

Uniform Subdivision of Omnidirectional Camera Space for Efficient Spherical Stereo Matching

Donghun Kang^{†*} Hyeonjoong Jang[†] Jungeon Lee[†] Chong-Min Kyung[†] Min H. Kim[†]

[†]KAIST

*Hyundai Motor Company

Abstract

Omnidirectional cameras have been used widely to better understand surrounding environments. They are often configured as stereo to estimate depth. However, due to the optics of the fisheye lens, conventional epipolar geometry is inapplicable directly to omnidirectional camera images. Intermediate formats of omnidirectional images, such as equirectangular images, have been used. However, stereo matching performance on these image formats has been lower than the conventional stereo due to severe image distortion near pole regions. In this paper, to address the distortion problem of omnidirectional images, we devise a novel subdivision scheme of a spherical geodesic grid. This enables more isotropic patch sampling of spherical image information in the omnidirectional camera space. By extending the existing equal-arc scheme, our spherical geodesic grid is tessellated with an equal-epiline subdivision scheme, making the cell sizes and in-between distances as uniform as possible, i.e., the arc length of the spherical grid cell's edges is well regularized. Also, our uniformly tessellated coordinates in a 2D image can be transformed into spherical coordinates via one-to-one mapping, allowing for analytical forward/backward transformation. Our uniform tessellation scheme achieves a higher accuracy of stereo matching than the traditional cylindrical and cubemap-based approaches, reducing the memory footage required for stereo matching by 20%.

1. Introduction

Fisheye lenses have been used to capture a wide field of view (FOV) in even more than 180° . Stereo matching from fisheye images is an appealing combination for a better and efficient understanding of scene geometry. However, owing to fisheye lenses' optical characteristics, the traditional stereo matching algorithm based on the pinhole camera model is invalid. For instance, the outermost edges of fisheye images are severely distorted as the angle of view increases widely. It disturbs accurate depth estimation from spherical stereo images. To enable epipolar geometry on fisheye images, projective im-

age transformation, such as equirectangular projection and latitude-longitude (LL) transformation, has been utilized to estimate depth information from very wide-angle fisheye images. This additional transformation of the input images warps images to increase the image size redundantly. Thus, the computational cost and memory increase accordingly to process spherical stereo matching from fisheye images.

To estimate depth information from a pair of fisheye images, fisheye images of an ultra-wide field of view need to be warped and transformed into a certain type of projective image models, such as equirectangular projection and latitude-longitude (LL) transformation [15]. We can apply the conventional stereo matching algorithm that searches stereo correspondence along epipolar lines through epipolar geometry after rectifying the transformed images. Due to the excessive distortion near the horizontal poles along the stereo baseline, the accuracy of stereo matching degrades, especially around the distorted area. In addition, excessive warping of the near pole regions in the LL-transformed images increases the memory footage.

In this paper, we propose a novel subdivision scheme of spherical image information in the omnidirectional camera space. Our spherical geodesic grid enables equal-epiline subdivision to achieve higher stereo matching accuracy, where the cell sizes and in-between distances are uniformly regularized. Our novel transformation from sphere geometry to a 2D tessellated array enables us to apply epipolar geometry to directly achieve fast, spherical stereo matching from the tessellated fisheye images. Our geodesic grid-based stereo matching runs directly on the uniformly tessellated stereo images so that our method can reduce the required memory footage by 20% without compromising depth accuracy.

2. Related Work

Omnidirectional Camera Model. The pinhole camera model is widely used with epipolar geometry for estimating depth from stereo images. The FOV of the camera is smaller than 180° [9]. However, when the field of view of the lens increases more than 120° , the pinhole camera model is invalid

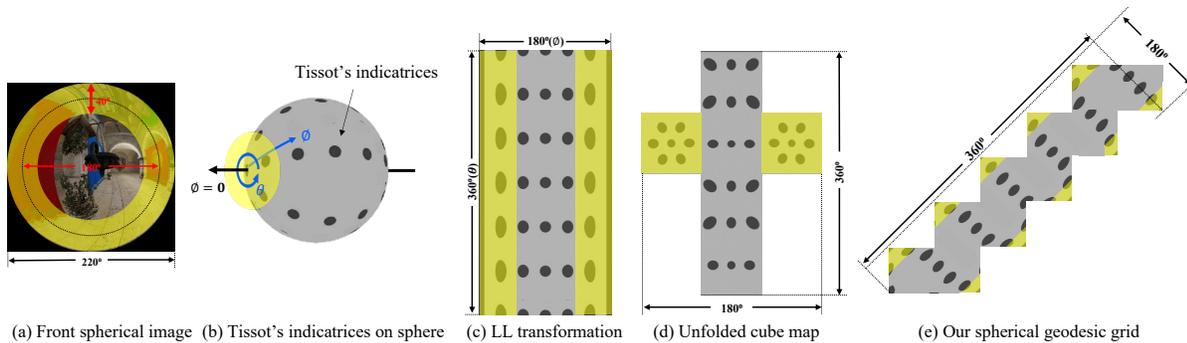


Figure 1. Geometric transformation options of a fisheye image. (a) Fisheye image. For all geometric domains with Tissot's indicatrices that indicate the same area size of patches: (b) the spherical domain in 3D, (c) LL-transformed 2D image (equivalent to the rotated equirectangular image), (d) unfolded cube map, and (e) our proposed spherical geodesic grid. Our geometric transform provides a more regularized transform of Tissot's indicatrices. Note that the regions marked with yellow color indicate the extreme image distortion near poles.

anymore, in particular, more than 180° . The omnidirectional camera model includes a virtual 3D spherical image by projecting points incident from omnidirectional surroundings on the surface of a sphere, whose focal length is the radius of the sphere as the origin of the sphere. In addition, the recently proposed calibration methods, where a 3D spherical image is obtained from a 2D fisheye image, provide a sufficient basis to implement a spherical stereo algorithm using a real fisheye lens [3, 11, 23, 24]. Figures 1a and 1b compare a fisheye and the spherical image coordinates.

Latitude-Longitude/Equirectangular Projection. Stereo matching cannot be processed directly on a 3D stereo spherical image generated from a spherical camera model. Therefore, we use either perspective projection or cylindrical projection that converts a 3D spherical image into a 2D planar image in order to apply a block-matching stereo algorithm to the spherical stereo image directly [1]. The stereo algorithm that applies perspective projection based on the pinhole camera model loses the advantage of using a wide FOV, so it is not suitable for spherical stereo.

