Journal of Zhejiang University SCIENCE A
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

New multi-camera calibration algorithm based on 1D objects*

Zi-jian ZHAO, Yun-cai LIU

(*Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200240, China*)

E-mail: zj_zhao@sjtu.edu.cn; whomliu@sjtu.edu.cn

Received Oct. 31, 2007; revision accepted Feb. 20, 2008; published online May 10, 2008

Abstract: A new calibration algorithm for multi-camera systems using 1D calibration objects is proposed. The algorithm integrates the rank-4 factorization with Zhang (2004)'s method. The intrinsic parameters as well as the extrinsic parameters are recovered by capturing with cameras the 1D object's rotations around a fixed point. The algorithm is based on factorization of the scaled measurement matrix, the projective depth of which is estimated in an analytical equation instead of a recursive form. For more than three points on a 1D object, the approach of our algorithm is to extend the scaled measurement matrix. The obtained parameters are finally refined through the maximum likelihood inference. Simulations and experiments with real images verify that the proposed technique achieves a good trade-off between the intrinsic and extrinsic camera parameters.

Key words: Multi-camera calibration, Homography, Factorization, Scaled measurement matrix, Projective depth
doi:10.1631/jzus.A071573 **Document code:** A **CLC number:** TP242.6+2

INTRODUCTION

Camera calibration (Faugeras, 1993; Hartley and Zisserman, 2003; Wong *et al.*, 2003; Cao and Foroosh, 2006) is an important and necessary issue in computer vision, especially in recovering the 3D structure of a scene. Classical camera calibration method is performed by capturing a 3D calibration object with a known Euclidean structure (Tsai, 1987). If the calibration object has a highly accurate geometrical structure, this type of technique yields the best results. However, it is an elaborate task for setting up a 3D reference object with high accuracy, especially for multi-camera calibration. To avoid such difficulties, plane-based camera calibration (Triggs, 1998; Sturm and Maybank, 1999; Zhang, 2000; Meng and Hu, 2003; Chen *et al*., 2004; Wu *et al*., 2004; Kim *et al*., 2005; Gurdjos *et al*., 2006a; 2006b) is becoming a hot research topic for its flexibility. Zhang (2000) uses a square pattern board as the calibration object and extracts the imaged circular points (Euclidean structure) of the plane through the world-to-image homography. Meng and Hu (2003) apply a calibration pattern that is made up of a circle and a set of lines through its center. Wu *et al.*(2004) describe the associated lines of two coplanar circles and give the quasi-affine invariance for camera calibration. Other researchers (Kim *et al*., 2005; Gurdjos *et al*., 2006a; 2006b) pay more attention to confocal conics (including concentric circles) which can be used to compute the Euclidean structure of the supporting plane. Although all plane-based camera calibration algorithms yield very good results in calibrating a single camera, they are used to simply repeat the calibration process independently for each camera when applied to multi-camera systems, thus the obtained calibration results may not be suitable for the whole system.

Zhang (2004) proposes a camera calibration algorithm using 1D objects (three collinear points), which is suitable for multi-camera calibration. However, his method focuses on single camera calibration and only determines the intrinsic camera parameters without computing the extrinsic camera parameters. In fact, the extrinsic camera parameters are very important for a multi-camera system. Re-

* Project supported by the National Natural Science Foundation of China (No. 60675017) and the National Basic Research Program of China (No. 2006CB303103)

cently, Zhang *et al*.(2007) propose a new camera calibration algorithm using images of spheres, which is also suitable for multi-camera calibration, but has the same shortcomings as Zhang (2004)'s algorithm. To find a trade-off between the intrinsic and extrinsic camera parameters in the multi-camera calibration, Svoboda *et al*.(2005) give a convenient multi-camera calibration algorithm, which computes the intrinsic and extrinsic camera parameters via the rank-4 factorization and the Euclidean stratification. However, this algorithm involves mass computation of fundamental matrices and epipoles for estimating the projective scales and needs some prior knowledge of the camera parameters.

We propose a practical multi-camera calibration algorithm of 1D calibration objects, which extends Zhang (2004)'s method and can solve the problems the above algorithms encounter. Our work is integrating the rank-4 factorization technique (Sturm and Triggs, 1996) with Zhang (2004)'s method into a factorization-based multi-camera calibration algorithm. According to the geometrical property of the 1D object, we can estimate the projective depths in a simple analytical form instead of a recursive form as described by Sturm and Triggs (1996).

This paper is organized as follows. Section 2 gives some preliminary information such as the camera model, the homography and the scaled measurement matrix. In Section 3, we introduce how to solve the multi-camera calibration problem with the 1D calibration object rotating around a fixed point. The results with both simulation and real image experiments are provided in Section 4. Section 5 covers the concluding remarks.

PRELIMINARIES

Camera model and homography

Let $\tilde{\boldsymbol{P}} = [X\ Y\ Z\ 1]^{\mathrm{T}}$ be the 3D homogeneous coordinates of a world point P, and $\tilde{\boldsymbol{p}} = [u\ v\ 1]^{\mathrm{T}}$ be the 2D homogeneous coordinates of its projection in the image plane. $\tilde{\boldsymbol{P}}$ and $\tilde{\boldsymbol{p}}$ are related by the following equation in the pinhole model:

$$z_{\tilde{p}}\tilde{\boldsymbol{p}} = \boldsymbol{K}\begin{bmatrix}\boldsymbol{R} & \boldsymbol{t}\end{bmatrix}\tilde{\boldsymbol{P}},\ \text{with}\ \boldsymbol{K} = \begin{bmatrix} f_u & s & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (1)$$

where $z_{\tilde{p}}$ is a scale factor (projective depth of P); $\boldsymbol{K}$ is the camera intrinsic matrix, with the focal lengths f_u,f_v, the principle point (u_0, v_0) and the skew factor s; [$\boldsymbol{R}$ $\boldsymbol{t}$] is the camera extrinsic matrix, i.e., the rotation and translation from the world frame to the camera frame; $\boldsymbol{H}$=$\boldsymbol{K}$[$\boldsymbol{R}$ $\boldsymbol{t}$] is referred to as the world-to-image homography.

