

Surface Normals and Shape From Water

Meng-Yu Jennifer Kuo, Satoshi Murai, Ryo Kawahara, Shohei Nobuhara, *Member, IEEE*, and Ko Nishino, *Senior Member, IEEE*

Abstract—In this paper, we introduce a novel method for reconstructing surface normals and depth of dynamic objects in water. Past shape recovery methods have leveraged various visual cues for estimating shape (e.g., depth) or surface normals. Methods that estimate both compute one from the other. We show that these two geometric surface properties can be simultaneously recovered for each pixel when the object is observed underwater. Our key idea is to leverage multi-wavelength near-infrared light absorption along different underwater light paths in conjunction with surface shading. Our method can handle both Lambertian and non-Lambertian surfaces. We derive a principled theory for this surface normals and shape from water method and a practical calibration method for determining its imaging parameters values. By construction, the method can be implemented as a one-shot imaging system. We prototype both an off-line and a video-rate imaging system and demonstrate the effectiveness of the method on a number of real-world static and dynamic objects. The results show that the method can recover intricate surface features that are otherwise inaccessible.

Index Terms—Computational Photography, Underwater Reconstruction, Near-infrared Light, Absorption.

1 INTRODUCTION

COMPUTER vision research has produced a variety of successful methods for 3D geometry reconstruction that exploit different visual cues ranging from focus to texture. Most of these shape-from-X approaches, however, recover either the 3D coordinates (*i.e.*, shape) or the surface normals, but not both. Surface normals are computed from the estimated shape or the shape is integrated from the surface normals. Although a number of methods have demonstrated combinations of these contrasting recovery approaches, they require multiple views or active depth sensors and fundamentally cannot be used in a single exposure setup, which precludes the possibility of dynamic surface reconstruction. Estimating both surface normals and 3D coordinates at each and every observable pixel simultaneously, but not as byproducts of each other, is crucial for 3D reconstruction of general surfaces that can have arbitrary intricate geometric features. Avoiding the use of multi-view appearance matching is also essential to handle textureless surfaces.

In this paper, we show that per-pixel surface normals and shape can be simultaneously but separately recovered for an object immersed in water. In other words, as shown in Fig. 1, we introduce a novel 3D sensing method to directly recover 3D geometry as oriented points. Underwater 3D reconstruction may sound peculiar and limiting, but it finds significant applications in a wide range of fields including medicine (*e.g.*, endoscopy), biology, oceanography, archaeology, as well as general surveillance and navigation. Moreover, immersing objects in water for measurements is non-invasive as long as the object is nonabsorbent and is as practical as other 3D reconstruction methods.

Our key idea is to leverage multi-wavelength near-infrared light absorption along different underwater light paths in conjunction with surface shading. The use of near-infrared light absorption for shape recovery builds upon recent work by Asano *et al.* [1]. The integration of shading cues to this infrared spectral imaging is,

however, nontrivial since unlike regular photometric stereo, light is attenuated. We derive a principled theory for this dense depth and surface normal recovery, which we refer to as *surface normals and shape from water*.

We show that surface normals and shape from water requires at least four near-infrared directional light sources, each illuminating the object surface whose radiance is captured with an orthographic camera. When using four light sources, the theory reveals that one of the light sources, which we refer to as the base light source, should lie within the convex cone spanned by other light sources, and that the remaining light sources can have the same polar angle with respect to the viewing direction as long as they realize different *effective absorption coefficients* and also span a 3D space. Most important, we show when and how the depth and surface normals can be separately and uniquely estimated, leading to the identification of preferred combinations of directions and wavelengths of light sources. Underwater creatures, such as swimming fish, often exhibit strong specularity in their appearances. For this, we also extend our method to handle non-Lambertian surfaces by jointly estimating the surface reflectance and 3D geometry. Furthermore, we derive a practical calibration method that automatically estimates light source directions and intensities.

We demonstrate the effectiveness of our method on a number of static and dynamic real-world objects with complex shape. We implement the method with two imaging systems, one for off-line capture using an off-the-shelf monochromatic camera and interchangeable near-infrared bandpass filters, and another for video-rate capture using a custom-built multi-wavelength camera. Experimental results demonstrate the method’s ability to recover intricate details of shape that dynamically change, which would be challenging for conventional methods.

