accurate method of image orientation and point determination in photogrammetry [4].

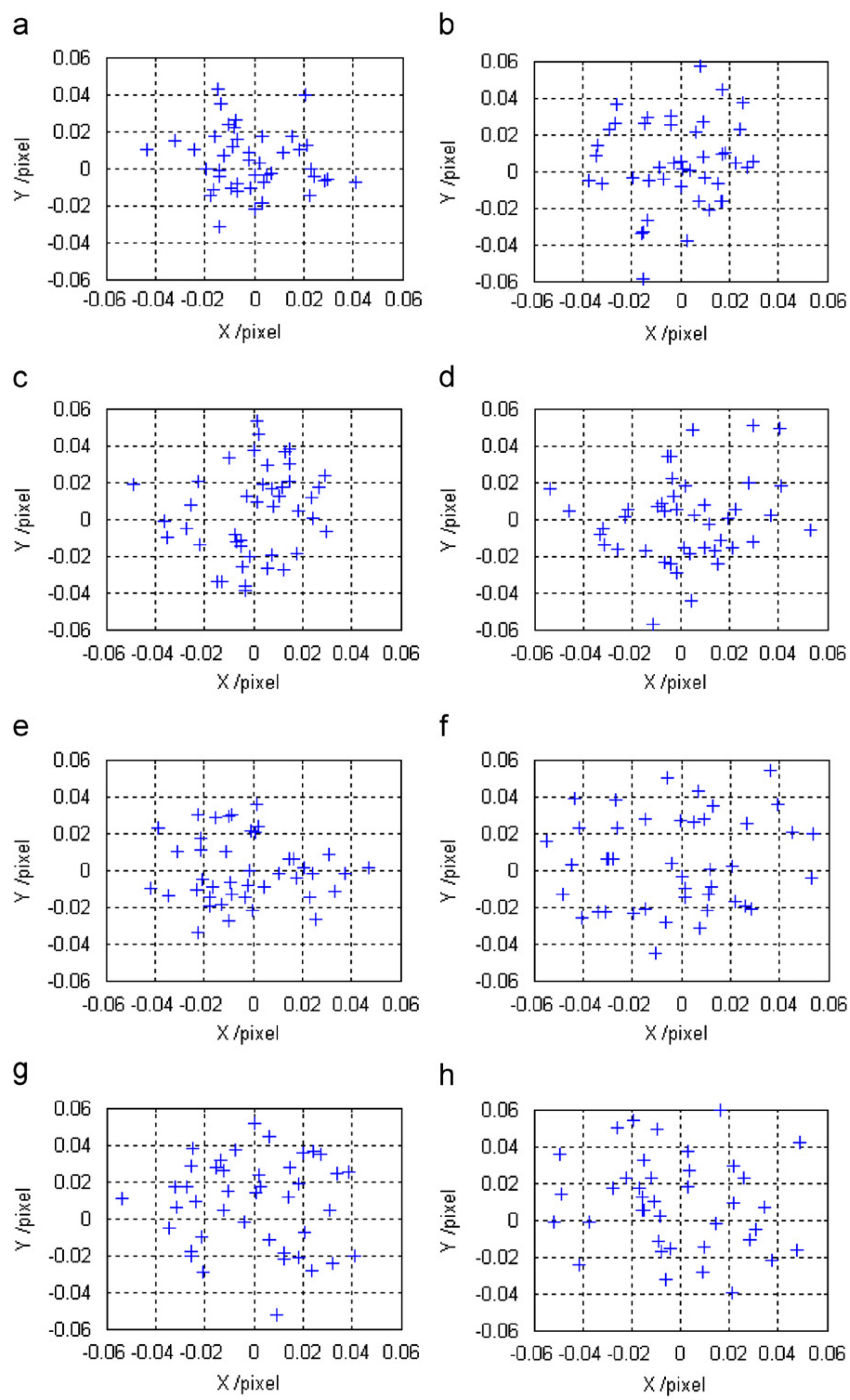

Fig. 10. Reprojection errors for each image: (a) left 1; (b) right 1; (c) left 2; (d) right 2; (e) left 3; (f) right 3; (g) left 4; (h) right 4; (i) left 5; (j) right 5; (k) left 6; (l) right 6; (m) left 7; (n) right 7; (o) left 8, and (p) right 8.