Scaled measurement matrix (Svoboda *et al*., 2005)

Let us consider m cameras and n object points $\tilde{\boldsymbol{P}}^{(i)} = [X^{(i)}\ Y^{(i)}\ Z^{(i)}\ 1]^{\mathrm{T}}$, i=1, 2, …, n. According to the pinhole camera model Eq.(1), we have

$$z_{\tilde{\boldsymbol{p}}_j^{(i)}}\tilde{\boldsymbol{p}}_j^{(i)} = \boldsymbol{K}_j[\boldsymbol{R}_j\ \boldsymbol{t}_j]\tilde{\boldsymbol{P}}^{(i)} = \boldsymbol{H}_j\tilde{\boldsymbol{P}}^{(i)},\ j=1, 2, \ldots, m, \quad (2)$$

where $z_{\tilde{\boldsymbol{p}}_j^{(i)}} \in (0,+\infty)$. We put all the points and the projective depths into one matrix $\boldsymbol{W}_{\mathrm{s}}$,

$$\begin{aligned}\boldsymbol{W}_{\mathrm{s}} &= \begin{bmatrix} z_{\tilde{\boldsymbol{p}}_1^{(1)}}\tilde{\boldsymbol{p}}_1^{(1)} & \cdots & z_{\tilde{\boldsymbol{p}}_1^{(n)}}\tilde{\boldsymbol{p}}_1^{(n)} \\ \vdots & & \vdots \\ z_{\tilde{\boldsymbol{p}}_m^{(1)}}\tilde{\boldsymbol{p}}_m^{(1)} & \cdots & z_{\tilde{\boldsymbol{p}}_m^{(n)}}\tilde{\boldsymbol{p}}_m^{(n)} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{H}_1 \\ \vdots \\ \boldsymbol{H}_m \end{bmatrix}_{3m\times 4} \begin{bmatrix} \tilde{\boldsymbol{P}}^{(1)} & \cdots & \tilde{\boldsymbol{P}}^{(n)} \end{bmatrix}_{4\times n} = \boldsymbol{P}\boldsymbol{Q},\end{aligned} \quad (3)$$

where $\boldsymbol{P} = [\boldsymbol{H}_1^{\mathrm{T}}\ \ldots\ \boldsymbol{H}_m^{\mathrm{T}}]^{\mathrm{T}}$, $\boldsymbol{Q} = [\tilde{\boldsymbol{P}}^{(1)}\ \ldots\ \tilde{\boldsymbol{P}}^{(n)}]$. $\boldsymbol{W}_{\mathrm{s}}$ is called the scaled measurement matrix. If we collect enough noiseless image points and know all the projective depths, $\boldsymbol{W}_{\mathrm{s}}$ has the rank 4 and can be factored into $\boldsymbol{P}$ and $\boldsymbol{Q}$ (Sturm and Triggs, 1996). The factorization of Eq.(3) recovers the motion and the shape up to a 4×4 projective transformation $\boldsymbol{T}$,

$$\boldsymbol{W}_{\mathrm{s}} = \hat{\boldsymbol{P}}\hat{\boldsymbol{Q}} = \boldsymbol{P}\boldsymbol{T}\boldsymbol{T}^{-1}\boldsymbol{Q}, \quad (4)$$

where $\hat{\boldsymbol{P}} = \boldsymbol{P}\boldsymbol{T}$ and $\hat{\boldsymbol{Q}} = \boldsymbol{T}^{-1}\boldsymbol{Q}$.

MULTI-CAMERA CALIBRATION ALGORITHM

Wu *et al*.(2005) propose that the 1D object based calibration is also suitable for a 1D object undergoing a general planar motion. However, it is in-

convenient to perform such a planar motion in real applications. For this reason, we consider the condition that a 1D object rotates around a fixed point only, as shown in Fig.1. As for the 1D object, it is not necessary to consider the occlusion problems.

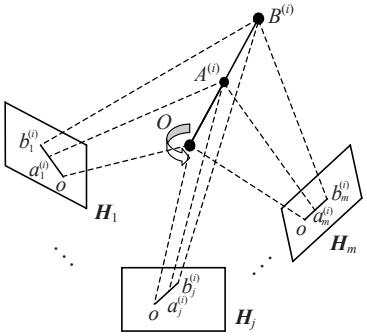

Fig.1 Multi-camera setup (*O* is the fixed point)

Constructing the scaled measurement matrix

Suppose there are m cameras and every camera captures n images of the 1D object. As shown in Fig.1, the image points of the points $B^{(i)}$, $A^{(i)}$, O on the 1D object are denoted by $b_j^{(i)}$, $a_j^{(i)}$, o_j (i=1, 2, …, n; j=1, 2, …, m), respectively. Because the three points $B^{(i)}$, $A^{(i)}$, O are collinear, the position of $A^{(i)}$ can be computed with respect to $B^{(i)}$ and O,

$$\tilde{\boldsymbol{A}}^{(i)} = \lambda_O \tilde{\boldsymbol{O}} + \lambda_B \tilde{\boldsymbol{B}}^{(i)}, \tag{5}$$

where λ_B and λ_O are known. According to Eqs.(1) and (5), we have

$$z_{\tilde{\boldsymbol{a}}_j^{(i)}} \tilde{\boldsymbol{a}}_j^{(i)} = \lambda_O z_{\tilde{\boldsymbol{o}}_j} \tilde{\boldsymbol{o}}_j + \lambda_B z_{\tilde{\boldsymbol{b}}_j^{(i)}} \tilde{\boldsymbol{b}}_j^{(i)}, \quad i = 1, 2, \cdots, n. \tag{6}$$

