# Single-camera panoramic stereo imaging system with a fisheye lens and a convex mirror

**Weiming Li and Y. F. Li***

*Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong*
*\*meyfli@cityu.edu.hk*

**Abstract:** This paper presents a panoramic stereo imaging system which uses a single camera coaxially combined with a fisheye lens and a convex mirror. It provides the design methodology, trade analysis, and experimental results using commercially available components. The trade study shows the design equations and the various tradeoffs that must be made during design. The system's novelty is that it provides stereo vision over a full 360-degree horizontal field-of-view (FOV). Meanwhile, the entire vertical FOV is enlarged compared to the existing systems. The system is calibrated with a computational model that can accommodate the non-single viewpoint imaging cases to conduct 3D reconstruction in Euclidean space.

**OCIS codes:** (150.6910) Three-dimensional sensing; (100.6890) Three-dimensional image processing.

## References and links

1. J. Kim, and M. Chungy, "SLAM with omni-directional stereo vision sensor," in *Proceedings of IEEE International Conference on Intelligent Robots and Systems* (Institute of Electrical and Electronics Engineers, New York, 2003), pp. 442–447.
2. H. Koyasu, J. Miura, and Y. Shirai, "Mobile robot navigation in dynamic environments using omnidirectional stereo," in *Proceedings of IEEE International Conference on Robotics and Automation* (Institute of Electrical and Electronics Engineers, New York, 2003), pp. 893–898.
3. Z. Zhu, D. R. Karuppiah, E. Riseman, and A. Hanson, "Adaptive panoramic stereovision," Robot. Autom. Magazine **11**, 69–78 (2004).
4. F. B. Valenzuela, and M. T. Torriti, Comparison of panoramic stereoscopic sensors based on hyperboloidal mirrors, in *Proceedings of the 6th Latin American Robotics Symposium*, (Institute of Electrical and Electronics Engineers, New York, 2009), pp. 1–8.
5. E. L. L. Cabral, J. C. S. Junior, and M. C. Hunold, "Omnidirectional stereo vision with a hyperbolic double lobed mirror," in *Proceedings of International Conference on Pattern Recognition*, J. Kittler, M. Petrou, and M. Nixon, ed. (IEEE Computer Society, Los Alamitos, Calif., 2004), pp.1–9.
6. G. Jang, S. Kim, and I. Kweon, "Single-camera panoramic stereo system with single-viewpoint optics," Opt. Lett. **31**(1), 41–43 (2006).
7. L. C. Su, C. J. Luo, and F. Zhu, "Obtaining obstacle information by an omnidirectional stereo vision system," Int. J. Robot. Autom. **24**, 222–227 (2009).
8. G. Caron, E. Marchand, and E. M. Mouaddib, "3D model based pose estimation for omnidirectional stereovision," in *Proceedings of IEEE International Conference on Intelligent Robots and Systems* (Institute of Electrical and Electronics Engineers, New York, 2009), pp. 5228–5233.
9. S. Yi, and N. Ahuja, "An omnidirectional stereo vision system using a single camera," in *Proceedings of International Conference on Pattern Recognition*, Y.Y. Tang, S.P. Wang, G. Lorette, D.S. Yeung, and H. Yan, ed. (IEEE Computer Society, Los Alamitos, Calif., 2006), pp. 861–865.
10. K. H. Tan, H. Hua, and N. Ahuja, "Multiview panoramic cameras using mirror pyramids," IEEE Trans. Pattern Anal. Mach. Intell. **26**(7), 941–946 (2004).
11. G. Krishnan, and S. K. Nayar, "Cata-fisheye camera for panoramic imaging," in *Proceedings of IEEE Workshop on Application of Computer Vision* (Institute of Electrical and Electronics Engineers, New York, 2008), pp. 1–8.
12. S. Baker, and S. K. Nayar, "A theory of single-viewpoint catadioptric image formation," Int. J. Comput. Vis. **35**(2), 175–196 (1999).
13. J. P. Tardif, P. Sturm, M. Trudeau, and S. Roy, "Calibration of cameras with radially symmetric distortion," IEEE Trans. Pattern Anal. Mach. Intell. **31**(9), 1552–1566 (2009).
14. R. Swaminathan, M. D. Grossberg, and S. K. Nayar, "Non-Single Viewpoint Catadioptric Cameras: Geometry and Analysis," Int. J. Comput. Vis. **66**(3), 211–229 (2006).
15. Z. Zhang, "A flexible new technique for camera calibration," IEEE Trans. Pattern Anal. Mach. Intell. **22**(11), 1330–1334 (2000).

16. ACCOWLE. Vision, *http://www.accowle.com/english/*
17. H. Hirschmüller, P. R. Innocent, and J. Garibaldi, "Real-time correlation-based stereo vision with reduced border errors," Int. J. Comput. Vis. **47**(1/3), 229–246 (2002).
18. Z. Li, Y. Shi, C. Wang, and Y. Wang, "Accurate calibration method for a structured light system," Opt. Eng. **47**(5), 053604 (2008).

