

# Generating Near-Spherical Range Panoramas by Fusing Optical Flow and Stereo from a Single-Camera Folded Catadioptric Rig

Igor Labutov, Carlos Jaramillo, and Jizhong Xiao

Received: date / Accepted: date

**Abstract** We design a novel “folded” spherical catadioptric rig (formed by two coaxially-aligned spherical mirrors of distinct radii and a single perspective camera) to recover near-spherical range panoramas (about  $360^\circ \times 153^\circ$ ) from the fusion of depth given by optical flow and stereoscopy. We observe that for rigid motion that is parallel to a plane, optical flow and stereo generate nearly complementary distributions of depth resolution. While optical flow provides strong depth cues in the periphery and near the poles of the view-sphere, stereo generates reliable depth in a narrow band about the equator instead. We exploit this dual-modality principle by modeling (separately) the depth resolution of optical flow and stereo in order to fuse them later on a probabilistic spherical panorama. We achieve a

---

This work is supported in part by U.S. Army Research Office under grant No. W911NF-09-1-0565, and U.S. National Science Foundation under grants No. CNS-0619577 and No. IIS-0644127

I. Labutov  
Computer Engineering Department, The City College, City University of New York (CUNY City College), Convent Ave & 140th Street, New York, NY 10031 USA  
Tel.: +001-646-239-9079

E-mail: igor.labutov@gmail.com  
Present address: Cornell University, Ithaca, NY

C. Jaramillo  
Computer Engineering Department, The City College, City University of New York (CUNY City College), Convent Ave & 140th Street, New York, NY 10031 USA  
Tel.: +001-914-500-8204

E-mail: cjarami00@ccny.cuny.edu

J. Xiao (*corresponding author*)  
Electrical Engineering Department, The City College, City University of New York (CUNY City College), Convent Ave & 140th Street, New York, NY 10031 USA  
Tel.: +001-212-650-7268

E-mail: jxiao@ccny.cuny.edu

desired vertical field-of-view and optical resolution by deriving a linearized model of the rig in terms of three parameters (radii of the two mirrors plus axial distance between the mirrors’ centers). We analyze the error due to the violation of the Single Viewpoint (SVP) constraint and formulate additional constraints on the design to minimize this error. We evaluate our proposed method via a synthetic model and with real-world prototypes by computing dense spherical panoramas of depth from cluttered indoor environments after fusing the two modalities (stereo and optical flow).

**Keywords** catadioptrics · sensor fusion · omnidirectional vision · stereoscopy · optical flow

## 1 Introduction

Omnidirectional catadioptric systems have been applied to a range of important problems in robotics including egomotion estimation[4], reactive obstacle avoidance, and SLAM[16][8]. A significant subset of these applications is the recovery of dense omnidirectional depth maps for occupancy grids and path planning. Nonholonomically constrained motion, such as unmanned aerial vehicle (UAV) navigation in a vertically-constrained cluttered environment, requires a complete sphere of range data for agile maneuverability. Although catadioptric stereo partially addresses this need, current implementations are limited to estimating depth in a fairly narrow vertical region that typically lies in the equatorial region of the view-sphere. In what follows, we outline the state-of-the-art in catadioptric stereo, and look ahead to the design of the catadioptric-stereo rig that is devised specifically for depth-imaging an entire sphere.

The main advantages of the catadioptric approach to stereo are twofold. First, catadioptric stereo can be implemented with a single camera, offering practical advantages for robotics, such as reduced cost, weight, and robust disparity matching, because a single imaging device does not introduce discrepancies between cameras' intrinsic parameters. Second, catadioptric stereo offers a richer array of topologies that can be adapted to a specific task. The rig we are presenting belongs to the family of radial imaging systems [9]. Of practical interest to mobile robotics are configurations that not only offer a wide field-of-view, but also exploit the spatially variant resolution of a mirror to an advantage of the unique dynamics of a robot. For example, the spatial distribution of depth resolution may be “tuned” to a particular azimuth and elevation, such as the robot's dominant direction of motion.

Many of the catadioptric stereo configurations that were proposed over the last decade, however, are primarily derived from the geometries outlined in the seminal treatment by Baker and Nayar [1], and aimed at satisfying the ubiquitous single-viewpoint (SVP) constraint. While the SVP guarantees that true perspective geometry can always be recovered from the original image, it limits the selection of mirror profiles to a set of conic sections [13]. The effect of this limit is twofold: 1) conic mirrors are typically not “generic” and need to be manufactured uniquely, thus costly, and 2) vertical field-of-view is limited in conic mirrors.

Despite these limitations, several SVP-compliant stereo rigs have been successfully implemented. The most recent implementation of a catadioptric stereo rig for robotics by Su and colleagues [19] uses coaxially aligned hyperbolic mirrors and a single perspective camera. Although the field-of-view (hereafter, we refer to *field-of-view* as *FOV* to mean *vertical field-of-view*) is not explicitly stated, it can be estimated from the specified geometry to be less than 90 degrees. In addition, while being suitable for ground vehicles, the system is too bulky for small UAVs. Corrêa et al. [5] developed a stereo rig with an exceptionally small form-factor by utilizing a double-lobed hyperbolic mirror with a single camera in a coaxial configuration. However, in addition to its limited field-of-view, it suffers from having a small baseline. A small form-factor together with a scalable baseline can be achieved with a “folded” configuration first introduced by Nayar and Peri [14] and successfully demonstrated in a number of applications (none used in robotics), such as [7][3]. Both utilize SVP mirrors.

Non-SVP configurations using spherical mirrors have addressed the issues of cost and limited field-of-view. The most relevant of such is the work of Derrien and Konolige in [6], while not being stereo, it explicitly

models for the error introduced by relaxing the SVP constraint in the projection function. Although non-SVP mirrors have been previously used in robotics, we consider their work seminal in its detailed study of a non-SVP mirror in its application to mobile robotics.

Another approach to depth-mapping is through the use of optical flow. As proved by Nelson and Aloimonos [15], omnidirectional optical flow offers a significant advantage in that it provides an unambiguous recovery of the system's extrinsic parameters given a sufficiently dense optical flow field. This permits a more robust de-rotation of the optical flow field, and thus, a more robust recovery of depth. McCarthy et al. [12] implemented Nelson and Aloimonos algorithm in a planar-moving robot using fish-eye optics. While this method offers a nearly hemispherical field-of-view, the depth is only recovered to a scale factor. In addition, it suffers from loss of depth resolution in the direction of the robot's motion, where it is most valuable. This is inherent to all depth-from-optical-flow approaches.

The system proposed in this paper addresses several of the aforementioned limits by generating a near-spherical depth panorama using generic, low-cost spherical mirrors. We outline the main contributions of our work:

1. We use spherical mirrors in a folded configuration to maximize image resolution near the poles of the view-sphere. For robots moving in a horizontal plane, this generates high-resolution relative depth from optical flow above and below the robot.
2. We exploit radial epipolar geometry of the spherical mirrors to compute dense metric-depth in the equatorial region of the view-sphere.
3. We fuse depth from optical-flow (poles) and stereo (equator) in a dense probabilistic depth panorama to obtain comparable depth resolution in every direction. The scale factor for depth-from-optical-flow is recovered by using weighted least-squares in regions where depth from optical flow and stereo overlap.

