A Novel Catadioptric Ray-Pixel Camera Model and its Application to 3D Reconstruction

Ryo Kawahara[1,a)] Shohei Nobuhara[1,b)]

Abstract: This paper proposes a new ray-pixel camera model for image-based 3D measurement by a catadioptric imaging system. The key idea of this paper is to employ a virtual camera model that describes ray-pixel mappings with exploiting an axially-symmetric structure of the ray distribution of the system. Our contributions include, structured ray-pixel camera models which handle refractive and reflective projection rays efficiently, and practical calibration algorithms for them. Evaluations with real images prove the concept of our measurement system.

1. Introduction

3D shape acquisition has been an important topic in computer vision as an essential factor for interacting with real world, and there are a large number of studies in particular for capturing human[1], [2], [3] and buildings[4], [5]. Applications of unconstrained and noninvasive image-based 3D shape capturing include digital archiving, navigation in surgery, SLAM, industrial inspection, virtual reality, surveying and measurement, etc.

Most of the 3D shape capture studies in literature utilize regular perspective cameras. Catadioptric system with additional lenses and mirrors, however, can also be a practical solution for particular targets and scenes. For example, catadioptric system has been widely studied for panoramic imaging[6], [7]. Besides, in the case of underwater 3D capture, measurement through waterproof housings or aquarium surfaces can also be considered as catadioptiric system.

For such catadioptric measurements, the rays captured by the system do not form a pencil due to reflections and refractions, and the optical system as a whole cannot be modeled as perspective. Therefore the 3D recovery via triangulation becomes a non-trivial process as a result. The goal of this paper is to propose a camera model that can handle such projections in catadioptric system efficiently.

Our approach is to employ a ray-pixel (*raxel*) camera model which focuses on the mapping from each pixel to the scene ray[8]. It implements 2D-3D projection just by storing the pixel-ray correspondences without tracing the optical path between them, and therefore it is flexible enough to model the light-field captured by the catadioptric system.

The general ray-pixel representation, however, cannot provide 3D-2D projection in a straightforward manner. Given a 3D point to be projected, it needs to find a ray stored in the model that intersects with the point. This can be an exhaustive search if the rays are unstructured, while the 3D-2D projection is an essential process in 3D recovery in handling occlusion and view dependent color consistency analysis as done in space carving[9] for example.

The key idea of this paper is to propose a structured ray-pixel model which focuses on modeling the distribution of the rays in catadioptric systems. For modeling a general distribution of the rays, Grossberg and Nayar utilized a unique ray-surface called caustics [8], to which all the incoming scene rays tangent. In contract, to realize a compact description and an efficient 3D-2D projection, we exploit an axially-symmetric structure of the rays and employ a virtual camera model to describe such a 1D pixel-to-ray mapping as a relationship between a virtual focal length and a virtual pixel.

Based on our new camera model, we propose two practical catadioptric systems for 3D measurement, one for an underwater active stereo with flat water-proof housings, and the other for a microscale 3D reconstruction with teleidoscopic system. In the first system, projectors are utilized as reverse cameras[10], [11], [12] in order to improve the stereo matching for poorly-textured underwater objects by attaching artificial textures onto the target surface. We show our ray-pixel camera can correctly handle the underwater scene ray geometry and show that our efficient 3D-2D projection realizes a practical 3D capture of underwater objects such as swimming fish.

The second teleidoscopic system has three planar mirrors and a monocentric lens similarly to teleidoscopes. The planar mirrors virtually define multiple viewpoints, and the monocentric lens realizes a high magnification with less blurry and surround view even in closeup imaging. We show our camera model can handle the refractive and reflective projection of the rays and show that the system realizes a 3D shape capture of microscale objects.

2. Related Work

Catadioptric system have been utilzed for capturing objects in

[1] Graduate School of Informatics, Kyoto University
[a)] kawahara@vision.ist.i.kyoto-u.ac.jp
[b)] nob@i.kyoto-u.ac.jp

a wide range of scales and media, such as capturing objects in water. Several studies have been proposed for handling scattering or absorption[13], [14], [15], [16], [17], [18], [19], refraction at water surface boundary[20], [21], [22] or refraction by housing of underwater camera[23], [24], [25], [26], [27], [28], [29], [30].

3D shape capturing with refraction

As as well known, Snell's law describes the refraction process at the boundary of medium. Therefore, the process of observing underwater scene by a perspective camera via refractive medium such as waterproof housing can be expressed by tracing the light paths by the law of refraction[24], [25], [26]. This is an analytical process, however, its 3D-2D projection is defined as a solution of a 12th-degree polynomial[24] and is time-consuming for dense 3D geometry recovery.

On the other hand, the ray-pixel approach proposed by Grossberg and Nayar describes such projection process as a mapping between them regardless of the intermediate projection process[8]. While this pixel-wise representation has a great flexibility to describe complicated projections[31], [32], [33], [34], 3D-2D projection cannot be provided explicitly. That is, it requires finding a ray in the mapping function that intersects with the 3D point in question.

Therefore, in both analytical and general ray-pixel approaches, 3D-2D projections involve a time-consuming numerical optimization process. Instead of directly using 3D-2D projection, Sedlazeck and Koch propose a virtual camera whose projection centers are on a ray-surface called caustic, and they provide 2D-3D based reprojection error modeling for handing refraction[27], [35].

3D shape capturing with mirrors

While catadioptric system has a wide variety of applications such as omnidirectional observation and panoramic stereo[6], [7], [36], [37], the following reviews studies 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[38], [39], [40], [41], [42], [43]. Takahashi *et al.* [40] have proposed a kaleidoscopic imaging system and demonstrated a 3D shape reconstruction using multiple reflections. Tagawa *et al.* [42] have proposed a multi-facet imaging system for that observes a target from an equally distributed virtual cameras reflectance analysis.

These studies can be categorized into two groups: virtual camera approaches or virtual object approaches. The former utilizes the mirrors to define virtual cameras capturing the original object from different directions. The latter considers that the mirrors define virtual objects in the original camera image. While the latter allows modeling the entire light field by a single ray-pixel camera, we follow the former approach to exploit an axially-symmetric structure found in the rays of each virtual view.

3D shape capturing with monocentric lens

In the context of imaging system, the monocentric lens is often used to obtain a wide field-of-view[44], in particular for endoscopes, as a short focal length lens. The monocentric lens, however, has additional useful characteristics: its symmetric structure

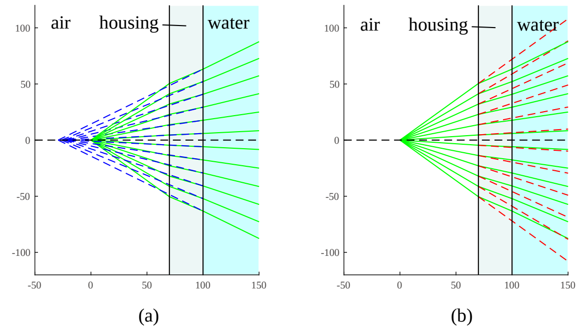


Fig. 1 Refraction caused by flat housing. Green lines illustrates the actual projection path. (a) Blue dashed lines denote the straightly extended rays in water and form a caustic structure. (b) Red dashed lines denote the rays which are straightly back-projected ignoring the refraction and obviously lead to a wrong measurement.

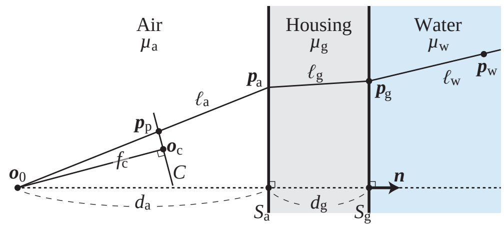


Fig. 2 Measurement model through planar refraction. A point $\boldsymbol{p}_w$ in water is projected to $\boldsymbol{p}_p$ and $\boldsymbol{o}_0$ along the segments ℓ_w-ℓ_g-ℓ_a. ©2016, Elsevier [28].

and magnifying power.

By exploiting the symmetric structure of the monocentric lens, Cossairt *et al.* [45] 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.* [46] 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× magnification[47].

3. 3D Reconstruction through Planar Refraction

Projection paths to a perspective camera via a flat housing is shown in **Fig. 1**-(a). Unlike the case in the air, the ray in water is not described as straight line from a common projection center because of the refraction. This fact indicates that conventional triangulation methods in the air cannot return the correct location and that refraction effects cannot simply be modeled by radial distortions.

The key idea of our approach on modeling such ray distribution is to exploit the axially-symmetric structure of the ray distribution by the flat housing.

3.1 Measurement through Planar Refraction

Measurement model through planar refraction is shown in **Fig. 2**. A pinhole camera C at $\boldsymbol{o}_0$ observes the underwater scene via flat housing. A point $\boldsymbol{p}_w$ in water is projected to $\boldsymbol{p}_p$ and $\boldsymbol{o}_0$ along the segments ℓ_w-ℓ_g-ℓ_a.

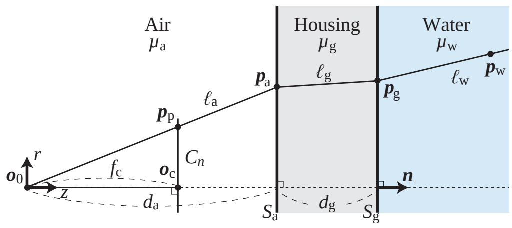


Fig. 3 Axial measurement model. The segments ℓ_w-ℓ_g-ℓ_a and housing normal $\boldsymbol{n}$ have axially symmetric structure around the normal $\boldsymbol{n}$. ©2016, Elsevier [28].

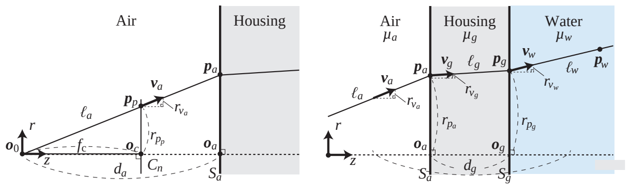


Fig. 4 Light path in axial measurement model. Left: the light path ℓ_a is derived from the observed point $\boldsymbol{p}_p$. Right: the segments $\ell_a - \ell_g - \ell_w$ is derived from ℓ_a by Snell's law. ©2016, Elsevier [28].

Assuming that the two surfaces S_a and S_g of the housing are flat and parallel, the segments ℓ_w-ℓ_g-ℓ_a and housing normal $\boldsymbol{n}$ are always on a single plane-of-refraction and have axially symmetric structure around the normal $\boldsymbol{n}$. Hence, we employ the axial camera concept proposed by Agrawal *et al.* [24] to simplify this model without loss of generality.

Consider a virtual camera C_n such that its projection center is placed at $\boldsymbol{o}_0$ and its direction of optical axis is identical to the normal $\boldsymbol{n}$ of the flat housing (**Fig. 3**). If the pose of camera C w.r.t. the housing is calibrated beforehand, the mapping from a pixel of virtual camera C_n to the corresponding pixel of C is given by a homography matrix H_C derived from the intrinsic and the pose of C, where we simply assume the cameras C and C_n share the same intrinsic parameter A and assume the radial distortion caused by the internal lens of the camera is already rectified.

Therefore, instead of C, we can use C_n which has a radially symmetric structure of refractive path around the Z-axis without loss of generality. In the coordinate of C_n, any continuous segments ℓ_w-ℓ_g-ℓ_a are observed as a single line on the image plane, because the segments are on a single plane-of-refraction. Hence we employ the $(r, z)^\top$ coordinate system hereafter.

Let r_α and z_α denote the r and z elements of 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 water from the camera, where X is each medium (Fig. 3).

