

Article

# Design and Analysis of a Single–Camera Omnistereo Sensor for Quadrotor Micro Aerial Vehicles (MAVs) <sup>†</sup>

Carlos Jaramillo <sup>1</sup>, Roberto G. Valenti <sup>2</sup>, Ling Guo <sup>3</sup> and Jizhong Xiao <sup>2,\*</sup>

<sup>1</sup> Department of Computer Science, The Graduate Center, The City University of New York (CUNY), 365 Fifth Avenue, New York, NY 10016, USA; cjaramillo@gradcenter.cuny.edu

<sup>2</sup> Electrical Engineering Department, The City College, City University of New York (CUNY City College), Convent Ave & 140th Street, New York, NY 10031, USA; rvalenti@ccny.cuny.edu

<sup>3</sup> Automation Department, Nanjing University of Science and Technology (NUST), Nanjing 210094, China; laura0955@live.cn

\* Correspondence: jxiao@ccny.cuny.edu; Tel.: +1-212-650-7268; Fax: +1-212-650-8249

<sup>†</sup> This paper is an extended version of our paper published in Proceedings of the IEEE Conference on Industrial Electronics and Applications (ICIEA). Jaramillo, C.; Guo, L.; Xiao, J. A Single-Camera Omni-Stereo Vision System for 3D Perception of Micro Aerial Vehicles (MAVs), Melbourne, Australia, 19–21 June 2013; Volume 10016.

Academic Editor: Xin Zhao

Received: 24 November 2015; Accepted: 29 January 2016; Published: 6 February 2016

**Abstract:** We describe the design and 3D sensing performance of an omnidirectional stereo (omnistereo) vision system applied to Micro Aerial Vehicles (MAVs). The proposed omnistereo sensor employs a monocular camera that is co-axially aligned with a pair of hyperboloidal mirrors (a vertically-folded catadioptric configuration). We show that this arrangement provides a compact solution for omnidirectional 3D perception while mounted on top of propeller-based MAVs (not capable of large payloads). The theoretical single viewpoint (SVP) constraint helps us derive analytical solutions for the sensor's projective geometry and generate SVP-compliant panoramic images to compute 3D information from stereo correspondences (in a truly synchronous fashion). We perform an extensive analysis on various system characteristics such as its size, catadioptric spatial resolution, field-of-view. In addition, we pose a probabilistic model for the uncertainty estimation of 3D information from triangulation of back-projected rays. We validate the projection error of the design using both synthetic and real-life images against ground-truth data. Qualitatively, we show 3D point clouds (dense and sparse) resulting out of a single image captured from a real-life experiment. We expect the reproducibility of our sensor as its model parameters can be optimized to satisfy other catadioptric-based omnistereo vision under different circumstances.

**Keywords:** catadioptrics; omnistereo; 3D perception; Micro Aerial Vehicles (MAVs)

## 1. Introduction

Micro aerial vehicles (MAVs), such as quadrotor helicopters, are popular platforms for unmanned aerial vehicle (UAV) research due to their structural simplicity, small form factor, vertical take-off and landing (VTOL) capability, and high omnidirectional maneuverability. In general, UAVs have plenty of military and civilian applications, such as target localization and tracking, 3-dimensional (3D) mapping, terrain and infrastructural inspection, disaster monitoring, environmental and traffic surveillance, search and rescue, deployment of instrumentation, and cinematography, among other uses. However, MAVs have size, payload, and on-board computation limitations, which involve the use of compact and lightweight sensors. The most commonly used perception sensors on MAVs are laser scanners and cameras in various configurations such as monocular, stereo, or omnidirectional.

We present a vision-based omnidirectional stereo (omnistereo) sensor motivated by several aspects of MAV robotics.

### 1.1. Sensor Motivation

We justify the need for the proposed omnistereo sensor after observing two basic differences in the sensor requirements between MAVs and ground vehicles:

1. Size and payload—In MAV applications, the sensor's physical dimensions and weight are always a great concern due to payload constraints. Generally, MAVs require fewer and lighter sensors that are compactly designed, while larger robots (including high-payload UAVs) have greater freedom of sensor choice.
2. Field-of-view (FOV)—Due to their omnidirectional motion model, MAVs require a simultaneous observation of the 3D surroundings. Conversely, most ground robots can safely rely upon narrow vision as their motion control on the plane is more stable.

### 1.2. Existing Range Sensors for MAVs

In addition to specifying our sensor requirements, it is important to note the most prevalent robot range sensors used today by MAVs and their limitations. For example, lightweight 2.5D laser scanners can accurately measure distances at fast rates, however, their instantaneous sensing is limited to plane sweeps, which in turn require the quadrotor to move vertically in order to generate 3D maps or to foresee obstacles and free space during navigation. More recently, 3D laser rangefinders and LiDARs are being developed, such as the sensor presented in [1], but this one is not compact enough for MAVs. Another disadvantage of laser-based technologies is their active sensing nature, which requires more power to operate and their measurements are more vulnerable to detection and corruption (e.g., due to dark/reflective surfaces) than vision-based solutions. Time-of-flight (ToF) cameras as well as red, green, blue plus depth (RGB-D) sensors like the Microsoft Kinect<sup>®</sup> are also very popular for robot navigation. They have been adopted for low-sunlight conditions and mainly indoor navigation of MAVs [2] due to its structured infrared light projection and short range sensing (under 5 m) [3]. Hence, a lightweight imaging system capable of instantly providing a large field of view (FOV) with acceptable resolutions is essential for MAV applications in 3D space. These state-of-the-art sensors' pitfalls motivate the design and analysis of our omnistereo sensor.

### 1.3. Related Work

Using omnidirectional images alone and motion—like the approaches taken in [4,5]—have been proposed to map and localize a robot. Omnidirectional vision using a single mirror for the flight of large UAVs was first attempted in [6]. In [7], Hrabar proposed the use of traditional horizontal stereo-based obstacle avoidance and path planing for AUVs, but these techniques were only tested in a scaled-down air vehicle simulator (AVS). Omnidirectional catadioptric cameras can be aided by structured light such as the prototypes presented in [8] and more flexible configurations demonstrated in [9]. Alternatively, stereo cameras can provide passive, instantaneous 3D information for robot mapping and navigation (including UAVs [10]). Intuitively, omnidirectional stereo (omnistereo) can be achieved through circular arrangements of multiple perspective cameras with overlapping views. Higher resolution panoramas can be achieved by rotating a linear camera as presented in [11], but this approach suffers from motion blur in dynamic environments. We point the reader to [12] for a detailed study of multiple view geometry, and [13] for a compendium of geometric computer vision concepts. Instead, our solution to omnistereo vision consists of a 'catadioptric' system by employing cameras and mirrors [14].

Throughout the years, [15–20] are some of the works that have applied various omnistereo catadioptric configurations for ground mobile robots. Unfortunately, these systems are not compact since they use separate camera-mirror pairs, which are known to experience synchronization issues.

In [21], Yi and Ahuja described a configuration using a mirror and a concave lens for omnistereo, but it rendered a very short baseline in comparison to the two-mirror configurations. Originally, Nayar and Peri [22] studied 9 possible folded-catadioptric configurations for a single-camera omnistereo imaging system. Eventually, a catadioptric system using two hyperbolic mirrors in a vertical configuration was implemented by He *et al.* [23]. Their omnistereo sensor provides a lengthy baseline at the expense of a very tall system. In the past [24], we developed a novel omnistereo catadioptric rig consisting of a perspective camera coaxially-aligned with two spherical mirrors of distinct radii (in a “folded” configuration). One caveat of spherical mirrors is their non-centrality; they do not satisfy the single effective viewpoint (SVP) constraint (discussed in Section 2.2) but rather a locus of viewpoints is obtained [25].

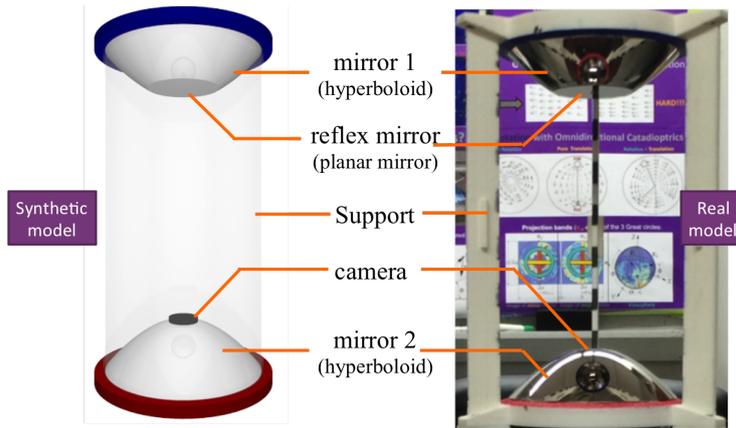
#### 1.4. Proposed Sensor

We design a SVP-compliant omnistereo system based on the folded, catadioptric configuration with hyperboloidal mirrors. Our approach resembles the work of Jang, Kim, and Kweon [26], who first implemented an omnistereo system using a pair of hyperbolic mirrors and a single camera. However, their sensor’s characteristics were not studied in order to justify their design parameters and capabilities, which we do in our case.

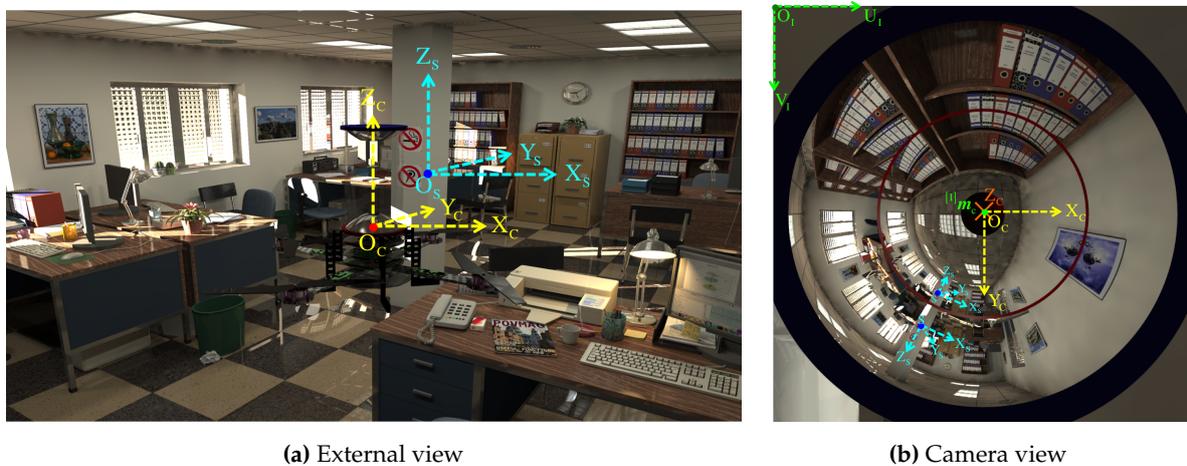
It is true that an omnidirectional catadioptric system sacrifices spatial resolution on the imaging sensor (analyzed in Section 3.4). However, our sensor offers practical advantages such as reduced cost, acceptable weight, and truly-instantaneous pixel-disparity correspondences since the same single camera-lens operates for both views, so mis-synchronization issues do not exist. In fact, we believe we are the first to present a single-camera catadioptric omnistereo solution for MAVs. The initial geometry of our model was proposed in [27]. Now, we perform an extensive analysis of our model’s parameters (Section 2) involving its geometric projection (Section 3) that are obtained as a constrained numerical optimization solution devising the sensor’s real-life application to MAVs passive range sensing (Section 4). We also show how the panoramic images are obtained, where we find correspondences and triangulate 3D points for which an uncertainty model is introduced (Section 5). Finally, we present our experimental results and evaluation for 3D sensing with the proposed omnistereo sensor (Section 6), and we discuss the future direction of our work in Section 7.

## 2. Sensor Design

Figure 1 shows the single-camera catadioptric omnistereo vision system that we specifically design to be mounted on top of our micro quadrotors (manufactured by Ascending Technologies [28]). It consists of (1) one hyperboloid-planar mirror at the top; (2) one hyperboloidal mirror at the bottom; and (3) a high-resolution USB camera also at the bottom (inside the bottom mirror and looking up). The components are housed and supported by a (4) transparent tube or plastic standoffs (for the real-life prototype shown in Figure 13). The choice of the hyperboloidal reflectors owes to three reasons: it is one of the four non-degenerated conic shapes satisfying the SVP constraint [29]; it allows a wider vertical FOV than elliptical and planar mirrors; and it does not require a telescopic (orthographic) lens for imaging as with paraboloidal mirrors (so our system can be downsized). In addition, the planar part of mirror 1 works as a reflex mirror, which in part reduces distortion caused by dual conic reflections. Based on the SVP property, the system obtains two radial images of the omnidirectional views in the form of an inner and an outer ring as illustrated in Figure 2a,b). Nevertheless, the unique set of parameters describing the entire system categorizes it as a “global camera model” given by [13] because changing the value of any parameter in the model affects the overall projection function of visible light rays in the scene as well as other computational imaging factors such as depth resolution and overlapping field of view, which we attempt to optimize with the following design subsections. Please, refer to Appendix A for clarification on our symbolic notation.



**Figure 1.** Synthetic and real prototypes for the catadioptric single-camera omnistereo system.



**Figure 2.** Photo-realistic synthetic scene: (a) Side-view of the quadrotor with the omnistereo rig in an office environment; (b) the image captured by the system's camera using this pose.

### 2.1. Model Parameters

In the configuration of Figure 3, mirror 1's real or primary focus is  $F_1$ , which is separated by a distance  $c_1$  from its virtual or secondary focus,  $F'_1$ , at the bottom. Without loss of generality, we make both the camera's pinhole and  $F'_1$  coincide with the origin of the camera's coordinate system,  $O_C$ . This way, the position of the primary focus,  $F_1$ , can be referenced by vector  ${}^{[C]}\mathbf{f}_1 = [0, 0, c_1]^T$  in Cartesian coordinates with respect to the camera frame,  $[C]$ . Similarly, the distance between the foci of mirror 2,  $F_2$  and  $F'_2$ , is measured by  $c_2$ . Here, we use the planar (reflex) mirror of radius  $r_{ref}$  and unit normal vector

$${}^{[C]}\hat{\mathbf{n}}_{ref} = [0, 0, -1] \quad (1)$$

in order to project the real camera's pinhole located at  $O_C$  as a virtual camera  $O_C'$  coinciding with the virtual focal point  $F'_2$  positioned at  ${}^{[C]}\mathbf{f}_{2v} = [0, 0, d]^T$ . We achieve this by setting  $d/2$  as the symmetrical distance from the reflex mirror to  $O_C$  and from the reflex mirror to  $O_C'$ . With respect to  $[C]$ , mirror 2's primary focus,  $F_2$ , results in position  ${}^{[C]}\mathbf{f}_2 = [0, 0, d - c_2]^T$ . It yields the following expression for the reflective plane:

$${}^{[C]}\hat{\mathbf{n}}_{ref}^T {}^{[C]}\mathbf{x} = -d/2 \quad (2)$$



field of view, spatial resolution, and depth resolution. In the upcoming subsections, we first derive the analytical solutions for the forward projection problem in our coaxial stereo configuration as a whole. In Section 3.2, we derive the back-projection equations for lifting 2D image points into 3D space.

## 2.2. Single Viewpoint (SVP) Configuration for OmniStereo

As a central catadioptric system, its projection geometry must obey the existence of the so-called single effective viewpoint (SVP). 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 conics. Generally, a circular hyperboloid of revolution (about its axis of symmetry) conforms to the SVP constraint as demonstrated by Baker and Nayar in [30]. Since a hyperboloidal mirror has two foci, the effective viewpoint is the primary focus  $F$  inside the physical mirror and the secondary (outer) focus  $F'$  is where the centre (pinhole) of the perspective camera should be placed for depicting a scene obeying the SVP configuration discussed in this section.

First of all, a hyperboloid  $i$  can be described by the following parametric equation:

$$\frac{(z_i - z_{0_i})^2}{a_i^2} - \frac{r_i^2}{b_i^2} = 1, \text{ with } a_i = \frac{c_i}{2} \sqrt{\frac{k_i - 2}{k_i}}, b_i = \frac{c_i}{2} \sqrt{\frac{2}{k_i}} \quad (5)$$

where  $z_{0_i} = \frac{c_i}{2}$  is the offset (shift) position of the focus along the Z-axis from the origin  $O_C$ , and  $r_i$  is the orthogonal distance to the axis of revolution / symmetry (i.e. the Z-axis) from a point  $P_i$  on its surface.

In fact, the position of a valid point  $P_i$  is constrained within the mirror's physical surface of reflection, which is radially limited by  $r_{i,min}$  and  $r_{i,max}$ , such that:

$$r_i = \sqrt{x_i^2 + y_i^2}, \text{ for } r_{i,min} \leq r_i \leq r_{i,max}, \forall i \in \{1, 2\} \quad (6)$$

and  $r_{1,min} = r_{ref}$ ,  $r_{1,max} = r_{sys}$ ,  $r_{2,min} = r_{cam}$ ,  $r_{2,max} = r_{sys}$ . Observe that the radius of the system is the upper bound for both mirrors (Figure 3). In addition, the hyperboloids profiled by Equation (5) must obey the following conical constraints:

$$\forall i \in \{1, 2\} (c_i > 0 \wedge k_i > 2) \quad (7)$$

$k$  is a constant parameter (unit-less) inversely related to the mirror's curvature or more precisely, the eccentricity  $\epsilon_c$  of the conic. In fact,  $\epsilon_c > 1$  for hyperbolas, yet a plane is produced when  $\epsilon_c \rightarrow \infty$  or  $k = 2$ .

