Automatic Spherical Panorama Generation with Two Fisheye Images

Xiaoming Deng, Fuchao Wu, Yihong Wu, and Chongwei Wan National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, P. O. Box 2728, Beijing 100190, P. R. China Email: idengxm@gmail.com

Abstract-Panoramas are often used to enhance immersive experiences. Former panorama generation methods with fisheve images either require pre-calibrating the rotation of fisheye cameras with a calibration rig or manually selecting correspondences between images. In this paper, we propose an automatic spherical panorama generation method with two fisheve images. The method contains fisheye camera calibration and panorama stitching. In the calibration, the field of view and circular boundary are used to initialize the focal length, and line perspective constraint on sphere is used to refine the intrinsic parameters. In the panorama stitching, we use Scale Invariant Feature Transform (SIFT) features and RANSAC to get the fisheye correspondences. Then the correspondences are used to refine the estimation of rotation and intrinsic parameters. Finally, a spherical panorama is obtained by warping the fisheye images. In the panorama stitching process, no user input is needed. Experimental results on both simulated and real images show that our method is effective.

I. INTRODUCTION

Spherical panorama can be widely used in virtual reality, immersive display and computer game environment settings [15][6]. Since fisheye cameras have the prominent advantage of wide field of view (FOV), they can greatly decrease the number of images used for panorama generation. For example, the FOV of a fisheye camera by FC-E8 fisheye lens and Nikon CoolPix 990 is 183°, two fisheye images are enough for panorama generation. In this paper, we give calibration and fully automatic panorama generation with fisheye images.

Brown and Lowe [4] proposed a fully automatic construction of panoramas with perspective images. They use object recognition techniques based on invariant local features to select matching images, and a probability model for verification. Panoramic generation with perspective images requires a large number of images. Since the image matching has quadratic time complexity with the image number, a full matching process is very time-consuming. Uyttendaele et al [15] developed an interactive exploration of real world environments. They arranged six wide-angle video cameras around a pentagon, and data is acquired by a helmet with the cameras mounted on top. The method can let users have immersive and interactive experience. Since the used data is video instead of images, this system requires a portable storage with rather great capacity. Bakstein and Pajdla [2] proposed a panoramic mosaic technique with two fisheye images. The mosaicing is obtained by rotating a fisheye camera, and the mosaicing requires manually selecting correspondences in fisheye images. Li [11] reported

a method to realize a spherical camera with two fisheye lens. The spherical field of view is divided into two hemispherical ones. Each of the hemispherical views is observed by a fisheye lens. The two cameras are fixed, and the rotation between them is pre-calibrated with a 3D calibration object.

In this paper, we propose an automatic spherical panorama generation method with two fisheye images. The method consists of calibration and stitching. The fisheye imaging model is supposed as a central catadioptric imaging model approximately like in [16][19]. We use the field of view and circular boundary to initialize the focal length, and refine intrinsic parameters by the line perspective constraint on sphere. With the calibrated fisheye camera, we propose an automatic panorama stitching algorithm with two fisheye images by generalizing the work in [4]. With Scale Invariant Feature Transform (SIFT) features and RANSAC based outlier removal, our stitching algorithm makes requirement for neither manually selecting correspondences between fisheye images nor pre-calibration of camera rotation.

This paper is organized as follows. In Section 2, the imaging model is introduced, and our calibration algorithm is proposed. In Section 3, we report a panorama stitching algorithm with two fisheye images. Experiments are shown in Section 4, and Section 5 concludes this paper.

II. IMAGING MODEL AND CALIBRATION

A. Imaging Model

For many computer vision tasks, the relationship between the pixels in the image and the 3D rays entering the camera has to be known. In order to find this relationship, a suitable imaging model has to be used. Although several imaging model for fisheye camera are proposed, the exact imaging model of fisheye camera is still unknown. Ying and Hu [19] proposed that catadioptric cameras and fisheve cameras can be within a unified imaging model. Central catadioptric cameras and fisheye cameras can be classified by the model's characteristic parameter, and a fisheye image can be regarded like a central catadioptric image. An important merit of that unified model is that existing calibration methods for central catadioptric cameras can be directly applied to some fisheye cameras, if rigorous accuracy is not required. Therefore, we propose a calibration method of central catadioptric camera, and then apply it to a fisheye camera.