The light path ℓ_a-ℓ_g-ℓ_w follows Snell's law which is expressed as $\mu_a r_{\boldsymbol{v}_a} = \mu_g r_{\boldsymbol{v}_g} = \mu_w r_{\boldsymbol{v}_w}$, where μ_a, μ_g, and μ_w are the refractive indices of the air, housing and water. Using Snell's law, the light path ℓ_a-ℓ_g-ℓ_w is given as

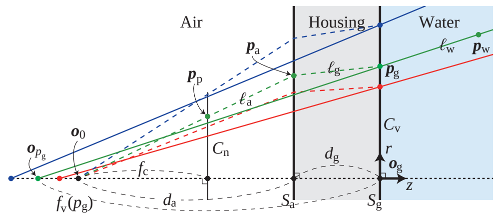


Fig. 5 Planar refraction ray-pixel camera model. The dashed lines illustrate the correct refractive paths in three random samples (red, green, blue). The straight lines illustrate the perspective projections by the virtual focal length f_v and virtual pixel $\boldsymbol{p}_g$ ©2016, Elsevier [28].

$$\boldsymbol{v}_a = \left(r_{\boldsymbol{p}_p} / \sqrt{r_{\boldsymbol{p}_p}^2 + f_c^2},\ f_c / \sqrt{r_{\boldsymbol{p}_p}^2 + f_c^2} \right)^\top, \tag{1}$$

$$\boldsymbol{p}_a = \frac{d_a}{z_{\boldsymbol{v}_a}} \boldsymbol{v}_a, \tag{2}$$

$$\boldsymbol{v}_g = \left(\frac{\mu_a}{\mu_g} r_{\boldsymbol{v}_a},\ \sqrt{1 - r_{\boldsymbol{v}_g}^2} \right)^\top, \tag{3}$$

$$\boldsymbol{p}_g = \boldsymbol{p}_a + \frac{d_g}{z_{\boldsymbol{v}_g}} \boldsymbol{v}_g, \tag{4}$$

$$\boldsymbol{v}_w = \left(\frac{\mu_g}{\mu_w} r_{\boldsymbol{v}_g},\ \sqrt{1 - r_{\boldsymbol{v}_w}^2} \right)^\top, \tag{5}$$

$$\boldsymbol{p}_w = \boldsymbol{p}_g + \frac{z_{\boldsymbol{p}_w} - z_{\boldsymbol{p}_g}}{z_{\boldsymbol{v}_w}} \boldsymbol{v}_w, \tag{6}$$

where f_c is the focal length of the camera. These equations allow computing direction of a ray in water $\boldsymbol{v}_w$ and a position $\boldsymbol{p}_g$, given a pixel $\boldsymbol{p}_p$ on the image plane. Similarly, computing $\boldsymbol{p}_p$ from $\boldsymbol{v}_w$ can be done by applying Snell's law inversely because $\boldsymbol{v}_a$ can be derived from $\boldsymbol{v}_w$ by Eqs. (3) and (5).

This indicates that the inverse process of Eq. (1)-Eq. (6) requires the direction $\boldsymbol{v}_w$, but in the case of 3D-2D projection, only given is 3D point $\boldsymbol{p}_w$. It requires solving a 12th degree equation and becomes a time-consuming process[24]. To realize a practical projection through planar refraction, we employ a new camera model that exploits the radial structure of the rays in C_n, based on the ray-pixel camera model[8] and Eqs. (1)-(5).

3.2 Planar Refraction Ray-Pixel Camera Model

Let us consider the representation of scene rays ℓ_w. On the assumption of the straightness of light, a scene ray ℓ_w is simply described by a set of a starting point and a direction. That is, when we set an arbitrary starting point $\boldsymbol{o}_{p_w}$ along the ℓ_w, the geometric representation of the ray ℓ_w is defined by the set $(\boldsymbol{o}_{p_w}, \boldsymbol{v}_w)$, where the point $\boldsymbol{o}_{p_w}$ is generally $\mathbb{R}^3$ and the direction $\boldsymbol{v}_w$ is $\mathbb{R}^2$ by two angles similarly to the geometric part of the light field of plenoptic camera by Adelson and Bergen[48]. In addition, the starting point $\boldsymbol{o}_{p_w}$ can be on a surface such as a unique caustic formed by the rays ℓ_w[35] or a 2D plane in front of the camera.

Fig. 5 shows our a new virtual camera model called planar refraction ray-pixel camera model.Notice that each extended rays of ℓ_w intersects a common axis whose direction is identical to the housing normal $\boldsymbol{n}$. This fact indicates that we can compactly represent the ray ℓ_w by the position of the intersection point on the axis.

In particular, we introduce a virtual camera whose image plane is on the outer housing surface S_g associated with a pixel-wise virtual focal length $f_v(\boldsymbol{p}_g) \in \mathbb{R}$ as follows.

Table 1 A pixel-ray mapping for generalized ray-pixel model. A pixel $\boldsymbol{p}_p$ is associated with a point on ray-surface $\boldsymbol{o}_{p_w} \in \mathbb{R}^3$ and the direction $\boldsymbol{v}_w \in \mathbb{R}^2$

pixel	$\boldsymbol{p}_{p0}$	$\boldsymbol{p}_{p1}$	$\cdots$	$\boldsymbol{p}_{pN}$
ray	$(\boldsymbol{o}_{p_w0}, \boldsymbol{v}_{w0})$	$(\boldsymbol{o}_{p_w1}, \boldsymbol{v}_{w1})$	$\cdots$	$(\boldsymbol{o}_{p_wN}, \boldsymbol{v}_{wN})$

Table 2 A pixel-ray mapping of our planar refraction ray-pixel camera model. A pixel $r_{\boldsymbol{p}_g}$ in $(r, z)^\top$ coordinate in C_v is associated with a pixel-wise focal length $f(r_{\boldsymbol{p}_g}) \in \mathbb{R}$. The 1D ordered structure easily provide the differential $\frac{\mathrm{d}f}{\mathrm{d}r_{\boldsymbol{p}_g}}$ for the mapping.

pixel	$r_{\boldsymbol{p}_{g0}}$	$r_{\boldsymbol{p}_{g1}}$	$\cdots$	$r_{\boldsymbol{p}_{gN}}$
ray	$f(r_{\boldsymbol{p}_{g0}})$	$f(r_{\boldsymbol{p}_{g1}})$	$\cdots$	$f(r_{\boldsymbol{p}_{gN}})$
$\frac{\mathrm{d}f}{\mathrm{d}r_{\boldsymbol{p}_g}}$	$f'(r_{\boldsymbol{p}_{g0}})$	$f'(r_{\boldsymbol{p}_{g1}})$	$\cdots$	$f'(r_{\boldsymbol{p}_{gN}})$

- The virtual image screen coincides with the housing surface S_g. The virtual pixel $\boldsymbol{p}_g$ is associated with a real pixel $\boldsymbol{p}_p$ of C by Eq. (4) and the homography H_C between C and C_n.
- The virtual optical axis (Z-axis) is identical to the housing normal $\boldsymbol{n}$.
- The pixel-wise projection center $\boldsymbol{o}_{p_g}$ is defined simply by connecting ℓ_w to the virtual optical axis (Fig. 5, the green straight line). The distance between $\boldsymbol{o}_{p_g}$ and S_g is denoted as the *pixel-wise focal length* $f_v(\boldsymbol{p}_g)$.

From the radially symmetric structure of the ray distribution, the pixel-wise focal length $f_v(\boldsymbol{p}_g)$ can be expressed as a function of the radial distance $r_{\boldsymbol{p}_g}$ of $\boldsymbol{p}_g$ from the optical axis without loss of generality:

$$f_v(\boldsymbol{p}_g) = f_v(r_{\boldsymbol{p}_g}). \tag{7}$$

In addition, it is obvious that $f_v(r_{\boldsymbol{p}_g})$ is a monotonically increasing function of $r_{\boldsymbol{p}_g}$ from Fig. 5.

By the definition, the pixel-wise virtual focal length $f_v(r_{\boldsymbol{p}_g})$ is derived as follows,

$$\boldsymbol{o}_{p_g} = t_o \boldsymbol{v}_w + \boldsymbol{p}_g, \tag{8}$$

$$\begin{pmatrix} 0 \\ -f_v(r_{\boldsymbol{p}_g}) \end{pmatrix} = t_o \begin{pmatrix} r_{\boldsymbol{v}_w} \\ z_{\boldsymbol{v}_w} \end{pmatrix} + \begin{pmatrix} r_{\boldsymbol{p}_g} \\ 0 \end{pmatrix},$$

$$\Leftrightarrow f_v(r_{\boldsymbol{p}_g}) = \left(\frac{r_{\boldsymbol{v}_w}}{z_{\boldsymbol{v}_w}} \right)^{-1} r_{\boldsymbol{p}_g}. \tag{9}$$

These representation leads to the compact description of the rays ℓ_w as a ray-pixel camera model.

3.2.1 The Ray-Pixel Mapping

As shown in Table 1, the generalized ray-pixel camera model such as [8] associates fully-described scene rays in $\mathbb{R}^5$ by the mapping of ray-surface such as caustic. Compared to such generalized ray-pixel camera model, our planar refraction ray-pixel camera model stores focal lengths for each radial distance in a 1D array as shown in Table 2.

In addition, our mapping also has the derivative $\frac{\mathrm{d}f}{\mathrm{d}r_{\boldsymbol{p}_g}}$ (the last row of Table 2), since $f(r_{\boldsymbol{p}_g})$ is a smooth and monotonic function. This $\frac{\mathrm{d}f}{\mathrm{d}r_{\boldsymbol{p}_g}}$ is the key to realize our efficient 3D-2D projection (Section 3.2.3).

In practice the mapping table can be obtained as follows.

- Sample the radial distance $r_{\boldsymbol{p}_g}$ of virtual pixels at equal intervals so that the interval is smaller than 1px in the image plane of corresponding real camera C.
- Compute the pixel-wise virtual focal length f_v given by $r_{\boldsymbol{p}_g}$ and $\boldsymbol{v}_w$ in Eq. (9), where $\boldsymbol{v}_w$ is obtained by using Snell's law (Eq. (4)).

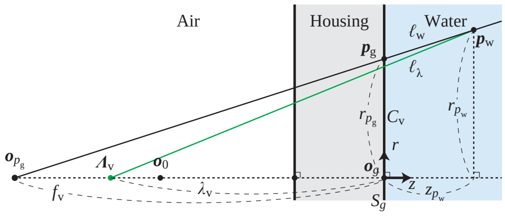


Fig. 6 3D-2D projection of planar refraction ray-pixel camera model. The black line ℓ_w is the correct projection path of the model. The search line ℓ_λ iteratively converges to the correct line ℓ_w by using $r_{\boldsymbol{p}_g}$-f_v correspondences. ©2016, Elsevier [28].

3.2.2 2D-3D Projection

For a pixel $\boldsymbol{p}_g$ on virtual image plane S_g, 2D-3D projection is trivially defined like a perspective projection from virtual focus $\boldsymbol{o}_{p_g}$ which corresponds to $r_{\boldsymbol{p}_g}$ by the Table 2. That is, the 2D-3D projection is represented by using virtual focus $\boldsymbol{o}_{p_g} = (0, -f_v(r_{\boldsymbol{p}_g}))^\top$ as

$$\ell_w : \boldsymbol{o}_{p_g} + t_w \boldsymbol{v}_w = \boldsymbol{o}_{p_g} + t_w \frac{\boldsymbol{p}_g - \boldsymbol{o}_{p_g}}{\|\boldsymbol{p}_g - \boldsymbol{o}_{p_g}\|}, \tag{10}$$

where t_w is the $\boldsymbol{p}_g$-$\boldsymbol{p}_w$ distance.

3.2.3 3D-2D projection

The 3D-2D projection process can be realized by the searching the mapping table for the ray intersecting with the point $\boldsymbol{p}_w$ to be projected. As shown in **Fig. 6**, let us hypothesize a ray ℓ_λ from $\boldsymbol{p}_w$ which intersects with the axis at $\Lambda_v(0, -\lambda_v)$. If the ray ℓ_λ is identical to ℓ_w stored in the mapping table, we can conclude that the hypothesized ray was the correct projection direction from $\boldsymbol{p}_w$. Otherwise, we can refine the hypothesis in the following two approaches.

By Gauss-Newton method

Utilizing the ray-pixel mapping and the derivative of f_v shown in Table2, we can solve the 3D-2D projection by Gauss-Newton method. When the ray ℓ_λ is identical to ℓ_w, following equation is satisfied,

$$G(r_{\boldsymbol{p}_g}) = f_v(r_{\boldsymbol{p}_g}) - \lambda_v(r_{\boldsymbol{p}_g}), \tag{11}$$

where, $G(r_{\boldsymbol{p}_g})$ is a monotonic function and hence we can compute $r_{\boldsymbol{p}_g}$ by iterative process as

$$r_{\boldsymbol{p}_g}^{(k+1)} = r_{\boldsymbol{p}_g}^{(k)} - G(r_{\boldsymbol{p}_g}^{(k)}) \frac{\Delta r_{\boldsymbol{p}_g}}{\Delta G} = r_{\boldsymbol{p}_g}^{(k)} - \frac{f_v(r_{\boldsymbol{p}_g}^{(k)}) - \lambda_v(r_{\boldsymbol{p}_g}^{(k)})}{f_v'(r_{\boldsymbol{p}_g}^{(k)}) - \lambda_v'(r_{\boldsymbol{p}_g}^{(k)})}, \tag{12}$$

where $r^{(X)}$ denotes the r element with the notation identifier X, in this case, $r^{(X)}$ denotes r element of the X-th iteration. The derivative f_v' is obtained by Table 2 computed beforehand. λ_v of $\Lambda_v(0, -\lambda_v)$ and its derivative are functions of $r_{\boldsymbol{p}_g}^{(k)}$:

$$\lambda_v = \frac{r_{\boldsymbol{p}_g}^{(k)} z_{\boldsymbol{p}_w}}{r_{\boldsymbol{p}_w} - r_{\boldsymbol{p}_g}^{(k)}}, \quad \lambda_v' = \frac{r_{\boldsymbol{p}_g}^{(k)} z_{\boldsymbol{p}_w}}{(r_{\boldsymbol{p}_w} - r_{\boldsymbol{p}_g}^{(k)})^2} + \frac{z_{\boldsymbol{p}_w}}{r_{\boldsymbol{p}_w} - r_{\boldsymbol{p}_g}^{(k)}}. \tag{13}$$

The global convergence of this process is ensured from the characteristics of $G(r_{\boldsymbol{p}_g})$. From Eq. (9), Eq. (11), and Eq. (13), the form of $G(r_{\boldsymbol{p}_g})$ becomes the same as Eq. (6) and known as a twice continuously differentiable, increasing, and convex function having a zero[28].