## 1. Introduction

Panoramic stereo imaging systems aim at obtaining visual information from a wide field of view (FOV) and meanwhile with stereo views that enable 3D depth perception. These features a re important for a variety of real-world applications such as surveillance and autonomous navigation missions. A number of panoramic stereo vision systems have been proposed in the literature. One category of such systems uses two or more omni-directional cameras, which are configured to produce either horizontal stereo pairs [1] or vertical stereo pairs [2]. A survey and a comparison of such systems can be found in [3] and [4]. In contrast to these, another category of systems uses only one single camera combined with optics to achieve stereo vision. Various optical components have been reported in these systems such as a hyperbolic double lobed mirror [5], multiple combined conic mirrors [6, 7], four parabolic mirrors (with an orthographic camera) [8], a concave lens with a convex mirror [9], or a mirror pyramid [10]. The implementation with a single camera brings various advantages such as simple data acquisition without multi-camera synchronization, compact system size, and reduced system cost. These features are especially attractive in robotics applications. A limitation of using a single camera is that the image resolution is reduced. This can be alleviated by using a high resolution image sensor.

In this work, we report a design for a single camera based panoramic stereo imaging system. Compared to existing systems, the proposed system simultaneously enhances the capability of image acquisition in two aspects. First, stereo vision is enabled over a 360 degree horizontal FOV. The vertical FOV for the stereo vision is centered near the horizontal plane, around which 3D information is essential for surface vehicles or robots that need to avoid obstacles on the ground and perform path planning. Second, besides the stereo FOV, the system's entire vertical FOV is enlarged. This eliminates the visual blindness region for a better monitoring of the surrounding environment in practical applications. A similar system was reported in [11] with a goal to enlarge the vertical FOV, however, without exploring the capability of stereo vision.

The proposed design consists of a fisheye lens camera and a convex reflective mirror as shown in Fig. 1. The optical axis of the fisheye lens and the optical axis of the mirror are aligned. An image captured by the system contains a fisheye image region and a catadioptric image region, which are captured through different optical paths. For formation of the fisheye image, the light ray is directed to the image sensor through the refraction of the fisheye lens. For formation the catadioptric image, the light ray is first reflected by the mirror and then refracted through the fisheye lens. Following this, a scene point in the overlapping FOV of the two image regions generates a stereo image pair in the captured image, which encodes its depth information.
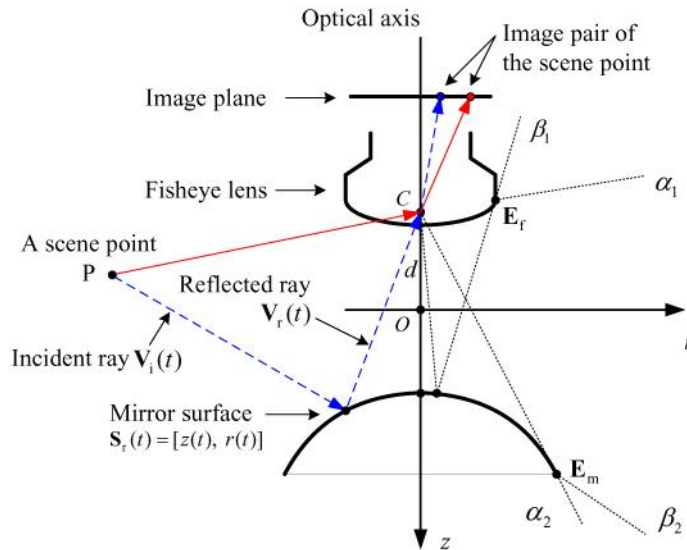
Fig. 1. An illustration of the proposed single-camera panoramic stereo vision system.

A theoretical analysis of the system geometry shows that the proposed design can be configured to balance a variety of performance requirements, such as the stereo FOV, the entire FOV, the equivalent baseline for stereo vision, and the image resolution. This provides design flexibility to meet application demands that may emphasize different performance aspects. The tradeoffs between the designs are presented. The design analysis also includes the configurations where the single-viewpoint (SVP) constraint [12] is relaxed. To work with systems with a non-SVP problem, we use a generic radial distortion model [13] to describe the imaging process. The model parameters can be estimated using a homography-based method [13], which simply requires a few images of a planar LCD panel in unknown positions.

An experimental setup has been built to demonstrate the feasibility of the proposed system using commercially available components. The setup captures a panorama with a FOV of 360° (horizontal) by 125° (vertical). The vertical FOV is enlarged by as much as a factor of 2.3 compared to the reported system [11] (55° vertical FOV). Stereo vision is enabled near the horizontal plane with a 21° vertical FOV. The system can provide a solution for obtaining 3D information for robotics missions such as obstacle avoidance and path planning while monitoring the surrounding environment with less visual blindness regions. Experiments of 3D reconstruction were conducted for an indoor environment. Further, positions of sample points displayed on an LCD panel with known positions in 3D were estimated for a quantitative evaluation of the 3D reconstruction result.

## 2. Image formation of the imaging system

Figure 1 shows the proposed system, which consists of a fisheye lens camera and a convex reflective mirror with coaxial installation. Since it is rotationally symmetric, the imaging system can be illustrated in its 2D vertical cross section. A coordinate frame $z$-$r$ is defined where the $z$ axis is aligned with the rotational axis (optical axis).