A binocular spherical stereo algorithm with various configurations has been proposed by applying the cylindrical projection method, which is also known as Mercator projection, used to represent the earth as a flat map.

The bipolar geometry in cylindrical projection is studied by an omnidirectional stereoscopic study of the horizontal [10] and vertical [4] alignment, and particularly it is shown that vertically aligned binocular stereo formed linear epipolar lines in the vertical direction, such as in the conventional planar stereo. The latitude-longitude transformation has proposed a stereo matching algorithm according to the linear epipolar geometry for a FOV of more than a hemisphere (Figure 1c) lying down vertical-aligned cylindrical projection transversely [14, 15, 17]. This implies that horizontal and vertical omnidirectional stereo image can be obtained with a low-cost stereo matching based algorithm with a system composed of two spherical stereo cameras [16, 18].

Figure 1c shows Tissot's indicatrices on different sphere projections to evaluate distortion in the 2D projection of

the sphere. Tissot's indicator in the LL transformed image, which should be close to the circle, is excessively distorted around the 180° of the LL transformed image, which is near the horizontal poles of the 3D spherical image. Obviously, this is critical for the accuracy of stereo matching from fish-eye images.

Cubemap Projection. Since cubemap projection (CMP) has been widely used to project omnidirectional environment into the 2D domain [5]. An enclosed 3D spherical image is projected onto six faces of a surrounding cube. As compared with the cylindrical projection, cubemap projection reduces the redundancy of the distortion near the pole regions, which significantly reduces the memory consumption and computational complexity. Therefore, cubemap projection is widely used in 3D graphical rendering, and omnidirectional video coding for VR [2, 8]. However, due to the following reasons, cubemap projection cannot be used in direct stereo matching. First, as a 3D spherical image is projected onto six divided planes, CMP performs complex computations referencing 3D coordinates for forward projection. Due to this referencing, CMP cannot be applied in the direct stereo matching, where 2D transformed image is used. Second, although CMP reduces the excessive distortion around the pole regions, it induces radial distortion that occurs outward from the center of each plane. This radial distortion is not parallel to the epipolar lines, reducing the matching accuracy. Third, the epipolar lines are formed radially on the two sides, including the horizontal poles. As a square matching block cannot represent the same sized area in the 3D spherical image for radial directions, the depth accuracy in two faces, including poles, is severely degraded (Figure 1d).

Geodesic Grid. Since Buckminster Fuller [20] introduced a geodesic dome composed of planar triangles, the geodesic grids have been widely used in fields such as geography and meteorology for modeling Earth's surface. The geodesic grid has been widely used with various methods of subdivision of icosahedron, for instance, the mid-arc and equal-arc subdivision methods of Class 1, depending on the conve-

nience of the configuration [21]. Compared to a cylindrical projection-based grid including the equirectangular and the LL, a 3D spherical geodesic grid becomes popular with the introduction of data structure using five sets of four triangles [22]. A geodesic grid has been used for feature detection and descriptor matching on a spherical image by applying equal-arc [25] and mid-arc [6, 7] subdivision. It is also shown that it is suitable for spherical SLAM based on the geodesic grid-based feature [12, 13] recently.

Existing 3D geodesic grids using several known subdivision methods cannot be directly converted into a 2D image processing plane. Due to this limitation, the geodesic grid cannot be used in direct stereo matching, but it was only used for feature detection and descriptor studies, which do not follow epipolar geometry. Figure 1e shows the distorted geodesic grid, which presents more uniform Tissot’s indicatrices. Our work is grounded on the geodesic grid.

3. Equal-epiline Subdivision on Geodesic Grid

To reduce distortion that can cause inaccuracies in spherical stereo matching, we begin with the geodesic grid, which represents the sphere most uniformly. When projecting a 3D spherical image into a 2D image processing plane using a geodesic grid, the sub-division methods do not have a mathematically defined transformation relationship in linear equation, which implies that the epipolar geometry cannot be defined in the 2D image processing plane like the LL-based spherical stereo method. The 3D spherical image coordinates must be referred to every pixel to perform stereo matching in the 2D projected images along an epipolar line represented by a great circle in a 3D spherical image.

3.1. Icosahedral Geodesic Polyhedron

A regular polyhedron can be used for a tessellation of the sphere, in which the surface is divided by the great arcs. Only the tetrahedron, octahedron, and icosahedron consisting of equilateral triangles are geodesic polyhedrons for spherical subdivision. The icosahedron with 20 faces, 30 edges, and 12 vertices has the highest number of identical regular polygons of any polyhedron. Since a higher number of faces can reduce the irregularity of a geodesic polyhedron when it is subdivided, the icosahedron is the most popular reference polyhedron for the spherical subdivision. Therefore, to achieve uniform subdivision, we utilize icosahedral geodesic polyhedron for the geodesic grid.

3.2. Geodesic Grid Subdivision

Each vertex from a subdivided spherical geodesic polyhedron is used as the center of pixels for image pixel representation of a spherical image, i.e., a pentagonal-hexagonal geodesic grid is obtained by taking the dual polyhedron of an icosahedral geodesic polyhedron. There are 12 reference vertices (A, B, C, D, E, F, ...) for the pentagonal grid and

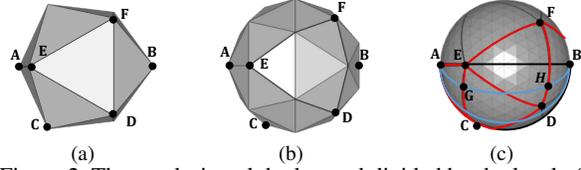


Figure 2. The geodesic polyhedron subdivided by the level of subdivision (n_d) (a) The icosahedron ($n_d = 1$). (b) Subdivision with $n_d = 2$. (c) Subdivision with $n_d = 8$.

complement vertices of a pentagonal vertex for hexagonal grid in an icosahedral geodesic grid (Figure 2b). To achieve a finer resolution of the geodesic grid, the principal polyhedral triangles (PPT) are subdivided, so the uniformity of geodesic grid cells will be increased when it is subdivided with a higher level of subdivision (n_d). Figure 2 shows the transverse of icosahedron and geodesic polyhedrons by the level of subdivision ($n_d = 1, 2, 8$). The data structure for image processing of the spherical geodesic is well defined in the previous research works [22, 25]. Separated five sets of four regular spherical PPTs are projected onto the planar geodesic grid consisting of regular triangle cells, and grid cells with unit vectors of the axes ($\mathbf{v}_1 = [1, 0]^T$ and $\mathbf{v}_2 = [\frac{1}{2}, \frac{\sqrt{3}}{2}]$) skewed to make axis vectors ($\mathbf{v}'_1 = [1, 0]$ and $\mathbf{v}'_2 = [0, 1]$) orthogonal (Figure 3). Here we only consider the skewed geodesic grid and refer to it as the *geodesic grid*. Refer to the supplemental document for comparison of existing subdivision schemes.