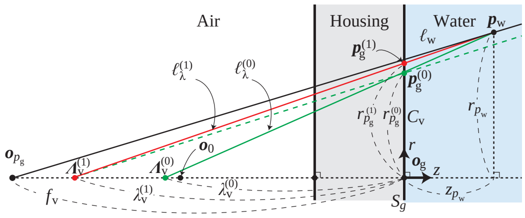


Fig. 7 3D-2D projection by a recurrence relation. The green line illustrates the initial guess. The virtual pixel $\boldsymbol{p}_g^{(0)}$ has the corresponding focal length $f_v(r_{\boldsymbol{p}_g}^{(0)})$ and then the line is updated for $\boldsymbol{\Lambda}_v^{(1)}$ (red line). As a result, the recurrence relation gradually converges to the correct projection point using calibrated relationships $\boldsymbol{p}_g \mapsto f_v$ for each virtual pixel $\boldsymbol{p}_g$.

By Recurrence Relation

As shown in Fig. 6, we can also obtain the corresponding ray without using the differential f_v' as follows:

$$r_{\boldsymbol{p}_g}^{(k)} = \frac{r_{\boldsymbol{p}_w} \lambda_v^{(k)}}{z_{\boldsymbol{p}_w} + \lambda_v^{(k)}}, \quad \lambda_v^{(k+1)} = f_v(r_{\boldsymbol{p}_g}^{(k)}) \tag{14}$$

$$\Leftrightarrow r_{\boldsymbol{p}_g}^{(k+1)} = \frac{r_{\boldsymbol{p}_w} f_v(r_{\boldsymbol{p}_g}^{(k)})}{z_{\boldsymbol{p}_w} + f_v(r_{\boldsymbol{p}_g}^{(k)})}. \tag{15}$$

The method guarantees global convergence because of the following reasons. From Eq. (9), f_v is a monotonic function of $r_{\boldsymbol{p}_g}$ satisfying

$$r_{\boldsymbol{p}_g}^{(k+1)} > r_{\boldsymbol{p}_g}^{(k)} \Leftrightarrow f_v^{(k+1)} > f_v^{(k)}. \tag{16}$$

Therefore, from Eq. (15),

$$r_{\boldsymbol{p}_g}^{(0)} < r_{\boldsymbol{p}_g}^{(1)} \Rightarrow r_{\boldsymbol{p}_g}^{(k+1)} < r_{\boldsymbol{p}_w} \frac{f_v(r_{\boldsymbol{p}_g}^{(k+1)})}{z_{\boldsymbol{p}_w} + f_v(r_{\boldsymbol{p}_g}^{(k+1)})} < r_{\boldsymbol{p}_g}. \tag{17}$$

That is, the following monotonicity is satisfied:

$$r_{\boldsymbol{p}_g}^{(1)} > r_{\boldsymbol{p}_g}^{(0)} \Rightarrow r_{\boldsymbol{p}_g}^{(k+1)} \geq r_{\boldsymbol{p}_g}^{(k)}, \quad r_{\boldsymbol{p}_g}^{(1)} < r_{\boldsymbol{p}_g}^{(0)} \Rightarrow r_{\boldsymbol{p}_g}^{(k+1)} \leq r_{\boldsymbol{p}_g}^{(k)}. \tag{18}$$

As a result, the iteration converges to

$$r_{\boldsymbol{p}_g}^{(k+1)} = r_{\boldsymbol{p}_g}^{(k)} \Leftrightarrow f_v^{(k+1)} = f_v^{(k)}. \tag{19}$$

Notice that a reasonable initial guess for $\lambda_v^{(0)}$ can be given by projecting the point in water to the projection center $\boldsymbol{o}_0$ of the real camera C without considering the refraction.

3.2.4 Efficiency of 3D-2D Projection of a Planar Refractive Ray-Pixel Camera

This section evaluates our 3D-2D projection computation in terms of efficiency. Notice that the proposed model does not improve the projection accuracy in comparison with the conventional analytical approach by definition.

Rate of Convergence

To evaluate the rate of convergence of our iterative methods for the 3D-2D projection in Section 3.2.3, **Fig. 8** shows the projection error E_p against the number of iterations k. By using a synthesized data set, the reprojection error is defined as

$$E_p = |P'(\hat{r}_{p_g}) - P'(r_{p_g}^{(k)})|, \tag{20}$$

where $\hat{r}_{p_g}$ is the ground-truth and $r_{p_g}^{(k)}$ is the value returned by the

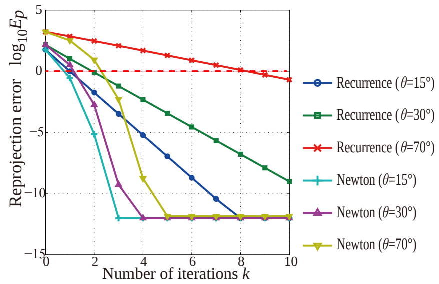


Fig. 8 Comparison of the rate of convergence. $\log_{10} E_p = 0$ (led dashed line) corresponds to the 1 pixel accuracy. Notice that errors are lower bounded by 10^{-12}, the default precision of the floating-point computations in our implementation. ©2016, Elsevier [28].

Table 3 Average computational costs of single 3D-2D projections.

	Analytical[24]	By Recurrence	By Newton
Runtime	1.39 msec	0.14 msec	0.27 msec

algorithm at the k-th iteration in C_v. $P'(r_{p_g}^{(k)})$ denotes the pixel position in the original image of C' corresponding to $r_{p_g}^{(k)}$ in C_v. Notice that $P'(\cdot)$ is employed only for evaluating E_p in pixels, and is not required for the 3D-2D projection to C_v.

From these results, we can observe that (1) the rates of convergence of the recurrence relation and the Newton-based ones are linear and quadratic respectively, and (2) our Newton-based algorithm with iteration $k = 3$ achieve a sub-pixel accuracy.

Computational Efficiency

Table 3 reports computational costs of our methods computing up to the subpixel accuracy and those of the state-of-the-art solving a 12th degree of equation analytically[24]. They are the average values of 6400 3D-2D projections run in Matlab on an Intel Core-i5 2.5GHz PC.

From these results, we can conclude that our method runs much faster than the analytical method while achieving the sub-pixel accuracy. In other words, our method achieves a comparable accuracy as other methods in practical image-based analysis.

The linear and quadratic rates of convergence shown in Fig. 8 do not immediately indicate that the Newton-based method is always better. That is because their computation costs for one iteration step is different as shown in Table 3. One practical option is that combining these two methods by updating with the recurrence relation method first and then by switching to the Newton-based one for fine tuning, because the computation cost of recurrence relation method is smaller but linear convergence requires many iteration steps in larger angle of refraction as shown in Fig. 8.

3.3 Underwater 3D Reconstruction

This section introduces our underwater active stereo system in which both cameras and projectors are modeled by our planar refraction ray-pixel camera model.

The main difficulty in underwater stereo is its depth-dependent distortion on the image plane caused by refraction. This distortion deforms the epipolar line and invalidates stereo methods which provide dense 3D shape by template matching.

We show that our model can realize an efficient implementation of underwater active stereo utilizing the 3D-2D projection introduced in the last section.

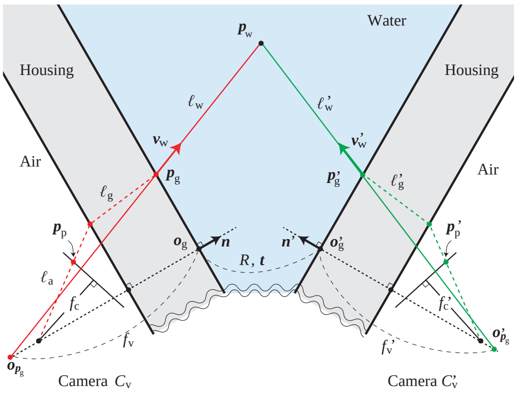


Fig. 9 Linear extrinsic calibration of underwater cameras. The constraint is that $\boldsymbol{p}_w$, $\boldsymbol{o}_{p_g}$, and $\boldsymbol{o}'_{p_g}$ form a triangle.

3.3.1 Calibration

Single Camera Calibration in Water

Single underwater camera is calibrated based on a conventional method using a reference object of known geometry (such as a chessboard) in water[24], [49]. Once the parameters describing the refraction process are obtained, we can build the ray-pixel mapping table as shown in Table 2.

Linear Extrinsic Calibration of Underwater Cameras

Relative posture of planar refraction ray-pixel cameras can be estimatd from corresponding pairs of projections of scene points.Suppose each of the cameras are calibrated as a planar refraction ray-pixel camera beforehand. Given a pair of corresponding points in two such cameras, we can back-project the rays in water as shown in **Fig. 9**. Since these two lines intersect, the following coplanarity constraint holds.

$$\boldsymbol{v_w}^\top \left(\left(R\boldsymbol{o}'_{p_g} + \boldsymbol{t} - \boldsymbol{o}_{p_g} \right) \times (R\boldsymbol{v}'_w) \right) = 0. \tag{21}$$

By rewriting this in a bilinear form, we have

$$\begin{pmatrix} x_{\boldsymbol{v}_w} \\ y_{\boldsymbol{v}_w} \\ z_{\boldsymbol{v}_w} \\ -f_v y_{\boldsymbol{v}_w} \\ f_v x_{\boldsymbol{v}_w} \\ 0 \end{pmatrix} \begin{pmatrix} [\boldsymbol{t}]_\times R & R \\ R & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} x'_{\boldsymbol{v}_w} \\ y'_{\boldsymbol{v}_w} \\ z'_{\boldsymbol{v}_w} \\ -f'_v y'_{\boldsymbol{v}_w} \\ f_v x'_{\boldsymbol{v}_w} \\ 0 \end{pmatrix} = 0, \tag{22}$$

where $[\boldsymbol{X}]_\times$ denotes the 3×3 skew-symmetric matrix defined by a 3D vector $\boldsymbol{X}$. Now we can rewrite the equation in a Plücker forms as

$$\ell_w E_v \ell'_w = 0. \tag{23}$$

Since ℓ_w and ℓ'_w are given by each of the corresponding pairs, we can linearly estimate 17 unknown (9 for $[\boldsymbol{t}_\times]R$, 8 for R except for $R_{3,3}$) elements of E_v up to a scale using 16 or more corresponding pairs for Eq. (23).

Notice that Eq. (23) has a trivial solution $R\boldsymbol{o}'_{p_g} + \boldsymbol{t} - \boldsymbol{o}_{p_g} = 0$. This indicate that two virtual optical centers coinside. As mentioned in [50],

$$E_v = \begin{pmatrix} \boldsymbol{0} & I \\ I & \boldsymbol{0} \end{pmatrix}. \tag{24}$$

exists iff $f_v = f'_v$ and these virtual focal lengths are changed for each pixel.

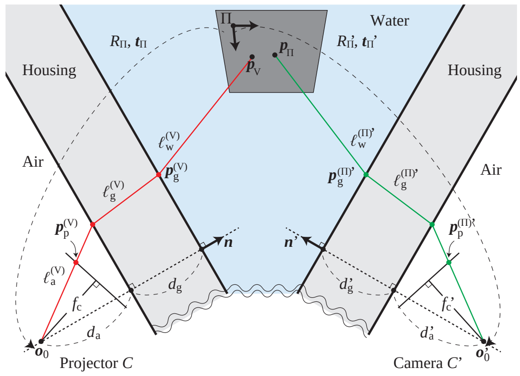


Fig. 10 Underwater projector-camera calibration. A projector C and a camera C' observe a point $\boldsymbol{p}_V$ on a plane Π in water via flat housings. The camera C' also observes a printed point $\boldsymbol{p}_\Pi$ to get a pose of the plane Π. The underwater projector calibration is conducted by the $\langle \boldsymbol{p}_p^{(V)}, \boldsymbol{p}_V \rangle$ pairs. ©2016, Elsevier [28].