## 2 Design

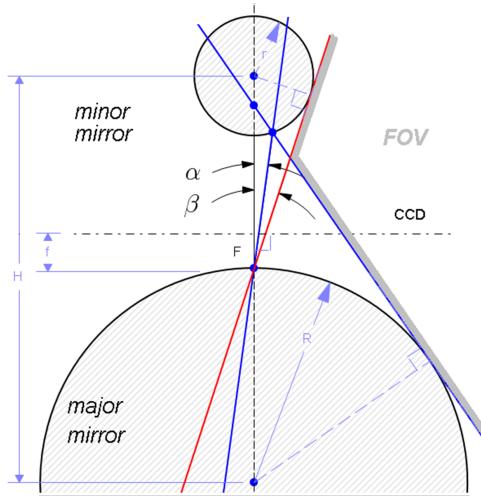
### 2.1 Model

We present a novel “folded” configuration as shown in Fig. 1. Two spherical mirrors of distinct radii  $R$  and  $r$ , termed *major* and *minor* mirrors respectively, are separated by a distance  $H$  from their centers. A perspective camera (aligned coaxially with the mirrors) is located near the surface of the major mirror ( $F$  in Fig.1) and observes the *minor* mirror within its field-of-view  $2\beta$ . Rays that lay within a cone bounded by

$\alpha$  image the *major* mirror through its reflection in the *minor* mirror, while rays bounded between  $\alpha$  and  $\beta$  image the *minor* mirror directly:

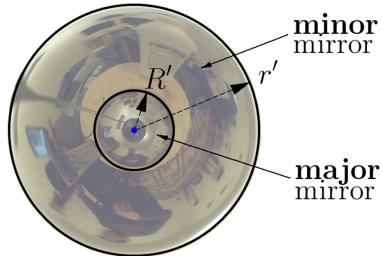
$$\alpha = r \tan^{-1} \left( \frac{2R}{\sqrt{4H^2 - 4Hr - 4R^2 + r^2}} \right) \quad (1)$$

$$\beta = 2H \tan^{-1} \left( \frac{r}{\sqrt{H^2 + 2rH}} \right) \quad (2)$$

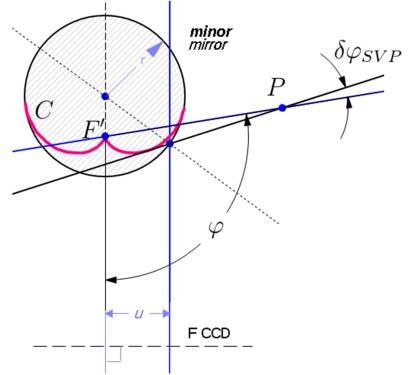


**Fig. 1** “Folded” catadioptric stereo system with coaxially-aligned spherical mirrors.  $F$  is the pinhole of the camera.

Note in (1) and (2) that  $\alpha$  and  $\beta$  are highly nonlinear functions of  $R$ ,  $r$ , and  $H$ . The vertical field-of-view (*FOV*) and the imaged radii  $R'$  and  $r'$  (Fig. 2) of the *major* and *minor* mirrors are of interest to us. Ideally, the imaged radii must be of comparable resolution so that sufficient detail is preserved in both mirrors for disparity matching. We define relative resolution as the ratio  $R'/r'$ , which we can approximate with  $\alpha/\beta$  given a sufficiently narrow field-of-view ( $2\beta$ ) of camera  $F$  (justified by the design constraints discussed next).



**Fig. 2** Imaged mirrors as observed by  $F$ .  $R'$  and  $r'$  are the radii of the imaged *major* and *minor* mirrors respectively.

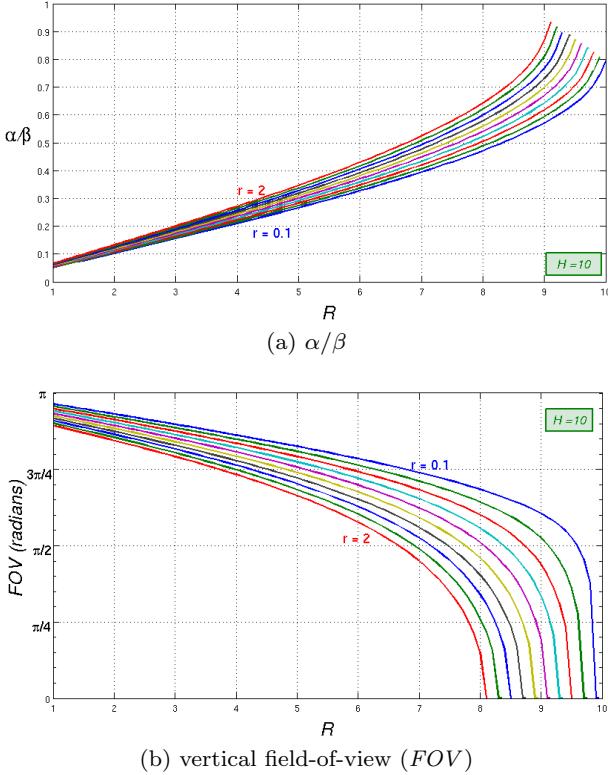


**Fig. 3** Virtual camera  $F'$  can be approximated to be at the midpoint between the center of the *minor* mirror and its surface (given  $H \gg r$ ).

It is convenient if the camera ( $F$  in Fig. 1) can be decomposed into two cameras:  $F$  itself and a second virtual camera  $F'$  (Fig. 3) that observes the *major* mirror directly. The two cameras could then be assumed to image the two spherical mirrors independently, thus, simplifying the analysis and calibration. Such decomposition is possible if the *major* mirror can be assumed to be imaged from a single viewpoint. While SVP is not satisfied by spherical mirrors in general, it can be approximated to arbitrary precision given that the locus of the effective viewpoint that images the *major* mirror alone is sufficiently compact. A caustic (locus of the effective viewpoint) for a spherical mirror was computed parametrically by Baker [2]. It can be shown that when the pinhole is sufficiently far from the *minor* mirror and the incoming rays are close to the axis of radial symmetry, the single effective viewpoint  $F'$  can be assumed to lie coaxially with the mirrors at a midpoint between the center and the surface of the *minor* mirror. This is illustrated in Fig. 3, where  $C$  (magenta) is the caustic of the *minor* mirror. Thus,  $F'$  is positioned at the cusp of caustic  $C$  when the conditions above are met. This translates into the design constraint requiring  $H$  (separation between the two mirrors’ centers) to be sufficiently larger than  $r$  (radius of *minor* mirror), and camera  $F$ ’s field-of-view ( $2\beta$ ) to be small enough to fit the entire *minor* mirror. We define what design constraints are “sufficient” when we return to the analysis of the viewpoint error ( $\delta\varphi_{SVP}$  in Fig. 3) introduced by this approximation (Subsection 3.2). Given that  $H \gg r$ , we can now approximate the real camera  $F$  (Fig. 1) to be orthographic, and define a virtual camera  $F'$  (Fig. 3) looking down, located at the cusp of the *minor* mirror’s caustic.