We devise  $M_i$  as the set of all the reflection points  $P_i$  with coordinates  $(x_i, y_i, z_i)$  laying on the surface of the respective mirror  $i$  within bounds. Formally,

$$M_i := \left\{ P_i \in \mathbb{R}^3 \left| \frac{(z_i - z_{0_i})^2}{a_i^2} - \frac{r_i^2}{b_i^2} = 1 \wedge \text{Equation (6)} \wedge \text{Equation (7)} \right. \right\} \quad (8)$$

In our model, we describe both hyperboloidal mirrors, 1 and 2, with respect to the camera frame [C], which acts as the common origin of the coordinate system. Therefore,

$$z_{0_1} = \frac{c_1}{2} \quad (9)$$

$$z_{0_2} = d - \frac{c_2}{2} \quad (10)$$

By expanding Equation (5) with their respective index terms, it becomes

$$\left(z_1 - \frac{c_1}{2}\right)^2 - r_1^2 \left(\frac{k_1}{2} - 1\right) = \frac{c_1^2}{4} \left(\frac{k_1 - 2}{k_1}\right) \quad (11)$$

$$\left(z_2 - d + \frac{c_2}{2}\right)^2 - r_2^2 \left(\frac{k_2}{2} - 1\right) = \frac{c_2^2}{4} \left(\frac{k_2 - 2}{k_2}\right) \quad (12)$$

Additionally, we define the function  $f_{z_i} : r \mapsto z_i$  to find the corresponding  $z_i$  component from a given  $r$  value as

$$f_{z_i}(r) := \begin{cases} z_{0_i} + \gamma_i & \text{if } i = 1 \wedge \text{Equation (6)} \\ z_{0_i} - \gamma_i & \text{if } i = 2 \wedge \text{Equation (6)} \\ \text{None} & \text{otherwise} \end{cases} \quad (13)$$

where  $\gamma_i = \frac{a_i}{b_i} \sqrt{b_i^2 + r_i^2}$ .

The inverse relation  $f_{r_i} : z \mapsto \{+r_i, -r_i\}$  can be also implemented as

$$f_{r_i}(z) := \begin{cases} \pm b_i \Gamma_i & \text{if } i \in \{1, 2\} \wedge \text{Equation (6)} \\ \text{None} & \text{otherwise} \end{cases} \quad (14)$$

where  $\Gamma_i = \sqrt{\frac{(z - z_{0_i})^2}{a_i^2} - 1}$ , so a valid input  $z$  can be associated with both positive and negative solutions  $r_i$ .

### 2.3. Rig Size

In the attempt to evaluate the overall system size, we consider the height and weight variables due to the primary design parameters,  $\theta$ .

First, the height of the system,  $h_{sys}$  can be estimated from the functional relationships  $f_{z_1}$  and  $f_{z_2}$  defined in Equation (13), which can provide the respective  $z$ -component values at the out-most point on the mirror's surface. More specifically, knowing  $r_{sys}$ , we get

$$h_{sys} = z_{max} - z_{min} \quad (15)$$

where  $z_{max} = f_{z_1}(r_{sys})$  and  $z_{min} = f_{z_2}(r_{sys})$ .

The rig's weight can be indicated by the total resulting mass of the main "tangible" components:

$$m_{sys} = m_{cam} + m_{tub} + m_{mir} \quad (16)$$

where the mass of the camera-lens combination is  $m_{cam}$ ; the mass of the support tube  $m_{tub}$  can be estimated from its cylindrical volume  $V_{tub}$  and material density  $\rho_{tub}$ , and the mass due to the mirrors

$$\begin{aligned} m_{mir} &= V_{mir} \rho_{mir} \\ &= (V_1 + V_{ref} + V_2) \rho_{mir} \end{aligned} \quad (17)$$

For computing the volume of the hyperboloidal shell,  $V_i$  for mirror  $i$ , we apply a "ring method" of volume integration. By assuming all mirror material has the same wall thickness  $\tau_m$ , we acquire  $V_i$  by integrating the horizontal cross-sections area along the Z-axis. Each ring area depends on its outer and inner circumferences that vary according to radius  $r|_z$  for a given height  $z$ . Equation (14)

establishes the functional relation  $r_i^+ = f_{r_i}(z)$ , from which we only need its positive answer. We let  $A$  be the function that computes the ring area of constant thickness  $\tau_m$  for a variable outer radius  $r_i$

$$\begin{aligned}
 A(r_i) &= \pi r_i^2 - \pi (r_i - \tau_m)^2 \\
 &= \pi \tau_m (2r_i - \tau_m)
 \end{aligned}
 \tag{18}$$

We consider the definite integral evaluated in the  $z$  interval bounded by its height limits, which are correlated with its radial limits Equation (6) and can be obtained via the  $f_{z_i}$  defined in Equation (13), such that

$$z_{i,min} = f_{z_i}(r_{i,min}) \quad \text{and} \quad z_{i,max} = f_{z_i}(r_{i,max})
 \tag{19}$$

Then, we proceed to integrate Equation (18), so the shell volume for each hyperboloidal mirror is defined as

$$V_i = \int_{z_{i,min}}^{z_{i,max}} A(r_i) dz
 \tag{20}$$

Finally, since the reflex mirror piece is just a solid cylinder of thickness  $\tau_m$ , its volume is simply

$$V_{ref} = \tau_m \pi r_{ref}^2
 \tag{21}$$

### 3. Projective Geometry

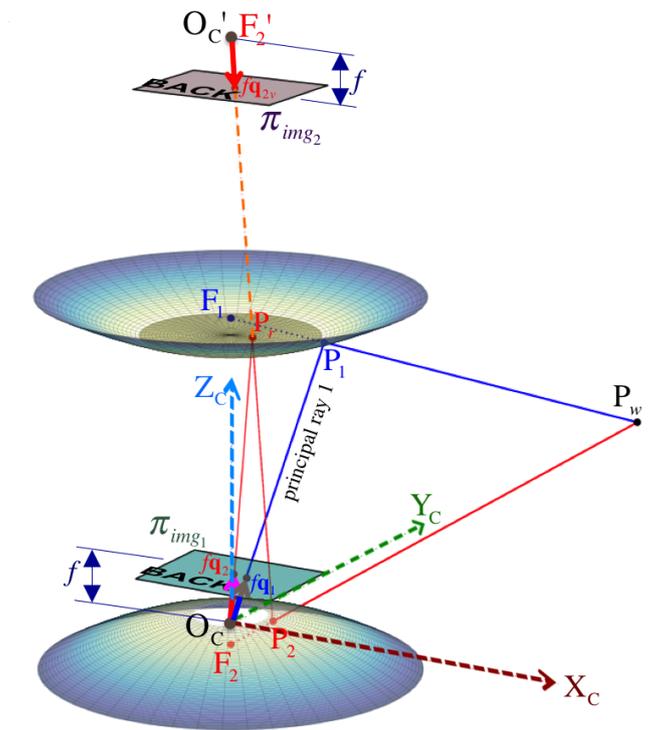


Figure 4. Omnistereo projection of a 3D point  $P_w$  to obtain image points  ${}^{[l]}m_1$  and  ${}^{[l]}m_2$ .

### 3.1. Analytical Solutions to Projection (Forward)

Assuming a central catadioptric configuration for the mirrors and camera system (Section 2.2), we derive the closed-form solution to the imaging process (forward projection) for an observable point  $P_w$ , positioned in three-dimensional Euclidean space,  $\mathbb{R}^3$ , with respect to the reference frame,  $[C]$ , as vector  ${}^{[C]}\mathbf{p}_w = [x_w, y_w, z_w]^T$ . In addition, we assume all reference frames such as  $[F_1]$  and  $[F_2]$  have the same orientation as  $[C]$ .

For mathematical stability, we must constrain that all projecting world points lie outside the mirror's volume:

$$f_{r_i}(z_w) < \rho_w, \text{ where } \rho_w = \sqrt{x_w^2 + y_w^2} \quad (22)$$

where  $f_{r_i}$  is defined by Equation (14) and  $\rho_w$  measures the horizontal range to  $P_w$ .

$P_w$  is imaged at pixel position  ${}^{[l]}\mathbf{m}_1$  after its reflection as point  $P_1$  on the hyperboloidal surface of mirror 1 (Figure 4). On the other hand, the second image point's position,  ${}^{[C]}\mathbf{m}_2$ , due to reflection point  $P_2$  on mirror 2 is rather obtained indirectly after an additional point  $P_r$  is reflected at  ${}^{[C]}\mathbf{p}_{ref}$  on the reflex mirror represented via Equation (32).

First, for  $P_w$ 's reflection point via mirror 1 at position vector  ${}^{[C]}\mathbf{p}_1$ , we use  $\lambda_1$  as the parametrization term for the line equation passing through  $F_1$  toward  $P_w$  with direction  ${}^{[F_1]}\mathbf{d}_1 = {}^{[C]}\mathbf{p}_w - {}^{[C]}\mathbf{f}_1$ . The position of any point  $P_1$  on this line is given by:

$${}^{[C]}\mathbf{p}_1 = {}^{[C]}\mathbf{f}_1 + \lambda_1 {}^{[F_1]}\mathbf{d}_1 \quad (23)$$

Substituting Equation (23) into Equation (11), we obtain:

$$\begin{aligned} \left( \lambda_1(z_w - c_1) + \frac{c_1}{2} \right)^2 - \left( \lambda_1^2 x_w^2 + \lambda_1^2 y_w^2 \right) \left( \frac{k_1}{2} - 1 \right) \\ - \frac{c_1^2}{4} \left( \frac{k_1 - 2}{k_1} \right) = 0 \end{aligned}$$

in order to solve for  $\lambda_1$ , which turns out to be

$$\lambda_1 = \frac{c_1}{\|{}^{[F_1]}\mathbf{d}_1\| \sqrt{k_1 \cdot (k_1 - 2)} - k_1(z_w - c_1)} \quad (24)$$

where  $\|{}^{[F_1]}\mathbf{d}_1\| = \sqrt{x_w^2 + y_w^2 + (z_w - c_1)^2}$  is the Euclidean norm between  $P_w$  and mirror 1's focus,  $F_1$ .

In practice, we represent the reflection point's position  ${}^{[C]}\mathbf{p}_1$  as a matrix-vector multiplication between the  $3 \times 4$  transformation matrix  $\mathbf{K}_1 = [\lambda_1 \mathbf{I}_{(3)}, (1 - \lambda_1) {}^{[C]}\mathbf{f}_1]$  and the point's position vector  ${}^{[C]}\mathbf{p}_{w,h} = [x_w, y_w, z_w, 1]^T$  in homogeneous coordinates:

$${}^{[C]}\mathbf{p}_1 = \mathbf{K}_1 {}^{[C]}\mathbf{p}_{w,h} \quad (25)$$

Note that  ${}^{[C]}\mathbf{p}_1$ 's elevation angle,  $\theta_1$ , must be bounded as

$$\theta_{1,min} \leq \theta_1 \leq \theta_{1,max} \quad (26)$$

where  $\theta_{1,min}$  and  $\theta_{1,max}$  are the angular elevation limits for the real reflective area of the hyperboloid.

Finally, the reflection point  $P_1$  with position  ${}^{[C]}\mathbf{p}_1$  can now be perspectively projected as a pixel point located at  ${}^{[I]}\mathbf{m}_1 = [u_1, v_1]^T$  on the image. In fact, the entire imaging process of  $P_w$  via mirror 1 can be expressed in homogeneous coordinates as:

$${}^{[I]}\mathbf{m}_{1,h} = \zeta_1 \mathbf{K}_c \mathbf{K}_1 {}^{[C]}\mathbf{p}_{w,h} \quad (27)$$

where the scalar  $\zeta_1 = 1/z_1 = 1/(c_1 + \lambda_1(z_w - c_1))$  is the perspective normalizer that maps the *principal ray* passing through  $\mathbf{p}_1$  onto a point  ${}^{[C]}\mathbf{q}_1 = [x_{q_1}, y_{q_1}, 1]^T$  on the normalized projection plane  $\hat{\pi}_{img_1}$ . The traditional  $3 \times 3$  intrinsic matrix of the camera's pinhole model is

$$\mathbf{K}_c = \begin{bmatrix} f_u & s & u_c \\ 0 & f_v & v_c \\ 0 & 0 & 1 \end{bmatrix} \quad (28)$$

in which  $f_u = f/h_x$  and  $f_v = f/h_y$  are based on the focal length  $f$  and the pixel dimension  $(h_x, h_y)$ ,  $s$  is the skew parameter, and  ${}^{[I]}\mathbf{m}_c = [u_c, v_c]^T$  is the optical center position on the image  $[I]$ . Figure 4 illustrates the projection point  $f{}^{[C]}\mathbf{q}_1$  on the respective image plane  $\pi_{img_1}$ .

Similarly, we provide the analytical solution for the forward projection of  $P_w$  via mirror 2 by first considering the position of reflection point  $P_2$ :

$${}^{[C]}\mathbf{p}_2 = \mathbf{K}_2 {}^{[C]}\mathbf{p}_{w,h} \quad (29)$$

where  $\mathbf{K}_2 = [\lambda_2 \mathbf{I}_{(3)}, (1 - \lambda_2) {}^{[C]}\mathbf{f}_2]$  is similar to the transformation matrix  $\mathbf{K}_1$ , but obviously it now uses  ${}^{[C]}\mathbf{f}_2$  and

$$\lambda_2 = \frac{c_2}{\|{}^{[E_2]}\mathbf{d}_2\| \sqrt{k_2 \cdot (k_2 - 2) + k_2(z_w - (d - c_2))}} \quad (30)$$

with direction vector's norm

$$\|{}^{[E_2]}\mathbf{d}_2\| = \|{}^{[C]}\mathbf{p}_w - {}^{[C]}\mathbf{f}_2\| = \sqrt{x_w^2 + y_w^2 + (z_w - (d - c_2))^2} \quad (31)$$

For completeness, note that the physical projection via mirror 2 is incident to the reflex mirror at

$${}^{[C]}\mathbf{p}_{ref} = {}^{[C]}\mathbf{f}_{2v} + \lambda_{ref} ({}^{[C]}\mathbf{p}_2 - {}^{[C]}\mathbf{f}_{2v}) \quad (32)$$

where  $\lambda_{ref} = \frac{d}{2(d - z_2)}$  according to Equation (2) in the theoretical model. Ultimately, ignoring any astigmatism and chromatic aberrations introduced by the reflex mirror, and because the same (and only) real camera with  $\mathbf{K}_c$  is used for imaging, we obtain the projected pixel position  ${}^{[I]}\mathbf{m}_{2,h} = [u_2, v_2, 1]^T$ :

$${}^{[I]}\mathbf{m}_{2,h} = \zeta_2 \mathbf{K}_c \mathbf{K}_{ref} \mathbf{K}_2 {}^{[C]}\mathbf{p}_{w,h} \quad (33)$$

where  $\zeta_2 = 1/(d - z_2)$  is the perspective normalizer to find  ${}^{[C]}\mathbf{q}_2$  on the normalized projection plane,  $\hat{\pi}_{img_2}$ .

Due to planar mirroring via the reflex mirror,  ${}^{[C]}\mathbf{K}_{ref}$  is used to change the coordinates of  $P_2$  from  $[C]$  onto the virtual camera frame,  $[C']$ , located at  ${}^{[C]}\mathbf{f}_{2v}$ . Hence,

$${}^{[C]}\mathbf{K}_{ref} = \begin{bmatrix} \mathbf{I}_{(3)} + 2\mathbf{D}\hat{\mathbf{n}}_{ref} & \\ & {}^{[C]}\mathbf{f}_{2v} \end{bmatrix} \quad (34)$$

where the  $3 \times 1$  unit normal vector of the reflex mirror plane,  ${}^{[C]}\hat{\mathbf{n}}_{ref}$  given in Equation (1), is mapped into its corresponding  $3 \times 3$  diagonal matrix  $\mathbf{D}_{\hat{\mathbf{n}}_{ref}}$ , via the relationship:

$$\mathbf{D}_{\hat{\mathbf{n}}_{ref}} \leftarrow \mathbf{I}_{(3)} \text{diag} \left( {}^{[C]}\hat{\mathbf{n}}_{ref} \right) \quad (35)$$

It is convenient to define the forward projection functions  $\mathbf{f}_{\varphi_1}({}^{[C]}\mathbf{p})$  and  $\mathbf{f}_{\varphi_2}({}^{[C]}\mathbf{p})$  for a 3D point  $\mathbf{P}$  whose position vector is known with respect to  $[C]$  and which is situated within the vertical field of view  $\alpha_i$  of mirror  $i$  (for  $i \in \{1, 2\}$ ) indicated in Figure 5. Function  $\mathbf{f}_{\varphi_i}({}^{[C]}\mathbf{p})$  maps  ${}^{[C]}\mathbf{p}$  to image point  ${}^{[I]}\mathbf{m}_i$  on frame  $[I]$ , such that  $\mathbf{f}_{\varphi_i}: \mathbb{R}^3 \mapsto \mathbb{R}^2$ ,

$$\mathbf{f}_{\varphi_i}({}^{[C]}\mathbf{p}) := \begin{cases} {}^{[C]}\mathbf{p} \xrightarrow{\text{Equation (27)}} {}^{[I]}\mathbf{m}_1 & \text{if } i = 1 \wedge \text{Equations (37) and (22)} \\ {}^{[C]}\mathbf{p} \xrightarrow{\text{Equation (33)}} {}^{[I]}\mathbf{m}_2 & \text{if } i = 2 \wedge \text{Equations (37) and (22)} \\ \text{None} & \text{otherwise} \end{cases} \quad (36)$$

In fact,  ${}^{[I]}\mathbf{m}_i$  is considered valid if it is located within the imaged radial bounds, such that:

$${}^{[I]}\|\mathbf{m}_{r_{i,min}}\| \leq {}^{[I]}\|\mathbf{m}_i\| \leq {}^{[I]}\|\mathbf{m}_{r_{i,max}}\| \quad (37)$$

where the frame of reference  $[I_i]$  implies that its origin is the image center  ${}^{[I_i]}\mathbf{m}_c = [u_{c_i}, v_{c_i}]^T$  of the  $[I_i]$  masked image (Figure 7). Therefore, the magnitude (norm) of any position  ${}^{[I_i]}\mathbf{m}$  in pixel space  $[I_i]$  can be measured as

$${}^{[I_i]}\|\mathbf{m}\| := \|\mathbf{m} - {}^{[I_i]}\mathbf{m}_c\| = \sqrt{(u - u_c)^2 + (v - v_c)^2} \quad (38)$$

In particular,  ${}^{[I_i]}\|\mathbf{m}_{r_{i,lim}}\|$  is the image radius obtained from the projection  ${}^{[I_i]}\mathbf{m}_{r_{i,lim}} \leftarrow \mathbf{f}_{\varphi_i}({}^{[C]}\mathbf{p}_{i,lim})$  corresponding to a particular point coincident with the line of sight of the radial limit  $r_{i,lim}$ —it being either  $r_{sys}$ ,  $r_{ref}$ , or  $r_{cam}$  as indicated by Equation (6).