Calibration of Projector-Camera System in Water

This section describes our underwater projector-camera calibration as originally presented in [51]. Same as the case of single underwater camera calibration, what required for the parameters of underwater projector is the corresponding pairs of a 3D scene point and its projection $\langle \boldsymbol{p}_p, \boldsymbol{p}_w \rangle$. However, since the position of 3D point $\boldsymbol{p}_w$ projected by the underwater projector cannot be obtained directly, we estimate $\boldsymbol{p}_w$ by an underwater camera calibrated beforehand.

As illustrated in **Fig. 10**, there are underwater projector C, underwater camera C', and reference plane Π in water. Let us assume underwater camera C' is calibrated as C_v' beforehand.

Suppose reference points of known geometry are printed on the plane Π as $\boldsymbol{p}_\Pi = (x_{\boldsymbol{p}_\Pi}, y_{\boldsymbol{p}_\Pi}, 0)^\top$, for the pose estimation of camera C_v' w.r.t. Π by capturing pattern $\boldsymbol{p}_\Pi$. On the other hand, unknown points are projected on the plane Π as $\boldsymbol{p}_V = (x_{\boldsymbol{p}_V}, y_{\boldsymbol{p}_V}, 0)^\top$ by the projector C. Our goal is to estimate $\boldsymbol{p}_V$ and calibrate underwater projector as C_v.

To sum up, our calibration consists of the following steps.

Step 1. Camera C_v' pose estimation w.r.t. Π by capturing pattern $\boldsymbol{p}_\Pi$.

Step 2. Estimation of 3D geometry of a pattern $\boldsymbol{p}_V$ projected by C on Π using C_v'.

Step 3. Projector calibration of C_v and its pose estimation w.r.t. Π using 2D-3D correspondence of pattern $\boldsymbol{p}_V$.

Step 1. Pose Estimation of Camera using Planner Pattern in Water

The camera pose R_Π' and $\boldsymbol{t}_\Pi'$ w.r.t. Π can be estimated using the flat refraction constraint[24]. That is, the direction $\boldsymbol{v}_w^{(\Pi)\prime}$ of the ray ℓ_w' as a backprojection is identical to the vector from the incident point $\boldsymbol{p}_g^{(\Pi)\prime}$ to $\boldsymbol{p}_w^{(\Pi)\prime}$, where the known point $\boldsymbol{p}_\Pi = (x_{\boldsymbol{p}_\Pi}, y_{\boldsymbol{p}_\Pi}, 0)^\top$ is described as $\boldsymbol{p}_w^{(\Pi)\prime}$ in the coordinate system of camera C_v' (Fig. 10).

$$\boldsymbol{v}_w^{(\Pi)\prime} \times \left((R_\Pi' \boldsymbol{p}_\Pi' + \boldsymbol{t}_\Pi') - \boldsymbol{p}_g^{(\Pi)\prime} \right) = 0 \tag{25}$$

$$\Leftrightarrow \begin{pmatrix} x_{\boldsymbol{p}_{(\Pi)}}'[\boldsymbol{v}_w^{(\Pi)\prime}]_\times \\ y_{\boldsymbol{p}_{(\Pi)}}'[\boldsymbol{v}_w^{(\Pi)\prime}]_\times \\ [\boldsymbol{v}_w^{(\Pi)\prime}]_\times \end{pmatrix}^\top \begin{pmatrix} \boldsymbol{r}_{\Pi,1}' \\ \boldsymbol{r}_{\Pi,2}' \\ \boldsymbol{t}_\Pi' \end{pmatrix} = [\boldsymbol{v}_w^{(\Pi)\prime}]_\times \boldsymbol{p}_g^{(\Pi)\prime},$$

where $\boldsymbol{r}_{X,i}$ denotes the ith column vector of R_X with notation identifier X. Since this equation provides three constraints for 9 unknowns $\boldsymbol{r}_{c,1}{}'$, $\boldsymbol{r}_{c,2}{}'$, and $\boldsymbol{t}_c{}'$, we can solve this system of equations linearly by using at least three points. Once $\boldsymbol{r}_{c,1}{}'$ and $\boldsymbol{r}_{c,2}{}'$ are obtained, $\boldsymbol{r}_{c,3}{}'$ is given by their cross product.

Step 2. Estimation of 3D Geometry of Projected Pattern

The goal here is to estimate $\boldsymbol{p}_V = (x_{\boldsymbol{p}_V}, y_{\boldsymbol{p}_V}, 0)^\top$ from its projection $\boldsymbol{p}_p^{(V)}{}'$ in the camera $C_v{}'$ image in order to establish 2D-3D correspondences between 2D projector pixels $\boldsymbol{p}_p^{(V)}$ and 3D points $\boldsymbol{p}_V$ on Π.

Since $\boldsymbol{p}_V$ is on $\ell_w{}'$, we can represent its 3D position to $C_v{}'$ by 2D-3D projection (mentioned in the section3.2.2) with a scale parameter $t_w{}'$ as

$$\begin{aligned}\boldsymbol{p}_V &= R_\Pi^\top(\boldsymbol{p}_w^{(V)}{}' - \boldsymbol{t}_\Pi{}') \\ &= R_\Pi^\top(t_w{}'\boldsymbol{v}_w^{(V)}{}' + \boldsymbol{o}_{\boldsymbol{p}_g}^{(V)}{}' - \boldsymbol{t}_\Pi{}').\end{aligned} \tag{26}$$

Since we know z of $\boldsymbol{p}_V$ is equal to 0, it is trivial to determine the unknowns $x_{\boldsymbol{p}_V}$, $y_{\boldsymbol{p}_V}$, and $t_w{}'$

Step 3. Calibration of Projector using 2D-3D Correspondences

Given a set of correspondences between 2D projector pixels and its projection on the plane Π ($\boldsymbol{p}_p^{(V)}$, $\boldsymbol{p}_V$) in the previous step, the pose of the real projector R_Π and $\boldsymbol{t}_\Pi$ w.r.t. Π, and its housing parameters can be calibrated by the conventional method[24]. Once obtained these parameters, we can build a table representing the virtual focal length as done for underwater camera.

Notice that the 3D points $\boldsymbol{p}_V$ are not necessarily from a single Π. In fact, by capturing the panel Π with different poses in water, they can cover a larger area of the scene and contribute to improve the accuracy and robustness of the parameter estimation as pointed out in [24].

3.3.2 Triangulation by Planar Refraction Ray-Pixel Cameras

Towards the 3D reconstruction, we introduce the triangulation method by multiple planar refraction ray-pixel camerass. The basic idea is to form triangle as with the case of linear extrinsic calibration of underwater cameras in Section 3.3.1.

Instead of using the plane-of-refraction constraint, the triangulation process utilizes the two way of a light path description. That is, in Fig. 9, outgoing ray direction $\boldsymbol{v}_w = (r_{v_w}, z_{v_w})^\top$ of ℓ_w is also described as $\boldsymbol{o}_{p_g}(0, -f_v)$-$\boldsymbol{p}_w(r_{p_w}, z_{p_w})$ direction.

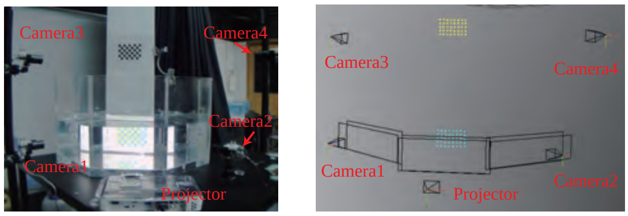
Therefore, the following relationship (flat refraction constraint) holds.

$$\begin{aligned}&\boldsymbol{v}_w \times (\boldsymbol{p}_w - \boldsymbol{o}_{p_g}) = 0 \\ \Leftrightarrow\ & r_{p_w} - \frac{r_{v_w}}{z_{v_w}} z_{p_w} = -\frac{r_{v_w}}{z_{v_w}} f_v.\end{aligned}$$

For multiple viewpoints, we rewrite this in (x, y, z) expression as

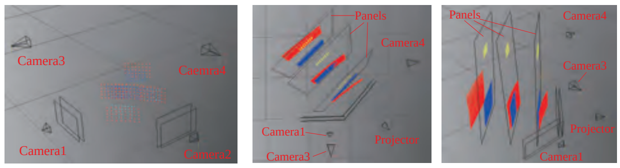
$$\begin{aligned}&\begin{pmatrix} 1 & 0 & -\frac{x_{v_w}}{z_{v_w}} \\ 0 & 1 & -\frac{y_{v_w}}{z_{v_w}} \end{pmatrix} \begin{pmatrix} x_{p_w} \\ y_{p_w} \\ z_{p_w} \end{pmatrix} = -\frac{f_q}{z_{v_w}} \begin{pmatrix} x_{v_w} \\ y_{v_w} \end{pmatrix}, \\ &\Leftrightarrow A_0 \boldsymbol{p}_w = \boldsymbol{b}_0,\end{aligned} \tag{27}$$

and for C'_{v1},



(a) (b)

Fig. 11 Evaluation environment of underwater projector-camera system. (a) Two cameras and one projector observing the underwater target (colored chess pattern) via a flat housing, and two cameras capturing the reference object (black chess pattern) in the air to provide the ground truth position of the colored chess pattern. (b) Calibration result of evaluation system. *Camera*1 and *Camera*2 define our underwater camera system. *Camera*1 and *Projector* define our underwater projector-camera system. ©2016, Elsevier [28].



(a) (b) (c)

Fig. 12 Evaluation Result. (a) 3D points estimated by the underwater camera pair. The blue points are ground truth. Our estimation with refraction modeling (cyan points) correctly provide the position comparing with the estimation without refraction computation (red points). (b) 3D points estimated by the underwater projector-camera pair. The panel planes with the yellow points are calibrated by the reference cameras in the air as the ground truth. The blue points are estimated by our underwater projector-camera system. The red points are estimated by assuming perspective projection without refraction. (c) Top view of (b). ©2016, Elsevier [28].

$$\begin{aligned}&A_1(R_1\boldsymbol{p}_w + \boldsymbol{t}_1) = \boldsymbol{b}_1, \\ \Leftrightarrow\ &(A_1R_1)\boldsymbol{p}_w = \boldsymbol{b}_1 - A_1\boldsymbol{t}_1.\end{aligned} \tag{28}$$

By combining these two, we obtain $\boldsymbol{p}_w$ by solving

$$\begin{pmatrix} A_0 \\ A_1R_1 \end{pmatrix} \boldsymbol{p}_w = \begin{pmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_1 - A_1\boldsymbol{t}_1 \end{pmatrix}. \tag{29}$$

In the case of three or more viewpoints, we can simply stack the same constraints in Eq. (29) for each viewpoint.

3.3.3 Quantitative Evaluation using Reference Cameras

As shown in **Fig. 11**, we used four SXGA cameras (Pointgrey CMLN-13S2C-CS) and one 1080p projector (BenQ MH680) around an octagonal water tank (900 mm diameter, 30mm thickness). The capture target is a flat panel having two chess patterns of known geometry on it: a colored pattern in water and a black pattern in the air. Notice that this is optically-equivalent to having an underwater projector and cameras with waterproof flat housings.

*Camera*1 and *Camera*2 and the projector observe the patterns of target in water. *Camera*3 and *Camera*4 are used as reference cameras calibrated beforehand by capturing the reference patterns in the air[52] for evaluation purpose. These two reference cameras can provide the ground truth of patterns' 3D geometry in water indirectly by capturing the both chess patterns in the air.

Evaluation of Underwater Cameras

Fig. 12-(a) shows the estimated 3D geometry of 40 chess corners in water on five panels at different distances, where these

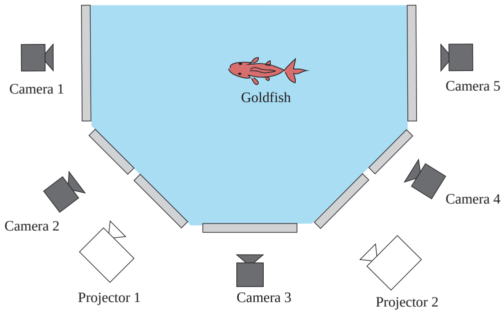


Fig. 13 Setup for dynamic 3D shape capture of *goldfish*. The cameras, the projectors, and the tank are the same ones used in Fig. 11. ©2016, Elsevier [28].

corners are different from the corners used for calibration step and the distance between the nearest and the farthest panels was roughly 400 mm.

The blue points are the ground truth calculated by the reference cameras in the air (*Camera*3, *Camera*4). The cyan ones are the points by our underwater camera system (*Camera*1, *Camera*2), and the average error of these 200 points was 2.43 mm. The red ones are points calculated by assuming the perspective projection without refraction, and its average error was 31.11 mm. The result shows our calibration of underwater cameras obviously provides a better result of 3D measurement quantitatively and qualitatively.

