

3D Reconstruction by Single Camera Omnidirectional Multi-Stereo System

Shuya Chen¹, Zhiyu Xiang^{*2}, Nan Zou¹, Yiman Chen¹ and Chengyu Qiao¹

Abstract—Omnidirectional catadioptric systems are popular in robotic applications thanks to their large field of view. For 3D scene reconstruction in a single shot, usually two different catadioptric cameras are needed. More cameras may contribute to better reconstruction while larger mounting space and higher power cost are required. In this paper, a single camera multi-stereo catadioptric system with vertical and horizontal baseline structure is proposed. It features achieving multi-pair of central or non-central omnidirectional stereos in a compact manner. To make the 3D reconstruction process general and adaptive to various types of system configurations, a flexible calibration and reconstruction algorithm pipeline is presented. The algorithm features approximating the system into multiple central sub-cameras and carrying out the stereo matching in a spherical representation. In addition, an effective 3D point cloud fusion algorithm is proposed to optimize the reconstruction results from multiple stereo pairs. The experiment carried out with synthetic and real data verified the feasibility and effectiveness of our system.

I. INTRODUCTION

Stereo vision has always been an important and cost-effective sensing module for robotics. Traditional stereo systems are configured with two or more perspective cameras. To achieve reconstruction for wider areas, cameras with larger or omnidirectional field of view could be used.

The common methods to obtain large field of view are multi-camera image stitching [1], fish-eye cameras [2] and catadioptric cameras [3], [4]. Among them, with an extra mirror, catadioptric cameras own the advantage of higher flexibility for constructing various omnidirectional vision systems in a cost-effective way. The characteristics of catadioptric cameras could be changed and determined by the types of the mirror and camera used, as well as their spatial alignment. For omnidirectional stereo, two catadioptric cameras can be employed.

More cameras can construct multiple stereo pairs, which may contribute to better 3D reconstruction. However, more cameras also bring the problems like higher cost, synchronized data acquisition and larger mounting space, which can be infeasible for some applications. Constraining only one camera, multiple mirrors could be used for 3D reconstruction. Mirrors are mostly laid out in the form of a horizontal planar array [3], [4], which is good for optical field acquiring. However, the limited baseline between mirrors usually leads

¹Shuya Chen, Nan Zou, Yiman Chen and Chengyu Qiao are with College of Information and Electronic Engineering, Zhejiang University, Hangzhou, China

²Zhiyu Xiang is with Zhejiang Provincial Key Laboratory of Information Processing, Communication and Networking, Zhejiang University, Hangzhou, China

*Corresponding author, email: xiangzy@zju.edu.cn

to inferior 3D reconstruction. Using mirrors with different profile and size, vertical alignment of two mirrors along the optical axis of the camera is also presented [5]. Although vertical alignment has a longer baseline, using non-uniform mirrors increases the complexity of the system.

This paper introduces a novel single-camera catadioptric multi-stereo system. By arranging the principal and assistant mirrors in two layers, the design combines the merits of vertical and horizontal structures to achieve a longer baseline in a compact manner. Each assistant mirror and part of the principal mirror can construct a stereo pair. The system owns at least two stereo pairs for each 3D scene point. To make the 3D reconstruction process general and adaptive to various types of system configuration, a flexible calibration and reconstruction algorithm pipeline is presented. An effective multi-stereo optimization algorithm is also proposed to improve the quality of the reconstructed 3D point cloud. By means of simulation and real experiments using multiple central and non-central systems, the effectiveness of the system is verified.

The work is built on top of our previous works on omnidirectional 3D reconstruction [6], [7], [8]. The main contributions of the paper are threefold. First, a compact omnidirectional multi-stereo visual system based on a single camera is proposed. The design includes combinations of different mirrors and cameras that form central or non-central imaging systems, which enhances the universality of the system. Second, a general processing pipeline for system calibration and 3D reconstruction is given. Third, an effective multi-stereo optimization algorithm is proposed to improve the quality of the reconstructed 3D point cloud.

II. RELATED WORK

Depending on whether the system poses a single viewpoint, catadioptric cameras can be classified as central or non-central systems. As for the former, the projection relationship is simple and can be transformed into a central camera. Most of the classical catadioptric cameras lie in this category [4], [8]. However, because of the existence of misalignment between the camera and mirrors, in practice most of catadioptric systems are non-central, let alone those general combinations of the camera and mirror. Non-central systems have been increasingly studied in robotic and computer vision fields in recent years [3], [5], [9], [10], [11].