### 3.2. Analytical Solutions to Back Projection

The back projection procedure establishes the relationship between the 2D position of a pixel point  ${}^{[I]}\mathbf{m}_i = [u, v]^T$  on the image  $[I_i]$  and its corresponding 3D projective direction vector  $\mathbf{v}_i$  toward the observed point  $\mathbf{P}_w$  in the world.

Initially, the pixel point  ${}^{[I]}\mathbf{m}_1$  (imaged via mirror 1) is mapped as  $\mathbf{Q}_1$  onto the normalized projection plane  $\hat{\pi}_{img_1}$  with coordinates  ${}^{[C]}\mathbf{q}_1 = [x_{q_1}, y_{q_1}, 1]^T$  by applying the inverse transformation of the camera intrinsic matrix Equation (28) as follows:

$${}^{[C]}\mathbf{q}_1 = {}^{[C]}_{[I]} \mathbf{K}_c^{-1} {}^{[I]}\mathbf{m}_{1,h} = \begin{bmatrix} \frac{1}{f_u} & -\frac{s}{f_u f_v} & \frac{sv_c - f_v u_c}{f_u f_v} \\ 0 & \frac{1}{f_v} & -\frac{v_c}{f_v} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} \quad (39)$$

For simplicity, we assume no distortion parameters exist, so we can proceed with the lifting step along the principal ray that passes through three points: the camera's pinhole  $\mathbf{O}_C$ , point  $\mathbf{Q}_1$  on the projection plane, and the reflection point  $\mathbf{P}_1$  (Figure 4). The vector form of this line equation can be written as:

$${}^{[C]}\mathbf{p}_1 = {}^{[C]}\mathbf{o}_c + t_1 ({}^{[C]}\mathbf{q}_1 - {}^{[C]}\mathbf{o}_c) = t_1 {}^{[C]}\mathbf{q}_1 \quad (40)$$

By substituting Equation (40) into Equation (11), we solve for the parameter  $t_1$ , to get

$$t_1 = \frac{c_1}{k_1 - \|[C] \mathbf{q}_1\| \sqrt{k_1 \cdot (k_1 - 2)}} \quad (41)$$

where  $\|[C] \mathbf{q}_1\| = \sqrt{x_{q_1}^2 + y_{q_1}^2 + 1}$  is the distance between  $Q_1$  and  $O_C$ .

Given  $^{[F_1]} \mathbf{v}_1$  as the direction vector leaving focal point  $F_1$  toward the world point  $^{[C]} P_w$ . Through frame transformation  $^{[F_1]} T_1^{[C]} \mathbf{p}_{1,h}$ , we get

$$^{[F_1]} \mathbf{v}_1 = \begin{bmatrix} ^{[F_1]} T_1^{[C]} \mathbf{p}_{1,h} \\ 1 \end{bmatrix}, \text{ where } ^{[F_1]} T_1^{[C]} = \begin{bmatrix} \mathbf{I}_{(3)}, & -[C] \mathbf{f}_1 \\ 0 & 1 \end{bmatrix} \quad (42)$$

for  $^{[C]} \mathbf{p}_{1,h}$  as the homogeneous form of Equation (40). In fact,  $^{[F_1]} \mathbf{v}_1$  provides the back-projected angles (elevation  $\theta_1$ , azimuth  $\psi_1$ ) from focus  $F_1$  toward  $^{[C]} P_w$ :

$$^{[F_1]} \theta_1 = \arcsin \left( \frac{z_{\mathbf{v}_1}}{\|^{[F_1]} \mathbf{v}_1\|} \right) = \arcsin \left( \frac{z_1 - c_1}{\|^{[F_1]} \mathbf{v}_1\|} \right) \quad (43)$$

$$^{[F_1]} \psi_1 = \arctan \left( \frac{y_{\mathbf{v}_1}}{x_{\mathbf{v}_1}} \right) = \arctan \left( \frac{y_1}{x_1} \right) \quad (44)$$

where  $\|^{[F_1]} \mathbf{v}_1\|$  is the norm of the back-projection vector up to the mirror surface.

Using the same approach, we lift a pixel point  $^{[I]} m_2$  imaged via mirror 2. Because the virtual camera  $O_C'$  located at  $^{[C]} \mathbf{f}_2 = [0, 0, d - c_2]^T$  uses the same intrinsic matrix  $\mathbf{K}_c$ , we can safely back-project pixel  $^{[I]} m_2$  to  $Q_{2v}$  on the normalized projection plane  $\hat{\pi}_{img_2}$  as follows:

$$^{[C]} \mathbf{q}_{2v} = [C] \mathbf{q}_2 = \mathbf{K}_c^{-1} [I] m_{2,h} \quad (45)$$

where the inverse transformation of the camera intrinsic matrix  $\mathbf{K}_c^{-1}$  is given by Equation (28). Since the reflection matrix  $\mathbf{K}_{ref}$  defined in Equation (34) is bidirectional due to the symmetric position of the reflex mirror about  $[C]$  and  $[C']$ , we can find the desired position of  $^{[C]} \mathbf{q}_{2v}$  with respect to  $[C]$ :

$$^{[C]} \mathbf{q}_{2v} = \begin{bmatrix} [C] \\ [C'] \end{bmatrix} \mathbf{K}_{ref} \begin{bmatrix} [C'] \\ [C] \end{bmatrix} \mathbf{q}_{2v,h} \quad (46)$$

which is equivalent to  $^{[C]} \mathbf{q}_{2v} = [x_{q_{2v}}, y_{q_{2v}}, d - 1]^T$ .

In Figure 4, we can see the principal ray that passes through the virtual camera's pinhole  $O_C'$  and the reflection point  $P_2$ , so this line equation can be written as:

$$^{[C]} \mathbf{p}_2 = [C] \mathbf{f}_{2v} + t_2 \left( ^{[C]} \mathbf{q}_{2v} - [C] \mathbf{f}_{2v} \right) \quad (47)$$

Solving for  $t_2$  from Equations (47) and (12), we get

$$t_2 = \frac{c_2}{k_2 - \|[C] \mathbf{q}_2\| \sqrt{k_2 \cdot (k_2 - 2)}} \quad (48)$$

where  $\|[C] \mathbf{q}_2\| = \sqrt{x_{q_2}^2 + y_{q_2}^2 + 1}$  is the distance between the normalized projection point  $Q_2$  and the camera  $O_C$  while considering Equation (46). Beware that the newly found location of  $P_2$  is given with respect to the real camera frame,  $[C]$ .

Again, we obtain the back-projection ray

$$^{[F_2]} \mathbf{v}_2 = \begin{bmatrix} ^{[F_2]} T_2^{[C]} \mathbf{p}_{2,h} \\ 1 \end{bmatrix}, \text{ where } ^{[F_2]} T_2^{[C]} = \begin{bmatrix} \mathbf{I}_{(3)}, & -[C] \mathbf{f}_2 \\ 0 & 1 \end{bmatrix} \quad (49)$$

in order to indicate the direction leaving from the primary focus  $F_2$  toward  $P_w$  through  $P_2$ . Here, the corresponding elevation and azimuth angles are respectively given by

$${}^{[F_2]}\theta_2 = \arcsin\left(\frac{z_{v_2}}{\|{}^{[F_2]}\mathbf{v}_2\|}\right) = \arcsin\left(\frac{d - t_2}{\|{}^{[F_2]}\mathbf{v}_2\|}\right) \quad (50)$$

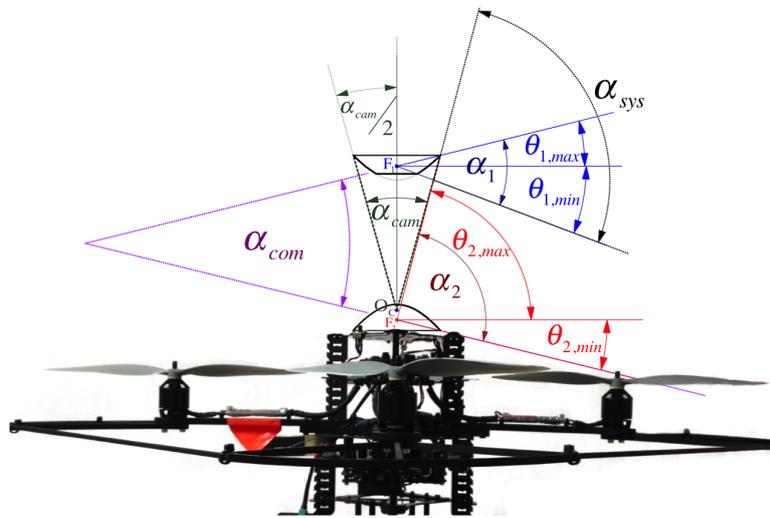
$${}^{[F_2]}\psi_2 = \arctan\left(\frac{y_{v_2}}{x_{v_2}}\right) = \arctan\left(\frac{y_2}{x_2}\right) \quad (51)$$

where  $\|{}^{[F_2]}\mathbf{v}_2\| = \sqrt{x_2^2 + y_2^2 + (c_2 - t_2)^2}$  is the magnitude of the direction vector from its reflection point  $P_2$ .

Like done for the (forward) projection, it is convenient to define the back-projection functions  $\mathbf{f}_{\beta_1}$  and  $\mathbf{f}_{\beta_2}$  for lifting a 2D pixel point  ${}^{[l]}\mathbf{m}$  within radial bounds validated by Equation (37) to their angular components  ${}^{[F_i]}(\theta_i, \psi_i)$  with respect to the respective foci frame  $[F_i]$  (oriented like  $[C]$ ) as indicated by Equations (43), (44), (50) and (51), such that  $\mathbf{f}_{\beta_i}: \mathbb{R}^2 \mapsto \mathbb{R}^2$ ,

$$\mathbf{f}_{\beta_i}({}^{[l]}\mathbf{m}) := \begin{cases} \left( \begin{array}{l} {}^{[l]}\mathbf{m} \xrightarrow{\text{Equation (43)}} {}^{[F_1]}\theta_1, {}^{[l]}\mathbf{m} \xrightarrow{\text{Equation (44)}} {}^{[F_1]}\psi_1 \\ {}^{[l]}\mathbf{m} \xrightarrow{\text{Equation (50)}} {}^{[F_2]}\theta_2, {}^{[l]}\mathbf{m} \xrightarrow{\text{Equation (51)}} {}^{[F_2]}\psi_2 \end{array} \right) & \text{if } i = 1 \\ \left( \begin{array}{l} {}^{[l]}\mathbf{m} \xrightarrow{\text{Equation (50)}} {}^{[F_2]}\theta_2, {}^{[l]}\mathbf{m} \xrightarrow{\text{Equation (51)}} {}^{[F_2]}\psi_2 \end{array} \right) & \text{if } i = 2 \\ \text{None} & \text{—Equation (37).} \end{cases} \quad (52)$$

### 3.3. Field-of-View



**Figure 5.** Vertical Field of View (vFOV) angles:  $\alpha_1$  and  $\alpha_2$  are the individual angles of the mirrors formed by their respective elevation limits  $\theta_{1/2,min/max}$ ;  $\alpha_{sys}$  is the overall vFOV angle of the system; and  $\alpha_{SROI}$  measures the overlapping region conceived between  $\alpha_1$  and  $\alpha_2$ .

The horizontal FOV is clearly  $360^\circ$  for both mirrors. In other words, azimuths  $\psi$  can be measured in the interval  $[0, 2\pi)$  rad. As discussed previously, there exists a positive correlation between the vertical field of view (vFOV) angle  $\alpha_i$  of mirror  $i$  and its profile parameter  $k_i$ , such that  $\alpha_i \rightarrow 180^\circ$  as  $k_i \rightarrow \infty$  (see Figure 9). As demonstrated in Figure 5,  $\alpha_i$  is physically bounded by its corresponding

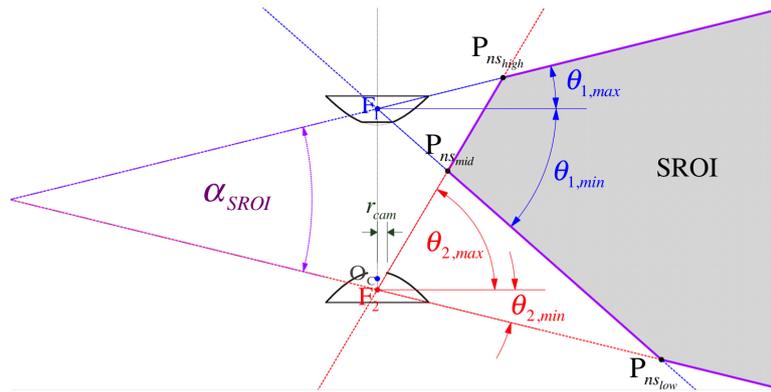
elevation angles:  $\theta_{i,max}$ ,  $\theta_{i,min}$ . Both vFOV angles,  $\alpha_1$  and  $\alpha_2$ , are computed from their elevation limits as follows:

$$\alpha_1 = \theta_{1,max} - \theta_{1,min} \quad (53a)$$

$$\alpha_2 = \theta_{2,max} - \theta_{2,min} \quad (53b)$$

The overall vFOV of the system is also given from these elevation limits:

$$\alpha_{sys} = \max(\theta_{1,max}, \theta_{2,max}) - \min(\theta_{1,min}, \theta_{2,min}) \quad (54)$$



**Figure 6.** A cross section of the SROI (shaded area) formed by the intersection of view rays for the limiting elevations  $\theta_{1/2,min/max}$ . The nearest stereo (ns) points are labeled  $P_{ns_{high}}$ ,  $P_{ns_{mid}}$  and  $P_{ns_{low}}$  since they are the vertices of the hull that near-bounds the set of usable points for depth computation from triangulation (Section 5.2). See Table 3 for the proposed sensor's values.

Figure 6 highlights the the so-called common vFOV angle,  $\alpha_{SROI}$ , for the stereo region of interest (SROI) where the same point can be seen from both mirrors so point correspondences can be found (Section 5). In our model,  $\alpha_{SROI}$  can be decided from the value of the three prevailing elevation angles ( $\theta_{1,max}$ ,  $\theta_{1,min}$ , and  $\theta_{2,min}$ ), such that:

$$\alpha_{SROI} = \theta_{SROI,max} - \theta_{SROI,min} \quad (55)$$

where generally,

$$\theta_{SROI,min} = \max(\theta_{1,min}, \theta_{2,min}) \quad (56a)$$

$$\theta_{SROI,max} = \min(\theta_{1,max}, \theta_{2,max}) \quad (56b)$$

The shaded area in Figure 6 illustrates the SROI that is far-bounded by the set of triangulated points found at the maximum range due to minimum disparity  $\Delta m_{12} = 1$  px in the discrete case (refer to Figure 17), such that

$$P_{fs} = \{ P_w \leftarrow \mathbf{f}_\Delta((\theta_1, \psi_1), (\theta_2, \psi_2)) \mid (\theta_1, \psi_1) \leftarrow \mathbf{f}_{\beta_1}(\mathbf{m}_1) \wedge (\theta_2, \psi_2) \leftarrow \mathbf{f}_{\beta_2}(\mathbf{m}_2) \wedge \Delta m_{12} = 1, \text{px} \} \quad (57)$$

where functions  $\mathbf{f}_{\beta_i}$  and  $\mathbf{f}_\Delta$ , are provided in Equations (52) and (89).

The SROI is near-bounded (to the Z-axis of radial symmetry) by its vertices  $P_{ns_{high}}$ ,  $P_{ns_{mid}}$  and  $P_{ns_{low}}$ , which result from the following ray-intersection cases:

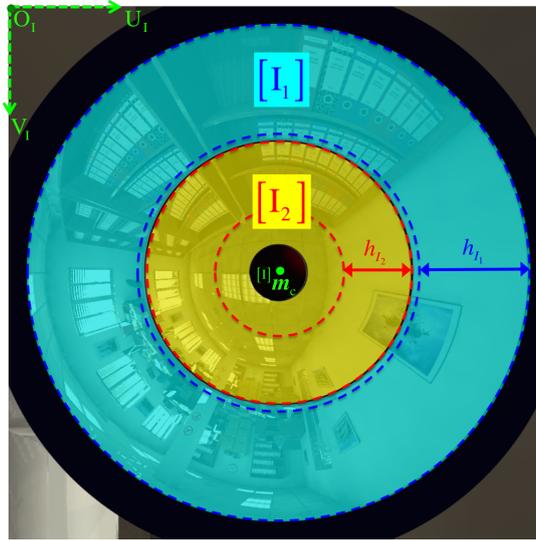
- (a)  $P_{ns_{high}} \leftarrow \mathbf{f}_{\Delta}((\theta_{1,max}, \psi_1), (\theta_{2,max}, \psi_2))$   
 (b)  $P_{ns_{mid}} \leftarrow \mathbf{f}_{\Delta}((\theta_{1,min}, \psi_1), (\theta_{2,max}, \psi_2))$   
 (c)  $P_{ns_{low}} \leftarrow \mathbf{f}_{\Delta}((\theta_{1,min}, \psi_1), (\theta_{2,min}, \psi_2))$

where the intersection function  $\mathbf{f}_{\Delta}$  is implemented for direction rays (or angles) as defined in the Triangulation Section 5.2.