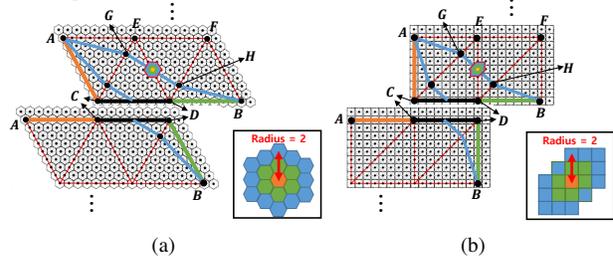


Figure 3. Illustration of the transformed geodesic grid. (a) The geodesic grid is separated into five sets of four triangles (PPTs). Two of the five sets are shown in the figure. (b) Two skewed partial geodesic grids for computational implementation. The shape of the matching block is also changed. The blue lines are the example of epipolar lines implemented by our method.

3.3. Sphere Projection Model

We propose a new subdivision scheme of the spherical geodesic grid that provides the transformation, which makes it possible to apply a stereo algorithm directly without any a look-up-table (LUT). The proposed scheme can be used for the application, which is related to a spherical image, which has a limitation of the low-cost implementation. The general function of transformation, which projects a spherical grid to a planar grid, is defined as¹ $\mathcal{F} : \mathbf{P}(\theta, \phi) \rightarrow \mathbf{p}(x, y)$.

¹We denote bold lower-case letters \mathbf{p} be vectors and light lower-case letters p be scalars. For the 3D coordinate system, the upper-case letter \mathbf{P} is used, and the projected 2D coordinates from the 3D coordinates are

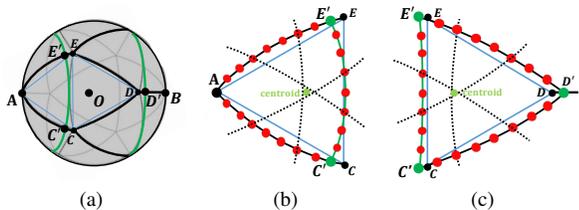


Figure 4. (a) To make epipolar subdivision proportional to the latitude angle ϕ , we first define the latitude arcs (green circle) with ϕ of 60° and 120° . (b) Subdivision scheme when $\phi < 60^\circ$ and $\phi > 120^\circ$. (c) Subdivision scheme when $60^\circ < \phi < 120^\circ$.

There are two important functional features of the image transformation for the low-cost spherical stereo algorithm. First, to generate a block for stereo matching, it is necessary to obtain a position of local neighboring pixels relative to the center pixel of the block on the 2D transformed image plane. Second, the absolute position on the spherical image of each pixel must be estimated on the transformed image. Conventional approaches assume that image distortion caused by transformation is small enough in the local block matching process. Thus neighboring pixels can be used for feature matching. However, There was no way to express the 3D geodesic grid into the 2D geodesic grid directly, the pre-calculated LUT based forward and backward projection is repeatedly performed. In other words, \mathcal{F} and \mathcal{F}^{-1} are not defined mathematically, but are defined as a form of LUT by corresponding relative position of respectively calculated spherical and projected geodesic grid. Because of the non-linearity of conventional methods, including equal-chord, mid-arc, and equal-arc, a set of analytic computations for stereo matching are required as follows: (1) Backprojection: defining the reference points (RPs) by subdividing PPT's three arcs in 3D and then defining three great circles (GCs) of three pairs of RPs to find three intersection points (IPs) of three GCs, resulting in the centroid of three IPs, (2) Shifting disparity candidates with an interval, (3) Forward projection: Defining the search range of the corresponding pixel and backprojecting of candidate pixels in the search range to find out the nearest points by computing distance candidates in 3D, and (4) Matching cost: computing the matching costs by searching the disparity candidates within the potential disparity range by repeating (1)–(3) for every pixel. In addition, considering that additional interpolation is required in the projection operation, it requires a larger memory footprint and computation cost. Therefore, the geodesic grid cannot be used in spherical stereo matching algorithm for low-cost implementation.

To solve this problem, we propose the new subdivision scheme that take the advantages of the uniform equal-arc geodesic grid on a spherical surface and equal-chord geodesic grid on a planar plane simultaneously, as opposed to the conventional LUT-based method, which refers to the origin of the sphere. First of all, the spherical geodesic grid

expressed as lower-case letter \mathbf{p} .

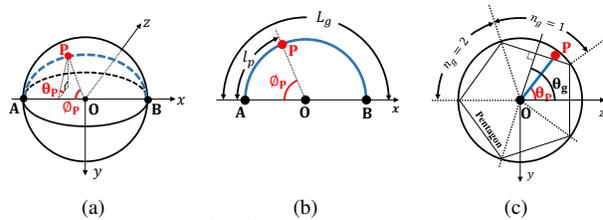


Figure 5. (a) Point $\mathbf{p}(\phi, \theta)$ on the 3D sphere with the epipolar line (blue line). (b) Intersection of the epipolar plane including ϕ_P . $\phi_P = l_P / L_g \cdot 180^\circ$. (c) Intersection of the $y-z$ plane including θ_P .

is transversely rotated to make the direction of two antipodal poles from north-south to west-east similar to the LL transformation (Figure 2). When the spherical geodesic grid is subdivided with an equal-arc method, it has nearly uniform cell size and in-between distance, i.e., the arc length of the spherical grid cell's edges is uniform.

To divide epipolar lines equally on the transformed geodesic grid, in our subdivision scheme, we change 10 arcs of the PPT along with the circles of $\phi = 60^\circ$ and 120° from great circles (Figure 4a). The points A and B indicate north/south poles of the 90° -rotated spherical coordinate system, and the black arcs lines are part of the great circles. The green circles indicate latitude angles at 60° and 120° , where E' and C' are intersection points between the 60° latitude and the great circle, D' is the intersection between the 120° latitude and the great circle. The blue triangle presents the icosahedron planar triangle. By changing the great circles into the latitude line, the equal-epiline subdivision method divides the epipolar line uniformly so that each subdivided point can linearly correspond to the 2D transformed geodesic grid composed of regular triangles.

