Photometric Stereo with Twin-Fisheye Cameras

Jordan Caracotte, Fabio Morbidi, El Mustapha Mouaddib
MIS Laboratory, Université de Picardie Jules Verne
33 Rue Saint-Leu, 80039 Amiens, France
Emails: {jordan.caracotte, fabio.morbidi, mouaddib}@u-picardie.fr

***Abstract*—In this paper, we introduce and solve for the first time, the photometric stereo problem for low-cost 360-degree cameras. In particular, we present a spherical image irradiance equation which is adapted to *twin-fisheye cameras*, and an original algorithm for the estimation of light directions based on the specular highlights observed on mirror balls. Extensive experiments with synthetic and real-world images captured by a Ricoh Theta V camera, demonstrate the effectiveness and robustness of the proposed 3D reconstruction pipeline. To foster reproducible research, the image dataset and code developed for this paper are made publicly available at the address: home.mis.u-picardie.fr/∼fabio/PhotoSphere.html**

I. Introduction

A. Motivation and related work

In recent years, 360-degree commodity cameras have become increasingly popular and affordable. These systems consist of multiple perspective cameras rigidly attached to the same support, or they combine wide-angle (fisheye) lenses and mirrors. Thanks to their ability to capture a scene from all around, these low-cost sensors offer new opportunities for image-based 3D reconstruction via multi-view [1], [2] or photometric stereo [3].

Photometric stereo methods rely on the image-intensity variations caused by illumination changes, to perform a per-pixel estimation of the normal vectors of an observed surface, from which the shape of the 3D object can be ultimately recovered. The seminal work of Woodham [4] has been extended in several directions in the last decades: for example, in [5]–[8], the authors have relaxed the lighting constraints, and used outdoor images under natural sunlight illumination, to perform the 3D reconstruction. The so-called "uncalibrated" photometric stereo problem arises when no information about the illumination and/or geometry, and reflectance of the surface is available. This problem is well-known to be ill-posed and it has been the subject of extensive research lately [9]–[13]. Other extensions include near-field (i.e. local and non-strictly directional lighting) [14]–[16], more sophisticated surface reflectance models (i.e. "Bidirectional Reflectance Distribution Functions" or BRDF, other than the classical Lambertian reflectance) [17]–[20], the relaxation of constraints on camera placement [21]–[23] and combinations thereof [24].

Alongside these recent developments, Woodham's original formulation has been also extended from orthographic to perspective projection in [25]. However, efforts have lagged behind research on consumer-grade vision sensors, and, to the best of our knowledge, no results exist in the literature for *twin-fisheye cameras*, such as the Samsung Gear 360, the LG 360 CAM, and the Ricoh Theta and Insta360 series. These cameras incorporate two fisheye lenses mounted back-to-back on the same support, and an array of prisms which direct the light beams towards two imaging sensors (see Fig. 1). Spherical cameras such as the Ricoh Theta V, are affordable and lightweight, they have a compact form factor, and they offer distinctive advantages over standard perspective cameras for the 3D reconstruction of large-scale environments. In fact, *a single shot* is sufficient to record the appearance of an entire complex scene. This obviates the need for stitching up images coming from multiple vantage points, a limiting factor for the use of perspective cameras. Robotics and consumer electronics have also recently benefited from the attractive features of twin-fisheye sensors, e.g. for image-based pose estimation [26]–[28].

B. Original contributions, organization and notation

In this paper, we adapt the general theory developed in our previous work [29] for central panoramic cameras (i.e. vision systems which preserve the uniqueness of the projection viewpoint), to twin-fisheye cameras. We make standard assumptions on the lighting and reflectance models (i.e. Lambertian surfaces), and we leverage the unifying projection model [30], [31], tailored to multi-fisheye cameras [32], [33], to define a spherical image irradiance equation. Differently from [29], the two sub-images corresponding to each fisheye lens of the camera, are processed separately to estimate two per-pixel gradient fields, which are then merged into a single one. A drift-free normal integration algorithm is designed to process the joint gradient field and recover the radial distances, and thus the 3D shape of the scene observed by the camera. To estimate the light directions in real images, we took advantage of the specular highlights observed on multiple mirror balls inside the scene, and modified the algorithm in [34] to comply with our spherical formulation. The overall 3D reconstruction pipeline has been validated with synthetic images, and with real-world images captured by the Ricoh Theta V camera in a laboratory environment.

The rest of this paper is organized as follows. In Sect. II, we present the mathematical model of a twin-fisheye camera and the spherical image irradiance equation, and describe our 3D reconstruction pipeline. In Sect. III, we present the results of the simulation experiments. In Sect. IV, we describe the algorithms developed for the calibration of the Ricoh Theta V camera and for the estimation of the light directions, and we discuss the results obtained with the real-world images. Finally, in Sect. V, the main contributions of the paper are summarized and some possible directions for future research are outlined.

Notation: Throughout this article, as generally accepted in the literature, we shall use the term *twin-fisheye* to refer to the camera, and *dual-fisheye* to refer to the images captured by the twin-fisheye camera (see Fig. 1(a) and Fig. 1(c), respectively). The symbol $\mathbb{R}^n$ denotes the n-dimensional Euclidean space, $\mathbb{R}^{m\times n}$ the space of $m \times n$ matrices, $\|\mathbf{x}\|$ the Euclidean norm of vector $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{R}_y(\theta)$ and $\mathbf{R}_z(\varphi)$, the 3×3 elementary rotations of an angle θ and φ about the y- and z-axis, respectively. Finally, $\langle \mathbf{x}, \mathbf{y} \rangle$ is the scalar product of $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\widehat{\mathbf{x}}$ indicates a normalized vector i.e. $\widehat{\mathbf{x}} = \mathbf{x}/\|\mathbf{x}\|$, $\overline{\mathbf{x}}$ an estimate of vector $\mathbf{x}$, and $\triangleq$ the equality by definition.

II. Estimation of Radial Distances from Photometric Ratios

A. Modeling of a twin-fisheye camera

In this work, we deal with *twin-fisheye cameras*, i.e. 360-degree cameras equipped with two identical fisheye lenses pointing in opposite directions (see Fig. 1), with special attention to the Ricoh Theta V camera, which will be used in our real-world experiments in Sect. IV.