## 2.2 Field-of-view and Resolution

We consider the vertical field-of-view (*FOV*) of the imaging system and the imaged mirrors' ratio ( $\alpha/\beta$ ) as a function of the design parameters  $H$ ,  $R$ ,  $r$ . From Fig. 1, it is clear that the *FOV* is maximized when the ratio  $r/H$  is minimized (the constraint  $H \gg r$  introduced in the previous section facilitates the approximation of the virtual camera  $F'$ ). While reducing  $r$  increases the *FOV*, it proportionally reduces  $\alpha/\beta$ . To compensate for this reduction, we increase the radius of the *major* mirror ( $R$ ). In Fig. 4, we demonstrate the effect of  $R$  and  $r$  on *FOV* and  $\alpha/\beta$ .



**Fig. 4** Relationships between (a)  $\alpha/\beta$  and  $R$ , and (b) vertical field-of-view (*FOV*) and  $R$  are approximately independent of  $r$  when  $R \gg r$  and  $H \gg r$ .  $H$  is fixed at 10 units as  $r$  varies from 0.1 units to 2 units (10 samples).

When  $R \gg r$ , the system's field-of-view (*FOV*) is approximately independent of  $r$  (*minor* mirror's radius):

$$\text{FOV} = \pi - \tan^{-1} \left( \frac{R}{\sqrt{H^2 - R^2}} \right) \quad (3)$$

Another useful design observation is that the *FOV* (4b) and the ratio  $\alpha/\beta$  (4a) behave linearly with  $R$ , as long as  $R$  is sufficiently smaller than  $H$ . Putting these constraints together ( $H \gg R \gg r$ ) and linearizing the

Taylor expansion, we end up with:

$$\text{FOV} = \pi - \frac{R}{H} \quad (4)$$

$$\frac{\alpha}{\beta} = \frac{R}{\sqrt{2H}} \quad (5)$$

Equations (4) and (5) have been numerically verified to yield less than 10% deviation from the non-linearized model when  $H \geq 2R$  and  $R \geq 2r$ . The model indicates that the design is fairly tolerant to a wide selection of  $r$ ,  $R$  and  $H$ , while still satisfying the underlying assumptions.

## 3 Omnidirectional Stereo Geometry

### 3.1 Triangulation Model

We adapt the model of triangulation error introduced in [11] to include the distortion introduced by the two spherical mirrors. As in [11], we assume a normally distributed error in measured pixel coordinates with a variance  $\sigma_{px}^2$  of one pixel. From this point the images of *minor* and *major* mirrors are assumed to be viewed directly by  $F$  and  $F'$ , respectively. Practically, this is achieved by cropping the image of the *major* mirror from the original image and resampling both images to a common frame size. Eventually, we perform stereo matching along the rectified epipolar lines of the undistorted-panoramic images in a traditional fashion (like done by Kuthirummal and Nayar in [9]).

In more detail, let  $u$  and  $v$  (Fig. 5) be the radial positions of the pixels (in polar coordinates) in the images of  $F$  and  $F'$ , respectively. Because of radial symmetry, the azimuths of  $u$  and  $v$  have the same azimuths as the rays to which they project. We define projection functions  $f(u)$  and  $g(v)$  that map the pixels from their respective images to elevation angles  $\varphi$  and  $\theta$ , relative to the axis of radial symmetry. Practically, we compute the projection functions through two separate calibration procedures. However,  $f(u)$  lends itself to a simple analytic description due to the approximate orthographic projection of  $F$ . This facilitates a more compact formulation of depth error in triangulation that we derive in Subsection 3.2.

The distance  $d$  from the *major* mirror's (approximate) viewpoint to point  $P$  (Fig. 5) is given by:

$$d = h \frac{\sin \varphi}{\sin(\varphi + \theta)} \quad (6)$$

where  $h$  is the baseline (distance between the approximate viewpoints of the mirrors), and is always less than  $H$ .

### 3.2 Uncertainty Model

The convex polygon which bounds the region of uncertainty around point P (Fig. 5) is described exactly by a three-dimensional non-Gaussian pdf [11]. However, it can be approximated by a Gaussian under the assumption that projected pixel uncertainties  $\delta\theta$  and  $\delta\varphi$  are sufficiently small to define a parallelogram. We can then write the uncertainty as a product of the two independent Gaussians:  $\mathcal{N}(\mu_d, \sigma_d^2)$  (depth uncertainty) and  $\mathcal{N}(\mu_\theta, \sigma_\theta^2)$  (elevation uncertainty). Assigning the origin to the viewpoint of the *minor* mirror, (7) shows that the depth uncertainty depends on the resolution of the *minor* mirror (imaged by camera  $F'$  through projection function  $f(u)$ ), while elevation uncertainty depends on the resolution of the *major* mirror (captured by camera  $F'$  through  $g(v)$ ):

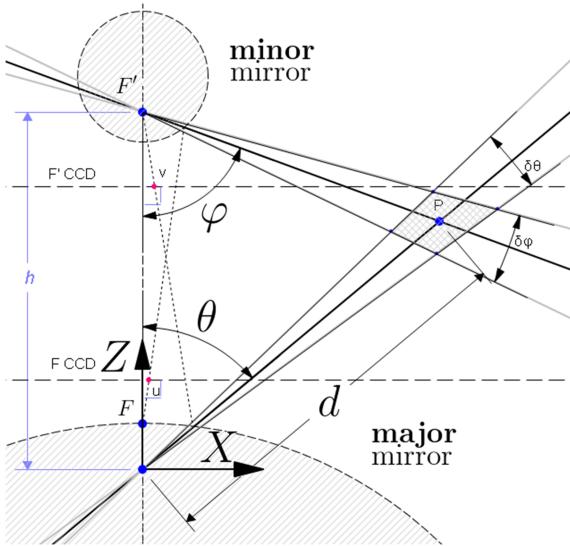
$$\begin{pmatrix} \sigma_d^2 \\ \sigma_\theta^2 \end{pmatrix} = \begin{pmatrix} \frac{\partial d}{\partial \varphi} \frac{df(u)}{du} & 0 \\ 0 & \frac{dg(v)}{dv} \end{pmatrix} \begin{pmatrix} \sigma_{px}^2 \\ \sigma_{py}^2 \end{pmatrix} \quad (7)$$

As outlined in Section 2, if the  $H \gg R \gg r$  constraint is satisfied, we can approximate camera  $F$  to have an orthographic projection. We can then analytically compute the depth uncertainty to be:

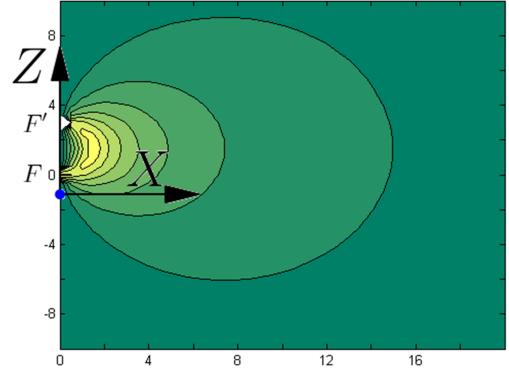
$$\sigma_d^2 = \left\{ h \frac{\sin \theta}{\sin(\varphi + \theta)} \right\} \left\{ \frac{2}{R'^2 - u^2} \right\} \sigma_{px}^2 \quad (8)$$

where  $R'$  is the imaged radius of the *major* mirror (in pixels) which depends on relative imaged resolution ratio ( $\alpha/\beta$ ).