By performing cross product on both sides of Eq.(6) with $\tilde{\boldsymbol{b}}_j^{(i)}$ and $\tilde{\boldsymbol{a}}_j^{(i)}$ respectively, we can get

$$\begin{cases} z_{\tilde{\boldsymbol{a}}_j^{(i)}} (\tilde{\boldsymbol{a}}_j^{(i)} \times \tilde{\boldsymbol{b}}_j^{(i)}) - \lambda_O z_{\tilde{\boldsymbol{o}}_j} (\tilde{\boldsymbol{o}}_j \times \tilde{\boldsymbol{b}}_j^{(i)}) = 0, \\ \lambda_B z_{\tilde{\boldsymbol{b}}_j^{(i)}} (\tilde{\boldsymbol{b}}_j^{(i)} \times \tilde{\boldsymbol{a}}_j^{(i)}) + \lambda_O z_{\tilde{\boldsymbol{o}}_j} (\tilde{\boldsymbol{o}}_j \times \tilde{\boldsymbol{a}}_j^{(i)}) = 0. \end{cases} \tag{7}$$

In turn, we obtain

$$\begin{cases} z_{\tilde{\boldsymbol{a}}_j^{(i)}} = \dfrac{\lambda_O z_{\tilde{\boldsymbol{o}}_j} (\tilde{\boldsymbol{o}}_j \times \tilde{\boldsymbol{b}}_j^{(i)}) \cdot (\tilde{\boldsymbol{a}}_j^{(i)} \times \tilde{\boldsymbol{b}}_j^{(i)})}{(\tilde{\boldsymbol{a}}_j^{(i)} \times \tilde{\boldsymbol{b}}_j^{(i)}) \cdot (\tilde{\boldsymbol{a}}_j^{(i)} \times \tilde{\boldsymbol{b}}_j^{(i)})}, \\ z_{\tilde{\boldsymbol{b}}_j^{(i)}} = -\dfrac{\lambda_O z_{\tilde{\boldsymbol{o}}_j} (\tilde{\boldsymbol{o}}_j \times \tilde{\boldsymbol{a}}_j^{(i)}) \cdot (\tilde{\boldsymbol{b}}_j^{(i)} \times \tilde{\boldsymbol{a}}_j^{(i)})}{\lambda_B (\tilde{\boldsymbol{b}}_j^{(i)} \times \tilde{\boldsymbol{a}}_j^{(i)}) \cdot (\tilde{\boldsymbol{b}}_j^{(i)} \times \tilde{\boldsymbol{a}}_j^{(i)})}. \end{cases} \tag{8}$$

Introduce the following notation for the image points in the ith image captured by the jth camera:

$$\boldsymbol{x}_{ij} = [z_{\tilde{\boldsymbol{b}}_j^{(i)}} \tilde{\boldsymbol{b}}_j^{(i)} \quad z_{\tilde{\boldsymbol{a}}_j^{(i)}} \tilde{\boldsymbol{a}}_j^{(i)} \quad z_{\tilde{\boldsymbol{o}}_j} \tilde{\boldsymbol{o}}_j].$$

Substituting $z_{\tilde{\boldsymbol{b}}_j^{(i)}}$, $z_{\tilde{\boldsymbol{a}}_j^{(i)}}$, $z_{\tilde{\boldsymbol{o}}_j}$ by Eq.(8) gives

$$\boldsymbol{x}_{ij} = z_{\tilde{\boldsymbol{o}}_j} \boldsymbol{\Lambda}_{\mathrm{I}}^{ij} \boldsymbol{\Lambda}_{\mathrm{I}} [\tilde{\boldsymbol{b}}_j^{(i)} \quad \tilde{\boldsymbol{a}}_j^{(i)} \quad \tilde{\boldsymbol{o}}_j], \tag{9}$$

where

$$\boldsymbol{\Lambda}_{\mathrm{I}}^{ij} = \mathrm{diag}(m_2^{ij}, m_1^{ij}, 1), \quad \boldsymbol{\Lambda}_{\mathrm{I}} = \mathrm{diag}(-\lambda_O / \lambda_B, \lambda_O, 1),$$

$$m_1^{ij} = [(\tilde{\boldsymbol{o}}_j \times \tilde{\boldsymbol{b}}_j^{(i)}) \cdot (\tilde{\boldsymbol{a}}_j^{(i)} \times \tilde{\boldsymbol{b}}_j^{(i)})] \Big/ [(\tilde{\boldsymbol{a}}_j^{(i)} \times \tilde{\boldsymbol{b}}_j^{(i)}) \cdot (\tilde{\boldsymbol{a}}_j^{(i)} \times \tilde{\boldsymbol{b}}_j^{(i)})],$$

$$m_2^{ij} = [(\tilde{\boldsymbol{o}}_j \times \tilde{\boldsymbol{a}}_j^{(i)}) \cdot (\tilde{\boldsymbol{b}}_j^{(i)} \times \tilde{\boldsymbol{a}}_j^{(i)})] \Big/ [(\tilde{\boldsymbol{b}}_j^{(i)} \times \tilde{\boldsymbol{a}}_j^{(i)}) \cdot (\tilde{\boldsymbol{b}}_j^{(i)} \times \tilde{\boldsymbol{a}}_j^{(i)})].$$

