Computer Vision and Image Understanding 116 (2012) 715–729

Contents lists available at SciVerse ScienceDirect

Computer Vision and Image Understanding

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

Self-calibration of hybrid central catadioptric and perspective cameras ☆

Xiaoming Deng [a,*], Fuchao Wu [b], Yihong Wu [b], Fuqing Duan [c], Liang Chang [c], Hongan Wang [d]

[a] Institute of Software, Chinese Academy of Sciences, P.O. Box 8718, Beijing 100190, China
[b] National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences, P.O. Box 2728, Beijing 100190, China
[c] College of Information Science and Technology, Beijing Normal University, No. 19, XinJieKouWai Street, Beijing 100875, China
[d] State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, P.O. Box 8718, Beijing 100190, China

ARTICLE INFO

Article history:
Received 16 December 2010
Accepted 8 February 2012
Available online 22 February 2012

Keywords:
Central catadioptric camera
3D imaging
Radial distortion
Camera calibration

ABSTRACT

Hybrid central catadioptric and perspective cameras are desired in practice, because the hybrid camera system can capture large field of view as well as high-resolution images. However, the calibration of the system is challenging due to heavy distortions in catadioptric cameras. In addition, previous calibration methods are only suitable for the camera system consisting of perspective cameras and catadioptric cameras with only parabolic mirrors, in which priors about the intrinsic parameters of perspective cameras are required. In this work, we provide a new approach to handle the problems. We show that if the hybrid camera system consists of at least two central catadioptric and one perspective cameras, both the intrinsic and extrinsic parameters of the system can be calibrated linearly without priors about intrinsic parameters of the perspective cameras, and the supported central catadioptric cameras of our method can be more generic. In this work, an approximated polynomial model is derived and used for rectification of catadioptric image. Firstly, with the epipolar geometry between the perspective and rectified catadioptric images, the distortion parameters of the polynomial model can be estimated linearly. Then a new method is proposed to estimate the intrinsic parameters of a central catadioptric camera with the parameters in the polynomial model, and hence the catadioptric cameras can be calibrated. Finally, a linear self-calibration method for the hybrid system is given with the calibrated catadioptric cameras. The main advantage of our method is that it cannot only calibrate both the intrinsic and extrinsic parameters of the hybrid camera system, but also simplify a traditional nonlinear self-calibration of perspective cameras to a linear process. Experiments show that our proposed method is robust and reliable.

1. Introduction

Catadioptric cameras have the prominent property of wide view angle, and they have been used in 3D reconstruction [30,22,19,11], robot navigation [7,23], virtual reality [24] and visual surveillance [8,29,26]. However, catadioptric cameras typically capture low-resolution images, which cannot meet the requirements in various applications. Perspective cameras can capture high-resolution images, but they have limited view angles. To overcome such disadvantages, the hybrid of catadioptric and perspective cameras is desired [29,33,25,8]. For example, in visual surveillance, catadioptric cameras can be used to detect moving objects, and active pan-tilt-zoom cameras (a perspective camera with a pan-tilt-zoom unit) can view the detected objects in detail [27]. However, the calibration of the hybrid camera system is challenging due to the heavy distortion in catadioptric cameras. In this paper, we aim to self-calibrate hybrid central catadioptric and perspective camera system.

Representative studies on the calibration of hybrid cameras have appeared in recent years, and they can be summarized into two categories:

(1) Methods based on part of information about radial distortion, intrinsic parameters or scene information. Micusik et al. [22] proposed a 3D metric reconstruction method from uncalibrated omnidirectional image. In this work, the intrinsic parameters were obtained via epipolar constraint, which was solved as a Quadratic Eigenvalue Problem (QEP), and then this calibration information was used for metric reconstruction. Barreto and Danilidis [2] proposed a practical approach for calibrating multiple cameras with radial distortion. The approach used the Fitzgibbon's division model [13], and could simultaneously calibrate the projection matrices and radial distortion. Chen and Yang [9] calibrated a hybrid camera network, which consisted of catadioptric and perspective cameras. The coordinates of a few 3D control points were required. Sturm [29] proposed a pioneer work on

☆ This paper has been recommended for acceptance by Antonis Argyros.

* Corresponding author.
E-mail address: xiaoming@iscas.ac.cn (X. Deng).

1077-3142/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2012.02.003

multi-view relations between perspective cameras and paracatadioptric system with an affine camera, while the method required prior knowledge on intrinsic parameters of the perspective cameras. The hybrid camera relations presented in [29] were later extended to cameras with lens distortion due to the similarities between the para-catadioptric and division models by Barreto and Daniilidis [3], and the epipolar relations concerning hyper-catadioptric cameras were also presented. Svoboda et al. [31] proposed an effective self-calibration approach of perspective camera networks using point object. However, it was designed for perspective cameras not for omnidirectional cameras with heavy radial distortions. Wang et al. [37] proposed a flexible multi-camera calibration method using 1D objects undertaking general rigid motion, and the method was designed for multiple perspective cameras.

(2) Methods concerning only the extrinsic parameters of cameras [14,25]. Ramalingam et al. [25] proposed a generic structure-from-motion framework to reconstruct scenes by cameras of different types, where the intrinsic and distortion parameters of the cameras were precalibrated. Ying and Hu used a sphere to calibrate the extrinsic parameters of catadioptric cameras [36]. Bazin et al. [5] proposed an effective motion estimation by decoupling rotation and translation in calibrated catadioptric vision, which estimated the absolute attitude (roll and pitch angles) at each image without error accumulation. A recent study on hybrid camera systems is reported in [4], which combines an automatic point matching step with the epipolar geometry extraction and 3D reconstruction. For the case of multiple perspective cameras, Sturm and Triggs [28] proposed several robust factorization algorithms that solved for the shape and motion parameters under both affine and perspective camera models.

In this paper, we propose a method to calibrate hybrid central catadioptric and perspective cameras. An approximated polynomial model of the catadioptric camera is derived and used for rectification of catadioptric images, and then we use epipolar geometry to estimate the distortion parameters in the polynomial model. Thus, we can calibrate the catadioptric cameras easily by using the derived relationship between the intrinsic parameters and distortion parameters in the polynomial model of central catadioptric cameras. Then, a self-calibration method for the hybrid camera system is given. It is shown that if the hybrid camera system consists of at least two central catadioptric cameras and one perspective camera, both the intrinsic and extrinsic parameters of the system can be calibrated linearly. The primary advantage of using self-calibration approach for hybrid camera system lies in its flexibility to calibrate the hybrid camera system at any place if a sufficient number of image correspondences are available. It is particularly noteworthy that our self-calibration approach does not rely on specific scene structure such as colinearity of points or straightness of lines [17]. While it does require image correspondences between multiple cameras, this is a small price to pay for its flexibility of use.

Our main contributions can be summarized as follows:

(1) A new method to estimate the intrinsic parameters of a central catadioptric camera under the polynomial model is introduced. An approximated polynomial model for central catadioptric camera is derived, and applied to the calibration of central catadioptric camera. With the polynomial model, the intrinsic parameter calibration problem for catadioptric camera can be converted to an easy distortion parameter estimation problem. Once the distortion parameters in the polynomial model are obtained, the intrinsic parameters of catadioptric camera can be explicitly computed. In addition, a classic calibration method for hybrid camera system [29] is generalized, and our method supports central catadioptric cameras with more generic mirrors.

(2) A nine-point method to estimate the distortion parameters under the polynomial model is proposed. Radial distortion solvers with less correspondences are important for autocalibration, which decreases the sampling complexity of RANSAC during the automatic point matching process [16,18]. Previous work used Gröbner basis to solve the minimal problems, while the generator of Gröbner basis solvers is not a trivial task. We proposed a nine-point calibration method of radial distortion. Our method is not a minimal solver for the problem, while it has the merit of easy implementation. The method only needs to solve equations in two unknown variables (two second order polynomials), and avoids solving complex third order polynomials equation by the singularity of fundamental matrix in the minimal solver. In addition, it can be solved easily with the `solve` function in Symbolic Math Toolbox of Matlab.

(3) A linear method to self-calibrate perspective cameras in the hybrid camera system is proposed, which does not require prior knowledge on the intrinsic parameters of perspective cameras. However, traditional self-calibration methods of perspective cameras involve either unstable nonlinear process without prior knowledge on camera intrinsic parameters [16], or linear calibration process with prior knowledge on cameras (e.g. aspect ratio, skew factor or principal point) [20].

This paper is organized as follows. Preliminaries are introduced in Section 2. Section 3 elaborates on our calibration method for hybrid central catadioptric and perspective cameras. Section 4 reports simulated and real experiments. Finally, Section 5 concludes this paper.

2. Preliminary

In this paper, the sign $\approx$ between vectors denotes the equality up to a nonzero scale.

2.1. Division model

Fitzgibbon [13] proposed a radial division model $L_1(u,v,l) = 1 + \sum_{i=1}^{l} \kappa_i (u^2+v^2)^i$ for omnidirectional cameras, where κ_i ($i = 1, \ldots, l$) are the radial distortion parameters, (u, v) is the inhomogeneous coordination of a point on a distorted image relative to the radial distortion center. The homogenous coordinate of (u, v) after rectification can be computed with division operation by

$$\begin{pmatrix} u \\ v \\ L_1(u,v,l) \end{pmatrix} \approx \begin{pmatrix} \frac{u}{L_1(u,v,l)} \\ \frac{v}{L_1(u,v,l)} \\ 1 \end{pmatrix} \tag{1}$$

Therefore, this model is called division model, and it is widely used in the calibration of omnidirectional cameras [22,33,32,18].

2.2. Central catadioptric camera

Geyer and Daniilidis [15] proposed a generalized image formation model for central catadioptric camera (see Fig. 1). Under the view sphere coordination system $O - xyz$, a 3D point $\mathbf{X}$ is projected to a point $\mathbf{X}^s$ on a unit sphere at the viewpoint $\mathbf{O}$ by $\mathbf{X}^s = \frac{\mathbf{RX}+\mathbf{t}}{\|\mathbf{RX}+\mathbf{t}\|}$, then projected to a point $\mathbf{m} = (u, v, 1)^T$ on the image plane $\mathbf{\Pi}$ by a virtual pinhole camera through the perspective center $\mathbf{O}^c$. $\mathbf{\Pi}$ is perpendic-

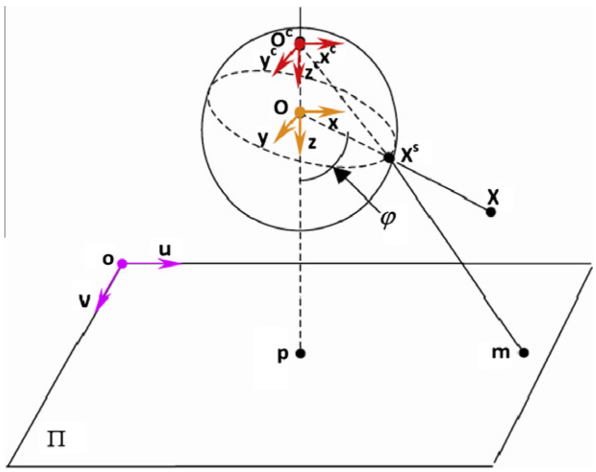

Fig. 1. Image formation of a central catadioptric camera.

ular to the line defined by the viewpoints $\mathbf{O}$ and $\mathbf{O}^c$. The angle φ is called the view angle of $\mathbf{m}$. The image formation process can be explicitly expressed as follows [39]:

$$\lambda\mathbf{m} = \mathbf{K}(\mathbf{X}^s + (0,0,\xi)^T) \tag{2}$$

$$\mathbf{K} = \begin{pmatrix} rf & s & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{X}^s = \frac{\mathbf{RX}+\mathbf{t}}{\|\mathbf{RX}+\mathbf{t}\|}, \quad \lambda = \frac{\xi + \sqrt{\xi^2 - \mathbf{m}^T\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{m}(\xi^2-1)}}{\mathbf{m}^T\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{m}} \tag{3}$$

where $(\mathbf{R},\mathbf{t})$, called the extrinsic parameters, is the rotation and translation which relates the world coordinate system to the view sphere coordination system $O-xyz$, and $\mathbf{K}$ is the camera intrinsic matrix, with f the focal length, r the aspect ratio, s the skew factor and $\mathbf{p} = (u_0, v_0, 1)^T$ is defined as the homogenous coordination of the principal point. ξ is usually called as the mirror parameter, which is the distance from $\mathbf{O}$ to $\mathbf{O}^c$. If $\xi = 1$, the used mirror is a paraboloid (i.e. the camera is paracatadioptric). If $0 < \xi < 1$, the mirror is an ellipsoid or a hyperboloid (i.e. the camera is hypercatadioptric). Since the calibration of a catadioptric camera with $\xi = 0$ is the same with that of a pinhole camera, in this work we only concentrate on the calibration of central catadioptric camera with $0 < \xi \leqslant 1$.