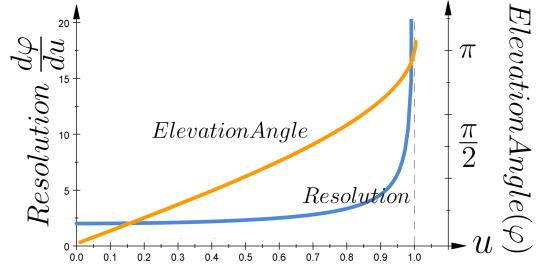
Fig. 6 illustrates the combined effect of the spatially-variant resolution and triangulation error, from which



**Fig. 5** Triangulation geometry model. The region of uncertainty around  $P$  is approximately a parallelogram when  $\delta\theta$  and  $\delta\varphi$  are small.



**Fig. 6** Error contours represent the combined effect of depth and elevation uncertainty  $\sqrt{\sigma_d^2 + \sigma_\theta^2}$ . Brighter areas represent lower uncertainty, and can be observed to lie dominantly in the equatorial plane of the view-sphere.



**Fig. 7** The approximately linear orthographic projection function  $f(u)$  provide nearly constant resolution for the majority of the image. The rapid loss of resolution near the edge of the image contributes to the poor depth resolution near the poles of the view-sphere (Fig. 6). Note that the  $u$ -axis appears normalized in this figure.

we conclude that the highest depth resolution lies in the equatorial region.

Note that because neither mirror satisfies the SVP, a point approximation of the effective viewpoint does not hold for points that are too close to the mirrors. Worst case depth error (greatest uncertainty  $\sigma_d^2$ ) will occur at the periphery of the mirror where  $\delta\varphi_{SVP}$  (angular difference between true projection and approximate-SVP projection) (Fig. 3) is greatest. The greatest angular (elevation) error of  $\delta\varphi_{SVP}$  is given by:

$$\delta\varphi_{SVP,max} = \frac{r}{h - \sqrt{d^2 - r^2}} \quad (9)$$

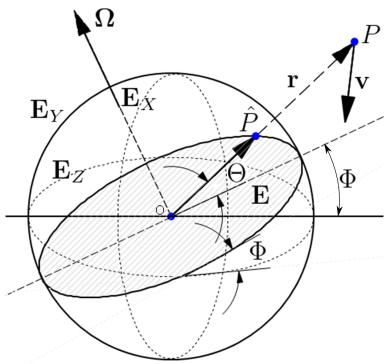
The error vanishes when the imaged point is much farther than the separation between the mirrors' viewpoints ( $d \gg h$ ). Numerical simulations result in  $\delta\varphi_{SVP,max}$  less than  $1^\circ$  for points with depth  $d > 3h$  and less than 5% error in depth at  $d = 3h$ . In practice, the effect of viewpoint error tends to be less significant as  $\delta\varphi_{SVP}$  is maximum near the periphery of the mirror, where depth error is dominated by deteriorating image resolution (Fig. 7).

## 4 Depth from Omnidirectional Optical Flow

Nelson and Aloimonos outline an algorithm for de-rotating and recovering depth from an omnidirectional optical flow field [15]. In what follows, we summarize the Nelson-Aloimonos algorithm, describe our method for recovering 3-D motion parameters and relative depth, and analyze the distribution of depth uncertainty on the view-sphere.

### 4.1 3-D De-rotation

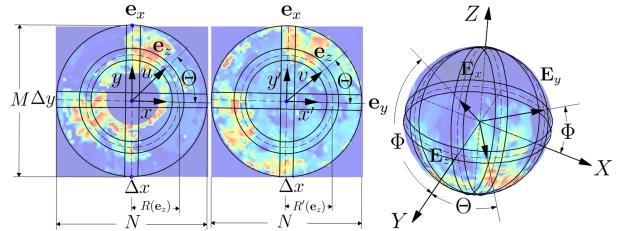
For a spherical camera moving with a linear velocity  $\mathbf{v}$  and rotational velocity  $\boldsymbol{\Omega}$ , a 3-D point  $P$  projects to a point  $\hat{P}$  on the view-sphere  $O$  and generates an optical flow field  $\mathbf{U}(\hat{P}) = [\dot{\Theta}, \dot{\Phi}]^T$  (Fig. 8).



**Fig. 8** View-sphere and the associated great circles for a spherical camera moving with linear velocity  $\mathbf{v}$  and angular velocity  $\boldsymbol{\Omega}$ . Depth  $\mathbf{r}$  to point  $P$  can be computed up to scale after the optical flow field is de-rotated.

As shown in [15], if  $\hat{P}$  is coplanar with a great circle  $\mathbf{E}_i$ , where  $i$  stands for either  $X$ ,  $Y$ , or  $Z$ , then the component of the optical flow generated by  $\hat{P}$  parallel to  $\mathbf{E}_i$  is independent of rotation and translation parallel to great circles orthogonal to  $\mathbf{E}_i$ . Thus, recovery of the 3-D vectors  $\mathbf{v}$  (up to scale factor) and  $\boldsymbol{\Omega}$  reduces to a sequential recovery of its components through a 2-D search for rotation and translation in the three mutually-orthogonal great circles (as opposed to a 3-D search on the entire sphere). To facilitate this recovery directly in the image space of  $F$  and  $F'$  (prior to unwrapping), it is convenient to choose a set of great circles  $\{\mathbf{E}_X, \mathbf{E}_Y, \mathbf{E}_Z\}$  because their projections yield a set of perpendicular lines  $\{\mathbf{e}_x, \mathbf{e}_y\}$ , which span the length and width of each image, and circles  $\mathbf{e}_z$  whose radii  $R(\mathbf{e}_z)$  and  $R'(\mathbf{e}_z)$  are fixed by the corresponding projection functions  $f(u)$  and  $g(v)$  (Fig. 9).