It is well known that some classes of fisheye cameras are approximately equivalent to a central catadioptric system [35], [36]. Therefore, the unified central projection model can be used [30], [31]. We describe each fisheye lens of the twin-fisheye camera by its own set of intrinsic parameters $\mathcal{P}_{c_j} = \{a_{u_j}, a_{v_j}, u_{0_j}, v_{0_j}, \xi_j\}$, $j \in \{1, 2\}$, where a_{u_j} and a_{v_j} are the focal lengths in pixels in the horizontal and vertical directions, respectively, and (u_{0_j}, v_{0_j}) are the coordinates of the principal point in pixels. The fifth parameter, ξ_j, is the distance between the unit sphere's first projection center and the perspective second projection center of fisheye lens j, as described in [31, Fig. 2]. Following [26], we assume that the translation vector between the reference frames $\mathcal{F}_{c_1}$ and $\mathcal{F}_{c_2}$ of the two fisheye lenses is zero, to ensure the uniqueness of point of view. Moreover, we assume that the camera frame $\mathcal{F}_c$ coincides with $\mathcal{F}_{c_1}$ (they are shown separately in Fig. 1(b), for illustration purposes only). The extrinsic parameters of the twin-fisheye camera are thus incorporated into the rotation matrix ${}^{c_2}\mathbf{R}_{c_1}$, the orientation of $\mathcal{F}_{c_1}$ relative to $\mathcal{F}_{c_2}$.

The spherical image captured by a twin-fisheye camera can be represented by using the spherical coordinate system (ρ, θ, φ) [37, Sect. 2.5]:

$$\begin{cases} x = \rho \sin\theta \cos\varphi, \\ y = \rho \sin\theta \sin\varphi, \\ z = \rho \cos\theta, \end{cases} \tag{1}$$

where $\rho \in [0, \infty)$ is the radial distance, $\theta \in (0, \pi)$ the polar angle and $\varphi \in [0, 2\pi)$ the azimuthal angle.

Our ultimate goal in this paper, will be to estimate the radial distance (or range) ρ from multiple images of a 3D scene observed from the same viewpoint, but under different illumination conditions.

B. Spherical image irradiance equation

To solve the photometric stereo problem for central panoramic cameras under a Lambertian reflectance model,

Fig. 1. (a) Front, side and top view of a twin-fisheye camera, the Ricoh Theta V (image courtesy of Ricoh), and schematic of the optical system with the two fisheye lenses, prisms and CMOS sensors; (b) $\mathcal{F}_c$ is the camera frame, and $\mathcal{F}_{c_1}$, $\mathcal{F}_{c_2}$, the reference frames associated to the two fisheye lenses (top view); (c) Example of dual-fisheye image taken by the Ricoh Theta V: the left and right sub-images correspond to the two fisheye lenses.

in our previous work [29, Th. 2], we have introduced the spherical image irradiance equation,

$$I(\theta, \varphi) = \frac{\alpha(\theta, \varphi)\, \mathbf{L}^T}{\|\mathbf{L}\| \sqrt{p^2 + \left(\dfrac{q}{\sin\theta}\right)^2 + 1}} \mathbf{R}_z(\varphi)\, \mathbf{R}_y(\theta) \begin{bmatrix} p \\ \dfrac{q}{\sin\theta} \\ -1 \end{bmatrix}, \tag{2}$$

where $I(\theta, \varphi)$ and $\alpha(\theta, \varphi) \in [0, 1]$, denote the intensity and the albedo at point (θ, φ), respectively, $\mathbf{L} \triangleq [p_d, q_d, -1]^T$ is the illumination vector (the light source without attenuation is set at infinity), and

$$p \triangleq \frac{1}{\rho}\frac{\partial \rho}{\partial \theta} = \frac{\partial \ln \rho}{\partial \theta}, \quad q \triangleq \frac{1}{\rho}\frac{\partial \rho}{\partial \varphi} = \frac{\partial \ln \rho}{\partial \varphi}. \tag{3}$$

Note that equation (2) only depends on the partial derivatives of $\ln \rho(\theta, \varphi)$ with respect to θ and φ, but not on $\ln \rho(\theta, \varphi)$ itself. As a consequence, the problem of recovering the radial distance $\rho(\theta, \varphi)$ from the image irradiance equation reduces to the problem of recovering $\ln \rho(\theta, \varphi)$ from (2). Since the natural logarithm is a bijective function and $\rho(\theta, \varphi) > 0$, estimate $\ln \rho(\theta, \varphi)$ amounts to recover $\rho(\theta, \varphi) = \exp(\ln \rho(\theta, \varphi))$.

Photometric stereo relies on multiple images of the same scene observed from the same point of view, under n different lighting conditions. If we define the images and the corresponding illumination vectors, $I_i(\theta,\, \varphi)$ and

$$\mathbf{L}_i \triangleq [p_{d_i},\, q_{d_i},\, -1]^T,\;\; i \in \{0,\, 1, \ldots,\, n-1\},$$

respectively, then from equation (2), the ith-image irradiance equation is:

$$I_i(\theta,\, \varphi) = \frac{\alpha(\theta,\, \varphi)\, \mathbf{L}_i^T}{||\mathbf{L}_i|| \sqrt{p^2 + \left(\frac{q}{\sin\theta}\right)^2 + 1}}\, \mathbf{R}_z(\varphi)\, \mathbf{R}_y(\theta) \begin{bmatrix} p \\ \frac{q}{\sin\theta} \\ -1 \end{bmatrix}.$$

By dividing the ith image by the kth (assuming that the latter is non-zero everywhere), we obtain:

$$\frac{I_i(\theta,\, \varphi)}{I_k(\theta,\, \varphi)} = \frac{\|\mathbf{L}_k\|\, \mathbf{L}_i^T\, \mathbf{R}_z(\varphi)\, \mathbf{R}_y(\theta) \Big[p,\, \frac{q}{\sin\theta},\, -1\Big]^T}{\|\mathbf{L}_i\|\, \mathbf{L}_k^T\, \mathbf{R}_z(\varphi)\, \mathbf{R}_y(\theta) \Big[p,\, \frac{q}{\sin\theta},\, -1\Big]^T}. \quad (4)$$

By expanding the numerator and the denominator, and by collecting the terms containing p and q, the photometric ratio (4) can be rewritten as,

$$A_{i,k}\, p \,+\, B_{i,k}\, q \,+\, C_{i,k} \,=\, 0,\;\; i,k \in \{0,1,\ldots,n-1\}, \quad (5)$$

where

$$A_{i,k} = I_i(\theta,\, \varphi)\, \|\mathbf{L}_i\| \langle \mathbf{L}_k,\, \mathbf{e}_\theta \rangle - I_k(\theta,\, \varphi)\, \|\mathbf{L}_k\| \langle \mathbf{L}_i,\, \mathbf{e}_\theta \rangle,$$

$$B_{i,k} = \frac{1}{\sin\theta} \big(I_i(\theta,\varphi) \|\mathbf{L}_i\| \langle \mathbf{L}_k,\, \mathbf{e}_\varphi \rangle - I_k(\theta,\varphi) \|\mathbf{L}_k\| \langle \mathbf{L}_i,\, \mathbf{e}_\varphi \rangle \big),$$

$$C_{i,k} = -I_i(\theta,\, \varphi)\, \|\mathbf{L}_i\| \langle \mathbf{L}_k, \mathbf{e}_\rho \rangle + I_k(\theta,\, \varphi)\, \|\mathbf{L}_k\| \langle \mathbf{L}_i,\, \mathbf{e}_\rho \rangle,$$

and $\{\mathbf{e}_\rho,\, \mathbf{e}_\theta,\, \mathbf{e}_\varphi\}$ is the basis set of unit vectors for the spherical coordinates [37, Sect. 2.3]. We can notice that system (5) is *linear* in p and q. Since two images are necessary to construct each of its equations, then three images are sufficient to recover the two unknowns p and q under *ideal conditions*. However, in the presence of image noise, the least-squares solution to system (5) is clearly preferable.