For the revolution conic section mirror, the mirror parameter ξ satisfies [35,1]

$$\xi = \frac{2\varepsilon}{1+\varepsilon^2} \tag{4}$$

where ε is the eccentricity of the conic. The relationship between eccentricity ε and the mirror parameter ξ for different types of central catadioptric cameras is shown in Table 1 [35]. Generally speaking, the mirror parameter ξ can be computed easily with the provided eccentricity ε from manufactures [1,15,12]. Therefore, ξ is assumed to be known in this paper.

For a central catadioptric camera, the principal point can be easily calibrated using the center of the bounding ellipse [13,35,12] or line images [32,39,38], and then the origin of the image can be translated to $\mathbf{p}$ by a linear transformation:

$$\mathbf{T}_p = \begin{pmatrix} 1 & 0 & -u_0 \\ 0 & 1 & -v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{5}$$

Hence, the image coordination $\mathbf{m}$ can be translated to $\tilde{\mathbf{m}}$ by $\tilde{\mathbf{m}} = \mathbf{T}_p\mathbf{m} = (u-u_0, v-v_0, 1)^T$, and $\mathbf{p}$ is translated to $\tilde{\mathbf{p}} = (0,0,1)^T$.

Denote $\widetilde{\mathbf{K}}$ as

$$\widetilde{\mathbf{K}} = \begin{pmatrix} rf & s & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{6}$$

The image formation process (2) can be described by

$$\lambda\tilde{\mathbf{m}} = \widetilde{\mathbf{K}}(\mathbf{X}^s + (0,0,\xi)^T) \tag{7}$$

The matrix $\widetilde{\mathbf{K}}^{-T}\widetilde{\mathbf{K}}^{-1}$ is in the following form:

$$\begin{pmatrix} k_1 & k_2/2 & 0 \\ k_2/2 & k_3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{8}$$

where

$$k_1 = 1/(r^2f^2), \quad k_2 = -2s/(r^2f^3), \quad k_3 = (r^2f^2+s^2)/(r^2f^4) \tag{9}$$

In this paper, we assume the principal point of catadioptric camera is precalibrated using the center of the bounding ellipse like in [39]. Hereafter, denote $u \triangleq u-u_0,\ v \triangleq v-v_0,\ \mathbf{m} \triangleq \tilde{\mathbf{m}}$ and $\mathbf{p} \triangleq \tilde{\mathbf{p}}$ for simplicity.

2.3. Projective reconstruction by factorization

Suppose there are N perspective cameras $\mathbf{P}_i$, $i = 1,\ldots,N$ and M 3D points $\mathbf{X}_j = (X_j, Y_j, Z_j, 1)^T$, $j = 1,\ldots,M$. The image coordinates are represented by $\mathbf{m}_{ij} = (u_{ij}, v_{ij}, 1)^T$. The image formation process can be described as follows:

$$\lambda_{ij}\mathbf{m}_{ij} = \mathbf{P}_i\mathbf{X}_j \tag{10}$$

where λ_{ij} is a non-zero scale factor, commonly called as the projective depth.

We stack Eq. (10) of all the cameras into a matrix $\mathbf{W}_{3N\times M}$, which can be factorized as follows:

$$\underbrace{\begin{pmatrix} \lambda_{11}\mathbf{m}_{11} & \lambda_{12}\mathbf{m}_{12} & \ldots & \lambda_{1M}\mathbf{m}_{1M} \\ \lambda_{21}\mathbf{m}_{21} & \lambda_{22}\mathbf{m}_{22} & \ldots & \lambda_{2M}\mathbf{m}_{2M} \\ \ldots & \ldots & \ldots & \ldots \\ \lambda_{N1}\mathbf{m}_{N1} & \lambda_{N2}\mathbf{m}_{N2} & \ldots & \lambda_{NM}\mathbf{m}_{NM} \end{pmatrix}}_{\mathbf{W}_{3N\times M}} = \underbrace{\begin{pmatrix} \mathbf{P}_1 \\ \ldots \\ \mathbf{P}_N \end{pmatrix}}_{\mathbf{M}_{3N\times 4}} \underbrace{(\mathbf{X}_1,\ldots,\mathbf{X}_M)}_{\mathbf{S}_{4\times M}} \tag{11}$$

$$\mathbf{W}_{3N\times M} = \mathbf{M}_{3N\times 4}\mathbf{S}_{4\times M} \tag{12}$$

where $\mathbf{W}$ is the scaled measurement matrix, $\mathbf{M}_{3N\times 4} = \left(\mathbf{P}_1^T, \mathbf{P}_2^T, \ldots, \mathbf{P}_N^T\right)^T$ is the projective matrix, and $\mathbf{S}_{4\times M} = (\mathbf{X}_1, \ldots, \mathbf{X}_M)$ is the projective shape matrix. We use Sturm and Triggs' method [28] for the computation of $\{\lambda_{ij}\}$, and recovery of projective reconstruction $\{\mathbf{P}_i\}_{i=1}^N$ by factorizing $\mathbf{W}$. Obviously, the factorization in (11) is a projective reconstruction up to a 4×4 homography matrix $\mathbf{H}_{4\times 4}$. That is, $\mathbf{W}$ can also be factorized as $\mathbf{W} = (\mathbf{M}_{3N\times 4}\mathbf{H}_{4\times 4})\left(\mathbf{H}_{4\times 4}^{-1}\mathbf{S}_{4\times M}\right)$. In order to get the Euclidean reconstruction, metric constraint should be used to recover the 4×4 homography matrix $\mathbf{H}$ [16].

3. Calibration method

In this paper, we assume the principal point of catadioptric camera is precalibrated using the center of the bounding ellipse

Table 1
The relationship between eccentricity ε and mirror parameter ξ.

	Ellipsoidal	Paraboloidal	Hyperboloidal	Planar
ε	$0 < \varepsilon < 1$	$\varepsilon = 1$	$\varepsilon > 1$	$\varepsilon \to \infty$
ξ	$0 < \xi < 1$	$\xi = 1$	$0 < \xi < 1$	$\xi = 0$

like in [35,39], and the origin of image coordination is translated to the principal point (see Section 2.1) i.e. $\mathbf{p} = (0,0,1)^T$.

3.1. Polynomial approximations for catadioptric camera models

In this section, we derive polynomial approximation model for central catadioptric camera, which will be applied to the calibration of central catadioptric camera in Section 3.2.

With Eq. (7), the following equation can be derived:

$$\underbrace{\mathbf{m}-\frac{\xi}{\lambda}\mathbf{p}}_{\mathbf{m}_u}=\begin{pmatrix} u \\ v \\ 1-\frac{\xi\mathbf{m}^T\widetilde{\mathbf{K}}^{-T}\widetilde{\mathbf{K}}^{-1}\mathbf{m}}{\xi+\sqrt{\xi^2+(1-\xi^2)\mathbf{m}^T\widetilde{\mathbf{K}}^{-T}\widetilde{\mathbf{K}}^{-1}\mathbf{m}}}\end{pmatrix}\approx\widetilde{\mathbf{K}}(\mathbf{R},\mathbf{t})\begin{pmatrix}\mathbf{X}\\1\end{pmatrix} \quad (13)$$

Comparing Eq. (13) with the imaging model of perspective camera (see Ref. [16]), we find that $\mathbf{m}_u = \mathbf{m} - \frac{\xi}{\lambda}\mathbf{p}$ is a rectification for $\mathbf{m} = (u,v,1)^T$ on the catadioptric image, which is the image of the space point $\mathbf{X}$ (or $\mathbf{X}^s$) under a pinhole camera with the optical center at the view sphere center $\mathbf{O}$.

Let $\mu = \mathbf{m}^T\widetilde{\mathbf{K}}^{-T}\widetilde{\mathbf{K}}^{-1}\mathbf{m} - 1 = k_1u^2 + k_2uv + k_3v^2$. Then the third component in $\mathbf{m}_u$ can be represented by μ as

$$f(\mu)=\frac{1-\xi\sqrt{1+(1-\xi^2)\mu}}{1-\xi^2} \quad (14)$$

Therefore, $\mathbf{m}_u$ in Eq. (13) can be simplified as

$$\mathbf{m}_u=\begin{pmatrix} u\\ v\\ f(\mu)\end{pmatrix} \quad (15)$$

By Taylor expansion of $f(\mu)$ at $\mu = 0$, we have

$$f(\mu)=\frac{1}{1+\xi}+\xi\sum_{n=1}^{\infty}\frac{(-1)^{n-1}(1-\xi^2)^{n-1}}{2^n n!}\prod_{i=1}^{n}(2i-3)\mu^n,\left(\mu\in\left[0,\frac{1}{1-\xi^2}\right)\right) \quad (16)$$

Multiply $(1+\xi)$ to $\mathbf{m}_u$ in Eq. (15), replace μ with $k_1u^2 + k_2uv + k_3v^2$ in $(1+\xi)f(\mu)$, and omit items of μ with more than l orders, then we can obtain an lth order series as follows:

$$\widetilde{L}(u,v,l)=1+\sum_{p=1}^{l}\sum_{\substack{i+j=2p,\\ i,j\geqslant 0}}b_{ij}u^iv^j \quad (17)$$

where

$$b_{ij}=(-1)^{p-1}\xi(1+\xi)(1-\xi^2)^{p-1}a_p\sum_{\substack{m,n\geqslant 0,2m+n=i,\\(m+n)\leqslant p}}\frac{p!k_1^mk_2^nk_3^{p-m-n}}{m!n!(p-m-n)!},$$
$$(i+j=2p) \quad (18)$$

k_1, k_2, k_3 in b_{ij} are defined in Eq. (9), and $a_p=\prod_{i=1}^{p}(2i-3)/(2^pp!)$.

The image rectification process in (15) is equivalent to the following:

$$\mathbf{m}_u(\mathbf{c})=((1+\xi)u,(1+\xi)v,\widetilde{L}(u,v,l))^T,(l\to\infty). \quad (19)$$

Hence, we have the following proposition, by which the intrinsic parameters of a central catadioptric camera can be computed with the distortion parameters b_{20}, b_{11}, b_{02} of the model $\widetilde{L}(u,v,l)$.

Proposition 1. *For a central catadioptric camera with the principal point $\mathbf{p} = (0,0,1)^T$ and mirror parameter $\xi (0 < \xi \leqslant 1)$, the parameters in (8) can be computed by $k_1=-\frac{2b_{20}}{\xi(\xi+1)}$, $k_2=-\frac{2b_{11}}{\xi(\xi+1)}$, and $k_3=-\frac{2b_{02}}{\xi(\xi+1)}$. Thus, the intrinsic matrix can be computed as follows:*

$$\widetilde{\mathbf{K}}=\begin{pmatrix} rf & s & 0\\ 0 & f & 0\\ 0 & 0 & 1\end{pmatrix} \quad (20)$$

where $f=2\sqrt{\frac{k_1}{4k_1k_3-k_2^2}}$, $r=\frac{\sqrt{4k_1k_3-k_2^2}}{2k_1}$, $s=-\frac{k_2}{\sqrt{k_1(4k_1k_3-k_2^2)}}$.

Proof. With Eqs. (17) and (18), we have $b_{20}=-\frac{\xi(\xi+1)}{2}k_1$, $b_{11}=-\frac{\xi(\xi+1)}{2}k_2$, $b_{02}=-\frac{\xi(\xi+1)}{2}k_3$, and then k_1, k_2, k_3 can be expressed by ξ, b_{20}, b_{11}, b_{02} as in the above.

With Eqs. (8) and (9), $\widetilde{\mathbf{K}}^{-T}\widetilde{\mathbf{K}}^{-1}$ can be computed by

$$\begin{pmatrix} k_1 & k_2/2 & 0\\ k_2/2 & k_3 & 0\\ 0 & 0 & 1\end{pmatrix}=\begin{pmatrix}\frac{1}{f^2r^2} & -\frac{s}{f^3r^2} & 0\\ -\frac{s}{f^3r^2} & \frac{s^2}{r^2f^4}+\frac{1}{f^2} & 0\\ 0 & 0 & 1\end{pmatrix} \quad (21)$$

And then, we have

$$k_1=\frac{1}{f^2r^2} \quad (22)$$
$$\frac{k_2}{2}=-\frac{s}{f^3r^2} \quad (23)$$
$$k_3=\frac{s^2}{r^2f^4}+\frac{1}{f^2} \quad (24)$$

With Eqs. (22) and (23), we have

$$r^2=\frac{1}{k_1f^2} \quad (25)$$
$$s=-\frac{k_2}{2k_1}f \quad (26)$$

By substituting (25) and (26) into (24), we have

$$k_3=\frac{k_2^2}{4k_1}+\frac{1}{f^2} \quad (27)$$

and then the focal length f can be recovered by

$$f=2\sqrt{\frac{k_1}{4k_1k_3-k_2^2}} \quad (28)$$

Finally, r, s can be recovered with the obtained f by Eqs. (25) and (26). □

Proposition 1 is important for our proposed calibration method of central catadioptric camera, which will be discussed in Section 3.2.