Similar to perspective system, two or more cameras can be used to construct omnidirectional stereo vision [9], [10]. However, using only one camera and multiple mirrors, it's also possible to reconstruct 3D scene with reflected image

of mirrors [3], [4], [5], [8], [10], [11]. Sagawa et al. [3] proposed a catadioptric system with a large mirror and seven small mirrors around. Caron et al. [4] utilized one camera equipped with four parabolic mirrors placed in a square to do the pose estimation. Besides horizontal layout, vertical arrangement of double-lobed mirror was also attempted in [5]. Ollis et al. [10] analyzed and compared five different physical configurations that using two vertically aligned mirrors. Yi et al. [11] proposed a single camera omnidirectional stereo system by using a hyperbolic mirror and a concave lens. This paper extended our previous work [8] and proposed a two-layer non-central multi-stereo system for 3D reconstruction.

Before reconstruction, optical systems have to be well calibrated by certain projection model. Geyer proposed a unified spherical projection model in [12] for central systems. Some modified calibration algorithms [13], [14] based on this model were also developed. For non-central systems the projection could be much more complex. Physical ray tracing for each incoming ray is usually employed in these models [15], [16]. In black box model [15] the calibration problem was converted to establishing the correspondence between individual image pixels and their 3D rays. Agrawal et al. [16] derived the forward and backward projection equations of the separate model for different mirrors, and solved a complex high-order equation for each 3D point in space. Xiang et al. [6] improved the unified projection model proposed by Geyer [12] and proposed a generalized unified model (GUM) for non-central systems caused by misalignment. They also developed the off-center spherical projection (OCSP) to deal with the non-central spherical mirror array [7].

With the calibrated projection model, stereo matching can be carried out. If only sparse 3D reconstruction is required, it would be easier since stereo matching is limited on distinct feature points [7]. For dense stereo reconstruction, it generally consists several steps like image rectification, stereo matching and reconstruction [17]. In omnidirectional stereo images, the image representation could be either cylindrical [18] or spherical [19]. For pure vertical alignment of the stereo, the epipolar-lines are vertical lines on the cylindrical image [18]. Spherical representation is more universal to general alignments and it can be further transformed to longitude-latitude representation for stereo rectification [19]. Upon reasonable rectifications, various stereo matching algorithms could be applied [20], [21].

As for multi-camera stereos, the algorithms of stereo fusion can be generally classified into three categories [22], i.e., voxel based methods [23], feature expansion methods [24] and depth maps based methods [25]. Voxel based methods [23] subdivide the entire 3D space into small voxels and remove inconsistent voxels with multi-view projection constraints from the initial volume. Feature expansion methods [24] first detect and match features in multiple images, then use an expansion algorithm to complete the dense reconstruction. The depth maps methods [25] utilize the consistency constraints between the maps to fuse the reconstruction results.

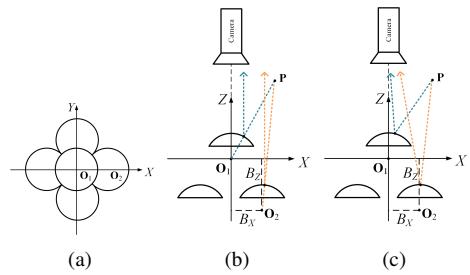


Fig. 1. General structure of the designed single camera omnidirectional multi-stereo system. (a) Top view of system. (b) Side view of orthographic camera with parabolic mirrors. (c) Side view of perspective camera with parabolic, hyperbolic or spherical mirrors. \mathbf{P} and the colored dashed lines represent a 3D object point and its incident or reflective rays into the camera respectively. Configuration (b) is a central system while (c) is non-central.

III. IMAGING SYSTEM STRUCTURE

The general structure of the proposed multi-stereo system is shown in Fig. 1. It consists of one camera and five mirrors with horizontal and vertical layouts. The top mirror with focus \mathbf{O}_1 is the principal mirror and the rest are the sub-mirrors. The sub-mirrors are placed symmetrically in a plane perpendicular to the optical axis of the principal mirror and the camera. The reflection image of five mirrors can be acquired in a single shot by the camera. Taking the reference frame at $\mathbf{O}_1 - XYZ$, the relative position of a sub-mirror \mathbf{O}_2 can be denoted by $\mathbf{T} = [B_X, 0, -B_Z]^T$, where B_X and B_Z are the horizontal and vertical baselines of the system, respectively.

The advantages of our system design lie in three aspects: (1) The principal mirror and four sub-mirrors constitute a four-pair binocular stereo vision system. Each object in the scene can be viewed by at least two stereo pairs, providing the possibility of achieving higher reconstructing accuracy by fusing stereo pairs; (2) The special layout between the principal and sub-mirrors achieves longer stereo baselines in a compact manner, comparing with the common pure horizontal [3], [4] or vertical layout [5], [9], [10]; (3) The type selection of the mirror and camera is flexible. Unlike in the traditional central catadioptric system, where only limited combination of camera and mirror type could be used, our system largely releases the constraint. The system can be built by central or non-central configurations, as shown in Fig. 1. An orthographic camera looking at five parabolic mirrors could be regarded as five different central cameras. A perspective camera looking at several parabolic, hyperbolic or spherical mirrors could be regarded as several non-central cameras. By carefully system modeling, the 3D reconstructing process can be unified and simplified.