The main contributions of this paper are

- Meng-Yu Jennifer Kuo, Satoshi Murai, Ryo Kawahara, Shohei Nobuhara, and Ko Nishino are with the Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University, Kyoto, JP. E-mail: see <https://vision.ist.i.kyoto-u.ac.jp/>
- A novel near-infrared multi-wavelength imaging method for simultaneous per-pixel depth and surface normal recovery of both Lambertian and non-Lambertian objects.
- Thorough theoretical analysis of the method including

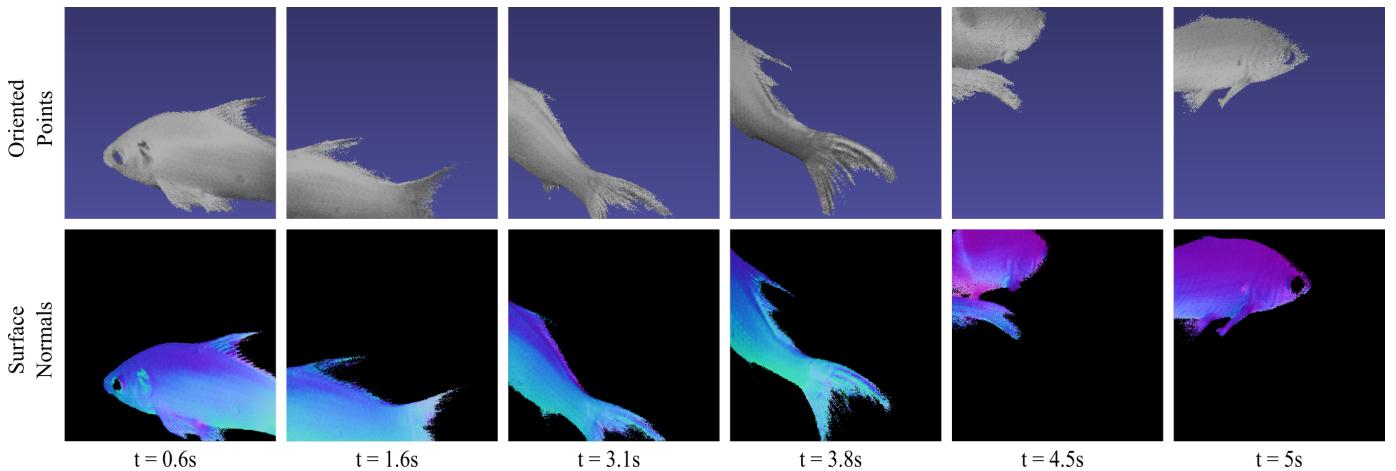


Fig. 1: Our method enables video-rate, per-pixel surface normal and depth recovery of dynamic underwater objects without any artificial constraints on one another such as smoothness. The recovered oriented points retain intricate surface details including sharp geometric features that would otherwise be hard to capture with conventional methods.

identification of minimum configuration conditions for independent shape and normal estimation.

- Demonstration of video-rate accurate normal and shape recovery of complex dynamic surfaces.

2 RELATED WORKS

We first discuss 3D reconstruction methods that focus on recovering both surface normals and shape, and then review other underwater surface normal or shape reconstruction methods.

2.1 Surface Normals and Shape Reconstruction in Air

A representative early work that achieves simultaneous pixel-wise surface normal and shape estimation is Helmholtz stereo [2], [3] which can also handle non-Lambertian surfaces. The method, however, requires interchanged viewing and lighting directions which is hard to achieve for dynamic targets. For non-Lambertian surface reconstruction, Zhou *et al.* [4] introduced a multi-spectral light field system that enables single-shot simultaneous recovery of surface normal and depth. Their depth estimation relies on multi-view photo-consistency which is further refined by normal integration. Multi-view image capture significantly restricts the recoverable frustum making it hard to realize for live underwater creatures. Furthermore, our goal is to recover depth and surface normal as independent quantities at each pixel.

In general, for image-based 3D reconstruction, spatial disparity in stereo and temporal disparity in shape-from-motion carry depth information of the scene, while shading (*e.g.*, photometric stereo and shape-from-shading), polarization, and distortion (*e.g.*, shape-from-texture) carry surface normal information. Naturally, most past methods for estimating both surface normals and shape combine these reconstruction cues.

Patch-based stereo methods such as PMVS [5] and Patch-Match stereo [6] explicitly model the scene as a collection of oriented points. They, however, cannot provide pixel-wise reconstruction of surface normals and shape as they require local support of a certain size for evaluating the stereo matching with slanted windows whose shape deforms according to the hypothesized depth and normal.