By assuming a radial symmetry on the camera's field of view  $\alpha_{cam}$ , it should allow for a complete view of the mirror surface at its outmost diameter of  $2r_{sys}$  according to Equation (6). Substantially, as depicted in Figure 6,  $\alpha_{cam}$  is upper-bounded by the camera hole radius  $r_{cam}$  selected according to Equation (78). The following inequality constraint emerges

$$2 \arctan\left(\frac{r_{sys}}{f_{z_1}(r_{sys})}\right) \leq \alpha_{cam} \leq 2 \arctan\left(\frac{r_{cam}}{f_{z_2}(r_{cam})}\right) \quad (58)$$

where the respective functions  $f_{z_i}$  are defined in Equation (13).

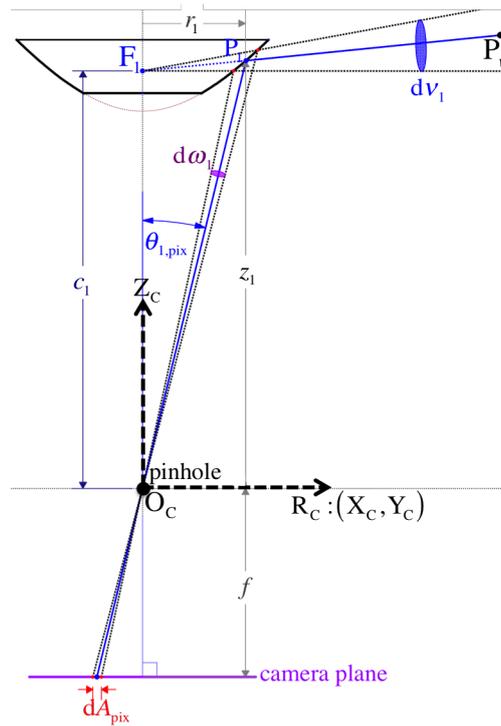


**Figure 7.** The omnidirectional image  $[I]$  shown in Figure 2b is now annotated for the separate regions of interest in  $[I_1]$  and  $[I_2]$ . In addition, we indicate the corresponding radial heights  $h_{I_1}$  and  $h_{I_2}$  of the SROI, so we can determine the imaging ratio  $\chi_{I_{1,2}} = \frac{h_{I_1}}{h_{I_2}}$ . For the optimal parameter values listed in Table 1, we find that  $\chi_{I_{1,2}} \approx 2$ .

Our specific viewing requirements when mounting the omnidirectional sensor along the central axis of the quadrotor ensure that objects located at 15 cm under the rig's base and at 1 meter away (from the central axis) can be viewed. Thus, angles  $\theta_{1,min}$  and  $\theta_{2,min}$  should only be large enough as to avoid occlusions from the MAV's propellers (Figure 5) and to produce inner and outer ring images at a useful ratio (Figure 7).

### 3.4. Spatial Resolution

The resolution of the images acquired by our system are not space invariant. In fact, an omnidirectional camera producing spatial resolution-invariant images can only be obtained through a non-analytical function of the mirror profile as shown in [31]. In this section, we study the effect our design has on its spatial resolution as it depends on position parameters like  $d$  and  $c_i$  introduced in Section 2.1 as well as a direct dependency on the characteristics (e.g., focal length  $f$ ) of the camera obtaining the image.



**Figure 8.** The spatial resolution for a central catadioptric sensor is the ratio between an infinitesimal image area  $dA$  and its corresponding solid angle  $dv_1$  that views a point  $P_w$ . (Note: infinitesimal elements are exaggerated in the figure for better visualization.)

Let  $\eta_{cam}$  be the spatial resolution for a conventional perspective camera as defined by Baker and Nayar in [25,29]. It measures the ratio between the infinitesimal solid angle  $d\omega_i$  (usually measured in steradians) that is directed toward a point  $P_i$  at an angle  $\theta_{i,pix}$  (formed with the optical axis  $Z_C$ ) and the infinitesimal element of image area  $dA_{pix}$  that  $d\omega_i$  subtends (as shown in Figure 8). Accordingly, we have:

$$\eta_{cam} = \frac{dA_{pix}}{d\omega_i} = \frac{f^2}{\cos^3 \theta_{i,pix}} \quad (59)$$

whose behavior tends to decrease as  $\theta_{pix} \rightarrow 0$ , so higher resolution areas on the sensor plane continuously increase the farther away they get from the optical center imaged at  ${}^{[1]}m_c$ . For ease of visualization, we plot only the  $u$  pixel coordinates corresponding to the 2D spatial resolution  $\eta_{2D}$ , which is obtained by projecting the solid angle  $\Omega$  onto a planar angle  $\theta_\Omega$  (the apex angle in 2D of the solid cone of view). This yields  $\theta_\Omega = 2 \arccos(1 - \Omega/2\pi)$ , and we reduce the image area into its circular diameter with  $2\sqrt{dA/\pi}$ . Generally, our conversion from 3D spatial resolution  $\eta$  in  $[m^2/sr]$  units to 2D proceeds as follows:

$$\eta_{2D} = \frac{2\sqrt{\eta/\pi}}{\theta_{\Omega=1 \text{ sr}}} \quad (60)$$

where  $\theta_{\Omega=1\text{sr}} \approx 1.14390752211$  rad. More specifically, Equation (59) is manipulated to provide  $\eta_{i,cam}$  as the indicative of spatial resolution toward any specific point in the mirror,  ${}^{[C]}P_i \in M_i$  according to Equation (8), as follows:

$$\eta_{i,cam} = \begin{cases} f^2 \left( \frac{\sqrt{r_1^2 + z_1^2}}{z_1} \right)^3 & \text{if } i = 1 \\ f^2 \left( \frac{\sqrt{r_2^2 + (d - z_2)^2}}{d - z_2} \right)^3 & \text{if } i = 2 \end{cases} \quad (61)$$

where  $r_i$  is the radial length defined in Equation (6) and its associated  $z_i$  coordinate,  $f$  is the camera's focal length, and the design parameters  $d$  and  $c_i$  that relate to the position of the mirror focal points  $F_i$  with respect to the camera frame [C].

Thus, for a conventional perspective camera,  $\eta_{i,cam}$  grows as  $\theta_{i,pix} \rightarrow \pi/2$  due to the foreshortening effect that stretches the image representation around the sensor plane's periphery where spatial information gets collected onto a larger number of pixels. Therefore, image areas farther from the optical axis are considered to have higher spatial resolutions.

Baker and Nayar also defined the resolution,  $\eta_i$ , of a catadioptric sensor in order to quantify the view of the world or  $dv_i$ , an infinitesimal element of the solid angle subtended by the mirror's effective viewpoint  $F_i$ , which is consequently imaged onto a pixel area  $dA_{pix}$ . Again, here we provide the resolution according to our model:

$$\eta_1 = \frac{dA_{pix}}{dv_1} = \left[ \frac{r_1^2 + (c_1 - z_1)^2}{r_1^2 + z_1^2} \right] \eta_{1,cam} \quad (62a)$$

$$\eta_2 = \frac{dA_{pix}}{dv_2} = \left[ \frac{r_2^2 + (c_2 - d + z_2)^2}{r_2^2 + (d - z_2)^2} \right] \eta_{2,cam} \quad (62b)$$

for our mirror-perspective camera configuration, where  $O_C$  is the origin of coordinates as shown in Figure 8 and  $\eta_{i,cam}$  is given in Equation (61).

As demonstrated by the plot of Figure 12 in Section 4.2.2,  $\eta_i$  grows accordingly towards the periphery of each mirror (the equatorial region). This aspect of our sensor design is very important because it indicates that the common field of view,  $\alpha_{SROI}$ , where stereo vision is employed (Section 5), is imaged at a relatively higher resolution than the unused polar regions closer to the optical axis (the  $Z_C$  axis).

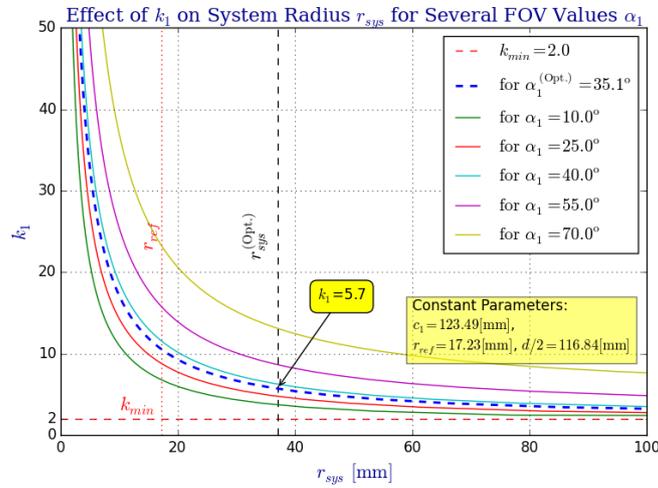
If we modify  $\eta_i$  by substituting  $r_i$  with its equivalent  $f_{r_i}(z_i)$  function defined in Equation (14), using mirror 1 for example, we get:

$$\begin{aligned} \eta_1 &= \left[ \frac{f_{r_1}^2(z_1) + (c_1 - z_1)^2}{f_{r_1}^2(z_1) + z_1^2} \right] \eta_{1,cam} \\ &= \frac{f^2 \sqrt{f_{r_1}^2(z_1) + z_1^2} \left[ f_{r_1}^2(z_1) + (c_1 - z_1)^2 \right]}{z_1^3} \end{aligned} \quad (63)$$

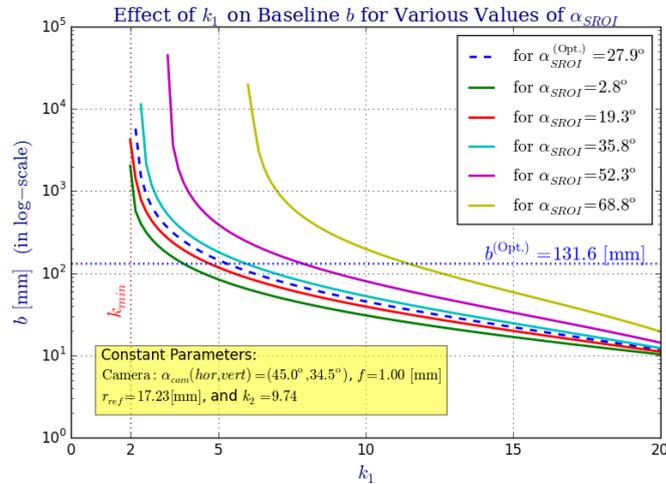
which is an inherent indicative of how the resolution  $\eta_i$  for a reflection point  $P_i$  increases with  $k_i \rightarrow \infty$  (Figure 11). Conversely, the smaller the  $k_i$  parameter gets (related to eccentricity as discussed in Section 2.2), the flatter the mirror becomes, so its resolution resembles more that of the perspective camera alone. Mathematically,  $\lim_{k_i \rightarrow 2} \eta_i \rightarrow \eta_{i,cam}$ .

As shown in Figure 9, a smaller  $k_i$  would require a wider radius  $r_{sys}$  in order to achieve the same omnidirectional vertical field of view,  $\alpha_{sys}$ . Even worse, in order to image such a wider reflector, either the camera's field of view,  $\alpha_{cam}$ , would have to increase (by decreasing the focal length  $f$  and perhaps requiring a larger camera hole  $r_{cam}$  and sensor size), or the distance  $c_i$  between the effective pinhole and the viewpoint would have to increase accordingly. Another consequence is the effect on the

baseline  $b$ , which must change in order to maintain the same vertical field of view (Figure 10). As a result, the depth resolution of the stereo system would suffer as well.



**Figure 9.** The effect that parameter  $k_i$  (showing mirror 1 only) has over the system radius  $r_{sys}$  for various values of the vertical field of view angle  $\alpha_i$ . In order to maintain a vertical field of view  $\alpha_i$  that is bounded by  $z_{max} |_{r_{sys}}$ , the value of  $r_{sys}$  must change accordingly. Inherently, the system’s height,  $h_{sys}$ , and its mass,  $m_{sys}$ , are also affected by  $k_i$  (see Section 2.3).



**Figure 10.** The effect that parameter  $k_1$  has over the omnistereo system’s baseline  $b$  for several common FOV angles ( $\alpha_{SROI}$ ) and a fixed camera with  $\alpha_{cam}$ . An inverse relationship exists between  $k$  and  $b$  as plotted here (using a logarithmic scale for the vertical axis). Intuitively, the flatter the mirror gets ( $k \rightarrow 2$ ), the farther  $F_1$  must be translated in order to fit within the camera’s view,  $\alpha_{SROI}$ , causing  $b$  to increase.

#### 4. Parameter Optimization and Prototyping

The nonlinear nature of this system makes it very difficult to balance among its desirable performance aspects. The optimal vector of design parameters,  $\theta^*$ , can be found by posing a constrained maximization problem for the objective function

$$f_b(\theta) = c_1 + c_2 - d \tag{64}$$

which measures the baseline according to Equation (3). Indeed, the optimization problem is subject to the set of constraints  $C$ , which we enumerate in Section 4.1. Formally,

$$\boldsymbol{\theta}^* = \arg \max_{\boldsymbol{\theta} \in \Theta} f_b(\boldsymbol{\theta}) \text{ subject to } C \quad (65)$$

where  $\Theta \subseteq \mathbb{R}^6$  is the 6-dimensional solution space for  $\boldsymbol{\theta} \in \mathbb{R}^6$  given in Equation (4) as  $\boldsymbol{\theta} = [c_1, c_2, k_1, k_2, d, r_{sys}]$ .

#### 4.1. Optimization Constraints

We discuss the constraints that the proposed omnistereo sensor is subject to. Overall, we mainly take the following into account:

- (a) *geometrical constraints* — including SVP and reflex constraints described by Equations (11), (12) and (2);
- (b) *physical constraints* — the rig's dimensions, which include the mirrors radii as well as by-product parameters such as system height  $h_{sys}$  and mass  $m_{sys}$ ;
- (c) *performance constraints* — the spatial resolution and range from triangulation determined by parameters  $k_1, k_2$ , and  $c_1$ ; the desired viewing angles for an optimal SROI field of view,  $\alpha_{SROI}$ .

Following the design model described throughout Section 2, we now list the pertaining linear and nonlinear constraints that compose the set  $C$ . We disjoint the linear constraints in a subset  $C_L$  and the non-linear constraints subset  $C_{NL}$ , so  $C = C_L \uplus C_{NL}$ . Within each subset, we generalize equality constraints as functions  $h: \mathbb{R}^6 \mapsto \mathbb{R}$  that obey

$$h(\boldsymbol{\theta}) = 0 \quad (66)$$

whereas inequality functions  $g: \mathbb{R}^6 \mapsto \mathbb{R}$  satisfy

$$g(\boldsymbol{\theta}) \leq 0 \quad (67)$$

##### 4.1.1. Linear Constraints

We have only setup linear inequalities for constraints in  $C_L$ . Specifically, we require the following:

- g<sub>1</sub>: In order to set the position of  $F_2$  below the origin  $O_C$  of the pinhole camera frame  $[C]$ , the focal distance  $c_2$  of mirror 2 must be larger than  $d$  (distance between  $O_C$  and  $F_{2v}$ ),

$$d \leq c_2 \quad (68)$$

- g<sub>2</sub>: Because the hyperboloidal mirror should reflect light towards its effective viewpoint  $F_1$  without being occluded by the reflex mirror, mirror 1's focal distance,  $c_1$ , needs to exceed the placement of the reflex mirror,

$$d/2 \leq c_1 \quad (69)$$

- g<sub>3</sub>: The empirical constraint

$$\frac{5}{3} \leq \frac{k_2}{k_1} \quad (70)$$

pertains our rig dimensions in order to assign a greater curvature to mirror 2's profile (located at the bottom), so its view is directed toward the equatorial region rather than up. Complementarily, this constraint flattens mirror 1's profile, so it can possess a greater view of the ground. This

curvature inequality allows the SROI to be bounded by a wider vertical field of view when the sensor must be mounted above the MAV's propellers as depicted in Figure 5.

#### 4.1.2. Non-Linear Constraints

For the non-linear design constraints, we establish the following inequalities:

- g4: The AscTec Pelican quadrotor has a maximum payload of 650 g (according to the manufacturer specifications [28]). Therefore, we must satisfy the system mass computed via Equation (16), such that

$$m_{sys} \leq 650 \quad (71)$$

- g5: Similarly, we limit the system's height obtained with Equation (15) by a height limit  $h_{sys,max}$ ,

$$h_{sys} \leq h_{sys,max} \quad (72)$$

For example, we set  $h_{sys,max} = 150$  mm for the 37 mm-radius rig.

- g6: The origin of coordinates for the camera frame is set at its viewpoint,  $O_C$ . In order to fit the camera enclosure under mirror 2, it is realistic to position the focus  $F_2$  on the vertical transverse axis at more than 5 mm away from  $O_C$ :

$$5 \leq z_{0_2} - a_2 \quad (73)$$

where  $z_{0_2}$  is defined in Equation (10), and  $a_2$  pertains to Equation (5).