C. 3D reconstruction pipeline

In the proposed 3D reconstruction pipeline (see the flowchart in Fig. 2), the processing steps from the inverse perspective projection to the estimation of the gradient field, are performed separately for each sub-image. This yields a per-pixel estimation of $(p,\, q)$ in $\mathcal{F}_{c_1}$ and $\mathcal{F}_{c_2}$. In the *inverse perspective projection* step, each pixel of the spherical image $I_i(\theta,\, \varphi)$ is mapped into a pixel in the image plane $I_i(u,\, v)$, and into a corresponding 3D point on the unit sphere $I_i(x_s,\, y_s,\, z_s)$, $i \in \{0, 1, \ldots, n-1\}$. The estimated gradient field on the image plane and unit sphere is denoted by $(p(u,\, v),\, q(u,\, v))$ and $(p(x_s,\, y_s,\, z_s),\, q(x_s,\, y_s,\, z_s))$, respectively. In the *fusion and resampling* step, the two gradient fields are merged into one on unit the sphere, each hemisphere corresponding to a fisheye lens of the camera. In particular, along the equator, we set,

$$p(\pi/2,\, \varphi) \,=\, \tfrac{1}{2}\big[p(\pi - \pi/\lambda,\, \varphi) \,+\, p(\pi + \pi/\lambda,\, \varphi)\big],$$

$$q(\pi/2,\, \varphi) \,=\, \tfrac{1}{2}\big[q(\pi - \pi/\lambda,\, \varphi) \,+\, q(\pi + \pi/\lambda,\, \varphi)\big],$$

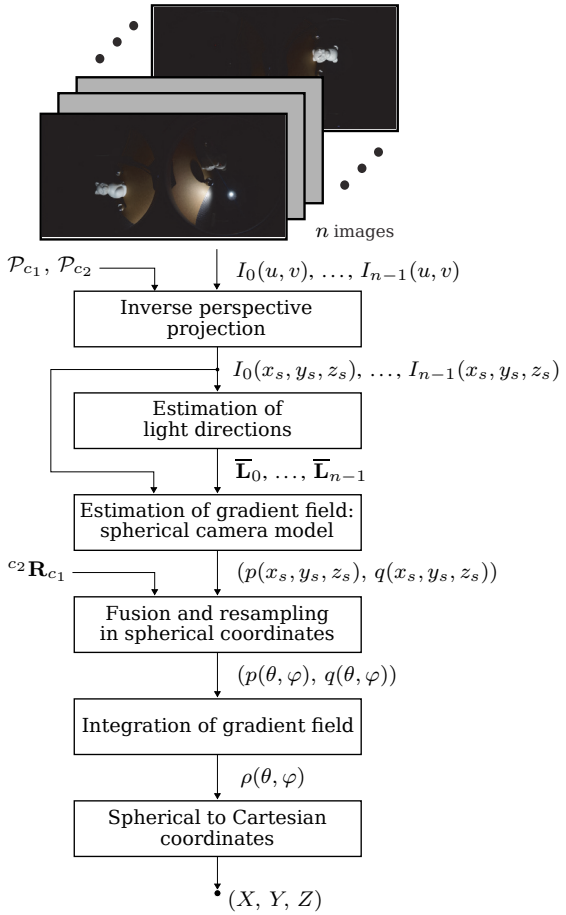


Fig. 2. Flowchart of the proposed photometric stereo algorithm, which takes n dual-fisheye images as input.

where $\varphi \in [0,\, 2\pi)$, and $1/\lambda$ is the sampling interval (we set $\lambda = 512$ in our tests in Sect. III and Sect. IV).

Next, as shown in Fig. 2, the unknown *radial map* $\rho(\theta,\, \varphi)$ is recovered by integration of the joint gradient field. Numerous methods exist in the literature to solve the integration problem [38]. The popular variational approach considers a least-squares cost function and solves the corresponding 2D Poisson-equation [39]. Iterative methods [40], path integrals and grid-based basis functions [3], have also been proposed for the numerical integration.

In this work, we use a drift-free method which takes the gradient field directly as input (cf. [29]). Two geometric constraints are enforced to improve the quality of the reconstructed surface. As a result of the *fusion and resampling* step, the joint gradient field $(p(\theta,\varphi),\, q(\theta,\varphi))$ is defined over the full sphere (the two poles excluded). The first constraint is a periodic boundary condition, and it is referred to as the "Circularity Constraint". Since the radial distance must take on the same values along opposite azimuth boundaries, we enforced $\rho(\theta,\, \varphi + \Delta\varphi) = \rho(\theta,\, 0)$ for $\varphi = 2\pi$ with $\theta \in (0,\, \pi)$, where $\Delta\varphi$ denotes the step-size along the φ-direction used in the finite-difference approximation of $\partial \ln \rho / \partial \varphi$ in (3). The second constraint, referred to as "Polar Constraint", is imposed on the contiguous grid points lying on the annuli