We can put all $\boldsymbol{x}_{ij}$ (i=1, 2, …, n; j=1, 2, …, m) into one matrix $\boldsymbol{W}_{\mathrm{s}}$,

$$\boldsymbol{W}_{\mathrm{s}} = \begin{bmatrix} \boldsymbol{x}_{11} & \cdots & \boldsymbol{x}_{n1} \\ \vdots & & \vdots \\ \boldsymbol{x}_{1m} & \cdots & \boldsymbol{x}_{nm} \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{H}_1 \\ \vdots \\ \boldsymbol{H}_m \end{bmatrix}}_{\boldsymbol{P}} \cdot \underbrace{\begin{bmatrix} \tilde{\boldsymbol{B}}^{(1)} & \tilde{\boldsymbol{A}}^{(1)} & \tilde{\boldsymbol{O}} & \cdots & \tilde{\boldsymbol{B}}^{(n)} & \tilde{\boldsymbol{A}}^{(n)} & \tilde{\boldsymbol{O}} \end{bmatrix}}_{\boldsymbol{Q}}, \tag{10}$$

which is the scaled measurement matrix we construct for our calibration algorithm. The scales $z_{\tilde{\boldsymbol{o}}_j}$ (j=1, 2, …, m) are unknown, but we can set them to 1. To satisfy the rank-4 condition of $\boldsymbol{W}_{\mathrm{s}}$, m ($m \geq 2$) cameras are needed. The factorization (Sturm and Triggs, 1996) of $\boldsymbol{W}_{\mathrm{s}}$ recovers the motion $\hat{\boldsymbol{P}} = \boldsymbol{PT}$ and the shape $\hat{\boldsymbol{Q}} = \boldsymbol{T}^{-1}\boldsymbol{Q}$ up to a 4×4 projective transformation $\boldsymbol{T}$. The calibration process computes such a matrix $\boldsymbol{T}$ that makes $\hat{\boldsymbol{P}}$ and $\hat{\boldsymbol{Q}}$ Euclidean structures.

If there are more than three points lying on the 1D object, the sub-matrix $\boldsymbol{x}_{ij}$ can be expanded easily and the scaled measurement matrix $\boldsymbol{W}_{\mathrm{s}}$ may lead to obtaining more accurate calibration results because of the data redundancy in combating noise in image points. For example, if we have k (k>3) collinear points $p_{jk}^{(i)}$ (i=1, 2, …, n; j=1, 2, …, m), we can have the sub-matrix $\boldsymbol{x}_{ij}$ as

$$\boldsymbol{x}_{ij} = \begin{bmatrix} z_{\tilde{\boldsymbol{p}}_{j1}^{(i)}} \tilde{\boldsymbol{p}}_{j1}^{(i)} & \cdots & z_{\tilde{\boldsymbol{p}}_{jk}^{(i)}} \tilde{\boldsymbol{p}}_{jk}^{(i)} \end{bmatrix},$$

and then put all $\boldsymbol{x}_{ij}$ into one matrix $\boldsymbol{W}_s$ as Eq.(10). This completes the construction of the scaled measurement matrix.

Linear solution of the calibration algorithm

If we set $\hat{\boldsymbol{P}}=[\hat{\boldsymbol{H}}_1^{\mathrm{T}} \cdots \hat{\boldsymbol{H}}_m^{\mathrm{T}}]^{\mathrm{T}}$ and $\hat{\boldsymbol{Q}}=[\hat{\boldsymbol{B}}^{(1)}\ \hat{\boldsymbol{A}}^{(1)}\ \hat{\boldsymbol{O}}^{(1)} \cdots \hat{\boldsymbol{B}}^{(n)}\ \hat{\boldsymbol{A}}^{(n)}\ \hat{\boldsymbol{O}}^{(n)}]$, by Eq.(10), we can have

$$\begin{cases} \hat{\boldsymbol{H}}_j = \alpha_j \boldsymbol{H}_j \boldsymbol{T},\ \hat{\boldsymbol{B}}^{(i)} = \beta \boldsymbol{T}^{-1}\tilde{\boldsymbol{B}}^{(i)}, \\ \hat{\boldsymbol{A}}^{(i)} = \beta \boldsymbol{T}^{-1}\tilde{\boldsymbol{A}}^{(i)},\ \hat{\boldsymbol{O}}^{(i)} = \beta \boldsymbol{T}^{-1}\tilde{\boldsymbol{O}}, \end{cases} \tag{11}$$

where α_j (j=1, 2, …, m) and β are unknown scales. Any non-singular 4×4 matrix may be the matrix $\boldsymbol{T}$. Without loss of generality, we can fix both projective and Euclidean coordinate frames to the first camera. Then the transformation matrix $\boldsymbol{T}$ can be restricted to the following form:

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{K}_1^{-1}\boldsymbol{F}_0 \\ \boldsymbol{\pi}^{\mathrm{T}} \end{bmatrix}, \tag{12}$$

where $\boldsymbol{F}_0=[\boldsymbol{I}\ \boldsymbol{0}]$, $\boldsymbol{\pi}^{\mathrm{T}}$ represents the plane at infinity. The task of finding the appropriate $\boldsymbol{T}$ can be accomplished by imposing certain geometric constraints. The most useful constraint is the geometric structure of the calibration object. We have known that $\|\tilde{\boldsymbol{B}}^{(i)}-\tilde{\boldsymbol{O}}\|=l$ and $\|\tilde{\boldsymbol{A}}^{(i)}-\tilde{\boldsymbol{O}}\|=\lambda_B l$. According to Eq.(11), we obtain

