Teleidoscopic Imaging System for Microscale 3D Shape Reconstruction

Ryo Kawahara
Kyushu Institute of Technology
rkawahara@ai.kyutech.ac.jp

Meng-Yu Jennifer Kuo
University of Minnesota
kuo00052@umn.edu

Shohei Nobuhara
Kyoto University
nob@i.kyoto-u.ac.jp

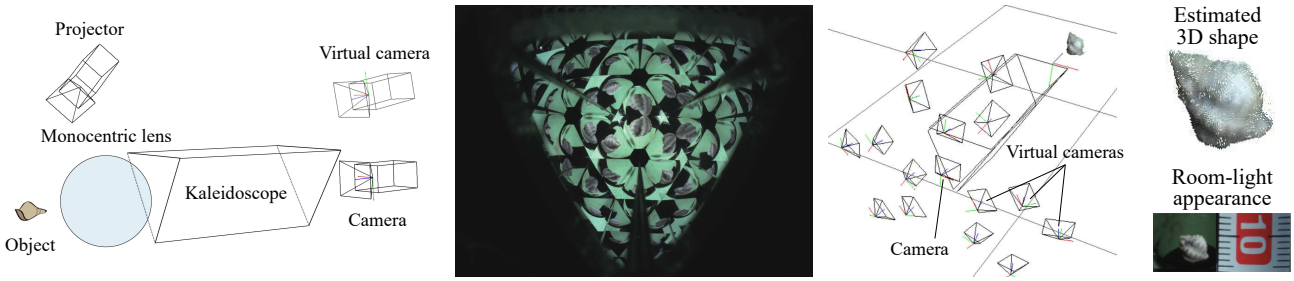

Figure 1. Teleidoscopic Imaging System. (a) We propose a novel system for microscale 3D imaging using an imaging setup that consists of a camera, a kaleidoscopic multi-facet mirror with a monocentric front lens, and a projector. (b) Our system realizes a dense close-up surrounding multi-view capture of the target. (c) Our method enables fuller and dense reconstruction of microscale objects.

Abstract

This paper proposes a practical method of microscale 3D shape capturing by a teleidoscopic imaging system. The main challenge in microscale 3D shape reconstruction is to capture the target from multiple viewpoints with a large enough depth-of-field. Our idea is to employ a teleidoscopic measurement system consisting of three planar mirrors and monocentric lens. The planar mirrors virtually define multiple viewpoints by multiple reflections, and the monocentric lens realizes a high magnification with less blurry and surround view even in closeup imaging. Our contributions include, a structured ray-pixel camera model which handles refractive and reflective projection rays efficiently, analytical evaluations of depth of field of our teleidoscopic imaging system, and a practical calibration algorithm of the teleidoscopic imaging system. Evaluations with real images prove the concept of our measurement system.

1. Introduction

Microscale 3D reconstruction has found profound applications in a wide range of domains including medical imaging, life science, and aquaculture, due to its non-constrained and non-invasive measurements. The main challenges in image-based microscopic 3D shape measurement is its shallow depth of field and camera arrangement in the closeup scenario. Applying conventional multiple camera system designed for human-size capture [14, 30] cannot be a feasible solution due to limitations on camera placement. Conventional multiple mirror system [33] also have difficulties inevitably in depth-of-focus due to differences in their optical path lengths with varying numbers of bounces.

In this paper, we show that the fuller 3D shape of a microscale object can be recovered. Our key idea is to employ a catadioptric imaging system which realizes a practical closeup multi-view imaging. The point of our design is that the system has a monocentric front lens like a teleidoscope, instead of using microscopic system in the camera side. That is, as shown in Fig. 1(a), we introduce a kaleidoscopic multi-facet mirror between the front lens and the camera, where the design realizes a deeper depth-of-field and results in less blurring imaging. Unlike conventional microscopic imaging system such as differential phase contrast (DPC) microscopy [4, 34] and multi-focus approaches [15, 24, 26], our method realizes a multi-view capture of the target from a single physical viewpoint which can contribute to free-viewpoint rendering, 3D shape reconstruction, and reflection analysis. Our system also has an advantage over existing imaging systems that build multiple views behind the main lens [7, 21] in a closeup environment. The wide FoV with a monocentric main lens and virtual multiple views with mirrors allow closeup and surrounding view capturing in focus.

We call our system *teleidoscopic imaging system* and show that the system can be compactly modeled by a structured ray-pixel camera model [11], which handles refractive and reflective projection rays efficiently. Based on our ray-pixel camera model, we derive a practical calibration algo-

rithm that estimates the positions of the monocentric lens and multi-facet mirror w.r.t. the camera by using a single reference planar pattern (*i.e.*, a checkerboard). Given the calibration parameters, a scene point can then be linearly triangulated from its teleidoscopic projection in a direct linear transform (DLT) manner [13].

We implement our method with an imaging system consisting of a camera, three planar mirrors, a monocentric lens, and a projector placed outside the mirrors for structured-light casting. We quantitatively evaluate the computation cost of numerical projections, the robustness of the calibration, and the depth of field of our teleidoscopic imaging system. We also validate the effectiveness of our method qualitatively on a number of real-world microscale objects. These results demonstrate that our method enables holistic and dense reconstruction of microscale objects. We believe our method expands the avenues of three-dimensional analysis of microscale objects and scenes in real world scenarios.

2. Related Work

Our teleidoscopic imaging system is a catadioptric system consisting of a monocentric lens and kaleidoscopic mirrors (Fig. 1(a)). While catadioptric system has a wide variety of applications such as omnidirectional observation and panoramic stereo [1, 10, 25, 27], this section reviews studies with multiple flat mirrors and monocentric lenses.

2.1. Imaging with Multiple Flat Mirrors

A fundamental motivation of introducing mirrors in observation system is to increase the number of viewpoints without installing additional cameras for multi-view capture of a target [3, 8, 20, 28, 32, 33]. Takahashi *et al*. [33] have proposed a kaleidoscopic imaging system and demonstrated a 3D shape reconstruction using multiple reflections. Tagawa *et al*. [32] have proposed a multi-facet imaging system that observes a target from an equally distributed virtual cameras for reflectance analysis.

2.2. Imaging with Monocentric Lens

In the context of imaging system, the monocentric lens is often used to obtain a wide field-of-view [17, 35], in particular for endoscopes, as a short focal length lens. The monocentric lens, however, has additional useful characteristics: its symmetric structure and magnifying power. Cossairt *et al*. [5] have proposed a camera array which captures a same scene through a single monocentric lens so that the images can be stitched into a single high-resolution image. Similarly, Dansereau *et al*. [6] have proposed a lightfield camera which captures omnidirectional lightfield images through a single monocentric lens with a camera orbiting around it.

A typical use of the monocentric lens as a magnifier can be found in the Leeuwenhoek's microscope in the 17th century. It utilizes a single monocentric lens and realized over $100\times$ magnifications [9].