Surface normal estimation combined with 3D shape reconstruction can also be found in methods for refining outputs of active depth sensors. Kadambi *et al.* [7] proposed a polarization-based depth enhancement method, and Wu *et al.* [8] and Yu *et al.* [9] have proposed those based on shape-from-shading. These methods, however, require an initial estimate of the scene geometry provided by the depth sensor (*e.g.*, ToF or active stereo), since shading or polarization themselves do not carry sufficient information to estimate the normals independently. As such, the normal estimates are dependent on the depth. One can interpret our method as using near-infrared light absorption in lieu of an active depth sensor. Our approach, however, directly recovers both the normals and depth as separate estimates.

Photometric stereo combined with depth sensors [10], [11], structured light [12], and structure-from-motion [13], [14] can also return both surface normals and shape. These methods, however, require multi-view measurement of the target and/or coarse-to-fine iterative alternating estimation, which renders single-shot extensions impossible.

For surface normal recovery, Hernández *et al.* [15] introduced the use of RGB lighting enabling single shot photometric stereo. Our method similarly uses multi-wavelength light that can be captured in a single exposure, but in the near-infrared spectrum.

2.2 3D Reconstruction in Water

Asano *et al.* [1] proposed a bispectral imaging approach called *shape from water* for underwater depth estimation. It estimates the scene depth based on the differences in absorption by water at different near-infrared wavelengths. A recent approach that achieves full (fuller) 3D shape reconstruction of underwater non-rigid objects [16] also leverages this bispectral near-infrared imaging principle. Their method, however, can only recover depth and relies on surface textures to establish correspondences. Other approaches that do not require reflection cues at the object surface such as 3D range scanning by fluorescence immersion [17] and tomographic reconstruction using visible light [18] also only recovers depth. In contrast, our method can recover both depth and surface normal at each pixel regardless of its texture.

Most other studies on 3D reconstruction in water such as underwater stereo [19], refractive structure-from-motion [20], un-

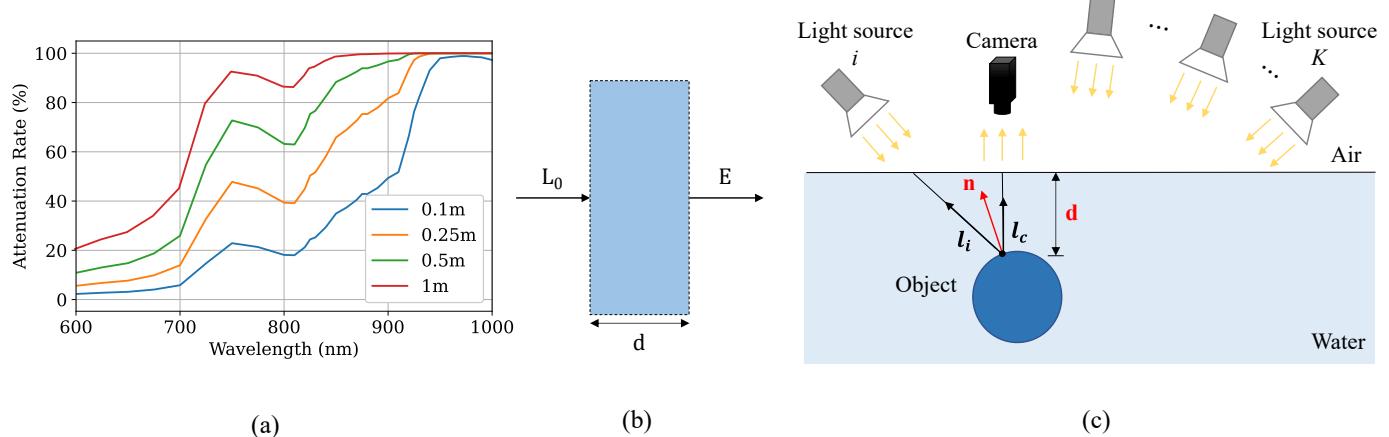


Fig. 2: We leverage (a) wavelength-dependent near-infrared light absorption by water (reprinted from [16]) which can be modeled with (b) the Beer-Lambert law that relates light attenuation with the absorption coefficient and light path length. We derive the theory for and implement (c) a multi-wavelength imaging system to achieve simultaneous estimation of depth and surface normals from multi-path light attenuation and surface shading due to multi-directional near-infrared lighting.

derwater photometric stereo [21], [22], shape from chromatic dispersion [23], [24], and underwater depth from light-field imaging [25] assume absorption by water plays little effect and focus on applying in-air conventional 3D reconstruction methods to underwater imagery. Some other works leverage multi-view image sets to recover both water surface [26] and underwater scene [27] using appearance matching. Their multi-view approaches, however, are fundamentally limited to well-textured scenes. Tsotsios *et al.* [28] explicitly model the attenuation by medium for photometric stereo. Their method, however, is an iterative approach involving a normal integration process for depth recovery.