$$(\hat{\boldsymbol{B}}^{(i)}-\hat{\boldsymbol{O}}^{(i)})^{\mathrm{T}}\boldsymbol{F}_0^{\mathrm{T}}\boldsymbol{\omega}_1\boldsymbol{F}_0(\hat{\boldsymbol{B}}^{(i)}-\hat{\boldsymbol{O}}^{(i)})=\beta^2 l^2, \tag{13}$$

$$(\hat{\boldsymbol{A}}^{(i)}-\hat{\boldsymbol{O}}^{(i)})^{\mathrm{T}}\boldsymbol{F}_0^{\mathrm{T}}\boldsymbol{\omega}_1\boldsymbol{F}_0(\hat{\boldsymbol{A}}^{(i)}-\hat{\boldsymbol{O}}^{(i)})=\beta^2\lambda_B^2 l^2, \tag{14}$$

where $\boldsymbol{\omega}_1=(\boldsymbol{K}_1^{-1})^{\mathrm{T}}\boldsymbol{K}_1^{-1}$, i=1, 2, …, n. Because Eq.(13) and Eq.(14) are correlative, we will have n equations for six unknowns of $\boldsymbol{\omega}_1$. Then n ($n\geq 6$) images are needed for every camera. After some manipulation, we can rewrite the constraints of Eqs.(13) and (14) into a set of linear equations and determine $\boldsymbol{\omega}_1$ in a least-squares sense. Once $\boldsymbol{\omega}_1$ is estimated, the matrix $\boldsymbol{K}_1$ can then be obtained by Cholesky factorization and matrix inversion. If the vanishing points of line $B^{(i)}A^{(i)}O$ (i=1, 2, ..., n) lie on a conic, we cannot solve $\boldsymbol{\omega}_1$ from Eqs.(13) and (14). That is the degenerate condition proposed by Hammarstedt *et al*.(2005).

Once $\boldsymbol{K}_1$ is determined, β can be solved according to Eqs.(13) and (14). Then we can get the following equation with respect to $\boldsymbol{\pi}^{\mathrm{T}}$:

$$\boldsymbol{\pi}^{\mathrm{T}}\hat{\boldsymbol{Q}}=\beta\begin{bmatrix}1 & \cdots & 1\end{bmatrix}_{1\times 3n}, \tag{15}$$

where $\boldsymbol{\pi}^{\mathrm{T}}$ is solved as a least-squares solution for the over-determined linear equation, and this completes the computing of $\boldsymbol{T}$. The shape matrix $\boldsymbol{Q}$ is recovered by $\boldsymbol{Q}=\boldsymbol{T}\hat{\boldsymbol{Q}}$. All homography matrices $\boldsymbol{H}_j=\boldsymbol{K}_j[\boldsymbol{R}_j\ \boldsymbol{t}_j]$ (j=1, 2, …, m) are recovered by $\boldsymbol{H}_j=\hat{\boldsymbol{H}}_j\boldsymbol{T}^{-1}$. The first 3×3 sub-matrix of $\boldsymbol{H}_j$ may be decomposed into the upper triangular calibration matrix $\boldsymbol{K}_j$ and the orthonormal matrix $\boldsymbol{R}_j$ by RQ matrix decomposition. Then the position vector $\boldsymbol{t}_j$ is computed according to Eq.(2).

Nonlinear optimization

Due to the existence of random noise, the above solution of camera parameters is not robust. Therefore, we refine it through the maximum likelihood inference. Generally, we assume that the image points are corrupted by independent and identically distributed Gaussian noise. Given n images of the calibration object for each camera, the maximum likelihood estimate can be obtained by minimizing the following function:

$$\begin{aligned} f=\sum_{j=1}^{m}\sum_{i=1}^{n}\Big(&\|\tilde{\boldsymbol{b}}_j^{(i)}-\tilde{\boldsymbol{b}}'(\boldsymbol{K}_j,\boldsymbol{R}_j,\boldsymbol{t}_j,\tilde{\boldsymbol{B}}^{(i)})\|^2 \\ &+\|\tilde{\boldsymbol{a}}_j^{(i)}-\tilde{\boldsymbol{a}}'(\boldsymbol{K}_j,\boldsymbol{R}_j,\boldsymbol{t}_j,\tilde{\boldsymbol{A}}^{(i)})\|^2 \\ &+\|\tilde{\boldsymbol{o}}_j-\tilde{\boldsymbol{o}}'(\boldsymbol{K}_j,\boldsymbol{R}_j,\boldsymbol{t}_j,\tilde{\boldsymbol{O}})\|^2\Big), \end{aligned} \tag{16}$$

where according to Eq.(2) $\tilde{\boldsymbol{b}}'(\boldsymbol{K}_j,\boldsymbol{R}_j,\boldsymbol{t}_j,\tilde{\boldsymbol{B}}^{(i)})$, $\tilde{\boldsymbol{a}}'(\boldsymbol{K}_j,\boldsymbol{R}_j,\boldsymbol{t}_j,\tilde{\boldsymbol{A}}^{(i)})$ and $\tilde{\boldsymbol{o}}'(\boldsymbol{K}_j,\boldsymbol{R}_j,\boldsymbol{t}_j,\tilde{\boldsymbol{O}})$ are the projections of the points $B^{(i)}$, $A^{(i)}$ and O, respectively. If we use the spherical coordinate (θ, ϕ) to parameterize $B^{(i)}$ and $A^{(i)}$, we have

$$\boldsymbol{B}^{(i)}=\boldsymbol{O}+l\begin{bmatrix}\sin\theta\cos\phi\\ \sin\theta\sin\phi\\ \cos\theta\end{bmatrix},\ \boldsymbol{A}^{(i)}=\boldsymbol{O}+\lambda_B l\begin{bmatrix}\sin\theta\cos\phi\\ \sin\theta\sin\phi\\ \cos\theta\end{bmatrix}. \tag{17}$$

We only need two additional parameters for each observation. Therefore, we have totally (14+2n)m unknowns (5m camera intrinsic parameters, 3m parameters for the coordinates of O, 6m camera extrinsic parameters, 2nm additional parameters to define the points $B^{(i)}$ and $A^{(i)}$). Minimizing Eq.(16) is an optimization problem. We can solve it using the Levenberg-Marquardt algorithm (Press *et al*., 1988). The initial guess of $\boldsymbol{K}_j$, $\boldsymbol{R}_j$, $\boldsymbol{t}_j$, $\tilde{\boldsymbol{B}}^{(i)}$, $\tilde{\boldsymbol{A}}^{(i)}$, $\tilde{\boldsymbol{O}}$ required can be obtained by using the technique described above.

