Fusing Optical Flow and Stereo in a Spherical Depth Panorama Using a Single-Camera Folded Catadioptric Rig

Igor Labutov, Carlos Jaramillo and Jizhong Xiao, *Senior Member, IEEE*

***Abstract*—We present a novel catadioptric-stereo rig consisting of a coaxially-aligned perspective camera and two spherical mirrors with distinct radii in a "folded" configuration. We recover a nearly-spherical dense depth panorama (360°x153°) by fusing depth from optical flow and stereo. We observe that for motion in a horizontal 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. We exploit this principle by modeling the depth resolution of optical flow and stereo in order to fuse them probabilistically in a spherical panorama. To aid the designer in achieving a desired field-of-view and resolution, we derive a linearized model of the rig in terms of three parameters (radii of the two mirrors plus axial separation from their centers). We analyze the error due to the violation of the Single Viewpoint (SVP) constraint and formulate additional constraints on the design to minimize the error. Performance is evaluated through simulation and with a real prototype by computing dense spherical panoramas in cluttered indoor settings.**

I. INTRODUCTION

Omnidirectional catadioptric systems have been applied to a range of important problems in robotics including egomotion estimation, reactive obstacle avoidance, and SLAM. 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 as 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. 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 [2]. 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 [3] uses coaxially aligned hyperbolic mirrors and a single perspective camera. Although the field-of-view (hereafter referring to *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. A small form-factor together with a scalable baseline can be achieved with a "folded" configuration first introduced by Nayar and Peri [4] and successfully demonstrated in a number of applications (none used in robotics), such as [5][6]. 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 being the work of Derrien and Konolige in [7], 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 optical flow. As proved by Nelson and Aloimonos [8], 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. [9] 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:

Manuscript received January 28, 2011. 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

Igor Labutov and Carlos Jaramillo are with the Computer Engineering Department, City College of New York (CUNY), New York, NY 10031 USA. Labutov's e-mail: igor.labutov@gmail.com — Jaramillo's e-mail: cjarami00@ccny.cuny.edu

Jizhong Xiao is with the Electrical Engineering Department, City College of New York, Convent Ave & 140th Street, New York, NY 10031 (e-mail: jxiao@ccny.cuny.edu; phone: 212-650-7268)

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.

II. Design

A. 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 2β. Rays that lay within a cone bounded by α image the *major* mirror through its reflection in the *minor* mirror, while rays bounded between α and β image the *minor* mirror directly. Note that α and β are highly nonlinear functions of R, r, and H.

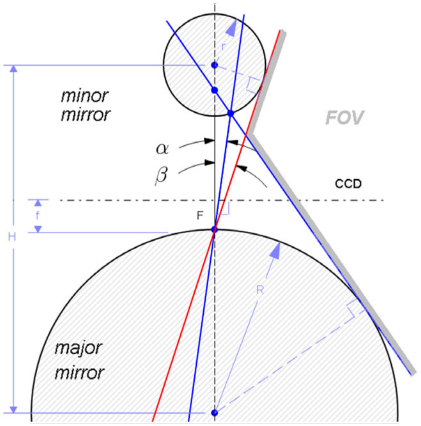

Fig. 1. "Folded" catadioptric stereo system with coaxially-aligned spherical mirrors. F is the pinhole of the camera.

The 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 α/β given a sufficiently narrow *FOV* (2β) of camera F (justified by the design constraints discussed next).

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. 4) 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 [10]. 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. 4 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 the field-of-view of the camera F to be small enough to fit the entire *minor* mirror in its FOV. We define what is "sufficient" when we return to the analysis of viewpoint error (Fig. 4) introduced by this approximation (end of Section III).