The system views the world through both the fisheye lens (from vertical view angle $\alpha_1$ to $\alpha_2$) and the reflection of the convex mirror (from vertical view angle $\beta_1$ to $\beta_2$). Here since all the view angles are discussed in the vertical cross section, we omitted the term 'vertical' before them for a neat presentation. Following this, the entire FOV of the system is from $\beta_1$ to $\alpha_2$. Within the entire FOV, a stereo FOV is from $\alpha_1$ to $\beta_2$. The boundaries of these views can be obtained based on an image formation model introduced as follows.

For simplicity, we consider the fisheye lens camera as a SVP imaging system. Here, the fisheye lens aperture is simplified to an optical center $C$ (see Fig. 1) and it is located along the axis of symmetry of the mirror at a distance $d$ from the origin $O$ ($d=|OC|$). The mirror used in the system can have a variety of shapes. To accommodate this, we use a generic parameterization [14] which gives the shape of the mirror by

$$Az^2 + r^2 + Bz = C. \tag{1}$$

This parameterization can accommodate parabolic ($A=0$), elliptic ($A>0$), hyperbolic ($A<0$), and spherical ($A=1$, $C+(B^2/4)>0$) mirrors. With the variation of $A$, systems with different mirror shapes can be analyzed in a unified framework. Following this approach, a point on the mirror surface is defined by $S_r(t) = [z(t), r(t)]$ where

$$\begin{cases} z(t) = t \\ r(t) = \sqrt{C - Bt - At^2} \end{cases} \tag{2}$$

This parameterization can be represented by mirror parameters of physical interpretations. When the mirror is a sphere centered at $[0, u]$ with a radius $R$. The parameters are given by $A = 1$, $B = -2u$, and $C = R^2 - (B^2/4)$. When the mirror is a conic with its focus located at $[0, p]$ and with an eccentricity $e$, the parameters are given by $A = 1 - e^2$, $B = -2p$, and $C = -p^2$. Here the mirror shape is ellipse ($e<1$), parabola ($e=1$), and hyperbola ($e>1$).

In the image formation process, a scene point $\mathbf{P}$ is imaged on the image sensor after it is reflected at some point $\mathbf{S}_r(t)$ on the mirror surface. Figure 1 shows the vector along the incident ray $\mathbf{V}_i(t)$ and the reflected ray $\mathbf{V}_r(t)$. Then the scene point $\mathbf{P}$ can be parameterized by its distance $k$ to its reflection point $\mathbf{S}_r(t)$ and it is expressed as:

$$\mathbf{P}(t,k) = \mathbf{S}_r(t) + k \cdot \mathbf{V}_i(t) = [p_z(t,k), p_r(t,k)] \tag{3}$$

Based on Eq. (2), all the terms in Eq. (3) can be analytically derived. In particular, a scene point $\mathbf{P}(t,k) = [p_z(t,k), p_r(t,k)]$ is given by:

$$\begin{cases} p_z(t,k) = N_z / D_z \\ p_r(t,k) = N_r / D_r \end{cases} \tag{4}$$

where

$$N_z = B^2(dk + t(3k + \Omega)) + 4(Ck(-d + t - 2At) + C\Omega t + At^2((1+A)dk + (-1+A)t(k+\Omega))) - \ldots$$
$$4B(Ck + t(-(1+A)dk + t(k - 2Ak - (-1+A)\Omega)));$$
$$N_r = \sqrt{C - t(B + At)}(B^2(k+\Omega) - 4B(-dk + t(\Omega - A(k+\Omega))) + 4(C(-k+\Omega) + \ldots$$
$$At(2dk + (-1+A)t(k+\Omega))));$$
$$D_z = D_r = \Omega(B^2 + 4(-1+A)Bt + 4(C + (-1+A)At^2));$$
$$\Omega = \sqrt{C + d^2 - 2dt - t(B + (-1+A)t)}.$$

Following the above formulation, the FOV boundary equations can be derived as follows, which forms the basis of the design trade analysis in section 3.

The upper FOV boundary of the fisheye image $\alpha_1$ is determined by the selection of the fisheye lens. The lower FOV boundary of the fisheye image $\alpha_2$ is determined by the half-diameter of the mirror $w$ and the position of the fisheye lens center $d$. A point at the edge of

the mirror can be expressed as $\mathbf{E}_m = [(-B + \sqrt{B^2 - 4A(w^2 - C)})/(2A), w]$ (see Fig. 1). By solving the equation $\mathbf{S}_r(t_m) = \mathbf{E}_m$, $\alpha_2$ can be obtained as:

$$\alpha_2 = \arctan(2wA/(d - B + \sqrt{B^2 - 4A(w^2 - C)})). \tag{5}$$

The upper FOV boundary of the catadioptric image $\beta_1$ is formed due to the occlusion of the fisheye lens. Let the half-diameter of the fisheye lens be $m$. Then a point at the edge of the fisheye lens can be expressed as $\mathbf{E}_f = [d + \tan(\alpha_1)*m, m]$ (see Fig. 1). The incident ray passing $\mathbf{E}_f$ can be determined by solving the equation $\mathbf{P}(t_f, k_f) = \mathbf{E}_f$. Then $\beta_1$ is determined by $\mathbf{V}_i(t_f)$. Similarly, the lower FOV boundary of the catadioptric image $\beta_2$ can be obtained as $\mathbf{V}_i(t_m)$, where $t_m$ is calculated by solving the equation $\mathbf{S}_r(t_m) = \mathbf{E}_m$. As $k$ is not related with the angle of light ray, it does not appear in the FOV boundary calculation results.