To our knowledge, our work is the first to achieve simultaneous underwater recovery of per-pixel surface normals and depth.

3 BISPECTRAL SHAPE FROM WATER

Although water in the visual spectrum has extremely small absorption coefficients (which is why water appears transparent), as shown in Fig. 2(a), it increases quickly in the sub-meter range for near-infrared light. As depicted in Fig. 2(b), light absorption in water can be accurately modeled with the Beer-Lambert law [29] that expresses transmitted radiance E as that of incident light L_0 which exponentially decays as a function of the distance d it travels

$$E = L_0 \exp[-\alpha(\lambda)d], \quad (1)$$

where λ is the wavelength of light, and $\alpha(\lambda)$ is the wavelength-dependent absorption coefficient.

Asano *et al.* [1] exploited this near-infrared light absorption for shape recovery, by capturing the same surface with two near-infrared wavelengths. In particular, they use one broad near-infrared directional light source and a coaxial camera system with two distinct near infrared filters (*e.g.*, at 905nm and 950nm). They showed through extensive measurements that the reflectance of object surfaces do not vary much in the near infrared range, and thus by taking the ratio of the bispectral near-infrared observations, light path length to and from the object surface and hence its depth can be estimated. We also leverage their findings that the spectral characteristics of surface reflectance can be considered invariant to wavelength in the near-infrared range.

4 SURFACE NORMALS AND SHAPE FROM WATER

Our goal is to simultaneously estimate the depth and surface normals of underwater objects without imposing unwanted constraints on their dependency such as spatial smoothness. We also aim to derive a method that can be implemented as a real-time 3D sensing system. We show that we can achieve these by capturing the target object under near-infrared directional light sources of different wavelengths and directions with an orthographic camera.

4.1 Near-Infrared Multi-Wavelength Imaging

As illustrated in Fig. 2(c), we consider an orthographic camera oriented fronto-parallel to the flat water surface. The object in water is illuminated by monochromatic directional light sources each from a different direction. Let \mathbf{l}_c and \mathbf{l}_i denote the viewing direction and the direction of light source i with wavelength λ_i and intensity L_i , respectively.

Incident light to a surface point x of depth $d(x)$ from the water surface travels distance $\frac{d(x)}{\mathbf{l}_c^\top \mathbf{l}_i}$ in water before reflecting at x and traveling $d(x)$ into the camera. Recall that near-infrared light is not absorbed in air, but only in water.

From Eq. 1, for a surface with a factorized reflectance function $s(\lambda)r(\omega)$ where $s(\lambda)$ is the spectral component and $r(\omega)$ is the geometric component of reflectance, the intensity of the light captured by the camera becomes

$$\begin{aligned} E_i &= s(\lambda_i)r(\omega_i)L_i \exp \left[- \left(1 + \frac{1}{\mathbf{l}_c^\top \mathbf{l}_i} \right) \alpha(\lambda_i)d(x) \right] \\ &= s(\lambda_i)r(\omega_i)L_i \exp [-\hat{\alpha}_i d(x)], \end{aligned} \quad (2)$$

where we have defined the *effective absorption coefficient*

$$\hat{\alpha}_i = \left(1 + \frac{1}{\mathbf{l}_c^\top \mathbf{l}_i} \right) \alpha(\lambda_i). \quad (3)$$

As we exploit surface shading, in addition to near-infrared light absorption, we assume Lambertian surfaces. Lambertian reflection has a factorized reflectance function where the albedo is the spectral term and the shading is the geometric term. Following Asano *et al.* [1], we assume that the spectral term, in other words

the per-pixel albedo, can be approximated to be invariant to the near-infrared wavelength of light

$$s(\lambda_i) = \rho. \quad (4)$$

The geometric term (*i.e.* shading) can be expressed as

$$r(\omega_i) = \mathbf{l}_i^\top \mathbf{n}(x), \quad (5)$$

using the surface normal $\mathbf{n}(x)$. Therefore we have

$$E_i = \rho \mathbf{l}_i^\top \mathbf{n}(x) L_i \exp[-\hat{\alpha}_i d(x)]. \quad (6)$$

Note that we do not make any assumptions on the spatial variation of surface albedo.