20 PPTs are made by reference points (including C' , D' & E') with an offset from the green circles. Then, we utilize the equal-arc subdivision scheme which takes subdivided points with centroid points (Figures 4b and 4c). On the 2D transformed grid, the length ratio of lines along the epipolar line is corresponding to the length of arcs on the 3D spherical geodesic grid. We call this as the *equal-epiline subdivision*. By using the proposed geodesic grid based on the equal-epiline scheme, transformation function (\mathcal{F}_{gg}) uniformly subdivide the spherical image by the advantage of the equal-arc, and also minimize the distortion simultaneously, which was not possible in previous research.

This is appropriate for the low-cost stereo matching algorithm since we do not need repetitive forward (\mathcal{F}) and backward projection (\mathcal{F}^{-1}) operation necessary to specify coordinates in the spherical geodesic grid and the pre-calculated LUT consuming large memory.

3.4. Derivation of the Transformation Function

We propose a novel image transformation method, where the stereo block matching algorithm is performed on the 2D images directly, similar to the LL method. A point \mathbf{P} on the unit spherical geodesic grid is expressed in Figure 5a. We

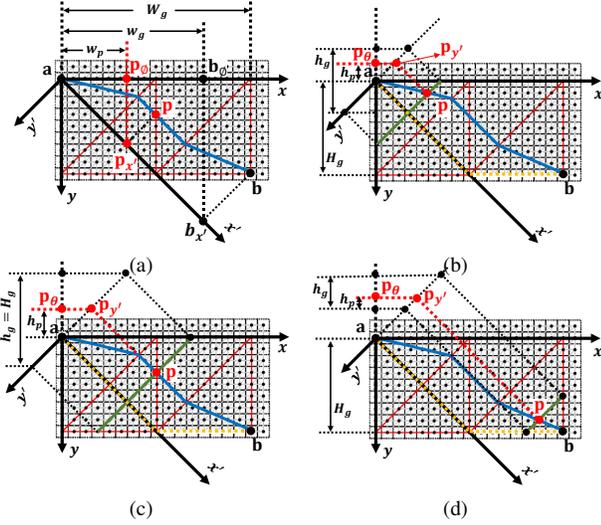


Figure 6. Transformation of the point $\mathbf{P}(\phi, \theta)$ in the 2D geodesic grid with the epipolar line (blue line). (a) x' -axis is used to calculate the ratio of w_g and w_p as $\phi_{\mathbf{P}}(^{\circ})$, and $\theta_{\mathbf{P}}$ is defined on the geodesic grid with the ratio of h_p and h_g for (b) $0^{\circ} \leq \phi_{\mathbf{P}}(^{\circ}) < 60^{\circ}$, (c) $60^{\circ} \leq \phi_{\mathbf{P}}(^{\circ}) < 120^{\circ}$, and (d) $120^{\circ} \leq \phi_{\mathbf{P}}(^{\circ}) < 180^{\circ}$.

can define $\phi_{\mathbf{P}}$ and $\theta_{\mathbf{P}}$ by intersections including a point \mathbf{P} and the origin \mathbf{O} . In Figure 5b, $\phi_{\mathbf{P}}$ is defined by the ratio of the arc length between the point $\mathbf{P}(l_{\mathbf{P}})$ and the length of the epipolar line (L_g).

Because x - y orthogonal coordinate system of the geodesic grid is rotated 45-degree compared to the, $\phi_{\mathbf{P}}$, $\theta_{\mathbf{P}}$ coordinates, it is impossible to calculate $\phi_{\mathbf{P}}$ and $\theta_{\mathbf{P}}$ with simple linear equation. We perform the derivation of $\phi_{\mathbf{P}}$ and $\theta_{\mathbf{P}}$ on the virtual x' - y' axis. The axes rotated by 45°, x' and y' , represent ϕ and θ , respectively. The projected points \mathbf{p}_{ϕ} and \mathbf{p}_{θ} from the point \mathbf{P} are defined as follows:

$$\mathbf{p}(x, y) \rightarrow \mathbf{p}_{\phi}(0.5x + 0.5y, 0), \quad (1)$$

$$\mathbf{p}(x, y) \rightarrow \mathbf{p}_{\theta}(0, -0.5x + 0.5y). \quad (2)$$

We proceed with the derivation of $\phi_{\mathbf{P}}$. The intersection of the sphere is shown in Figure 5b. As it is shown, $\phi_{\mathbf{P}}$ is defined as the ratio of the length between the line connecting pole a and the point \mathbf{p} along the epipolar line and epipolar line ($\overline{\mathbf{a}\mathbf{b}_{x'}}$). Although the epipolar line which expresses 180 degrees on the geodesic grid (shown as the blue line) has different length, the epipolar lines projected on the x' -axis have the same length of $\overline{\mathbf{a}\mathbf{b}_{x'}}$. Therefore, the ratio for $\phi_{\mathbf{P}}$ can be calculated as the length ratio of $\overline{\mathbf{a}\mathbf{p}_{x'}}$ and $\overline{\mathbf{a}\mathbf{b}_{x'}}$. Finally, as shown in Figure 6a, by taking w_g and $w_p (= 0.75W_g)$ which are x -coordinates of points (\mathbf{b}_{ϕ} , \mathbf{p}_{ϕ}) re-projected to x -axis, ϕ of \mathbf{P} can be defined as,

$$\phi_{\mathbf{P}}(^{\circ}) = \frac{w_p}{w_g} \cdot 180^{\circ} = \frac{120(x+y)}{W_g}, \quad (3)$$

where W_g is the width of the partial geodesic grid in pixel.

Next, we describe the derivation of $\theta_{\mathbf{P}}$. The projection of the point \mathbf{p} onto the y' -axis is used to calculate the ratio of

length between h_p and h_g . Because the sphere's projected geodesic grid is separated into five partial geodesic grids, we have to define reference angles for each partial geodesic grid. The reference angle of a partial geodesic grid is $\theta_g(^{\circ}) = 72^{\circ} \cdot n_g$, when n_g is defined sequentially as a number of partial grids (0 – 4) (Figure 5c). The length of the projected line, h_p , represents the angle between the reference angle line (yellow line) and the point \mathbf{p} .