In this paper, the calibration is based on the generalized image formation model for central catadioptric cameras proposed by Geyer and Daniilidis [8]. The image formation is shown in Fig.1(a): a 3D point \mathbf{M} is projected to a point \mathbf{M}_s on a unit sphere centered at the viewpoint \mathbf{O} , then projected to a point \mathbf{m} on the image plane $\mathbf{\Pi}$ by a pinhole camera through the perspective center \mathbf{O}_c . The image plane is perpendicular to the line defined by the viewpoints \mathbf{O} and \mathbf{O}_c . The image formation process can be explicitly expressed as [16]

$$\mathbf{m} \approx \mathbf{K} \left(\frac{\mathbf{R}\mathbf{M} + \mathbf{t}}{\|\mathbf{R}\mathbf{M} + \mathbf{t}\|} + [0, 0, \xi]^T \right), \mathbf{K} = \begin{bmatrix} rf & s & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$
(1)

where \mathbf{R} is the rotation matrix and \mathbf{t} is the translation vector, and \mathbf{K} is the camera intrinsic matrix, with f the focal length, r the aspect ratio, s the skew factor and $[u_0, v_0]$ the principal point. As illustrated in Fig. 1(a), the principal point $\mathbf{p} = [u_0, v_0, 1]^T$ is the intersecting point of the optical axis \mathbf{OO}_c with the image plane $\mathbf{\Pi}$. $\xi = \|\mathbf{O} - \mathbf{O}_c\|$ is commonly called the mirror parameter. The mirror is a paraboloid if $\xi = 1$, an ellipsoid or hyperboloid if $0 < \xi < 1$, and a plane if $\xi = 0$.

In recent years, various calibration methods for the central catadioptric cameras have been proposed. Geyer and Daniilidis [9], Barreto and Araujo [3] and Deng et al. [5] proposed a linear calibration method for central catadioptric cameras with line images. Ying and Hu [17] used the geometric properties of a sphere to calibrate central catadioptric cameras. Wu et. al. [16] introduced a transformation from the central catadioptric model to the pinhole model and then used it to calibrate paracatadioptric-like cameras. Kang [10], Mincusik and Pajdla [13] gave self-calibration methods for omnidirectional cameras. In the following, we propose a practical and efficient calibration for central catadioptric as well as fisheye cameras.

B. Calibration Algorithm

In this paper, we assume both the mirror parameter ξ and the field of view (FOV) are known, which are usually provided by manufacturers. Our method is also suitable for some fisheye cameras by setting ξ as a constant (e.g. $\xi=1.0$) [19], and the calibration is capable of the applications like panorama generation.

Initialization of Intrinsic Parameters In our calibration method, we use the center of mirror boundary image (or circular boundary of fisheye image) as the principal point, and initialize the skew factor and aspect ratio to 0 and 1 respectively by prior.

As shown in Fig.1(b), \mathbf{m} is a point on the mirror boundary image and FOV of the catadioptric camera is 2ϕ . With known ξ and FOV, we give the following proposition to initialize the focal length f [6].

Proposition If the skew factor and aspect ratio of the camera are 0 and 1 respectively, the focal length can be determined

with a point [u, v] on the boundary image as

$$f = \sqrt{\frac{(u - u_0)^2 + (v - v_0)^2}{\eta - 1}}$$
 (2)

where $\eta = \frac{2\xi\cos\phi + \xi^2 + 1}{(\cos\phi + \xi)^2}$, $\phi = \frac{\text{FOV}}{2}$.

Nonlinear Optimization The initialized intrinsic parameters in Section 2.1 are optimized with the line perspective constraints on sphere (i.e. *the line projection on unit sphere is a great circle*) [18]. Therefore, the refinement of intrinsic parameters is to find intrinsic parameters to minimize the sum of squares of distances from the back-projected spherical points to the corresponding planes fitted by them.

Let N be the number of selected line images, $N_j(j = 1, 2, ..., N)$ the number of sample points on the jth line image, $\mathbf{m}_{i,j}(i = 1, 2, ..., N_j)$ the ith sample point on the jth line image, and $\mathbf{M}_{i,j}^s$ the back-projected point on the unit sphere for $\mathbf{m}_{i,j}$.