Suppose another observation of the same point (*i.e.*, the same pixel position in the camera image) is given for another near-infrared light source j

$$E_j = \rho \mathbf{l}_j^\top \mathbf{n}(x) L_j \exp[-\hat{\alpha}_j d(x)]. \quad (7)$$

If we assume $L_i = L_j$ by using light sources of equal intensities, we obtain

$$\begin{aligned} \frac{E_j}{E_i} &= \frac{\mathbf{l}_j^\top \mathbf{n}(x) \exp[-\hat{\alpha}_j d(x)]}{\mathbf{l}_i^\top \mathbf{n}(x) \exp[-\hat{\alpha}_i d(x)]}, \\ \frac{\mathbf{l}_j^\top \mathbf{n}(x)}{\mathbf{l}_i^\top \mathbf{n}(x)} &= \frac{E_j}{E_i} \exp[(\hat{\alpha}_j - \hat{\alpha}_i)d(x)]. \end{aligned} \quad (8)$$

Given observations taken under K distinct light sources, we obtain

$$\begin{aligned} \frac{1}{\mathbf{l}_1^\top \mathbf{n}(x)} \mathbf{L} \mathbf{n}(x) &= \frac{1}{E_1} \begin{bmatrix} E_2 \exp[(\hat{\alpha}_2 - \hat{\alpha}_1)d(x)] \\ \vdots \\ E_K \exp[(\hat{\alpha}_K - \hat{\alpha}_1)d(x)] \end{bmatrix}, \\ &= \mathbf{g}(d(x)), \end{aligned} \quad (9)$$

where the light source matrix $\mathbf{L} = [\mathbf{l}_2 \dots \mathbf{l}_K]^\top \in \mathbb{R}^{(K-1) \times 3}$. We refer to \mathbf{l}_1 as the base light source and all the other light sources (*i.e.*, those in the light matrix \mathbf{L}) as auxiliary light sources. Table 1 summarizes primary notation.

4.2 Depth and Surface Normal Recovery

Eq. 9 is a system of $K - 1$ non-linear equations of both depth $d(x)$ and surface normal $\mathbf{n}(x)$. Although it does not have a closed-form solution, we can recover a unique depth value $d(x)$ and surface normal $\mathbf{n}(x)$ for the surface point x from it. Let us first left-multiply the Moore-Penrose pseudo-inverse matrix $\mathbf{L}^+ = (\mathbf{L}^\top \mathbf{L})^{-1} \mathbf{L}^\top$ to both sides of Eq. 9:

$$\frac{1}{\mathbf{l}_1^\top \mathbf{n}(x)} \mathbf{n}(x) = \mathbf{L}^+ \mathbf{g}(d(x)). \quad (10)$$

Then multiplying \mathbf{l}_1^\top from the left yields

$$\begin{aligned} \frac{\mathbf{l}_1^\top \mathbf{n}(x)}{\mathbf{l}_1^\top \mathbf{n}(x)} &= \mathbf{l}_1^\top \mathbf{L}^+ \mathbf{g}(d(x)), \\ 1 &= \mathbf{b}\mathbf{g}(d(x)). \end{aligned} \quad (11)$$

where $\mathbf{b} = \mathbf{l}_1^\top \mathbf{L}^+ (\in \mathbb{R}^{1 \times (K-1)})$. Note that Eq. 11 is a sum-of-exponents function of $d(x)$ and does not have an analytical solution.

We derive conditions on Eq. 11 for a unique and global $d(x)$, which turns out to be numerical root finding of a monotonic function as we show in Section 4.3. This means its global optimal solution can be obtained with conventional numerical optimizers

TABLE 1: Nomenclature

Calibration parameters	
$\alpha(\lambda_i)$	Absorption coefficient
\mathbf{l}_i	Light source direction
L_i	Light source intensity
Surface parameters	
$d(x)$	Depth
$\mathbf{n}(x)$	Surface normal
ρ_d, r_s	Diffuse albedo and specular reflectance

such as Newton-Raphson, efficiently. Moreover, the optimization can utilize the first and second-order derivatives of $\mathbf{b}\mathbf{g}(d(x))$, thanks to its simple sum-of-exponents form.