EXPERIMENTS

The multi-camera calibration method described in previous sections has been implemented using simulated data and has been tested on real images.

Simulation results

In the simulation, six simulated cameras were used. The intrinsic parameters of the cameras were set with the common values: f_u=f_v=900, u_0=512, v_0=384, s=0.01. The calibration object (with three collinear points B, A and O) performed a series of rotations around the fixed point O and was simulated by the computer. The length of the simulated line-segment BAO was 60 mm, and λ_B=λ_O=0.5. Thirty synthetic images of BAO were captured with every camera. Independent Gaussian noise with 0 mean is added to the captured image points while varying its standard deviation σ at various levels.

We vary the noise level σ from 0.1 to 1 pixel. For each of the 10 noise levels, we perform 100 independent trials and estimate the average relative errors of intrinsic parameters. Figs.2a and 2b show the average relative errors of the linear solution and the nonlinear optimization solution for all six cameras in intrinsic parameters, respectively. From Fig.2, we can see that errors of each camera increase almost linearly with the noise level, and that the nonlinear optimization produces better results than the linear solution. We also compute the extrinsic parameters of all six cameras with final optimization at each noise level with 100 independent trials. After having the extrinsic parameters, the absolute errors of the norms of the rotation axes (see Rodrigue's formula) and the translation vectors are computed. Figs.3a and 3b respectively show the average rotation errors and average translation errors of the rest five cameras relative to the first camera with respect to the noise level σ. The average errors of both rotation and translation increase almost linearly with σ and are small when σ<0.5 pixel. If σ>0.5 pixel, the calibration results of extrinsic parameters will not be reliable for real applications.

Experimental results with real images

To show the validity of our algorithm for multi-camera calibration, we do a triple-camera calibration experiment (with two Point Grey FLEA cameras and a WATEC 902H camera). In the experiment, we use a 20 cm long stick with three markers. Fifty images are recorded for each camera. The sample images captured by three cameras are shown in Fig.4. We apply our calibration algorithm to these images and estimate the parameters of the three cameras with final optimization. The calibration results of the intrinsic parameters are shown in Table 1. For comparison, we also use the plane-based camera calibration algorithm (Zhang, 2000) to calibrate the three cameras, respectively. Nine images of a planar pattern are taken for each camera calibration. The calibration results are also shown in Table 1. After getting the extrinsic parameters of the three cameras, we do the pose estimation for all cameras. Figs.5a and 5b show the reconstructed camera poses according to the extrinsic parameters from our algorithm and the plane-based algorithm, respectively. In Fig.5, we do not see much difference between the reconstructed camera poses according to the extrinsic parameters from the two algorithms.

As can be seen from Table 1, the results of our calibration algorithm have some differences from those of the plane-based method. However, our algorithm can solve all camera parameters simultaneously instead of repeating the calibration process as the plane-based algorithm does. Therefore, our multi-camera calibration algorithm is more convenient and suitable for real multi-camera systems than the plane-based method.