Furthermore, careful observation of the y' -axis of the partial geodesic grid shows that the resolution of the height for $\theta_{\mathbf{P}}$ representation is proportional to $\phi_{\mathbf{P}}$. Therefore, the length of h_g of the 72° line on the geodesic grid (shown in green) is defined as $h_g = \frac{H_g \cdot \phi_h}{60}$, when $\phi_h(^{\circ})$ is $\phi_{\mathbf{P}}$ (for $0^{\circ} \leq \phi_{\mathbf{P}} < 60^{\circ}$) or 60 (for $60^{\circ} \leq \phi_{\mathbf{P}} < 120^{\circ}$) or $180^{\circ} - \phi_{\mathbf{P}}$ (for $120^{\circ} \leq \phi_{\mathbf{P}} \leq 180^{\circ}$).

As shown in Figures 6b, 6c and 6d, $\theta_{\mathbf{P}}$ is formulated by using h_p and h_g as follows:

$$\theta_{\mathbf{P}}(^{\circ}) = \theta_g + \frac{\vec{h}_p}{h_g} \cdot 72^{\circ}. \quad (4)$$

Combining the definitions of h_g and \vec{h}_p with Equation (4) yields a simple linear equation which expresses $\theta_{\mathbf{P}}$ in terms of $\phi_{\mathbf{P}}(^{\circ})$:

$$\frac{\vec{h}_p}{h_g} = \begin{cases} \frac{(-x+y)}{2(x+y)} & \text{if } 0^{\circ} \leq \phi_{\mathbf{P}}(^{\circ}) < 60^{\circ} \\ \frac{(-x+y)}{2H_g} & \text{if } 60^{\circ} \leq \phi_{\mathbf{P}}(^{\circ}) < 120^{\circ} \\ \frac{y-H_g}{3H_g-(x+y)} & \text{if } 120^{\circ} \leq \phi_{\mathbf{P}}(^{\circ}) < 180^{\circ} \end{cases}, \quad (5)$$

where H_g is the height of the partial geodesic grid in pixel.

4. Spherical Stereo Matching

We use the sum of absolute differences (SAD) as a cost function of correspondence for block stereo matching. The cost function C by angle disparity ϕ_d in 3D spherical is defined as follows:

$$C(\phi, \theta, \phi_d) = \sum_{(\phi_{\mathbf{P}}, \theta_{\mathbf{P}}) \in \mathcal{N}} |I_L(\phi_{\mathbf{P}}, \theta_{\mathbf{P}}) - I_R(\phi_{\mathbf{P}} - \phi_d, \theta_{\mathbf{P}})|, \quad (6)$$

where \mathcal{N} is the matching block centered at $\mathbf{P}(\phi, \theta)$.

For spherical stereo matching on the 2D projected plane, we need to define a cost function with $\mathbf{p}(x, y)$. Because the $\theta_{\mathbf{P}}$ is the same in the epipolar line, we can define cost function c as follows:

$$C(x, y, \phi_d) = \sum_{(i,j) \in \mathcal{N}} |I_L(i, j) - I_R(i', j')|, \quad (7)$$

where (i', j') is the position of pixels in moved matching block by the angle disparity ϕ_d centered at $\mathbf{p}'(x', y')$.

Based on the proposed projection model, we implement the spherical stereo algorithm (Figure 7). By the definition

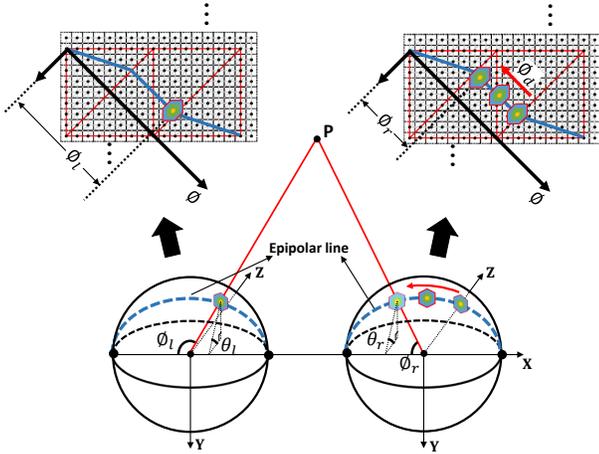


Figure 7. Illustration of the proposed spherical stereo using a geodesic grid. Along the epipolar line (blue line), we calculate the cost of each matching blocks (Equation (7)). For the position of the minimum cost block, we have ϕ and θ on the left and right geodesic grid (Equations (3) and (4)). Using Equation (8), the stereo matching algorithm can be implemented on the projected geodesic grid in terms of x and y without referring to the 3D spherical geodesic grid.

of the proposed projection model, an epipolar line is represented by three connected straight lines. Using the conditions of linearly expressed epipolar line and ϕ_P , the position of moved center pixel position $\mathbf{p}'(x', y')$ of matching block π' by ϕ_P can be expressed as follows:

$$\mathbf{p}'_{\phi'_d}(x', y') = \begin{cases} \left(\frac{xk}{x+y}, \frac{yk}{x+y} \right) & \text{if } 0^\circ \leq \phi_P(^\circ) < 60^\circ \\ \left(\frac{x-y+k}{x}, \frac{-x+y+k}{2} \right) & \text{if } 60^\circ \leq \phi_P(^\circ) < 120^\circ \\ \left(\frac{k-mx+y}{m+1}, \frac{-mk-mx+y}{m-1} \right) & \text{if } 120^\circ \leq \phi_P(^\circ) < 180^\circ \end{cases}, \quad (8)$$

where $\phi'_d = \phi_l - \phi_d$ and $k = \frac{H_g \phi'_d}{60}$ and $m = \frac{H_g - y}{2H_g - x}$.

The shape of the hexagonal block for stereo matching is changed when the geodesic grid is skewed (Figure 3). For radius r , the number of pixels in matching block can be defined as: $N_r = 3r^2 + 3r + 1$. The hexagonal-shaped matching block is appropriate for the stereo matching because a pixel in a hexagonal block only meets with the wall's neighboring pixels, i.e., no directional dependency than rectangular block shape, which has eight neighboring pixels with four walls and 4 points. Moreover, it has 20% smaller block size compared with rectangular blocks used in the LL, which eliminates the redundancy problem, reducing the memory consumption and the computational complexity of the algorithm. Once the cost-volume is generated in accordance with Equations (7) and (8), the matched block is decided in the winner-takes-all manner. According to the sine theorem, the distance of the point \mathbf{P} with $P_l(\theta_l, \phi_l)$ and matched point $P_r(\theta_r, \phi_r)$ of the left and right normalized spherical images can be expressed as follows: $d_l = b \frac{\sin \phi_r}{\sin(\phi_l - \phi_r)}$, $d_r = b \frac{\sin(\pi - \phi_l)}{\sin(\phi_l - \phi_r)}$, where b is the baseline of spherical stereo camera.