IV. SYSTEM IMAGING MODEL AND CALIBRATION

The multi-stereo vision system may be configured by different types of camera and mirrors. To make the 3D reconstruction process uniform and effective to all types of configurations, a general imaging and calibration model is necessary. After calibration, the parameters could be further optimized by jointly calibrating the whole array.

A. Virtual Sub-camera Model

We adopted our Generalized Unified Model (GUM) [6] to describe the imaging process for each virtual sub-camera. This model is suitable not only for central cameras, but also for lots of non-central systems. By treating each mirror and its corresponding sub-image as a virtual camera, the whole system can be regarded as a composition of multiple virtual cameras. For detailed information please refer to [6]. With a series of calibration images of the checkerboard, the intrinsic and extrinsic parameters of each sub-camera can be calculated separately.

B. Joint Calibration of Multiple Mirror Positions

In the concept of virtual camera array [7], mirror parameters are integrated into the virtual sub-camera, and the relative position between mirrors is transformed into the rigid body transformation between the virtual sub-cameras. After calibration of each virtual sub-camera independently, the relative positions among them should be jointly optimized to improve the subsequent stereo reconstruction process.

Let c_1 be the reference coordinates of the principal camera, the rigid body transformation of $c_k (k = 2, 3, 4, 5)$ relative to c_1 can be represented by $\mathbf{T}_{c_k}^{c_1}$:

$$\mathbf{T}_{c_k}^{c_1} = \begin{bmatrix} \mathbf{R}_{3 \times 3} & \mathbf{T}_{3 \times 1} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}, k = 2, 3, 4, 5 \quad (1)$$

Given each 3D point \mathbf{X}_{ij}^w in world coordinate system and its corresponding imaging pixel \mathbf{p}_{ij}^1 in principal camera and \mathbf{p}_{ij}^k in k th camera, the procedure of calculating the reprojection error is as follows:

- 1) The world point \mathbf{X}_{ij}^w is first transformed to $\mathbf{X}_{ij}^{c_1}$ in c_1 coordinate system by using $\mathbf{T}_{c_1}^w$;

$$\mathbf{X}_{ij}^{c_1} = (\mathbf{T}_{c_1}^w)^{-1} \times \mathbf{X}_{ij}^w \quad (2)$$

- 2) $\mathbf{X}_{ij}^{c_1}$ is then transformed to $\mathbf{X}_{ij}^{c_k}$ with the matrix $\mathbf{T}_{c_k}^{c_1}$;

$$\mathbf{X}_{ij}^{c_k} = (\mathbf{T}_{c_k}^{c_1})^{-1} \times \mathbf{X}_{ij}^{c_1} \quad (3)$$

- 3) $\mathbf{X}_{ij}^{c_k}$ converts to reprojection pixel coordinate $\mathbf{p}_{ij}^{k'}$ with the intrinsic parameters \mathbf{K}_k of k th sub-camera;

$$\mathbf{p}_{ij}^{k'} = \mathbf{K}_k \times \mathbf{X}_{ij}^{c_k} \quad (4)$$

- 4) Calculate the reprojection error $\|\mathbf{p}_{ij}^{k'} - \mathbf{p}_{ij}^k\|^2$.

Let function G represent the entire procedure of getting $\mathbf{p}_{ij}^{k'}$ from \mathbf{X}_{ij}^w , then the optimal rigid body transformation $\mathbf{T}_{c_k}^{c_1}$ can be computed by minimizing the reprojection error shown in (5):

$$\arg \min_{\mathbf{T}_{c_k}^{c_1}} \sum_{i,j} \|G(\mathbf{K}_k, \mathbf{T}_{c_1}^w, \mathbf{T}_{c_k}^{c_1}, \mathbf{X}_{ij}^w) - \mathbf{p}_{ij}^k\|^2 \quad (5)$$

Since each virtual camera has been calibrated in Section IV-A, the initial values of parameters in G have been obtained. By using non-linear optimization algorithms such as Levenberg-Marquardt algorithm, (5) can be well solved.

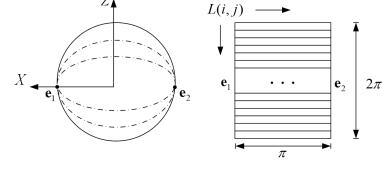


Fig. 2. Longitude-latitude representation of spherical image.

V. SPHERICAL RECTIFICATION AND DISPARITY COMPUTING

In order to achieve dense 3D reconstruction efficiently, the original image of each virtual camera has to be rectified to make the epipolar lines parallel with each other. Considering the omnidirectional nature of the proposed multi-stereo system, rectification and disparity computing are carried out in a spherical manner.

