

# Monocular multi-view stereo imaging system

**W. Jiang**  
[jiangwei@ipc.zju.edu.cn](mailto:jiangwei@ipc.zju.edu.cn)

State Key Lab. of Industrial Control Technology, Zhejiang University, Hangzhou, Zhejiang 310027, China

**M. Shimizu**

College of Science and Technology, Nihon University, Chiyoda-ku, Tokyo 101-8308, Japan

**M. Okutomi**

Graduate School of Science and Engineering, Tokyo Institute of Technology, Meguro-ku, Tokyo, 152-8550, Japan

*This research was done while the authors were at the Tokyo Institute of Technology.*

In this study, we present a single-camera, multi-view stereo imaging system for capturing three-dimensional (3D) information. First, we design a monocular, multi-view stereo imaging device composed of a fisheye lens, and planar mirrors placed around the lens. The fisheye lens has a wide view-angle. The captured image includes a centered region of direct observation and surrounding regions of mirrored observations. These regions can be considered as images captured by multiple cameras at different positions and orientations. Therefore, the proposed device is equivalent to a synchronous multiple-cameras configuration, in which all the cameras share the same physical characteristics. In addition, we show how to place the mirrors in order to maximize the common view-angles, which is an important design consideration. Then, after calibrating the projection function of the fisheye lens, we obtain the positions and orientations of the virtual cameras from the external parameters. We also develop two multi-baseline stereo algorithms for the 3D measurement system. The first algorithm transforms the captured image to perspective images, and uses the traditional method to perform stereo determination. The second algorithm directly uses the original captured image along with an analysis of the epipolar geometry. Experimental results show that our system is more effective than traditional stereo methods that use a stereo pair, and it can achieve robust 3D reconstruction. [DOI: <http://dx.doi.org/10.2971/jeos.2011.11051>]

**Keywords:** monocular stereo imaging , multi-view stereo, catadioptric mirrors , wide-angle lens

## 1 INTRODUCTION

Range estimation, or three-dimensional (3D) shape measurement using a non-contact method, is a fundamental technique used in the fields of security, intelligent transport systems (ITS), and robotic navigation. This technique has been widely studied, and many commercially available products have been developed through its adaptation. Nonetheless, the field of study remains active.

Triangulation, which includes passive stereo vision and active stereo vision [6] [15], is a basic method of range estimation. The active stereo vision is commonly used to achieve dense shape reconstruction. For this technique, a special device (e.g. projector) is employed to emit special patterns onto the identified object, which will be detected by the camera. The active method is advantageous for robust and accurate 3D scene reconstruction, but is also expensive and noncompact. The passive stereo vision requires multiple synchronized cameras and observations from different camera positions. However, subtle differences in the camera system parameters such as focal length, gain, and exposure time are unavoidable when using multiple cameras, making it difficult to establish correspondences among these images. In addition, the synchronization mechanism among the multiple cameras can be complicated.

Therefore, there is a strong demand for the use of a single camera and a single image.

Image layering and divided field-of-view (FOV) methods are the two major methods used to capture images from different camera positions in a single shot. Both methods incorporate the multi-view information into a single image, and can be used for dynamic objects and real-time applications. As the image layering method, single-lens aperture [1], coded aperture [13], and reflection stereo [26] have been proposed. Nevertheless, the efficiency of the use of incident light is not good in the single-lens aperture and coded aperture methods. The baseline length for stereo measurement is also very narrow (approximately 10 mm) in all of the methods, and the results are limited to a low resolution of range. A combination of camera mirrors is often used in the divided FOV method. It is possible to obtain identical brightness, contrast, and color from the mirrored cameras at different positions, enabling simple matching between camera observations. Used together, these methods can provide a relatively long baseline length. However, due to the limits of resolution and the view-angle of the camera in all of these methods, the FOV of the camera is only

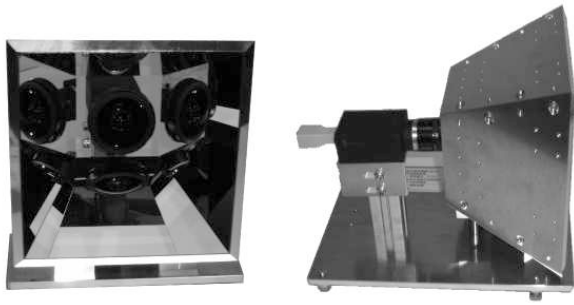


FIG. 1 Prototype of the monocular multi-view stereo imaging system.

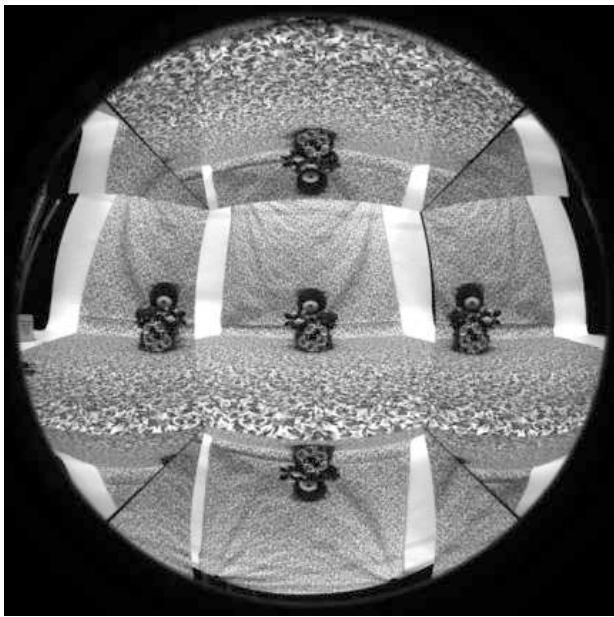


FIG. 2 Circular fisheye image captured by the proposed system. The image is divided into five regions :a center region projected by direct light, and four surrounding regions projected by light reflected from the mirrors.

divided into two views. Thus, two-view stereo matching suffers from the problems of image noise and occlusion.

In this study, we attempt to solve the above problems by proposing a monocular multi-view stereo imaging device, composed of a fisheye lens and planar mirrors that are placed around the lens. The fisheye lens has a wide view-angle, and the captured image includes a centered region of direct observation and surrounding regions of mirrored observations. These regions can be considered as views captured by multiple synchronized cameras at different positions and orientations. Therefore, the proposed device is equivalent to a multiple-cameras configuration in which the cameras all have the same physical characteristics. After calibrating the proposed device, we develop two multi-baseline stereo algorithms [22] for 3D measurement. Therefore, when compared to previous approaches [9] using two views from a single camera, our approach can reduce matching errors due to image noise and occlusions by employing multi-baseline stereo techniques with four image pairs.