5. Results

We validate our method quantitatively and qualitatively with a synthetic rendering dataset and a real image dataset.

Datasets. For scientific evaluation of the proposed method, we created and made use of a synthetic 360-degree image dataset. Synthetic datasets with ground truth depth were used to generate spherical stereo images. The fisheye stereo datasets *San-miguel*, *Sponza*, and *Bedroom* were rendered with a virtual camera of 220° FOV [19]. See Table 1 for more details. In the dataset, three different scenes with different positions and illumination conditions were generated by rendering 3D scene models with a synthetic stereo camera. Each dataset includes omnidirectional stereo images of color and depth (Table 1). Real datasets were captured by our custom-built stereo camera system with two commercial 360 cameras (Ricoh Theta) calibrated using [24].

Table 1. Environmental conditions of synthetic datasets

	Place	Distance	Light source
San Miguel	Outdoor	~20m	Sunlight
Sponza	Outdoor	~50m	Sunlight
Bedroom	Indoor	~4m	Point

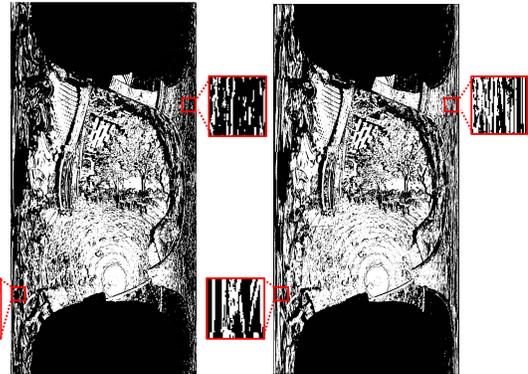


Figure 8. The disparity confidence map of *Sponza 1*. Left (LL rectification) and right (ours) map shows valid/error pixels in binary color (white/black). In the near pole region, our method provides more valid pixels significantly.

Experiments. The accuracy of the proposed spherical stereo algorithm based on the equal-epiline subdivision scheme is compared with the LL-based and the CMP-based stereo matching algorithm in terms of the error-rate and the RMSE.

For fairness in experimental comparison, the geodesic grid resolution and matching block size are set to keep the ratios of the vertical image resolution and matching block radius equal, as shown in Figures 1 and 3b. The optimal parameters of the LL-based and our methods were determined with consideration of resolution and computational cost. Both methods have a similar matching block size with a radius (r): $(4r^2 + 4r + 1)$ for LL and $(3r^2 + 3r + 1)$ for ours. Also, both methods have the same vertical image resolution: (1280×2560) for LL and (1024×2560) for ours. Note that our method actually uses a 25% less block size and a 20%

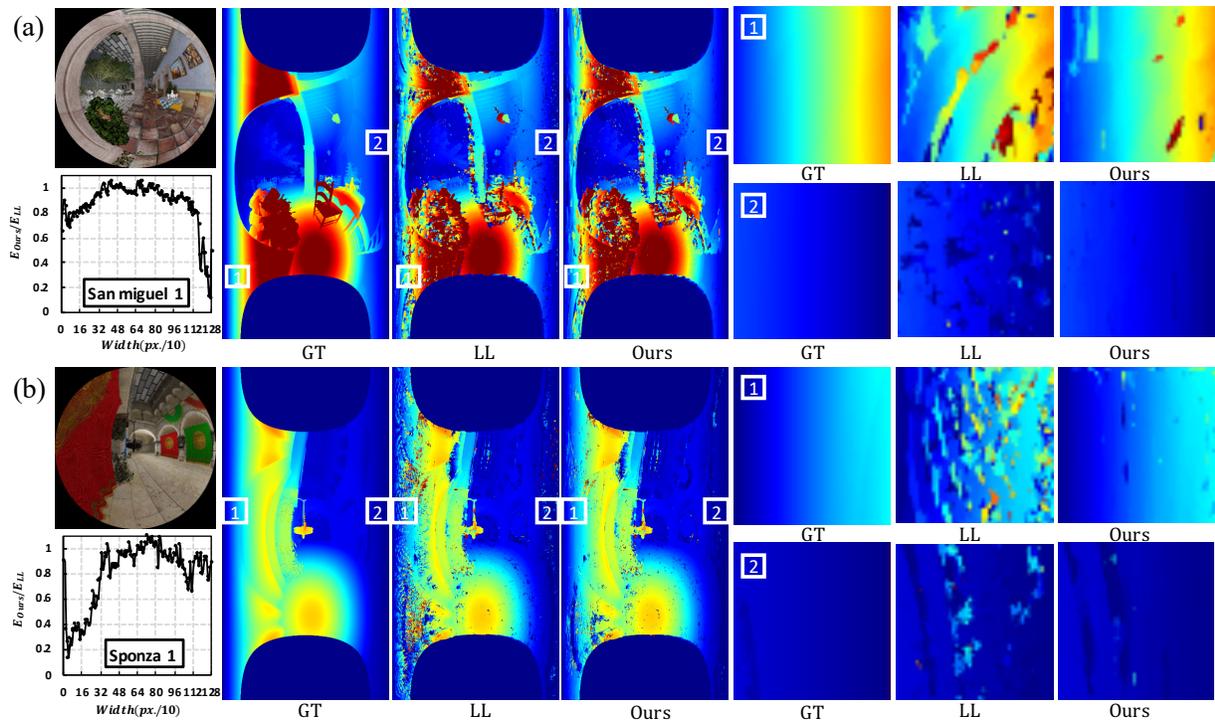


Figure 9. Results of synthetic datasets. Left-top is the left image of fisheye stereo; left-bottom is the ratio of error pixels by 10-pixel-step in x-axis; GT disparity map; LL-based disparity map; Our disparity map; Comparison of disparity results at two identical positions. Refer to the supplemental for more results.