Evaluation of Underwater Projector and Camera

Figs. 12(b) and (c) show the estimated chess corner positions on three panels at every 200 mm using the Structured Lighting[53], [54] conventional Gray code pattern. The yellow points and the plane are the ground truth calculated by the reference cameras in the air (*Camera*3, *Camera*4). The blue dots denote the points by our underwater projector and camera system (*Camera*1, *Projector*). The red dots denote the points by assuming the perspective projection without refraction. These figures qualitatively visualize that our method better reconstructs the 3D points of different distances from the camera and the projector.

The quality of the calibration is assessed by measuring the distance from the ground truth plane to the estimated 3D points. The average errors of the blue points on the three panels were, from near to far, 1.90 mm, 1.59 mm, and 4.01 mm respectively. Those of the red ones were 34.36 mm, 9.53 mm, and 89.06 mm respectively.

From these results we can conclude that our method realized a practical underwater projector-camera calibration in a reasonable accuracy for a wide range of distance from the cameras.

3.3.4 Dynamic 3D Capture of Swimming Fish

As shown in **Fig. 13**, we used five underwater cameras and two underwater projectors, and captured a swimming goldfish. Each of the projector casts a pattern in different color channels (red and blue) for avoiding interference. The system ran at 15 fps in recording, and took about 30 sec per frame to reconstruct the 3D shape by our underwater space carving using 4 mm voxel resolution.

The three columns in the left of **Fig. 14** show the captured images, and the three columns in the right show rendered images of the reconstructed 3D shapes by our method and by the conventional space carving with perspective projection[9]. As the left column of the rendered images shows, we can virtually observe the object appearance even from the top-side of the object where the real camera does not exist. This well demonstrates the accuracy of our 3D shape estimation quality. On the other hand, the conventional space carving cannot produce a comparable result since it ignores refraction and results in poor 3D estimations due to wrong photo-consistency evaluation. These points prove the concept of our image-based full 3D shape reconstruction of underwater dynamic objects.

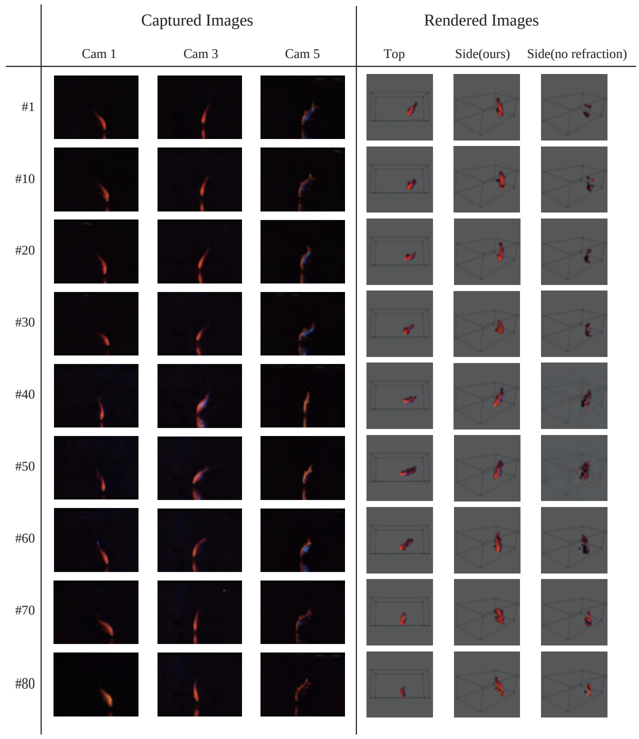


Fig. 14 Result of 3D shape estimation of *goldfish*. Each row shows images of the same frame indicated in the left most column. We can virtually observe the object appearance even from the viewpoint where the real camera does not exist (left column of Rendered Images), and the conventional space carving cannot produce a comparable result since it ignores refraction (right most column). ©2016, Elsevier [28].

4. Catadioptric Ray-Pixel Camera Model

In this section, we introduce the idea and description of our new ray-pixel camera model for catadioptric system with a front lens and planar mirrors.

4.1 Monocentric Lens

Monocentric lens is a spherical and homogeneous optical lens which often has a high refraction index. As shown in **Fig. 15**, rays from a single point (blue) towards a monocentric lens diverge at a wide angle on the other side.

This indicates that a perspective camera located at the point can use the monocentric lens as a conversion lens to obtain a wider field-of-view. It approximately has a short focal length as a thick lens, however, it does not have a single focus strictly[45], as illustrated as the caustic by the red lines in Fig. 15. In order to model such rays efficiently, we first model the ray-pixel mapping through the monocentric lens.

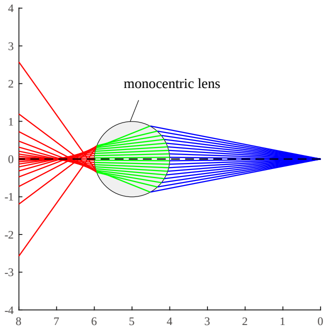


Fig. 15 Refraction by monocentric lens. The blue, green, and red lines indicate incident, refracted, and emergent rays through a monocentric lens respectively. Notice that the emergent rays have a wider field-of-view than that of the incident rays, while they form a caustic.

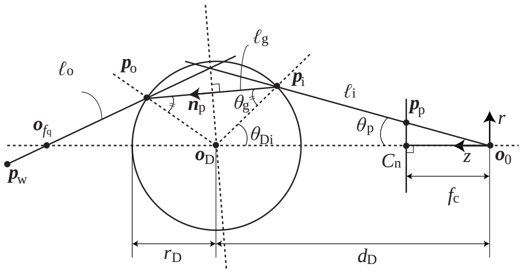


Fig. 16 Measurement through monocentric lens. The segments of projection path ℓ_o-ℓ_g-ℓ_i have an axially symmetric structure around the axis which directs to the monocentric lens center $\boldsymbol{o}_D$.

4.1.1 Measurement through a Monocentric Lens

Suppose a perspective camera whose camera center is located at $\boldsymbol{o}_0$ observes the scene through a monocentric lens located at $\boldsymbol{o}_D$ as illustrated in Fig. 15. Here the perspective camera can be assumed to be directed to the 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 back-projected 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[55]. Therefore, we can describe the ray through a single pixel by a 2D $(r, z)^\top$ coordinate system centered at $\boldsymbol{o}_0$ as shown in Fig. 15.

Consider the projection path ℓ_o-ℓ_g-ℓ_i from $\boldsymbol{o}_0$ through the point $\boldsymbol{p}_p$ in Fig. 15. 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

$$\begin{aligned} z_{pi} &= \frac{d_D - \sqrt{d_D^2 - (1+\tan^2\theta_\mathrm{p})(\mathrm{d}_\mathrm{D}^2 - \mathrm{r}_\mathrm{D}^2)}}{1+\tan^2\theta_\mathrm{p}}, \\ r_{pi} &= z_{pi}\tan\theta_\mathrm{p}. \end{aligned} \tag{30}$$

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:

$$\begin{aligned} \mu_g \sin\theta_g &= \mu_a \sin(\theta_p + \theta_{Di}), \\ \Leftrightarrow \sin\theta_g &= \frac{\mu_a}{\mu_g}\frac{d_D r_{pi}}{r_D\sqrt{r_{pi}^2 + z_{pi}^2}}. \end{aligned} \tag{31}$$

The outgoing point $\boldsymbol{p}_o$ is derived as the intersection of the path ℓ_g and sphere boundary. 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.

Table 4 Pixel-ray mapping of a 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	$f_q(\theta_{q0})$	$f_q(\theta_{q1})$	$\cdots$	$f_q(\theta_{qN})$
$\frac{\mathrm{d}f_q(\theta_q)}{\mathrm{d}\theta_q}$	$f_q'(\theta_{q0})$	$f_q'(\theta_{q1})$	$\cdots$	$f_q'(\theta_{qN})$

As shown in **Fig. 16**, 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

$$\begin{aligned} \boldsymbol{p}_o &= H_p \boldsymbol{p}_i + \boldsymbol{t}_p, \\ \Leftrightarrow \boldsymbol{p}_o &= \left(\boldsymbol{I} - 2\boldsymbol{n}_p\boldsymbol{n}_p^\top\right)\boldsymbol{p}_i + 2(\boldsymbol{o}_D^\top \boldsymbol{n}_p)\boldsymbol{n}_p, \end{aligned} \tag{32}$$

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

$$\boldsymbol{n}_p = \begin{pmatrix} \cos(\frac{\pi}{2} + \theta_{Di} - \theta_g) \\ \sin(\frac{\pi}{2} + \theta_{Di} - \theta_g) \end{pmatrix} = \begin{pmatrix} -\sin(\theta_{Di} - \theta_g) \\ \cos(\theta_{Di} - \theta_g) \end{pmatrix}, \tag{33}$$

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

$$\begin{aligned} H_p &= \begin{pmatrix} \cos 2(\theta_{Di} - \theta_g) & \sin 2(\theta_{Di} - \theta_g) \\ \sin 2(\theta_{Di} - \theta_g) & -\cos 2(\theta_{Di} - \theta_g) \end{pmatrix}, \\ \boldsymbol{t}_p &= \begin{pmatrix} -d_D \sin 2(\theta_{Di} - \theta_g) \\ d_D \cos 2(\theta_{Di} - \theta_g) + d_D \end{pmatrix}. \end{aligned} \tag{34}$$

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

$$\boldsymbol{v}_o = H_p \boldsymbol{v}_i. \tag{35}$$

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:

$$\boldsymbol{o}_{f_q} = t_o \boldsymbol{v}_o + \boldsymbol{p}_o, \tag{36}$$

$$\begin{aligned} \begin{pmatrix} 0 \\ f_q \end{pmatrix} &= t_o \begin{pmatrix} r_{\boldsymbol{v}_o} \\ z_{\boldsymbol{v}_o} \end{pmatrix} + \begin{pmatrix} r_{\boldsymbol{p}_o} \\ z_{\boldsymbol{p}_o} \end{pmatrix}, \\ \Leftrightarrow f_q &= z_{\boldsymbol{p}_o} + \begin{pmatrix} -r_{\boldsymbol{v}_o} \\ z_{\boldsymbol{v}_o} \end{pmatrix}^{-1} r_{\boldsymbol{p}_o}. \end{aligned} \tag{37}$$

The inverse process is similarly obtained for a given path ℓ_o. The analytical 3D-2D projection, however, requires solving a 10th-degree equation[55]. 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.

4.2 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 center to the lens center. As illustrated in Fig. 15, given a pixel $\boldsymbol{p}_p$ by specifying θ_p, the corresponding ray ℓ_o can be determined uniquely.

This section introduces our spherical refraction ray-pixel camera model by representing the ray-pixel mapping as follows.

Fig. 17 illustrates the light path from a pixel $\boldsymbol{p}_p$ on the plane-of-refraction. Suppose the emergent ray ℓ_o intersects with the optical

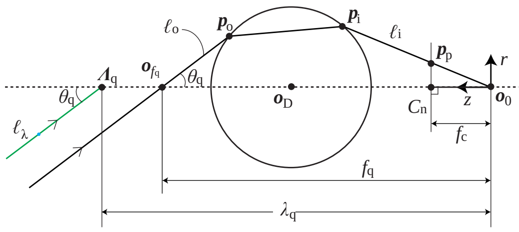


Fig. 17 3D-2D projection of spherical refraction ray-pixel camera model. The black line ℓ_o is the correct projection path that intersects with the optical axis at the camera center $\boldsymbol{o}_0$. Our 3D-2D projection hypothesizes a line ℓ_λ as an initial guess of projection, and then optimizes it by verifying if the point and the angle of intersection with the optical axis is consistent with the ray-pixel mapping $\langle \theta_q, f_q \rangle$.

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 ℓ_o, *i.e.*, $\langle f_q, \theta_q \rangle$ pair.

That is, since the mapping between the pixel $\boldsymbol{p}_p$ and the ray ℓ_o 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. 17 as a whole can be seen as a pixel-wise varifocal camera which changes the focal length f_p 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 4 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 later.

4.2.1 2D-3D Projection

Given a virtual pixel θ_q, the corresponding light path ℓ_o is simply given by

$$\ell_o : \boldsymbol{o}_{f_q} + t_w \boldsymbol{v}_o = \boldsymbol{o}_{f_q} + t_w \begin{pmatrix} -\sin(\theta_q) \\ \cos(\theta_q) \end{pmatrix}, \tag{38}$$

where the parameter t_w represents the depth. The mapping between θ_q and the camera pixel can be provided by the measurement model in Section.4.1 , which can also be stored in the ray-pixel mapping table (Table 4) in practice.

4.2.2 3D-2D Projection

Instead of solving a 10th-degree equation[55] analytically, this section introduces a numerical 3D-2D projection using our ray-pixel mapping table (Table 4.