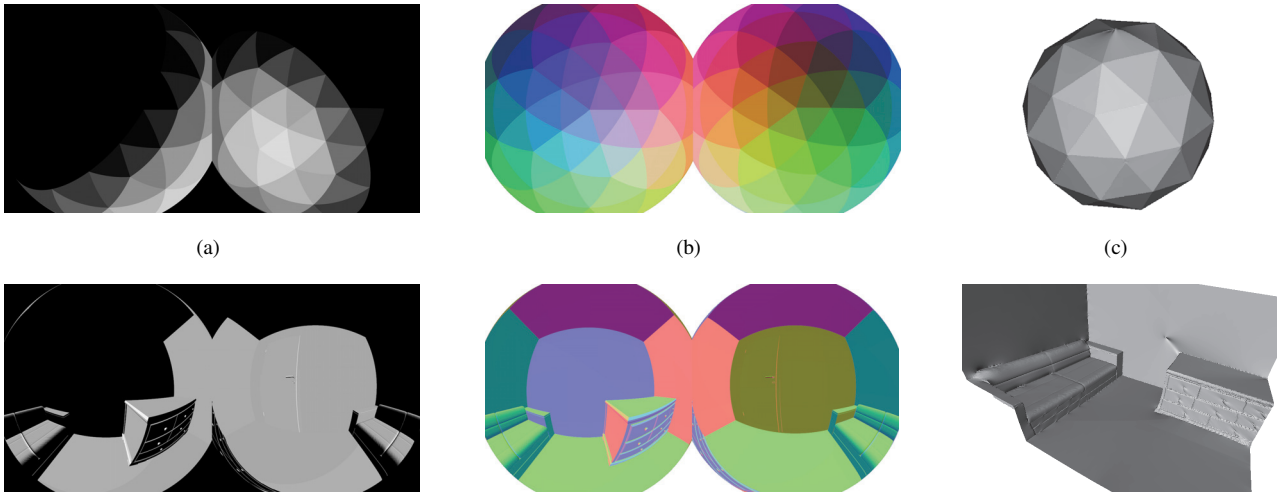


Fig. 3. Synthetic images: 1st row, *Geodesic Polyhedron*, 2nd row, *Room*. (a),(d) Sample input dual-fisheye images (grayscale); (b),(e) Estimated normal maps in false colors; (c),(f) 3D reconstructions: for the sake of clarity, the external surface of the reconstructed *Geodesic Polyhedron* is shown in (c), and only one side of the reconstructed *Room* is shown in (f) (the sofa is visible on the left and the chest of drawers on the right).

around the North or South poles (i.e. the points with $\theta = 0$ or $\theta = \pi$, which are the singularities of the spherical coordinate representation (1)). As such, all the points on the North and South annulus are considered as neighbors.

III. Simulation Experiments

In this section, we present the results of our experiments with synthetic images, which allowed us to easily verify how critical our assumptions on the geometry of the scene and on the lighting conditions are.

To test the proposed approach, we used Blender[1] to create a 3D virtual scene that we considered as our ground truth. We generated grayscale dual-fisheye images using the estimated calibration parameters of the Ricoh Theta V camera (see Sect. IV-A for more details), under 14 different lighting conditions. The illumination vectors $\mathbf{L}_i$, $i \in \{0, 1, \ldots, 13\}$, were positioned so as to light up each 3D point of the scene at least three times. To simplify the simulation process and reduce the number of input images to the 3D reconstruction pipeline, the self-occlusions were neglected. The quantization effect on 8-bit gray images was simulated by mapping $I_i(\theta, \varphi) \in [0, 1] \subset \mathbb{R}_{\geq 0}$, $i \in \{0, 1, \ldots, 13\}$, onto the discrete set $\{0, 1, \ldots, 255\}$.

To assess the 3D reconstruction quality, we considered the angular deviation between the ground-truth and estimated normal vectors,

$$\varepsilon(\theta, \varphi) = \arccos(\langle \mathbf{N}(\theta, \varphi), \overline{\mathbf{N}}(\theta, \varphi) \rangle),$$

being $\mathbf{N}(\theta, \varphi)$ the actual normal at (θ, φ), and $\overline{\mathbf{N}}(\theta, \varphi)$ the normal estimated by our method.

To validate the overall image-processing pipeline, in our first test we considered a *Geodesic Polyhedron* with 80 faces and circumradius r (a 2-frequency subdivided icosahedron). Fig. 3(a) shows one of the 14 input dual-fisheye images, and Fig. 3(b) the estimated normal map (in false colors) computed using the formula (cf. equation (2)):

$$\overline{\mathbf{N}}(\theta, \varphi) = \frac{1}{\sqrt{p^2 + \left(\frac{q}{\sin\theta}\right)^2 + 1}} \mathbf{R}_z(\varphi)\, \mathbf{R}_y(\theta) \begin{bmatrix} p \\ \frac{q}{\sin\theta} \\ -1 \end{bmatrix}.$$

Note that the *Mean Angle Error* (MAE) is $0.35°$ and not zero, because of the quantization error. Finally, Fig. 3(c) shows the external surface of the *Geodesic Polyhedron* reconstructed with the direct geometry-informed method described in Sect. II-C.

In our second and third test, we focused on two phenomena which occur when an image is recorded by a real camera. In fact, we studied the effect of gamma correction and of non-directional lighting on the quality of surface reconstruction. The second row of Table I reports the MAE for an increasing value of γ, and with (shaded columns) or without (white columns) correction. Finally, the fourth row of the table reports the MAE for a growing distance between a virtual point light source and the center of the *Geodesic Polyhedron*.

TABLE I

Geodesic Polyhedron: (1st and 2nd row) Mean angular error (MAE) in degrees for an increasing value of γ: the shaded columns report the values obtained with the gamma correction. (3rd and 4th row) MAE for a growing distance between the point light source and the camera.

γ	1.2	1.2	1.5	1.5	2.2	2.2
MAE [deg.]	0.75	0.36	2.02	0.37	4.22	0.41
Distance	2 r	5 r	10 r	15 r	20 r	25 r
MAE [deg.]	21.53	3.67	1.54	1.14	0.89	0.70

[1]Blender 2.79b: 3D Computer Graphics Software, `www.blender.org`

To define such a distance, we considered integer multiples of the circumradius r of the polyhedron

In our last test with synthetic images, we considered a more challenging 3D scene, called *Room*, which contains several depth discontinuities. The images for this scene were generated using the same procedure described above for the *Geodesic Polyhedron*. Fig. 3(d) shows one of the 14 grayscale dual-fisheye images, while the estimated normal map and final 3D reconstruction of one side of the scene are reported in Fig. 3(e) and Fig. 3(f), respectively.

IV. Real-world experiments

In this section, we describe the calibration method used to estimate the intrinsic and extrinsic parameters of the Ricoh Theta V camera, and present the algorithm for the estimation of the light directions in real images. We conclude the section with a discussion of the results obtained, and of some possible future improvements.

A. Calibration of the twin-fisheye camera

The Ricoh Theta V is an omnidirectional camera consisting of a twin-lens folded optical system coupled with two photosensitive sensors (1/2.3" 12 MP CMOS). The two fisheye lenses point in opposite directions and provide a spherical field of view (see Fig. 1(a)).

The quality and resolution of the dual-fisheye images recorded by the Ricoh Theta V in the default *live* and *video* modes is relatively poor. To overcome this limitation, we developed a plug-in module to capture dual-fisheye images with the maximum possible resolution (5792 pixels $\times$ 2896 pixels) in *still* mode. Thanks to the plug-in, we can also manually change the camera settings, such as the exposure time, aperture, ISO sensitivity, and quality (Q factor) of the output image.