## 1.1 Previous works of monocular stereo with mirrors

Several researchers have demonstrated the use of planar mirrors to acquire stereo data with a single camera. Mitsumoto *et al.* [19] presented an approach using one camera and a mirror for reconstructing the 3D shape of polyhedral objects. In their paper, they obtained a stereo image by imaging an object and its mirror reflection, using only a single camera. They reconstructed the 3D shape by finding corresponding points between the direct image and the mirrored image. For acquiring a stereo image pair in a single shot, Gosthasby and Gruver [11] proposed an approach with a single camera and two planar mirrors. Here, they placed the two planar mirrors symmetrically to the optical axis in front of the camera to produce two virtual cameras. However, aligning the mirrors with the camera is complicated, and it is also difficult to handle and relocate the system during observations, while at the same time preventing parts of the system from moving.

Based on the above single-camera stereo system using two plane mirrors, Gluckman and Nayar [10] developed a theory of stereo imaging with planar mirrors. They demonstrated how two mirrors in an arbitrary configuration can be self-calibrated and used for creating a single-camera stereo system. They also analyzed the geometry and calibration of catadioptric stereo with two planar mirrors, and assert that the catadioptric single-camera stereo, in addition to simplifying data acquisition, provides both geometric and radiometric advantages over the traditional two-camera stereo system. Epipolar geometry of catadioptric stereo systems has also been intensively studied [7] [28]. Another stereo system, using a single camera and four planar mirrors, was also proposed by Inaba *et al.* [14]. The system is a more mature architecture for two-view stereo imaging, and has been applied in stereo imaging techniques [2]. Other systems use curved mirrors [21] to reconstruct 3D information. These systems can provide wider *FOVs* than the planar mirror system, but the calibration of the system is very complicated; also, curved mirrors are expensive and need high precision.

## 1.2 Outline of This Paper

The remainder of the paper is organized as follows. Section 2 presents the proposed system and its equivalent multi-camera system. Section 3 explains fisheye lens calibration and system calibration. In Section 4, we present two stereo algorithms that we use for range estimation using the proposed system. Section 5 presents a description of the experimental results, and we conclude this paper with some relevant remarks in Section 6.

## 2 PROPOSED MONOCULAR MULTI-VIEW STEREO IMAGING SYSTEM

In this section, we describe the design procedure of a monocular multi-view stereo imaging system. We show that the system is equivalent to a synchronous multiple-cameras configuration, in which all the cameras share the same physical characteristics.

## 2.1 System configuration

The monocular multi-view stereo imaging system is composed of a fisheye lens and four identical isosceles trapezoidal planar mirrors. As shown in Figure 1, the four mirrors are placed around the fisheye lens, and form a frustum of a pyramid. The reflective side of the mirrors faces towards the inside of the frustum. The fisheye lens fits into the small, port-side of the frustum, and its optical axes coincide with the axis of the frustum.

In the system, the view-angle of the fisheye lens is  $360^\circ \times 185^\circ$ ; therefore, the light rays incident on the lens include direct light and the light reflected by the mirrors. By adjusting the size and shape of the mirrors, it is possible to make the direct light and reflected light come from the same object. Figure 2 presents a circular fisheye image, as captured by the proposed system. The circular image is divided into five regions: a center region, projected by direct light, and four surrounding regions, projected by light reflected from the mirrors. The center region can be considered as a regular (but with a heavy lens distortion) camera observation, and views the object directly. Moreover, the surrounding regions can be considered as *mirrored camera* observations, and can be used to measure the 3D shape of the object by using the multi-baseline stereo algorithm. The optical axes and FOV of the mirrored cameras can be determined by the placement angle and size of the mirrors (we will describe this in more detail in the following subsections).

In the prototype system in this study, we use four mirrors for the horizontal view-angle of the mirrored cameras; however, the number of mirrors ( $\geq 1$ ) is arbitrary.

## 2.2 Equivalent multi-camera system

Figure 3 portrays a sectional side view of the proposed system. A camera coordinate is set with its origin at the optical center of the lens, and the Z-axis equates to the optical axis. As shown in Figure 3, light rays, which intersect with the optical center  $O$  of the camera after being reflected by the up mirror, will intersect at the point  $O'$  if the rays go straight, without being reflected by the mirror. The point  $O'$  represents the optical center of a virtual camera, which we call a mirrored camera in this study. The vertical view-angle of the center region,  $\alpha$ , the vertical view-angle of the upper mirrored camera,  $\beta$ , and the angle between the axes of the two cameras,  $\theta$ , are functions of the mirror position,  $b$ , the mirror size,  $m$ , and the mirror angle,  $\gamma$ , as follows<sup>1</sup>:

$$\alpha = \pi - 2 \tan^{-1} \frac{\hat{m} \sin \gamma}{1 + \hat{m} \cos \gamma}, \quad (1)$$

$$\beta = \frac{1}{2}(\pi - \alpha), \quad (2)$$

$$\theta = \pi - 2\gamma, \quad (3)$$

$$O'(y, z) = (b(1 - \cos 2\gamma), -b \sin 2\gamma). \quad (4)$$

Here,  $\hat{m} = m/b$  denotes a normalized mirror size. Furthermore,  $O'(y, z)$  denotes the position of the mirrored camera.

<sup>1</sup> The horizontal view-angle of the upper mirrored camera is almost the same as that of the center region in the prototype system.

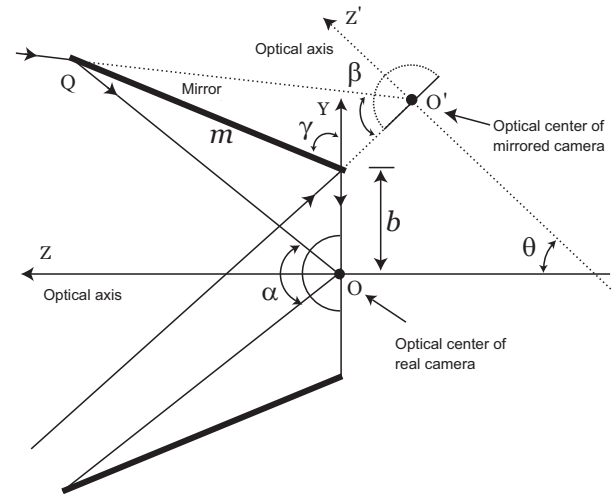


FIG. 3 Sectional side view of the proposed system. The figure represents the displacement relationship between the real camera and the mirrored camera.