Next, we determine the bounds for the limiting angles that partake in the computation of the system's vertical field of view  $\alpha_{sys}$ , which is based on equation Equation (54). Our application has specific viewing requirements that can be achieved with the following application conditions:

- g7: Let  $\Lambda_{1,max} = 14^\circ$  be an acceptable upper-bound for angle  $\theta_{1,max}$ , such that

$$\theta_{1,max} \leq \Lambda_{1,max} \quad (74)$$

- g8: Because we desire a larger view towards the ground from mirror 1, we empirically set  $\Lambda_{1,min} = -25^\circ$  as a lower-bound for the minimum elevation  $\theta_{1,min}$ ,

$$\Lambda_{1,min} \leq \theta_{1,min} \quad (75)$$

- g9: In order to avoid occlusions with the MAV's propellers while being capable to image objects located about 5 cm under the rig's base and 20 cm away (horizontally) from the central axis, we limit mirror 2's lowest angle by a lower-bound  $\Lambda_{2,min} = -14^\circ$ ,

$$\Lambda_{2,min} \leq \theta_{2,min} \quad (76)$$

Finally, we restrict the radius of the system,  $r_{sys}$ , to be identical for both hyperboloids by satisfying the following equality condition:

- h1: With functions  $f_{r_1}$  and  $f_{r_2}$  defined in Equation (14), we set

$$r_{sys} = r_{i,max} = f_{r_i}(z_{i,max}), \forall i \in \{1, 2\}$$

where we imply that  $z_{i,max} \leftarrow f_{z_i}(r_{sys})$  using Equation (13). Thus, the entire function composition for this equality becomes

$$f_{r_1} \left( f_{z_1}(r_{sys}) \right) = f_{r_2} \left( f_{z_2}(r_{sys}) \right) \quad (77)$$

#### 4.2. Optimal Results

Applying the aforementioned constraints (Section 4.1) and using an iterative nonlinear optimization method such as one of the surveyed in [32], a bounded solution vector  $\theta^*$  converges to the values shown in Table 1 for two rig sizes. Table 2 contains the by-product parameters corresponding to the dimensions listed in Table 1.

**Table 1.** Optimal System Design Parameters.

Parameter	Big Rig	Small Rig
$b = \max f_b(\theta^*)$	131.61	108.92
$r_{sys}$ [mm]	37.0	28.0
$c_1$ [mm]	123.49	104.59
$c_2$ [mm]	241.80	204.34
$d$ [mm]	233.68	200.00
$k_1$	5.73	6.88
$k_2$	9.74	11.47

**Table 2.** By-product Length Parameters.

Parameter	Big Rig	Small Rig
$r_{ref}$ [mm]	17.23	11.74
$r_{cam}$ [mm]	7	7
$h_{sys}$ [mm]	150.00	120.00

As Figure 3 illustrates, a realistic dimension for the radius of the camera hole,  $r_{cam}$ , must consider the maximum value between a physical micro-lens radius ( $r_{lens}$ ) and the radius  $r_{\alpha_{cam}|r_{sys}}$  for an unoccluded field of view of the camera  $\alpha_{cam}$  imaging the complete surface of mirror 1. Practically,

$$r_{cam} = \max \left( r_{lens}, r_{\alpha_{cam}|r_{sys}} \right) \quad (78)$$

For both rigs, the expected vertical field of views are  $\alpha_{sys} = 75^\circ - (-21^\circ) \approx 96^\circ$  according to Equation (54), and  $\alpha_{SROI} = 14^\circ - (-14^\circ) \approx 28^\circ$  using Equation (55). Note that  $\theta_{2,max}$  may be actually limited by the camera hole radius, so in reality  $\theta_{cam} \rightsquigarrow 59^\circ$ , and  $\alpha_{sys} \rightsquigarrow 80^\circ$ . For the big rig, Table 3 shows the nearest vertices of the SROI that result from these angles (Figure 6).

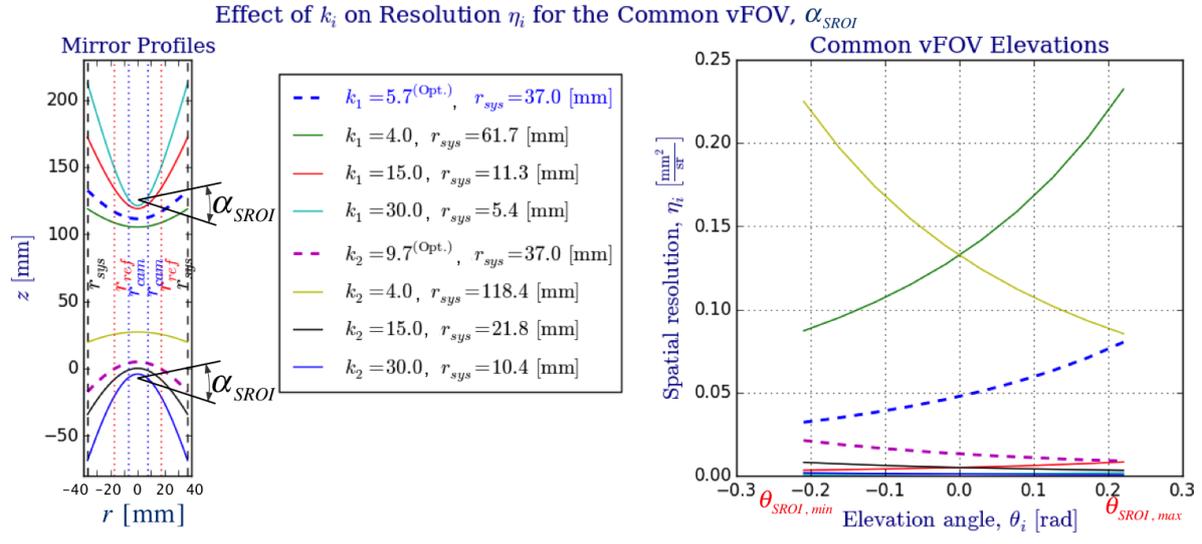
**Table 3.** Near Vertices of the SROI for the Big Rig.

Vertex	${}^{[c]}\rho_w$ [mm]	${}^{[c]}z_w$ [mm]
$Pn^{shigh}$	93.5	144.4
$Pn^{mid}$	65.2	98.4
$Pn^{slow}$	763.4	-170.3

##### 4.2.1. Optimality of Parameters $k_1$ and $k_2$

Finally, we study the effect parameter  $k_i$  has over the system radius  $r_{sys}$  (Figure 9), the omnistereo baseline  $b$  (Figure 10), and the spatial resolution (Figures 11 and 12). Figure 9 addresses the relation between  $k_i$  and radius  $r_{sys}$  (recall the rig size specified in Section 2.3). In Figure 11, it can be seen that

for the same  $r_{sys}$ , realistic values for  $k_1$  fall in the range  $3 < k_1 < 13$ , and the vertical field of view  $\alpha_1 \rightarrow 0$  as  $k \rightarrow 2$ , which is expected according to the SVP property specified in Section 2.2. In fact, the left part of Figure 11 also demonstrates the necessary  $r_{sys}$  to maintain  $\alpha_{SROI} \approx 28^\circ$  for various values of  $k_i$ .



**Figure 11.** Comparison of  $k_i$  values and their effect on spatial resolution  $\eta_i$  for  $i = \{1, 2\}$ . For the big rig, the optimal focal dimensions  $c_1$  and  $c_2$  (from Table 1) were used as well as the angular span on the common vertical FOV,  $\alpha_{SROI} \approx 28^\circ$ . Although resolution  $\eta_i^{(Opt.)}$  for the optimal values of  $k_i$  could be improved by employing smaller  $k$  values (lower curvature profiles indicated on the left plot of the figure), this would in turn increase the system radius,  $r_{sys}$ , as to maintain  $\alpha_i$  (Figure 9). As expected, the plot on the right help us appreciate how the spatial resolutions,  $\eta_i$ , increase towards the equatorial regions ( $\theta_1 \rightarrow \theta_{SROI,max}$  and  $\theta_2 \rightarrow \theta_{SROI,min}$ ).

Figure 10 shows the inverse relationship between values of  $k_1$  and the baseline,  $b$ , as we attempt to fit the view of a wider/narrower mirror profile (due to  $k_1$ ) on the constant camera field of view,  $\alpha_{cam}$ . In order to make a fair comparison, let

$$k'_1 = k_1 + \epsilon_k, \quad \forall k_1 > 2, \epsilon_k > 0$$

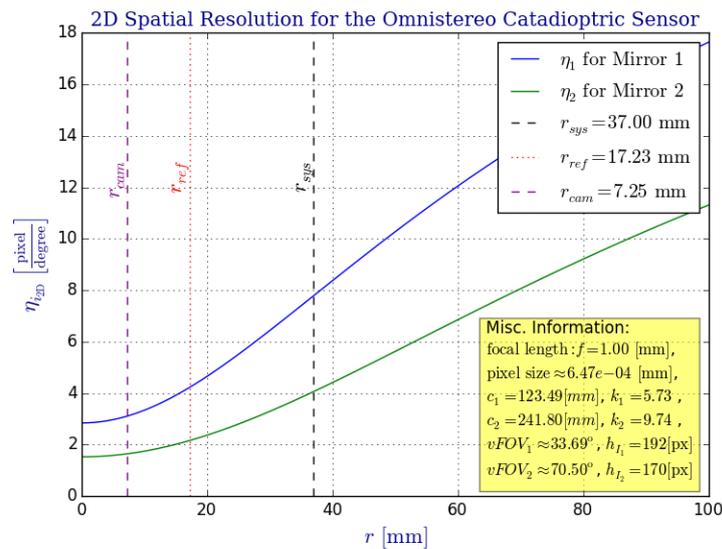
for which we find its new focal length  $c'_1$  while solving for the new  $r'_{sys}$  and  $z'_{max}$ . Provided with a function such that  $c_1 \leftarrow f_{c_1}(k_1)$ , we perform the analysis for a given  $\alpha_{SROI}$  and  $\alpha_{cam}$  shown in Figure 10. Given the baseline function  $f_b$  defined in Equation (64), the following implication holds true:

$$f_b |_{c_1 \leftarrow f_{c_1}(k_1)} > f_b |_{c'_1 \leftarrow f_{c_1}(k_1 + \epsilon_k)}, \quad \forall k_1 > 2, \epsilon_k > 0 \tag{79}$$

Notice that  $k_2$ ,  $c_2$  and  $d$  are kept constant through this last analysis, and we ignore possible occlusions from the reflex mirror fixed at  $d/2$ .

#### 4.2.2. Spatial Resolution Optimality

In this section, we compare the sensor’s spatial resolution,  $\eta_i$ , defined in Section 3.4 for the optimal parameters listed in Table 1 (for the big rig, only). In Figure 12, we verify how both resolutions  $\eta_1$  and  $\eta_2$  increase towards the equatorial region according to the spatial resolution theory presented in [29]. Indeed, the increase in spatial resolution within the SROI that covers the equatorial region (as indicated in Figure 6) justifies our model’s coaxial configuration intended for omnistereo applications.



**Figure 12.** Using the formula given in Equation (60), we plot the 2D version of the spatial resolution of our proposed omnistereo catadioptric sensor (37 mm-radius rig). Both resolutions  $\eta_1$  and  $\eta_2$  increase towards the equatorial region where they are physically limited by  $r_{sys}$ . This verifies the spatial resolution theory given in [29], and it justifies our coaxial configuration useful for omnistereo sensing within the SROI indicated in Figure 6.

In Figure 11, we compare the effect on  $\eta_i$  for various mirror profiles, which depend directly on  $k_i$ . We illustrate the change in curvature due to parameters  $k_1$  and  $k_2$  and also show (in the legend) the respective  $r_{sys}$  achieving a common vFOV of  $\alpha_{SROI} \approx 28^\circ$  as for the optimal parameters of the big rig. From this plot, we appreciate the compromise due to optimal parameters,  $k_1^{(Opt.)} = 5.7$  and  $k_2^{(Opt.)} = 9.7$ , for a realistic system size due to  $r_{sys}$  and a suitable range of spatial resolutions,  $\eta_i$ , within the SROI.

### 4.3. Prototypes

We validate our design with both synthetic and real-life models.

#### 4.3.1. Synthetic Prototype (Simulation)

After converging to an optimal solution  $\theta^*$ , we employ these parameters (Table 1) to describe synthetic models using POV-Ray, an open-source ray-tracer. We render 3D scenes via the camera of the synthetic omnistereo sensor like the example shown in Figure 2b. The simulation stage plays two important roles in our investigation:

- (1) to acquire ground-truth 3D-scene information in order to evaluate the computed range by the omnistereo system (as explained in Section 5); and
- (2) to provide an almost accurate geometrical representation of the model by discounting some real-life computer vision artifacts such as assembly misalignments, glare from the support tube (motivating the use of standoffs on the real prototype), as well as the camera's shallow depth-of-field. All of these artifacts can affect the quality of the real-life results shown in Section 6.

### 4.3.2. Real-Life Prototypes



(a) Omnistereo rig using 37 mm-radius mirrors mounted on an AscTec Pelican quadrotor.



(b) Omnidirectional image captured by the real-life prototype in Figure 13a.

**Figure 13.** Real-life prototype of the omnistereo sensor.

We have also produced two physical prototypes that can be installed on the Pelican quadrotor (made by Ascending Technologies [28]). Figure 13a shows the rig constructed with hyperboloidal mirrors of  $r_{sys} \approx 37$  mm, and a Logitech<sup>®</sup> HD Pro Webcam C910 camera capable of  $(2592 \times 1944)$  pixel images at 15~20 FPS. We decided to skip the use of the acrylic glass tube to separate the mirrors at the specified  $h_{sys}$  distance, and instead we constructed a lighter 3-standoff mount in order to avoid glare and cross-reflections. This support was designed in 3D-CAD and printed for assembly. The three areas of occlusion due to the 3 mm-wide standoffs are non-invasive for the purpose of omnidirectional sensing and can be ignored with simple masks during image processing. In fact, we stamped fiducial markers to the vertical standoffs to aid with the panoramas generation (Section 5.1) and future calibration methods. To image the entire surface of mirror 1, we require a camera with a (minimum) field of view of  $\alpha_{cam} > 31^\circ$ , which is achieved by  $r_{cam} > 1.4$  mm. In practice, as noted by Equation (78), microlenses measure around  $r_{lens} \approx 7$  mm. Therefore, we set  $r_{cam} > 7$  mm, as a safe specification to fit a standard microlens through the opening of mirror 2 as shown in Figure 3.

Recall that  $m_{sys}$  is limited by the maximum 650 g-payload that the AscTec Pelican quadrotor is capable of flying with (according to the manufacturer specifications [28]). The camera with lens weights approximately 25 g. A cylindrical tube made of acrylic has an average density  $\rho_{tub} \approx 1.18 \text{ g}\cdot\text{cm}^{-3}$ , whereas the mirrors machined out of brass have a density  $\rho_{mir} \approx 8.5 \text{ g}\cdot\text{cm}^{-3}$ . Empirically, we verify a close estimate of the entire system's mass, such that  $m_{sys} \approx 550$  g for the big rig, and  $m_{sys} \approx 150$  g for the small rig.

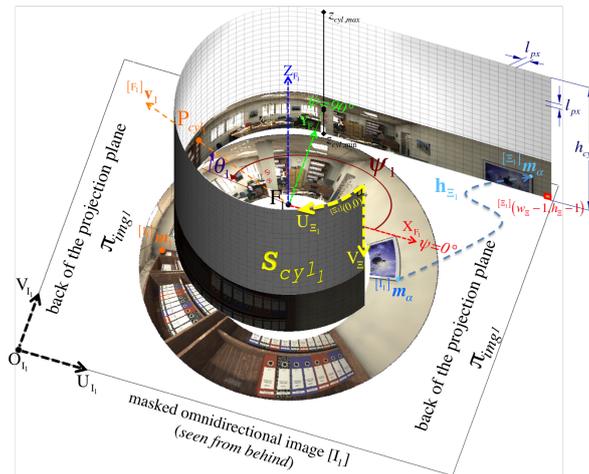
## 5. 3D Sensing from Omnistereo Images

Stereo vision from point correspondences on images at distinct locations is a popular method for obtaining 3D range information via triangulation. Techniques for image point matching are generally divided between dense (area-based scanning [32]) and sparse (feature description [33]) approaches. Due to parallax, the disparity in point positions for objects close to the vision system must be larger than for objects that are farther away. As illustrated in Figure 6, the nearsightedness of the sensor is determined mainly by the common observable space (a.k.a. SROI) acquired by the limiting elevation angles of the mirrors (Section 3.3). In addition, we will see next (Section 5.2) that the baseline  $b$  also plays a major role in range computation.

Due to our model’s coaxial configuration, we could scan for pixel correspondences radially between a given pair of warped images ( $[I_1], [I_2]$ ) like in the approach taken by similar works such as [34]. However, it seems more convenient to work on a rectified image space, such as with panoramic images, where the search for correspondences can be performed using any of the various existing methods for perspective stereo views. Hence, we first demonstrate how these rectified panoramic images are produced (Section 5.1) and used for establishing point correspondences. Then, we proceed to study our triangulation method for the range computation from a given set of point correspondences (Section 5.2). Last, we show preliminary 3D point clouds as the outcome from such procedure.