Given the depth $d(x)$ from Eq. 11, we can compute the surface normal scaled by $\mathbf{l}_1^\top \mathbf{n}(x)$ using the right-hand side of Eq. 10. The surface normal $\mathbf{n}(x)$ is then given by

$$\mathbf{n}(x) = \frac{\mathbf{L}^+ \mathbf{g}(d(x))}{\|\mathbf{L}^+ \mathbf{g}(d(x))\|}. \quad (12)$$

As this derivation shows, we can recover the surface normal and depth at each pixel in the overlapping area of light sources captured from the single viewpoint. It is important to note that auxiliary geometric constraints on the object surface, such as smoothness, are not assumed.

4.3 Conditions for Unique Recovery

Let us now analyze the conditions for unique depth and surface normal recovery. Once the depth is estimated, from Eq. 12, we observe that the rank of matrix \mathbf{L} must be at least 3. This condition on light source directions, that they should span the 3D space, is the same as in regular photometric stereo. This, however, also means that in addition to the base light source, we need at least three auxiliary light sources, making the total number of light sources to be at least four $K \geq 4$.

When recovering depth, from Eq. 11, we observe that the effective absorption coefficients of the auxiliary light sources should all differ from that of the base light source

$$[\hat{\alpha}_2 - \hat{\alpha}_1 \dots \hat{\alpha}_K - \hat{\alpha}_1] \neq \mathbf{0}. \quad (13)$$

Recall that the effective absorption coefficient (Eq. 3) is a function of both the absorption coefficient α_i and shading $\mathbf{l}_i^\top \mathbf{l}_i$. This requirement of the effective absorption coefficient does not necessarily mean that all light sources should have different wavelength or that they should have different polar angles with respect to the viewing direction as they can be combined to satisfy Eq. 13.

The right hand side of Eq. 11 is a general exponential-sum function of $d(x)$ and therefore does not have a closed-form solution. It, however, is a monotonic function of $d(x)$, if all signs of pairs of b_i (*i.e.*, the i -th element of \mathbf{b}) and $\hat{\alpha}_i - \hat{\alpha}_1$ match. In reality, the light sources should be all on the same hemisphere, which makes the elements of \mathbf{b} either all non-negative or negative. In this case, monotonicity holds when $\hat{\alpha}_i - \hat{\alpha}_1$ are all non-negative or negative for $i = 2, \dots, K$, respectively, since a positive sum of monotonic functions is also monotonic. Requiring that the effective absorption coefficient of the base light source $\hat{\alpha}_1$ be the minimum of those of all light sources instead of the maximum also has the advantage for robust depth estimation as

discussed in Section 4.4. In this case, since all differences in effective absorption coefficient from the base light source becomes positive, we can obtain a unique solution when all elements of \mathbf{b} are non-negative.

By the definition of pseudo inverse matrix [30], $\mathbf{b} = \mathbf{l}_1^\top \mathbf{L}^+$ is the minimum norm solution for reconstructing \mathbf{l}_1 as a linear combination of $\mathbf{l}_2 \cdots \mathbf{l}_k$. In the case of $K = 4$, i.e., when we use three auxiliary light sources, this suggests that if the base light source lies within the cone spanned by the other three light sources, (as shown in Fig. 3(a)), all the elements of \mathbf{b} will be non-negative and we can attain a global unique solution. This also intuitively means that making all elements of vector \mathbf{b} be negative is not physically feasible, although it can also guarantee the monotonicity, since it corresponds to each of the auxiliary light sources illuminating the target from the opposite side of the base light source.

In the case of $K > 4$, identifying the space in which $\mathbf{b} \geq 0$ holds is not trivial. In practice, however, since $\hat{\alpha}_i$ and \mathbf{b} depend only on the absorption coefficients and the lighting directions, and are shared by all pixels, once the illumination directions are calibrated, we can immediately verify minimum effective absorption coefficient of the base light source $\hat{\alpha}_1$ and the non-negativity of elements \mathbf{b} . That is, once the calibrated light source directions satisfy the non-negative constraint, the depth estimation by function Eq. 11 is guaranteed to be a monotonic function that results in a unique and global solution.

In summary, the following conditions should be satisfied when implementing surface normals and shape from water.