B. Field-of-view and Resolution

We consider the vertical field-of-view (*FOV*) of the imaging system and the imaged mirrors' ratio (α/β) 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 $H >> r$ introduced in the previous section facilitates the approximation of the virtual camera F'). While reducing r increases the *FOV*, it proportionally reduces α/β. To compensate for this reduction, we increase the radius of the *major* mirror R. In Fig. 3, we demonstrate the effect of R and r on the *FOV* and α/β. When $R>>r$, the system's *FOV* is approximately independent of r:

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

Another useful design observation is that the *FOV* and the ratio α/β behave linearly with R, as long as R is sufficiently smaller than H. Putting these constraints together

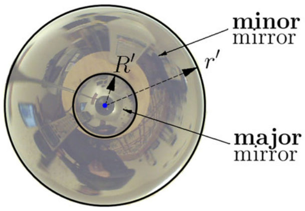

Fig. 2. Imaged mirrors as observed by F. R' and r' are the radii of the imaged *major* and *minor* mirrors respectively.

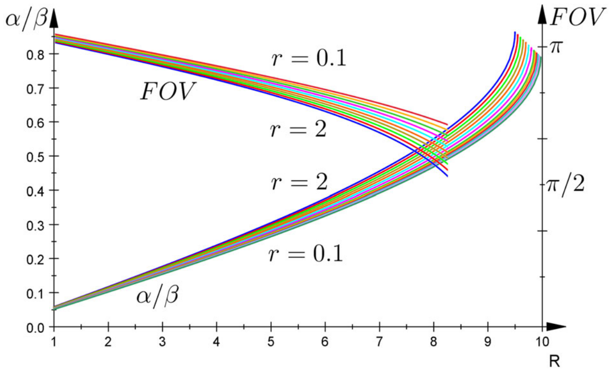

Fig. 3. Relationship between field-of-view (*FOV*) and α/β is approximately independent of r when $R >> r$ and $H >> r$. H is fixed at 10 units, as r varies from 0.1 units to 2 units (10 samples)

($H >> R >> r$) and linearizing the Taylor expansion, we get:

$$FOV = \pi - \frac{R}{H} \tag{2}$$

$$\frac{\alpha}{\beta} = \frac{R}{\sqrt{2}H} \tag{3}$$

Equations (2) and (3) 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.

III. Omnidirectional Stereo Geometry

A. 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 σ^2_{px} 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.

Let u and v (Fig. 4) be the radial pixel positions (polar coordinates) in the images of F and F', respectively. Because of radial symmetry, u and v have the same azimuths as the rays to which they project. We define projection functions $\mathrm{f}(u)$ and $\mathrm{g}(v)$ that map the pixels from their respective images to elevation angles φ and θ , relative to the axis of radial symmetry. Practically, we compute the projection functions through two separate calibration procedures. However, $\mathrm{f}(u)$ lends itself to a simple analytic description due to the approximate orthographic projection of F.

The distance d from the *major* mirror's (approximate) viewpoint to point P (Fig. 4) is given by:

$$d = h\frac{\sin\varphi}{\sin(\varphi + \theta)} \tag{4}$$

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

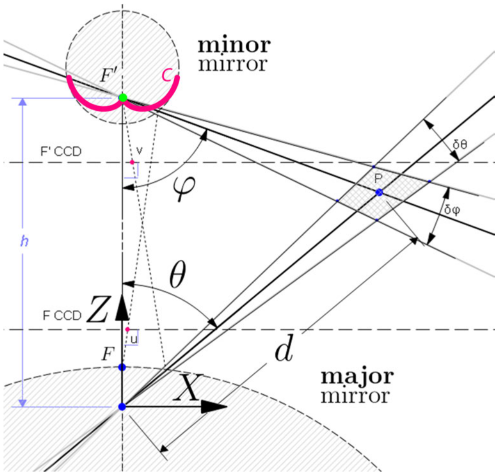

Fig. 4. Triangulation geometry model. The region of uncertainty around P is approximately a parallelogram when $\delta\theta$ and $\delta\varphi$ are small.