Rotation can be recovered in each great circle  $\mathbf{E}_i$  by finding a rotation  $\Omega_i$ , which after being factored



**Fig. 9** Projection of great circles onto image flow-fields  $F$  and  $F'$ . Great circles  $\mathbf{E}_X$  and  $\mathbf{E}_Y$  project to horizontal and vertical lines  $\mathbf{e}_x$ ,  $\mathbf{e}_y$ , but  $\mathbf{E}_Z$  projects to circles  $\mathbf{e}_z$  whose radii  $R(\mathbf{e}_z)$  and  $R'(\mathbf{e}_z)$  are fixed by projection functions  $f(u)$  and  $g(v)$ .

out, partitions the flow field into symmetric halves of clockwise and counterclockwise flows. For each great circle  $\mathbf{E}_i$ , we adapt the distance metric  $d_i$  (defined in [15]) to recover the direction of translation  $\psi_i$  and rotation  $\Omega_i$ , parallel to  $\mathbf{E}_i$ , such that:

$$d_i = \int_{-\pi}^{\pi} \sigma(\mathbf{U}_i(\theta_i) - \Omega_i, \psi_i - \theta_i) d\theta \quad (10)$$

where we define  $\theta_i$  to be the polar angle in the plane of  $\mathbf{E}_i$ , and  $\mathbf{U}_i(\theta_i)$  to be the component of optical flow in the direction  $\theta_i$ , also parallel to  $\mathbf{E}_i$ . In (10),  $\sigma$  is defined to be:

$$\sigma(a, \varrho) = \begin{cases} a & \text{if } \operatorname{sgn}(a) = \operatorname{sgn}(\varrho) \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

The direction of translation and rotation (parallel to  $\mathbf{E}_i$ ) are recovered when the “distance” between a purely translational flow and the current flow are minimized. The distance metric  $d_i$  in (10) is general and does not account for the projection of optical flow from image space to the view-sphere. The only exception is the equator  $\mathbf{E}_Z$ , which projects to an isocontour ( $\mathbf{e}_z$ ) of the projection functions  $f(u)$  and  $g(v)$  (due to the radial symmetry of  $\mathbf{e}_z$ ). Incidentally, the only known implementation of (10) de-rotates exclusively about the equator and thus utilizes the equation directly [12].

Adapting (10) to the remaining great circles  $\mathbf{E}_X$  and  $\mathbf{E}_Y$  is complicated by the fact that lines  $\mathbf{e}_x$ ,  $\mathbf{e}_y$  project on the images of  $F$  and  $F'$  as complementary portions of the great circles. Therefore, we essentially formulate (12) only for great circle  $\mathbf{E}_Z$  in the images of  $F$  and  $F'$  by splitting and discretizing (10) into two sums (corresponding to each half of view-sphere  $O$ ). Although we require both distance metrics  $d_x$  and  $d_y$  for the  $X$  and  $Y$  components, respectively, we only show

the discretized distance metric for the  $X$ -component:

$$d_x = \left\{ \sum_{u=0}^{R(\mathbf{e}_z)} \sigma \left( \frac{df}{du} \mathbf{F}_x \left( \frac{\pi}{2}, u \right) - \Omega_x, \psi_x - \mathbf{f}(u) \right) + \dots \right. \\ \left. + \sum_{v=0}^{R'(\mathbf{e}_z)} \sigma \left( \frac{dg}{dv} \mathbf{F}'_x \left( \frac{\pi}{2}, v \right) - \Omega_x, \psi_x - \mathbf{g}(v) \right) \right\} \quad (12)$$

On the other hand,  $d_z$  must be computed as the tangential component of the flow within  $\mathbf{e}_z$ .

We define  $[\theta, u]$  and  $[\theta', v]$  to be the polar image coordinates of  $F$  and  $F'$  with the origin in the center of the images (Fig. 9). Let  $\mathbf{F}(\theta, u)$  and  $\mathbf{F}'(\theta', v)$  be the optical flow fields defined as  $[\dot{x}, \dot{y}]$  and  $[\dot{x}', \dot{y}']$ , similar to  $\mathbf{U}$ , but in the image space of  $F$  and  $F'$ , respectively.  $\mathbf{F}_x$  and  $\mathbf{F}_y$  are components of optical flow field  $\mathbf{F}$  parallel to  $\mathbf{e}_x, \mathbf{e}_y$  that project to spherical optical flow  $\mathbf{U}_X$  and  $\mathbf{U}_Y$ , respectively.

$\mathbf{F}$  and  $\mathbf{F}'$  are computed densely using a block-matching optical flow algorithm implemented in the OpenCV library (`cv::CalcOpticalFlowBM`). While (12), in general, must be computed about an infinitesimally thin great circle, practically, we compute the integral as a sum along a strip of non-zero width (we choose  $\Delta x$  and  $\Delta y$  to be 10 pixels). Nelson and Aloimonos calculate the error introduced by this offset and report it to have a linear relationship with the offset. In our case, for  $\Delta x = 10$  px,  $\Delta y = 10$  px, in an 800x600 pixels image, the mean angular error in recovered ego-motion vectors is  $< 2\%$ . In practice, the error is partially compensated by the increased number of samples along the width of the strip (some of which may be null due to lack of texture).

## 4.2 Depth Error Analysis

After  $\hat{\mathbf{v}}$  and  $\Omega$  have been recovered, relative depth  $\tau$  is computed on the view-sphere by applying (13) to  $\mathbf{F}$  and  $\mathbf{F}'$ :

$$\tau = \frac{\|\hat{\mathbf{v}} \times \hat{\mathbf{r}}\|}{\|\mathbf{U}\|} \quad (13)$$

where  $\tau$  is also defined as  $\|\hat{\mathbf{r}}\| / \|\hat{\mathbf{v}}\|$  and is often referred to as time-to-contact in robotics [18] and biology literature.  $\hat{\mathbf{r}}$  is a unit vector in the direction of ray  $[\Theta, \Phi]$  on view-sphere  $O$  (Fig. 8), and  $\|\mathbf{U}\|$  is the magnitude of the spherical optical flow and can be computed from image flow  $\mathbf{F}$  and  $\mathbf{F}'$  with the Jacobian of the projection functions  $f$  and  $g$ , respectively. The source of depth error computed from optical flow is attributed to the quantization of the flow vectors in the image. We can

thus formulate the error as a function of optical flow magnitude:

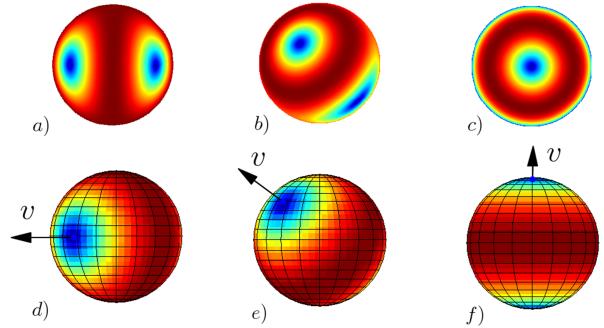
$$\frac{\partial \tau}{\partial \|\mathbf{F}\|} = \frac{\|\hat{\mathbf{v}} \times \hat{\mathbf{r}}\|}{\|\mathbf{U}^2\|} \cdot \frac{\partial \|\mathbf{U}\|}{\partial \|\mathbf{F}\|} \quad (14)$$

We remind the reader that the projection functions of *minor* and *major* mirrors are approximately linear for the majority of the image (Fig. 7), thus, making it possible to neglect the transformation between image flow and spherical flow within that region. If relative depth is computed with (13), depth error can be approximated to be:

$$\frac{\partial \tau}{\partial \|\mathbf{F}\|} = \frac{\tau}{\|\mathbf{F}\|} \quad (15)$$

From numerical simulations, error introduced by the approximation (15) is less than 10% for pixels bounded by  $\mathbf{e}_z$ , and since the entire view-sphere is contained within  $\mathbf{e}_z$  of  $F$  and  $F'$ , we suffice by computing depth from the regions bounded by  $\mathbf{e}_z$ . As in (7), we formulate the uncertainty in terms of pixel variance to be  $\sigma_\tau^2 = (\tau / \|\mathbf{F}\|) \sigma_{px}^2$ .