Similar to the case of planar refraction ray-pixel cameras in Section3.2.3, the key idea of our numerical 3D-2D projection is to hypothesize a projection line ℓ_λ in Fig. 17 first, and check if it intersects with the optical axis at the identical virtual focal length stored in the ray-pixel mapping table. 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. 17, let us consider the ray ℓ_λ from $\boldsymbol{p}_w$ which intersects the axis at $\Lambda_q(0, \lambda_q)$ with 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:

$$\theta_q = \operatorname*{argmin}_{\theta_q} G(\theta_q) = \operatorname*{argmin}_{\theta_q} \Big(f_q(\theta_q) - \lambda_q(\theta_q) \Big). \tag{39}$$

Here $G(\theta_q)$ is a monotonic function and hence we can refine θ_q iteratively as

$$\begin{aligned} \theta_q^{(k+1)} &= \theta_q^{(k)} - G \frac{\Delta\theta_q}{\Delta G} \\ &= \theta_q^{(k)} - \frac{f_q(\theta_q^{(k)}) - \lambda_q(\theta_q^{(k)})}{f_q'(\theta_q^{(k)}) - \lambda_q'(\theta_q^{(k)})}, \end{aligned} \tag{40}$$

where $\theta_q^{(k)}$ denotes θ_q of kth iteration.

Therefore, if we compute the derivative f_q' beforehand as shown in the third row of Table 4, then this 3D-2D projection can be computed efficiently.

Computational Efficiency

Table 5 Average computational costs of single 3D-2D projections.

	Analytical[55]	By Newton with LUT (ours)
Runtime	3.93 msec	0.62 msec

As shown in Table 5, our 3D-2D projection is much faster than the analytical way while maintaining the sub-pixel accuracy. They are the average values of 100 trial, 10K points 3D-2D projections run in Matlab on an Intel Core-i7 2.6GHz PC. The same as the case of planar refraction, the result shows our Spherical Refraction Ray-Pixel Camera compactly realize the 3D-2D projection, while the analytical way requires the process that choosing the solution from the roots of the higher-degree equation for each point.

4.3 Multifacet Mirror

As is well known, observing the scene via a multifacet 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[56], [57]. In our teleidoscopic imaging system, we use a three-facet mirror which defines reflections of the spherical refraction ray-pixel camera introduced in Section 4.2.

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

$$\begin{aligned} \boldsymbol{p}' &= H_i \boldsymbol{p} + \boldsymbol{t}_i \\ \Leftrightarrow \boldsymbol{p}' &= \left(\boldsymbol{I} - 2\boldsymbol{n}_i \boldsymbol{n}_i^\top \right) \boldsymbol{p} + 2 d_i \boldsymbol{n}_i, \end{aligned} \tag{41}$$

where H_i the Householder transformation matrix.

In the case of kaleidoscopic imaging, the mirrors generate bouncing reflections as shown in **Fig. 18**. The reflection of $\boldsymbol{p}'$ by another mirror of normal $\boldsymbol{n}_j$ and distance d is given simply by

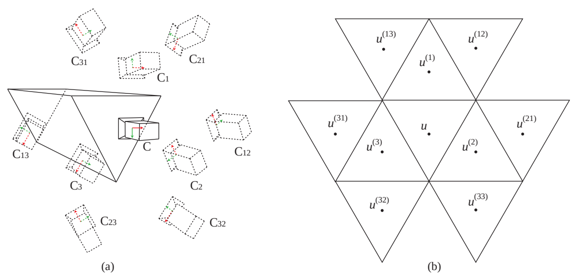


Fig. 18 Kaleidoscopic imaging. (a) Mirrored Cameras. (b) Chambers in the real camera image. The label i of C_i and $\boldsymbol{u}^{(i)}$ denote the reflections of C and u by the ith mirror respectively.

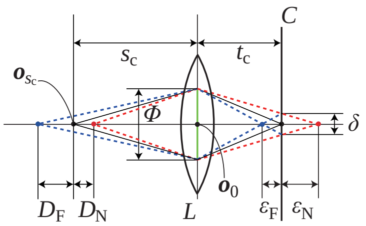


Fig. 19 Depth-of-field of thin lens camera with aperture. The red and blue lines show the backprojection of the permissible circle-of-confusion δ through the aperture Φ. D_F and D_N denote the near and the far depth-of-fields.

$$\begin{aligned} \boldsymbol{p}'' &= H_{ij}\boldsymbol{p} + \boldsymbol{t}_{ij} \\ &= H_j H_i \boldsymbol{p} + 2d_j\boldsymbol{n}_j + 2H_j d_i \boldsymbol{n}_i. \end{aligned} \tag{42}$$

As a result, catadioptric imaging systems can be modeled as a 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. (42) if the mirror parameters are given. The later section describes our calibration algorithm to estimate such parameters.

4.4 Depth of field

This section describes analytical evaluations on the depth-of-field of a catadioptric imaging system. We first review the depth-of-field of the thin lens camera model, and then introduces the monocentric lens.

4.4.1 Depth-of-Field of Thin Lens Camera

Ideally all the incident light rays from a subject point to the lens focuses at a common point. As shown in **Fig. 19**, suppose that a point $\boldsymbol{o}_{s_c}$ is focused on the image plane of a camera C through its lens L. Then the following thin lens formula for paraxial ray holds:

$$\frac{1}{s_c} + \frac{1}{t_c} = \frac{1}{f_c}. \tag{43}$$

The depth-of-field is defined as the backprojection of the permissive circle-of-confusion centered at the focused point. If the subject distance s_c is not long enough, the near and the far depth-of-focus ϵ_N and ϵ_F are given by aperture size Φ and t_c as

$$\begin{aligned} \epsilon_N &= \frac{\delta t_c}{\Phi - \delta} = \frac{\delta s_c f_c}{(\Phi - \delta)(s_c - f_c)}, \\ \epsilon_F &= \frac{\delta t_c}{\Phi + \delta} = \frac{\delta s_c f_c}{(\Phi + \delta)(s_c - f_c)}. \end{aligned} \tag{44}$$

The near and the far depth-of-field D_N and D_F corresponding to ϵ_N and ϵ_F are then obtained by Eq. (43) as

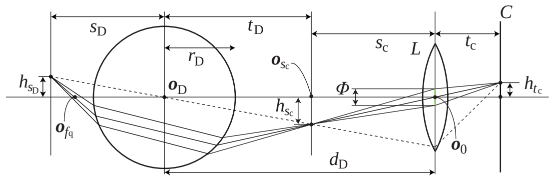


Fig. 20 Light paths through the monocentric lens. Notice that the center of the beam is identical to the light path in Fig. 16.

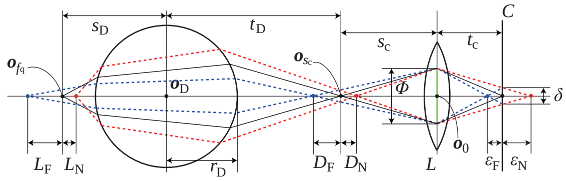


Fig. 21 Depth-of-field with monocentric lens. D_F and D_N denote the depth-of-field of the camera itself without the monocentric lens. The effective depth-of-field of the entire system L_F and L_N can be obtained by projecting D_N and D_F through the monocentric lens (the red and blue dashed lines).

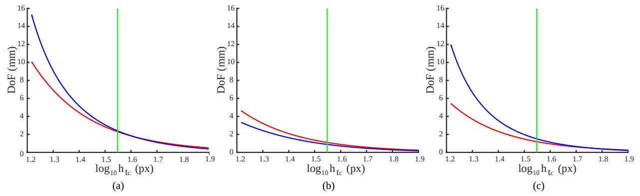


Fig. 22 Depth-of-field with and without monocentric lens. (a), (b) and (c) shows the total, near, and far depth-of-field with and without a monocentric lens. The red and blue plots indicate the results with and without the monocentric lens respectively. The green lines indicate the subject distance where the subject and the focal distances of the monocentric lens itself are same.

$$D_N = \frac{\epsilon_N (s_c - f_c)^2}{f_c^2 + \epsilon_N (s_c - f_c)}, \quad D_F = \frac{\epsilon_F (s_c - f_c)^2}{f_c^2 - \epsilon_F (s_c - f_c)}. \tag{45}$$

As a result, D_N and D_F are described as follows:

$$D_N = \frac{\delta s_c f_c (s_c - f_c)}{(\Phi - \delta) f_c^2 + \delta s_c f_c}, \quad D_F = \frac{\delta s_c f_c (s_c - f_c)}{(\Phi + \delta) f_c^2 - \delta s_c f_c}. \tag{46}$$

Eq. (46) indicates that the depth-of-field of camera C depends on the subject distance s_c, the aperture size Φ, and the permissible circle-of-confusion δ.

4.4.2 Depth of Field with a Monocentric Lens

Fig. 20 illustrates the back-projection of the permissible circle-of-confusion through a thin lens and a monocentric camera. The light path through the lens center $\boldsymbol{o}_0$ is identical to the path illustrated in Fig. 16.

The key point is that an on-focus scene point at the distance s_D is also on-focus at the distance s_c between the two lenses. That is, depth-of-field with a monocentric lens can be simply defined as an extension of the path of the thin lens camera. This is because the points in L_N and L_F 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 led and blue dash lines in **Fig. 21**) and the backprojection path through the lens center $\boldsymbol{o}_0$.

Fig. 22 shows the changes in the depth-of-field by the mono-

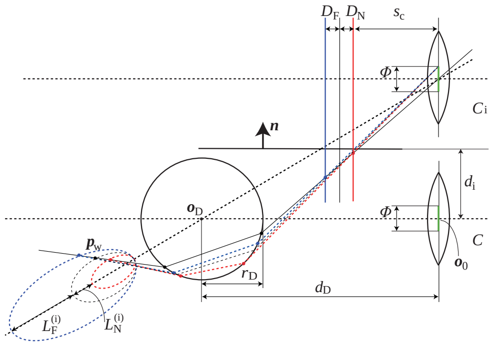


Fig. 23 Depth-of-field of catadioptric imaging system. D_N and D_F denote the near and far depth-of-field of the thin lens camera C respectively. All the points once imaged within D_N-D_F should be originally emitted from the points within L_N-L_F.

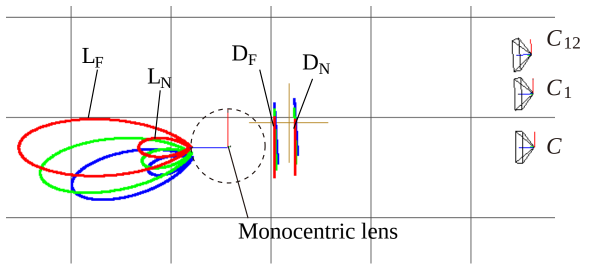


Fig. 24 Intersection of depth-of-fields in catadioptric imaging system. The red curves illustrate the plots of L_N-L_F of the original camera, and the green and the blue curves illustrate those of the first and the second reflections. Notice that the mirrors are at off-perpendicular angle of 1.4° from the camera image plane and hence surrounding viewpoints are generated.

centric lens, in the case of capturing an object of 0.5mm height at the same magnification.

Fig. 22(a) shows the total depth-of-field with and without the monocentric lens $L_N + L_F$ (red and blue respectively). Fig. 22(b) and (c) shows 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 Section 5.4.

In the both cases, the depth-of-field decreases by increasing the magnification in general, while the depth-of-field with the monocentric lens becomes clearly deeper from a specific subject distance. This distance is near from the point where the subject distance of the monocentric lens $s_D - r_D$ is equal to its focal distance $t_D - r_D$. That is, if the target is located closer than this distance, the monocentric lens can contribute to achieve a deeper depth-of-field and hence less blurry imaging.

4.4.3 Depth of Field of a Catadioptric Camera

Fig. 23 illustrates the near and the far depth-of-fields of a virtual camera C_i defined as a mirror of the original camera. 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 catadioptric imaging system has multiple virtual cameras as shown in Fig. 18, each of them has a different depth-of-field according to their relative pose to the monocentric lens as shown in **Fig. 24**. That is, the intersections of such depth-of-fields can be used for multi-view stereo reconstruction for example.

5. Teleidoscopic Imaging System

In this section, we aimed at proposing a catadioptric imaging

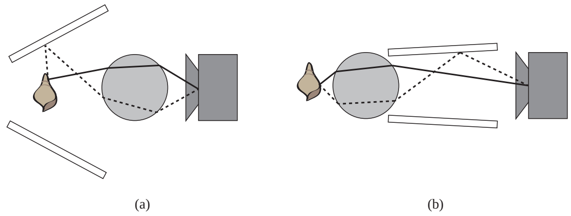


Fig. 25 Two types of catadioptric system for multi-view microscopy. Teleidoscopic imaging system shown in (b) provides semi-surrounding views and less blurry images compared with the system shown in (a), which provides virtually fully surrounding views with larger differences in optical path lengths in closed-up scenarios.

system for microscopic object capture. Unlike conventional microscopic imaging system such as differential phase contrast microscopy[58], [59] and multi-focus approaches[60], [61], [62], 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.