B. Uncertainty Model

The convex polygon which bounds the region of uncertainty around point P (Fig. 4) 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: $\mathcal{N}(\mu_d, \sigma^2_d)$ (depth uncertainty) and $\mathcal{N}(\mu_\theta, \sigma^2_\theta)$ (elevation uncertainty). Assigning the origin to the viewpoint of the *major* mirror, (5) shows that the depth uncertainty depends on the resolution of the *minor* mirror, while elevation uncertainty depends on the resolution of the *major* mirror:

$$\begin{pmatrix} \sigma^2_d \\ \sigma^2_\theta \end{pmatrix} = \begin{pmatrix} \frac{\partial d}{\partial \varphi}\frac{d\,\mathrm{f}(u)}{du} & 0 \\ 0 & \frac{d\,\mathrm{g}(v)}{dv} \end{pmatrix} \begin{pmatrix} \sigma^2_{px} \\ \sigma^2_{px} \end{pmatrix} \tag{5}$$

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 σ^2_d) will occur at the periphery of the mirror where $\delta\varphi_{SVP}$ (angular difference between true projection and approximate-SVP projection) (Fig. 4) is greatest. However, the SVP error vanishes when the imaged point is much farther than the separation between the mirrors' viewpoints ($d >> h$).

IV. Depth from Optical Flow

Nelson and Aloimonos outline an algorithm for de-rotating and recovering depth from an omnidirectional optical flow field [8]. In what follows, we summarize the Nelson- Aloimonos algorithm, describe our method for recovering 3-D motion parameters and relative depth.

For a spherical camera moving with a linear velocity $\mathbf{v}$ and rotational velocity $\mathbf{\Omega}$, 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. 5).

As shown in [8], 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

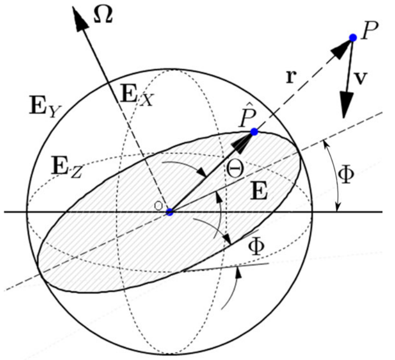

Fig. 5. View-sphere and the associated great circles for a spherical camera moving with linear velocity $\mathbf{v}$ and angular velocity $\mathbf{\Omega}$. Depth $\mathbf{r}$ to point P can be computed up to scale after the optical flow field is de-rotated.

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 $\mathbf{\Omega}$ 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 $\mathrm{f}(u)$ and $\mathrm{g}(v)$ (Fig. 6).

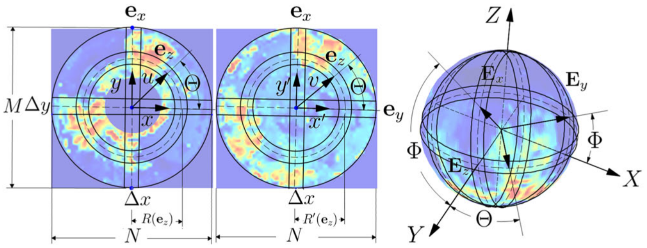

Fig. 6. Projection of great circles onto image flow-fields $\mathbf{F}$ and **F'**. Great circles $\mathbf{E}_X$ and $\mathbf{E}_Y$ project to horizontal and vertical lines $\mathbf{e}_x$, $\mathbf{e}_y$, but $\mathbf{E}_Z$ projects to circles $\mathbf{e}_z$ whose radii $R(\mathbf{e}_z)$ and $R'(\mathbf{e}_z)$ are fixed by projection functions $\mathrm{f}(u)$ and $\mathrm{g}(v)$.

Rotation can be recovered in each great circle $\mathbf{E}_i$ by finding a rotation Ω_i, which after being factored out, partitions the flow field into symmetric halves of clockwise and counter-clockwise flow. For each great circle $\mathbf{E}_i$, we adapt the distance metric d_i (defined in [8]) to recover the direction of translation ψ_i and rotation Ω_i, parallel to $\mathbf{E}_i$, such that:

$$d_i = \int_{-\pi}^{\pi} \sigma(\mathbf{U}_i(\theta_i) - \Omega_i, \psi_i - \theta_i) d\theta \tag{6}$$

where we define θ_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 θ_i, also parallel to $\mathbf{E}_i$. In (6), σ is defined to be:

$$\sigma(a, \varrho) = \begin{cases} a & \text{if } \mathrm{sgn}(a) = \mathrm{sgn}(\varrho) \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

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 (6) 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 $\mathrm{f}(u)$ and $\mathrm{g}(v)$ (due to the radial symmetry of $\mathbf{e}_z$). Incidentally, the only known implementation of (6) de-rotates exclusively about the equator and thus utilizes the equation directly [9].