To calibrate the Ricoh Theta V with *HySCaS* [41], we modeled it as a stereo system with two fisheye cameras sharing the same projection center (the translation vector between $\mathcal{F}_{c_1}$ and $\mathcal{F}_{c_2}$ was then set to zero). As in [26], we used a custom-made calibration rig consisting of six circle patterns glued inside two half cubes. Three images of the rig, one of which is reported in Fig. 4, were taken by the hand-held camera. This yielded,

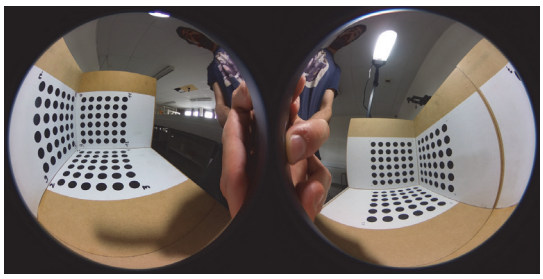


Fig. 4. One of the dual-fisheye images used for the calibration of the Ricoh Theta V camera. The calibration rig comprises six orthogonal circle patterns and it is visible in the two sub-images (the image is rotated by 180° to improve readability).

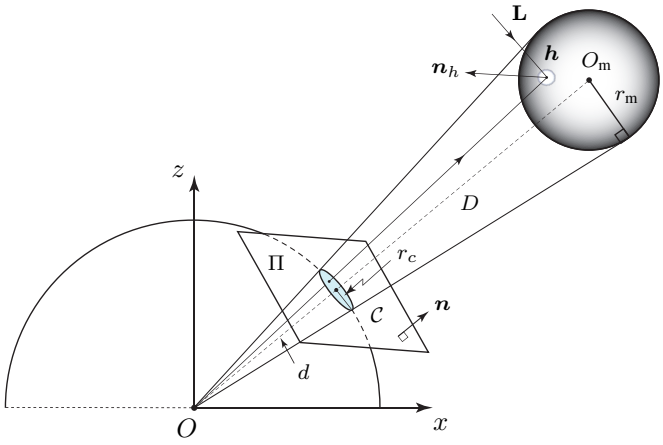


Fig. 5. Estimation of the light direction $\mathbf{L}$ from the specular highlight observed on a mirror ball of radius r_m and center O_m (top right). The unit sphere of the unified camera model, is depicted at the bottom left.

$\mathcal{P}_{c_1} = \{2492.8, 2496.0, 4311.0, 1464.1, 1.866\}$, $\mathcal{P}_{c_2} = \{2484.1, 2505.5, 1457.2, 1466.3, 1.848\}$, and

$$ {}^{c_2}\mathbf{R}_{c_1} = \begin{bmatrix} -1.000 & -0.001 & 0.025 \\ 0.000 & 1.000 & 0.031 \\ -0.025 & 0.031 & -0.999 \end{bmatrix} \simeq \mathbf{R}_y(\pi). $$

B. Estimation of light directions

To estimate the illumination vectors $\mathbf{L}_0, \ldots, \mathbf{L}_{n-1}$ using real dual-fisheye images, we adapted the light calibration method proposed in [34] for perspective cameras, to the spherical formulation considered in this paper. The method consists in observing the specular highlight generated by a light source on a mirror ball of *known* radius r_m located at a distance D from the center O of the unit sphere of the unified central projection model (see Fig. 5). **Algorithm 1** summarizes the main steps of our method for the estimation of a generic illumination vector $\mathbf{L}$. The circle $\mathcal{C}$ corresponds to the intersection between the unit sphere and the cone with apex at O and tangent to the mirror ball. The light direction $\mathbf{L}$, the viewing direction to the specular highlight, and the normal $\boldsymbol{n}_h$ to the mirror at the point of incidence $\boldsymbol{h}$, lie on the same plane. For the sake of simplicity, we approximated the specular highlight by a point, by considering its centroid.

C. Experimental results

To evaluate the performance of our 3D reconstruction pipeline in a real setting, we built a 2 $\times$ 1.25 $\times$ 1.12 m^3 booth made of an aluminum frame covered by a black curtain, to suppress the parasitic ambient light (see Fig. 6(a)). Two expanded polystyrene foam objects, *Cat* (23 cm tall and 14 cm at the widest point) and *Teddy Bear* (19 cm tall and 12 cm at the widest point), were placed inside the booth. The two objects, whose white matte surface complies with the Lambertian model, were illuminated with a 220 V, 5.3 W OSRAM 5-LED lamp with the following optical specifications: luminous flux 350 lm, CIE R_a 80, temperature 4000 K (cool white), and beam angle 36°.

Algorithm 1: Estimation of the light direction $\mathbf{L}$ from the specular highlight on a mirror ball (see Fig. 5).

Input : a) The apparent contour of the mirror ball, visible in one of the dual-fisheye sub-images,
b) The coordinates of the specular highlight in the dual-fisheye sub-image considered in a),
c) The radius r_{m} of the mirror ball.

Output: The estimated light direction $\overline{\mathbf{L}}$.

1 The pixels of the apparent contour are projected to the unit sphere at a circle $\mathcal{C}$, using the inverse projection model. Let $\mathbf{x}_{\mathcal{C}_j} \in \mathbb{R}^3$ be the coordinates of point j on $\mathcal{C}$, and $\mathbf{x}_{\text{cnt}} \in \mathbb{R}^3$ the coordinates of the centroid of points $\mathbf{x}_{\mathcal{C}_j}$, $j \in \{1, \ldots, P\}$;

2 To avoid poor numerical conditioning, the points $\mathbf{x}_{\mathcal{C}_j}$ are centered at the origin by subtracting their mean and scaled in the $[-1, 1]$ interval. The transformed points are denoted by $\mathbf{x}^*_{\mathcal{C}_j}$;

3 Let $\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ be the Singular Value Decomposition of $\mathbf{M} = [\mathbf{x}^*_{\mathcal{C}_1}, \ldots, \mathbf{x}^*_{\mathcal{C}_P}] \in \mathbb{R}^{3\times P}$ where $\mathbf{\Sigma}$ contains the singular values of $\mathbf{M}$ on the main diagonal, in descending order. The 3rd column of $\mathbf{U}$ corresponds to the normal $\boldsymbol{n} = [n_x, n_y, n_z]^T$ to the plane Π containing $\mathcal{C}$ (the direction of least surface variance). If $n_z < 0$, replace $\boldsymbol{n}$ with $-\boldsymbol{n}$;