Based on the image formation process, we have

$$\mathbf{M}_{i,j}^{s} = \lambda_{i,j} \mathbf{K}^{-1} \mathbf{m}_{i,j} - \mathbf{L}$$
 (3)

where $\mathbf{L} = [0,0,\xi]^T$. Since $\|\mathbf{M}_{i,j}^s\|_2 = 1$ and $\mathbf{M}_{i,j}^s$ is in front of the perspective center \mathbf{O}_c , i. e. $\lambda_{i,j} > 0$, we can obtain

$$\lambda_{i,j} = \frac{\xi + \sqrt{\xi^2 - \mathbf{m}_{i,j}^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{m}_{i,j} (\xi^2 - 1)}}{\mathbf{m}_{i,j}^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{m}_{i,j}}$$
(4)

From (3) and (4), if a sample point $\mathbf{m}_{i,j}$ is given, the back-projected 3D point $\mathbf{M}_{i,j}^s$ on a unit sphere can be computed.

With the line perspective constraint on sphere [18], $\mathbf{M}_{i,j}^s(i=1,...,N_j)$ should be on a great circle. Let \mathbf{n}_j be the normal of the plane containing the great circle, \mathbf{n}_j and $\mathbf{M}_{i,j}^s$ satisfy $\mathbf{n}_j^T \mathbf{M}_{i,j}^s = 0 (i=1,...,N_j)$. With at least two points on the unit sphere, \mathbf{n}_j can be estimated by the following equation

$$\mathbf{A}_i \mathbf{n}_i = 0 \tag{5}$$

where $\mathbf{A}_j = [\mathbf{M}_{1,j}^s, \mathbf{M}_{2,j}^s, ..., \mathbf{M}_{N_j,j}^s]^T$. Doing the singular value decomposition for \mathbf{A}_j , we can obtain the solution of \mathbf{n}_j , which is corresponding to the unit right singular vector with the smallest singular value.

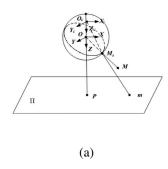
In order to refine \mathbf{K} , we minimize the following cost function

$$\varsigma(\mathbf{K}) = \sum_{i,j} (\mathbf{n}_j^T \mathbf{M}_{i,j}^s)^2$$
 (6)

Here, $\varsigma(\mathbf{K})$ is the sum of square distances between the given points $\mathbf{M}_{i,j}^s (i=1,...,N_j)$ and the fitted plane with the normal $\mathbf{n}_i (j=1,...,N)$.

Since the initial estimation of the principal point is always very close to the true value, the principle point can be fixed during the optimization. Therefore, the cost function (6) is with respect to the intrinsic parameters f, r, s. It can be solved by nonlinear optimization techniques, such as the Levenberg-Marquardt algorithm.

Implementation of Our Calibration Algorithm:



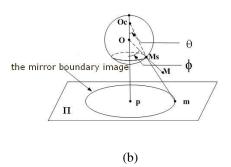


Fig. 1. (a) Image formation of a central catadioptric camera. (b)The field of view (FOV) of a central catadioptric camera.

- 1) Use the image of mirror boundary (or circular boundary of a fisheye image) and field of view to initialize the intrinsic parameters by (2);
- 2) Select the points $\mathbf{m}_{i,j} (i = 1, ..., N_j)$ on the jth line image j = 1, ..., N;
- 3) Minimize the cost function (6) with Levenberg-Marquardt algorithm. In each iteration, compute the back-projected points $\mathbf{M}_{i,j}^s$ on the unit sphere from $\mathbf{m}_{i,j}$ by (3) using the current intrinsic parameters, and get \mathbf{n}_j with (5).

Next, we propose a panorama generation algorithm with two fisheye images, which is inspired by the effective panorama generation algorithm from a large number of unordered perspective images in [4].

III. AUTOMATIC PANORAMA GENERATION ALGORITHM

In order to generate panorama, the calibrated fisheye camera should be rotated nearly oppositely about it optical center. The key problems for automatic panorama generation are robust image matching and rotation estimation. Since Scale Invariant Feature Transform (SIFT) is geometrically invariant under similarity transforms and invariant under affine changes in intensity, we adopt it to get potentially correspondences between two fisheye images. Although SIFT features are not distortion invariant, many panorama experiments with both indoor and outdoor images show its robustness.