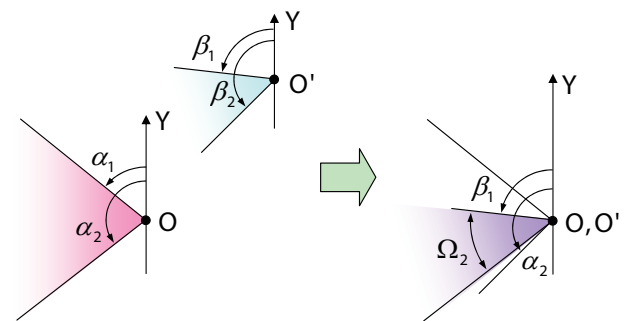


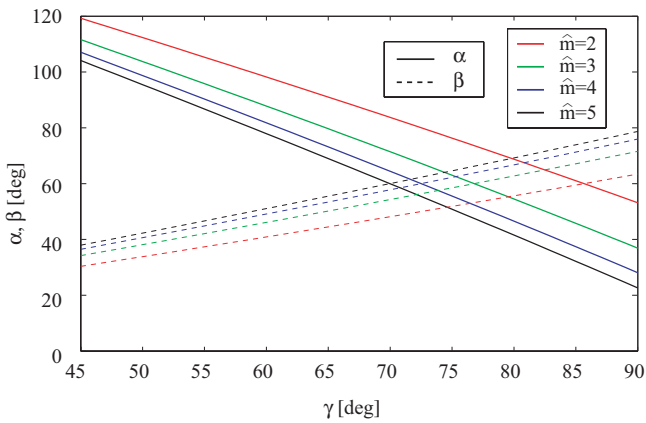
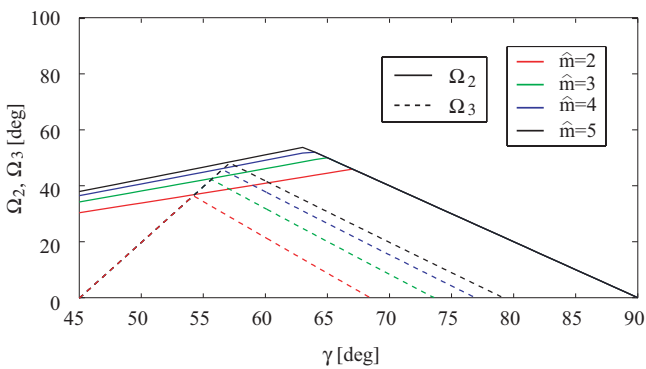
FIG. 4 Common angle for a far distant object.

The other three mirrored cameras can be represented in a similar way. Thus, the proposed system is equivalent to a five-camera configuration, in which all the cameras are synchronized, and share the same physical characteristics. This is the advantage of the proposed system over a configuration with five real cameras.

## 2.3 Dimensional design of mirrors

As described above, the view-angles  $\alpha$  and  $\beta$ , and the position of the mirrored camera can be determined by the normalized mirror size,  $\hat{m}$ , and mirror angle,  $\gamma$ , that depend on the dimension of the isosceles trapezoidal planar mirror. For the proposed monocular stereo imaging system, the size of a common view-angle between  $\alpha$  and  $\beta$  is an important design consideration. We need to design the mirrors so as to maximize the common view-angle for a given mirror size,  $\hat{m}$ . This subsection presents a design guideline for the mirror by evaluating the respective common view-angles of two cameras (center and upper),  $\Omega_2$ , and of three cameras (center, upper, and lower),  $\Omega_3$ . The common view-angle  $\Omega_2$  indicates that at least two of the three cameras can detect a distant object in this angle range, whereas the angle  $\Omega_3$  indicates that all three cameras can see the object.

For describing the common view-angles  $\Omega_2$  and  $\Omega_3$ , as shown in Figure 4, we define four new angle variables,  $\alpha_1$ ,  $\alpha_2$ ,  $\beta_1$  and  $\beta_2$ , all of which refer to the same direction of the Y-axis.

FIG. 5 View angle  $\alpha$  and  $\beta$ .FIG. 6 Common angle  $\Omega_2$  and  $\Omega_3$ .

Thus the view-angles  $\alpha$  and  $\beta$  mentioned in Section 2.2 can be rewritten as  $\alpha = \alpha_2 - \alpha_1$ , and  $\beta = \beta_2 - \beta_1$ . According to the trigonometric relationship shown in Figure 3, the four angles are the functions of the mirror angle,  $\gamma$ , as shown in Equation (5).

$$\begin{cases} \alpha_1 = \tan^{-1} \frac{\hat{m} \sin \gamma}{1 + \hat{m} \cos \gamma} \\ \alpha_2 = \pi - \alpha_1 \end{cases} \quad \begin{cases} \beta_1 = 2 \sin^{-1} \frac{\sin \alpha_1}{\hat{m}} + \alpha_1 \\ \beta_2 = 2\gamma \end{cases} \quad (5)$$

The formulae in Equation (5) show that angle  $\beta_1$  is always larger than angle  $\alpha_1$  across the  $\alpha_1$  domain,  $\{0, \pi/2\}$ . Without considering the case of multiple reflections from two mirrors, the angle range of the common angle  $\Omega_2$  is  $\beta_1$  to  $\min\{\alpha_2, \beta_2\}$ . The common angle  $\Omega_2$  has a maximum at a specific mirror angle, because the magnitude relation of  $\alpha_2$  and  $\beta_2$ , and that of  $\beta_1$  change with the mirror angle  $\gamma$ . We can obtain the common angle for  $\Omega_3$  in a similar way.

Figures 5 and 6 portray view-angles and common angles respectively, and may be used as a design reference. Larger view-angles are preferred, but the view-angle of the center region,  $\alpha$ , has priority over the view-angle of the upper camera,  $\beta$ . A larger common angle  $\Omega_2$  is preferred, because the object in this angle range can be measured. Moreover, a smaller normalized mirror size,  $\hat{m}$ , is preferred for a smaller mirror system.

For the prototype system, we chose the mirror angle  $\gamma = 65.0^\circ$  and the normalized mirror size  $\hat{m} = 3.0$ . In this case, the view-angle of the center region is  $\alpha \approx 80^\circ$ , the view-angle of the

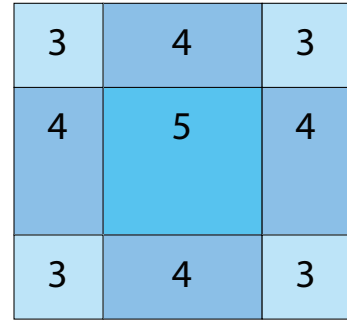


FIG. 7 The number of available cameras for stereo measurement.