A. Spherical Epipolar Rectification

In our multi-stereo system, the principal mirror and one sub-mirror construct one stereo pair. Traditional rectification method for perspective images is not suitable for omnidirectional cameras with large field of view. Since the spherical model (GUM) is adopted for each virtual camera in Section IV-A, it is natural to carry out rectification in spherical image, where the rectified epipolar lines are big circles in the sphere. In the new rectified frame, the X -axis is defined along the baseline between two spherical sub-cameras.

In order to simplify the calculation, the longitude-latitude representation of the spherical image is adopted [19] which transforms the disparity search on a 3D arc to the search on a one-dimensional parallel line, as shown in Fig. 2. Epipoles e_1 and e_2 are equivalent to north and south poles on the earth, then epipolar lines can be considered as longitude lines. After transforming the image to longitude-latitude representation, the epipolar circles become parallel lines.

B. Spherical Disparity and 3D Reconstruction

With the longitude-latitude representation of the pixels, traditional dense disparity matching algorithms could be used. Semi-global matching (SGM) [26] is adopted for its good performance in most of the stereo applications. 3D reconstruction can be obtained from the spherical disparity. A profile example of epipolar plane after rectification is shown in Fig. 3.

Normalized spherical disparity d_n at point \mathbf{p}_r is defined as the difference between arc length α_l and α_r starting from the positive X -axis.

$$d_n = \alpha_l - \alpha_r = \theta_l - \theta_r \quad (6)$$

It is obvious that the normalized spherical disparity d_n is equivalent to angle γ in Fig. 3. Based on this condition, the distance of the spatial point \mathbf{P} in two sub-cameras can

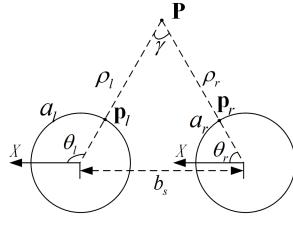


Fig. 3. Definition of spherical disparity. X -axes of the two sub-cameras are coaxial with baseline b_s . The spatial point \mathbf{P} corresponds to points \mathbf{p}_l and \mathbf{p}_r on the sphere, and the polar angles of these spherical points are θ_l and θ_r , respectively.

be computed. Then the coordinate of \mathbf{P} can be obtained by either ρ_l or ρ_r in Fig. 3, where

$$\rho_l = b_s \frac{\sin(\theta_r)}{\sin(\gamma)} \quad (7)$$

$$\rho_r = b_s \frac{\sin(\pi - \theta_l)}{\sin(\gamma)} \quad (8)$$

VI. MULTI-STEREO FUSION AND OPTIMIZATION

In our multi-stereo system, each object can at least be observed by three image patches, i.e., one image patch from top principal mirror and the other two from two mirrors in the bottom, which forms at least two stereo pairs. With the limited image resolution and non-central distortion nature of the system, position errors exist in the reconstructed 3D point clouds. Therefore, it's necessary to fuse and optimize the 3D point clouds produced by different stereo pairs and obtain an improved quality for reconstruction results.

We propose an effective algorithm for multi-stereo fusion and optimization. It is composed of two parts: outlier filtering and point cloud fusion.

A. Outlier Filtering

Some outliers may exist in disparity image for each stereo pair. They can be removed by consistency checking between stereo pairs. Given an image point \mathbf{p} in the principal camera, and its corresponding reconstructed 3D point by one stereo pair, the virtual disparity of this 3D point in another stereo pair can be obtained. By calculating the difference between this virtual disparity and the real one, the 3D points with large errors can be eliminated. An example of this process is illustrated in Fig. 4.

In next step color error is adopted as the consistency clue to further remove the outliers. Similar to the previous step, we first get the matching points \mathbf{p}_1 and \mathbf{p}_2 in sub-cameras corresponding to point \mathbf{p} in principal camera C_1 , as shown in Fig. 4. Then we calculate the color square error of points \mathbf{p}_1 and \mathbf{p}_2 within 3×3 neighboring image window. The less the error is, the higher the consistency. The points with large color errors are then removed.

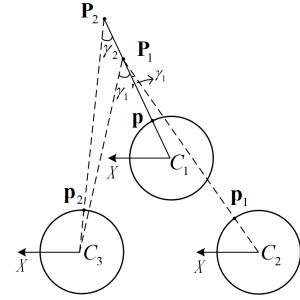


Fig. 4. An example of consistency checking between two stereo pairs that composed of the camera pair C_1-C_2 and C_1-C_3 respectively. For the same pixel point \mathbf{p} in principal camera C_1 , the stereo pair C_1-C_2 and C_1-C_3 have the disparity γ_1 and γ_2 , corresponding to spatial point \mathbf{P}_1 and \mathbf{P}_2 , respectively. The disparity consistency can be checked by computing the difference between disparity γ_2 and the virtual disparity transformed from γ_1 (denoted as γ'_1), given the point \mathbf{P}_1 and relative position of C_3 to C_1 .