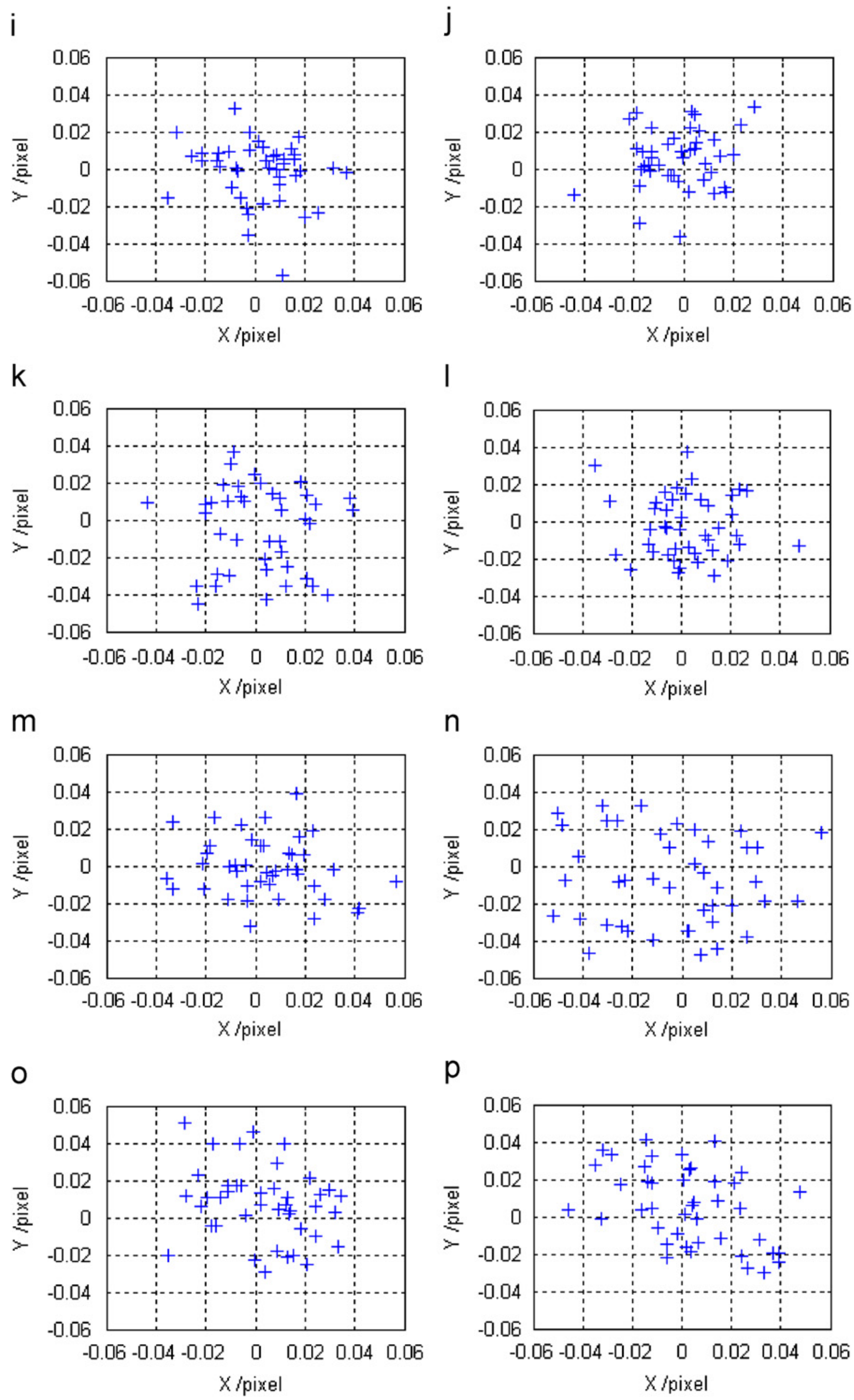

Fig. 10. (*Continued*)

The observation equations of the bundle adjustment method are obtained through the linearization of Eq. (1)

$$V = AX_1 + BX_2 + CX_3 - L \tag{9}$$

where V is residuals of reprojection, X_1, X_2, and X_3 are the partial derivative of the internal parameters (include distortion parameters), external parameters, and object points coordinates, respectively. A, B, and C are derivative matrix, respectively, and L is the observation vector.

The task of bundle adjustment algorithm is to determine all the unknown parameters from a number of observed values. If more observations are available than required for the determination of the unknowns, there is normally no unique solution and the unknown parameters are estimated according to functional and stochastic models. The condition for the residuals V of Eq. (9) is

$$\min V^T V$$

The accuracy of the observations and the adjusted unknowns are important when analyzing the quality in an adjustment procedure. Here the residuals V represents the RMS error of the reprojection.

3. System setup and experiments

Our stereo vision system software is developed using VC++6.0 in the Windows XP environment. The hardware platform consists of two 1.3 million-pixel CCD cameras and a DLP projector. The CCD size is 6.6 mm × 5.3 mm, the pixel size is 5.2 μm × 5.2 μm, the resolution is 1280 × 1024, and the frame rate is 15 frames/s. The lens we used is Schneider Lens with focal length fixed at 16 mm. As shown in Fig. 8, the measuring devices are installed on a heavy frame, the two cameras are located on both ends of a beam and the projector is in the middle.

Eight images of the calibration target under different orientations were taken for each camera, which are shown in Fig. 9. The parameters can be estimated by applying the calibration algorithm described in Section 2.4.

The estimated parameters of the left camera are shown in Table 1.