5.1. Panoramic Images

Figure 14 illustrates how we form the respective panoramic image  $[\Xi_1]$  out of its warped omnidirectional image  $[I_1]$ . As illustrated in Figure 7,  $[I_i]$  is simply the region of interest out of the full image  $[I]$  where projection occurs via mirror  $i$ . However, we can safely refer to  $[I]$  because it will never be the case that projections via different mirrors overlap on the same pixel position  $^{[i]}m$ . In a few words, we obtain a panorama  $[\Xi_i]$  by reverse-mapping each discretized 3D point  $P_{cyl_i} \in S_{cyl_i}$  to its projected pixel coordinates  $^{[i]}m$  on  $[I]$  according to Section 3.2.



**Figure 14.** An example for the formation of panoramic image  $[\Xi_1]$  out of the omnidirectional image  $[I_1]$  (showing only the masked region of interest on the back of image plane  $\pi_{img_1}$ ). Any particular ray,  $\mathbf{v}_1$  indicated by its elevation and azimuth such as  $^{[i]}(\psi_1, \theta_1)$  that is directed towards the focus  $F_1$  must traverse the projection cylinder  $S_{cyl_1}$  at point  $P_{cyl_1}$ . More abstractly, the figure also shows how a pixel position  $^{[i]}m_\alpha$  on the panoramic pixel space gets mapped from its corresponding pixel position  $^{[i]}m_\alpha$  via function  $\mathbf{h}_{\Xi_1}$  defined in Equation (85). Although not up to scale, it’s crucial to notice the relative orientation between  $S_{cyl_1}$  and the back of the projection plane  $\pi_{img_1}$  where the omnidirectional image  $[I_1]$  is found.

More thoroughly, for  $i = \{1, 2\}$ ,  $S_{cyl_i}$  is the set of all valid 3D points  $P_{cyl_i}$  that lie on an imaginary unit cylinder centered along the Z-axis and positioned with respect to the mirror’s primary focus  $F_i$ . Recall that the radius of a unit cylinder is  $r_{cyl} = 1$ , so its circumference becomes  $w_{cyl} = 2\pi r_{cyl} = 2\pi$ . Noticed that the imaging ratio,  $\chi_{I_1:2} = \frac{h_{I_1}}{h_{I_2}}$ , illustrated in Figure 7 provides a way of inferring the scale between pairs of point correspondences. However, we achieve conforming scales among both panoramic representations by simply setting both cylinders to an equal height  $h_{cyl}$ , which is determined

from the system's elevation limits,  $(\theta_{sys,min}, \theta_{sys,max})$ , since they partake in the measurement of the system's vertical field of view given by Equation (54). Hence, we obtain

$$h_{cyl} = z_{cyl,max} - z_{cyl,min}, \text{ where } \begin{cases} z_{cyl,max} &= \tan(\theta_{sys,max}) \\ z_{cyl,min} &= \tan(\theta_{sys,min}) \end{cases} \quad (80)$$

Consequently, to achieve panoramic images  $[\Xi_i]$  of the same dimensions by maintaining a true aspect ratio  $w_{\Xi} : h_{\Xi}$ , it suffices to indicate either the width (number of columns)  $w_{\Xi}$  or the height (number of rows)  $h_{\Xi}$  as number of pixels. Here, we propose a custom method for resolving the panoramic image dimensions by setting the equality for the length  $l_{px}$  of an individual "square" pixel in the cylinder (behaving like a panoramic camera sensor):

$$l_{px} = \frac{w_{cyl}}{w_{\Xi}} = \frac{h_{cyl}}{h_{\Xi}} \quad (81)$$

For instance, if the width  $w_{\Xi}$  is given, then the height is simply  $h_{\Xi} = w_{\Xi} h_{cyl} / w_{cyl}$ .

To increase the processing speed for each panoramic image  $[\Xi_i]$ , we fill up its corresponding look-up-table  $LUT_{\Xi_i}$  of size  $w_{\Xi} \times h_{\Xi}$  that encodes the mapping for each panoramic pixel coordinates  $^{[\Xi_i]}m = ^{[\Xi_i]}[u, v]^T$  to its respective projection  $^{[i]}m = ^{[i]}[u, v]^T$  on the distorted image  $[I_i]$ . Each pixel  $^{[\Xi_i]}m$  gets associated with its cylinder's 3D point positioned at  $^{[i]}p_{cyl_i}$ , which can inherently be indicated by its elevation  $^{[i]}\theta_i$  and azimuth  $^{[i]}\psi_i$  (relative to the mirror's primary focus  $F_i$ ) as illustrated in Figure 4. Thus, the ray  $^{[i]}v_i$  of a particular 3D point directed about  $^{[i]}(\psi_i, \theta_i)$  must pass through  $P_{cyl_i}$  in order to get imaged as pixel  $^{[i]}m_i$ .

Since the circumference of the cylinder,  $w_{cyl}$ , is discretized with respect to the number of pixel columns or width  $w_{\Xi}$ , we use the pixel length  $l_{px}$  as the factor to obtain the arc length  $l_{\psi_i}$  spanned by the azimuth  $^{[i]}\psi_i$  out of a given  $^{[\Xi_i]}u$  coordinate on the panoramic image. Generally,

$$^{[i]}\psi_i = \frac{l_{\psi_i}}{r_{cyl}} = \frac{w_{cyl} - ^{[\Xi_i]}u l_{px}}{r_{cyl}} \quad (82)$$

or simply  $^{[i]}\psi_i = 2\pi - ^{[\Xi_i]}u l_{px}$  for the unit cylinder case.

An order reversal in the columns of the panorama is performed by Equation (82) because we account for the relative position between  $S_{cyl_i}$  and the projection plane  $\pi_{img}$ . For  $[\Xi_1]$ , Figure 14 depicts the unrolling of the cylindrical panoramic image onto a planar panoramic image. However, note that  $\pi_{img}$  is shown from above (or its back) in Figure 14, so the panorama visualization places the viewer inside the cylinder at  $F_1$ .

Similarly, the elevation angle  $^{[i]}\theta_i$  is inferred out the row or  $^{[\Xi_i]}v$  coordinate, which is scaled to its cylindrical representation by  $l_{px}$ . Recall that both cylinders have the same height,  $h_{cyl}$ , computed by Equation (80). By taking into account any row offset from the maximum height position,  $^{[i]}z_{cyl,max}$ , of the cylinder, we get

$$^{[i]}\theta_i = \arctan\left(^{[i]}z_{cyl,max} - ^{[\Xi_i]}v l_{px}\right) \quad (83)$$

Given these angles and assuming coaxial alignment, we evaluate the position vector  $^{[C]}p_{cyl_i}$  for a point on the panoramic cylinder with respect to the camera frame  $[C]$ :

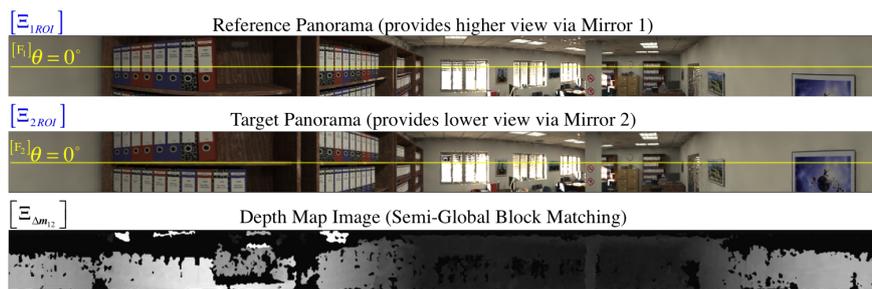
$$^{[C]}p_{cyl_i} = r_{cyl} \begin{bmatrix} \cos(^{[i]}\psi_i) \\ \sin(^{[i]}\psi_i) \\ \tan(^{[i]}\theta_i) \end{bmatrix} + ^{[C]}f_i \quad (84)$$

where  $r_{cyl}$  cancels out for a unit cylinder. The direction equations Equations (82) and (83) leading to Equation (84) as a process:  ${}^{[E_i]}m \xrightarrow[\text{Equation (83)}]{\text{Equation (82)}} {}^{[E_i]} \begin{pmatrix} \psi \\ \theta \end{pmatrix} \xrightarrow{\text{Equation (84)}} {}^{[C]}p_{cyl_i}$ , which is eventually used as the input argument to Equation (36) in order to determine pixel  ${}^{[I_i]}m$  via the mapping function  $\mathbf{h}_{E_i} : \mathbb{R}^2 \mapsto \mathbb{R}^2$ ,

$${}^{[I_i]}m \leftarrow \mathbf{h}_{E_i}({}^{[E_i]}m) := \mathbf{f}_{\varphi_i} \left( {}^{[C]}p_{cyl_i} \Big|_{[E_i]m} \right) \quad (85)$$

### 5.1.1. Stereo Matching on Panoramas

We understand that the algorithm chosen for finding matches is crucial to attain correct pixel disparity results. We refer the reader to [35] for a detailed survey of stereo correspondence methods. After comparing various block matching algorithms, we were able to obtain acceptable disparity maps with the semi-global block matching (SGBM) method introduced by [36], which can find subpixel matches in real time. As a result of this stereo block matcher among the pair of panoramic images ( $[E_1], [E_2]$ ), we get the dense disparity map  $[E_{\Delta m_{12}}]$  visualized as an image in Figures 15 and 21a. Note that valid disparity values must be positive ( $\Delta m_{12}|_{[E_i]m_1} > 0$ ) and they are given with respect to the reference image, in this case,  $[E_1]$ . In addition, recall that no stereo matching algorithm (as far as we are aware) is totally immune to mismatches due to several well-known reasons in the literature such as ambiguity of cyclic patterns.



**Figure 15.** For the synthetic omnidirectional image  $[I]$  shown in Figure 2b, we generate its pair of panoramic images ( $[E_1], [E_2]$ ) using the procedure explained in Section 5.1. Note that we only work on the SROI (shown here) to perform a semi-global block match between the panoramas as indicated in Section 5.1.1. The resulting disparity map,  $[E_{\Delta m_{12}}]$ , is visualized at the bottom as a gray-scale panoramic image normalized about its 256 intensity levels, where brighter colors imply larger disparity values. To distinguish the relative vertical view of both panoramas, we have annotated the row position of the zero-elevation.

An advantage of the block (window) search for correspondences is that it can be narrowed along epipolar lines. Unlike the traditional horizontal stereo configuration, our system captures panoramic images whose views differ in a vertical fashion. As shown in [14], the unwrapped panoramas contain vertical, parallel epipolar lines that facilitate the pixel correlation search. Thus, given a pixel position  ${}^{[E_1]}m_1$  on the reference panorama  $[E_1]$  and its disparity value  $\Delta m_{12}|_{[E_1]m_1}$ , we can resolve the correspondence  ${}^{[E_2]}m_2$  pixel coordinate on the target image,  $[E_2]$ , by simply offsetting the  $v$ -coordinate with the disparity value:

$${}^{[E_2]}m_2 = \begin{bmatrix} u_1 \\ v_1 + \Delta m_{12}|_{[E_1]m_1} \end{bmatrix} \quad (86)$$

## 5.2. Range from Triangulation

Recall the duality that states a point  $P_w$  as the intersection of a pair of lines. Regardless of the correspondence search technique employed, such as block stereo matching between panoramas  $[\Xi_i]$  (Section 5.1.1) or feature detection directly on  $[I]$ , we can resolve for  ${}^{[I]}(\mathbf{m}_1, \mathbf{m}_2)$ . From Equations (42) and (49), we obtain the respective pair of back-projected rays  $({}^{[F_1]}\mathbf{v}_1, {}^{[F_2]}\mathbf{v}_2)$ , emanating from their respective physical viewpoints,  $F_1$  and  $F_2$ , which are separated by baseline  $b$ . We can compute elevation angles  $\theta_1$  and  $\theta_2$  using equations Equations (43) and (50). Then, we can triangulate the back-projected rays in order to calculate the horizontal range  $\rho_w$  defined in Equation (22), as follows:

$$\rho_w = \left| \frac{b \cos(\theta_1) \cos(\theta_2)}{\sin(\theta_1 - \theta_2)} \right| \quad (87)$$

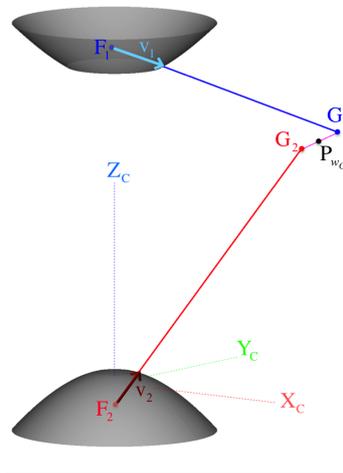
Finally, we obtain the 3D position of  $P_w$ :

$${}^{[C]}\mathbf{p}_w = \begin{bmatrix} -\rho_w \cos(\psi_{12}) \\ -\rho_w \sin(\psi_{12}) \\ c_1 - \rho_w \tan(\theta_1) \end{bmatrix} \quad (88)$$

where  $\psi_{12}$  is the common azimuthal angle (on the  $XY$ -plane) for coplanar rays, so it can be determined either by Equation (44) or Equation (51). Functionally, we define the “naive” intersection function that implements Equations (87) and (88) such that

$${}^{[C]}\mathbf{p}_w \leftarrow \mathbf{f}_\Delta((\theta_1, \psi_1), (\theta_2, \psi_2), \boldsymbol{\theta}) \quad (89)$$

where  $\boldsymbol{\theta}$  is the model parameters vector defined in Equation (4) and can be omitted when calling this function because the model parameters should not change (ideally).



**Figure 16.** The more realistic case of skew back-projection rays  $(\mathbf{v}_1, \mathbf{v}_2)$  approximates the triangulated point  $P_w$  by getting the midpoint  $P_{w_G}$  on the common perpendicular line segment  $\overline{G_1G_2} : \lambda_{1\perp 2} \hat{\mathbf{v}}_{1\perp 2}$ . Note that the visualized skew rays were formed from a pixel correspondence pair  ${}^{[I]}(\mathbf{m}_1, \mathbf{m}_2)$  and by offsetting the coordinate  $u_2$  by 15 pixels.

### 5.2.1. Common Perpendicular Midpoint Triangulation Method

Because the coplanarity of these rays cannot be guaranteed (skew rays case), a better triangulation approximation while considering coaxial misalignments is to find the midpoint of their common perpendicular line segment (as attempted in [23]). As illustrated in Figure 16, we define the common

perpendicular line segment  $\overline{G_1G_2}$  as the parametrized vector  $\mathbf{v}_{1\perp 2} = \lambda_{1\perp 2}\hat{\mathbf{v}}_{1\perp 2}$ , for the unit vector normal to the back-projected rays,  $\mathbf{v}_1$  and  $\mathbf{v}_2$ , such that:

$$\hat{\mathbf{v}}_{1\perp 2} = \frac{\mathbf{v}_1 \otimes \mathbf{v}_2}{\|\mathbf{v}_1 \otimes \mathbf{v}_2\|} \tag{90}$$

If the rays are not parallel ( $\|\mathbf{v}_1 \otimes \mathbf{v}_2\| \neq 0$ ), we can compute the “exact” solution,  $\lambda = [\lambda_{G_1}, \lambda_{G_2}, \lambda_{1\perp 2}]^T$ , of the well-determined linear matrix equation

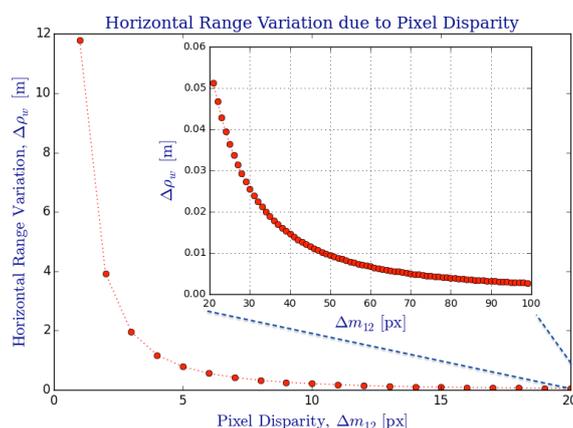
$$\mathbf{V}\lambda = \mathbf{b}, \text{ where } \mathbf{V} = \begin{bmatrix} \mathbf{v}_1, & -\mathbf{v}_2, & \hat{\mathbf{v}}_{1\perp 2} \end{bmatrix} \text{ and } \mathbf{b} = {}^{[c]}\mathbf{f}_2 - {}^{[c]}\mathbf{f}_1 \tag{91}$$

It follows that the location of the midpoint  $P_{w_G}$  on the common perpendicular  $\mathbf{v}_{1\perp 2}$  with respect to the common frame  $[C]$  is

$${}^{[c]}\mathbf{p}_{w_G} = {}^{[c]}\mathbf{f}_1 + \lambda_{G_1} {}^{[f_1]}\mathbf{v}_1 + \frac{1}{2}\lambda_{1\perp 2} {}^{[G_1]}\hat{\mathbf{v}}_{1\perp 2} \tag{92}$$

### 5.2.2. Range Variation

Before we introduce an uncertainty model for triangulation (Section 5.3), we briefly analyze how range varies according to the possible combinations of pixel correspondences,  ${}^{[I]}(\mathbf{m}_1, \mathbf{m}_2)$  on the image  $[I]$ . Here, we demonstrate how a radial variation of discretized pixel disparities,  $\Delta m_{12}$ , affects the 3D position of a point obtained from triangulation (Section 5.2). Figure 17 demonstrates the nonlinear characteristics of the variation in horizontal range,  $\Delta\rho_w$ , from the discrete relation between pixel positions  ${}^{[I]}\mathbf{m}_i$  and their respective back-projected (direction) rays obtained from  $\mathbf{f}_{\beta_i}$  and triangulated via function  $\mathbf{f}_\Delta$  defined in Equation (89). It can be observed that the horizontal range variation,  $\Delta\rho_w$ , increases quadratically as  $\Delta m_{12} \rightarrow 1\text{px}$ , which is the minimum *discrete* pixel disparity, which provides a maximum horizontal range  $\rho_{w,max} \approx [18, 28]$  m (computed analytically). The main plot of Figure 17 shows the small disparity values in the interval  $\overline{\Delta m_{12}} = [1, 20]$  px, whereas the subplot is a zoomed-in extension of the large disparity cases in the interval  $\overline{\Delta m_{12}} = [20, 100]$  px.