The series $\widetilde{L}(u,v,l)$ (see Eq. (17)) does not assume prior knowledge about the aspect ratio r, skew factor s and mirror parameters ξ. Inspired by the formation of $\widetilde{L}(u,v,l)$, we generalize the division model (see Eq. (1)) into a more generic polynomial model as follows:

$$L(u,v,l)=1+\sum_{p=1}^{l}\sum_{\substack{i+j=2p\\ i,j\geqslant 0}}c_{ij}u^iv^j \quad (29)$$

Hereafter, we call $L(u,v,l)$ as the lth-order polynomial division model for central catadioptric cameras, and $\mathbf{c} = (c_{ij})$ are the distortion parameters to be calibrated.

Inspired by the image rectification process (19), the image point $\mathbf{m} = (u,v,1)^T$ can be rectified with the polynomial division model $L(u,v,l)$ as follows:

$$\hat{\mathbf{m}}_u(\mathbf{c}) = ((1+\xi)u, (1+\xi)v, L(u,v,l))^T \tag{30}$$

For a central paracatadioptric camera ($\xi = 1.0$), the algebraic formation of the 1st-order polynomial division model $L(u, v, 1)$ is the same as that of the true image rectification model (17) with $l \to \infty$, because the parameter $b_{ij}(i+j>2)$ for $l>1$ is zero with $\xi = 1.0$ (see Eq. (18)). Therefore, the 1st-order polynomial division model is suitable for image rectification of the central paracatadioptric camera during camera calibration; for hypercatadioptric cameras ($0<\xi<1.0$), the 1st-order polynomial division model is a close approximation to the true image rectification model in practice (see experiments in Section 4, with the 1st-order polynomial model, fairly good results of intrinsic parameters in the catadioptric cameras can be estimated.). Therefore, we use the 1st-order model to rectify the catadioptric images.

In this paper, we use the polynomial model $L(u, v, l)$ to approximate the true model $\widetilde{L}(u,v,l)$ of central catadioptric camera (see Eq. (17)), and the corresponding parameters in the two models are shown close in lots of our calibration experiments. With the parameters $\{b_{ij}\}$, the ground truth intrinsic parameters of central catadioptric cameras can be computed with Proposition 1. In the experiments (see Section 4.1), the calibrated parameter c_{ij} in $L(u, v, l)$ is used to replace the parameter b_{ij} in Proposition 1, and the estimated intrinsic parameters of the catadioptric cameras are very close to the ground truth. Therefore, the parameter c_{ij} in $L(u, v, l)$ by our calibration method is close to the true coefficient b_{ij} in $\widetilde{L}_l(u,v,l)$ for central catadioptric cameras.

3.2. Intrinsic parameter calibration for central catadioptric cameras

In this section, we use epipolar geometry between the perspective and rectified central catadioptric images to estimate the distortion parameters in the polynomial model, and then estimate the intrinsic parameters of catadioptric camera by Proposition 1.

Since the eccentricity of the mirror in the catadioptric camera is usually known, and the mirror parameter ξ can be obtained from the eccentricity of the mirror by Eq. (4), the mirror parameter ξ in our work is assumed to be known like in [15]. In this paper, the skew factor s of central catadioptric cameras is assumed to be zero.

Epipolar geometry between the perspective and rectified central catadioptric images is

$$\mathbf{m}_1^T \mathbf{F} \hat{\mathbf{m}}_{u2}(\mathbf{c}) = 0 \tag{31}$$

where $\mathbf{m}_i = (u_i, v_i, 1)^T$ ($i = 1,2$) are the image correspondences on the perspective and catadioptric images respectively, $\mathbf{c} = (c_{ij})$ is the distortion parameters in the polynomial division model, $\hat{\mathbf{m}}_{u2}(\mathbf{c})$ is a rectification of $\mathbf{m}_2$ with (30) and $\mathbf{c}$, $\mathbf{F} = (f_{ij})_{3\times3}$ is the fundamental matrix. Assuming $f_{33} \neq 0$, and we set $f_{33} = 1$ like in [18].

Next, we will use the constraint (31) of $\mathbf{c}$ and $\mathbf{F} = (f_{ij})_{3\times3}$ to calibrate the distortion parameters in the polynomial division model.

Since the central catadioptric camera is skew-free (i.e. $s = 0$), we can get $b_{11} = 0$ (see Proposition 1) and the distortion parameter c_{11} can be set to be zero (i.e. $c_{ij} = 0$). Therefore, we have

$$\mathbf{D}_1 \mathbf{g} = 0 \tag{32}$$

where

$$\mathbf{D}_1 = (u_1u_2^2, u_1v_2^2, u_1u_2, u_1v_2, u_1, v_1u_2^2, v_1v_2^2, v_1u_2, v_1v_2, v_1, u_2^2, v_2^2, u_2, v_2, 1) \tag{33}$$

$$\mathbf{g} = (c_{20}f_{13}, c_{02}f_{13}, (1+\xi)f_{11}, (1+\xi)f_{12}, f_{13}, c_{20}f_{23}, c_{02}f_{23}, (1+\xi)f_{21}, (1+\xi)f_{22}, f_{23}, c_{20}, c_{02}, (1+\xi)f_{31}, (1+\xi)f_{32}, 1)^T \tag{34}$$

where $\mathbf{g}$ is denoted as the vector $\mathbf{g} = (g_1, g_2, \ldots, g_{15})^T = (c_{20}f_{13}, c_{02}f_{13}, (1+\xi)f_{11}, (1+\xi)f_{12}, f_{13}, c_{20}f_{23}, c_{02}f_{23}, (1+\xi)f_{21}, (1+\xi)f_{22}, f_{23}, c_{20}, c_{02}, (1+\xi)f_{31}, (1+\xi)f_{32}, 1)^T$.

The epipolar constraints give one equation with 15 monomials ($c_{20}f_{13}$, $c_{02}f_{13}$, $(1+\xi)f_{11}$, $(1+\xi)f_{12}$, f_{13}, $c_{20}f_{23}$, $c_{02}f_{23}$, $(1+\xi)f_{21}$, $(1+\xi)f_{22}$, f_{23}, c_{20}, c_{02}, $(1+\xi)f_{31}$, $(1+\xi)f_{32}$, 1) and 11 variables (f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{23}, f_{31}, f_{32}, c_{20}, c_{02}).

For a pair of correspondences, we can establish a set of homogeneous linear equations by piling up the vector $\mathbf{D}_1$. If we regard the 15 monomials in $\mathbf{g}$ as independent variables, then at least 14 correspondences are required for calibration. In the following, we use singular value decomposition (SVD) to compute c_{20}, c_{02} and $\mathbf{F} = (f_{ij})_{3\times3}$. It should be noted that the coordinates of image correspondences should be normalized like in [16]. In Appendix A, we derive a normalization approach for the distortion parameter estimation problem of the hybrid camera system, which is inspired by the method in [4].

Denote $\{g_i\}_{i=1}^{15}$ to be the ith item in the monomials $\mathbf{g}$, $Cand(c_{20})$,-$Cand(c_{02})$ to be the sets with candidate solutions of c_{20}, c_{02} respectively. We can use SVD to estimate $\{g_i\}_{i=1}^{15}$. $Cand(c_{20})$ can be computed by

$$Cand(c_{20}) = \left\{\frac{g_1}{g_5}, \frac{g_6}{g_{10}}, \frac{g_{11}}{g_{15}}\right\} \tag{35}$$

and $Cand(c_{02})$ can be computed by

$$Cand(c_{02}) = \left\{\frac{g_2}{g_5}, \frac{g_7}{g_{10}}, \frac{g_{12}}{g_{15}}\right\} \tag{36}$$

From Eqs. (35) and (36), there are nine candidate solutions for the combination of c_{20} and c_{02}.

From Eqs. (22) and (24), we have $k_1 > 0$, $k_3 > 0$, and obtained $b_{20} = -\frac{\xi(\xi+1)}{2}k_1$ and $b_{02} = -\frac{\xi(\xi+1)}{2}k_3$ are both negative (see in Proposition 1). With the discussion in Section 3.1, we know that the parameter c_{ij} of $L(u, v, l)$ is close to the true coefficient b_{ij} of $\widetilde{L}_l(u,v,l)$ for the central catadioptric camera. Therefore, if there exist negative values in a candidate solution for c_{20}, c_{02}, then the candidate solution should be removed. In the rest candidate solutions, we choose a pair of c_{20}, c_{02} as the distortion parameter calibration result, which can minimize the following cost function:

$$cost(\mathbf{c}) = \sum_{q=1}^{M} \frac{\mathbf{m}_{1,q}^T \mathbf{F} \hat{\mathbf{m}}_{u2,q}(\mathbf{c})}{(\mathbf{F}\hat{\mathbf{m}}_{u2,q}(\mathbf{c}))_{(1)}^2 + (\mathbf{F}\hat{\mathbf{m}}_{u2,q}(\mathbf{c}))_{(2)}^2} \tag{37}$$

where $\mathbf{c} = (c_{20}, c_{02})$, M is the number of correspondences, and $(\mathbf{F}\hat{\mathbf{m}}_{u2,q}(\mathbf{c}))_{(i)}$ is the ith element of $\mathbf{F}\hat{\mathbf{m}}_{u2,q}(\mathbf{c})$. This cost function is an evaluation to the epipolar cost in the perspective camera, which is the sum square distance between the image point and the epipolar line in the perspective camera.

Hence, the intrinsic parameters of a central catadioptric camera can be estimated with the distortion parameters $\mathbf{c} = (c_{ij})$ of the polynomial model by Proposition 1. Our method is linear, because the epipolar geometry relationship (31) gives rise to linear constraints on the distortion parameters of the polynomial model. In sum, with the polynomial model, the intrinsic parameter calibration problem for catadioptric camera is converted to an easy distortion parameter estimation problem.

We would like to point out that the calibration method of central catadioptric camera becomes invalid, if the mirror parameter ξ is equal to 0.0. This is because the distortion parameters b_{20}, b_{11}, b_{02} in $L(u, v, l)$ (see Eq. (29)) will be zeros in the case of $\xi = 0$, and the intrinsic parameters with the distortion parameters cannot be computed with Proposition 1.

Remark 1. In Eq. (32), the distortion parameters are over-constrained, and distortion parameter estimation with the minimal number of correspondences is possible with minimal problem solvers such as [18]. It is well known that for standard uncalibrated case without considering distortion, seven point correspondences are sufficient to estimate the epipolar geometry. In Eq. (32), we have two more parameters, the distortion parameters $\{c_{20}, c_{02}\}$. Therefore, nine point correspondences are sufficient to estimate the unknown parameters $(f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{23}, f_{31}, f_{32}, c_{20}, c_{02})$. In Appendix B, we give a nine-point solver for the distortion parameter estimation problem.

3.3. Calibration of the hybrid camera system

In this section, we provide a simplified derivation on how to calibrate the hybrid camera system without measuring coordinates of 3D points. We aim to calibrate both intrinsic and extrinsic parameters of the hybrid camera system with calibrated catadioptric cameras, and we do not make assumptions on the intrinsic parameters of perspective cameras.

After the central catadioptric cameras are calibrated, the catadioptric images can be rectified to calibrated perspective images. The calibration of the hybrid camera system is converted to a special multi-perspective camera calibration problem, in which part of the perspective cameras (i.e. calibrated central catadioptric cameras) are calibrated, and the rest perspective cameras are uncalibrated.

Firstly, Sturm and Triggs' matrix factorization [28] is used for recovery of projective reconstruction $\{\mathbf{P}_i\}_{i=1}^N, \{\mathbf{X}_j\}_{j=1}^M$ with catadioptric and perspective cameras. For a central catadioptric camera, we use the rectified image points with Eq. (30) for projective reconstruction.

Secondly, we will get Euclidean reconstruction of the hybrid central catadioptric and perspective camera system.