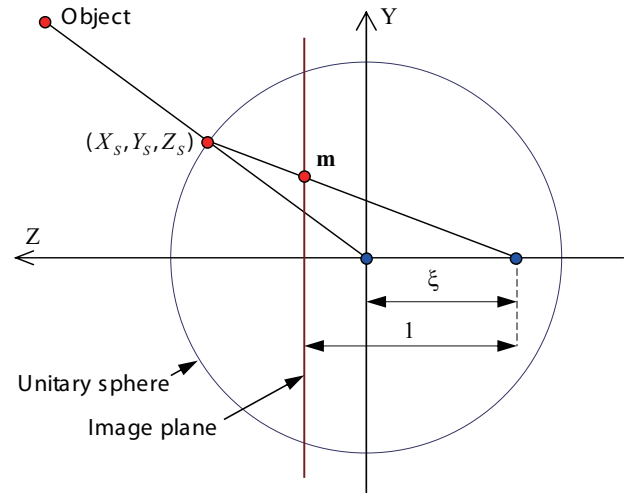


FIG. 8 Unified projection model proposed to model the projection of omni-directional cameras, such as a camera with a normal lens and hyperbolic or parabolic mirrors, and a camera with a fisheye lens.

upper camera is  $\beta \approx 50^\circ$ , and the common angle of the two cameras is  $\Omega_2 \approx 50^\circ$ .<sup>2</sup>

The proposed system captures an image of five regions: a center region with a real camera, and four surrounding regions with mirrored cameras. The common angles described above are the angles in a horizontal or vertical direction. The number of available cameras for the multi-baseline stereo differs with respect to the image position in the center region. Figure 7 shows the number of cameras available for stereo measurement in the center region.

### 3 SYSTEM CALIBRATION

#### 3.1 Unified projection model for a fisheye lens

The unified projection model [8] was proposed to model the projection of omni-directional cameras, such as a camera with a normal lens and hyperbolic or parabolic mirrors, and a camera with a fisheye lens. With respect to the model, others then proposed a calibration method [18] with a lens distortion model [27].

<sup>2</sup> The common angles  $\Omega_2$  and  $\Omega_3$  are useful for far distant objects; they vary according to the object distance. We are seeking a better mirror design for actual measuring situations.



This subsection briefly describes the fisheye lens calibration method with the lens distortion model. In the unified projection model, an object in 3D space is projected in the image plane according to the following four steps (refer to Figure 8):

1. Projection onto a unitary sphere: An object  $\mathbf{X} = (X, Y, Z)$  is projected onto a unitary sphere surface with its center at the coordinate origin. The projection origin is also the center of the sphere.

$$(X_s, Y_s, Z_s) = \frac{\mathbf{X}}{\|\mathbf{X}\|} \quad (6)$$

2. Projection onto a normalized plane: The origin of the coordinate system is set to  $(0, 0, -\xi)$ . The projected object on the unitary sphere is then projected onto a normalized plane that is orthogonal to the  $Z$ -axis at a unit distance from the new origin.

$$\mathbf{m} = (x, y, z)^\top = \left( \frac{X_s}{Z_s + \xi}, \frac{Y_s}{Z_s + \xi}, 1 \right)^\top \quad (7)$$

3. Lens distortion: We then consider the following radial and tangential lens distortions.

$$\rho \rightarrow \rho(1 + k_1\rho^2 + k_2\rho^4 + k_5\rho^6), \quad \rho = \sqrt{x^2 + y^2} \quad (8)$$

$$x \rightarrow x + 2k_3xy + k_4(3x^2 + y^2) \quad (9)$$

$$y \rightarrow y + k_3(x^2 + 3y^2) + 2k_4xy \quad (10)$$

4. Projection onto an image plane: The distorted object is then projected onto an image plane with the following intrinsic camera parameters:

$$\mathbf{p} = \mathbf{K}\mathbf{m} = \begin{bmatrix} f_u & 0 & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{m} \quad (11)$$

### 3.2 Lens calibration

We employ a MATLAB implementation [20] of the calibration method using the unified projection model [18] described previously. Figure 9 shows an example image of the calibration. The whole process of calibration is as follows:

1. Initialize the calibration parameter as  $\{\xi, k_1, k_2, k_3, k_4, k_5, f_u, f_v, u_0, v_0\} = \{1, 0, 0, 0, 0, 0, f, f, \hat{u}_0, \hat{v}_0\}$ , where  $(\hat{u}_0, \hat{v}_0)$  is the half size of the image.
2. Estimate an initial focal length,  $f$ , using at least three user-defined points<sup>3</sup> on a line.
3. The user specifies four corners of the calibration target in the image.
4. Estimate the fisheye lens projection function (intrinsic parameter) and the relative position and orientation of the calibration target using an optimization method.

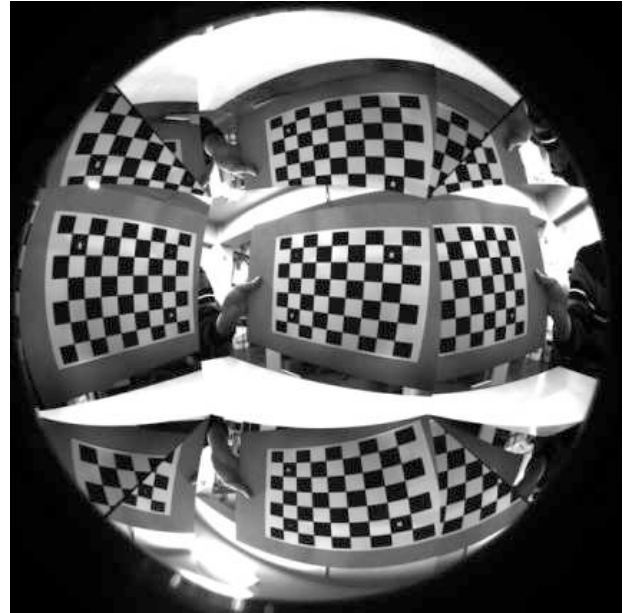


FIG. 9 A captured checkerboard image for system calibration.

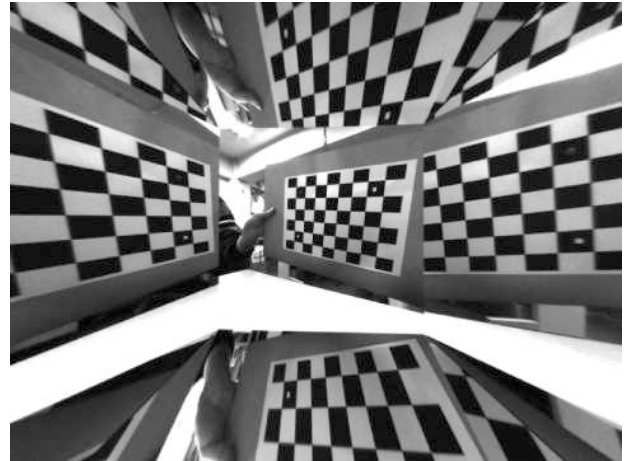


FIG. 10 Perspective reprojection of Figure 9.