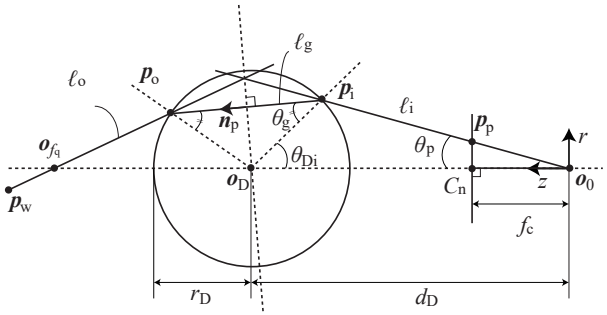

Figure 2. Measurement through Monocentric Lens. The segments of projection path ℓ_o-ℓ_g-ℓ_i have an axially symmetric structure around the axis which directs to the monocentric lens center $\boldsymbol{o}_D$.

2.3. Ray-pixel Camera Model

In order to handle light paths captured via complex imaging system such as reflection and refraction, Grossberg and Nayar [11] have proposed a ray-pixel (or raxel) camera model which describes the projection by end-to-end mappings between the pixels and their corresponding emergent rays. While this pixel-wise representation has a great flexibility to describe complicated projections [22, 31], it is also possible to exploit a structure of the rays to simplify the model [16, 22, 36–38]. Kawahara *et al*. [16] have proposed a ray-pixel camera model for cameras observing the scene through flat surfaces. Compared with analytical forward projections through flat surfaces which requires solving a high order equation [2], it realized a practical numerical forward projection computation by exploiting the axially symmetric structure of the rays.

3. Ray-Pixel Camera Model for Teleidoscopic Imaging System

In our teleidoscopic system, the cameras mirrored by the kaleidoscopic multi-facet mirrors observe the scene through the single monocentric lens (Fig. 1(a)) as done in [5, 6], and each of the monocentric lens and the camera pairs form a microscope capturing the object from a different direction.

In this section, we show that this design realizes a microscale multi-view capturing with a deep depth-of-field, and that the projections by the teleidoscopic imaging system can be modeled compactly by ray-pixel cameras.

3.1. Measurement through Monocentric Lens

As illustrated in Fig. 2, suppose a perspective camera whose camera center is located at $\boldsymbol{o}_0$ observes the scene through a monocentric lens located at $\boldsymbol{o}_D$. Here the perspective camera can be assumed to be directed to the center

of the monocentric lens $\boldsymbol{o}_D$ without loss of generality, since we can calibrate the rotation to align the optical axis of the camera to the line $\boldsymbol{o}_D - \boldsymbol{o}_0$ as described later. We denote this normalized camera by C_n.

Obviously, the rays backprojected through pixels of C_n have an axially symmetric structure around the optical axis, and an incident ray ℓ_i through a pixel $\boldsymbol{p}_p$ and its refraction ℓ_g and emergent ray ℓ_o are on a single plane-of-refraction [2]. Therefore, we can describe the ray through a single pixel by a 2D $(r, z)^\top$ coordinate system centered at $\boldsymbol{o}_0$.

Hereafter, let r_α and z_α denote the r and z elements of the vector $\boldsymbol{\alpha}$ in general. For example, point $\boldsymbol{p}_p$ is described as $\boldsymbol{p}_p = (r_{\boldsymbol{p}_p}, z_{\boldsymbol{p}_p})^\top$. Also let $\boldsymbol{v}_X = (r_{\boldsymbol{v}_X}, z_{\boldsymbol{v}_X})^\top$ denote the direction vector of line ℓ_X towards the scene from the camera, where $X = \{\mathrm{i}, \mathrm{g}, \mathrm{o}\}$ represents the label of the light path segment.

Consider the projection path ℓ_o-ℓ_g-ℓ_i from $\boldsymbol{o}_0$ through the point $\boldsymbol{p}_p$. The incident point $\boldsymbol{p}_i$ on the sphere is described as a function of $\theta_p = \tan^{-1}(r_{\boldsymbol{p}_p}/f_c)$ as

$$\begin{aligned} z_{p_i} &= \frac{d_D - \sqrt{d_D^2 - (1+\tan^2\theta_p)(d_D^2 - r_D^2)}}{1+\tan^2\theta_p}\,, \\ r_{p_i} &= z_{p_i}\tan\theta_p\,, \end{aligned} \tag{1}$$

where d_D and r_D are the distances to the monocentric lens and the lens radius.

The refraction angle θ_g is then given by Snell's law with assuming the refractive indices of the air μ_a and the lens μ_g are known:

$$\begin{aligned} &\mu_g \sin\theta_g = \mu_a \sin(\theta_p + \theta_{Di})\,, \\ \Leftrightarrow\ &\sin\theta_g = \frac{\mu_a}{\mu_g}\frac{d_D r_{p_i}}{r_D\sqrt{r_{p_i}^2 + z_{p_i}^2}}\,, \end{aligned} \tag{2}$$

where θ_{D_i} is the angle of $\boldsymbol{o}_0$-$\boldsymbol{o}_D$-$\boldsymbol{p}_i$.

The point of emergence $\boldsymbol{p}_o$ is the intersection of the path ℓ_g and the sphere. It is also obtained as the mirror of the point $\boldsymbol{p}_i$ because of the symmetrical relationship between incident path ℓ_i and outgoing ℓ_o w.r.t. the monocentric lens. As shown in Fig. 2, ℓ_i and ℓ_o are in line symmetry to the line that is perpendicular to the vector $\boldsymbol{n}_p$ through the point $\boldsymbol{o}_D$. That is, given the point $\boldsymbol{p}_i$, its reflection $\boldsymbol{p}_o$ is described as

$$\begin{aligned} &\boldsymbol{p}_o = H_p\boldsymbol{p}_i + \boldsymbol{t}_p\,, \\ \Leftrightarrow\ &\boldsymbol{p}_o = \left(\boldsymbol{I} - 2\boldsymbol{n}_p\boldsymbol{n}_p^\top\right)\boldsymbol{p}_i + 2(\boldsymbol{o}_D^\top\boldsymbol{n}_p)\boldsymbol{n}_p\,, \end{aligned} \tag{3}$$

where H_p is the Householder matrix and $\boldsymbol{t}_p$ denotes the center of reflection. Besides, the direction $\boldsymbol{n}_p$ is given as

$$\boldsymbol{n}_p = \begin{bmatrix}\cos(\pi/2 + \theta_{Di} - \theta_g) \\ \sin(\pi/2 + \theta_{Di} - \theta_g)\end{bmatrix} = \begin{bmatrix}-\sin(\theta_{Di} - \theta_g) \\ \cos(\theta_{Di} - \theta_g)\end{bmatrix}\,, \tag{4}$$