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[63], [64] cannot be a feasible solution due to limitations on camera placement. Conventional multiple mirror system[56] also have difficulties inevitably in depth-of-focus due to differences in their optical paths with varying numbers of bounces.

The key idea to solve these problems 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. 25**, we introduce a kaleidoscopic multi-facet mirror between the front lens and the camera. As discussed later, this design realizes a deeper depth-of-field and results in less blurring imaging.

We call our system *teleidoscopic imaging system* and show that the system can be compactly modeled by a structured ray-pixel camera model[8].

5.1 Calibration of Teleidoscopic Imaging System

This section introduces our calibration algorithm of teleidoscopic imaging system which requires capturing a single reference planar patterns. **Fig. 26** 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 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[56], while the observation in the teleidoscopic system does not include such mirrored points that follow Eq. (42) due to the refraction by the monocentric lens. Similarly, the rays reflected and then refracted through the projection in teleidoscopic imag-

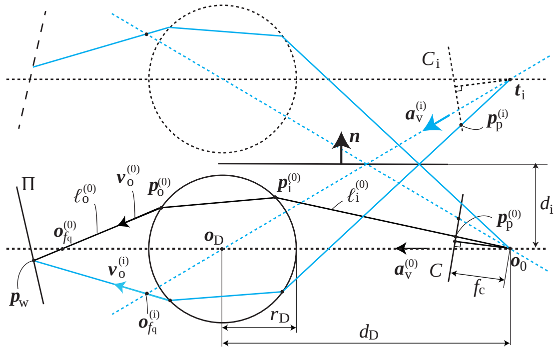


Fig. 26 Calibration of teleidoscopic imaging system. Cyan lines denote the light paths for a mirrored camera. The calibration can be conducted by capturing a reference points on a plane with a known geometry, even if the reference points are captured through a refractive medium.

ing system do not satisfy the coplanarity constraint. In addition, as mentioned in [55], the estimation of the monocentric lens parameters requires multiple viewpoints or multiple monocentric lenses.

These points indicate that this is a chicken-and-egg problem of the following form:

(1) estimation of the monocentric lens parameters requires the mirror parameters,
(2) estimation of the mirror parameters requires at least one mirrored point pair, and
(3) estimation of a mirrored point pair requires the monocentric lens parameters to handle the refraction.

To solve this problem, we utilize the fact that the positions of the centers of the monocentric lens and its mirrors are always captured without refraction by definition, and satisfy Eq. (42). 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.

5.1.1 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 [55], we estimate $\boldsymbol{a}_v^{(0)}$ by the coplanarity constraint of the three vectors $\boldsymbol{a}_v^{(0)}$, $\boldsymbol{p}_p^{(0)}$, $\boldsymbol{p}_w$ on a plane-of-refraction:

$$\begin{aligned} \boldsymbol{p}_p^{(0)\top}\left(\boldsymbol{a}_v^{(0)} \times (R_\Pi \boldsymbol{p}_w + \boldsymbol{t}_\Pi)\right) &= 0, \\ \Leftrightarrow \boldsymbol{p}_p^{(0)\top}(E_\Pi \boldsymbol{p}_w + \boldsymbol{s}_\Pi)) &= 0, \end{aligned} \tag{47}$$

where R_Π, t_Π 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

$$\boldsymbol{a}_v^{(0)} = \frac{E_\Pi(:,1) \times E_\Pi(:,2)}{\|E_\Pi(:,1) \times E_\Pi(:,2)\|}. \tag{48}$$

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

5.1.2 Mirror Normals

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

$$\boldsymbol{a}_v^{(0)\top}[\boldsymbol{n}_i]_\times \boldsymbol{a}_v^{(i)} = 0, \tag{49}$$

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 $\boldsymbol{n}_i$[56]. Therefore $\boldsymbol{n}_i$ can be obtained linearly only from the axes to the centers of the monocentric lenses.

5.1.3 Mirror Distances

Once the mirror normals are estimated, we can utilize the kaleidoscopic triangulation[56] to obtain linear constraints on the mirror distances d_i. That is, for the first and the second reflections such as C_i and C_{ij}, following equation holds:

$$\begin{aligned} (\boldsymbol{a}_v^{(0)} \times (H_{ij}\boldsymbol{a}_v^{(ij)}))^\top(\boldsymbol{0} - \boldsymbol{t}_{ij}) &= 0 \\ \Leftrightarrow (\boldsymbol{a}_v^{(0)} \times (H_j H_i \boldsymbol{a}_v^{(ij)}))^\top(\boldsymbol{0} - 2d_j\boldsymbol{n}_j - 2H_j d_i \boldsymbol{n}_i) &= 0. \end{aligned} \tag{50}$$

By integrating Eq. (50) for $ij = \{12, 13, 21, 23, 31, 32\}$ as a set of linear equations of d_i $(i = 1,2,3)$, we linearly obtain d_1, d_2, d_3 up to scale.

5.1.4 Pose of Reference Plane

Similarly to Eq. (47), the plane-of-refraction constraint holds for the mirrored cameras C_i:

$$(\boldsymbol{p}_p^{(i)})^\top\left(\boldsymbol{a}_v^{(i)} \times (H_i(R_\Pi \boldsymbol{p}_w + \boldsymbol{t}_\Pi) + \boldsymbol{t}_i)\right) = 0. \tag{51}$$

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

5.1.5 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:

$$\boldsymbol{v}_o^{(i)} \times (\boldsymbol{p}_w^{(i)} - \boldsymbol{p}_o^{(i)}) = 0. \tag{52}$$

This is a nonlinear constraint for the monocentric lens parameters as described in Section 4.1.1 and we solve this as a nonlinear optimization problem with assuming their rough estimates are available in practice.

5.1.6 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. On computing the 3D-2D projection, we used the analytical solution by [55].

5.2 Evaluation

In this section, we evaluate the calibration of our catadioptric ray-pixel camera.

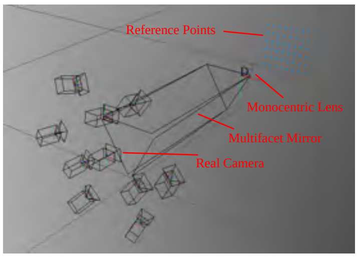


Fig. 27 Teleidoscopic system for quantitative evaluation.

5.2.1 Quantitative Evaluation using Synthesized Data

Fig. 27 shows the measurement environment which simulates the real capture system used in Section 5.4. 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.

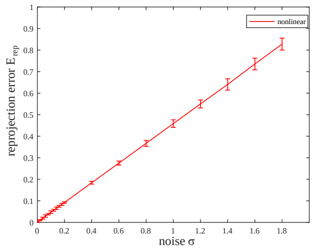


Fig. 28 Reprojection errors at different noise levels. The bars denote the standard deviation of the errors.

The system captures 48 reference points (blue dots in Fig. 27) to calibrate its parameters. By injecting Gaussian noise of different standard deviations σ to the 2D positions of their projections, we evaluate the robustness of our calibration procedure.

Fig. 28 shows average reprojection errors in pixel of 100 trials at each pixel noise level σ. We can observe that the reprojection errors increases linearly against the pixel noise level.

Figs. 29, **30**, **31** show 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.

5.3 Teleidoscopic Triangulation

Since our Teleidoscopic Imaging System has multiple ray-pixel cameras, the manner of the triangulation is the same as described in Section 3.3.2.

In the case of three or more viewpoints, depending on the number of the cameras sharing an intersection of the depth-of-fields (Fig. 24), we can add equations in the same form into this system for triangulation.

5.4 Teleidoscopic 3D Shape Reconstruction

To evaluate the proposed teleidoscopic system as a multi-view

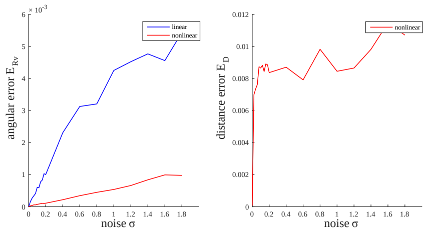


Fig. 29 Estimation errors of the monocentric lens parameters. Left: Angular error of the axis to the monocentric lens center (degree). Right: Distance error normalized by the ground truth.

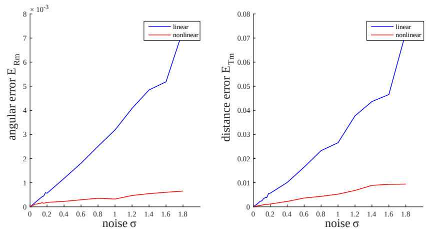


Fig. 30 Estimation errors of the mirror poses. Left: Angular error of the mirror normal (degree). Right: Distance error normalized by the ground truth.

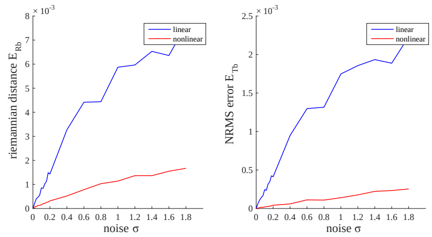


Fig. 31 Estimation errors of the reference plane pose. Left: Rotation error (Riemannian distance). Right: Distance error (RMS normalized by the ground truth).

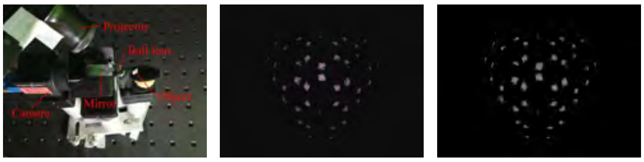
Fig. 32 Teleidoscopic imaging. (a) Captured object (approximately 5mm size). (b) Image by our system. (c) Image Without the monocentric lens. The refraction by the monocentric lens provides the surround view.

camera system for 3D shape reconstruction, this section demonstrates a 3D reconstruction of a small object of approximately 5mm size shown in **Fig. 32**(a).

The system consists of a FLIR Flea3 FL3-U3-88S2C-C camera (4000×3000 resolution, pixel size 1.55μm) with an S-mount lens (focal length 3mm, F8), three 10×30mm mirrors, and a monocentric lens of 10mm diameter whose refraction index is $\mu_g = 2.0$.

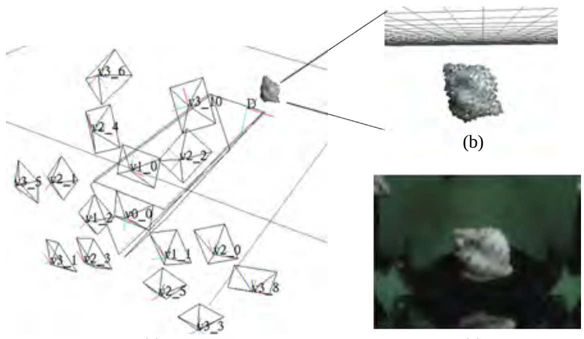
Figs. 32(b) and (c) show images captured by our teleidoscopic system and by its kaleidoscopic part without the monocentric lens. These images clearly demonstrates that the use of the monocentric lens realizes a denser close-up surrounding multi-view capture of the target.

In order to demonstrate the idea of our teleidoscopic multi-view capture, we introduce a focus-free laser projector Sony MP-CL1A to cast structured light patterns[53], [54] to minimize er-



(a) (b) (c)

Fig. 33 Teleidoscopic imaging system with a projector. (a) Overview. (b) Captured image under a structured illumination. (c) Decoded result.



(a) (b) (c)

Fig. 34 Result of 3D shape estimation. (a) Overall view of the result. (b) Enlarged view of the result. (c) A part of captured image under the scene light (in base chamber).

rors in the stereo correspondence search process as shown in **Figs. 34**(b) and (c).

Figs. 34(a) and (b) show the result of our 3D reconstruction as a point cloud. This result demonstrates that our system realizes a closeup and surround-view capturing successfully.

6. Conclusion

This paper proposed a catadioptric ray-pixel camera model exploiting an axially-symmetric structure of the rays captured by the system. Our model describes the ray associated with each pixel by a simple 1D mapping, and realizes an efficient forward 3D-2D projection.

The proposed system realized a closeup, semi-surround view, and less blurring imaging system for microscale objects. Our future work includes an extension to fully-surround view 3D capture of microscale underwater objects, with immersing the front monocentric lens in water.

Acknowledgments

This research is partially supported by JSPS KAKENHI 26240023, 15J07706, and 18K19815.

References

[1] Starck, J., Hilton, A. and Miller, G.: Volumetric stereo with silhouette and feature constraints, *Proc. BMVC*, pp. 1189–1198 (2006).

[2] Furukawa, Y. and Ponce, J.: Accurate, dense, and robust multi-view stereopsis, *Proc. CVPR*, pp. 1–8 (2007).