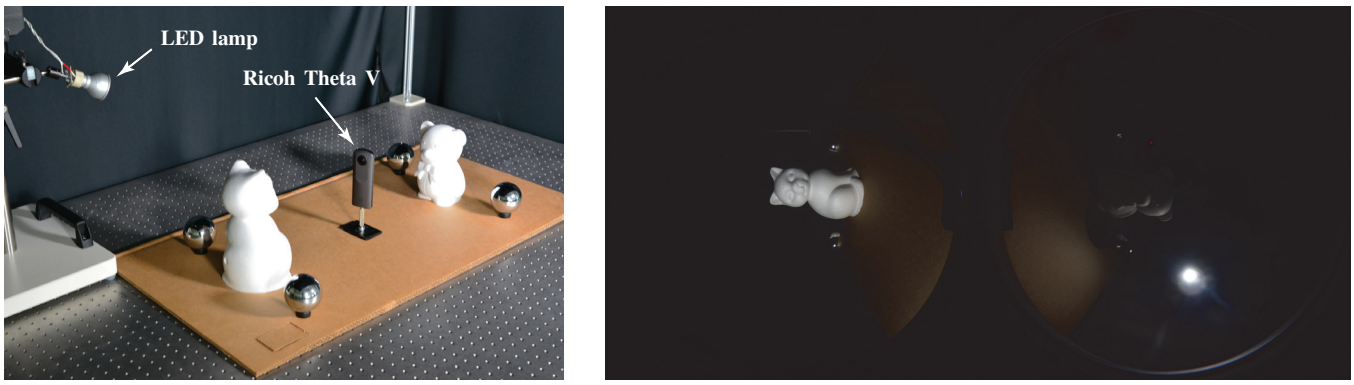
4 The distance d between the plane Π and the center of the unit sphere is computed as $d = |\langle \widehat{\boldsymbol{n}}, \mathbf{x}_{\text{cnt}} \rangle|$;

5 The coordinates of the center of $\mathcal{C}$ and its radius are given by $\boldsymbol{o}_{\text{c}} = -d\,\widehat{\boldsymbol{n}}$ and $r_c = \sqrt{1 - d^2}$;

6 The coordinates of the center O_{m} of the mirror ball are computed as $\boldsymbol{o}_{\text{m}} = \boldsymbol{o}_{\text{c}}\, r_{\text{m}}/r_c$;

7 The unit surface normal at $\boldsymbol{h}$ is given by $\widehat{\boldsymbol{n}}_h = (\boldsymbol{h} - \boldsymbol{o}_{\text{m}})/r_{\text{m}}$;

8 The light direction $\mathbf{L}$ is recovered as $\overline{\mathbf{L}} = 2\langle \widehat{\boldsymbol{h}}, \widehat{\boldsymbol{n}}_h \rangle \widehat{\boldsymbol{n}}_h - \widehat{\boldsymbol{h}}$.



(a) (b)

Fig. 6. (a) Experimental setup: LED lamp, Ricoh Theta V camera, and test objects (*Cat* and *Teddy Bear*). The four mirror balls used to estimate the light directions via **Algorithm 1**, are visible on both sides of the test objects; (b) One of the dual-fisheye images captured by the Ricoh Theta V.

A binary mask was applied to the dual-fisheye images of *Cat* and *Teddy Bear*, to subtract the background. We considered the three RGB channels separately, and normalized the corresponding pixel intensities using the formula,

$$\widehat{I}(\theta,\, \varphi) \;=\; \frac{I(\theta,\, \varphi) \;-\; \min I(\theta,\, \varphi)}{\max I(\theta,\, \varphi) \;-\; \min I(\theta,\, \varphi)},$$

where the min and max are evaluated over $\theta \in (0,\, \pi)$ and $\varphi \in [0,\, 2\pi)$. Overall, 11 images of *Teddy Bear* and 12 images of *Cat* were processed.

The experimental setup and one of the images captured by the Ricoh Theta V are shown in Fig. 6(a) and Fig. 6(b), respectively. The estimated normal and albedo maps of *Cat* and *Teddy Bear* are reported in Fig. 7(a) (the albedo map was computed with the method described in [42, Sect. 4.1]). Fig. 7(b) and Fig. 7(c), which show the 3D reconstruction of the test objects, demonstrate the effectiveness of the Circularity and Polar Constraints in the normal integration step. Note that a deformation can be observed on the top of *Cat*'s head, which is due to an incorrect estimation of the gradient field $(p(\theta, \varphi),\, q(\theta, \varphi))$ in this area.

D. Discussion

In this section, we provide some remarks on the obtained results, and discuss some possible improvements of our 3D reconstruction pipeline.

The inaccuracies observed in our 3D reconstructions of *Cat* and *Teddy Bear*, can be attributed to several factors: uncertainty in the estimation of light directions, cast shadows, etc. However, the main sources of error appear to be the non-ideal illumination conditions, the photometric response of the Ricoh Theta V, and the spatial resampling of the gradient field. The results reported in Table I provide evidence for these errors, and support this conclusion.