**Figure 17.** Variation of horizontal range,  $\Delta\rho_w$ , due to change in pixel disparity  $\Delta m_{12}$  on the omnidirectional image,  $[I]$ . There exists a “nonlinear & inverse” relation between the change in depth from triangulation ( $\Delta\rho_w$ ) and the number of disparity pixels ( $\Delta m_{12}$ ) available from the omnistereo image pair ( $[I_1], [I_2]$ ), which are exclusive subspaces of  $[I]$ .

The current analysis is an indicative that triangulation error (e.g., due to false pixel correspondences) may have a severe effect on range accuracy that increases quadratically with distance as it can be appreciated with the 8 m variation on the disparity interval  $\overline{\Delta m_{12}} = [1, 2]\text{px}$ . Also, observe the example of Figure 20 for a reconstructed point cloud, where this range sensing characteristic is

more noticeable for faraway points. In fact, the following uncertainty model provides a probabilistic framework for the triangulation error (uncertainty) that agrees with the current numerical claims.

### 5.3. Triangulation Uncertainty Model

Let  $\mathbf{f}_{P_w}$  be the vector-valued function that computes the 3D coordinates of point  $P_{w_G}$  with respect to [C] as the common perpendicular midpoint defined in Equation (92). We express this triangulation function component-wise as follows:

$${}^{[C]}\mathbf{p}_{w_G} \leftarrow \mathbf{f}_{P_w}(\mathbf{m}_{12}) := \begin{bmatrix} \mathbf{f}_{x_w}(\mathbf{m}_{12}) \\ \mathbf{f}_{y_w}(\mathbf{m}_{12}) \\ \mathbf{f}_{z_w}(\mathbf{m}_{12}) \end{bmatrix} \quad (93)$$

where  $\mathbf{m}_{12} = [u_1, v_1, u_2, v_2]$  is composed by the pixel coordinates of the correspondence  ${}^{[I]}(\mathbf{m}_1, \mathbf{m}_2)$  upon which to base the triangulation (Section 5.2).

Without loss of generality, we model a *multivariate Gaussian* uncertainty model for triangulation, so that the position vector  ${}^{[C]}\mathbf{p}_{w_G}$  of any world point is centered at its mean  ${}^{[C]}\boldsymbol{\mu}_{\mathbf{f}_{P_w}}$  with a  $3 \times 3$  covariance matrix  $\boldsymbol{\Sigma}_{\mathbf{f}_{P_w}}$ :

$${}^{[C]}\boldsymbol{\mu}_{\mathbf{f}_{P_w}} = \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix}, \boldsymbol{\Sigma}_{\mathbf{f}_{P_w}} = \begin{bmatrix} \sigma_{f_{x_w}}^2 & \sigma_{f_{x_w}} \sigma_{f_{y_w}} & \sigma_{f_{x_w}} \sigma_{f_{z_w}} \\ \sigma_{f_{x_w}} \sigma_{f_{y_w}} & \sigma_{f_{y_w}}^2 & \sigma_{f_{y_w}} \sigma_{f_{z_w}} \\ \sigma_{f_{x_w}} \sigma_{f_{z_w}} & \sigma_{f_{y_w}} \sigma_{f_{z_w}} & \sigma_{f_{z_w}}^2 \end{bmatrix} \quad (94)$$

However, since  $\mathbf{f}_{P_w}$  is a non-linear vector-valued function, we linearize it by approximation to a first-order Taylor expansion and we use its Jacobian matrix to propagate the uncertainty (covariance) as in the linear case as follows:

$$\boldsymbol{\Sigma}_{\mathbf{f}_{P_w}} = \mathbf{J}_{\mathbf{f}_{P_w}} \boldsymbol{\Omega}_{\mathbf{m}_{12}} \mathbf{J}_{\mathbf{f}_{P_w}}^\top \quad (95)$$

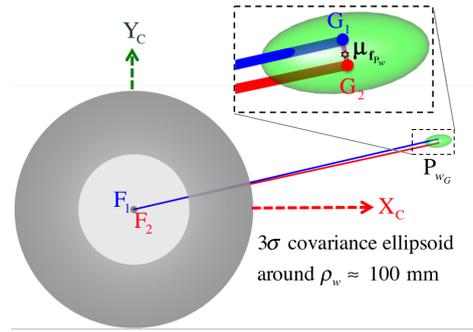
where the  $3 \times 4$  Jacobian matrix for the triangulation function is

$$\mathbf{J}_{\mathbf{f}_{P_w}} = \begin{bmatrix} \frac{\partial f_{x_w}}{\partial u_1} & \frac{\partial f_{x_w}}{\partial v_1} & \frac{\partial f_{x_w}}{\partial u_2} & \frac{\partial f_{x_w}}{\partial v_2} \\ \frac{\partial f_{y_w}}{\partial u_1} & \frac{\partial f_{y_w}}{\partial v_1} & \frac{\partial f_{y_w}}{\partial u_2} & \frac{\partial f_{y_w}}{\partial v_2} \\ \frac{\partial f_{z_w}}{\partial u_1} & \frac{\partial f_{z_w}}{\partial v_1} & \frac{\partial f_{z_w}}{\partial u_2} & \frac{\partial f_{z_w}}{\partial v_2} \end{bmatrix} \quad (96)$$

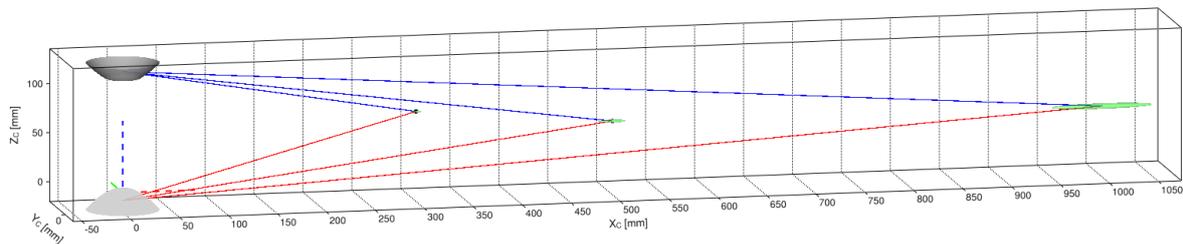
and the  $4 \times 4$  covariance matrix of the pixel arguments being

$$\boldsymbol{\Omega}_{\mathbf{m}_{12}} = \sigma_{px}^2 \mathbf{I}_4 \quad (97)$$

where we assume  $\sigma_{px} = 1$  px for the standard deviation of each pixel coordinate in the *discretized* pixel space. The complete symbolic solution of  $\boldsymbol{\Sigma}_{\mathbf{f}_{P_w}}$  is too involved to appear in this manuscript. However, in Figure 18, we show the top-view of the covariance ellipsoid drawn at a three- $\sigma_{\mathbf{f}_{P_w}}$  level for a point triangulated nearly around  $\rho_w \approx 100$  mm. Figure 19 visualizes uncertainty ellipsoids drawn at a one- $\sigma_{\mathbf{f}_{P_w}}$  level for several triangulation ranges. We refer the reader to the end of Section 6.3 where we validate the safety of this 1 pixel deviation assumption through experimental results using subpixel precision.



**Figure 18.** Top-view of the three-sigma level ellipsoid for the triangulation uncertainty of a pixel pair  ${}^{[l]}(m_1, m_2)$  with an assumed standard deviation  $\sigma_{px} = 1$  px.



**Figure 19.** Uncertainty ellipsoids for triangulated points at ranges  $\rho_w \approx \{0.3, 0.5, 1.0\}$  m.

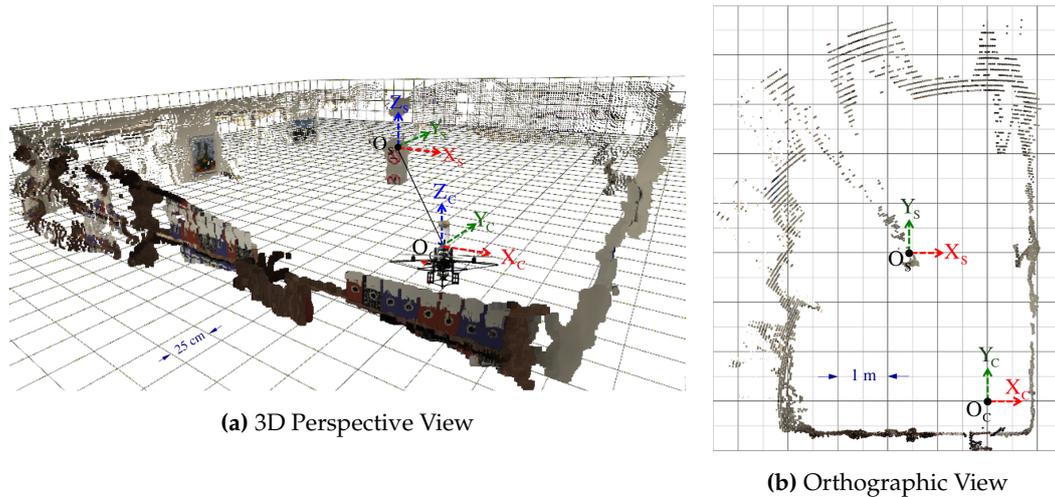
## 6. Experiment Results

In this section, we demonstrate the capabilities of the omnistereo sensor to provide 3D information either as dense point clouds or as for the registration of sparse 2D features and 3D points. We also evaluate the precision of both projection and triangulation of a few detected corners from a chessboard whose various 3D poses are given as ground-truth.

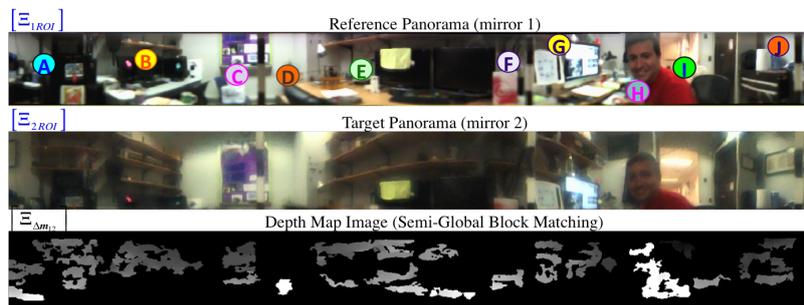
### 6.1. Dense 3D Point Clouds

By implementing the process described in Section 5, we begin by visualizing the dense point-cloud obtained from the omnidirectional synthetic image given in Figure 2b, whose actual size is  $1280 \times 960$  pixels. The associated panoramic images,  $[\Xi_i]$ , were obtained using function  $h_{\Xi_i}$  defined in Equation (85) and are shown in Figure 15. Pixel correspondences ( ${}^{[2i]}m_1, {}^{[2i]}m_2$ ) on the panoramic representations are mapped via  $h_{\Xi_i}$  into their respective image positions  ${}^{[l]}(m_1, m_2)$ . Then, these are triangulated with  ${}^{[c]}f_{pw}$  given in Equation (93), resulting in the set (cloud) of color 3D points  $P_{\Delta}$  visualized in Figure 20. Here, the synthetic scene (Figure 2a) is for a room 5.0 m wide (along its X-axis), 8.0 m long (along its Y-axis), and 2.5 m high (along its Z-axis). With respect to the scene center of coordinates,  $[S]$ , the catadioptric omnistereo sensor,  $[C]$ , is positioned at  ${}^{[S]}t = [1.60, -2.85, 0.16]^T$  in meters.

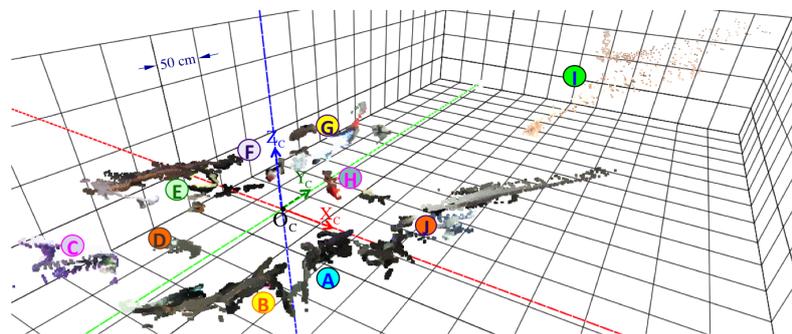
We also present results from a real experiment using the prototype described in Section 4.3.2 and shown in Figure 13a. The panoramic images and dense point cloud shown in Figure 21 are obtained by implementing the pertinent functions described throughout this manuscript and by holding the SVP assumption of an ideal configuration. We provide these qualitative results as preliminary proof of concept for the proposed sensor after employing a calibration procedure based on the generalized unified model proposed in [37].



**Figure 20.** A 3-D dense point cloud computed out of the synthetic model that rendered the omnidirectional image shown in Figure 2b. Pixel correspondences are established via the panoramic depth map visualized in Figure 15. The 3D point triangulation implements the common perpendicular midpoint method indicated in Section 5.2.1. The position of the omnistereo sensor mounted on the quadrotor is annotated as frame [C] with respect to the scene’s coordinates frame [S]. (a) 3D visualization of the point cloud (the quadrotor with the omnistereo rig has been added for visualization only); (b) Orthographic projection of the point cloud to the XY-plane of the visualization grid.



(a) Panoramic Images

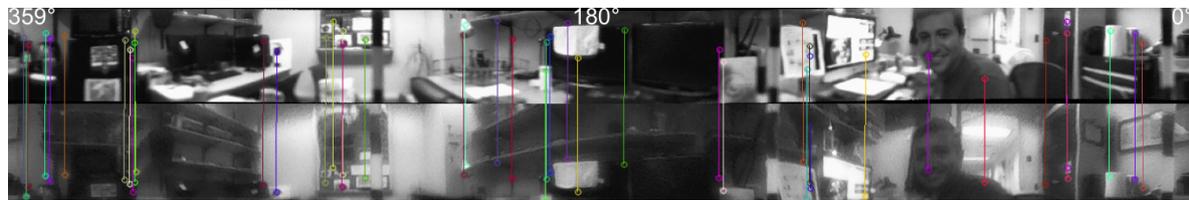


(b) 3D Point Cloud

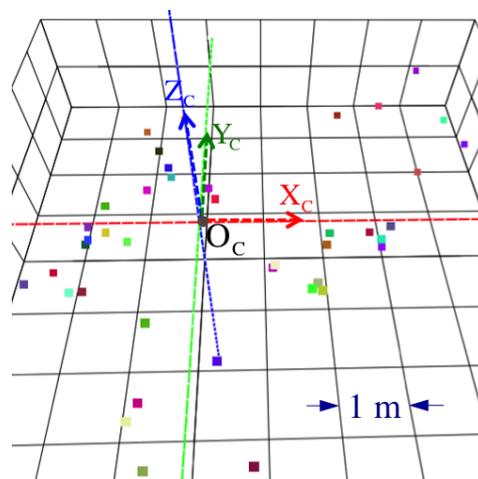
**Figure 21.** Real-life experiment using the 37 mm-radius prototype and a single  $2592 \times 1944$  pixels image where the rig was positioned in the middle of the room observed in Figure 13a. Some landmarks of the scene are annotated as following: (A) appliances, (B) monitors and shelf, (C) back wall, (D) chair, (E) monitors and shelf, (F) book, (G) monitors, (H) person, (I) hallway, (J) supplies. For the point cloud, the grid size is 0.50 m in all directions and points are thickened for clarity.

## 6.2. Sparse 3D Points from Features

Using the SURF feature detector and descriptors [38], Figure 22 demonstrates 44 correct matches that are triangulated with Equation (93). Sparse 3D points can be useful for applications of visual odometry where the sensor changes poses and those registered point features can be matched against new images. Please, refer to [39] for a tutorial on visual odometry.



(a) SURF Features Matched on Panoramic Images



(b) 3D Sparse Point Cloud

**Figure 22.** Sparse point correspondences for the real-life image from Figure 13b. Point correspondences are identifiable by random colors that persist in both the panoramic image and the respective triangulated 3D points (scaled-up for visualization).

## 6.3. Triangulation Evaluation

### 6.3.1. Evaluation of Synthetic Rig

Due to the unstructured nature of the dense point clouds previously discussed, we proceed to triangulate sets of sparse 3D points whose positions with respect to the omnistereo sensor camera frame,  $[C]$ , are known in advance. We synthesize a calibration chessboard pattern  $[G]$  containing  $m \times n$  square cells for various predetermined poses  ${}^{[C]}T_h$ . Since the sensor is assumed to be rotationally symmetric, it suffices to experiment with groups of  $L = 4$  chessboard patterns situated at a given horizontal range. A total of  $Lmn$  3D points are available for each range group. Each corner point's position  ${}^{[C]}p_j$  is found with respect to  $[C]$  via the frame transformation  ${}^{[C]}p_{j,g} = {}^{[C]}T_h^{[G_g]} p_j$  for all indices  $j \in \{1, \dots, mn\}, g \in \{1, \dots, L\}$ .