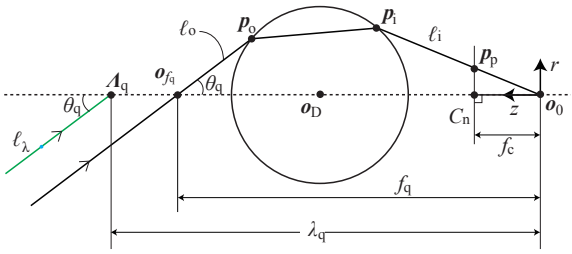

Figure 3. 3D-2D Projection. The black line ℓ_o is the correct projection path that intersects with the optical axis at the camera center $\boldsymbol{o}_0$. Our 3D-2D projection hypothesizes a line ℓ_λ as an initial guess of projection, and then optimizes it by verifying if the point and the angle of intersection with the optical axis is consistent with the ray-pixel mapping $\langle\theta_q, f_q\rangle$.

therefore H_p and $\boldsymbol{t}_p$ are rewritten to

$$\begin{aligned} H_p &= \begin{bmatrix}\cos 2(\theta_{Di} - \theta_g) & \sin 2(\theta_{Di} - \theta_g) \\ \sin 2(\theta_{Di} - \theta_g) & -\cos 2(\theta_{Di} - \theta_g)\end{bmatrix}\,, \\ \boldsymbol{t}_p &= \begin{bmatrix}-d_D \sin 2(\theta_{Di} - \theta_g) \\ d_D\cos 2(\theta_{Di} - \theta_g) + d_D\end{bmatrix}\,. \end{aligned} \tag{5}$$

The direction $\boldsymbol{v}_o$ is also given by Householder matrix as

$$\boldsymbol{v}_o = H_p\boldsymbol{v}_i\,. \tag{6}$$

As a result, the intersection $\boldsymbol{o}_{f_q} = (0, f_q)^\top$ of the ray ℓ_o and the optical axis is given as follows:

$$\boldsymbol{o}_{f_q} = t_o\boldsymbol{v}_o + \boldsymbol{p}_o\,, \tag{7}$$

$$\Leftrightarrow t_o = -\frac{r_{\boldsymbol{p}_o}}{r_{\boldsymbol{v}_o}}, \quad f_q = z_{\boldsymbol{p}_o} + \left(\frac{-r_{\boldsymbol{v}_o}}{z_{\boldsymbol{v}_o}}\right)^{-1} r_{\boldsymbol{p}_o}\,. \tag{8}$$

This equation describes the ray ℓ_o backprojecting from the pixel $\boldsymbol{p}_p$ as a function of θ_p. Inversely, the forward projection can be obtained for a given 3D point in front of the monocentric lens. The analytical 3D-2D projection, however, requires solving a 10th-degree equation [2]. The next section introduces a ray-pixel camera which exploits the axial symmetric structure of the rays to provide an efficient numerical 3D-2D projection.

3.1.1 Spherical Refraction Ray-Pixel Camera Model

Up to this point, we observed that the rays emitted from a perspective camera through a monocentric lens show an axially symmetric structure around the line from the camera to the lens center. As illustrated in Fig. 2, given a pixel $\boldsymbol{p}_p$ by specifying θ_p, the corresponding ray ℓ_o can be determined uniquely by Eq. (8). This section introduces our spherical refraction ray-pixel camera model by representing the ray-pixel mapping as follows.

Fig. 3 illustrates the light path from a pixel $\boldsymbol{p}_p$ on the plane-of-refraction. Suppose the emergent ray ℓ_o intersects

Table 1. The Pixel-Ray Mapping in Spherical Refraction Ray-Pixel Camera. A virtual pixel parameterized by θ_q is associated with a pixel-wise focal length $f_q(\theta_q) \in \mathbb{R}$. The derivative $\frac{\mathrm{d}f_q(\theta_q)}{\mathrm{d}\theta_q}$ is also stored for our numerical 3D-2D projection.

pixel	θ_{q0}	θ_{q1}	$\cdots$	θ_{qN}
ray	$f_q(\theta_{q0})$	$f_q(\theta_{q1})$	$\cdots$	$f_q(\theta_{qN})$
$\frac{\mathrm{d}f_q(\theta_q)}{\mathrm{d}\theta_q}$	$f'_q(\theta_{q0})$	$f'_q(\theta_{q1})$	$\cdots$	$f'_q(\theta_{qN})$

with the optical axis at $\boldsymbol{o}_{f_q} = (0, f_q)^\top$ with angle θ_q. Obviously, changing the position of $\boldsymbol{p}_p$ in r direction results in changing the corresponding $\langle f_q, \theta_q \rangle$ pair. That is, since the mapping between the pixel $\boldsymbol{p}_p$ *i.e.*, ℓ_i, and the ray ℓ_o, *i.e.*, $\langle f_q, \theta_q \rangle$ is bijective because of the reversibility of light, representing the light paths from the pixels in r space is identical to knowing all possible $\langle f_q, \theta_q \rangle$ pairs. In other words, the system of Fig. 3 as a whole can be seen as a pixel-wise varifocal camera which changes the focal length f_q for each virtual pixel parameterized by θ_q. In fact, the mapping $\theta_q \mapsto f_q$ is a monotonic function due to the spherical structure of the lens.

Table 1 shows our ray-pixel mapping. Due to the spherical structure of the lens, this is a discretization of the monotonic function $\theta_q \mapsto f_q$ by θ_q. In practice, we sample θ_q so that their interval results in a sub-pixel sampling in the original image domain. Notice that the derivative $\frac{\mathrm{d}f_q(\theta_q)}{\mathrm{d}\theta_q}$ is also stored for our numerical 3D-2D projection as described below.

3.1.2 3D-2D Projection

Instead of the analytical forward projection which requires solving a 10th-degree equation [2], this section introduces a numerical forward projection using our ray-pixel mapping table (Tab. 1). The key idea of our numerical forward projection is to hypothesize a projection line ℓ_λ in Fig. 3 first, and check if it intersects with the optical axis at the identical virtual focal length stored in the ray-pixel mapping table (Tab. 1). If the focal lengths are not identical, then ℓ_λ, *i.e.*, the virtual pixel θ_q equivalently, is refined to minimize the difference. We can formulate this process as a Gauss-Newton optimization as follows.

Gauss-Newton Method As shown in Fig. 3, let us consider the ray ℓ_λ from a 3D point to be projected. By hypothesizing that ℓ_λ intersects the optical axis at an angle θ_q, *i.e.*, by hypothesizing that the 3D point is projected to a virtual pixel θ_q, we can compute the intersection $\Lambda_q(\theta_q) = (0, \lambda_q(\theta_q))^\top$ of ℓ_λ and the optical axis. If λ_q is equal to the virtual focal length $f_q(\theta_q)$ stored in the ray-pixel mapping, the ray ℓ_λ is identical to ℓ_o, and hence that can intersect with the optical axis at the camera center $\boldsymbol{o}_0$.