B. Point Clouds Fusion

By the process of outlier filtering, the 3D points with large error have been removed. However, the point clouds produced from different stereos may remain separated because of the system error of the non-central system, as shown in Fig. 9(b). To improve the quality of final result, the point clouds have to be fused. Given two sets of point clouds in the frame of principal camera, the direct way of point fusion is averaging the pair of 3D point produced by different stereos. The advantage of this error smoothing method is its simplicity. However, this point-pair averaging method does not consider the global information of the point cloud. Furthermore, it requires the existence of matching pairs and is not applicable for those constructed only in one stereo.

We propose an effective point fusion and optimization method that can utilize the global information of the both raw point clouds. It first transforms the two separated point clouds into a “middle” place and then fuses it. The initial smoothed point cloud obtained by direct averaging of matching points is taken as the reference of “middle”. Because of the approximation of the central model and the non-linear nature of the system, the projection between the reference points and the original ones can not be represented by a simple rigid body transformation. For each point cloud, an affine transformation for 3D points, i.e., a general projection matrix \mathbf{A} and a translation vector \mathbf{T} , is used to transform the raw point cloud to the reference position. The transformation can be computed by minimizing the following loss function,

$$E(\mathbf{A}, \mathbf{T}) = \frac{1}{N} \sum_{j=1}^N \|\mathbf{M}_j - (\mathbf{A}_i \mathbf{S}_{ij} + \mathbf{T}_i)\| \quad (9)$$

where N is the number of matched point pairs, namely the number of reference points, \mathbf{M}_j represents the j th reference point, and \mathbf{S}_{ij} ($i = 1, 2$) represents the j th point in i th point cloud matching with \mathbf{M}_j .

The minimization of (9) can be effectively solved in two steps:

- 1) Based on the matching relationship between \mathbf{S}_i and

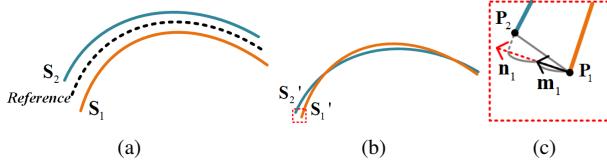


Fig. 5. Procedure of point clouds fusion. (a) The raw point clouds S_1 and S_2 and the obtained reference position. (b) Two projected point clouds after affine transformation. (c) Zoom in view of motion vector \mathbf{m}_1 for \mathbf{P}_1 in the local window of (b).

the reference point cloud, singular value decomposition (SVD) is used to solve the initial value of \mathbf{A}_i and \mathbf{T}_i .

- 2) The final solution of \mathbf{A}_i and \mathbf{T}_i can be obtained by minimizing the value of the cost function $E(\mathbf{A}, \mathbf{T})$ with Levenberg-Marquardt algorithm.

The affine transformation obtained in the above steps is then applied to the both raw point clouds respectively. The scope of the fused point cloud is consequently expanded. However, it only transforms the point clouds approximately to the reference region, and the two point clouds are still not completely integrated into a whole, as shown in Fig. 5(b). Therefore further operation is necessary to obtain a more accurate reconstruction results.

In this step, we carefully move 3D points in space according to the information of normal vector, as shown in Fig. 5(c). Given the transformed 3D point cloud S_1' , normal vector \mathbf{n}_1 of each point \mathbf{P}_1 is firstly computed. Then search for a matched point of \mathbf{P}_1 in point cloud S_2' . A matched point \mathbf{P}_2 is confirmed if the distance and the difference of normal direction between \mathbf{P}_1 and \mathbf{P}_2 are both less than a predefined threshold. Projecting the vector of $\overrightarrow{\mathbf{P}_1\mathbf{P}_2}$ on the direction of \mathbf{n}_1 , and taking half of the amplitude value as the motion vector (denoted \mathbf{m}_1) for \mathbf{P}_1 . After processing all of the 3D points in S_1' and S_2' , a more integrated point cloud considering the entire raw point clouds can be achieved.

VII. EXPERIMENTAL RESULTS

To verify the effectiveness of our single camera multi-stereo system, both synthetic and real data experiments were designed. Experiments based on synthetic data were carried out on two different configurations, which represent multi-central and multi-non-central system respectively. Then a real prototype system was built for further experiment. It is composed of spherical mirrors with a perspective camera, which is a more challenging non-central configuration.

A. Experiments on Synthetic Data

The synthetic data were produced in POV-Ray [27], a software for generating virtual scene images given a certain configuration. Two virtual systems were built, i.e., a multi-central system with an Orthographic camera looking at Parabolic mirrors (OP), and a multi-non-central system with a Perspective camera looking at Parabolic mirrors (PP), as shown in Fig. 1(b) and 1(c). Resolution of the camera was set as 1600×1200 . Other parameters of the system are: focal length [12] of parabolic $p = 40$ mm, mirror diameter $d = 50$ mm, baselines $B_X = 50$ mm and $B_Z = 80$ mm.