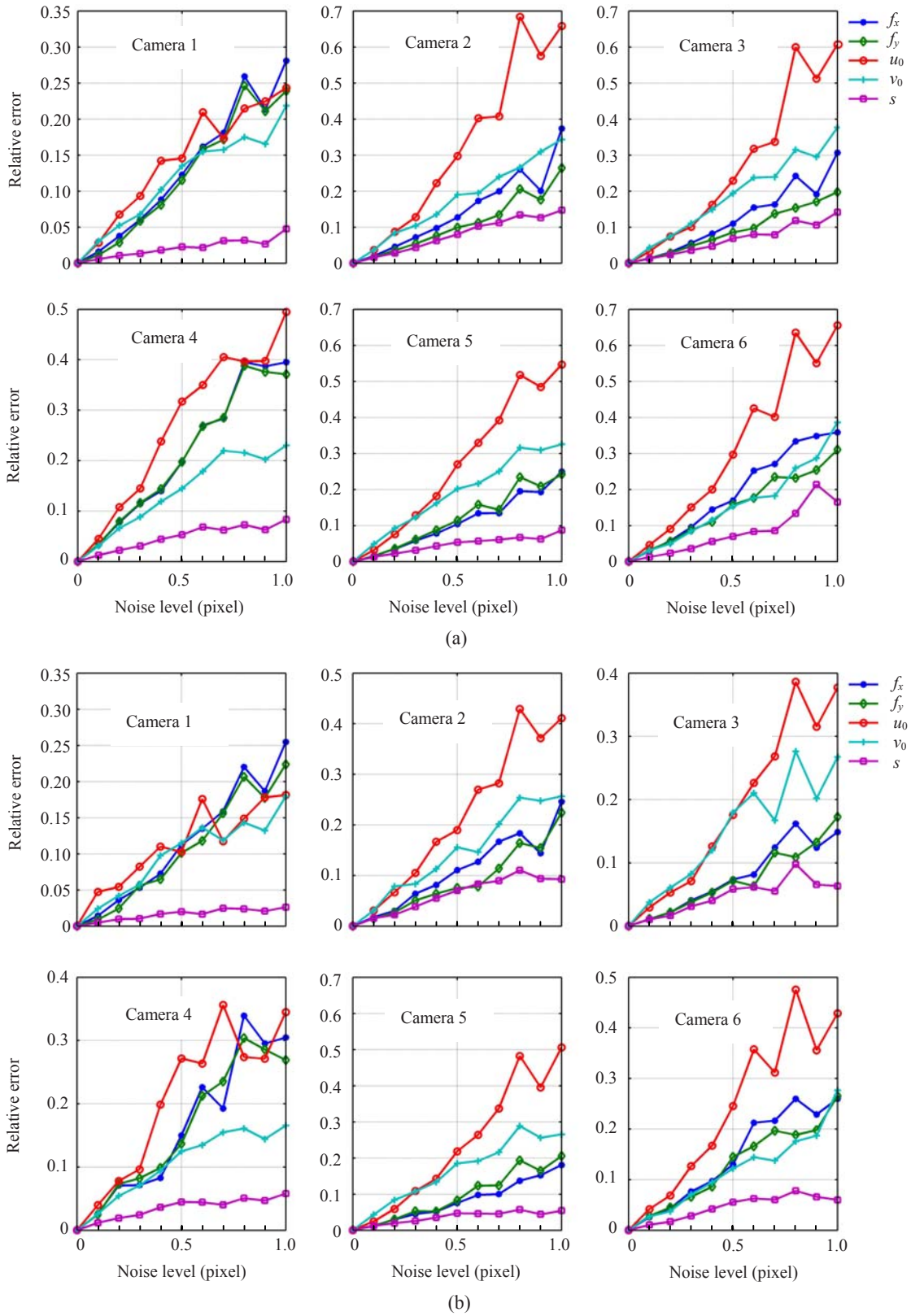

Fig.2 Average relative calibration errors of the linear and nonlinear optimization solutions for all six cameras in intrinsic parameters at 10 different noise levels for the focal lengths f_u, f_v, the principle point (u_0, v_0), and the skew factor s. For each noise level, we performed 100 independent trials. (a) Linear solution; (b) Nonlinear optimization

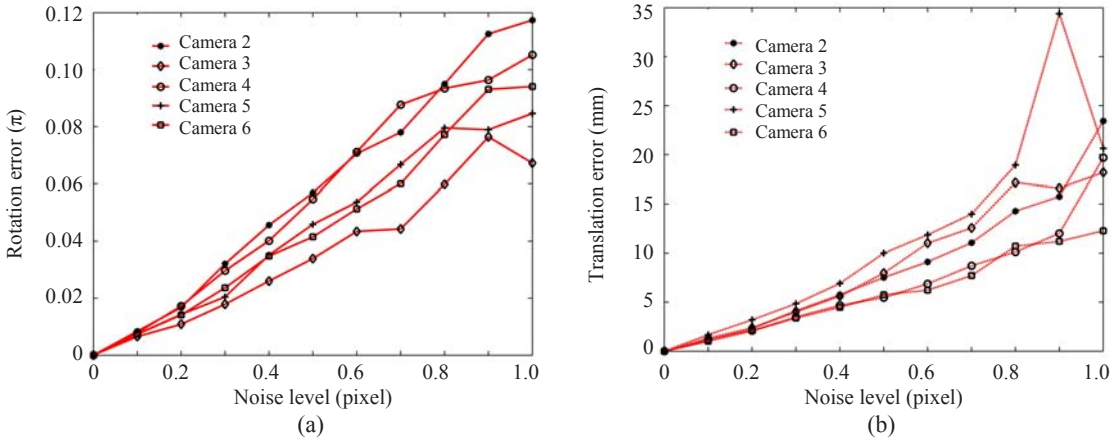

Fig.3 Average rotation errors (a) and average translation errors (b) of the rest five cameras relative to the first camera at 10 different noise levels. For each noise level, we performed 100 independent trials

Fig.4 Sample images of three cameras in the calibration. (a), (b) and (c) are sample images of Cameras 1, 2 and 3 respectively. Cameras 1 and 2 are Point Grey FLEA, and Camera 3 is WATEC 902H

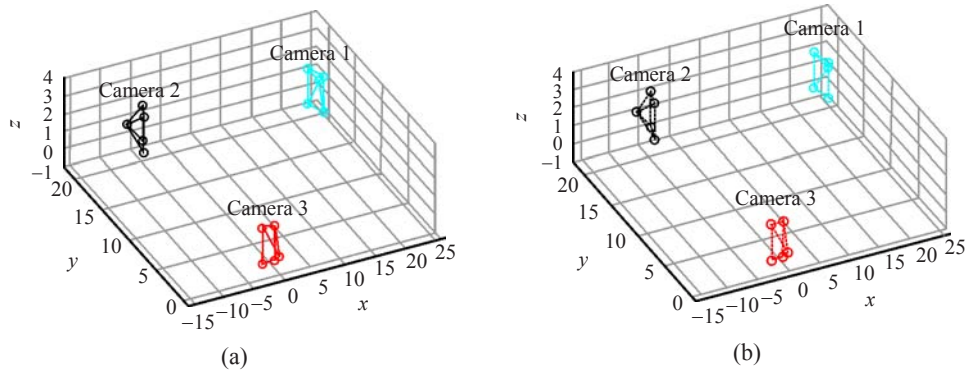

Fig.5 Reconstructed camera poses according to the extrinsic parameters from our algorithm (a) and from the plane-based algorithm (b) (unit: cm)

Table 1 Calibration results of our algorithm and the plane-based algorithm for three cameras in intrinsic parameters for the focal lengths f_u, f_v, the principle point (u_0, v_0), and the skew factor s