That is, the numerical 3D-2D projection can be achieved by solving the following optimization:

$$\theta_q = \underset{\theta_q}{\operatorname{argmin}}\, G(\theta_q) = \underset{\theta_q}{\operatorname{argmin}}\, \left(f_q(\theta_q) - \lambda_q(\theta_q)\right) . \quad (9)$$

Here $G(\theta_q)$ is a monotonic function and hence we can refine θ_q iteratively by Gauss-Newton method as

$$\begin{aligned} \theta_q^{(k+1)} &= \theta_q^{(k)} - G\frac{\Delta\theta_q}{\Delta G} \\ &= \theta_q^{(k)} - \frac{f_q(\theta_q^{(k)}) - \lambda_q(\theta_q^{(k)})}{f'_q(\theta_q^{(k)}) - \lambda'_q(\theta_q^{(k)})} , \end{aligned} \quad (10)$$

where $\theta_q^{(k)}$ denotes θ_q of the k-th iteration.

Therefore, if we compute the derivative f'_q beforehand as shown in the third row of Tab. 1, then this 3D-2D projection can be computed efficiently.

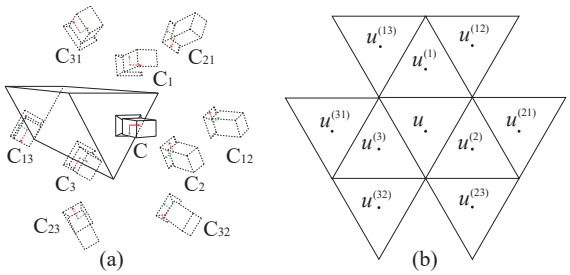

Figure 4. Kaleidoscopic Imaging. (a) Mirrored Cameras. (b) Chambers in the real camera image. The label i of C_i and $\boldsymbol{u}^{(i)}$ denote the reflections of C and u by the ith mirror respectively.

3.2. Multifacet Mirror

As is well known, observing the scene via a multi-facet mirror or a kaleidoscope is identical to observing the scene by virtual multi-view cameras, and in particular, kaleidoscopes with three mirrors are known to be reasonable in terms of less overlaps of mirrored images called discontinuities [29, 33]. In our teleidoscopic imaging system, we use a three-facet mirror which defines reflections of the spherical refraction ray-pixel camera introduced in Sec. 3.1.1.

The reflection $\boldsymbol{p}'$ of a 3D point $\boldsymbol{p}$ by a mirror of normal $\boldsymbol{n}_i$ and distance d is given by

$$\begin{aligned} &\boldsymbol{p}' = H_i\boldsymbol{p} + \boldsymbol{t}_i \\ \Leftrightarrow\ &\boldsymbol{p}' = \left(\boldsymbol{I} - 2\boldsymbol{n}_i\boldsymbol{n}_i^\top\right)\boldsymbol{p} + 2d_i\boldsymbol{n}_i , \end{aligned} \quad (11)$$

where H_i the Householder transformation matrix.

In the case of kaleidoscopic imaging, the mirrors generate bouncing reflections as shown in Fig. 4. The reflection of $\boldsymbol{p}'$ by another mirror of normal $\boldsymbol{n}_j$ and distance d is given simply by

$$\begin{aligned} \boldsymbol{p}'' &= H_{ij}\boldsymbol{p} + \boldsymbol{t}_{ij} \\ &= H_jH_i\boldsymbol{p} + 2d_j\boldsymbol{n}_j + 2H_jd_i\boldsymbol{n}_i . \end{aligned} \quad (12)$$

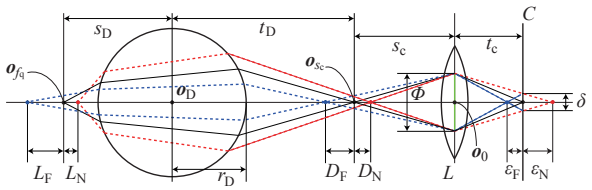

Figure 5. Depth-of-Field with Monocentric Lens. D_F and D_N denote the depth-of-field of the camera itself without the monocentric lens. The effective depth-of-field of the entire system L_F and L_N can be obtained by projecting D_N and D_F through the monocentric lens (the red and blue dashed lines).

As a result, our teleidoscopic imaging system can be modeled as multi-view spherical refraction ray-pixel cameras. They share the single monocentric lens, and the pose of each virtual cameras can be computed by Eq. (12) if the mirror parameters are given. The later section describes our calibration algorithm to estimate such parameters.

4. Depth-of-Field of Teleidoscopic Imaging System

This section describes analytical evaluations on the depth-of-field of our teleidoscopic imaging system. We first review the depth-of-field of monocentric lens.

4.1. Depth-of-Field with Monocentric Lens

As shown in Fig. 5, depth-of-field with a monocentric lens can be simply defined as an extension of the path of the thin lens camera (*cf*. supplemental material). This is because the points in the near and the far depth-of-field D_N and D_F of thin lens are projected within the permissible circle-of-confusion even though the monocentric lens itself introduces the spherical aberration. Hence, we define the depth-of-field with a monocentric lens L_N and L_F as the intersection of the backprojection path through the edge of aperture (the red and blue dash lines in Fig. 5) and the backprojection path through the lens center $\boldsymbol{o}_0$.

Because of the symmetric structure of the monocentric lens, the depth-of-field $L_N + L_F$ can be deeper than the original depth-of-field $D_N + D_F$ if the subject distance of the monocentric lens $s_D - r_D$ is larger than its focal distance $t_D - r_D$. That is, there is a trade-off between magnification and depth-of-field even in a monocentric lens. The reason for the deep depth-of-field is due to the significant contribution of the monocentric lens to the far depth of field L_F, and thus imaging with less blurring even when designed for high magnification (*cf*. Sec. 6.3).

4.2. Depth-of-Field of Teleidoscopic Imaging System

Fig. 6 illustrates the near and the far depth-of-fields of a virtual camera C_i defined as a mirror of the original camera.

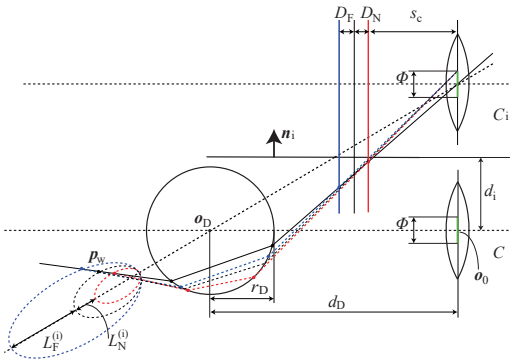

Figure 6. Depth of Field of Teleidoscopic Imaging System. D_N and D_F denote the near and far depth-of-field of the thin lens camera C respectively. All the points once imaged within D_N-D_F should be originally emitted from the points within L_N-L_F.

Due to the aperture, the near and the far depth-of-fields form a curve respectively (the red and the blue dashed lines).