Adapting (6) 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. We essentially formulate d_i in the images of F and F' by splitting and discretizing the distance metric (6) into two sums (corresponding to each half of view-sphere O).

Optical flow fields $\mathbf{F}$ and $\mathbf{F}'$ are computed densely using a block-matching optical flow algorithm implemented in the OpenCV library (`cv::CalcOpticalFlowBM`). Practically, we compute (6) as a sum along a strip of non-zero width (we choose Δx and Δy 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 (for Δx,Δy = 10 pixels in an 800x600 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).

After $\hat{\mathbf{v}}$ and $\mathbf{\Omega}$ have been recovered, relative depth τ is computed on the view-sphere by applying (8) to $\mathbf{F}$ and $\mathbf{F}'$:

$$\tau = \frac{\|\hat{\mathbf{v}} \times \hat{\mathbf{r}}\|}{\|\mathbf{U}\|} \tag{8}$$

where τ is also defined as $\|\hat{\mathbf{r}}\| / \|\hat{\mathbf{v}}\|$ and is often referred to as time-to-contact in robotics [12] and biology literature. $\hat{\mathbf{r}}$ is a unit vector in the direction of ray $[\mathbf{\Theta}, \mathbf{\Phi}]$ on view-sphere O (Fig. 5), 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 $\mathbf{f}$ and $\mathbf{g}$ respectively. If relative depth is computed with (8), depth error can be approximated to be:

$$\frac{\partial \tau}{\partial \|\mathbf{F}\|} = \frac{\tau}{\|\mathbf{F}\|} \tag{9}$$

As in (5), we formulate the uncertainty in terms of pixel variance to be $\sigma_\tau^2 = (\tau / \|\mathbf{F}\|)\sigma_{px}^2$.

V. 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. 5). 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 σ_r^2 are computed for every pixel in S.

S_{OF} is obtained by mapping the computed value of relative depth τ using (8) 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 (5) and (9), we compute stereo and optical-flow depth variances $\sigma^2_{d,ster}(\Theta_m, \Phi_n)$ and $\sigma^2_{\tau,OF}(\Theta_m, \Phi_n)$, respectively.

We can recover the scale for optical flow depth $S_{OF}(\Theta_m, \Phi_n)$ by searching for a scale factor ρ 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. 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[13]. For space economy, we let $S_{ster}(\Theta_m, \Phi_n)$, $S_{OF}(\Theta_m, \Phi_n)$, $\sigma^2_{d,ster}(\Theta_m, \Phi_n)$ and $\sigma^2_{\tau,OF}(\Theta_m, \Phi_n)$, be d_{ster}, τ_{OF}, $\sigma^2_{d,ster}$, $\sigma^2_{\tau,OF}$, respectively. Estimated depth r and variance σ_r^2 are:

$$r = \frac{d_{ster}\sigma^2_{\tau,OF} + \tau_{OF}\rho^2\sigma^2_{d,ster}}{\sigma^2_{d,ster} + \rho^2\sigma^2_{\tau,OF}} \tag{10}$$

$$\sigma_r^2 = \frac{1}{\sigma^{-2}_{d,ster} + \rho^{-2}\sigma^{-2}_{\tau,OF}} \tag{11}$$

VI. EXPERIMENTS

A. Simulations

Simulations were conducted with synthetic imagery rendered with *POV-Ray* (open-source ray-tracer). The simulated rig was designed with the parameters in Table I (first row), where values satisfy the design constraints $H \geq 2R$ and $R \geq 2r$ for the model.