One calibration target (a checkerboard) can be taken by the proposed system as five targets with different positions and orientations. These five targets can be considered as five observations of a single target, and a single image including five targets is sufficient to estimate all calibration parameters. However, the target position for five such observations is limited to a specific region in the image, as described in Section 2.3. The calibration accuracy is insufficient because biased regions are used in the circular fisheye image.

In our calibration, the target was placed evenly in the image, irrespective of whether it was a direct or a mirrored observation. The lens calibration was performed using 23 such observations.

The estimated parameters are

$$\begin{aligned} & \{\xi, k_1, k_2, k_3, k_4, k_5, f_u, f_v, u_0, v_0\} \\ & = \{1.6988, -0.06093, 0.18404, -0.00015, -0.00017, 0, \\ & \quad 871.54278, 868.49105, 791.49429, 595.47177\} \end{aligned}$$

<sup>3</sup>Four points are minimum in the MATLAB implementation.

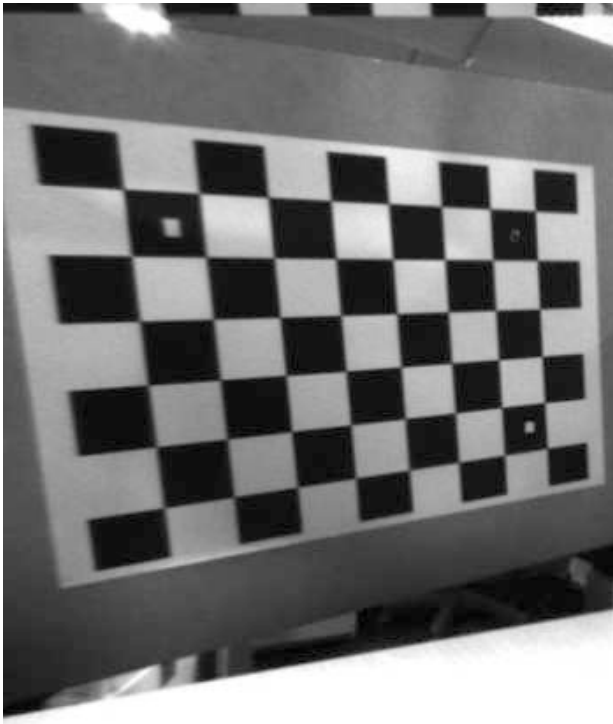


FIG. 11 Another perspective reprojection of Figure 9.

for an image size of  $1600 \times 1200$  pixels.

Figure 10 shows a perspective reprojection of the image shown in Figure 9 using the estimated lens parameters. Figure 11 shows another perspective reprojection, on a different plane to the one in Figure 10. It is readily apparent that the fisheye lens distortions are perfectly compensated for a heavily distorted part of Figure 9.

### 3.3 Position and orientation of the mirrored cameras

The proposed system can capture the images of one real camera and four mirrored cameras in a single image. This subsection explains the extrinsic parameter calibration between these cameras.

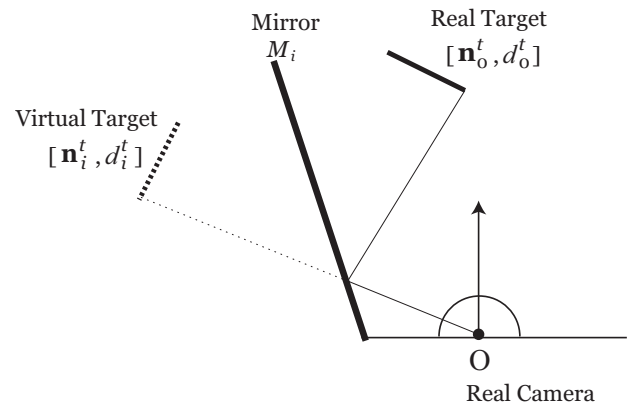
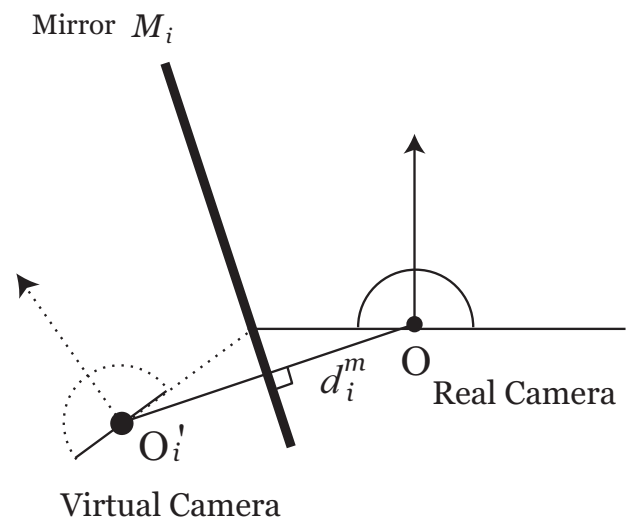
#### 3.3.1 Calibration target

In a similar fashion to that of a well-known calibration tool [4] for a perspective camera, we estimate the extrinsic parameters iteratively using intrinsic parameters, by observing a stationary target from different camera locations.

#### 3.3.2 Mirrors

Using the extrinsic parameters, we can determine the mirror position and orientation between a real and mirrored target. As depicted in Figure 12, the mirror position and orientation  $[\mathbf{n}_i^m, d_i^m]$  that reflect the  $i$ -th mirrored target are obtainable as follows.

$$[\mathbf{n}_i^m, d_i^m] = \left[ \frac{\mathbf{n}_0^t - \mathbf{n}_i^t}{|\mathbf{n}_0^t - \mathbf{n}_i^t|}, \frac{(\mathbf{n}_0^t - \mathbf{n}_i^t) \cdot \mathbf{n}_i^t}{|\mathbf{n}_0^t - \mathbf{n}_i^t| (1 - \mathbf{n}_0^t \cdot \mathbf{n}_i^t)} (d_i^t - d_0^t) \right] \quad (12)$$

FIG. 12 Displacement relationship between the real target (0-th) and the mirrored ( $i$ -th) target.FIG. 13 Displacement relationship between the real camera  $O$  and the mirrored camera  $O_i'$ .

In Eq. (12),  $[\mathbf{n}_i^t, d_i^t]$  denotes a set containing the target normal,  $\mathbf{n}_i^t$ , and the distance from the optical center,  $d_i^t$ , as the position and orientation of the  $i$ -th target. As described previously, we only need a single image, including the five targets, to estimate the four mirror positions and orientations, but they are estimated by minimizing the sum of the reprojection error of many target positions.