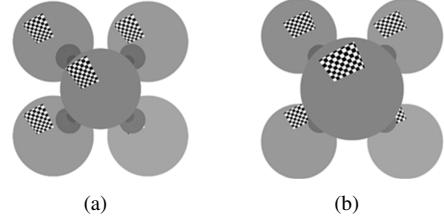


Fig. 6. Synthetic example images of checkerboards acquired by (a) Orthographic camera looking at Parabolic mirrors (OP) and (b) Perspective camera looking at Parabolic mirrors (PP). The bottom mirrors look smaller than the principal one in the perspective image because of the longer distance to the camera.

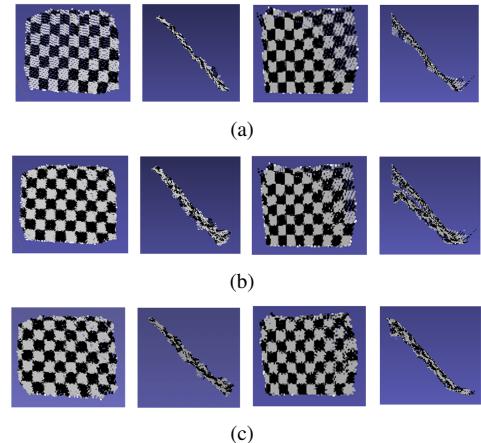


Fig. 7. 3D reconstruction of the synthetic multi-stereo system. (a) Results by one stereo pair, (b) Results by direct superposition of two point clouds and (c) Results by fusion of two point clouds. The columns from left to right: front, side view for the OP system, front, side view for the PP system, respectively.

Using a synthetic checkerboard, several images can be acquired by the virtual systems. Some examples of the images are shown in Fig. 6. Parts of these images were used to calibrate the systems, while the rest were used for accuracy verification of 3D reconstruction. During calibration individual GUM model was adopted for each sub-camera. To achieve a better consistency on relative position between sub-cameras, joint optimization for intrinsic and extrinsic parameters was applied, as introduced in Section IV-B. After spherical epipolar line rectification, SGM was applied for dense stereo matching. Finally 3D point clouds for each stereo pair can be computed by triangulation, as shown in Fig. 7.

As observed from Fig. 7(a), both systems can successfully reconstruct the checkerboard with acceptable planarity. For the non-central PP system, the planarity was a little bit worse than that of central OP system. This is due to the system model error by approximating the non-central system to the central one.

Then for each system, reconstructed point clouds from two stereo pairs were transformed into a global frame and displayed together, as shown in Fig. 7(b). Comparing with the results of one point cloud only in Fig. 7(a), obvious non-planar errors can be observed in Fig. 7(b). This is due to both

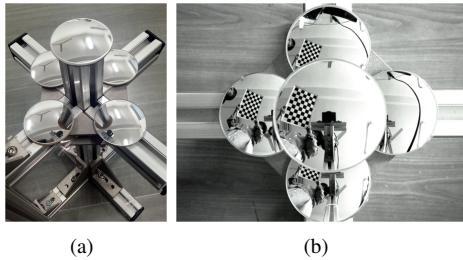


Fig. 8. The system assembling and the captured image. (a) The mirror layouts. (b) An example of the images viewing at a checkerboard.

the approximation of the system model and the calibration residual error of the different stereo.

Finally, the two point clouds of the system were fused by the algorithm described in Section VI. The results are shown in Fig. 7(c), where the two point clouds were transformed into a more unified whole. The entire planarity of the reconstructed plane was largely improved compared to Fig. 7(b). In addition, the basic color pattern was well reserved. The quantitative results are listed in Table I, where μ denotes the mean absolute distance error of reconstructed points to the fitted plane, and e denotes the relative percentage error of μ with respect to the distance of the plane.

In our previous work [8], we have validated that vertical stereo pair with longer baseline achieved better performance than horizontal pair. In this work, it is beneficial to obtain further improved results by fusing multiple vertical stereo pairs.

B. Experiments on Real Prototype System

1) *System structure*: To further verify the proposed idea, a more challenging multi-non-central system with a perspective camera viewing at spherical mirrors was built. In the design, five identical spherical mirrors with the radius of 120 mm and the diameter of the bottom circle 51 mm were assembled together under the condition of $B_X = 50$ mm and $B_Z = 80$ mm. An MV-CA030-10GC perspective camera produced by Hikvision with a resolution of 1920×1440 was employed. The mounting of the mirror array and an example of the captured images are shown in Fig. 8.

2) *3D reconstruction and optimization*: In this part, the checkerboard and one orthogonal trihedral object were employed to evaluate the 3D reconstruction performance. The initial 3D reconstruction by one stereo pair is shown in Fig. 9(a). For the results of checkerboard, some noises can be observed in edge area while the overall planarity was guaranteed. For the trihedral, some variation of the planes can be noticed in the side view.