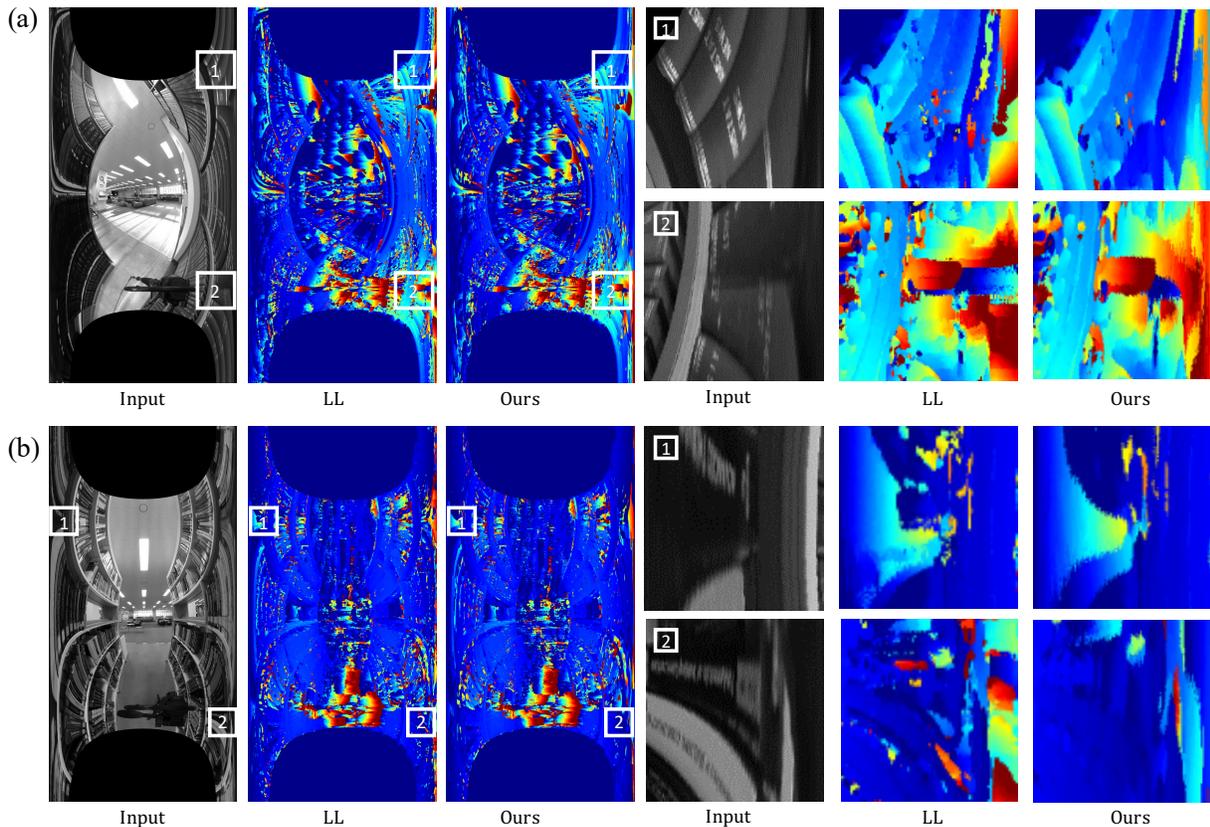


Figure 10. Results of real datasets captured by our 360 stereo camera setup with 7.5cm baseline. Leftmost column shows one of input stereo images, which is followed by LL-based disparity map and ours. Right-half columns compare disparity maps in closeups.

Table 2. Quantitative comparison with the LL- and CMP-based stereo matching. For the given 1, 3, and 5 of δ_d , the proposed method shows improved depth accuracy in terms of the error rate and RMSE for all datasets. Red numbers mean the best.

	<1		<3		<5		RMSE					
	LL	CMP	Ours	LL	CMP	Ours	LL	CMP	Ours			
San miguel 1	44.89	41.96	38.52	30.53	28.66	26.07	26.14	24.32	22.99	31.21	31.11	30.89
San miguel 2	44.99	42.01	39.46	35.91	33.13	31.51	31.47	30.21	27.88	29.71	29.55	28.45
Sponza 1	33.79	29.76	22.37	26.28	22.89	16.31	22.65	19.98	13.20	19.26	17.43	13.57
Sponza 2	39.41	38.15	35.71	29.69	28.03	26.32	25.26	23.33	22.30	30.51	30.15	29.36
Bedroom 1	53.41	52.41	48.34	41.11	39.56	37.53	36.52	34.42	33.91	39.13	38.88	37.79
Bedroom 2	51.51	46.41	38.99	38.79	37.65	35.35	34.21	32.85	31.25	39.96	37.77	28.66

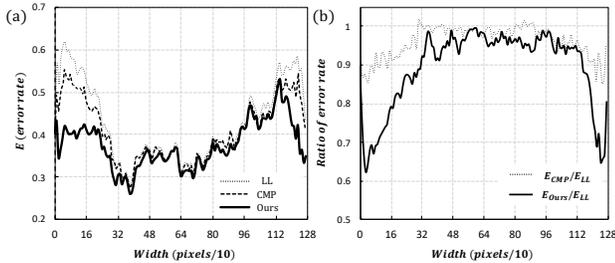


Figure 11. (a) The average of the error rates for six datasets. (b) The error rate ratio between the LL-based method and the proposed method. Results are sampled by 10 pixels along the x -axis to handle the outliers in results.

less vertical resolution of the input image than the LL-based method.

Figure 8 shows the disparity confidence map, which is the expression of valid/error pixels in binary color. An error pixel is defined if the disparity difference of the pixel between the result disparity map and the ground truth is larger than δ_d , and vice versa for a valid pixel. For the given threshold of error ($\delta_d = 1$), we discriminate valid/error pixels by comparing disparity with the ground truth. As shown in Figure 8, error pixels (black-colored) are remarkably found in the near pole region. The quantitative comparison is performed in terms of error rate and root-mean-squared error (RMSE) of the disparity. The percentage of error pixels (error rate, E) is defined as $E = \frac{1}{N} \sum_{(x,y)} (|d(x,y) - d_{GT}(x,y)| < \delta_d)$, where N is the total number of pixels.

Table 2 is the table of the quantitative results for six datasets. For all datasets, we have improved results in terms of the error rate and RMSE on the full image. The proposed method targets the pole problem, which is related to the degradation of depth accuracy. The average error rate of six datasets by the x -axis position ϕ is expressed in Figure 11(a). The result is sampled by 10 pixels in horizontal resolution. As can be observed in the figure, as farther from the center, results of the LL- and CMP-based methods show a larger error rate than the proposed method. To evaluate the number of error pixels according to the ϕ , we plot the ratio of error rate (Figure 11b). For the position of the x -axis sampled by 10 pixels, especially on the near pole region, we have significantly improved results in error rate up to 40%. In the center region, which means the near pole region’s complement region, the value of E is almost the same.