Note that the distribution of error on a spherical image has a regular pattern (Fig. 10) that is shaped by two factors: 1) direction of translation  $\hat{\mathbf{v}}$ , and 2) non-linearity of the projection functions  $f$  and  $g$ . Qualitatively, depth error is largest in the direction of  $\hat{\mathbf{v}}$  and near the edge of the image (Fig. 10). When  $\hat{\mathbf{v}}$  is parallel to the equatorial plane  $\mathbf{E}_Z$ , the error is largest at the antipodal points on the equator joined by  $\hat{\mathbf{v}}$ . Note that the error distribution is complementary to triangulation error as described in Section 3, where it is greatest near the poles (Fig. 6). Next we describe the method for fusing depth-from-optical-flow and stereo.



**Fig. 10** Depth error from optical flow as a function of  $\hat{\mathbf{v}}$  on the view-sphere (d,e,f) and projected on the image plane (a,b,c). Greatest error (blue) is found at the antipodes (blue) joined by  $\hat{\mathbf{v}}$ . For a camera moving horizontally (a,d), the greatest error is found in the equatorial plane  $\mathbf{E}_Z$ . Compare this to stereo error distribution in Fig. 6, where greatest error lies near the poles.

## 5 Fusing Depth from Optical Flow and Stereo

Let  $[\Theta_m, \Phi_n] \in \mathbb{R}^2$  be a pixel in an  $M \times N$  spherical-panorama image  $S$  that projects to ray  $[\Theta, \Phi]$  on the view sphere  $O$  (Fig. 8). For each pixel in  $S$ , we define depth  $r$  with a normal error distribution  $\mathcal{N}(\mu_r, \sigma_r^2)$  which is obtained by fusing stereo and optical-flow depth measurements. Fusion proceeds in three steps: 1) stereo and optical flow depth measurements (and their variances) are mapped into their respective spherical panorama images  $S_{ster}$  and  $S_{OF}$ , 2) scale factor for optical-flow depth is recovered, and 3) metric depth  $r$  and variance  $\sigma_r^2$  are computed for every pixel in  $S$ .

$S_{OF}$  is obtained by mapping the computed value of relative depth  $\tau$  using (13) from the optical-flow images  $\mathbf{F}$  and  $\mathbf{F}'$  to pixel  $[\Theta_m, \Phi_n]$  in  $S_{OF}$ . The computation of  $S_{ster}$  is a prerequisite to the process of disparity matching, and therefore need not be computed again. Using (7) and (15), we compute stereo and optical-flow depth variances  $\sigma_{d,ster}^2(\Theta_m, \Phi_n)$  and  $\sigma_{\tau,OF}^2(\Theta_m, \Phi_n)$ , respectively.

We can recover the scale for optical flow depth  $S_{OF}(\Theta_m, \Phi_n)$  by searching for a scale factor  $\rho$  that minimizes the Euclidean distance between  $\rho S_{OF}(\Theta_m, \Phi_n)$  and  $S_{ster}(\Theta_m, \Phi_n)$  for pixels in panoramas  $S_{ster}$  and  $S_{OF}$  where both measurements are available. We perform weighted least squares regression to account for the spatially-variant depth resolution in different regions of the view-sphere:

$$\rho = \arg \min_{\rho} \sum_{\Theta_m}^M \sum_{\Phi_n}^N d_{\Theta_m, \Phi_n}^2(\rho) \quad (16)$$

where the residual  $d_{\Theta_m, \Phi_n}(\rho)$  is given by:

$$d_{\Theta_m, \Phi_n}(\rho) = \frac{S_{ster}(\Theta_m, \Phi_n) - \rho S_{OF}(\Theta_m, \Phi_n)}{\sigma_{d,ster}^2(\Theta_m, \Phi_n) \sigma_{\tau,OF}^2(\Theta_m, \Phi_n)} \quad (17)$$

Finally, we fuse  $S_{ster}(\Theta_m, \Phi_n)$  and  $\rho S_{OF}(\Theta_m, \Phi_n)$  in  $S$  by assuming independence between stereo and optical flow error [10]. For space economy, we let  $S_{ster}(\Theta_m, \Phi_n)$ ,  $S_{OF}(\Theta_m, \Phi_n)$ ,  $\sigma_{d,ster}^2(\Theta_m, \Phi_n)$  and  $\sigma_{\tau,OF}^2(\Theta_m, \Phi_n)$ , be  $d_{ster}$ ,  $\tau_{OF}$ ,  $\sigma_{d,ster}^2$ ,  $\sigma_{\tau,OF}^2$ , respectively. Estimated depth  $r$  and variance  $\sigma_r^2$  are then given by:

$$r = \frac{d_{ster} \sigma_{\tau,OF}^2 + \tau_{OF} \rho^2 \sigma_{d,ster}^2}{\sigma_{d,ster}^2 + \rho^2 \sigma_{\tau,OF}^2} \quad (18)$$

$$\sigma_r^2 = \frac{1}{\sigma_{d,ster}^{-2} + \rho^{-2} \sigma_{\tau,OF}^{-2}} \quad (19)$$

## 6 Experiments

### 6.1 Simulations

Simulations were conducted with synthetic imagery rendered with *POV-Ray* (open-source ray-tracer), which in addition to photorealistic rendering, is able to accurately model complex catadioptric systems such as ours. The simulated rig was designed with the parameters in Table 1, where values satisfy the design constraints  $H \geq 2R$  and  $R \geq 2r$  for the model.