Note that the above discussion is based on the assumption that the lateral shift of the entrance pupil of the fisheye lens is negligible. When the entrance pupil shift is significant, high-order calculations are needed.

## 3. System design

The proposed design can be configured to emphasize different aspects of application requirements by variation of a set of design parameters. In particular, several interested performance requirements are the entire FOV, the stereo FOV, the image resolution, and the length of (equivalent) baseline for stereo vision. The system size may also need to be considered when a compact system is desired.

The entire FOV and the stereo FOV are as defined in section 2. The image resolution is evaluated as the image pixels per unit vertical FOV, which can be expressed as $\Delta N / \Delta\Theta$, where $\Delta N$ is the number of pixels along a radial image interval and $\Delta\Theta$ is the vertical FOV observed by these pixels. Since the resolution of the fisheye image will be fixed once the fisheye lens camera is chosen, the main concern will be the catadioptric image resolution. For ease of analysis, we assume that the fisheye image has a constant image resolution in the radial direction and it is associated with an image resolution index (IRI) of 1. By using this average IRI, the image resolution analysis is normalized to the fisheye image. As the FOV is extended after the mirror reflection, the catadioptric image has a lower image resolution. We consider the average IRI of the catadioptric image and it is determined by $(90^\circ + \alpha_2)/(\beta_1 - \beta_2)$. Note that since the image resolution is not uniform along the image radius, the average IRI may not be linear when it is compared between systems. In practice, the resolution difference between the fisheye image and the catadioptric image reduces the stereo matching accuracy. A partial solution is to match the resolution by applying Gaussian smoothing with different kernel sizes to the two image regions.

As in many catadioptric systems [12], there is also the focusing range setting problem in the proposed system. The proposed system consists of a fisheye lens system and a catadioptric system which requires different focusing range settings. A partial solution is to enlarge the depth of field of the fisheye lens by using a small aperture size at the cost of reducing the image acquisition frame rate. In this paper, the aperture size of the fisheye lens is assumed to be sufficiently small the IRI is simply defined based on the angular FOV.

For stereo vision, the baseline is important for the accuracy of 3D construction. When the catadioptric image is SVP, the baseline of the proposed system is the distance between the optical center of the fisheye lens and its virtual point reflected behind the mirror. In contrast to this, when the catadioptric image is non-SVP, the extended incident light rays intersect the optical axis within an interval instead of a single virtual viewpoint. In this case, an equivalent baseline is defined based on the average of these intersection points.

As shown in section 2, the design parameters in the image formation include the distance of the fisheye lens $d$, the selection of the fisheye lens, and the parameters of the mirror. When

an off-the-shelf fisheye lens is selected, the parameters of the fisheye lens are determined. With respect to the mirror, the mirror shapes can be spherical or conic. The spherical mirror parameters are the radius $R$ and the half-diameter $w$. The conic mirror parameters are the eccentricity $e$, the focus of the mirror $p$ and the half-diameter of the mirror $w$.

For a wide FOV, conic mirrors are useful. In this work, we first consider the case when a hyperbola mirror is used with a SVP constraint. Here all the light rays captured are directed to the hyperbola's focus point inside the mirror [12]. The SVP constraint requires the center of the fisheye lens to locate at the second focus of the hyperbola mirror and thus we have

$$d = -p + 2e^2 p / (e^2 - 1). \tag{6}$$

By expressing $d$ in this specific form using $e$ and $p$, $d$ does not appear in the design parameters. When we consider a fixed mirror size, the number of design parameters is further reduced to two: $e$ and $p$, where $e$ satisfies $e>1$. Here we set the mirror half-diameter to be 100 mm and the fisheye lens half-diameter to be 25 mm.

Figure 2(a) shows the variation of the entire FOV and stereo FOV with respect to $e$ and $p$. For a specific $e$, as $p$ varies, a larger entire FOV is coupled with a smaller stereo FOV and vice versa. When $p$ is fixed, a decrease of $e$ would increase both the entire FOV and the stereo FOV. However, it can be seen in Fig. 2(b) that, the decrease of $e$ would lead to a degradation of the catadioptric image resolution. Figure 2(b) shows the catadioptric image resolution (measured as average IRI) and the stereo baseline, for both of which larger values are related with better 3D reconstruction precision. It can be seen that for a fixed $e$, the two factors work in opposition. For a fixed $p$, a decrease in $e$ would obtain a desired larger baseline; however, this improvement in baseline length must be traded against an undesired decrease in image resolution.
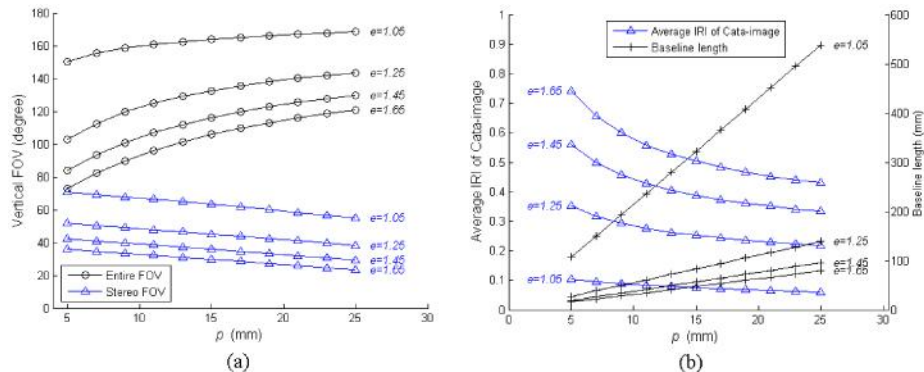