#### 3.3.3 Mirrored cameras

As depicted in Figure 13, the optical center,  $O_i'$ , and the rotation matrix,  $\mathbf{R}_i$ , of the  $i$ -th mirrored camera is obtainable using the estimated  $i$ -th mirror position and orientation, as follows [9]:

$$O_i' = -2d_i^m \mathbf{n}_i^m \quad (13)$$

$$\mathbf{R}_i = \mathbf{I} - 2\mathbf{n}_i^m \mathbf{n}_i^{m\top} \quad (14)$$

## 4 SINGLE-CAMERA MULTI-BASELINE STEREO

It is acceptable to use the five observations from the different positions and orientations that we estimated during calibration with the stereo method, especially the multi-

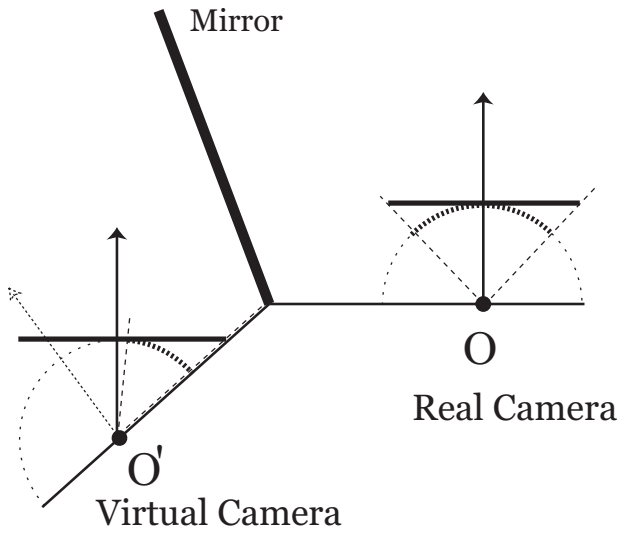


FIG. 14 Projection planes for the reprojection.

baseline stereo method [22]. This section presents two range-estimation methods using the proposed system.

#### 4.1 Perspective reprojection of the fisheye image

In the first approach, we create five perspective projection images to use for the multi-baseline stereo. The view-angle of a fisheye lens is very wide ( $180^\circ$ ), and so it is not possible to convert the entire image taken by a fisheye lens to a single perspective projection image. Each view of the real and mirrored cameras should be converted separately to perspective projection images. We divide the whole circular fisheye image into the five observations manually, although we only need to this once, because the mirror positions and orientations are stationary relative to the fisheye lens.

The conversion involves two steps. The first step converts the fisheye projected image to incident and azimuthal angles using the estimated projection function of the lens. The second step reprojects the angles to a perspective projected image.

The optical axis of the mirrored camera has an offset angle to the axis of the real camera, as described in Eq. (3). Moreover, the central direction of the view-angle of the mirrored camera differs from that of the real camera. In our reprojection, the perspective projection plane is placed parallel to the projection plane for the real camera, as shown in Figure 14. The placement of the five cameras is therefore a parallel stereo, with an anteroposterior offset.

The epipolar constraints for the four mirrored cameras in their reprojection images are derived from the extrinsic parameters estimated in the calibration. The multi-baseline stereo method [22] is applicable to the proposed system by evaluating  $SSSD$  (sum of the sum of squared differences) between a small  $ROI$  (region of interest) set in the centered real image and small  $ROIs$  on the constraint lines in the converted mirrored camera images, with respect to the object range.

Specifically, Let  $I_c(x_0, y_0)$  be the pixel value (intensity) of the

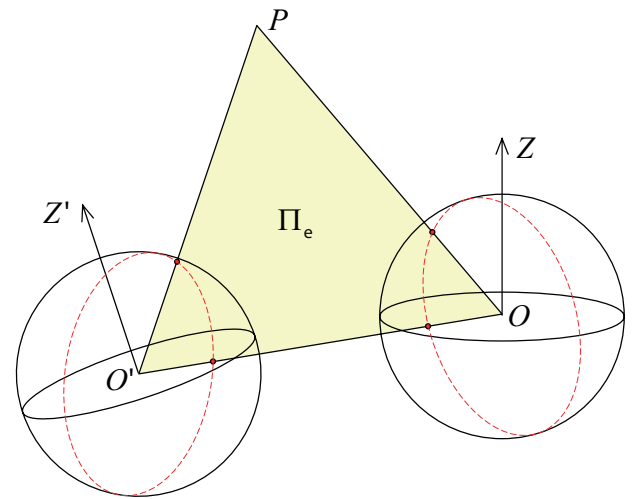


FIG. 15 Unitary spheres for the fish-eye projection and their epipolar plane.

reference pixel at  $(x_0, y_0)$  in the centered real camera image,  $I_c$ . The value of a corresponding pixel in the  $i$ -th converted mirrored camera image,  $I_{m_i}$ , is denoted by  $I_{m_i}(x, y; l)$ .

We define the  $SSD$  (sum of squared differences) value that represents the similarity between  $I_c(x_0, y_0)$  and  $I_{m_i}(x, y; l)$  at a depth  $l$ , as follows:

$$SSD_{I_c, I_{m_i}}(l; x_0, y_0) = \sum_W (I_c(x_0, y_0) - I_{m_i}(x_0, y_0; l))^2 \quad (15)$$

where  $W$  indicates a matching window size.

An actual depth can minimize all four  $SSD$  values defined by Eq. (15). We compute the  $SSSD$  (Sum of  $SSD$ ) to average out noise and reduce ambiguities [22]. The  $SSSD$  is defined as follows:

$$SSSD(l; x_0, y_0) = \sum_{i \in S(x_0, y_0)} SSD_{I_c, I_{m_i}}(l; x_0, y_0) \quad (16)$$

Here,  $S(x_0, y_0)$  denotes an index set of mirrored cameras that can achieve stereo measurement with the central real camera on point  $(x_0, y_0)$ .

For each point  $(x_0, y_0)$  in the centered real camera image,  $I_c$ , we estimate a depth value  $\hat{l}$ , which minimizes the  $SSSD$  over all possible of  $l$ . This process results in a dense depth map which covers all pixels in the image,  $I_c$ .

#### 4.2 Matching in the fisheye image

The preceding subsection described how to convert images. This subsection explains how, not to convert the images, but to convert the epipolar constraint in the fisheye image for direct matching.