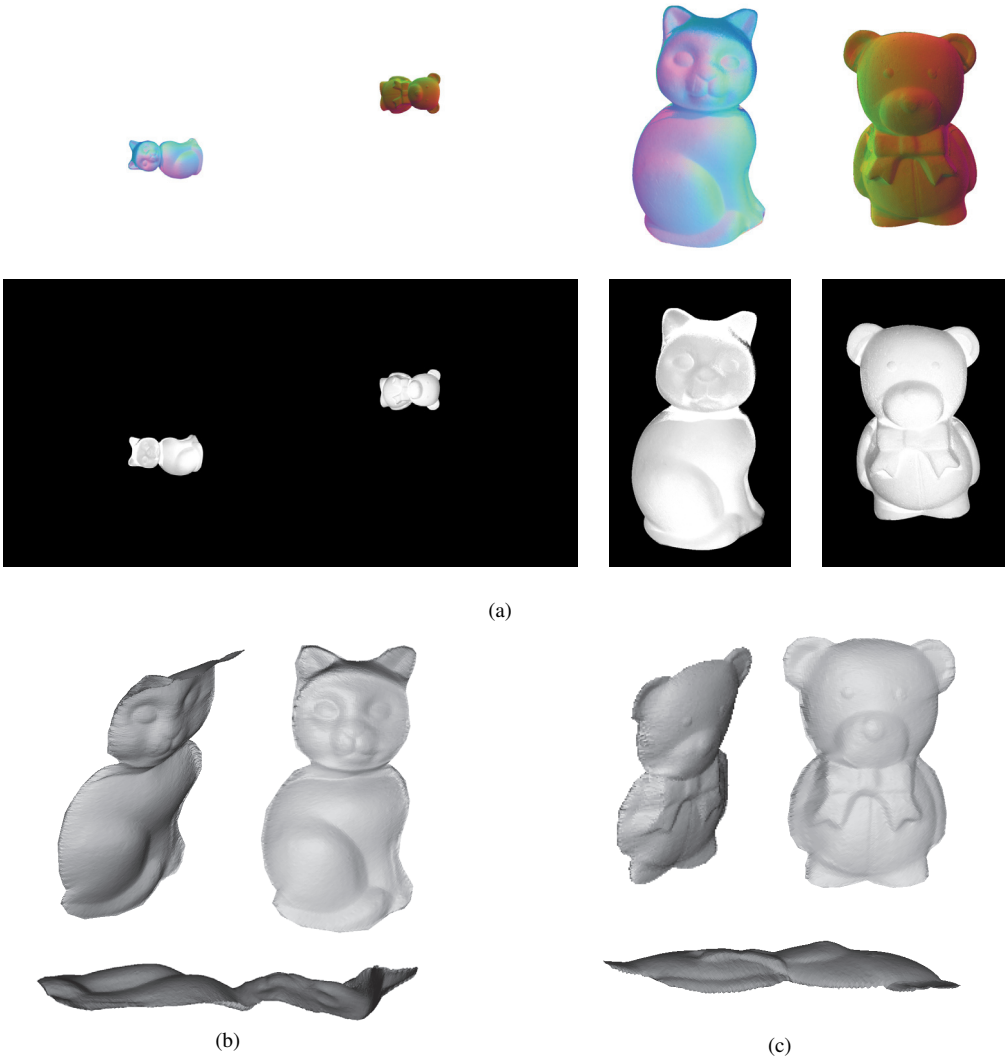


Fig. 7. Real-world images: (a) Estimated normal map in false colors (top), and estimated albedo map (bottom). Results in dual-fisheye format (left), and close-up views (right). (b),(c) 3D reconstruction of *Cat* and *Teddy Bear* (clockwise from top left: three-quarter, full face, and profile view).

In our tests with the synthetic images, the camera response curve was known and the assumption of directional lighting was strictly satisfied, as opposed to the experiments with real-world images. In the second test in Sect. III, we studied the effect of gamma correction, in an effort to reproduce the nonlinear behavior of the Ricoh Theta V with the virtual camera. In the third test of Sect. III, we moved a point light source progressively away from a known scene, to quantify its impact on the estimation of $(p(\theta, \varphi), q(\theta, \varphi))$. As evident from Table I, the farther the point light source, the smaller the MAE. In the real setup of Fig. 6(a), the distance between the LED lamp and the 3D scene ranges from 6.1r to 8.7r, where r = 23 cm is half the minimum distance between *Cat* and *Teddy Bear*.

In the reconstruction pipeline adapted to the dual-fisheye images, the interpolation phase occurs at a later stage than in [29], i.e. during the *fusion and resampling* step (cf. Fig. 2). This yields a per-pixel estimation of the gradient field $(p(\theta, \varphi), q(\theta, \varphi))$, which more accurately reflects the input data. In future works, we plan to take advantage of the free tangent directions introduced in [29, Th. 2], to mitigate the effect of spatial resampling (or to remove this processing step altogether).

As detailed in Sect. II-C, the gradient fields estimated from the dual-fisheye sub-images processed separately, are merged together (by considering a pure rotation ${}^{c_2}\mathbf{R}_{c_1}$ between $\mathcal{F}_{c_1}$ and $\mathcal{F}_{c_2}$). The same procedure could be applied, in principle, to other generic multi-camera arrangements having a unique effective viewpoint, as long as the intrinsic and extrinsic parameters are known.

Our experiments corroborated the proposed theory and successfully validated the whole 3D reconstruction pipeline. The results obtained with relatively few real-world images, provided further evidence in support of our method.

To reproduce the results of Sect. III and Sect. IV, our image dataset, plug-in for the Ricoh Theta V, and Matlab/Blender

code, can be freely downloaded at the following address: home.mis.u-picardie.fr/∼fabio/PhotoSphere.html

V. Conclusion and future works

In this paper, we have studied the photometric stereo problem for consumer-grade twin-fisheye cameras, which provide 360-degree images of a scene. Building upon our previous work and a novel spherical image irradiance equation, we have proposed a full 3D reconstruction pipeline, whose effectiveness has been demonstrated via experiments with synthetic and real images taken by a Ricoh Theta V camera.

This project is currently underway, and in future works we would like to validate our photometric stereo algorithm using dual-fisheye images of an outdoor scene under natural illumination conditions. The preliminary results presented in this paper being promising, our ultimate goal is to apply our 3D reconstruction pipeline to digital cultural heritage [43].

Acknowledgment

The work of J. Caracotte at the Université de Picardie Jules Verne was funded by the French "*Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation*", contract number: MIN/2016-17.

References

[1] Y. Furukawa and C. Hernández, "Multi-View Stereo: A Tutorial," *Found. Trends Comput. Graphics and Vision*, vol. 9, no. 1–2, pp. 1–148, 2015.

[2] L. Barazzetti, M. Previtali, and F. Roncoroni, "3D modelling with the Samsung Gear 360," in *Proc. 3D Virtual Reconstruction and Visualization of Complex Architectures*, vol. XLII-2/W3, 2017, pp. 85–90.

[3] J. Ackermann and M. Goesele, "A Survey of Photometric Stereo Techniques," *Found. Trends Comput. Graphics and Vision*, vol. 9, no. 3–4, pp. 149–254, 2015.

[4] R. Woodham, "Photometric method for determining surface orientation from multiple images," *Opt. Eng.*, vol. 19, no. 1, pp. 139–144, 1980.

[5] A. Abrams, C. Hawley, and R. Pless, "Heliometric Stereo: Shape from Sun Position," in *Proc. Eur. Conf. Comp. Vis.*, 2012, pp. 357–370.

[6] J. Ackermann, F. Langguth, S. Fuhrmann, and M. Goesele, "Photometric Stereo for Outdoor Webcams," in *Proc. IEEE Conf. Comp. Vis. Pattern Recogn.*, 2012, pp. 262–269.

[7] A. Abrams, K. Miskell, and R. Pless, "The Episolar Constraint: Monocular Shape from Shadow Correspondence," in *Proc. IEEE Conf. Comp. Vis. Pattern Recogn.*, 2013, pp. 1407–1414.

[8] J. Jung, J.-Y. Lee, and I. Kweon, "One-day outdoor photometric stereo via skylight estimation," in *Proc. IEEE Conf. Comp. Vis. Pattern Recogn.*, 2015, pp. 4521–4529.

[9] H. Hayakawa, "Photometric stereo under a light source with arbitrary motion," *J. Opt. Soc. Am. A*, vol. 11, no. 11, pp. 3079–3089, 1994.

[10] R. Basri, D. Jacobs, and I. Kemelmacher, "Photometric Stereo with General, Unknown Lighting," *Int. J. Comput. Vision*, vol. 72, no. 3, pp. 239–257, 2007.

[11] Y. Quéau, F. Lauze, and J.-D. Durou, "Solving Uncalibrated Photometric Stereo Using Total Variation," *J. Math. Imaging Vis.*, vol. 52, no. 1, pp. 87–107, 2015.

[12] Y. Quéau, T. Wu, F. Lauze, J.-D. Durou, and D. Cremers, "A Non-Convex Variational Approach to Photometric Stereo under Inaccurate Lighting," in *Proc. IEEE Conf. Comp. Vis. Pattern Recogn.*, 2017, pp. 350–359.

[13] T. Papadhimitri and P. Favaro, "A New Perspective on Uncalibrated Photometric Stereo," in *Proc. IEEE Conf. Comp. Vis. Pattern Recogn.*, 2013, pp. 1474–1481.