The rotation matrix is

$$R = \begin{pmatrix} 9.003298\text{e-}001 & 1.417599\text{e-}002 & -4.349772\text{e-}001 \\ -4.112756\text{e-}003 & 9.997019\text{e-}001 & 2.406775\text{e-}002 \\ 4.351887\text{e-}001 & 9.997019\text{e-}001 & 9.001197\text{e-}001 \end{pmatrix}$$

and the translation matrix is

$$T = \begin{pmatrix} -4.712901\text{e}+002 \\ 2.542105\text{e}+000 \\ -1.097902\text{e}+002 \end{pmatrix}$$

As shown in Fig. 10, the RMS error of the reprojection in each image is <0.05 pixels.

To demonstrate the effectiveness of the calibration procedure, an 803-MCP scale bar bought from American Brunson Inc. is used for testing in our system, as shown in Fig. 11. The scale bar has two circle points at both ends. It has a length of 221.001 ± 0.003 mm between the two circle points' center. In the experiment, the scale bar is put at seven different orientations and positions within the measurement volume. At each position, the length of the scale bar is measured by the vision system using our calibration result. The measuring results are shown in Table 2.

Fig. 11. The 803-MCP scale bar.

Table 2
The measuring result of the scale bar.

Position	Result (mm)	Error (mm)
1	221.038	0.037
2	221.022	0.021
3	220.981	−0.02
4	221.056	0.055
5	221.024	0.023
6	221.039	0.038
7	221.062	0.061
Mean	221.032	0.031
Standard deviation		0.026

4. Discussions

In most conventional calibration methods,stereo vision system calibration is in the way that calibrating two cameras separately. The camera calibration is accomplished based on the reference data composed of the 3D reference points and their 2D camera correspondences extracted from the images. In this way, the calibration accuracy rely more on the precision of the 3D reference points' world coordinates. It is difficult to get highly accurate 3D reference points. The error of the 3D reference points' world coordinates is neglected in conventional methods. The bundle adjustment algorithm is employed in our method, which adjusts the 3D reference points' world coordinates, image orientation parameters, and additional model parameters altogether. If more observations are available, the more optimal results are estimated according to stochastic principles. So calibrating two cameras in one procedure can get optimal results theoretically.

5. Conclusions

We have developed an accurate stereo vision system using cross target self-calibration method. Major issues of the vision system have been discussed in detail. A calibration procedure is proposed by using a cross-shaped calibration target. The cross-shaped calibration target is portable and easy to be manufactured. It is more convenient than the planar pattern used in conventional methods. A set of experiments has been designed to demonstrate the effectiveness of the proposed approach. The experiment result shows that the RMS error of the reprojection in our method is <0.05 pixels and the distance measurement error is 0.031 mm with an 803-MCP scale bar which length is 221.001 ± 0.003 mm. This study may provide some valuable references for stereo vision and camera self-calibration research.

Acknowledgments

The authors acknowledge the support of the National Natural Science Foundation of China (Grant no. 50975219) and the National High Technology Research and Development Program of China (863 Program, Grant no. 2007AA04Z124).

References

[1] Tsai RY. A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. IEEE J Robot Automat 1987;RA-3(40):323–44.
[2] Weng J, Cohen P, Herniou M. Camera calibration with distortion models and accuracy evaluation. IEEE Trans Pattern Anal Mach 1992;10:965–80.
[3] Zhang Z. A flexible new technique for camera calibration. IEEE Trans Pattern Anal Mach Intell 2000;22(11):246–53.

[4] Heyden A, Astrom K. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. CVPR 1997: 438–40.
[5] Bougnoux S. From projective to Euclidean space under any practical situation, a criticism of self-calibration. Bombay: ICCV; 1998.
[6] Luhmann T. Close range photogrammetry. 2000.
[7] Feng W. Close range photogrammetry. Wuhan University Press; 2002. 125–130.
[8] Fraser C. Digital camera self-calibration. Photogrammetry Remote Sensing 1997:149–59.
[9] Motta S. Robot calibration using a 3D vision-based measurement system with a single camera. Robot Comput Integrat Manuf 2001:487–97.
[10] Bouguet J. Camera calibration toolbox for Matlab 〈http://www.vision.caltech.edu/bouguetj/calib_doc/〉.
[11] Canny J. A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell 1986;8(6):679–714.
[12] Mai F. A hierarchical approach for fast and robust ellipse extraction. Pattern Recognition 2008;41(8):2512–24.
[13] Honkavaara E, Ahokas E. Geometric test field calibration of digital photogrammetric sensors. Photogrammetry Remote Sensing 2006;60(4):387–99.
[14] Zhu J, Li Y. Design and calibration of a single-camera-based stereo vision sensor. Opt Eng 2006;45(8).
[15] Grammatikopoulos L, Karras G. An automatic approach for camera calibration from vanishing points. Photogrammetry Remote Sensing 2007;62(6):64–76.
[16] Tian J, Ding Y, Peng X. Self-calibration of a fringe projection system using epipolar constraint. Opt Laser Technol 2008;40(7):538–44.
[17] Song L, Wang M. High precision camera calibration in vision measurement. Opt Laser Technol 2007;39(8):1413–420.