Since the teleidoscopic imaging system has multiple virtual cameras as shown in Fig. 4, each of them has a different depth-of-field according to their relative pose to the monocentric lens. That is, the intersections of such depth-of-fields can be used for multi-view stereo reconstruction.

5. Calibration of Teleidoscopic Imaging System

This section introduces our calibration algorithm of teleidoscopic imaging system which requires capturing a single reference planar patterns. Fig. 7 shows the measurement model where the real camera C observes reference points $\boldsymbol{p}_w$ such as chessboard corners on a reference board Π via a monocentric lens and three mirrors. Our calibration estimates the mirror normals $\boldsymbol{n}_i$, their distances d_i from the camera, and the position and the radius of the monocentric lens, with assuming that the 2D positions of the reference points $\boldsymbol{p}_w$ on the reference plane Π, the intrinsic parameters of the camera, and the refraction index of the lens are given beforehand.

A challenge in this calibration is the fact that the mirrors require observing 3D points and their reflections to estimate their poses [33], while the observation in the teleidoscopic system does not include such mirrored points that follow Eq. (12) due to the refraction by the monocentric lens. Similarly, the rays reflected and then refracted through the projection in teleidoscopic imaging system do not satisfy the coplanarity constraint. In addition, as mentioned in [2], the estimation of the monocentric lens parameters requires multiple viewpoints or multiple monocentric lenses. These points result in a chicken-and-egg problem.

To solve this problem, we utilize the fact that the positions of the centers of the monocentric lens and its mirrors

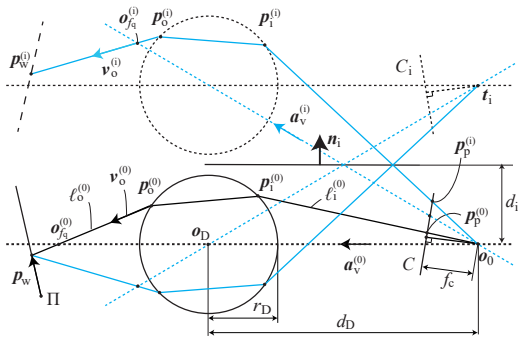

Figure 7. Calibration of Teleidoscopic Imaging System. Cyan lines denote the light paths for a mirrored camera. The calibration can be conducted by capturing reference points on a plane of a known geometry through the monocentric lens.

are always captured without refraction by definition, and satisfy Eq. (12). Therefore, we start by estimating the center of the monocentric lens from the captured reference points.

In what follows, $\boldsymbol{p}_p^{(0)}$ denotes the real image of a reference point $\boldsymbol{p}_w$ in the camera C. Similarly $\boldsymbol{p}_p^{(i)}$ $(i = \{1,2,3\})$ denotes the image of its first reflection by the mirror i, and $\boldsymbol{p}_p^{(ij)}$ $(ij = \{12,13,21,23,31,32\})$ denotes its second reflection by the mirrors i and j.

Axis of Monocentric Lens The axis to the center of monocentric lens $\boldsymbol{a}_v^{(0)}$ from the camera C can be estimated by $\boldsymbol{p}_w$-$\boldsymbol{p}_p^{(0)}$ correspondences. Similarly to [2], we estimate $\boldsymbol{a}_v^{(0)}$ by the coplanarity constraint of the three vectors $\boldsymbol{a}_v^{(0)}$, $\boldsymbol{p}_p^{(0)}$, and $\boldsymbol{p}_w$ on a single plane-of-refraction:

$$\begin{aligned} \boldsymbol{p}_p^{(0)\top}\left(\boldsymbol{a}_v^{(0)} \times (R_\Pi \boldsymbol{p}_w + \boldsymbol{t}_\Pi)\right) = 0\,, \\ \Leftrightarrow \boldsymbol{p}_p^{(0)\top}\left(E_\Pi \boldsymbol{p}_w + \boldsymbol{s}_\Pi\right)) = 0\,, \end{aligned} \tag{13}$$

where $R_\Pi, \boldsymbol{t}_\Pi$ are the pose of the reference plane in the camera coordinate system and $E_\Pi = \boldsymbol{a}_v^{(0)} \times R_\Pi$ and $\boldsymbol{s}_\Pi = \boldsymbol{a}_v^{(0)} \times \boldsymbol{t}_\Pi$. Since this is a linear equation with 9 unknown parameters of E_Π and $\boldsymbol{s}_\Pi$, we can obtain E_Π and $\boldsymbol{s}_\Pi$ up to scale by observing at least 8 points on the reference plane. Using the estimated E_Π, the axis $\boldsymbol{a}_v^{(0)}$ is given by

$$\boldsymbol{a}_v^{(0)} = \frac{E_\Pi(:,1) \times E_\Pi(:,2)}{||E_\Pi(:,1) \times E_\Pi(:,2)||}\,. \tag{14}$$

Similarly, the axis to the center of the mirrored monocentric lens $\boldsymbol{a}_v^{(i)}$ can be obtained by the mirrored points $\boldsymbol{p}_p^{(i)}$.

Mirror Normals The axis to the center of the monocentric lens $\boldsymbol{a}_v^{(0)} = (x_{a_v}^{(0)}, y_{a_v}^{(0)}, z_{a_v}^{(0)})^\top$ and its mirror $\boldsymbol{a}_v^{(i)} = (x_{a_v}^{(i)}, y_{a_v}^{(i)}, z_{a_v}^{(i)})^\top$ satisfies

$$\boldsymbol{a}_v^{(0)\top} [\boldsymbol{n}_i]_\times \boldsymbol{a}_v^{(i)} = 0\,, \tag{15}$$

where $[\boldsymbol{n}_i]_\times$ denotes the skew-symmetric matrix defined by the normal $\boldsymbol{n}_i = (x_{n_i}, y_{n_i}, z_{n_i})^\top$ of the mirror i.

The same constraint holds for each of the first-second reflection pairs $\boldsymbol{a}_v^{(i)}$-$\boldsymbol{a}_v^{(ij)}$ about the same mirror normal $\boldsymbol{n}_i$ [33]. Therefore $\boldsymbol{n}_i$ can be obtained linearly only from the axes to the centers of the monocentric lenses.

Mirror Distances Once the mirror normals are estimated, we can utilize the kaleidoscopic triangulation [33] to obtain linear constraints on the mirror distances d_i (See supp. Sec. 2 for details).

Pose of Reference Plane Similarly to Eq. (13), the plane-of-refraction constraint holds for the mirrored cameras C_i:

$$(\boldsymbol{p}_p^{(i)})^\top \left(\boldsymbol{a}_v^{(i)} \times (H_i(R_\Pi \boldsymbol{p}_w + \boldsymbol{t}_\Pi) + \boldsymbol{t}_i)\right) = 0\,. \tag{16}$$

This constraint allows us estimating the pose of the reference plane R_Π and $\boldsymbol{t}_\Pi$ linearly.