The 3D point clouds obtained by different stereo pairs can be transformed to the frame of principal mirror and laid up together. As shown in Fig. 9(b), the two point clouds were separated in space and large drift error can be observed, especially for the trihedral. This is due to the relatively large non-central nature of the system. By approximating those non-central sub-cameras as approximated central ones, the 3D reconstruction process is greatly simplified. Fusion and

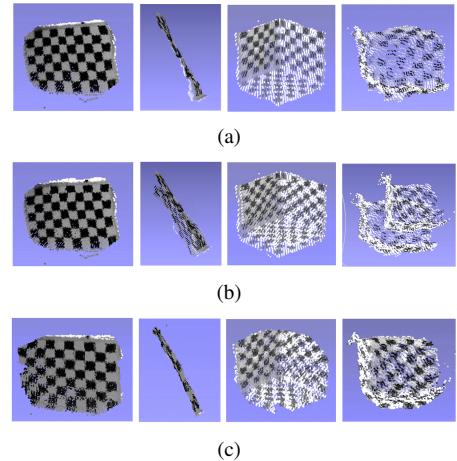


Fig. 9. 3D reconstruction of the multi-stereo system. (a) Results by one stereo pair, (b) Results by direct superposition of two point clouds and (c) Results by fusion of two point clouds. The columns from left to right: front, side view for the planar checkerboard, front, side view for the orthogonal trihedral, respectively.

optimization process were applied to decrease the drift error of the point clouds, as described in Section VI. The final point clouds were well integrated with much smaller drift error comparing with Fig. 9(b), as shown in Fig. 9(c).

The quantitative evaluation results are shown in Table II, where μ and e are still the absolute and percentage error respectively. For the trihedral object, the angles between normal vectors of adjacent planes $\theta_1, \theta_2, \theta_3$ were further computed after plane fitting. The average absolute angle error θ_e defined in (10) is adopted for quantitative evaluation.

$$\theta_e = \frac{|\theta_1 - 90^\circ| + |\theta_2 - 90^\circ| + |\theta_3 - 90^\circ|}{3} \quad (10)$$

For both cases of the checkerboard and trihedral, the fusion results of two point clouds were better than those of one stereo pair. In planar case the improvement was about 30% on the percentage error. Due to the system error of the imaging model and the disadvantageous viewpoint for some planes, accurate angle reconstruction is a very challenging task. In our experiments, the average angle error of the fusion result was about 7° , which was about 15% better than the original result.

VIII. CONCLUSION

A single camera omnidirectional multi-stereo catadioptric system is proposed in this paper. The system features reasonable stereo baselines in a compact spatial structure with the combination of vertical and horizontal mirror layouts. With the advantages of the multi-stereo, each object in the scene can at least be observed by two stereo pairs. This characteristic provides possibility for further 3D point cloud optimization for better accuracy. To make the 3D reconstruction process adaptive to various central or non-central system configurations, a general algorithm pipeline is presented. By approximating the non-central systems as central ones, calibration, disparity computation and 3D reconstruction can all be completed in a unified spherical representation. In

TABLE I
QUANTITATIVE EXPERIMENTAL RESULTS ON SYNTHETIC DATA.

	OP system		PP system	
	$\mu(\text{mm})$	$e(\%)$	$\mu(\text{mm})$	$e(\%)$
Reconstruction results of one stereo pair	4.70	0.52	8.31	0.90
Direct superposition of two point clouds	6.28	0.69	17.3	1.92
Fusion results of two point clouds	4.64	0.51	5.77	0.64

TABLE II
QUANTITATIVE EXPERIMENTAL RESULTS ON REAL PROTOTYPE SYSTEM.

	Checkerboard		Trihedral Object		
	$\mu(\text{mm})$	$e(\%)$	$\theta_1(\circ)$	$\theta_2(\circ)$	$\theta_3(\circ)$
Reconstruction results of one stereo pair	6.38	0.99	83.48	80.12	81.68
Fusion results of two point clouds	4.60	0.70	85.45	80.43	83.02

particular, an effective 3D point cloud fusion algorithm is proposed for reconstruction optimization. The experiments based on synthetic and more challenging real data verified the effectiveness of the system.

ACKNOWLEDGMENT

The work is supported by National Natural Science Foundation of China (NSFC) (61571390) and National Natural Science Foundation of China-Zhejiang Joint Fund for the Integration of Industrialization and Informatization (U1709214).