Fig. 2. (a) Vertical FOV with respect to p and e. (b) Average IRI of the catadioptric image and the stereo baseline length with respect to p and e.

As the multiple system performances cannot be obtained simultaneously, some compromise has to be made in the performance to meet the needs in real-world applications. Following this, we do not intend to obtain an optimal system design. Instead, we list several design examples in Table 1, where each design focuses on a different aspect of performance requirement. Design 1 has the widest stereo FOV. Design 2 has the largest equivalent baseline length. Design 3 has the finest catadioptric image resolution. The geometrical profiles of these designs are shown in Fig. 3.

**Table 1. Several examples of system design.**

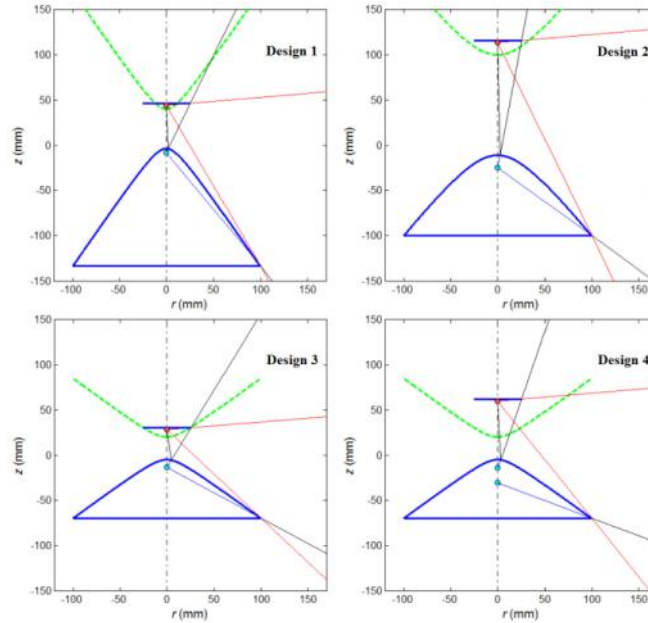|  | $e$ | $p$ (mm) | $d$ (mm) | Entire FOV (degree) | Stereo FOV (degree) | Equivalent baseline (mm) | Average IRI of cata-image |
|---|---|---|---|---|---|---|---|
| Design 1 | 1.20 | 8 | 44.36 | 125.8 | **53.1** | 52.4 | 0.25 |
| Design 2 | 1.25 | 25 | 113.89 | 144.7 | 38.4 | **138.9** | 0.21 |
| Design 3 | 1.65 | 13 | 28.09 | 104.4 | 31.0 | 41.1 | **0.51** |
| Design 4 | 1.65 | 13 | 60 | 124.2 | 23.0 | 82.2 | 0.40 |

Fig. 3. Four examples of possible system designs.

Note that design 1, 2 and 3 are under the SVP constraint. By relaxing this constraint, more design flexibility is enabled. For example, in design 3, the catadioptric image resolution reaches about half of that in the fisheye image. However, the baseline length of design 3 is small. We relax the SVP constraint in design 3 to get design 4, in which the fisheye lens is moved farther from the mirror. This effectively enlarges the equivalent baseline. A cost for this design flexibility is that the resultant system does not meet the SVP constraint. This requires a camera model that can accommodate the non-SVP nature of the system, which we shall introduce in section 4.

## 4. A computational model of the system and its calibration

To conduct 3D reconstruction, a computational model is needed to establish a mapping between the light rays in 3D space and the pixels in the image. In order to accommodate the non-SVP nature of the imaging process, we adopt the generic radial distortion model [13] as illustrated in Fig. 4. This model assumes that the imaging geometry is radially symmetric to the optical axis. Then an image *distortion center* can be defined as the intersection of the optical axis with the image plane. With regard to this distortion center, the image is decomposed into a series of *distortion circles*. The light rays associated with pixels lying on the same distortion circle lie on a right 3D viewing cone centered on the optical axis. Each viewing cone can be considered as an individual perspective camera and it can be defined by a vertex position $v(c)$ on the optical axis and a focal length $f(c)$, where $c$ is the image radial distance to the distortion center. A viewing cone can alternatively be parameterized by $v(c)$ and $\theta(c)$, where $\theta(c)$ is a half of the opening angle of the viewing cone. A mapping between $v(c)$ and $\theta(c)$ can be expressed as $\tan(\theta(c)) = c / f(c)$.