For potentially correspondences in two images, we have a set of feature matches that are geometrically inconsistent (outliers). In order to remove outliers, we use an effective robust estimate technique – RANSAC [7]. Given a correspondence $\mathbf{m}_i^1 \leftrightarrow \mathbf{m}_i^2(\mathbf{m}_i^1, \mathbf{m}_i^2)$ are the ith corresponding inliers in the 1st and 2nd images), $\{\tilde{\mathbf{M}}_{i,1}^s, i=1,2,3,...\}$, $\{\tilde{\mathbf{M}}_{i,2}^s, i=1,2,3,...\}$ are their back-reprojections on unit sphere by (3). Since the geometrically consistent inliers satisfy $\tilde{\mathbf{M}}_{i,2}^s = \hat{\mathbf{R}} \mathbf{M}_{i,1}^s$ ($\hat{\mathbf{R}}$ is a 3D rotation matrix), the rotation matrix can be computed with the rotation estimation method between two sets of 3D inlier points [1].

Next, we refine the rotation $\hat{\mathbf{R}}$ and intrinsic parameters \mathbf{K} by nonlinear optimization. This step is of great importance for eliminating the structure misalignments in overlapping regions. The residue is defined as $||\mathbf{m}_i^1 - \hat{\mathbf{m}}_i(\mathbf{K}, \hat{\mathbf{R}}, \xi, \tilde{\mathbf{M}}_{i,2}^s)||$ where $\hat{\mathbf{m}}_i(\mathbf{K}, \hat{\mathbf{R}}, \xi, \tilde{\mathbf{M}}_{i,2}^s)$ is the projection of $\tilde{\mathbf{M}}_{i,2}^s$ onto the

lst fisheye image. $\hat{\mathbf{R}}$ and \mathbf{K} can be refined by minimizing the sum over all inliers of the residual error:

$$\min_{\hat{\mathbf{R}}, \mathbf{K}} \sum_{i=1}^{n} ||\mathbf{m}_{i}^{1} - \hat{\mathbf{m}}_{i}(\mathbf{K}, \hat{\mathbf{R}}, \xi, \tilde{\mathbf{M}}_{i,2}^{s})||^{2}$$
 (7)

where $\hat{\mathbf{R}}$ is parameterized by the Rodrigues formula. The cost function (7) is optimized by Levenberg-Marquardt algorithm.

Implementation of Panorama Generation Algorithm:

Input: Two nearly oppositely-rotated fisheye images, ${\bf K}$ and ξ

- 1) Extract SIFT features and get the matches [12] from the two fisheye images, and get two sets of back-projected 3D points on a unit sphere with (3);
- 2) Select geometric consistent inliers using RANSAC [7] from the two 3D point sets obtained in Step 1, and estimate the rotation $\hat{\mathbf{R}}$ between the two 3D inlier sets;
- 3) Refine $\hat{\mathbf{R}}$ and \mathbf{K} by minimizing (7);
- 4) For each pixel in the spherical map [14], compute its corresponding 3D point M^s on the unit sphere;
- 5) For each \mathbf{M}^s , determine its mapping into a fisheye image by $\mathbf{m} \approx \mathbf{K}(\tilde{\mathbf{R}}_i \mathbf{M}^s + [0, 0, \xi]^T)(i = 1, 2)$, where $\tilde{\mathbf{R}}_1 = \text{diag}(1, 1, 1)$ and $\tilde{\mathbf{R}}_2 = \hat{\mathbf{R}}$;
- 6) Form a blended image from the above warped images.

Output: A spherical panoramic image.

Remarks

- 1. If the camera is strictly rotated oppositely with respect to its optical center, there is no need to use this algorithm to find corresponding inliers between two images and merge them together. However, this setting is quite hard for mechanical designing. Our algorithm only requires that the images are nearly opposite-rotated, which can be attained easily for a hand-held camera with a tripod.
- 2. Panoramic generation with perspective images in [4] requires a large number of images, which is often a time-consuming matching process. The image matching in our algorithm is between only two images, therefore this algorithm is computationally efficient.