Denote $\widehat{\mathbf{P}}_i = \mu_i \mathbf{K}_i(\mathbf{R}_i, \mathbf{t}_i)$ to be the Euclidean projection matrix of the ith camera, μ_i is a non-zero scale factor, and let the corresponding intrinsic parameters to be $\mathbf{K}_i$, the rotation matrix $\mathbf{R}_i$ and translation vector $\mathbf{t}_i$. The parameters $\mathbf{K}_i$, $\mathbf{R}_i$, $\mathbf{t}_i$ are denoted as follows:

$$\mathbf{K}_i = \begin{pmatrix} r_i f_i & s_i & u_{0i} \\ 0 & f_i & v_{0i} \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{R}_i = \begin{pmatrix} \mathbf{i}_i^T \\ \mathbf{j}_i^T \\ \mathbf{k}_i^T \end{pmatrix}, \quad \mathbf{t}_i = \begin{pmatrix} t_{xi} \\ t_{yi} \\ t_{zi} \end{pmatrix}$$

And thus, the Euclidean projection matrix $\widehat{\mathbf{P}}_i$ can be expressed by [31]

$$\widehat{\mathbf{P}}_i = \begin{pmatrix} \mathbf{m}_{xi} & T_{xi} \\ \mathbf{m}_{yi} & T_{yi} \\ \mathbf{m}_{zi} & T_{zi} \end{pmatrix} \tag{38}$$

where $\mathbf{m}_{xi} = \mu_i r_i f_i \mathbf{i}_i^T + \mu_i s_i \mathbf{j}_i^T + \mu_i u_{0i} \mathbf{k}_i$, $\mathbf{m}_{yi} = \mu_i f_i \mathbf{j}_i^T + \mu_i v_{0i} \mathbf{k}_i^T$, $\mathbf{m}_{zi} = \mu_i \mathbf{k}_i^T$, $T_{xi} = \mu_i r_i f_i t_{xi} + \mu_i s_i t_{yi} + \mu_i u_{0i} t_{zi}$, $T_{yi} = \mu_i f_i t_{yi} + \mu_i v_{0i} t_{zi}$ and $T_{zi} = \mu_i t_{zi}$.

Define the 4×4 projective transformation $\mathbf{H}$ as $\mathbf{H} = (\mathbf{A}_{4\times3}, \mathbf{B}_{4\times1})$, and denote $\mathbf{M}_{3N\times4} = \left(\mathbf{P}_1^T, \mathbf{P}_2^T, \ldots, \mathbf{P}_N^T\right)^T$ and $\widehat{\mathbf{M}}_{3N\times4} = \left(\widehat{\mathbf{P}}_1^T, \widehat{\mathbf{P}}_2^T \ldots, \widehat{\mathbf{P}}_N^T\right)^T$. Since $\mathbf{M}$ recovers the motion and structure up to a 4×4 linear projective transformation $\mathbf{H}$ (i.e. $\widehat{\mathbf{M}} = \mathbf{MH}$ and $\widehat{\mathbf{S}} = \mathbf{H}^{-1}\mathbf{S}$), we have [31,16]

$$\widehat{\mathbf{M}} = \mathbf{M}(\mathbf{A}_{4\times3}, \mathbf{B}_{4\times1}) \tag{39}$$

In order to upgrade the projective reconstruction $\{\mathbf{P}_i\}_{i=1}^N$, $\{\mathbf{X}_j\}_{j=1}^M$ to Euclidean reconstruction, the projective transformation $\mathbf{H} = (\mathbf{A}, \mathbf{B})$ should be recovered.

Firstly, The column vector $\mathbf{B}$ can be computed with linear least squares solutions of linear system by $\frac{T_{xi}}{T_{zi}}$ and $\frac{T_{yi}}{T_{zi}}$ [31].

Secondly, the matrix $\mathbf{A}$ can be linearly estimated as follows.

Multiplying the first three columns of $\widehat{\mathbf{M}}$ in (39) with its transposed matrix, we have

$$\widehat{\mathbf{M}}^{(1:3)}\widehat{\mathbf{M}}^{(1:3)T} = \mathbf{MAA}^T\mathbf{M}^T \tag{40}$$

where $\widehat{\mathbf{M}}_i^{(1:3)}$ is the first three column of $\widehat{\mathbf{M}}_i$.

Substituting $\mathbf{M} = \left(\mathbf{P}_1^T, \mathbf{P}_2^T, \ldots, \mathbf{P}_N^T\right)^T$ and $\widehat{\mathbf{M}} = \left(\widehat{\mathbf{P}}_1^T, \widehat{\mathbf{P}}_2^T, \ldots, \widehat{\mathbf{P}}_N^T\right)^T$ into (40), we have

$$\widehat{\mathbf{P}}_i^{(1:3)}\widehat{\mathbf{P}}_i^{(1:3)T} = \mathbf{P}_i \underbrace{\mathbf{AA}^T}_{\mathbf{Q}_{4\times4}} \mathbf{P}_i^T \tag{41}$$

where $\widehat{\mathbf{P}}_i^{(1:3)}$ is the first three column of $\widehat{\mathbf{P}}_i$, $\widehat{\mathbf{P}}_i^{(1:3)}\widehat{\mathbf{P}}_i^{(1:3)T} = \mu_i \mathbf{K}_i \mathbf{R}_i (\mu_i \mathbf{K}_i \mathbf{R}_i)^T = \mu_i^2 \mathbf{K}_i \mathbf{K}_i^T$, and $\mathbf{Q} = \mathbf{AA}^T$ is a 4×4 unknown symmetric matrix.

Thus, we have

$$\mu_i^2 \mathbf{K}_i \mathbf{K}_i^T = \mathbf{P}_i \mathbf{Q} \mathbf{P}_i^T \tag{42}$$

Let us denote $\mathbf{C}_i = \mathbf{P}_i \mathbf{Q} \mathbf{P}_i^T$. With the intrinsic parameters $\mathbf{K}_i$ of calibrated catadioptric cameras i, we can derive the following linear constraints on the matrix $\mathbf{Q}$

$$\begin{aligned} \mathbf{C}_i(1,1) &= (r_i^2 f_i^2 + s_i^2 + u_{0i}^2)\mathbf{C}_i(3,3) \\ \mathbf{C}_i(2,2) &= (f_i^2 + v_{0i}^2)\mathbf{C}_i(3,3) \\ \mathbf{C}_i(1,2) &= (f_i s_i + u_{0i} v_{0i})\mathbf{C}_i(3,3) \\ \mathbf{C}_i(1,3) &= u_{0i}^2 \mathbf{C}_i(3,3) \\ \mathbf{C}_i(2,3) &= v_{0i}^2 \mathbf{C}_i(3,3) \end{aligned}$$

Since the origin of image coordinate systems for the catadioptric cameras is transformed to the principal point and the catadioptric cameras are assumed to be skew-free, we have $u_{0i} = v_{0i} = 0$, $s_i = 0$.

$$\mathbf{C}_i(1,1) = r_i^2 f_i^2 \mathbf{C}_i(3,3) \tag{43}$$

$$\mathbf{C}_i(2,2) = f_i^2 \mathbf{C}_i(3,3) \tag{44}$$

$$\mathbf{C}_i(1,2) = 0 \tag{45}$$

$$\mathbf{C}_i(1,2) = 0 \tag{46}$$

$$\mathbf{C}_i(2,3) = 0 \tag{47}$$

After some simple manipulations, we can rewrite the constraints (43)–(47) into a set of linear equations about $\mathbf{Q}$, which can be solved by using singular value decomposition (SVD). For each central catadioptric camera, we have five linear equations of the 10 unknown elements in $\mathbf{Q}$. Therefore, $\mathbf{Q}$ can be estimated with at least two central catadioptric cameras, which have been calibrated with the method in Section 3.2. The matrix $\mathbf{A}$ can be further computed from $\mathbf{Q}$ by rank 3 matrix decomposition [16], thus the matrix $\mathbf{H} = (\mathbf{A}, \mathbf{B})$ can be recovered. With the transformation matrix $\mathbf{H}$, the projective reconstruction of the hybrid camera system $\{\mathbf{P}_i\}_{i=1}^N$, $\{\mathbf{X}_j\}_{j=1}^M$ can be upgraded to the Euclidean reconstruction by $\{\mathbf{P}_i\mathbf{H}\}_{i=1}^N, \{\mathbf{H}^{-1}\mathbf{X}_j\}_{j=1}^M$.

Finally, for each camera i (perspective or catadioptric) in the hybrid camera system, the intrinsic parameter matrix $\mathbf{K}_i$ and the extrinsic parameters $(\mathbf{R}_i, \mathbf{t}_i)$ of the ith camera can be extracted by the RQ decomposition [16] from the Euclidean projection matrix of the ith camera $\widehat{\mathbf{P}}_i = \mathbf{P}_i\mathbf{H}$, and the scene points can be recovered by $\widehat{\mathbf{X}}_j = \mathbf{H}^{-1}\mathbf{X}_j$.

Remark 2. The above calibration process is linear without prior knowledge about the intrinsic parameters of perspective cameras. In contrast, traditional self-calibration methods for perspective cameras either need to recover the plane at infinity, which

Table 2
Intrinsic parameters of the four simulated cameras.

Camera	f	r	s	u_0	v_0
1	2100	$\frac{2000}{2100}$	5.0	512	512
2	1710	$\frac{1700}{1710}$	0.0	700	750
3	1800	$\frac{1800}{1800}$	0.0	750	810
4	1910	$\frac{1900}{1910}$	0.0	850	880

usually involve nonlinear process and lead to unstable numerical computations [16], or need prior knowledge on intrinsic parameters (for example, the principal point, the aspect ratio and the skew factor) to get linear constraints of the image of absolute conic [20,16].

3.4. Bundle adjustment

The above calibration results are obtained by minimizing an algebraic distance, and they can be further refined through bundle adjustment.

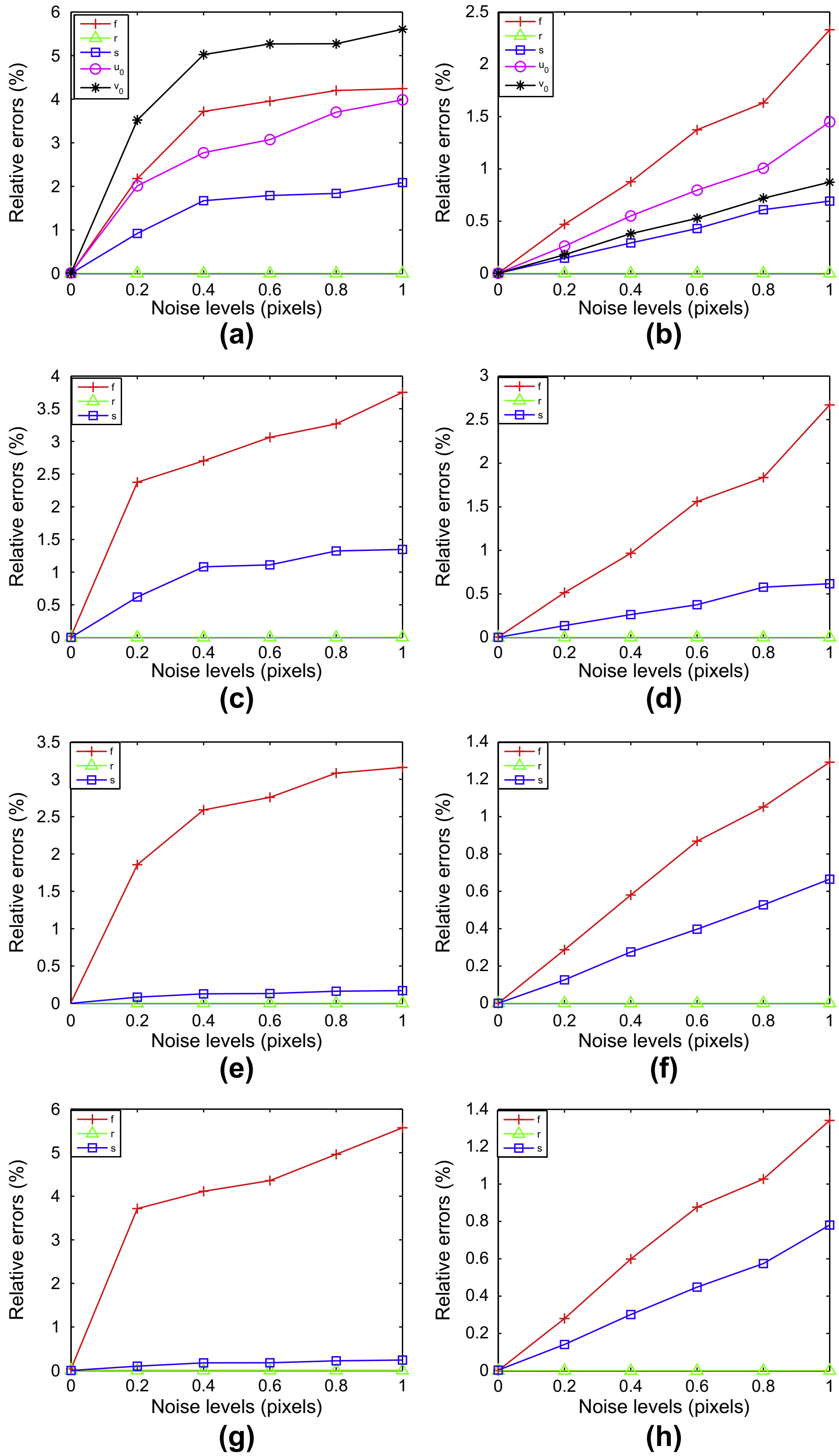

Fig. 2. The relative errors of four cameras' intrinsic parameters with 36 correspondences. (a), (c), (e) and (g) The linear calibration results of the 1st, 2nd, 3rd and 4th cameras respectively. (b), (d), (f) and (h) The results with bundle adjustment.