Figure 15 depicts a stereo camera pair with a fisheye lens, which sees a point  $P$  from the optical centers  $O$  and  $O'$ . The following epipolar plane  $\Pi_e$  includes the three points  $P$ ,  $O$ , and  $O'$  for this situation.

$$(\vec{OP} \times \vec{OO'}) \cdot \mathbf{x} = (\mathbf{p} \times (-2d^m \mathbf{n}^m)) \cdot \mathbf{x} = 0 \quad (17)$$

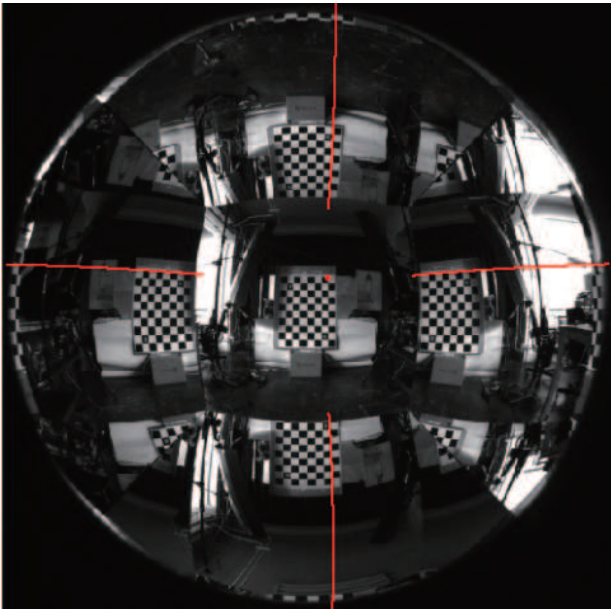


FIG. 16 Epipolar curves in mirrored images.

Here,  $\mathbf{x} = (x, y, z)$  signifies a point on the plane  $\Pi_e$ , and  $\mathbf{p}$  represents the position of  $P$ .

Consider that the camera,  $O$ , detects an object,  $P$ , that changes in distance from the camera. As the distance changes, the spherical projection of point  $P$  moves along the intersection of the epipolar plane,  $\Pi_e$ , and the unitary sphere with its center at  $O'$  [3]. We can represent a point  $\mathbf{x} = (x, y, z)$  on the unitary sphere with its center at  $O'$  as follows:

$$(\mathbf{x} - O') \cdot (\mathbf{x} - O') = (\mathbf{x} + 2d^m \mathbf{n}^m) \cdot (\mathbf{x} + 2d^m \mathbf{n}^m) = 1 \quad (18)$$

Then we can obtain the epipolar constraint curve as the projection by the calibrated fisheye projection function of  $\mathbf{x}$ , which satisfies both Eqs. (17) and (18). Figure 16 presents examples of the epipolar curves.

Specifically, the epipolar curve corresponding to a reference point can be determined as follows:

(1) We can calculate the projection light ray of the reference point according to the fisheye projection function and the coordinates of the reference point in the central camera image. We can use the point of intersection between the projection light ray and the unitary sphere of the reference camera as point  $\mathbf{p}$  in Eq. (17) to calculate the epipolar plane  $\Pi_e$ .

(2) We can obtain an intersection circle between the epipolar plane  $\Pi_e$  and unitary sphere of the mirrored camera from Eqs. (17) and (18). According to triangle geometry, the incident and azimuthal angle of point  $\mathbf{x} = (x, y, z)$  on the intersection circle is obtainable, as

$$\begin{cases} \phi = \tan^{-1} \frac{y}{x} \\ \theta = \frac{\pi}{2} - \sin^{-1} \frac{z}{\sqrt{x^2 + y^2 + z^2}} \end{cases}, \quad (19)$$

where  $\theta$  denotes the angle to the  $Z$  axis.

(3) Then, we project Eq. 19 onto the image using the calibrated

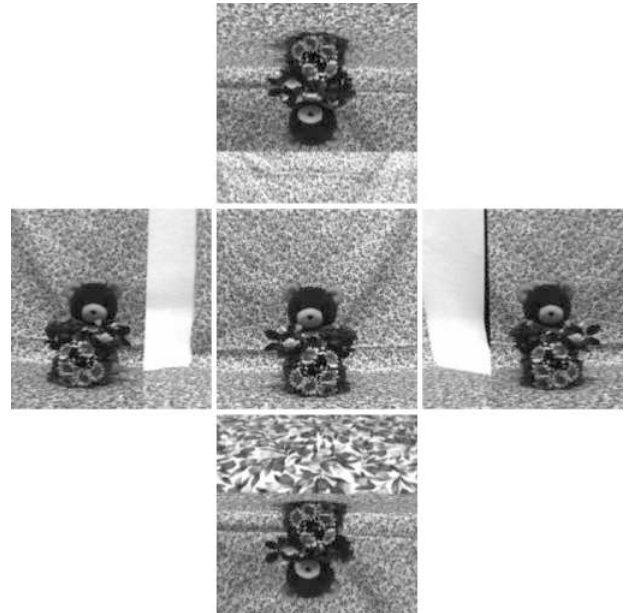


FIG. 17 Perspective reprojected images from Figure 2.

projection function  $\hat{\rho}(\theta)$  and image center  $(u_0, v_0)$ , as follows:

$$\begin{cases} u = u_0 + \hat{\rho}(\theta) \cos \phi \\ v = v_0 + \hat{\rho}(\theta) \sin \phi \end{cases}, \quad (20)$$

where  $\hat{\rho}(\theta)$  includes the lens distortion compensation.

We can obtain an epipolar curve for each mirrored camera. As described in the preceding subsection, we can apply the multi-baseline stereo method [22] to the proposed system by evaluating the *SSSD* between a small *ROI* set in the centered real image, and small.

## 5 EXPERIMENTAL RESULTS

Figure 17 shows the perspective reprojected images converted from Figure 2 using the method described in 4.1.

Figure 18 (left) shows the perspective reprojected image of the real camera. Figure 18 (center) shows the range estimation results using the perspective reprojected method described in 4.1, and Figure 18 (right) shows the constraint in the fisheye image method described in Section 4.2. The *ROI* size is  $7 \times 7$  pixels for both methods. The results are almost the same, because the only difference is the reprojected image or the projection of the constraint. In both results, the detailed 3D shapes are clearly recovered, such as the edges of the leaves in front of the bear.

To demonstrate the effectiveness of the proposed system, we also performed another experiment. Figure 19 shows a comparison of experimental results performed by the proposed method and a two-view stereo vision method, separately. These results verify that the proposed system using a camera and four mirrors can greatly improve the accuracy in depth estimation.



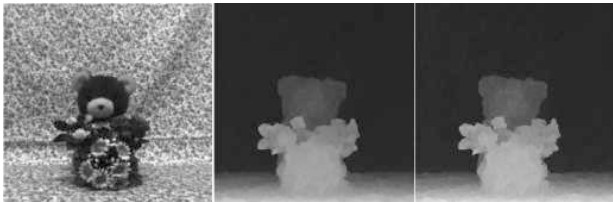


FIG. 18 Perspective reprojected image of the real camera (left), and range estimation results (center and right). The center and right images are results obtained using the methods described in 4.1 and 4.2, respectively.