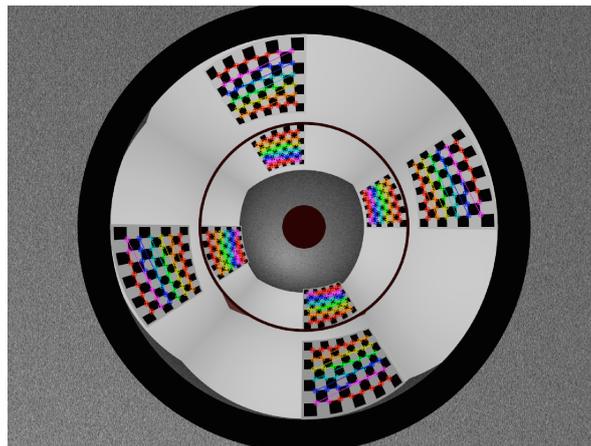
Figure 23 shows the set of detected corner points on the image from the group of patterns set to a range of  ${}^{[C]}p_G = 2$  m. We adjust the pattern's cell sizes accordingly so its points can be safely discerned by an automated corner detector [35]. We systematically establish correspondences of pattern points on the omnidirectional image, and proceed to triangulate with Equation (93). For each range group of points, we compute the root-mean-square of the 3D position errors (RMSE) between the observed

**Table 4.** Results of RMSE from Synthetic Triangulation Experiment.

${}^{[C]}\rho_G$ [m]	RMSE [mm]	SD [mm]
0.25	0.46	0.31
0.50	1.20	0.71
1.0	4.62	2.55
2.0	14.85	9.06
4.0	57.67	31.34
8.0	219.09	129.92

(triangulated) points  ${}^{[C]}\tilde{\mathbf{p}}_j \leftarrow \mathbf{f}_{P_w}(\tilde{\mathbf{m}}_1, \tilde{\mathbf{m}}_2)$  and the true (known) points  ${}^{[C]}\mathbf{p}_j$  that were used to describe the ray-traced image. Table 4 compiles the RMSE results and the standard deviation (SD) for some group of patterns whose frames  $[G_g]$ , are located at specified horizontal ranges  ${}^{[C]}\rho_G \in [0.25, 8.0]$  m away from  $[C]$ .

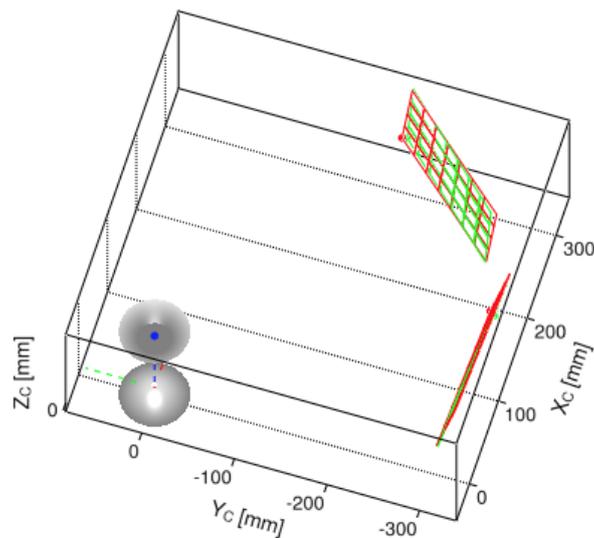
We notice that for all the 3D points in the synthetic patterns, we obtained an average error of 0.1 px with a standard deviation  $\tilde{\sigma}_{px} = 0.05$  px for the subpixel detection of corners on the image versus their theoretical values obtained from  $\mathbf{f}_{\varphi_i}$  defined in Equation (36). This last experiment helps us validate the pessimistic choice of  $\sigma_{px} = 1$  px for the discrete pixel space in the triangulation uncertainty model proposed in Section 5.3.



**Figure 23.** Example of sparse point correspondences detected with subpixel precision from corners on the chessboard patterns around the omnistereo sensor. The size of the rendered images for this experiment is  $1280 \times 960$  pixels. For this example's patterns, the square cell size is 140 mm. The RMSE for this set of points at  ${}^{[C]}\rho_G = 2$  m is approximately 15 mm (Table 4).

### 6.3.2. Evaluation of Real-Life Rig

The following experiment uses  $L = 5$  different poses of a real chessboard pattern with  $5 \times 8$  corner points where the square cell size is 24 mm. As done in Section 6.3.1, the evaluated error is the Euclidean norms between the triangulated points and the ground-truth positions of the chessboard poses captured via a motion capture system. The RMSE for all projected points in this set of chessboard patterns is 2.5 pixels with a standard deviation of 1.5 pixels. The RMSE for all triangulated points in this set is 3.5 mm with a standard deviation of 1.4 mm. Figure 24 visually confirms the proximity of the triangulated chessboard poses against the ground-truth pose information.



**Figure 24.** Visualization of estimated 3D poses for some chessboard patterns using the real-life omnistereo rig. Color annotations: ground-truth poses (green), estimated triangulated poses (red).

## 7. Discussion and Future Work

The portable aspect of the proposed omnistereo sensor is one of its greatest advantages, as discussed in the introduction section. The total weight of the big rig using 37 mm-radius mirrors is about 550 g, so it can be carried by the AscTec Pelican quadrotor under its payload limitations of 650 g. The mirror profiles maximize the stereo baseline while obeying the various design constraints such as size and field of view. Currently, the mirrors are custom-manufactured out of brass using CNC machining. However, it is possible to reduce the system's weight dramatically by employing lighter materials.

In reality, it is almost impossible to assemble a perfect imaging system that fulfills the SVP assumption and avoids the triangulation uncertainty studied in Section 5.3 on top of the error already introduced by any feature matching technique. The coaxial misalignment of the folded mirrors-camera system, defocus blur of the lens, and the inauspicious glare from the support tube are all practical caveats we need to overcome for better 3D sensing tasks. As described in the text for the real-life rig, we have avoided the traditional use of a support cylinder in order to workaround the cross-reflections and glare issues. Possible vibrations caused by the robot dynamics are reduced by vibration pads placed on the sensor-body interface. Details about our tentative calibration method for vertically-folded omnistereo systems has not been included in the current study since we would like the reader's attention to be devoted to the sensor characteristics defended by this analysis.

Our ongoing research is also focusing on the development of efficient software algorithms for real-time 3D pose estimation from point clouds. Bear in mind that all the experimental results demonstrated in this manuscript rely upon a single camera snapshot. We understand that the narrow vertical field-of-view where stereo vision operates is a limiting factor for dense scene reconstruction from a single image, so we have also considered non-optimal geometries for the quadrotor's view. In fact, increasing the region of interest for stereo (SROI) while maintaining the wide baseline implies an enlargement of each mirror's radius. We believe that our omnidirectional system is more advantageous than forward-looking sensors because it can provide a robust pose estimation by extracting 3D point features from all around the scene at once. As in our past work [24], fusing multiple modalities (e.g., stereo and optical-flow) is a possibility in order to resolve the scale-factor problem inherent while performing structure from motion over the non-stereo regions of each mirror (near the poles).

In this work, we performed an extensive study of the proposed omnistereo sensor's properties, such as its spatial resolution and triangulation uncertainty. We validated the projection accuracy of the

synthetic model (the ideal case) where 3D points in the world are given exactly. In order to validate the precision of the real sensor, we require a perfectly constructed and assembled device so point projections can be accepted as the ultimate truth. This is hard to achieve at a low-cost prototyping stage. Although we acquired ground-truth 3D points via a position capture system alone, we deem this insufficient to validate the imaging accuracy of the real sensor because the precision of the calibration method is truly what is being accounted for. For reproducibility purposes, source code is available for the implementation of the theoretical omnistereo model, optimization, plots and figures presented in this analysis [40].

**Acknowledgments:** This work was supported in part by U.S. Army Research Office grantNo. W911NF-09-1-0565, U.S. National Science Foundation grant No. IIS-0644127, and a Ford Foundation Pre-doctoral Fellowship awarded to Carlos Jaramillo.

**Author Contributions:** The work presented in this paper is a collaborative development by all of the authors. C. Jaramillo wrote this manuscript, carried out all the experiments and conceived the extensive analysis of the omnistereo sensor studied here. R.G. Valenti contributed with the analytical derivation of various equations and manuscript revisions. L. Guo established the geometrical model and rules for the constrained optimization of design parameters. J. Xiao funded and guided this entire study and helped with revisions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Symbolic Notation

$P_i$ : a point  $\in \mathbb{R}^3$  where post-subscript  $i$  as a unique identifier.

$[A]$ : a reference frame or image space with origin  $O_A$ .

$^{[A]}\mathbf{p}_i$ : The position vector of  $P_i$  with respect to reference frame  $[A]$ .

$^{[A]}\mathbf{p}_{i,h}$ : for homogeneous coordinates.

$^{[I]}\mathbf{m}_i$ : a 2D point or pixel position on image frame  $[I]$ .

$\|\mathbf{p}_i\|$ : the magnitude (Euclidean norm) of  $\mathbf{p}_i$ .

$\hat{\mathbf{q}}$ : A unit vector so  $\|\hat{\mathbf{q}}\| = 1$ .

$\mathbf{M}_i$ : a  $3 \times 3$  matrix, or  $\mathbf{M}_{i,h}$  in homogeneous coordinates.

$f_s$ : a scalar-valued function that outputs some  $s$ .

$\mathbf{f}_v$ : a vector-valued function for the computation of  $\mathbf{v}$ .

All coordinate systems obey the right-hand rule unless otherwise indicated.

## References

1. Marani, R.; Renò, V.; Nitti, M.; D'Orazio, T.; Stella, E. A Compact 3D Omnidirectional Range Sensor of High Resolution for Robust Reconstruction of Environments. *Sensors* **2015**, *15*, 2283–2308.
2. Valenti, R.G.; Dryanovski, I.; Jaramillo, C.; Strom, D.P.; Xiao, J. Autonomous quadrotor flight using onboard RGB-D visual odometry. In Proceedings of the International Conference on Robotics and Automation (ICRA 2014), Hong Kong, China, 31 May–7 June 2014; pp. 5233–5238.
3. Khoshelham, K.; Elberink, S.O. Accuracy and resolution of Kinect depth data for indoor mapping applications. *Sensors* **2012**, *12*, 1437–1454.
4. Payá, L.; Fernández, L.; Gil, A.; Reinoso, O. Map building and monte carlo localization using global appearance of omnidirectional images. *Sensors* **2010**, *10*, 11468–11497.
5. Berenguer, Y.; Payá, L.; Ballesta, M.; Reinoso, O. Position Estimation and Local Mapping Using Omnidirectional Images and Global Appearance Descriptors. *Sensors* **2015**, *15*, 26368–26395.
6. Hrabar, S.; Sukhatme, G. Omnidirectional vision for an autonomous helicopter. In Proceedings of the International Conference on Robotics and Automation (ICRA), Taipei, Taiwan, 14–19 September 2003; pp. 3602–3609.
7. Hrabar, S. 3D path planning and stereo-based obstacle avoidance for rotorcraft UAVs. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Nice, France, 22–26 September 2008; pp. 807–814.

8. Orghidan, R.; Mouaddib, E.M.; Salvi, J. Omnidirectional depth computation from a single image. In Proceedings of the IEEE International Conference on Robotics and Automation, Barcelona, Spain, 18–22 April 2005; pp. 1222–1227.
9. Paniagua, C.; Puig, L.; Guerrero, J.J. Omnidirectional structured light in a flexible configuration. *Sensors* **2013**, *13*, 13903–13916.
10. Byrne, J.; Cosgrove, M.; Mehra, R. Stereo based obstacle detection for an unmanned air vehicle. In Proceedings of the International Conference on Robotics and Automation, Orlando, FL, USA, 15–19 May 2006.
11. Smadja, L.; Benosman, R.; Devars, J. Hybrid stereo configurations through a cylindrical sensor calibration. *Mach. Vis. Appl.* **2006**, *17*, 251–264.
12. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*, 2nd ed.; Volume 2, Cambridge University Press: Cambridge, UK, 2004.
13. Sturm, P.; Ramalingam, S.; Tardif, J.P.; Gasparini, S.; Barreto, J.P. Camera models and fundamental concepts used in geometric computer vision. *Found. Trends®Comput. Graph. Vis.* **2010**, *6*, 1–183.
14. Gluckman, J.; Nayar, S.K.; Thoresz, K.J. Real-Time Omnidirectional and Panoramic Stereo. *Comput. Vis. Image Underst.* **1998**.
15. Koyasu, H.; Miura, J.; Shirai, Y. Realtime omnidirectional stereo for obstacle detection and tracking in dynamic environments. In Proceedings of the International Conference on Intelligent Robots and Systems (IROS), Maui, HI, USA, 29 October–3 November 2001; Volume 1, pp. 31–36.
16. Bajcsy, R.; Lin, S.S. High resolution catadioptric omni-directional stereo sensor for robot vision. In Proceedings of the 2003 IEEE International Conference on Robotics and Automation, Taipei, Taiwan, 14–19 September 2003; pp. 1694–1699.
17. Cabral, E.E.; de Souza, J.C.J.; Hunold, M.C. Omnidirectional stereo vision with a hyperbolic double lobed mirror. In Proceedings of the 17th International Conference on Pattern Recognition (ICPR), Cambridge, UK, 23–26 August 2004; pp. 0–3.
18. Su, L.; Zhu, F. Design of a novel stereo vision navigation system for mobile robots. In Proceedings of the IEEE Robotics and Biomimetics (ROBIO), Hong Kong, China, 5–9 July 2005; pp. 611–614.
19. Mouaddib, E.M.; Sagawa, R. Stereovision with a single camera and multiple mirrors. In Proceedings of the International Conference on Robotics and Automation, Barcelona, Spain, 18–22 April 2005; pp. 800–805.
20. Schönbein, M.; Kitt, B.; Lauer, M. Environmental Perception for Intelligent Vehicles Using Catadioptric Stereo Vision Systems. In Proceedings of the European Conference on Mobile Robots (ECMR), Örebro, Sweden, 7–9 September 2011; pp. 1–6.
21. Yi, S.; Ahuja, N. An Omnidirectional Stereo Vision System Using a Single Camera. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, 20–24 August 2006; pp. 861–865.
22. Nayar, S.K.; Peri, V. Folded catadioptric cameras. In Proceedings of the 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Fort Collins, CO, USA, 23–25 June 1999; pp. 217–223.
23. He, L.; Luo, C.; Zhu, F.; Hao, Y. Stereo Matching and 3D Reconstruction via an Omnidirectional Stereo Sensor. In *Motion Planning*; Number 60575024; In-Tech Education and Publishing: Vienna, Austria, 2008; pp. 123–142.
24. Labutov, I.; Jaramillo, C.; Xiao, J. Generating near-spherical range panoramas by fusing optical flow and stereo from a single-camera folded catadioptric rig. *Mach. Vis. Appl.* **2011**, *24*, 1–12.
25. Swaminathan, R.; Grossberg, M.D.; Nayar, S.K. Caustics of catadioptric cameras. In Proceedings of the Eighth IEEE International Conference on Computer Vision (ICCV 2001), Vancouver, BC, Canada, 7–14 July 2001; Volume 2, pp. 2–9.
26. Jang, G.; Kim, S.; Kweon, I. Single camera catadioptric stereo system. In Proceedings of the Workshop on Omnidirectional Vision, Camera Networks and Nonclassical Cameras (OMNIVIS2005), Beijing, China, 21 October 2005.
27. Jaramillo, C.; Guo, L.; Xiao, J. A Single-Camera Omni-Stereo Vision System for 3D Perception of Micro Aerial Vehicles (MAVs). In Proceedings of the IEEE Conference on Industrial Electronics and Applications (ICIEA), Melbourne, Australia, 19–21 June 2013; Volume 10016.

28. Ascending Technologies (AscTec). Available online: <http://www.ascotec.de/en/uav-uas-drones-rpas-roav/> (accessed on 23 May 2014).
29. Baker, S.; Nayar, S.K. A theory of single-viewpoint catadioptric image formation. *Int. J. Comput. Vis.* **1999**, *35*, 175–196.
30. Nayar, S.K.; Baker, S. Catadioptric Image Formation. In Proceedings of the 1997 DARPA Image Understanding Workshop, New Orleans, LA, USA, May 1997; pp. 1431–1437.
31. Gaspar, J.; Deccó, C.; Okamoto, J.J.; Santos-Victor, J.; Sistemas, I.D.; Pais, A.R.; Brazil, S.P. Constant resolution omnidirectional cameras. In Proceedings of the OMNIVIS'02 Workshop on Omni-directional Vision, Copenhagen, Denmark, 2 June 2002.
32. Forsgren, A.; Gill, P.; Wright, M. Interior Methods for Nonlinear Optimization. *Soc. Ind. Appl. Math. (SIAM Rev.)* **2002**, *44*, 525–597.
33. Tuytelaars, T.T.; Mikolajczyk, K. Local Invariant Feature Detectors- A Survey. *Found. Trends® Comput. Graph. Vis.* **2008**, *3*, 177–280.
34. Spacek, L. Coaxial Omnidirectional Stereopsis. *Computer Vision-ECCV 2004*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2004; pp. 354–365.
35. Bradski, G.; Kaehler, A. *Learning OpenCV: Computer vision with the OpenCV library*; O'Reilly Media, Inc.: Sebastopol, California, 2008.
36. Hirschmüller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 328–341.
37. Xiang, Z.; Dai, X.; Gong, X. Noncentral catadioptric camera calibration using a generalized unified model. *Opt. Lett.* **2013**, *38*, 1367–1369.
38. Bay, H.; Ess, A.; Tuytelaars, T.; Vangool, L. Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359.
39. Scaramuzza, D.; Fraundorfer, F. Visual Odometry Part 1: The First 30 Years and Fundamentals. *IEEE Robot. Autom. Mag.* **2011**, *18*, 80–92.
40. Source Code Repository. Available online: [https://github.com/ubuntuSlave/omnistereo\\_sensor\\_design](https://github.com/ubuntuSlave/omnistereo_sensor_design) (accessed on 5 February 2016).



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).