IV. EXPERIMENTS

A. Experiments on Simulated Data

In order to test the stability and accuracy of our calibration algorithm, we perform a number of simulated experiments. The settings of the simulated camera are

$$\mathbf{K} = \begin{bmatrix} 710 & 0.8 & 700 \\ 0 & 700 & 750 \\ 0 & 0 & 1 \end{bmatrix}, \xi = 0.966, \text{FOV} = 176.2^{\circ}. (8)$$

Since the FOV of a catadioptric camera or fisheye camera provided by manufactures is often an approximation, we deliberately choose FOV = 160° . We selected 5 line images, and on each line image 50 points are sampled. Gaussian noise with zero mean and σ standard deviation varying from 0.0 to 3.0 pixels at the step of 0.5 pixels is added to the sample points and the principal point. At each noise level, we perform 50 independent trials. Here, the RMS of f, r, s are demonstrated in Fig.2. It can be observed that, although the initial focal length with FOV = 160° deviates about 70 pixels from ground truth, our results after optimization are closer to ground truth than those with [3] as noise level increases. We think that it is because our algorithm use priors of the camera, while the algorithm in [3] don't require them.

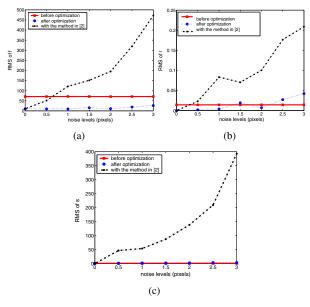


Fig. 2. RMS of the intrinsic parameters with our algorithm (before and after optimizations) and the algorithm [3].

B. Experiments on Real Images

The first experiment is carried out with a Nikon CoolPix 990 camera, combined with a hyperboloidal mirror. The mirror parameter ξ is 0.966 and the FOV provided by the manufacturer is 217.2°. Five images of 2048×1536 pixels are captured with the camera and one of them is shown in Fig. 3(a). In each image, we manually select 50 points on the longest line image. The calibration results with our algorithm are $f=503.3, r=1.09, s=8.8, u_0=1019.3, v_0=753.8$.

As a comparison, the results with the algorithm [3] are $f = 486.5, r = 1.01, s = -21.7, u_0 = 1112.2, v_0 = 755.4.$

The second experiment is with a Nikon CoolPix 990 camera mounted with a FC-E8 fisheye lens. The mirror parameter ξ can be regarded as 1.0 and the FOV is about 183° . Five images of 1024×768 pixels are captured with the camera and one of them is shown in Fig.3(d). In each image, we manually select 30 points on the longest line image. The calibration results are $f=349.2, r=0.982, s=0.0098, u_0=510.3, v_0=390.0$.

With our calibration results, the catadioptric and fisheye images are rectified as shown in Fig.3(b) and Fig.3(e). In both the images, heavy distortions are almost removed. The rectification with results by [3] is shown in Fig.3(c), where the distortions are still visible.

After the fisheye camera is calibrated, spherical panorama generation experiments with both indoor and outdoor images are carried out. Fig.4 and Fig.5 are typical operations of panorama generation algorithm. The fisheye images in Fig.4(a)(b) and Fig.5(a)(b) are taken with a tripod in a coffee room and a garden respectively, they are all obtained with a fisheye camera and a tripod, and no other equipments are required. Fig.4(c) and Fig.5(c) show the potential correspondences between input images, Fig.4(d) and Fig.5(d) are the geometrically consistent inliers by RANSAC, which are all correct. Fig.4(e) and Fig.5(e) are the spherical panoramas, and no structure misalignments are in them. The results validate the proposed calibration and panorama generation algorithms.

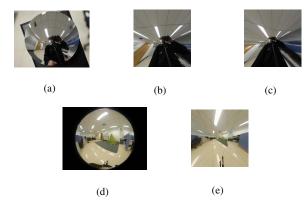


Fig. 3. (a) A catadioptric image. (b)(c) The rectified images with our algorithm and the algorithm [3]. (d) A fisheye image. (e) The rectified image with our algorithm. Here, we don't give the rectification with the results by [3], because the image of absolute conics in their algorithm is not positive definite and the intrinsic parameters can not be estimated.