Suppose M correspondences are captured across N camera views. Denote $\mathbf{I}_{pers}$ to be the set of perspective cameras and $\mathbf{I}_{cata}$ the set of central catadioptric cameras. Denote $\widehat{\mathbf{X}}_q$ to be the qth reconstructed 3D point, $\mathbf{m}_{pq}$ the observed projection of $\widehat{\mathbf{X}}_q$ on the pth view.

For the set of central catadioptric cameras $\mathbf{I}_{cata}$, we can refine the parameters by minimizing the following function:

$$E_1 = \sum_{p \in \mathbf{I}_{cata}} \sum_{q=1}^{M} \left\| \mathbf{m}_{pq} - \check{\mathbf{m}}(\mathbf{K}_p, \xi_p, \mathbf{R}_p, \mathbf{t}_p, \widehat{\mathbf{X}}_q) \right\|^2 \tag{48}$$

where $\check{\mathbf{m}}(\mathbf{K}_p, \xi_p, \mathbf{R}_p, \mathbf{t}_p, \widehat{\mathbf{X}}_q)$ is the projection of $\widehat{\mathbf{X}}_q$ on the pth view with Eq. (2).

For the set of perspective cameras $\mathbf{I}_{pers}$, we can refine the parameters by minimizing the following function:

$$E_2 = \sum_{p \in \mathbf{I}_{pers}} \sum_{q=1}^{M} \left\| \mathbf{m}_{pq} - \bar{\mathbf{m}}(\mathbf{K}_p, \mathbf{R}_p, \mathbf{t}_p, \widehat{\mathbf{X}}_q) \right\|^2 \tag{49}$$

where $\bar{\mathbf{m}}(\mathbf{K}_p, \mathbf{R}_p, \mathbf{t}_p, \widehat{\mathbf{X}}_q)$ is the perspective projection of $\widehat{\mathbf{X}}_q$ on the pth view [16].

For the hybrid central catadioptric and perspective camera system, the intrinsic and extrinsic parameters can be refined by minimizing the following cost function:

$$\min E_1 + E_2 \tag{50}$$

where the variables are $\{\widehat{\mathbf{X}}_q | q = 1, \ldots, M\}$, $\{\mathbf{K}_p, \mathbf{R}_p, \mathbf{t}_p | p \in \mathbf{I}_{cata}\}$, $\{\mathbf{K}_p, \mathbf{R}_p, \mathbf{t}_p | p \in \mathbf{I}_{pers}\}$.

The nonlinear optimization problem can be solved with numerical techniques such as Levenberg–Marquardt [16]. It requires an initial values of decision variables. The intrinsic parameters of central catadioptric cameras can be determined with the method in Section 3.2, the intrinsic parameters of perspective cameras and the extrinsic parameters of all cameras and $\{\widehat{\mathbf{X}}_q | q = 1, \ldots, M\}$ can be initialized with the method in Section 3.3.

3.5. Method summary

An outline of our self-calibration method for hybrid central catadioptric and perspective cameras is summarized as follows:

Step 1: An approximated polynomial model is used for the rectification of catadioptric images. For each pair of catadioptric and perspective cameras, we establish the epipolar geometry between the perspective and rectified catadioptric images by (31) and (32), and the distortion parameters of the polynomial model can be estimated by singular value decomposition (SVD) and equations (35) and (36).

Step 2: For each central catadioptric camera, the intrinsic parameters can be recovered with the estimated distortion parameters of the polynomial model and Proposition 1.

Step 3: Self-calibrate the hybrid camera systems with the intrinsic parameters of calibrated central catadioptric cameras (see Section 3.3), and recover the intrinsic parameters of perspective cameras as well as extrinsic parameters of all the cameras.

Step 4: Refine all parameters by minimizing the nonlinear cost function (50) with numerical techniques such as Levenberg–Marquardt.

4. Experiments

In both simulated and real experiments, no prior knowledge on the intrinsic parameters of perspective cameras is used.

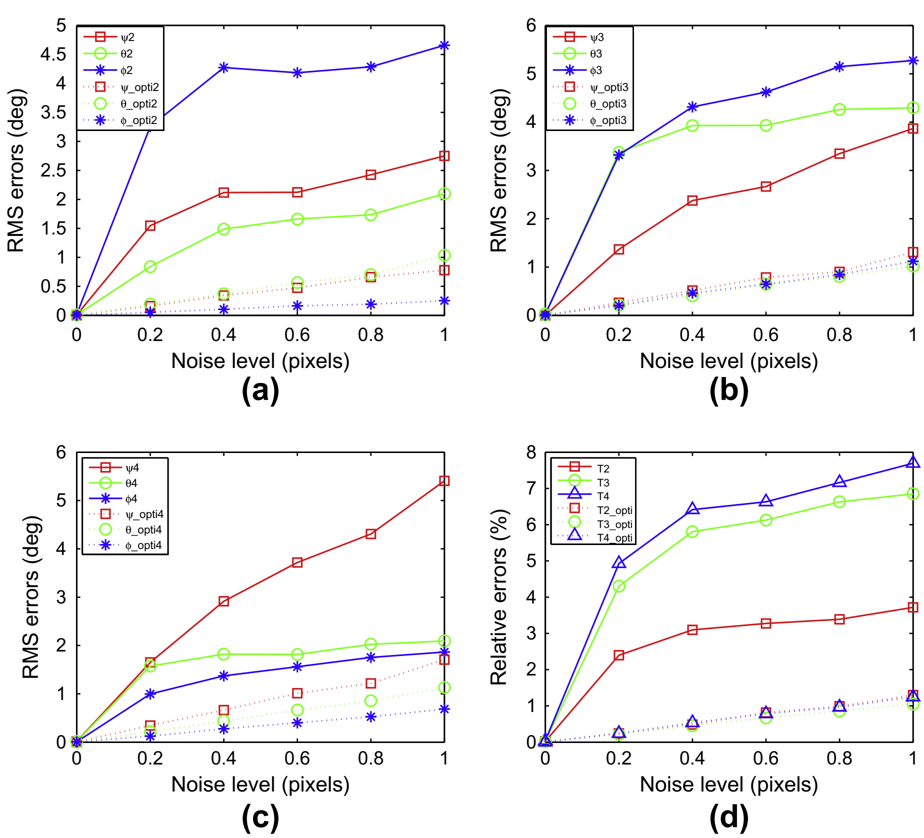

Fig. 3. The errors of three catadioptric cameras' rotation matrices and normalized translations with 36 correspondences.

4.1. Simulated experiments

We use four simulated cameras, the first one is a perspective camera, and the other three cameras are central catadioptric cameras with the mirror parameters $\xi = 0.96$, 0.98, 1.00 respectively. The intrinsic parameters of the simulated cameras are in Table 2. Simulated 3D points are projected onto the image planes, and the image correspondences are used for calibration.

4.1.1. Noise influence

Simulated 36 3D points are projected onto the image planes. Gaussian noise with mean 0 and standard deviation ranging from 0 to 1 pixel at the step 0.2 pixels is added to image points. For central catadioptric cameras, the mirror parameter ξ and principal points are assumed to be known. For each noise level, 200 independent trails are performed. Here, the relative errors of intrinsic parameters with respect to f are measured, which was suggested by Triggs [34]. Relative errors of the intrinsic parameters with linear and nonlinear estimations are given in Fig. 2. Errors increase almost linearly with the noise level increases, and bundle adjustment can produce fairly better results than the linear estimation. Fig. 3a–c display the errors of the other cameras' Euler angles relative to the first camera, Fig. 3d shows the errors of the other cameras' normalized translations. The errors also increase almost linearly with the noise level, and bundle adjustment can improve the results of the linear estimation. The experiment also shows that the 1st-order polynomial model is valid for the simulated central catadioptric cameras.

4.1.2. Influence of the number of image correspondences

We investigate the calibration performance with respect to the number of image correspondences, which varies from 9 to 36. In the experiments with 18, 27 and 36 correspondences, the catadioptric and perspective cameras are calibrated with the same set of correspondences. In the experiments with 9 correspondences, only catadioptric camera parameters are estimated with the nine-point solver, and the perspective camera are self-calibrated with 18 correspondences for the numerical robustness of self-calibration [16].

Gaussian noise with mean 0 and standard deviation 0.2 pixels is added to image points, 200 independent trails are performed for each correspondence number, and intrinsic and extrinsic camera

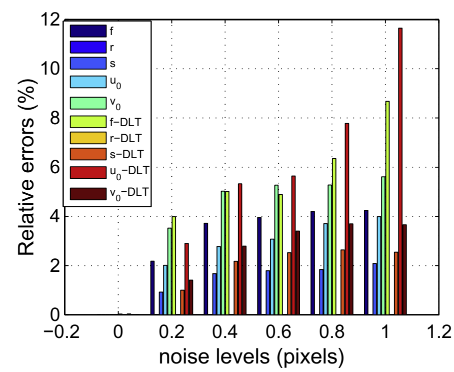

Fig. 5. The errors of perspective intrinsic parameters with our method and with the method by the reconstructed scene points via Direct Linear Transformation (DLT). (For interpretation of the references to colour in this figure the reader is referred to the web version of this article).

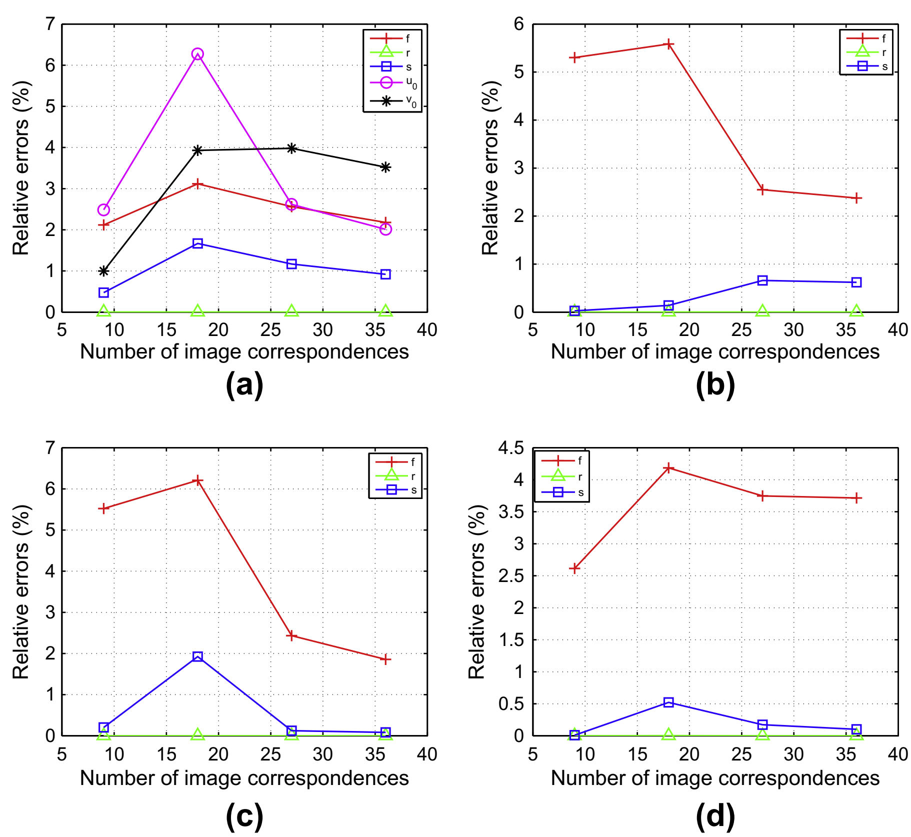

Fig. 4. Calibration errors of four cameras' intrinsic parameters with different correspondence numbers: 9, 18, 27 and 36. (a) Perspective camera. (b–d) Three catadioptric cameras. In experiments with 9 correspondences, only catadioptric camera parameters are estimated with the nine-point solver, and the perspective camera is self-calibrated with the same correspondences as in the experiments with 18 correspondences.