Monocentric Lens Parameters The calibration algorithm up to this point does not require the monocentric lens parameters d_D, r_D, and μ_g. We estimate these parameters by the coplanarity constraint of the ray through $\boldsymbol{p}_p^{(i)}$ of C_i:

$$\boldsymbol{v}_o^{(i)} \times (\boldsymbol{p}_w^{(i)} - \boldsymbol{p}_o^{(i)}) = 0\,. \tag{17}$$

This is a nonlinear constraint for the monocentric lens parameters as described in Sec. 3.1 and we solve this as a nonlinear optimization problem with assuming their rough estimates are available in practice.

Bundle Adjustment The last step of our calibration is to refine the parameters $\boldsymbol{a}_v^{(0)}, d_D, r_D, \mu_g, \boldsymbol{n}_i, d_i, R_\Pi, \boldsymbol{t}_\Pi$ $(i = 1,2,3)$ by minimizing the reprojection errors of the reference points $\boldsymbol{p}_w$ as a nonlinear optimization problem. Here the forward projection is computed using the method described in Sec. 3.1.2.

Once the teleidoscopic imaging system parameters are calibrated, a scene point can be linearly triangulated in a DLT manner [13] from its teleidoscopic projections (See supp. Sec. 3 for details).

6. Evaluation

This section provides quantitative evaluations of the proposed numerical 3D-2D projections, the robustness of the calibration, and the depth of field of our teleidoscopic imaging system. We also demonstrate the effectiveness of our method by reconstructing real-world microscale objects.

6.1. Computational Efficiency of 3D-2D Projection

The average computation cost of a single forward projection by our method and by the analytical solution solving a 10th-degree equation for each projection [2] are 0.62 ms and 3.93 ms respectively. These runtime costs are the average values of 100 trials of 10K points forward projections

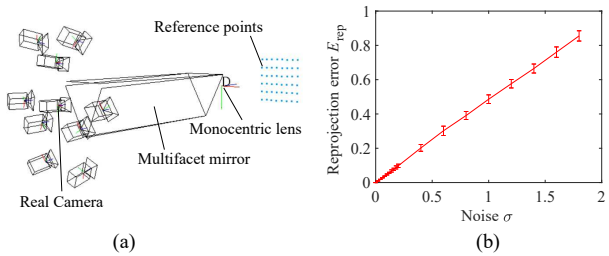

Figure 8. Teleidoscopic System for Quantitative Evaluation. (a) Simulated measurement environment. (b) Reprojection errors at different noise levels. The bars denote the standard deviation of the errors.

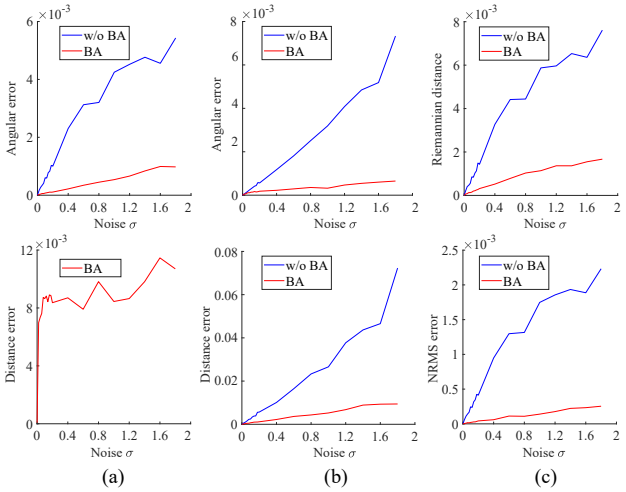

Figure 9. Calibration Errors. From left to right, the upper figures show the (a) angular error of the axis to the monocentric lens center (degree), (b) angular error of the mirror normal (degree), and (c) rotation error (Riemannian distance) of reference plane. The lower ones are the distance errors normalized by the ground truth. Here the distance of the monocentric lens is only estimated in a nonlinear manner (Sec. 5).

run in Matlab on an Intel Core-i7 2.6GHz PC. This result clearly verifies that our numerical forward projection based on our ray-pixel model can realize an efficient projection.

6.2. Calibration

6.2.1 Quantitative Evaluation using Synthesized Data

Fig. 8(a) shows the measurement environment which simulates the real capture system used in Sec. 6.2.2. The system has a kaleidoscope with three 10 × 30mm mirrors in front of the camera C. The mirrors are at slightly off-perpendicular angle of 1.4° to the camera image plane so that the mirrors define virtual cameras around the target with less overlaps of the mirrored images. The system also has a monocentric lens of 10mm diameter in front of the mirrors, at 40mm distance from the camera. The refraction index μ_g is set to 2.0. The system captures 48 reference points (blue dots in Fig. 8(a)) to calibrate its parameters. By injecting Gaussian

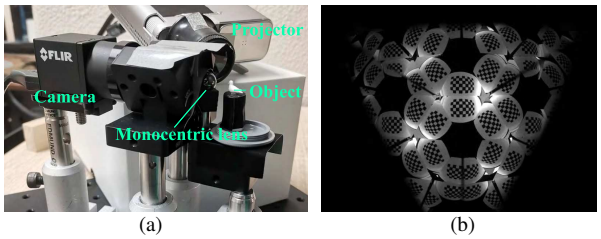

Figure 10. Experimental Setup and Calibration. (a) Our teleidoscopic imaging system consists of a camera, a projector, three mirrors (10 × 30mm), and a monocentric lens (5mm radius, refraction index $\mu_g = 2.0$). (b) Input image for calibration.

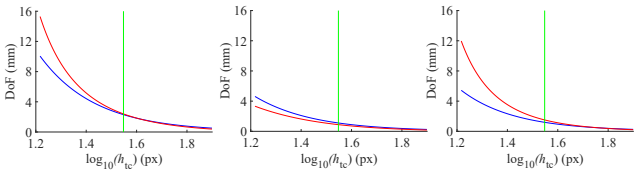

Figure 11. Changes in Depth-of-Field by the Monocentric Lens. (a), (b) and (c) show the total, near, and far depth-of-field with a monocentric lens. The red and blue plots indicate the results with and without the monocentric lens respectively. The green lines indicate the subject distance where the subject and the focal distances of the monocentric lens itself are same.

noise of different standard deviations σ to the 2D positions of their projections, we evaluate the robustness of our calibration procedure. Fig. 8(b) shows average reprojection errors in pixel of 100 trials at each pixel noise level σ. We can observe that the reprojection errors increase linearly against the pixel noise level.

Fig. 9 shows the estimation errors of the monocentric lens parameters, the mirror parameters, and the reference plane parameters respectively. These results indicate that our calibration algorithm performs reasonably under realistic observation noise.