FIG. 19 Comparison of estimation results: perspective reprojected image of the real camera (left), estimation result by the proposed method (center), and estimation result by a two-view stereo method (right).

## 6 CONCLUSIONS

This paper presented our proposed monocular range measurement system with a fisheye lens and planar mirrors, which are placed around the lens. The proposed system is equivalent to a multiple-camera configuration in which the cameras all have the same physical characteristics, including intrinsic parameters and color system. We have shown that the system is conducive to stereo matching. Compared with a real multiple-camera configuration, our system will be much easier to maintain in terms of synchronization between the cameras, the bandwidth needed to capture the images, and the memory and CPU power needed to handle the images. A captured image captured by our system can be used for the direct observation of a target with the centered region. Simultaneously, it can also be used for Stereoscopic scene capture to achieve 3D information display or 3D reconstruction.

Experimental results demonstrate the effectiveness of a real working system.

In future work, we would like to explore the benefits of using more complex mirror profiles. Another interesting direction is the use of different types of mirrors within a system. We believe that the use of various types of mirrors would yield even greater flexibility in terms of the imaging properties of the system, and at the same time enable us to optically fold the system to make it more compact.

## References

- [1] E. H. Adelson, and J. Y. A. Wang, "Single lens stereo with a plenoptic camera" *IEEE T. Pattern Anal.* **14** (2), 99–106 (1992).
- [2] O. Avni, T. Baum, G. Katzir, and E. Rivlin, "Recovery of 3D animal motions using cameras and mirrors" *Mach. Vision Appl.* **21**, 879–888 (2009).
- [3] J. P. Barreto, and K. Daniilidis, "Epipolar geometry of central projection systems using Veronese maps" in *CVPR'06: IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1258–1265 (IEEE Computer Society, New York, 2006).
- [4] [http://www.vision.caltech.edu/bouguetj/calib\\_doc/index.html](http://www.vision.caltech.edu/bouguetj/calib_doc/index.html)
- [5] J. Courbon, Y. Mezouar, L. Eck, and P. Martinet, "A generic fisheye camera model for robotic applications" in *Proceedings to IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1683–1688 (IEEE/RSJ, San Diego, 2007).
- [6] J. Diebel, K. Reutersward, S. Thrun, J. Davis, and R. Gupta, "Simultaneous Localization and Mapping with Active Stereo Vision" in *Proceedings to 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3436–3443 (IEEE/RSJ, Sendai, 2004).
- [7] A. R. J. Francois, G. G. Medioni, and R. Waupotitsch, "Mirror symmetry  $\Rightarrow$  2-view stereo geometry" *Image Vision Comput.* **21** (2), 137–143 (2003).
- [8] C. Geyer, and K. Daniilidis, "A unifying theory for central panoramic systems and practical implications" in *Proceedings to the European Conference on Computer Vision*, 445–461 (ECCV, Dublin, 2000).
- [9] J. Gluckman, and S. K. Nayar, "Catadioptric stereo using planar mirrors" *Int. J. Comput. Vision* **44** (1), 6–79 (2001).
- [10] J. Gluckman, and S. K. Nayar, "Rectified Catadioptric Stereo Sensors" *IEEE T. Pattern Anal.* **24** (2), 224–236 (2002).
- [11] A. Goshtasby, and W. A. Gruver, "Design of a single-lens stereo camera system" *Pattern Recogn.* **26** (6), 923–937 (1993).
- [12] M. D. Grossberg, and S. K. Nayar, "A general imaging model and a method for finding its parameters" in *Proceedings to the Eighth International Conference on Computer Vision*, 108–115 (IEEE Computer Society, Vancouver, 2001).
- [13] S. Hiura, and T. Matsuyama, "Depth measurement by the multi-focus camera" in *Proceedings to IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 953–959 (IEEE Computer Society, Santa Barbara, 1998).
- [14] M. Inaba, T. Hara, and H. Inoue, "A stereo viewer based on a single camera with view-control mechanism" in *Proceedings to the International Conference on Robots and Systems*, 1857–1864 (IEEE, Yokohama, 1993).
- [15] H. Kawasaki, Y. Ohsawa, R. Furukawa, and Y. Nakamura, "Dense 3D Reconstruction with an Uncalibrated Active Stereo System" in *Proceedings to the 7th Asian Conference on Computer Vision*, 882–891 (ACCV, Hyderabad, 2006).
- [16] S. Kuthurummal, and S. K. Nayar, "Multiview radial catadioptric imaging for scene capture" *ACM T. Graphic.* **25** (3), 916–923 (2006).
- [17] D.-H. Lee, and I.-S. Kweon, "A novel stereo camera system by a biprism" *IEEE T. Robot. Autom.* **16** (5), 528–541 (2000).
- [18] C. Mei, and P. Rives, "Single view point omnidirectional camera calibration from planar grids" in *Proceedings to 2007 IEEE International Conference on Robotics and Automation*, 3945–3950 (IEEE Computer Society, Rome, 2007).
- [19] H. Mitsumoto, S. Tamura, K. Okazaki, N. Kajimi, and Y. Fukui, "3-D reconstruction using mirror images based on a plane symmetry recovering method" *IEEE T. Pattern Anal.* **14** (9), 941–946 (1992).
- [20] <http://www.robots.ox.ac.uk/~cmei/>
- [21] S. A. Nene, and S. Nayar, "Stereo with Mirrors" in *Proceedings to the Eighth International Conference on Computer Vision*, 1087–1094 (ICCV, Bombay, 1998).
- [22] M. Okutomi, and T. Kanade, "A multiple-baseline stereo" *IEEE T. Pattern Anal.* **15** (4), 353–363 (1993).

- [23] M. Okutomi, Y. Katayama, and S. Oka, "A simple stereo algorithm to recover precise object boundaries and smooth surfaces" *Int. J. Comput. Vision* **47**, 261–273 (2002).
- [24] S. Ramalingam, P. Sturm, and S. Lodha, "Towards complete generic camera calibration" in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1093–1098 (IEEE Computer Society, New York, 2005).
- [25] R. Sagawa, N. Kurita, T. Echigo, and Y. Yagi, "Compound catadioptric stereo sensor for omnidirectional object detection" in *Proceedings to IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2612–2617 (IEEE/RSJ, Sendai, 2004).
- [26] M. Shimizu, and M. Okutomi, "Calibration and rectification for reflection stereo" in *Proceedings to IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1–8 (IEEE Computer Society, Anchorage, 2008).
- [27] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation" *IEEE T. Pattern Anal.* **14**, 965–980 (1992).
- [28] H. H. P. Wua, M. T. Lee, P. K. Weng, and S. L. Chen, "Epipolar geometry of catadioptric stereo systems with planar mirrors" *Image Vision Comput.* **27**, 1047–1061 (2009).