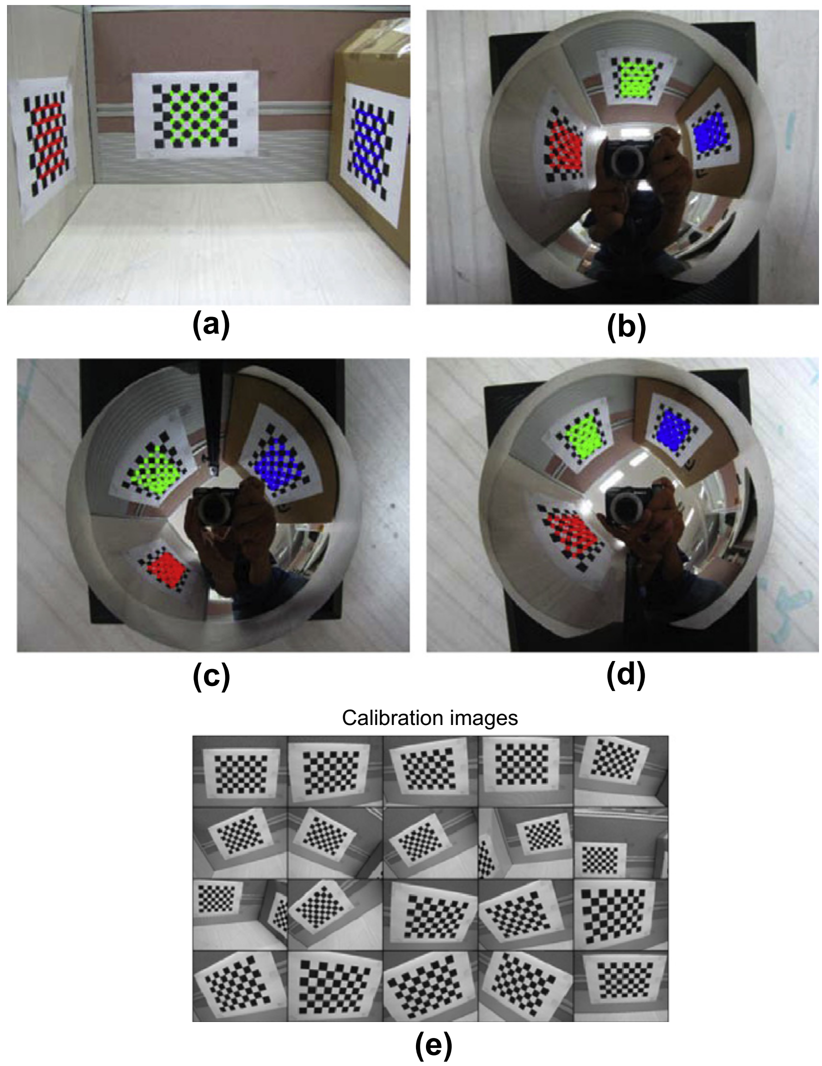

Fig. 6. The images used for calibration. (a) A perspective image. (b–d) Three catadiopotric images. (e) Images of calibration pattern captured by the perspective camera, and they are used to calibrate the camera by Bouguet Toolbox [6]. (For interpretation of the references to colour in this figure the reader is referred to the web version of this article).

Table 3
Intrinsic parameters of four cameras with 108 correspondences. The 1st camera is a perspective camera, and the other three cameras are all central catadioptric cameras.

Camera	f	r	s	u_0	v_0
1	1870.00	0.99	1.88	995.65	824.86
2	1174.50	0.94	6.46	1866.80	1372.60
3	1069.70	1.06	27.43	1827.90	1372.60
4	1056.00	0.98	13.92	1845.50	1372.10

parameters for perspective and catadioptric cameras are estimated. The relative errors of intrinsic parameters with respect to f are shown in Fig. 4. In the experiments with 18, 27 and 36 correspondences, the calibration results with more correspondences are better. This is due to the fact that more correspondences can improve the numerical robustness in the calibration and a better catadioptric camera calibration improves the perspective camera calibration results. The errors of catadioptric parameters with 9 correspondences are smaller than those with 18 correspondences (This case appears special), which is probably due to the following fact that prior knowledge ($c_{20} = c_{02}$) of central catadioptric camera is used in the nine-point solver; However, in the experiments with 18 correspondences, no such prior knowledge of central catadioptric camera is used. These experimental results indicate that the availability of some camera prior knowledge is more important than by merely increasing the number of correspondences sometimes. Experiment results show that the calibration results of the perspective camera in experiments with 9 correspondences are also better than those with 18 correspondences.

Table 4
Calibration results of three catadioptric cameras' intrinsic parameters with different numbers of correspondences (without bundle adjustment).

Cameras	Point number	f	r	s	u_0	v_0
2	9	1117.10	1.00	0.00	1866.80	1372.60
	18	1335.10	0.93	0.00	1866.80	1372.60
	27	1198.40	0.87	0.00	1866.80	1372.60
	36	1280.20	0.99	0.00	1866.80	1372.60
3	9	900.70	1.00	0.00	1827.90	1372.60
	18	1254.90	1.02	0.00	1827.90	1372.60
	27	1052.40	0.98	0.00	1827.90	1372.60
	36	1131.00	0.98	0.00	1827.90	1372.60
4	9	1012.70	1.00	0.00	1845.50	1372.10
	18	1250.60	1.02	0.00	1845.50	1372.10
	27	1060.60	0.99	0.00	1845.50	1372.10
	36	1126.60	1.05	0.00	1845.50	1372.10

4.1.3. Comparison of the estimated perspective camera parameters with DLT method

We compare the calibration results of perspective cameras with our method and the method by the reconstructed scene points with Direct Linear Transformation (DLT) [16]. In theory, the perspective camera can be calibrated with the reconstructed scene points with DLT, while this method is weak in combating image noises. In order to reconstruct scene points, motions between cameras in the same coordinate system should be estimated, while accurate motion estimation of multiple camera itself is not a trivial task for image correspondences contaminated with noises [14,5]. As for this work, if the intrinsic parameters of two catadioptric cameras are calibrated, the essential matrix can be recovered with image correspondences, and the camera motion can be extracted. Generally speaking, this motion estimation method does not exploit the constraints across multiple cameras, and it just uses two camera geometric constraints, which is weak in combating image noises [28]. After the motion between cameras are estimated, the 3D scene point can be reconstructed with triangulation, and thus the projection matrix of perspective camera can be calibrated with reconstructed 3D scene points and their images by DLT. Finally, the intrinsic and extrinsic parameters of perspective camera can be recovered with RQ decomposition.

The obtained experiment results show that our method is more reliable than DLT method as the noise level increases. Here, the relative errors of perspective intrinsic parameters with respect to f are measured. Fig. 5 shows the errors of intrinsic parameters with our method and with Direct Linear Transformation (DLT) as the noise level increases. The errors of focal length with our method are smaller than those with DLT, and the performance of other parameters is close.

4.2. Real experiments

In this section, we use real hybrid camera image data to validate our calibration method.

For the experiment with real data, a perspective cameras and a central catadioptric camera are used. The image resolution of the perspective cameras is of 2048×1536 pixels, and the resolution of the catadioptric camera is of 3648×2736 pixels. The used central catadioptric camera is a Canon Powershot digital camera, combined with a hyperboloid mirror designed by the Center for Machine Perception (CMP), Czech Technical University.[1] The eccentricity of the hyperbolic mirror is 1.302, and the corresponding mirror parameter ξ is 0.966.

An image with the perspective camera and three images by moving the central catadioptric camera three times are shown in Fig. 6a–d. The principal points of three catadioptric cameras are precalibrated using the center of the bounding ellipse like in [35,39]. We manually select 108 correspondences across the four views. The intrinsic parameters of perspective camera and catadioptric cameras with 108 correspondences are shown in Table 3. The cameras' parameters are calibrated with bundle adjustment. The estimations of skew factors s for camera 3 and 4 are a bit high. However, the relative error of skew factor with respect to f could be a better calibration evaluation measure, which was suggested by Triggs [34]. By this measure, the relative errors of skew factors for camera 3 and 4 are fairly small, and the maximal relative error is 2.56%.

We investigate calibration performance of three catadioptric cameras' intrinsic parameters with respect to the number of used image correspondences for distortion parameter estimation. Here, we choose 9, 18, 27 and 36 correspondences evenly from the 108 correspondences. As shown in Table 4, the intrinsic parameters of catadioptric cameras are close as the number of correspondences varies.

The calibration results of central catadioptric camera are evaluated with rectified images. The rectified images of Fig. 6a and b are shown in Fig. 7a and b. The rectified lines are straight, and the heavy distortion is removed. In addition, the three catadioptric images are captured with the same camera, and the calibrated intrinsic parameters for camera 2, 3, 4 are all close. We notice a MATLAB toolbox in [21], which estimates the mirror parameter ξ and the other parameters using a planar object. When performing the calibration by [21], we use the calibration toolbox that the author provided. The calibration results are $f = 1087.1$, $r = 0.9557$, $s = 0.0$, $u_0 = 1856.52$, $v_0 = 1487.09$, and $\xi = 0.9866 \pm 0.4525$. Details on Mei's calibration method can be seen in [21]. From the experiments, we can see that the results by our method is close to the result by Mei's method. The differences of our method and Mei's method lie in that our method does not use planar structure information of calibration pattern while Mei's method requires such information. The experiments validate the correctness of our calibration results.

The calibration results of the perspective camera are compared with the results by the 2D plane-based calibration method [6]. In the experiment, the perspective camera is calibrated by using Bouguet Calibration Toolbox [6], and images used for calibration are shown in Fig. 6e and f. The intrinsic parameters by our method and 2D plane-based calibration method are shown in Table 5, the first row shows the estimation by our method, the second row is the result by the 2D plane-based method, and the third row is the relative error of the results with our method to the focal length

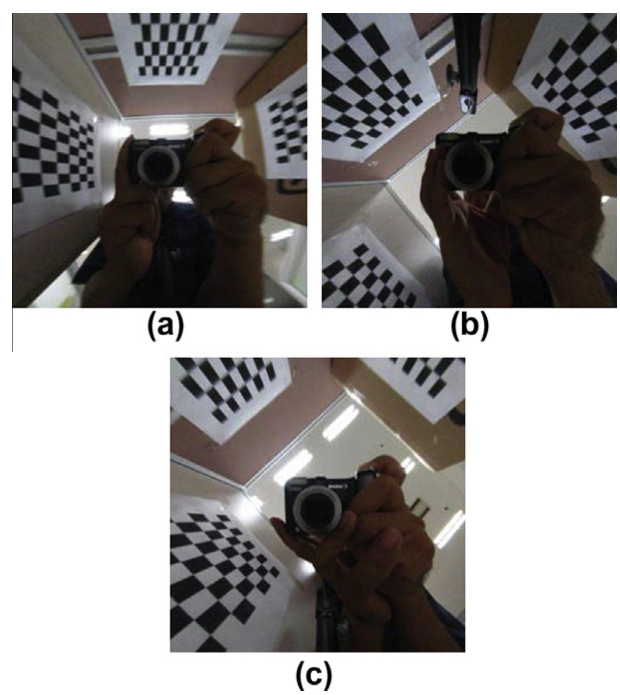

Fig. 7. Rectification of images in Fig. 6(b–d). Only a part of the rectified images are shown due to the fact that the original transformed images are very large.

Table 5
Calibration results of the perspective camera.

Method	f	r	s	u_0	v_0
Our method	1870.00	0.99	1.88	995.65	824.86
2D method [6]	2160.90	1.00	0.00	1047.72	758.25
Relative error (%)	13.50	0.00	0.08	2.41	3.08

[1] http://www.neovision.cz/prods/panoramic/mirrfoc.html.

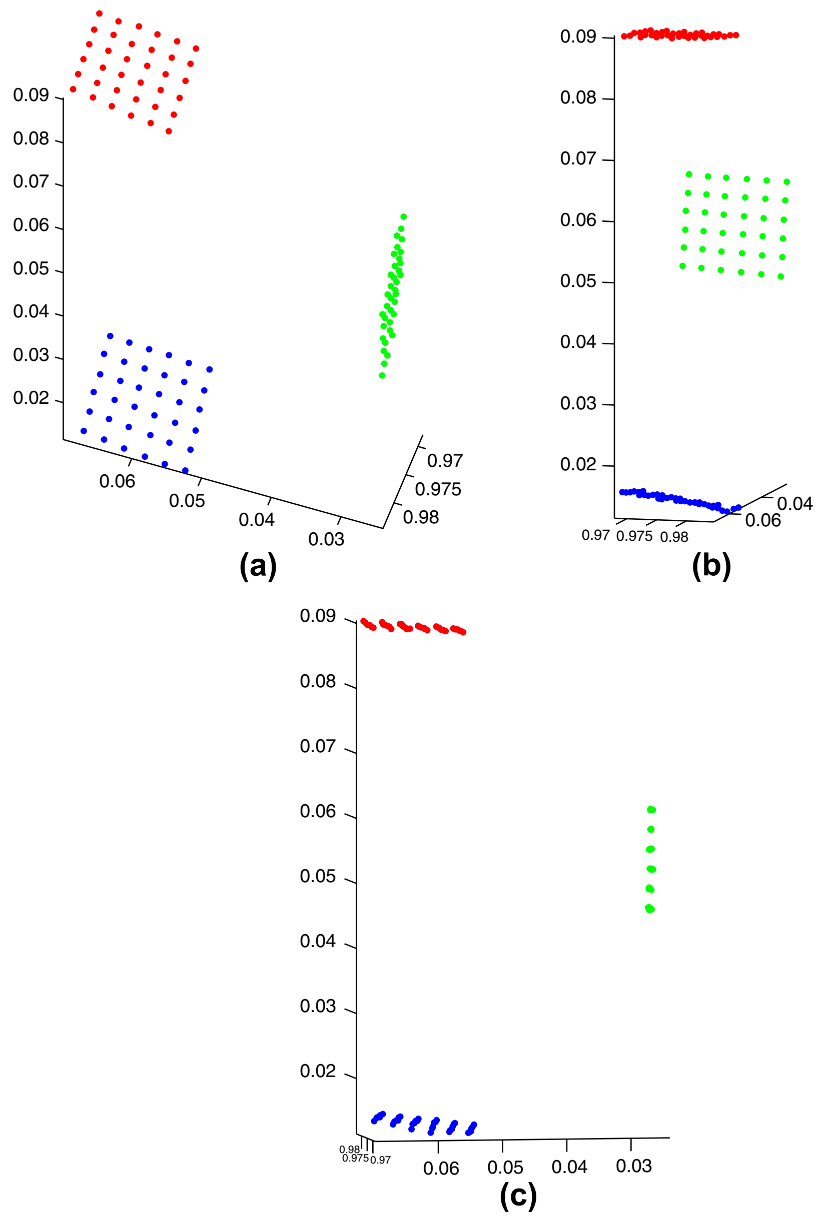

Fig. 8. Three views of the reconstructed scene points. (a) Side view of the reconstructed grids. (b) Front view (rotated) of the reconstructed grids. (c) Top view of the reconstructed grids. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Table 6
Reconstruction errors with the results by our method.