[3] Matsuyama, T., Nobuhara, S., Takai, T. and Tung, T.: *3D Video and Its Applications*, Springer Publishing Company, Incorporated (2012).

[4] Ikeuchi, K., Oishi, T., Takamatsu, J., Sagawa, R., Nakazawa, A., Kurazume, R., Nishino, K., Kamakura, M. and Okamoto, Y.: The great buddha project: digitally archiving, restoring, and analyzing cultural heritage objects, *IJCV*, Vol. 75, pp. 189–208 (2007).

[5] Thomas, D. and Sugimoto, A.: A flexible scene representation for 3d reconstruction using an RGB-D camera, *Proc. ICCV*, pp. 2800–2807 (2013).

[6] Gluckman, J., Nayar, S. K. and Thoresz, K. J.: Real-Time Omnidirectional and Panoramic Stereo, *In Proceedings of the 1998 DARPA Image Understanding Workshop*, Morgan Kaufmann, pp. 299–303 (1998).

[7] Aggarwal, R., Vohra, A. and Namboodiri, A. M.: Panoramic Stereo Videos with a Single Camera, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3755–3763 (online), DOI: 10.1109/CVPR.2016.408 (2016).

[8] Grossberg, M. D. and Nayar, S. K.: The Raxel Imaging Model and Ray-Based Calibration, *IJCV*, Vol. 61, No. 2, pp. 119–137 (2005).

[9] Kutulakos, K. N. and Seitz, S. M.: A theory of shape by space carving, *Proc. ICCV*, pp. 307–314 (1999).

[10] Kawasaki, H., Furukawa, R., Sagawa, R. and Yagi, Y.: Dynamic scene shape reconstruction using a single structured light pattern, *Proc. CVPR*, pp. 1–8 (2008).

[11] Furuse, T., Hiura, S. and Sato, K.: 3-D shape measurement method with modulated slit light robust for interreflection and subsurface scattering, *Proc. PROCAMS*, pp. 1–2 (2009).

[12] Zhang, Z.: Microsoft Kinect Sensor and Its Effect, *IEEE Multimedia*, Vol. 19, No. 2, pp. 4–10 (2012).

[13] Schechner, Y. and Karpel, N.: Clear underwater vision, *Proc. CVPR*, Vol. 1, pp. I–536–I–543 Vol.1 (2004).

[14] Narasimhan, S., Nayar, S., Sun, B. and Koppal, S.: Structured light in scattering media, *Proc. ICCV*, Vol. I, pp. 420–427 (2005).

[15] Kocak, D. M., Dalgleish, F. R., Caimi, F. M. and Schechner, Y. Y.: A Focus on Recent Developments and Trends in Underwater Imaging, *Marine Technology Society Journal*, Vol. 42, pp. 52–67 (2008).

[16] Mukaigawa, Y., Raskar, R. and Yagi, Y.: Analysis of Scattering Light Transport in Translucent Media, *IPSJ Transactions on Computer Vision and Applications*, Vol. 3, pp. 122–133 (2011).

[17] Akkaynak, D. and Treibitz, T.: A Revised Underwater Image Formation Model, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018).

[18] Fujimura, Y., Iiyama, M., Hashimoto, A. and Minoh, M.: Photometric Stereo in Participating Media Considering Shape-Dependent Forward Scatter, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018).

[19] Asano, Y., Zheng, Y., Nishino, K. and Sato, I.: Shape from Water: Bispectral Light Absorption for Depth Recovery, *ECCV* (2016).

[20] Alterman, M., Schechner, Y. Y., Perona, P. and Shamir, J.: Detecting Motion through Dynamic Refraction, *TPAMI*, Vol. 35, No. 1, pp. 245–251 (2013).

[21] Qian, Y., Zheng, Y., Gong, M. and Yang, Y.-H.: Simultaneous 3D Reconstruction for Water Surface and Underwater Scene, *The European Conference on Computer Vision (ECCV)* (2018).

[22] Alterman, M., Schechner, Y. and Swirski, Y.: Triangulation in random refractive distortions, *Proc. ICCP*, pp. 1–10 (2013).

[23] Kunz, C. and Singh, H.: Hemispherical refraction and camera calibration in underwater vision, *OCEANS 2008*, pp. 1–7 (2008).

[24] Agrawal, A., Ramalingam, S., Taguchi, Y. and Chari, V.: A theory of multi-layer flat refractive geometry, *Proc. CVPR*, pp. 3346–3353 (2012).

[25] Treibitz, T., Schechner, Y. Y. and Singh, H.: Flat refractive geometry, *Proc. CVPR* (2008).

[26] Pedersen, M., Hein Bengtson, S., Gade, R., Madsen, N. and Moeslund, T. B.: Camera Calibration for Underwater 3D Reconstruction Based on Ray Tracing Using Snell's Law, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* (2018).

[27] Jordt-Sedlazeck, A. and Koch, R.: Refractive structure-from-motion on underwater images, *Proc. ICCV*, pp. 57–64 (2013).

[28] Kawahara, R., Nobuhara, S. and Matsuyama, T.: Dynamic 3D capture of swimming fish by underwater active stereo, *Methods in Oceanography*, Vol. 17, pp. 118 – 137 (online), available from ⟨http://www.sciencedirect.com/science/article/pii/S2211122015300074⟩ (2016).

[29] Chadebecq, F., Vasconcelos, F., Dwyer, G., Lacher, R. M., Ourselin, S., Vercauteren, T. and Stoyanov, D.: Refractive Structure-from-Motion Through a Flat Refractive Interface, *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 5325–5333 (2017).

[30] Shibata, A., Fujii, H., Yamashita, A. and Asama, H.: Scale-reconstructable Structure from Motion using refraction with a single camera, *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5239–5244 (2015).

[31] Sturm, P.: Multi-view geometry for general camera models, *Proc. CVPR*, Vol. 1, pp. 206–212 (2005).

[32] Miraldo, P. and Araujo, H.: Calibration of Smooth Camera Models, *TPAMI*, Vol. 35, No. 9, pp. 2091–2103 (2013).

[33] Yano, T., Nobuhara, S. and Matsuyama, T.: 3D Shape from Silhouettes in Water for Online Novel-view Synthesis, *IPSJ Transactions on Computer Vision and Applications*, Vol. 5, pp. 65–69 (2013).

[34] Nishimura, M., Nobuhara, S., Matsuyama, T., Shimizu, S. and Fujii, K.: A Linear Generalized Camera Calibration from Three Intersecting Reference Planes, *Proc. of ICCV* (2015).

[35] Sedlazeck, A. and Koch, R.: Calibration of Housing Parameters for Underwater Stereo-Camera Rigs, *Proc. BMVC*, pp. 118.1–118.11 (2011).

[36] Nayar, S. K.: Omnidirectional Video Camera, *In Proceedings of the 1997 DARPA Image Understanding Workshop*, pp. 235–241 (1997).

[37] Nishi, R., Aoto, T., Kawai, N., Sato, T., Mukaigawa, Y. and Yokoya, N.: Ultra-Shallow DoF Imaging Using Faced Paraboloidal Mirrors, *Computer Vision - ACCV 2016 - 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part III*, pp. 115–128 (online), DOI: 10.1007/978-3-319-54187-7_8 (2016).

[38] Forbes, K., Nicolls, F., Jager, G. D. and Voigt, A.: Shape-from-silhouette with two mirrors and an uncalibrated camera, *Proc. ECCV* (2006).

[39] Lanman, D., Crispell, D. and Taubin, G.: Surround structured lightning: 3-D scanning with orthographic illumination, *CVIU*, pp. 1107–1117 (2009).

[40] Takahashi, K., Miyata, A., Nobuhara, S. and Matsuyama, T.: A Linear Extrinsic Calibration of Kaleidoscopic Imaging System From Single 3D Point, *Proc. CVPR* (2017).

[41] Tahara, T., Kawahara, R., Nobuhara, S. and Matsuyama, T.: Interference-free epipole-centered structured light pattern for mirror-based multi-view active stereo, *Proc. of International Conference on 3D Vision (3DV)*, pp. 153–161 (2015).

[42] Tagawa, S., Mukaigawa, Y., Kim, J., Raskar, R., Matsushita, Y. and Yagi, Y.: Hemispherical Confocal Imaging, Vol. 3, pp. 222–235 (2011).

[43] Reshetouski, I. and Ihrke, I.: *Mirrors in Computer Graphics, Computer Vision and Time-of-Flight Imaging*, pp. 77–104, Springer Berlin Heidelberg (2013).

[44] Krishnan, G. and Nayar, S. K.: Towards a true spherical camera, *Proc. SPIE*, Vol. 7240, 724002, pp. 1–13 (2009).

[45] Cossairt, O. S., Miau, D. and Nayar, S. K.: Gigapixel Computational Imaging, *2011 IEEE International Conference on Computational Photography (ICCP)*, pp. 1–8 (2011).

[46] Dansereau, D. G., Schuster, G., Ford, J. and Wetzstein, G.: A Wide-Field-Of-View Monocentric Light Field Camera, *Proc. CVPR* (2017).

[47] Ford, B. J.: Single Lens: The Story of the Simple Microscope, *Journal of the History of Biology*, Vol. 19, No. 2, pp. 320–321 (1986).

[48] Adelson, E. H. and Bergen, J. R.: The Plenoptic Function and the Elements of Early Vision, *Computational Models of Visual Processing*, MIT Press, pp. 3–20 (1991).

[49] Kawahara, R., Nobuhara, S. and Matsuyama, T.: 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*, pp. 819–824 (2013).

[50] Li, H., Hartley, R. and Kim, J.-H.: A linear approach to motion estimation using generalized camera models, *Proc. CVPR*, pp. 1–8 (2008).

[51] Kawahara, R., Nobuhara, S. and Matsuyama, T.: Underwater 3D Surface Capture Using Multi-view Projectors and Cameras with Flat Housings, *IPSJ Transactions on Computer Vision and Applications*, Vol. 6, pp. 43–47 (2014).

[52] Zhang, Z.: A flexible new technique for camera calibration, *TPAMI*, Vol. 22, No. 11, pp. 1330–1334 (2000).

[53] Moreno, D. and Taubin, G.: Simple, accurate, and robust projector-camera calibration, *Proc. 3DIMPVT*, pp. 464–471 (2012).

[54] Gupta, M., Agrawal, A., Veeraraghavan, A. and Narasimhan, S. G.: Structured light 3D scanning in the presence of global illumination, *CVPR 2011*, pp. 713–720 (online), DOI: 10.1109/CVPR.2011.5995321 (2011).

[55] Agrawal, A. and Ramalingam, S.: Single Image Calibration of Multi-axial Imaging Systems, *Proc. CVPR*, pp. 1399 –1406 (2013).

[56] Takahashi, K., Miyata, A., Nobuhara, S. and Matsuyama, T.: A Linear Extrinsic Calibration of Kaleidoscopic Imaging System From Single 3D Point, *Proc. CVPR* (2017).

[57] Reshetouski, I., Manakov, A., Seidel, H.-P. and Ihrke, I.: Three-Dimensional Kaleidoscopic Imaging, *Proc. CVPR*, pp. 353–360 (2011).

[58] Tian, L., Wang, J. and Waller, L.: 3D differential phase-contrast microscopy with computational illumination using an LED array, *Optics letters*, Vol. 39, pp. 1326–9 (2014).

[59] Chen, M., Phillips, Z. F. and Waller, L.: Quantitative differential phase contrast (DPC) microscopy with computational aberration correction, *Opt. Express*, No. 25, pp. 32888–32899.

[60] Niederöst, M., Niederöst, J. and Scucka, J.: 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. vol. XXXIV-5/W10, pp. 5–10 (2002).

[61] Karthik, S. and Rajagopalan, A. N.: Underwater Microscopic Shape from Focus, *2014 22nd International Conference on Pattern Recognition*, pp. 2107–2112 (2014).

[62] Mullen, A. D., Treibitz, T., Roberts, P. L., Kelly, E. L. A., Horwitz, R., Smith, J. E. and Jaffe, J. S.: Underwater microscopy for in situ studies of benthic ecosystems, *Nature Communications*, Vol. 7 (online), DOI: 10.1038/ncomms12093 (2016). n/a.

[63] Starck, J., Maki, A., Nobuhara, S., Hilton, A. and Matsuyama, T.: The Multiple-Camera 3-D Production Studio, *TCSVT*, Vol. 19, No. 6, pp. 856 –869 (2009).

[64] Joo, H., Simon, T., Li, X., Liu, H., Tan, L., Gui, L., Banerjee, S., Godisart, T., Nabbe, B., Matthews, I., Kanade, T., Nobuhara, S. and Sheikh, Y.: Panoptic Studio: A Massively Multiview System for Social Interaction Capture, *TPAMI*, Vol. 41, No. 1, pp. 190–204 (2019).