Figure 9 qualitatively compares representative results of our method with those of the LL-based stereo matching. Compared with the ground truth maps, our method outperforms the LL-based stereo matching method particularly near pole regions at the leftmost and rightmost ends, showing less disparity errors and noise.

Figure 10 compares disparity maps computed by the LL-based matching and our method from stereo images captured by a real camera setup (as shown in Figure 12). Near pole regions, our method can provide more stable depth estimation than the LL-based approach. Note that we are using the 20% smaller size of the transformation resolution and 25% smaller matching block; good or bad indicators of a very small difference in the center region is acceptable.



Figure 12. Real spherical stereo camera setup.

6. Conclusion

We have proposed a uniformly tessellated geodesic grid for spherical stereo. Our geodesic grid achieves more accurate disparity maps near the pole regions in spherical stereo than the conventional LL-based rectification. Our equal-epiline subdivision scheme can be used for low-cost direct block matching stereo. Our direct stereo matching method just requires a linear shift after only one projection of fisheye images using a linearly defined projection function without requiring any LUT and forward/backward projection. While our method presents improved disparity in terms of the error rate and RMSE in a full 360 image, including the near pole region, it can be performed with a 20% smaller size of image transformation resolution and a 25% smaller matching block size.

Acknowledgments

The authors acknowledge Korea NRF grants (CISS-2013M3A6A6073718), MSIT/IITP of Korea (2017-0-00072), National Research Institute of Cultural Heritage of Korea, Samsung Electronics, and MSRA, with partial support from the Samsung Research Funding Center (SRFC-IT2001-04) for developing partial 3D imaging algorithms.

References

- [1] Steffen Abraham and Wolfgang Förstner. Fish-eye-stereo calibration and epipolar rectification. *ISPRS Journal of photogrammetry and remote sensing*, 59(5):278–288, 2005. 2
- [2] Zhenzhong Chen, Yiming Li, and Yingxue Zhang. Recent advances in omnidirectional video coding for virtual reality: Projection and evaluation. *Signal Processing*, 146:66–78, 2018. 2
- [3] Christopher Geyer and Kostas Daniilidis. A unifying theory for central panoramic systems and practical implications. In *European conference on computer vision*, pages 445–461. Springer, 2000. 2
- [4] Joshua Gluckman, Shree K Nayar, and Keith J Thoresz. Real-time omnidirectional and panoramic stereo. In *Proc. of Image Understanding Workshop*, volume 1, pages 299–303. Citeseer, 1998. 2
- [5] Ned Greene. Environment mapping and other applications of world projections. *IEEE Computer Graphics and Applications*, 6(11):21–29, 1986. 2
- [6] Hao Guan and William AP Smith. BRISKS: Binary features for spherical images on a geodesic grid. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4516–4524, 2017. 3
- [7] Hao Guan, William AP Smith, and Peng Ren. Corner detection in spherical images via the accelerated segment test on a geodesic grid. In *International Symposium on Visual Computing*, pages 407–415. Springer, 2013. 3
- [8] Philippe Hanhart, Xiaoyu Xiu, Yuwen He, and Yan Ye. 360° video coding based on projection format adaptation and spherical neighboring relationship. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 9(1):71–83, 2018. 2
- [9] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 1
- [10] Hiroshi Ishiguro, Masashi Yamamoto, and Saburo Tsuji. Omni-directional stereo. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (2):257–262, 1992. 2
- [11] Juho Kannala and Sami S Brandt. A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. *IEEE transactions on pattern analysis and machine intelligence*, 28(8):1335–1340, 2006. 2
- [12] Wonjin Kim and Seong-Won Lee. Stereo matching validation using spherical image mapping. In *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 1127–1130. IEEE, 2018. 3
- [13] Jianfeng Li, Xiaowei Wang, and Shigang Li. Spherical-model-based SLAM on full-view images for indoor environments. *Applied Sciences*, 8(11):2268, 2018. 3
- [14] Shigang Li. Real-time spherical stereo. In *18th International Conference on Pattern Recognition (ICPR’06)*, volume 3, pages 1046–1049. IEEE, 2006. 2
- [15] Shigang Li. Binocular spherical stereo. *IEEE Transactions on intelligent transportation systems*, 9(4):589–600, 2008. 1, 2
- [16] Kaiwen Lin and Toby P Breckon. Real-time low-cost omnidirectional stereo vision via bi-polar spherical cameras. In *International Conference Image Analysis and Recognition*, pages 315–325. Springer, 2018. 2
- [17] Chuiwen Ma, Liang Shi, Hanlu Huang, and Mengyuan Yan. 3d reconstruction from full-view fisheye camera. *arXiv preprint arXiv:1506.06273*, 2015. 2
- [18] Kevin Matzen, Michael F Cohen, Bryce Evans, Johannes Kopf, and Richard Szeliski. Low-cost 360 stereo photography and video capture. *ACM Transactions on Graphics (TOG)*, 36(4):1–12, 2017. 2
- [19] Morgan McGuire. Computer graphics archive, July 2017. <https://casual-effects.com/data>. 6
- [20] Hiroaki Miura and Masahide Kimoto. A comparison of grid quality of optimized spherical hexagonal–pentagonal geodesic grids. *Monthly weather review*, 133(10):2817–2833, 2005. 2
- [21] Edward S Popko. *Divided spheres: Geodesics and the orderly subdivision of the sphere*. CRC press, 2012. 3
- [22] David A Randall, Todd D Ringler, Ross P Heikes, Phil Jones, and John Baumgardner. Climate modeling with spherical geodesic grids. *Computing in Science & Engineering*, 4(5):32–41, 2002. 3
- [23] Davide Scaramuzza, Agostino Martinelli, and Roland Siegwart. A toolbox for easily calibrating omnidirectional cameras. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5695–5701. IEEE, 2006. 2
- [24] Vladyslav Usenko, Nikolaus Demmel, and Daniel Cremers. The double sphere camera model. In *2018 International Conference on 3D Vision (3DV)*, pages 552–560. IEEE, 2018. 2, 6
- [25] Qiang Zhao, Wei Feng, Liang Wan, and Jiawan Zhang. SPHORB: A fast and robust binary feature on the sphere. *International journal of computer vision*, 113(2):143–159, 2015. 3