Method	$\bar{d}$ (mm)	σ_d (mm)	ψ (deg)	σ_ψ (deg)
Reconstruction error	22.80	0.80	89.60	0.78
	d (mm)		ψ (deg)	
Ground truth	22.00		90.00	

by [6]. We can see that the results with our method and the method [6] are close, and the errors of intrinsic parameters by our method are small.

Reconstruction experiments with the calibration results are also carried out to test the calibration results. In the reconstruction experiments, the sphere model in Eq. (2) is used. After self-calibration, the 3D positions of correspondences are recovered up to a common scale, and we use the lengths of line segments to estimate the scale. The images of three planar calibration patterns with black and white squares are shown in Fig. 6, and the reconstruction results are shown in Fig. 8. Table 6 allows a qualitative evaluation of the reconstruction by measuring the deviations from the right angles and from distances between points on a facet of the calibration object. Angles between vertical and horizonal line segments θ (whose ground truth is 90°) are computed, and the ground truth of distance d between each pair of neighbor corner points is 22 mm. The mean θ and standard deviation of σ_θ, and the mean $\bar{d}$ and standard deviation σ_d of distances computed by the reconstructed 3D points are shown in Table 6. The results by our method are very close to the ground truth.

5. Conclusions

In this paper, we proposed a linear self-calibration method of hybrid central catadioptric and perspective cameras. Our calibration method is flexible, because it does not require any special calibration pattern, nor does it assume any prior knowledge of camera motion or scene geometry. The proposed method cannot only calibrate the intrinsic and extrinsic parameters of the hybrid camera system but also simplify the traditional nonlinear self-calibration of perspective cameras to a linear process. Traditional self-calibration methods of perspective cameras usually involve nonlinear minimization which could lead to numerical instability, or require prior knowledge on the intrinsic parameters for a linear solution. In this work, for a hybrid camera system with at least two central catadioptric cameras and one perspective camera, we can linearly calibrate the system without any assumption of the perspective intrinsic parameters, which makes our method more practical. Experiments show that our method is robust and reliable. Hybrid camera systems are currently available at low cost, and it is necessary that such systems are deployed rapidly with minimal preparation [3]. Our method shows its potential in those systems because it relies merely on correspondences without any further prior knowledge on the scene and only little prior knowledge on cameras.

Acknowledgments

This work is supported by National Natural Science Foundation of China (Nos. 61005039 and 60872127), National Key Basic Research and Development Program of China (No. 2009CB320804), the open projects program of National Laboratory of Pattern Recognition of Institute of Automation, Chinese Academy of Sciences under Grant No. 20090096, and the Fundamental Research Funds for the Central Universities. The authors are grateful to the editors and reviewers for their careful review and inspiring suggestions.

Appendix A. Normalization for hybrid central catadioptric and perspective images

Hartley [16] suggested that the proper normalization of the image data is the key to success with the 8-point algorithm for fundamental matrix estimation. The normalization suggested by Hartley is a translation and scaling of each image so that the centroid of the input data points is at the origin of the coordinates and the RMS distance of the points from the origin is equal to $\sqrt{2}$. The normalized data leads to an enormous improvement in the conditioning of the problem and hence the stability of the estimation. In this section, we propose a normalization method for hybrid central catadioptric and perspective images.

Epipolar geometry between the perspective and rectified central catadioptric images is

$$\mathbf{m}_1^T \mathbf{F} \hat{\mathbf{m}}_{u2}(\mathbf{c}) = 0 \tag{A.1}$$

where $\mathbf{m}_1 = (u_1, v_1, 1)^T$ and $\hat{\mathbf{m}}_{u2}(\mathbf{c}) = ((1+\xi)u_2, (1+\xi)v_2, 1 + c_{20}u_2^2 + c_{02}v_2^2)^T$.

$\hat{\mathbf{m}}_{u2}(\mathbf{c})$ can be transformed into the following:

$$\hat{\mathbf{m}}_{u2}(\mathbf{c}) = \begin{pmatrix} (1+\xi)u_2 \\ (1+\xi)v_2 \\ 1 + c_{20}u_2^2 + c_{02}v_2^2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1+\xi & 0 & 0 \\ 0 & 0 & 0 & 1+\xi & 0 \\ c_{20} & c_{02} & 0 & 0 & 1 \end{pmatrix} \underbrace{\begin{pmatrix} u_2^2 \\ v_2^2 \\ u_2 \\ v_2 \\ 1 \end{pmatrix}}_{\tilde{\mathbf{m}}_2} \tag{A.2}$$

where $\tilde{\mathbf{m}}_2 = (u_2^2, v_2^2, u_2, v_2, 1)^T$.

Incorporating (A.2) into (A.1), we have

$$(u_1, v_1, 1) \underbrace{\begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1+\xi & 0 & 0 \\ 0 & 0 & 0 & 1+\xi & 0 \\ c_{20} & c_{02} & 0 & 0 & 1 \end{pmatrix}}_{\widetilde{\mathbf{F}}} \begin{pmatrix} u_2^2 \\ v_2^2 \\ u_2 \\ v_2 \\ 1 \end{pmatrix} = 0 \tag{A.3}$$

where

$$\widetilde{\mathbf{F}} = \begin{pmatrix} c_{20}f_{13} & c_{02}f_{13} & (1+\xi)f_{11} & (1+\xi)f_{12} & f_{13} \\ c_{20}f_{23} & c_{02}f_{23} & (1+\xi)f_{21} & (1+\xi)f_{22} & f_{23} \\ c_{20} & c_{02} & (1+\xi)f_{31} & (1+\xi_2)f_{32} & 1 \end{pmatrix} \tag{A.4}$$

Therefore, we have

$$\mathbf{m}_1 \widetilde{\mathbf{F}} \tilde{\mathbf{m}}_2 = 0 \tag{A.5}$$

Like the normalized 8-point algorithm for perspective cameras, we can normalize the correspondences of perspective and central catadioptric cameras as follows (inspired by [4]):

Normalization for perspective image as suggested by Hartley [16] is in the following:

$$\mathbf{T}_1 = \begin{pmatrix} d_1 & 0 & -d_1c_{x1} \\ 0 & d_1 & -d_1c_{y1} \\ 0 & 0 & 1 \end{pmatrix} \tag{A.6}$$

where d_1 is the value used for scale normalization and (c_{x1}, c_{y1}) be the centroid of the points in the perspective image. After the transform of the image coordinates $\mathbf{m}_1$ using $\mathbf{m}_1' = \mathbf{T}_1\mathbf{m}_1$, we get the transformed image coordinates $\mathbf{m}_1'$.

Normalization for central catadioptric image is in the following:

$$\mathbf{T}_2 = \begin{pmatrix} d_2^2 & 0 & -2c_{x2}d_2^2 & 0 & c_{x2}^2d_2^2 \\ 0 & d_2^2 & 0 & -2c_{y2}d_2^2 & c_{y2}^2d_2^2 \\ 0 & 0 & d_2 & 0 & -c_{x2}d_2 \\ 0 & 0 & 0 & d_2 & -c_{y2}d_2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{A.7}$$

where d_2 is the value used for scale normalization and (c_{x2}, c_{y2}) be the centroid of the points in the central catadioptric image, lifting a normalized point leads to the vector of five dimensions. In our experiments, d_1 and d_2 are chosen so that the RMS distance of the points (u_1, v_1) and (u_2, v_2) from the centroids are equal to $\sqrt{2}$. The geometric meaning of the normalization transformation $\mathbf{T}_2$ for $\tilde{\mathbf{m}}_2$ is that the RMS distances from (u_2^2, v_2^2) to the centroid $\left(c_{x2}^2, c_{y2}^2\right)$ and from (u_2, v_2) to the centroid (c_{x2}, c_{y2}) are both equal to $\sqrt{2}$.

After the transform of the image coordinates $\tilde{\mathbf{m}}_2$ using $\tilde{\mathbf{m}}_2' = \mathbf{T}_2\tilde{\mathbf{m}}_2$, we get the transformed image coordinates $\tilde{\mathbf{m}}_2'$.

Thus we can obtain a set of new point correspondences $\{\mathbf{m}_1' \leftrightarrow \tilde{\mathbf{m}}_2'\}$ and use the SVD method to estimate the fundamental matrix $\widetilde{\mathbf{F}}'$.

Therefore, we can recover the fundamental matrix $\widetilde{\mathbf{F}}$ by the following equation:

$$\widetilde{\mathbf{F}} = \mathbf{T}_1^T \widetilde{\mathbf{F}}' \mathbf{T}_2 \tag{A.8}$$

With the estimated $\widetilde{\mathbf{F}}$, we can get the vector $\mathbf{g}$ in (34) easily, and thus the distortion parameter (c_{20}, c_{02}) can be recovered by Eqs. (35)–(37).

Appendix B. A nine-point solver for the distortion parameter estimation problem

In this section, we derive a nine-point solver for the distortion parameter estimation problem.

It is well known that for standard uncalibrated case without considering radial distortion, 7 point correspondences are sufficient to estimate the epipolar geometry [16]. We have two more parameters, the radial distortion parameters $\{c_{20}, c_{02}\}$. Therefore, 9 point correspondences are sufficient to estimate the unknown $(f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{23}, f_{31}, f_{32}, f_{33}, c_{20}, c_{02})$, and the solution can be obtained by Gröbner basis solvers [18].[2] However, Gröbner basis solvers in the algebraic geometry field were generally designed for concrete problems, and cannot be applied to new problems [10].

In this section, we use distortion priors to derive a nine-point solver for the distortion parameter estimation problem, the solution can be a close approximation to the true value, and then the distortion solution could be further refined with optimizations. The approach is not a minimal solver for the problem, while it has the merit of easy implementation.

Firstly, we use camera distortion prior knowledge of central catadioptric camera ($c_{20} = c_{02}$) to simplify the problem. The geometric meaning $c_{20} = c_{02}$ is that the aspect ratio of the camera is 1.0 (see Proposition 1), which is a close approximation for most of off-the-shelf cameras. Thus, $\hat{\mathbf{m}}_{u2}(c)$ can be transformed into the following form [4]:

$$\hat{\mathbf{m}}_{u2}(c) = \begin{pmatrix} (1+\xi)u_2 \\ (1+\xi)v_2 \\ 1 + c_{20}(u_2^2+v_2^2) \end{pmatrix} = \begin{pmatrix} 0 & 1+\xi & 0 & 0 \\ 0 & 0 & 1+\xi & 0 \\ c_{20} & 0 & 0 & 1 \end{pmatrix} \underbrace{\begin{pmatrix} u_2^2+v_2^2 \\ u_2 \\ v_2 \\ 1 \end{pmatrix}}_{\check{\mathbf{m}}_2} \tag{B.1}$$

The fundamental matrix $\check{\mathbf{F}}$ between $\mathbf{m}_1 = (u_1, v_1, 1)^T$ and $\check{\mathbf{m}}_2 = (u_2^2 + v_2^2, u_2, v_2, 1)^T$ is in the following form:

$$\check{\mathbf{F}} = \begin{pmatrix} c_{20}f_{13} & (1+\xi)f_{11} & (1+\xi)f_{12} & f_{13} \\ c_{20}f_{23} & (1+\xi)f_{21} & (1+\xi)f_{22} & f_{23} \\ c_{20} & (1+\xi)f_{31} & (1+\xi)f_{32} & 1 \end{pmatrix} \tag{B.2}$$

Denoted the vector $\check{\mathbf{g}}$ to be the item vector of $\check{\mathbf{F}}$ as $\check{\mathbf{g}} = (\check{g}_1, \check{g}_2, \ldots, \check{g}_{12})^T = (c_{20}f_{13}, (1+\xi)f_{11}, (1+\xi)f_{12}, f_{13}, c_{20}f_{23}, (1+\xi)f_{21}, (1+\xi)f_{22}, f_{23}, c_{20}, (1+\xi)f_{31}, (1+\xi)f_{32}, 1)^T$. We pile up equations of epipolar geometry constraint $\mathbf{m}_1^T\check{\mathbf{F}}\check{\mathbf{m}}_2 = 0$ by using nine correspondences, and we have the following equations:

$$\underbrace{\left(\check{\mathbf{D}}_{1,1}^T, \check{\mathbf{D}}_{1,2}^T, \ldots, \check{\mathbf{D}}_{1,9}^T\right)^T}_{\check{\mathbf{D}}} \check{\mathbf{g}} = 0 \tag{B.3}$$

where $\check{\mathbf{D}} = \left(\check{\mathbf{D}}_{1,1}^T, \check{\mathbf{D}}_{1,2}^T, \ldots, \check{\mathbf{D}}_{1,9}^T\right)^T$, and $\check{\mathbf{D}}_{1,i}$ is the coefficient vector of $\check{\mathbf{g}}$ by the ith pair correspondences.