6.2.2 Quantitative Evaluation using Real Data

Fig. 10(a) shows an overview of our teleidoscopic imaging system. The system consists of a FLIR Blackfly BFS-U3-89S6C-C camera (4096×2160 resolution, pixel size 3.45μm) with a f=4mm / F8 lens, three 10 × 30mm mirrors, a monocentric lens of 10mm diameter whose refraction index is $\mu_g = 2.0$, and a LED projector ASUS E1-R for casting Gray code patterns onto the target surface to identify corresponding points [12, 23]. Note that the effective resolution of each "virtual camera" image is approximately 300×200 (16 "virtual cameras" in total), which is calculated from the area inside the base chamber with one horizontal edge of the triangle and inside the spherical lens.

We used a single image (Fig. 10(b)) capturing 5×8 corners of a checkerboard for calibration, and used up to second reflections in the image. The average reprojection er-

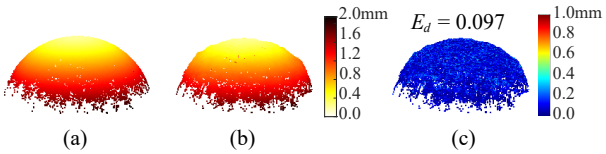

Figure 12. Evaluation with a sphere (radius: 3mm). (a) Ground truth. (b) Recovered 3D shape. The colors visualize the z-coordinate of the reconstructed shape. (c) Error map. E_d is the mean absolute error in millimeter.

ror is 0.62px, and the average distance error of the triangulated 3D points to their ground truth positions was 0.17mm. The size of reference checkerboard is around 15mm width, hence the relative error of the triangulation is around 1%. These result shows our method provide a practical calibration and the valid scale of the target object also depends on size of the checkerboard.

6.3. Depth-of-Field

Fig. 11 shows the changes in the depth-of-field by the monocentric lens capturing an object of 0.5mm height at the same magnification given the configuration same as the real capture in Sec. 6.2.2. Fig. 11(a) shows the total depth-of-field with and without the monocentric lens $L_N + L_F$ (red and blue respectively). Fig. 11(b) and (c) show L_N and L_F in the same manner. In these figures, the horizontal axis indicates the apparent size of the target in pixel, with assuming the measurement model used in Sec. 6.2. These plots clearly indicate that the monocentric lens contributes to increase the depth-of-field successfully.

6.4. Teleidoscopic 3D Shape Reconstruction

To validate the proposed teleidoscopic system as a multi-view camera system for 3D shape reconstruction, this section demonstrates 3D reconstruction of small objects of approximately 6mm size.

We used the system calibrated in Sec. 6.2.2. Fig. 12 visualizes our 3D reconstruction of the front side of a real sphere of 3mm radius with the ground-truth and the error map. The mean absolute error E_d of the estimated depth is 0.097mm. Compared with its size, the relative error is about 1.6%.

Fig. 13 shows additional 3D reconstructions. The 1st and the 2nd rows demonstrate that our imaging system has the ability to recover complex shapes. The proposed imaging system well recovers the complex bumpy parts of the shell and the sharp edges and corners of the object. The 2nd and the 3rd rows compare "with" and "w/o" the monocentric lens cases. As the input image (b) shows, the monocentric lens creates a surrounding view while preserving the area of each chamber, resulting in a fuller and more accurate geometry, as shown in (c). We can verify this quantitatively by comparing the numbers of pixels covering the object area in the captured image (145,217 *vs*. 12,730) and the numbers

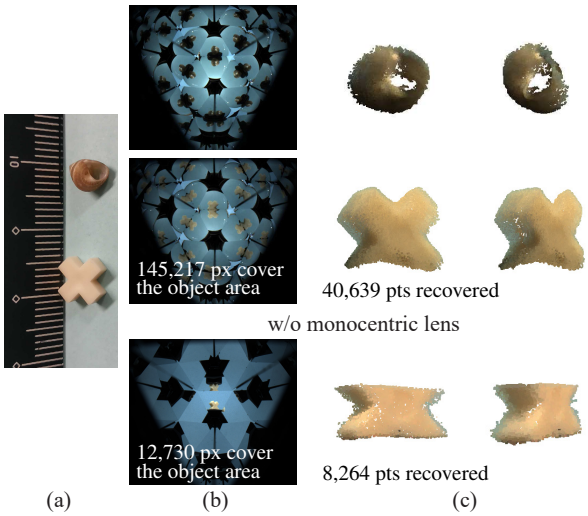

Figure 13. Results of 3D Shape Estimation. The second and the third row compares "with" and "w/o" the monocentric lens. (a) Room-light appearance. (b) Teleidoscopic image. (c) Reconstructed 3D shapes.

of triangulated 3D points using the ground-truth correspondences as shown in the figure (40,639 *vs*. 8,264). See the supplementary material for more results.

Limitations The major limitations of our method are twofold. The monocentric lens has a relatively large spherical aberration, and the optical resolution of the entire system is limited by such aberrations. One possible solution is to introduce nested concentric glass spheres with different refractive indices for the monocentric lens [6, 35]. The other limitation is the trade-off between the number of virtual viewpoints and the effective resolution of each view because the use of a kaleidoscope requires zooming-out to capture the reflected images. While larger number of the cameras contributes to robustify the triangulation, it inevitably reduces the resolution of the reconstruction as the number of pixels per view decreases.

7. Conclusion

This paper proposed the concept of image-based 3D measurement by a teleidoscopic imaging system for microscale objects. By combining a monocentric lens with a kaleidoscope, our system realized a closeup and surrounding multi-view capture of the target. Our contributions are twofold. We proposed a ray-pixel camera model that realized an efficient numerical 3D-2D projection computation, and also proposed a calibration method for the teleidoscopic imaging system. Our future work includes aberration and diffraction blur handling such as the inverse filter proposed in [5], and also refraction/attenuation modeling in the case of capturing objects in water [18, 19].

Acknowledgements

This work is in part supported by JSPS KAKENHI JP22K17914.

References

[1] Rajat Aggarwal, Amrisha Vohra, and Anoop M. Namboodiri. Panoramic stereo videos with a single camera. In *CVPR*, pages 3755–3763, 2016. 2

[2] Amit Agrawal and Srikumar Ramalingam. Single image calibration of multi-axial imaging systems. In *CVPR*, pages 1399 –1406, 2013. 2, 3, 4, 5, 6

[3] Byeongjoo Ahn, Ioannis Gkioulekas, and Aswin C Sankaranarayanan. Kaleidoscopic structured light. *ACM TOG*, 40(6):1–15, 2021. 2

[4] Michael Chen, Zachary F. Phillips, and Laura Waller. Quantitative differential phase contrast (dpc) microscopy with computational aberration correction. *Optics Express*, (25):32888–32899. 1

[5] O. S. Cossairt, D. Miau, and S. K. Nayar. Gigapixel computational imaging. In *ICCP*, pages 1–8, 2011. 2, 8

[6] Donald G. Dansereau, Glenn Schuster, Joseph Ford, and Gordon Wetzstein. A wide-field-of-view monocentric light field camera. In *CVPR*, 2017. 2, 8