**Table 1** Design Parameters for a Simulated System

Parameter	$R$	$r$	$H$	$\alpha/\beta$	$FOV$
Value	7	1	15	1/3	153°

Values satisfy the design constraints  $H > R >> r$  for the model

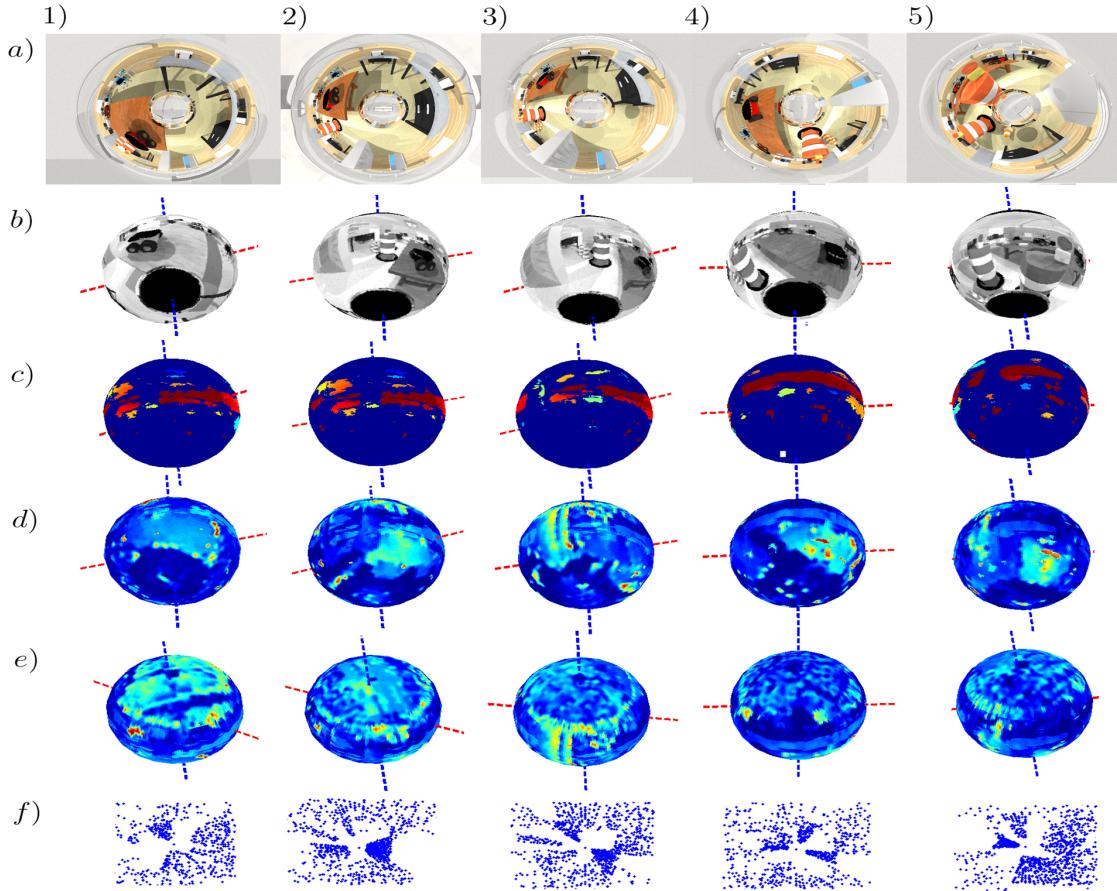
We follow guidelines set forth in Section 2 ( $H \geq 2R$  and  $R \geq 2r$ ) and derive the parameters in Table 1 using (4) and (5). As outlined in the model, both  $F$  and  $F'$  are treated independently, allowing us to calibrate them separately using *OCamCalib*, an omnidirectional camera calibration toolbox developed by Scaramuzza [17] in order to obtain  $f(u)$  and  $g(v)$ . To evaluate the accuracy of estimated depth (in the fused panorama image  $S$ ), we generate a fly-through sequence in a simulated cluttered lab environment (Fig. 11). Translation and rotation are dominant in the equatorial plane with pitch ranging within 20° from the equator as the camera completes a loop around the table. For each pixel  $[\Theta_m, \Phi_n]$  in spherical panorama  $S$ , we compute a normalized Euclidean distance to ground truth depth  $z(\Theta_m, \Phi_n)$  extracted from the simulated scene:

$$d(\Theta_m, \Phi_n) = \sqrt{\frac{(S(\Theta_m, \Phi_n) - z(\Theta_m, \Phi_n))^2}{\sigma_r^2(\Theta_m, \Phi_n)}} \quad (20)$$

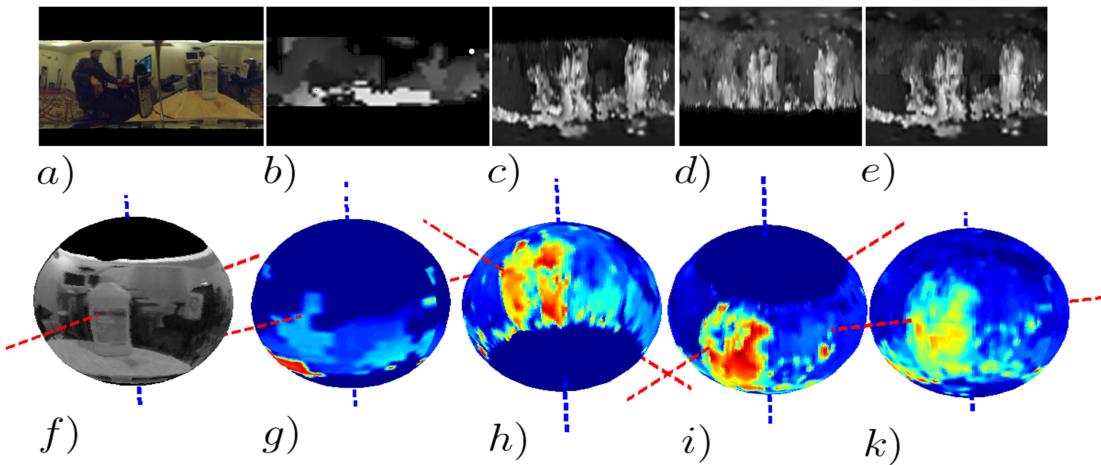
For the simulated fly-through sequence (50 frames), we tabulate (Table 2) the standard deviation of rotational error (defined as the absolute angular differences between estimated and ground truth orientation) and translational error (defined as the absolute angular difference between estimated and ground truth translation directions) recovered using the method outlined

**Table 2** Egomotion and Depth Error (Simulation)

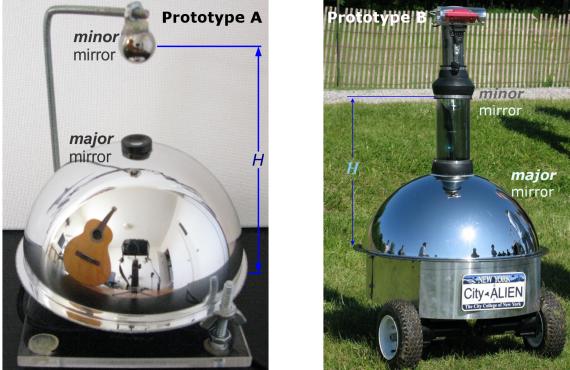
Trans. Dir. (elevation) (degrees)	Trans Dir. Error $\sigma$ (degrees)	Rotational Error $\sigma$ (degrees)	Average Depth Error (%)
0	2.4	7.5	5.6
5	2.5	7.7	10.1
10	3.2	5.6	16.7
15	2.9	6.5	23.6



**Fig. 11** A simulated fly-through sequence (1-5) in a cluttered lab environment. a) original images as observed by  $F$  b) projection of  $F$  on view-sphere c) projection of depth panorama  $S_{ster}$  from stereo  $S_{ster}$  on view-sphere. d) projection of fused depth panorama  $S$  on view-sphere as observed from below and e)  $S$  as observed from above. f) projection of the point cloud ( $50^\circ$  section in elevation about the equator) on the equatorial plane. Red and blue dashed lines are aligned with recovered motion vectors  $\hat{v}$  and  $\Omega$  respectively. Notice the following: as predicted (Sec. 3), stereo (c) generates depth mainly near the equator, while when fused with  $S_{OF}$  (d, e) generates depth near the poles (notice light fixture on the ceiling in 2.e)



**Fig. 12** Spherical depth from a real-world cluttered room environment. From left to right: a) unwrapped spherical panorama of  $F$ , b) stereo depth panorama  $S_{ster}$ , c) partial  $S_{OF}$  from  $F$  image only, d) partial  $S_{OF}$  from  $F'$  image only, e) fused depth and stereo panorama  $S$ . (f,g,h,i,k) are projections of (a,b,c,d,e) on the view-sphere. Red and blue dashed lines are aligned with recovered motion vectors  $\hat{v}$  and  $\Omega$ , respectively. Notice the following: as predicted in Sec. 3, stereo depth (g) is available only near the equator, while (h) and (i) generate depth in the opposite poles of the view-sphere (for motion in the equatorial plane) Note the foreground objects from left to right in first row: person, telephone handset, canister.