For the proposed system, the model consists of two clusters of viewing cones. One cluster describes the fisheye image and the other describes the catadioptric image. Based on these viewing cones, the imaging geometry can be fully described. Assuming variation of the optics along the image radius is smooth in the fisheye image and the catadioptric image, we can parameterize the opening angles and vertex positions of the viewing cones by polynomials. In particular, the opening angle $\theta(c)$ can be expressed as:

$$\theta(c) = \begin{cases} \theta_m(c) = a_0 + a_1 \cdot c + a_2 \cdot c^2 + \ldots + a_{N_1} \cdot c^{N_1}, c \le C_b \\ \theta_f(c) = b_0 + b_1 \cdot c + b_2 \cdot c^2 + \ldots + b_{N_2} \cdot c^{N_2}, c > C_b \end{cases} \quad (7)$$

Here $C_b$ is a radius corresponding to the circular boundary between the fisheye image and the catadioptric image. Similarly, the position of a viewing cone's vertex can be expressed as:

$$v(c) = \begin{cases} v_m(c) = g_0 + g_1 \cdot c + g_2 \cdot c^2 + \ldots + g_{N_3} \cdot c^{N_3}, c \le C_b \\ v_f(c) = h_0 + h_1 \cdot c + h_2 \cdot c^2 + \ldots + h_{N_4} \cdot c^{N_4}, c > C_b \end{cases} \quad (8)$$

Following this, the generic radial distortion model is fully expressed as the location of the distortion center, the focal length function $f(c)$ (equivalently $\theta(c)$), and the vertex position function $v(c)$ for the set of viewing cones. Notice that the SVP model corresponds to the case when only the constant coefficients are non-zero.

In order to determine the coefficients in Eq. (7) and (8), we use the method of [13]. Each distortion circle in the generic radial distortion model is regarded as a perspective camera and it is calibrated by observing points on a calibration plane at several unknown poses. A dense plane-image correspondence is involved, which is obtained by using an LCD panel as an active calibration plane. This calibration procedure can be further optimized by a maximum-likelihood method [15].

Since the proposed system is rotationally symmetric, the epipolar geometry is greatly simplified. Figure 4 shows that the corresponding image pair of an object point **P** is located on a radial line that passes the distortion center. By a latitude-longitude unwrapping, an original image can be rectified to a panoramic stereo image pair. Therefore the stereo matching can be performed by one-dimensional search along the vertically collinear scan line. Once the stereo image correspondence of a scene point is established, its position in 3D space can be estimated by triangulation based on the calibrated computational model. Further, once the computational model is calibrated, the FOV boundary parameters can be computed as the directions of the light rays related with the pixels on the boundaries of the fisheye image and the catadioptric image. Notice that this does not require knowing the system parameters.
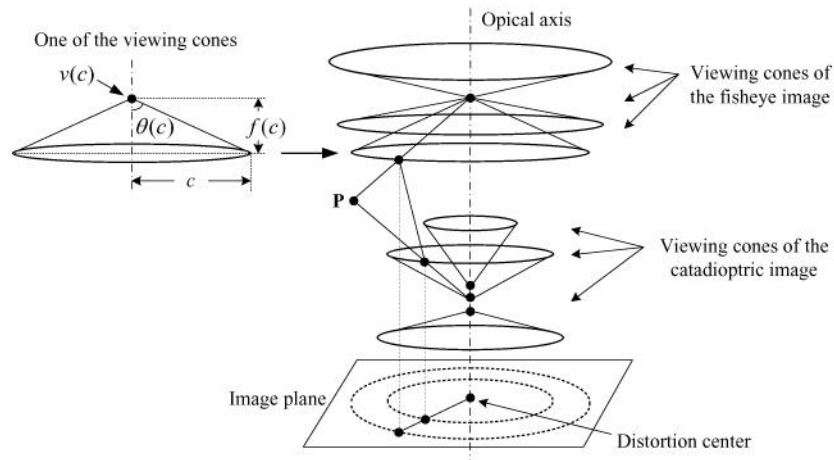


Fig. 4. An illustration of the computational model and the epipolar geometry.

## 5. An experimental setup

To demonstrate the feasibility of the proposed system, an experimental setup has been built as shown in Fig. 5. All the components are off-the-shelf and are connected to an optical mounting stage. A color video camera with a 2/3" CCD and a resolution of 1360 pixels by 1024 pixels is used. A Fujinon 1.4 mm fisheye lens (FE185C046HA-1) is mounted to the

camera. Its aperture is set to F16. The half-diameter of the lens is 28.3 mm. A hyperbolic mirror is used, which is part of a commercial omnidirectional sensor [16]. The half-diameter of the mirror is 44 mm. The vertex to vertex distance between the fisheye lens and the mirror surface was measured with a ruler and it is 52.2 mm with an uncertainty value of 0.1 mm. The eccentricity of the mirror is 1.44 and the focus is 20.1 mm. Following the similar method in section 3, the system parameters can be predicted, for which the entire FOV is 127.2°, the stereo FOV is 19.7°, the equivalent baseline is 63.3 mm, and the average IRI is 0.35.
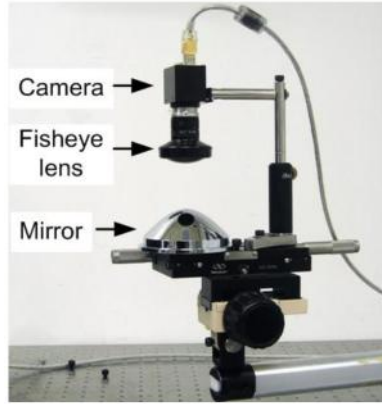


Fig. 5. Experimental setup of the single-camera panoramic stereo system.