[7] Michael De Zeeuw and Aswin C Sankaranarayanan. Wide-baseline light fields using ellipsoidal mirrors. 2022. 1

[8] Keith Forbes, Fred Nicolls, Gerhard De Jager, and Anthon Voigt. Shape-from-silhouette with two mirrors and an uncalibrated camera. In *ECCV*, 2006. 2

[9] Brian J. Ford. Single lens: The story of the simple microscope. *Journal of the History of Biology*, 19(2):320–321, 1986. 2

[10] Joshua Gluckman, Shree K. Nayar, and Keith J. Thoresz. Real-time omnidirectional and panoramic stereo. In *Proceedings of the 1998 DARPA Image Understanding Workshop*, pages 299–303. Morgan Kaufmann, 1998. 2

[11] Michael D. Grossberg and Shree K. Nayar. The raxel imaging model and ray-based calibration. *IJCV*, 61(2):119–137, 2005. 1, 2

[12] Mohit Gupta, Amit Agrawal, Ashok Veeraraghavan, and Srinivasa G. Narasimhan. Structured light 3d scanning in the presence of global illumination. In *CVPR*, pages 713–720, 2011. 7

[13] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000. 2, 6

[14] Hanbyul Joo, Tomas Simon, Xulong Li, Hao Liu, Lei Tan, Lin Gui, Sean Banerjee, Timothy Godisart, Bart Nabbe, Iain Matthews, Takeo Kanade, Shohei Nobuhara, and Yaser Sheikh. Panoptic studio: A massively multiview system for social interaction capture. *IEEE TPAMI*, 41(1):190–204, 2019. 1

[15] Srinivasa Karthik and Ambasamudram N. Rajagopalan. Underwater microscopic shape from focus. In *ICPR*, pages 2107–2112, 2014. 1

[16] Ryo Kawahara, Shohei Nobuhara, and Takashi Matsuyama. A pixel-wise varifocal camera model for efficient forward projection and linear extrinsic calibration of underwater cameras with flat housings. In *Proc. of ICCV 2013 Underwater Vision Workshop*, pages 819–824, 2013. 2

[17] Gurunandan Krishnan and Shree K. Nayar. Towards a true spherical camera. In *SPIE*, volume 7240, 724002, pages 1–13, 2009. 2

[18] Meng-Yu Jennifer Kuo, Ryo Kawahara, Shohei Nobuhara, and Ko Nishino. Non-rigid shape from water. *IEEE TPAMI*, 43(7):2220–2232, 2021. 8

[19] Meng-Yu Jennifer Kuo, Satoshi Murai, Ryo Kawahara, Shohei Nobuhara, and Ko Nishino. Surface normals and shape from water. *IEEE TPAMI*, 44(12):9150–9162, 2021. 8

[20] Douglas Lanman, Daniel Crispell, and Gabriel Taubin. Surround structured lightning: 3-d scanning with orthographic illumination. In *CVIU*, pages 1107–1117, 2009. 2

[21] A. Manakov, J. Restrepo, O. Klehm, R. Hegedus, E. Eisemann, H. P. Seidel, and I. Ihrke. A reconfigurable camera add-on for high dynamic range, multispectral, polarization, and light-field imaging. *ACM TOG*, 32(4):47–1, 2013. 1

[22] P. Miraldo and H. Araujo. Calibration of smooth camera models. *IEEE TPAMI*, 35(9):2091–2103, 2013. 2

[23] Daniel Moreno and Gabriel Taubin. Simple, accurate, and robust projector-camera calibration. In *3DPVT*, pages 464–471, 2012. 7

[24] Andrew D. Mullen, Tali Treibitz, Paul LD Roberts, Emily L. A. Kelly, Rael Horwitz, Jennifer E. Smith, and Jules S. Jaffe. Underwater microscopy for in situ studies of benthic ecosystems. *Nature Communications*, 7, 2016/07 2016. 1

[25] Shree K. Nayar. Omnidirectional video camera. In *Proceedings of the 1997 DARPA Image Understanding Workshop*, pages 235–241, 1997. 2

[26] Markus Niederöst, Jana Niederöst, and Jiri Scucka. Automatic 3d reconstruction and visualization of microscopic objects from a monoscopic multifocus image sequence. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXIV-5/W10:5–10, 01 2002. 1

[27] Ryoichiro Nishi, Takahito Aoto, Norihiko Kawai, Tomokazu Sato, Yasuhiro Mukaigawa, and Naokazu Yokoya. Ultra-shallow dof imaging using faced paraboloidal mirrors. In *ACCV*, pages 115–128, 2016. 2

[28] Ilya Reshetouski and Ivo Ihrke. *Mirrors in Computer Graphics, Computer Vision and Time-of-Flight Imaging*, pages 77–104. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. 2

[29] Ilya Reshetouski, Alkhazur Manakov, Hans-Peter Seidel, and Ivo Ihrke. Three-dimensional kaleidoscopic imaging. In *CVPR*, pages 353–360, 2011. 4

[30] J. Starck, A. Maki, S. Nobuhara, A. Hilton, and T. Matsuyama. The multiple-camera 3-d production studio. *IEEE TCSVT*, 19(6):856 –869, 2009. 1

[31] Peter Sturm. Multi-view geometry for general camera models. In *CVPR*, volume 1, pages 206–212, 2005. 2

[32] Seiichi Tagawa, Yasuhiro Mukaigawa, Jaewon Kim, Ramesh Raskar, Yasuyuki Matsushita, and Yasushi Yagi. Hemispherical confocal imaging. *IPSJ Transactions on Computer Vision and Applications*, 3:222–235, 2011. 2

[33] Kosuke Takahashi, Akihiro Miyata, Shohei Nobuhara, and Takashi Matsuyama. A linear extrinsic calibration of kaleidoscopic imaging system from single 3d point. In *CVPR*, 2017. 1, 2, 4, 5, 6

[34] Lei Tian, Jingyan Wang, and Laura Waller. 3d differential phase-contrast microscopy with computational illumination using an led array. *Optics letters*, 39:1326–9, 2014. 1

[35] Eric J Tremblay, Daniel L Marks, David J Brady, and Joseph E Ford. Design and scaling of monocentric multiscale imagers. *Applied Optics*, 51(20):4691–4702, 2012. 2, 8

[36] Borislav Trifonov, Derek Bradley, and Wolfgang Heidrich. Tomographic reconstruction of transparent objects. In *Proc. of Eurographics Conference on Rendering Techniques*, pages 51–60, 2006. 2

[37] David Uhlig and Michael Heizmann. A calibration method for the generalized imaging model with uncertain calibration target coordinates. In *ACCV*. 2

[38] Frank Verbiest, Marc Proesmans, and Luc Van Gool. Modeling the effects of windshield refraction for camera calibration. In *ECCV*, pages 397–412, 2020. 2