TABLE I
DESIGN PARAMETERS

Parameter	R	r	H	α/β	FOV
Simulation:	7	1	15	$1/3$	153°
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					

TABLE II
EGOMOTION AND DEPTH ERROR (SIMULATION)

Trans. Dir. (elevation) (degrees)	Trans Dir. Error σ (degrees)	Rotational Error σ (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

We follow guidelines set forth in Section II ($H \geq 2R$ and $R \geq 2r$) and derive the parameters in Table I using (2) and (3). 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 [14] in order to obtain $\mathrm{f}(u)$ and $\mathrm{g}(v)$. 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. VI-A). 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 (by σ_r^2) Euclidean distance to ground truth depth $z(\Theta_m, \Phi_n)$ extracted from the simulated scene. For the simulated fly-through sequence (50 frames), we tabulate (Table II) 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 in Section IV. 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 II by the elevation angle of translation to analyze its effect on depth fusion.

The average depth error for panorama S grows with increased elevation angle in the direction of translation. As explained in Sections III and IV, depth error from optical flow and stereo generate most efficient spherical coverage when the motion is in the equatorial plane.

B. Real-world experiments

Two prototypes (A and B) were constructed (Fig. 8) with the parameters listed in Table I.

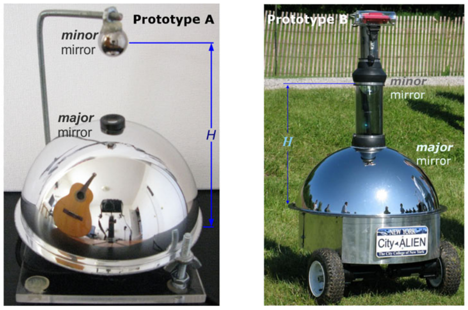

Fig. 8. Rig prototypes. (Only Prototype A was used for experimentation).

While prototype A was designed for experimentation, prototype B is a novel concept of catadioptric "enclosure" for an entire robot (the robot's body acts as a one giant "folded" catadioptric system). Prototype B debuted at *The 18th Annual Intelligent Ground Vehicle Competition* sponsored in-part by UAVSI and the U.S. Department of Defense. Our team won the