Through calibration, the parameters of the computational model in section 4 were determined. For the presented setup, we assume a SVP model for the fisheye image and a non-SVP model for the catadioptric image. Then for the catadioptric image, $a_0 = 0$, $a_1 = 1.4742$, $a_2 = -0.0035$, $g_0 = 70.3342$, $g_1 = -0.1441$. For the fisheye image, $b_0 = 0$, $b_1 = 0.2830$, $b_2 = -0.0001$, $h_0 = 0$. Based on the calibrated computational model, the FOV boundaries can be estimated: $\alpha_1 = 7°$, $\alpha_2 = -65°$, $\beta_1 = 60°$, and $\beta_2 = -14°$. All angles are defined with respect to the horizontal direction. Therefore the entire FOV is 125° (calculated as $\beta_1 - \alpha_2$), within which the stereo FOV is 21° (calculated as $\alpha_1 - \beta_2$). With the calibrated model, all the vertexes of the viewing cones are known. Following this, the equivalent baseline is calculated as the distance between the optical center of the fisheye lens and the average position of the vertexes of the catadioptric viewing cones, which is 63.2 mm. The average IRI of the catadioptric image is 0.33. The system parameters obtained through calibration are consistent with the design prediction.

Figure 6 shows an image taken in the middle of a square shaped room by the experimental setup. The image can be decomposed into four concentric circular image regions as marked in the figure. Region A and B are captured through the mirror and they form the catadioptric image. Region C and D are captured through the fisheye lens and they form the fisheye image. Region B and D have an overlapping FOV and they form the stereo image pair. It can be seen that the corresponding points are radially collinear. A latitude-longitude unwrapping of Fig. 6 is shown in Fig. 7. Here region B and region D are overlaid and only region D is displayed. It can be seen that the entire vertical FOV is extended. This enables both the ceiling and floor to be visible in the image.

Fig. 6. Image of an indoor environment captured by the system.



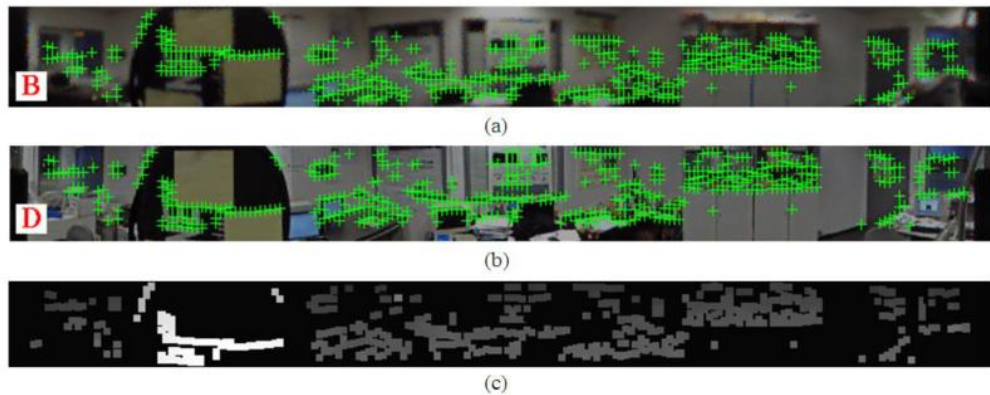Fig. 7. A latitude-longitude unwrapping of the image shown in Fig. 6.



Fig. 8. (a) and (b) show a panoramic stereo image pair from a latitude-longitude unwrapping of the image in Fig. 6. A set of matching points are displayed over the image. Fig. 8(c) shows a visualization of the disparity values of the matching point pairs.

The latitude-longitude unwrapping of region B and region D as a panoramic stereo pair is shown in Fig. 8(a) and (b). Correspondences between a set of sample points were established by a window-based cross-correlation search [17] along the vertical scan lines. The established corresponding points are displayed as the crosses over the image pair. Then the disparity between a matching point pair can be computed as the vertical position difference in the stereo pair. The disparity encodes the depth information of that point in 3D space. Figure 8(c) shows

a visualization of the disparity values of the matching image pairs. For each image pair, a square block is plotted with color. A brighter color indicates a larger disparity, which indicates a closer object to the system. In the figure, a nearby object near the left part of the figure results in blocks of brighter color. In practice, this can be used to locate nearby objects for obstacle detection. Note that there are still some mismatches in Fig. 8(c). By using more advanced matching methods with optimization, correspondence matching can be achieved with denser points and improved precision. Such efforts are beyond the focus of this work.