**Fig. 13** Real-world prototypes. (Only Prototype A was used for experimentation in an indoor cluttered setting).

**Table 3** Design Parameters for the Real-World Prototypes

Parameter	$R$	$r$	$H$	$\alpha/\beta$	FOV
Prototype A	5.25cm	0.7cm	10.5cm	$0.35 \approx 1/3$	151°
Prototype B	40.6cm	5.25cm	71.1cm	$0.40 \approx 2/5$	147.3°

Values satisfy the design constraints  $H > R >> r$  for the models

in Section 4. Depth error for the entire panorama  $S$  is measured by computing the mean of  $d(\Theta_m, \Phi_n)$  over all the available measurements in the panorama  $S$ . We parameterize Table 2 by the elevation angle of translation to analyze the effect of translation direction on depth fusion.

The average depth error for panorama  $S$  grows with increased elevation angle in the direction of translation. As explained in Sections 3 and 4, depth error from optical flow and stereo generate most efficient spherical coverage when the motion is in the equatorial plane. Deviations cause greater errors in areas where neither modality provides sufficient depth resolution.

## 6.2 Real-world experiments

Two prototypes (A and B) were constructed (Fig. 13) with the parameters listed in Table 3.

Both prototypes were calibrated with the *OCam-Calib Toolbox* [17]. Prototype A was tested in a cluttered room environment, and spherical depth was computed in three discrete locations in the room. The sequential generation of  $S$  for one location is depicted in Fig. 12. While no ground truth data is available, the resulting depth panoramas as well as variance distributions appear similar qualitatively to the results obtained in simulation.

While prototype A was designed for experimentation, prototype B is a novel concept of catadioptric “enclosure” for an autonomous ground vehicle where the robot’s body acts as a one giant “folded” catadioptric system.

Prototype B was employed during *The 18th Annual Intelligent Ground Vehicle Competition* sponsored in part by the Association for Unmanned Vehicle Systems International (AUVSI) and the U.S. Department of Defense. Our mobile robot won the first-place in the event’s Design Competition. Since real-time fusion of the two modalities could not be achieved, the system was put into practice during the autonomous competition by avoiding nearby obstacles using only the stereoscopic modality that is mostly reliable around the equator of the view-sphere.

## 7 Conclusion

The presented work constructs a low-cost catadioptric rig by using one perspective camera and two spherical mirrors of different radii (in a folded configuration). We have demonstrated (theoretically and with off-line imagery) that the fusion of these two modalities (stereo and optical flow) can generate an almost-entire view-sphere of probabilistic range information (with exception of self-occluded areas on the poles). However, computing dense optical flow (used for de-rotation) has not allowed us to achieve real-time performance. In future work, we believe that sparse optical flow fields could be applied by adapting the Nelson-Aloimonos’ de-rotation algorithm. Also, the near-sightedness of the rig (which depends on the baseline as in any other stereo system) could be addressed via software manipulation of the horopter. We have noticed that deviations cause greater errors in areas where neither modality provides sufficient depth resolution (no-overlapping regions) and where information is not existing at all (self-occlusion regions). A deeper analysis of uncertainty in each modality will follow as well as achieving real-time performance of the system.

## References

1. S. Baker and S. Nayar. A theory of catadioptric image formation. In *Computer Vision, 1998. Sixth International Conference on*, pages 35–42, January 1998.
2. Simon Baker and K. Nayar. A tutorial on catadioptric image formation, 1998.
3. Eduardo L. L. Cabral, Jos? C. de Souza Junior, and Marcos C. Hunold. Omnidirectional stereo vision with a hyperbolic double lobed mirror. *Pattern Recognition, International Conference on*, 1:1–4, 2004.
4. Joseph Conroy, Gregory Gremillion, Badri Ranganathan, and J. Sean Humbert. Implementation of wide-field integration of optic flow for autonomous quadrotor navigation. *Autonomous Robots*, 27(3):189–198, 2009.
5. F. R Corrêa, V. C Guizilini, and J. O Junior. Omnidirectional stereovision system with two-lobe hyperbolic mirror for robot navigation. *ABCM Symposium Series in Mechatronics*, 2:653–660, 2006.

6. S. Derrien and K. Konolige. Approximating a single viewpoint in panoramic imaging devices. In *Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE International Conference on*, volume 4, pages 3931–3938 vol.4, 2000.
7. G. Jang, S. Kim, and I. Kweon. Single camera catadioptric stereo system. In *Proc. of Workshop on Omnidirectional Vision, Camera Networks and Nonclassical cameras (OMNIVIS2005)*, 2005.
8. J. H Kim and M. J Chung. Slam with omni-directional stereo vision sensor. *Intelligent Robots and Systems, Las Vegas, Nevada*, 2003.
9. Sujit Kuthirummal and Shree K. Nayar. Multiview radial catadioptric imaging for scene capture. *ACM Transactions on Graphics*, 25(3):916, July 2006.
10. R. Li and S. Sclaroff. Multi-scale 3d scene flow from binocular stereo sequences. *Computer vision and image understanding*, 110(1):75–90, 2008.
11. L. Matthies and S. Shafer. Error modeling in stereo navigation. *IEEE Journal of Robotics and Automation*, 3(3):239–248, 1987.
12. C. McCarthy, N. Barnes, and M. Srinivasan. Real time biologically-inspired depth maps from spherical flow. In *2007 IEEE International Conference on Robotics and Automation*, pages 4887–4892, 2007.
13. S. Nayar. Omnidirectional vision. In *Robotics Research International Symposium*, volume 8, pages 195–202, 1998.
14. S. K Nayar and V. Peri. Folded catadioptric cameras. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 217–223, 1999.
15. R. C. Nelson and J. Aloimonos. Finding motion parameters from spherical motion fields (or the advantages of having eyes in the back of your head). *Biological Cybernetics*, 58(4):261–273, 1988.
16. Proceedings of TARO. *Omnidirectional vision simulation and robot localisation*, volume 2006, 2006.
17. D. Scaramuzza. Omnidirectional vision: from calibration to robot motion estimation. *ETH Zurich, PhD Thesis*, 17635, 2008.
18. K. Souhila and A. Karim. Optical flow based robot obstacle avoidance. *International Journal of Advanced Robotic Systems*, 4(1):13–16, 2007.
19. L. Su, C. Luo, and F. Zhu. Obtaining obstacle information by an omnidirectional stereo vision system. In *2006 IEEE International Conference on Information Acquisition*, pages 48–52, 2006.