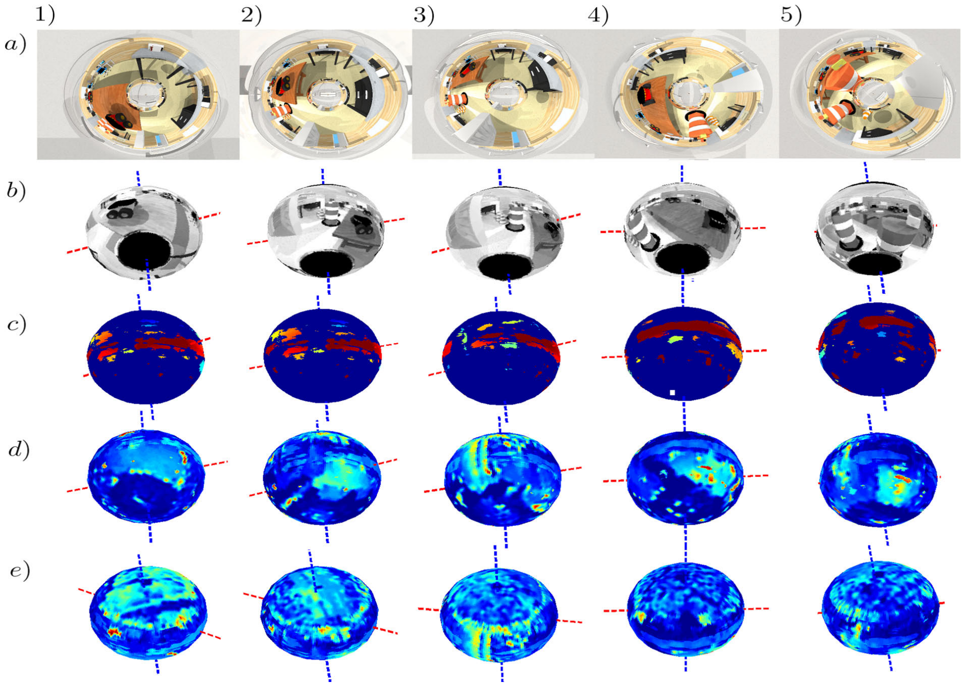

Fig. 7. A simulated fly-through sequence (1-5) in a cluttered lab environment. a) original images as observed by F b) projection of F on view-sphere c) projection of depth panorama from stereo S_{ster} on view-sphere. d) projection of fused depth panorama S on view-sphere as observed from below and e) S as observed from above. Red and blue dashed lines are aligned with recovered motion vectors $\hat{\mathbf{v}}$ and $\mathbf{\Omega}$ respectively. Notice the following: as predicted (Sec. III), stereo (c) generates depth mainly near the equator, while when fused with S_{OF} (d, e) generates depth near the poles.

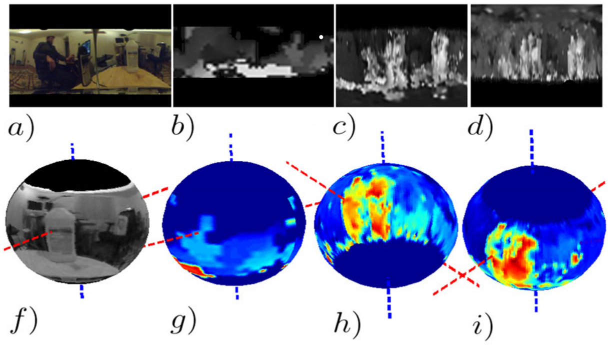

Fig. 9. Spherical depth from a real-world cluttered environment. From left to right: a) unwrapped spherical panorama of F, b) stereo depth panorama S_{ster}, c) partial S_{OF} from F image only, d) partial S_{OF} from F' image only. (f,g,h,i) are projections of (a,b,c,d) on the view-sphere. Red and blue dashed lines are aligned with recovered motion vectors $\hat{\mathbf{v}}$ and $\mathbf{\Omega}$, respectively. Notice the following: as predicted in Sec. III, 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).

first place in the event's Design Competition, and the system was also used during the autonomous navigation challenge.

Both prototypes were calibrated with *OCamCalib* [14]. 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.9. 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.

References

[1] S. Baker and S. Nayar, "A theory of catadioptric image formation," in *Computer Vision. Sixth International Conference on*, 1998, pp. 35–42.

[2] S. Nayar, "Omnidirectional vision," in *Robotics Research International Symposium*, vol. 8, 1998, pp. 195–202.

[3] L. Su, C. Luo, and F. Zhu, "Obtaining obstacle information by an omnidirectional stereo vision system," in *2006 IEEE International Conference on Information Acquisition*, 2006, pp. 48–52.

[4] S. K. Nayar and V. Peri, "Folded catadioptric cameras," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 1999, pp. 217–223.

[5] 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.

[6] E. L. L. Cabral, J. C. de Souza Junior, and M. C. Hunold, "Omnidirectional stereo vision with a hiperbolic double lobed mirror," *Pattern Recognition, International Conference on*, vol. 1, pp. 1–4, 2004.

[7] S. Derrien and K. Konolige, "Approximating a single viewpoint in panoramic imaging devices," in *International Conference on Robotics and Automation Proceedings. ICRA '00.*, vol. 4, 2000, pp. 3931–3938.

[8] R. 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*, vol. 58, pp. 261–273, 1988.

[9] 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*, 2007, pp. 4887–4892.

[10] S. Baker and K. Nayar. (1998) A tutorial on catadioptric image formation.

[11] L. Matthies and S. Shafer, "Error modeling in stereo navigation," *IEEE Journal of Robotics and Automation*, vol. 3, no. 3, pp. 239–248, 1987.

[12] K. Souhila and A. Karim, "Optical Flow based robot obstacle avoidance," *International Journal of Advanced Robotic Systems*, vol. 4, no. 1, pp. 13–16, 2007.

[13] R. Li and Sclaroff, "Multi-scale 3d scene flow from binocular stereo sequences," *Computer vision and image understanding*, pp. 75–90, 2008.

[14] D. Scaramuzza, "Omnidirectional vision: from calibration to robot motion estimation," *ETH Zurich, PhD Thesis*, vol. 17635, 2008.