## 6. 3D reconstruction experiment

A quantitative 3D reconstruction experiment was conducted in a controlled test environment. The test environment is setup by positioning an LCD panel at three orthogonal positions on an optical table as shown in Fig. 9. The positions and angles of the panels were measured and used as ground truth. The distance from the LCD to the system is set to 250 mm for each position in the test environment. 3D reconstruction was performed for a set of grid sample points on the LCD at each position. Figure 10 shows the image correspondences of the sample points, which were automatically established by displaying structured spatial-temporal coding patterns on the LCD and matching the points with the same encoding. Based on a phase shift method [18], the correspondences were established with sub-pixel localization precision.
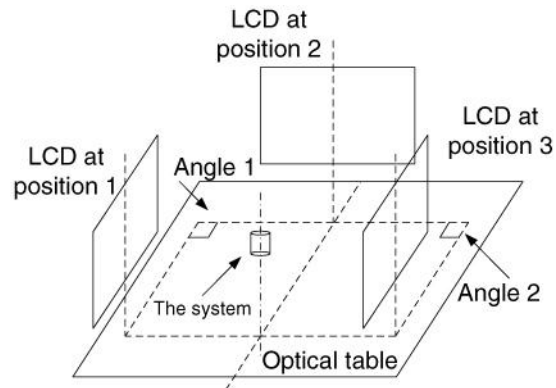


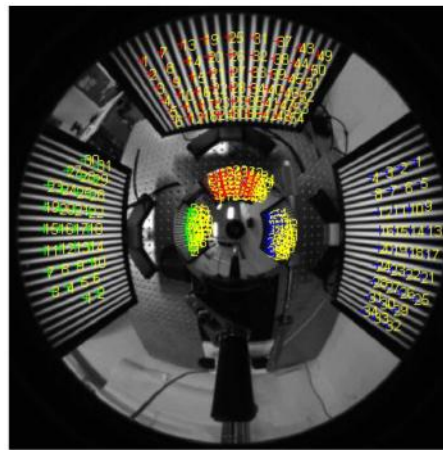Fig. 9. Setup of the test environment for 3D reconstruction.



Fig. 10. Sample points on the LCD used in the 3D reconstruction experiment.

A total of 119 sample points were used. As the ground truth of the LCD points are known, the reconstructed 3D points can be transformed to align with the LCD coordinate and be compared with the ground truth data. The reconstructed Angle 1 and Angle 2 (see Fig. 9) are

88.4° and 91.7° respectively, where the ground truth is 90° for both. Thus the angular error is about 1.8%. For the positions of the sample points on the LCD panels, compared to their ground truth values by vector magnitude error, the average reconstruction error is 8.3 mm. As the LCD panels were set to 250 mm from the system, the average error is about 3.3%. Since the image resolution is not uniform in the image, the position reconstruction error varies in space. The maximum error for position reconstruction is 7.9%. Figure 11 shows a visualization of the 3D reconstruction result. The distributions of the reconstructed sample points in 3D are consistent with the planar structures of the LCD panels without significant distortion. This result shows that within the tested working range, acceptable results in 3D reconstruction can be achieved with the system.
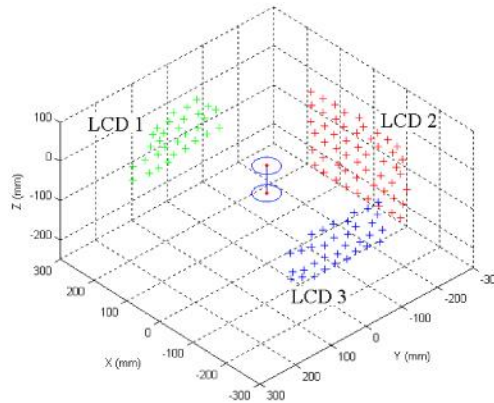


Fig. 11. Result of 3D reconstruction of the tested sample points.

It is worth noting that the precision of the 3D reconstruction is dependent on a variety of factors such as the length of equivalent baseline, resolution of the input image, object distance, and calibration precision. For the presented experimental setup, the effective working range for depth perception is limited to indoor environment as the equivalent baseline is only 63.2 mm. As we have shown in the design analysis, this can be alleviated by selecting more appropriate system design parameters. An increase in image pixel resolution of the camera can also improve the 3D reconstruction result as each pixel would have with a finer angular resolution.

## 7. Summary and future work

We have presented a design for a panoramic stereo vision system. The results with an experimental setup show that this system can produce acceptable 3D reconstruction results within a certain depth range. This design can provide a solution for a variety of applications where panoramic 3D depth is desired. In particular, this design provides a large vertical FOV for monitoring the surrounding environment.

In this work, we have investigated the design and trade analysis for the proposed setup. Currently, the analysis assumes that all the components are coaxially aligned. High-order calculations regarding the non-coaxial alignment may be considered to improve the analysis. Also, an in-depth study about the quantitative relationship between the design parameters and the precision of 3D reconstruction can be performed in the future work. A limiting feature of the work is the focusing range setting issue as mentioned in section 3. Although using a small aperture can alleviate this problem, the shutter speed has to be compromised, which is undesired in speed-demanding applications. In future work, the focusing issue can also be included in the design analysis to give an optimized solution to the system design.

## Acknowledgments

reviewers for their many constructive comments that greatly improved the quality of this work.