V. CONCLUSIONS

We propose a calibration algorithm for fisheye and central catadioptric cameras with some reasonable priors, then give an automatic spherical panorama generation method with two fisheye images. For applications like image based walking-through over large-scale environments (such as culture heritages, museums etc.), panoramas at many hotspots are needed. Traditional panorama generation methods either require lots of perspective images or manually selecting correspondences between fisheye images, while our method needs only two nearly



Fig. 4. Indoor panorama generation with our algorithm. (a)(b) Two fisheye images of a coffee room. Since each image has a FOV of 183° and the camera undergoes an nearly opposite rotation, the images have a 3° overlapping region around the circular boundary. (c) The potential correspondences with SIFT features in the overlapping region (non-black region). (d) Corresponding inliers selected by RANSAC. (e) A $360^{\circ} \times 180^{\circ}$ panorama with our algorithm.

oppositely-rotated fisheye images at a hotspot, and no user inputs during panorama generation are required. Therefore, our method makes quick and easy installation of walking-through over large-scale environments possible.

ACKNOWLEDGMENT

The authors would like to thank National Natural Science Foundation of P. R. China (60575019), National Hi-Tech Research Development Program (863 Program) (2006AA01Z116) and the CASIA Innovation Fund for Young Scientists.

REFERENCES

- K. S. Arun, T. Huang, and S. D. Blostein: Least-squares fitting of two 3-D point sets. IEEE Trans. Pattern Anal. Mach. Intell. 9 (1987) 698–700
- [2] H. Bakstein and T. Pajdla: Panoramic mosaicing with a 180 degrees field of view lens. Workshop on Omnidirectional Vision (2002) 60–67
- [3] J. P. Barreto and H. Araujo: Geometry properties of central catadioptric line images and application in calibration. IEEE Trans. Pattern Anal. Mach. Intell. 27(2005) 1327–1333
- [4] M. Brown and D. G. Lowe: Recognising panoramas. proceedings of ICCV (2003) 1218–1225
- [5] X. Deng, F. Wu, F. Duan, Y. Wu and Z. Hu: Catadioptric camera calibration with one-dimensional object. Chinese Journal of Computers 30(2007) 737-747
- [6] X. Deng, F. Wu, and Y. Wu: An easy calibration method for central catadioptric cameras. Acta Automatica Sinica 33(2007) 801-808

- Fig. 5. Outdoor panorama generation with our algorithm. (a)(b) Two fisheye images of a garden. (c) The potential correspondences with SIFT features in the overlapping region (non-black region). (d) Corresponding inliers selected by RANSAC. (e) A $360^\circ \times 180^\circ$ panorama with our algorithm.
- [7] M. A. Fischler and R. C. Bolles: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Comm. of the ACM 22 (1981) 381–395
- [8] C. Geyer and K. Daniilidis: Catadioptric projective geometry. Int. J. Comput. Vis. 43 (2001) 223–243
- [9] C. Geyer and K. Daniilidis: Paracatadioptric camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 687–695
- [10] S. Kang: Catadioptric self-calibration Proceedings of CVPR (2000) 201– 207
- [11] S. Li: Full view spherical image camera. Proceedings of ICPR (2006) 386-390
- [12] D. G. Lowe: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60 (2004) 91–110
- [13] B. Mincuski and T. Pajdla: Structure from motion with wide circular field of view cameras. IEEE Trans. Pattern Anal. Mach. Intell. 28 (2006) 1135–1149
- [14] R. Szeliski and H. Y. Shum: Creating full view panoramic image mosaics and environment maps. SIGGRAPH (1997) 251–258
- [15] M. Uyttendaele, A. Criminisi, S.B. Kang, S. Winder, R. Szeliski, and R. Hartley: IEEE Computer Graphics and Applications 24 (2004) 52 – 63
- [16] Y. Wu, Y. Li, and Z. Hu: Easy calibration for para-catadioptric-like camera. IEEE Int. Intell. Robots and Systems. (2006) 5719-5724
- [17] X. Ying and Z. Hu: Catadioptric camera calibration using geometric invariants. IEEE Trans. Pattern Anal. Mach. Intell. 26 (2004) 1260–1271
- [18] X. Ying, Z. Hu, and H. Zha: Fisheye lenses calibration using straight-line spherical perspective projection constraint, Proceedings of ACCV (2006) 61–70
- [19] X. Ying, Z. Hu: Can we consider central catadioptric cameras and fisheye cameras within a unified imaging model? Proceedings of ECCV 1 (2004) 442–455