With nine correspondences between $(u_1, v_1, 1)^T$ and $(u_2^2 + v_2^2, u_2, v_2, 1)^T$, we can compute three vectors $\check{\mathbf{g}}_i (i = 0, 1, 2)$, which form an orthogonal basis for the null space of $\check{\mathbf{D}}$, and the solution $\check{\mathbf{g}}$ should be in the form of

$$\check{\mathbf{g}} = x_0\check{\mathbf{g}}_0 + x_1\check{\mathbf{g}}_1 + x_2\check{\mathbf{g}}_2 \tag{B.4}$$

where $x_i (i = 1,2)$ are three unknown scalars. Since the nine equations generated by $\mathbf{m}_1^T\check{\mathbf{F}}\check{\mathbf{m}}_2 = 0$ are homogeneous in $\check{\mathbf{g}}$, we set $x_0 = 1$.

Secondly, we can get the following three constraints on $x_i (i = 1,2)$ by using the above parameterizations of $\check{\mathbf{g}}$, the rank constraint for fundamental matrix and the relationships between the variables in $\check{\mathbf{g}}$

$$\check{g}_4\check{g}_9 - \check{g}_1 = 0 \tag{B.5}$$

$$\check{g}_8\check{g}_9 - \check{g}_5 = 0 \tag{B.6}$$

$$\det(\mathbf{F}) = 0 \iff \det\begin{pmatrix} \check{g}_2 & \check{g}_3 & \check{g}_4 \\ \check{g}_6 & \check{g}_7 & \check{g}_8 \\ \check{g}_{10} & \check{g}_{11} & \check{g}_{12} \end{pmatrix} = 0 \tag{B.7}$$

The above problems for estimating the parameters $x_i (i = 1,2)$ all lead to solving systems of algebraic equations. The first two equations leads to the following symbolic equations in two unknown variables x_i $(i = 1,2)$ (two second order polynomials)

$$\begin{pmatrix} \check{g}_1^{(4)}\check{g}_1^{(9)} & \check{g}_2^{(4)}\check{g}_1^{(9)} + \check{g}_1^{(4)}\check{g}_2^{(9)} & \check{g}_2^{(4)}\check{g}_2^{(9)} & \check{g}_0^{(4)}\check{g}_1^{(9)} + \check{g}_1^{(4)}\check{g}_0^{(9)} - \check{g}_1^{(1)} & \check{g}_0^{(4)}\check{g}_2^{(9)} + \check{g}_2^{(4)}\check{g}_0^{(9)} - \check{g}_2^{(1)} & \check{g}_0^{(4)}\check{g}_0^{(9)} - \check{g}_0^{(1)} \\ \check{g}_1^{(8)}\check{g}_1^{(9)} & \check{g}_2^{(8)}\check{g}_1^{(9)} + \check{g}_1^{(8)}\check{g}_2^{(9)} & \check{g}_2^{(8)}\check{g}_2^{(9)} & \check{g}_0^{(8)}\check{g}_1^{(9)} + \check{g}_1^{(8)}\check{g}_0^{(9)} - \check{g}_1^{(5)} & \check{g}_0^{(8)}\check{g}_2^{(9)} + \check{g}_2^{(8)}\check{g}_0^{(9)} - \check{g}_2^{(5)} & \check{g}_0^{(8)}\check{g}_0^{(9)} - \check{g}_0^{(5)} \end{pmatrix} \begin{pmatrix} x_1^2 \\ x_1x_2 \\ x_2^2 \\ x_1 \\ x_2 \\ 1 \end{pmatrix} = \mathbf{0} \tag{B.8}$$

where $\check{g}_i^{(j)} (i = 0,1,2; j = 1,2,\ldots,12)$ is the jth item of the vector $\check{\mathbf{g}}_i$.

The solutions of symbolic Eq. (B.8) can be solved accurately and easily with the `solve` function in Symbolic Math Toolbox of Matlab, and no initializations for $x_i (i = 1,2)$ are required. Notice that multiple solutions of $x_i (i = 1,2)$ can be computed from the two symbolic equations, we choose the solution which can reach the minimal value of cost function like Eq. (37).

Since the $\check{\mathbf{F}}$ has one more parameter c_{20} than a conventional fundamental matrix, $\check{\mathbf{F}}$ can be recovered with a minimal number of eight correspondences. Our approach uses nine correspondences, which is not a minimal solver for the problem, while our approach only need to solve equations in two unknown variables $x_i (i = 1,2)$ (two second order polynomials) and avoids solving the more complex third order polynomials equation via $\det(\check{\mathbf{F}}) = 0$ in the case of minimal solver with eight correspondences. This approach has the advantage of easy implementation.

Finally, we can refine the catadioptric parameters by nonlinear optimization techniques. The cost function is chosen as the sum of the residual squares of the Eq. (37), and the distortion parameter c_{02} is initialized with c_{20}.

[2] A good link of minimal problems in computer vision, http://cmp.felk.cvut.cz/minimal/index.php.

References

[1] S. Baker, S. Nayar, A theory of single-viewpoint catadioptric image formation, Int. J. Comput. Vision 35 (2) (1999) 175–196.
[2] J.P. Barreto, K. Danilidis, 2004. Wide area multiple camera calibration and estimation of radial distortion. In: Proceedings of OMNIVIS'2004 – Workshop on Omnidirectional Vision.
[3] J.P. Barreto, K. Daniilidis, Epipolar geometry of central projection systems using 1066 veronese maps, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2006, pp. 1258–1265.
[4] Y. Bastanlar, A. Temizel, Y. Yardimci, P. Sturm, Effective structure-from-motion for hybrid camera systems, in: Proceedings of International Conference on Pattern Recognition, 2010.
[5] J.C. Bazin, C. Demonceaux, P. Vasseur, I.S. Kweon, Motion estimation by decoupling rotation and translation in catadioptric vision, Comput. Vision Image Understand. 114 (2) (2010) 254–273.
[6] Bouguet Calibration Toolbox, 2010. <http://www.vision.caltech.edu/bouguetj/>.
[7] J. Chahl, M. Srinivasan. A complete panoramic vision system, incorporating imaging, ranging, and three dimensional navigation, in: Proceedings of IEEE Workshop on Omnidirectional Vision, 2000, pp. 104–111.

[8] C.H. Chen, Y. Yao, D. Page, B. Abidi, A. Koschan, M. Abidi, Camera handoff and placement for automated tracking systems with multiple omnidirectional cameras, Comput. Vision Image Understand. 114 (2) (2010) 179–197.
[9] X. Chen, J. Yang, A. Waibel, Calibration of a hybrid camera network, in: Proceedings of the 9th International Conference on Computer Vision, 2003, pp. 150–155.
[10] D.A. Cox, J.B. Little, D. O'Shea, Using Algebraic Geometry, Graduate Texts in Mathematics, second ed., vol. 18, Springer, 2005.
[11] X. Deng, F. Wu, Y. Wu, F. Duan, Visual metrology with uncalibrated radial distorted images, in: Proceedings of International Conference on Pattern Recognition, 2008, pp. 1–4.
[12] F. Duan, F. Wu, M. Zhou, X. Deng, Y. Tian, Calibrating effective focal length for central catadioptric cameras using one space line, Pattern Recogn. Lett. 33 (5) (2012) 646–653.
[13] A.W. Fitzgibbon, Simultaneous linear estimation of multiple view geometry and lens distortion, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2001.
[14] T. Gandhi, M.M. Trivedi, Reconfigurable omnidirectional camera array calibration with a linear moving object, Image Vision Comput. 24 (9) (2006) 935–948.
[15] C. Geyer, K. Daniilidis, Catadioptric projective geometry, Int. J. Comput. Vision 43 (2001) 223–243.
[16] R.I. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, Cambridge, 2003.
[17] S.B. Kang, Catadioptric camera self-calibration, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2000, pp. 201–207.
[18] Z. Kukelova, M. Bujnak, T. Pajdla, Automatic generator of minimal problem solvers, in: Proceedings of European Conference on Computer Vision, 2008, pp. 302–315.
[19] M. Lhuillier, Toward flexible 3D modeling using a catadioptric camera, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
[20] S. Maybank, O. Faugeras, A theory of self-calibration of a moving camera, Int. J. Comput. Vision 8 (2) (1992) 123–151.
[21] C. Mei, P. Rives, Single view point omnidirectional camera calibration from planar grids, in: Proceedings of International Conference on Robotics and Automation, 2007, pp. 3945–3950.
[22] B. Micusik, D. Martinec, T. Pajdla, 2004, 3D metric reconstruction from uncalibrated omnidirectional image, in: Proceedings of Asia Conference on Computer Vision.
[23] D. Michel, A.A. Argyros, M.I.A. Lourakis, Horizon matching for localizing unordered panoramic images, Comput. Vision Image Understand. 114 (2) (2010) 274–285.
[24] H. Nagahara, Y. Yagi, M. Yachida, Wide field of view head mounted display for tele-presence with an omnidirectional image sensor, in: Proceedings of Computer Vision and Pattern Recognition Workshop, vol. 7, 2003..
[25] S. Ramalingam, S.K. Lodha, P.F. Sturm, A generic structure-from-motion framework, Comput. Vision Image Understand. 103 (3) (2006) 218–228.
[26] G. Scotti, L. Marcenaro, C. Coelho, F. Selvaggi, C. Regazzoni, Dual camera intelligent sensor for high definition 360 degrees surveillance, IEE Proceedings of Vision, Image and Signal Processing 152 (2) (2005) 250–257.
[27] S.N. Sinha, M. Pollefeys, Pan-tilt-zoom camera calibration and high-resolution mosaic generation, Comput. Vision Image Understand. 103 (3) (2006) 170–183.
[28] P.F. Sturm, B. Triggs, A factorization based algorithm for multi-image projective structure and motion, in: Proceedings of European Conference on Computer Vision, 1996, pp. 709–720.
[29] P.F. Sturm, Mixing catadioptric and perspective cameras, in: Proceedings of OMNIVIS'2002 – Workshop on Omnidirectional Vision, 2002.
[30] P.F. Sturm, S. Ramalingam, J.-P. Tardif, S. Gasparini, J. Barreto, Camera models and fundamental concepts used in geometric computer vision, Found. Trends Comput. Graph. Vision 6 (2010) 1–183.
[31] T. Svoboda, D. Martinec, T. Pajdla, A convenient multi-camera self-calibration for virtual environments, PRESENCE: Teleoperat. Virtual Environ. 14 (4) (2005) 407–422.
[32] J.P. Tardif, P.F. Sturm, S. Roy, Self-calibration of a general radially symmetric distortion model, in: Proceedings of the 9th European Conference on Computer Vision, 2006.
[33] S. Thirthala, M. Pollefeys, The radial trifocal tensor: a tool for calibrating radial 1150 distortion of wide-angle cameras, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 321–328.
[34] B. Triggs, Autocalibration from planar scenes, in: Proceedings of European Conference on Computer Vision, 1998, pp. 89–105.
[35] X. Ying, Z. Hu, Catadioptric camera calibration using geometric invariants, IEEE Trans. Pattern Anal. Mach. Intell. 26 (10) (2004) 1260–1271.
[36] X. Ying, Z. Hu, Spherical objects based motion estimation for catadioptric cameras, in: Proceedings of International Conference on Pattern Recognition, 2004, pp. 231–234.
[37] L. Wang, F. Wu, Z. Hu, Multi-camera calibration with one-dimensional object under general motions, in: Proceedings of International Conference on Computer Vision, 2007, pp.1–7.
[38] F. Wu, F. Duan, Z. Hu, Y. Wu, A new linear algorithm for calibrating central catadioptric cameras, Pattern Recogn. 41 (2008) 3166–3172.
[39] Y. Wu, Y. Li, Z. Hu, Easy calibration for para-catadioptric-like camera, in: Proceedings of IEEE International Conference on Intelligent Robots and Systems(IROS), 2006, pp. 5719–5724.