REFERENCES

- [1] M. Brown, D. G. Lowe, *et al.*, “Recognising panoramas,” in *Proceedings of the 9th International Conference on Computer Vision*, vol. 3. IEEE, 2003, p. 1218.
- [2] X. Deng, F. Wu, Y. Wu, and C. Wan, “Automatic spherical panorama generation with two fisheye images,” in *Proceedings of 7th World Congress on Intelligent Control and Automation*. IEEE, 2008, pp. 5955–5959.
- [3] R. Sagawa, N. Kurita, T. Echigo, and Y. Yagi, “Compound catadioptric stereo sensor for omnidirectional object detection,” in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3. IEEE, 2004, pp. 2612–2617.
- [4] G. Caron, E. Marchand, and E. M. Mouaddib, “3d model based pose estimation for omnidirectional stereovision,” in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 5228–5233.
- [5] D. Southwell, A. Basu, M. Fiala, and J. Reyda, “Panoramic stereo,” in *Proceedings of the 13th International Conference on Pattern Recognition*, vol. 1. IEEE, 1996, pp. 378–382.
- [6] Z. Xiang, X. Dai, and X. Gong, “Noncentral catadioptric camera calibration using a generalized unified model,” *Optics Letters*, vol. 38, no. 9, pp. 1367–1369, 2013.
- [7] Z. Xiang, Y. Zhou, and X. Gong, “A novel sub-camera array model for calibrating multi-mirror catadioptric systems,” *Measurement Science and Technology*, vol. 26, no. 8, p. 085402, 2015.
- [8] Z. Xiang, S. Chen, L. Luo, and N. Zou, “Compact omnidirectional multi-stereo vision system for 3d reconstruction,” *Applied Optics*, vol. 57, no. 34, pp. 9929–9935, 2018.
- [9] W. L. D. Lui and R. Jarvis, “Eye-full tower: A gpu-based variable multibaseline omnidirectional stereovision system with automatic baseline selection for outdoor mobile robot navigation,” *Robotics and Autonomous Systems*, vol. 58, no. 6, pp. 747–761, 2010.
- [10] M. Ollis, H. Herman, and S. Singh, *Analysis and design of panoramic stereo vision using equi-angular pixel cameras*. Carnegie Mellon University, The Robotics Institute, 1999.
- [11] S. Yi and N. Ahuja, “A novel omnidirectional stereo vision system with a single camera,” in *Proceedings of International Conference Image Analysis and Recognition*. Springer, 2006, pp. 146–156.
- [12] C. Geyer and K. Daniilidis, “A unifying theory for central panoramic systems and practical implications,” in *Proceedings of European Conference on Computer Vision*. Springer, 2000, pp. 445–461.
- [13] C. Mei and P. Rives, “Single view point omnidirectional camera calibration from planar grids,” in *Proceedings of IEEE International Conference on Robotics and Automation*. IEEE, 2007, pp. 3945–3950.
- [14] H. Duan and Y. Wu, “A calibration method for paracatadioptric camera from sphere images,” *Pattern Recognition Letters*, vol. 33, no. 6, pp. 677–684, 2012.
- [15] M. D. Grossberg and S. K. Nayar, “The raxel imaging model and ray-based calibration,” *International Journal of Computer Vision*, vol. 61, no. 2, pp. 119–137, 2005.
- [16] A. Agrawal, Y. Taguchi, and S. Ramalingam, “Analytical forward projection for axial non-central dioptric and catadioptric cameras,” in *Proceedings of European Conference on Computer Vision*. Springer, 2010, pp. 129–143.
- [17] D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 7–42, 2002.
- [18] L. He, C. Luo, F. Zhu, and Y. Hao, “Stereo matching and 3d reconstruction via an omnidirectional stereo sensor,” in *Motion Planning*. InTech, 2008.
- [19] S. Li, “Binocular spherical stereo,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 4, pp. 589–600, 2008.
- [20] G. Caron and E. M. Mouaddib, “Vertical line matching for omnidirectional stereovision images,” in *Proceedings of IEEE International Conference on Robotics and Automation*. IEEE, 2009, pp. 2787–2792.
- [21] H. Kim and A. Hilton, “3d scene reconstruction from multiple spherical stereo pairs,” *International Journal of Computer Vision*, vol. 104, no. 1, pp. 94–116, 2013.
- [22] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, “A comparison and evaluation of multi-view stereo reconstruction algorithms,” in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1. IEEE, 2006, pp. 519–528.
- [23] K. N. Kutulakos and S. M. Seitz, “A theory of shape by space carving,” *International Journal of Computer Vision*, vol. 38, no. 3, pp. 199–218, 2000.
- [24] Y. Furukawa and J. Ponce, “Accurate, dense, and robust multiview stereopsis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1362–1376, 2010.
- [25] P. Merrell, A. Akbarzadeh, L. Wang, P. Mordohai, J.-M. Frahm, R. Yang, D. Nistér, and M. Pollefeys, “Real-time visibility-based fusion of depth maps,” in *Proceedings of the 11th International Conference on Computer Vision*. IEEE, 2007, pp. 1–8.
- [26] H. Hirschmuller, “Stereo processing by semiglobal matching and mutual information,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 328–341, 2008.
- [27] Pov-ray – the persistence of vision raytracer. [Online]. Available: <http://www.povray.org/>