Camera No.*	Solution	f_u	f_v	u_0	v_0	s
1	Ours	1717.413	1791.421	596.013	348.056	−0.009
	Plane-based	1720.386	1716.052	547.439	396.746	0.011
2	Ours	1714.051	1702.126	538.511	435.802	0.016
	Plane-based	1719.499	1714.439	557.613	373.096	0.002
3	Ours	941.134	940.417	407.561	330.450	−0.150
	Plane-based	963.692	960.345	367.053	298.263	−0.102

* 1, 2: Point Grey FLEA; 3: WATEC 902H

CONCLUSION

In this paper we have proposed a practical multi-camera calibration algorithm based on 1D objects. This calibration algorithm integrates the rank-4 factorization with the standard 1D camera calibration method (Zhang, 2004). However, unlike the rank-4 factorization method (Svoboda *et al*., 2005), our algorithm does not need to compute the fundamental matrices and epipoles for estimating the projective depths and does not need prior knowledge of the camera parameters. Our algorithm gives an analytical equation for estimation of the projective depths according to the geometric structure of the 1D calibration object instead of a recursive form (Sturm and Triggs, 1996). The calibration algorithm proposed in this paper is also suitable for the 1D calibration object with more than three points when the scaled measurement matrix is expanded according to the current image points captured by all cameras.

Our calibration algorithm has been tested with both simulated data and real image experiments, and come out very encouraging results, which show that our algorithm achieves a good trade-off between the intrinsic and extrinsic camera parameters and is more convenient for real calibration of a multi-camera system than other algorithms.

References

Cao, X., Foroosh, H., 2006. Camera calibration using symmetric objects. *IEEE Trans. on Image Processing*, **15**(11): 3614-3619. [doi:10.1109/TIP.2006.881940]

Chen, Q., Wu, H., Wada, T., 2004. Camera Calibration with Two Arbitrary Coplanar Circles. European Conf. on Computer Vision. Prague, p.521-532.

Faugeras, O., 1993. Three-Dimensional Computer Vision: A Geometric Viewpoint. MIT Press, p.156-200.

Gurdjos, P., Sturm, P., Wu, Y., 2006a. Euclidean Structure Form *N*>=2 Parallel Circles: Theory and Algorithms. European Conf. on Computer Vision. Graz, p.238-252.

Gurdjos, P., Kim, J., Kweon, I., 2006b. Euclidean Structure from Confocal Conics: Theory and Application to Camera Calibration. Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. New York, p.1214-1221. [doi:10.1109/CVPR.2006.115]

Hammarstedt, P., Sturm, P., Heyden, A., 2005. Degenerate Cases and Closed-Form Solution for Camera Calibration with One-Dimensional Objects. Proc. 10th IEEE Int. Conf. on Computer Vision. Beijing, p.317-324. [doi:10.1109/ICCV.2005.68]

Hartley, R., Zisserman, A., 2003. Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge, p.60-300.

Kim, J., Gurdjos, P., Kweon, I., 2005. Geometric and algebraic constraints of projected concentric circles and their applications to camera calibration. *IEEE Trans. on PAMI*, **27**(4):637-642. [doi:10.1109/TPAMI.2005.80]

Meng, X., Hu, Z., 2003. A new easy camera calibration technique based on circular points. *Pattern Recognition*, **36**(5):1155-1164. [doi:10.1016/S0031-3203(02)00225-X]

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., 1988. Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press, p.33-60.

Sturm, P., Triggs, B., 1996. A Factorization Based Algorithm for Multi-Image Projective Structure and Motion. European Conf. on Computer Vision. Cambridge, p.709-720.

Sturm, P., Maybank, S., 1999. On Plane-Based Camera Calibration: A General Algorithm, Singularities, Applications. IEEE Conf. on Computer Vision and Pattern Recognition. Ft. Collins, p.1432-1437.

Svoboda, T., Martinec, D., Pajdla, T., 2005. A convenient multi-camera self-calibration for virtual environments. *Presence: Teleoperators and Virtual Environments*, **14**(4): 407-422. [doi:10.1162/105474605774785325]

Triggs, B., 1998. Auto-Calibration Form Planar Scenes. European Conf. on Computer Vision. Freiburg, p.89-105.

Tsai, R.Y., 1987. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf cameras and lens. *IEEE Robotics and Automation*, **3**(4):323-344.

Wong, K.Y.K., Mendonca, P.R.S., Cipolla, R., 2003. Camera calibration from surfaces of revolution. *IEEE Trans. on PAMI*, **25**(2):147-161. [doi:10.1109/TPAMI.2003.1177148]

Wu, F., Hu, Z., Zhu, H., 2005. Camera calibration with moving one-dimensional objects. *Pattern Recognition*, **38**(5):755-765. [doi:10.1016/j.patcog.2004.11.005]

Wu, Y., Zhu, H., Hu, Z., Wu, F., 2004. Camera Calibration form Quasi-Affine Invariance of Two Parallel Circles. European Conf. on Computer Vision. Prague, p.190-202. [doi:10.1007/b97865]

Zhang, H., Wong, K.Y.K., Zhang, G., 2007. Camera calibration from images of spheres. *IEEE Trans. on PAMI*, **29**(3):499-502. [doi:10.1109/TPAMI.2007.45]

Zhang, Z., 2000. A flexible new technique for camera calibration. *IEEE Trans. on PAMI*, **22**(11):1330-1334. [doi:10.1109/34.888718]

Zhang, Z., 2004. Camera calibration with one-dimensional objects. *IEEE Trans. on PAMI*, **26**(7):892-899. [doi:10.1109/TPAMI.2004.21]