[14] R. Mecca, A. Wetzler, A. Brustein, and R. Kimmel, "Near Field Photometric Stereo with Point Light Sources," *SIAM J. Imag. Sci.*, vol. 7, no. 4, pp. 2732–2770, 2014.

[15] F. Logothetis, R. Mecca, and R. Cipolla, "Semi-calibrated near field photometric stereo," in *Proc. IEEE Conf. Comp. Vis. Pattern Recogn.*, 2017, pp. 941–950.

[16] F. Logothetis, I. Budvytis, R. Mecca, and R. Cipolla, "A CNN Based Approach for the Near-Field Photometric Stereo Problem," in *Proc. 31st British Machine Vis. Conf.*, 2020, article 55.

[17] H. Tagare and R. Defigueiredo, "A Theory of Photometric Stereo for a Class of Diffuse Non-Lambertian Surfaces," *IEEE Trans. Pattern Anal.*, vol. 13, no. 2, pp. 133–152, 1991.

[18] T. Higo, Y. Matsushita, N. Joshi, and K. Ikeuchi, "A Hand-held Photometric Stereo Camera for 3-D Modeling," in *Proc. IEEE Int. Conf. Comp. Vis.*, 2009, pp. 1234–1241.

[19] M. Weinmann, A. Osep, R. Ruiters, and R. Klein, "Multi-View Normal Field Integration for 3D Reconstruction of Mirroring Objects," in *Proc. IEEE Int. Conf. Comp. Vis.*, 2013, pp. 2504–2511.

[20] F. Lu, Y. Matsushita, I. Sato, T. Okabe, and Y. Sato, "From Intensity Profile to Surface Normal: Photometric Stereo for Unknown Light Sources and Isotropic Reflectances," *IEEE Trans. Pattern Anal.*, vol. 37, no. 10, pp. 1999–2012, 2015.

[21] C. Hernández, G. Vogiatzis, and R. Cipolla, "Multiview Photometric Stereo," *IEEE Trans. Pattern Anal.*, vol. 30, no. 3, pp. 548–554, 2008.

[22] D. Vlasic, P. Peers, I. Baran, P. Debevec, J. Popović, S. Rusinkiewicz, and W. Matusik, "Dynamic Shape Capture using Multi-View Photometric Stereo," *ACM Trans. Graphic*, vol. 28, no. 5, 2009, article n. 174.

[23] F. Logothetis, R. Mecca, and R. Cipolla, "A Differential Volumetric Approach to Multi-View Photometric Stereo," in *Proc. IEEE Int. Conf. Comp. Vis.*, 2019, pp. 1052–1061.

[24] B. Shi, Z. Mo, Z. Wu, D. Duan, S.-K. Yeung, and P. Tan, "A Benchmark Dataset and Evaluation for Non-Lambertian and Uncalibrated Photometric Stereo," *IEEE Trans. Pattern Anal.*, vol. 41, no. 2, pp. 271–284, 2019.

[25] A. Tankus and N. Kiryati, "Photometric Stereo under Perspective Projection," in *Proc. IEEE Int. Conf. Comp. Vis.*, vol. 1, 2005, pp. 611–616.

[26] G. Caron and F. Morbidi, "Spherical Visual Gyroscope for Autonomous Robots using the Mixture of Photometric Potentials," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2018, pp. 820–827.

[27] S. Pathak, A. Moro, H. Fujii, and H. Asama, "Distortion-Robust Spherical Camera Motion Estimation via Dense Optical Flow," in *Proc. 25th IEEE Int. Conf. Image Proc.*, 2018, pp. 3358–3362.

[28] Y. Shan and S. Li, "Discrete Spherical Image Representation for CNN-Based Inclination Estimation," *IEEE Access*, vol. 8, pp. 2008–2022, 2020.

[29] J. Caracotte, F. Morbidi, and E. Mouaddib, "Photometric stereo with central panoramic cameras," *Comput. Vis. Image Und.*, vol. 201, 2020, article 103080.

[30] C. Geyer and K. Daniilidis, "A Unifying Theory for Central Panoramic Systems and Practical Implications," in *Proc. Eur. Conf. Comp. Vis.*, 2000, pp. 445–461.

[31] J. Barreto, "A unifying geometric representation for central projection systems," *Comput. Vis. Image Und.*, vol. 103, no. 3, pp. 208–217, 2006.

[32] A. Jamaluddin, C. Jiang, O. Morel, R. Seulin, and D. Fofi, "3D Reconstruction from Specialized Wide Field of View Camera System Using Unified Spherical Model," in *Proc. Int. Conf. Image Anal. Process.*, 2017, pp. 495–506.

[33] W. Gao and S. Shen, "Dual-fisheye omnidirectional stereo," in *Proc. IEEE/RSJ Int. Conf. Intel. Robots Syst.*, 2017, pp. 6715–6722.

[34] D. Schnieders and K.-Y. Wong, "Camera and light calibration from reflections on a sphere," *Comput. Vis. Image Und.*, vol. 117, no. 10, pp. 1536–1547, 2013.

[35] X. Ying and Z. Hu, "Can We Consider Central Catadioptric Cameras and Fisheye Cameras within a Unified Imaging Model?" in *Proc. Eur. Conf. Comp. Vis.*, 2004, pp. 442–455.

[36] J. Courbon, Y. Mezouar, and P. Martinet, "Evaluation of the Unified Model of the Sphere for Fisheye Cameras in Robotic Applications," *Adv. Robotics*, vol. 26, no. 8–9, pp. 947–967, 2012.

[37] H. Weber and G. Arfken, *Essential Mathematical Methods for Physicists*, 6th ed. Academic Press, 2003.

[38] Y. Quéau, J.-D. Durou, and J.-F. Aujol, "Normal Integration: A Survey," *J. Math. Imaging Vis.*, vol. 60, no. 4, pp. 576–593, 2018.

[39] ——, "Variational Methods for Normal Integration," *J. Math. Imaging Vis.*, vol. 60, no. 4, pp. 609–632, 2018.

[40] I. Horovitz and N. Kiryati, "Depth from gradient fields and control points: bias correction in photometric stereo," *Image Vision Comput.*, vol. 22, no. 9, pp. 681–694, 2004.

[41] G. Caron and D. Eynard, "Multiple camera types simultaneous stereo calibration," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2011, pp. 2933–2938.

[42] W. Smith and F. Fang, "Height from photometric ratio with model-based light source selection," *Comput. Vis. Image Und.*, vol. 145, pp. 128–138, 2016.

[43] E. Mouaddib, G. Caron, D. Leclet-Groux, and F. Morbidi, "E-Cathedral: *In Silico* Monumental Heritage," 2015-2025, [web] https://mis.u-picardie.fr/e-cathedrale.