- At least four near-infrared directional light sources.
- Auxiliary light source directions should be independent from each other.
- The effective absorption coefficient of auxiliary light sources should be different from that of the base light source.
- All elements of \mathbf{b} should be non-negative.

These conditions leave room for different combinations of wavelengths and directions of the light sources. For instance, if we choose auxiliary light sources to have the same wavelength, that wavelength should be different from the base light source and each of their directions should differ from one another as well as the viewing direction. If the auxiliary light source wavelengths are all different, they can all be situated such that they make the same polar angle with respect to the viewing direction. Note that the latter case enables image capture of all the necessary information for surface normal and shape recovery in a single exposure, as the multi-wavelength near-infrared lights do not interfere with each other.

4.4 Depth Accuracy Analysis

If we hypothetically set all lights to have the same direction $\mathbf{l}_i = \mathbf{l}_1$, the right hand side of Eq. 9 becomes an all-one vector and the depth estimate becomes

$$d(x) = -\frac{1}{\hat{\alpha}_k - \hat{\alpha}_1} \ln \left(\frac{E_k}{E_1} \right), \quad (14)$$

for every k . This is equivalent to the depth estimate from bispectral shape from water [1]. Eq. 14 suggests that the estimation of depth becomes more accurate as the ratio of the observed radiance becomes larger. As this condition needs to apply to all pairs of

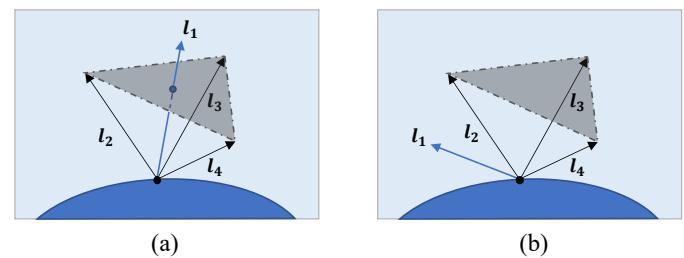


Fig. 3: For the minimal image configuration case of $K = 4$, we show that the base light source depicted as \mathbf{l}_1 must lie within the convex cone of the auxiliary light sources (a). If this condition is not met (b), a global and unique estimate is not guaranteed.

light sources with respect to the base light source, in practice this means that the difference in observed radiance from that by the base light source should be as large as possible for all auxiliary light sources.

If we consider additive noise ϵ to the input images, the absolute depth error Δd becomes

$$\begin{aligned} \Delta d &= -\frac{1}{\hat{\alpha}_k - \hat{\alpha}_1} \left(\ln \frac{E_k + \epsilon}{E_1 + \epsilon} - \ln \frac{E_k}{E_1} \right) \\ &= -\frac{1}{\hat{\alpha}_k - \hat{\alpha}_1} \ln \left(\frac{1 + \frac{\epsilon}{E_k}}{1 + \frac{\epsilon}{E_1}} \right). \end{aligned} \quad (15)$$

Eq. 15 shows that the absolute error decreases (*i.e.*, becomes robust to noise) as the observed radiances become larger.

These results suggest that we should choose the wavelengths and light source directions so that the differences of effective absorption coefficients between auxiliary light sources and the base light source are maximized, while the observed radiance due to each corresponding light source is also made as large as possible. As the attenuation rate changes rapidly from 700nm to 1000nm for the sub-meter range (Fig. 2(a)), we can use the proposed multi-wavelength imaging principle in this near-infrared range for the recovery.

4.5 Calibration

Up to this point, we have assumed that the projection model of the camera is strictly orthographic, the light sources are directional and have equal intensities. In practice, these requirements are not necessarily met, and we must account for any deviation, for instance, by scaling observations by the light source intensities. Although light source directions can be estimated by placing a chrome ball at where the target object will be situated and by using the highlights, the estimates can be erroneous. Besides, if the light sources emit different intensities, it will lead to inevitable errors in our shape recovery since we cannot cancel out those unknown incident intensities. We derive a practical calibration method that simultaneously estimates both light source directions and intensities to achieve robust and accurate estimation of shape and surface normals.

In particular, we immerse a Lambertian sphere at different known depths (at least two positions) from the camera, within the space the target objects will cover. As the radius of the sphere is known beforehand, we can compute the ground truth of depth $\hat{d}(x)$ and surface normal $\hat{\mathbf{n}}(x)$ at any given point x on the sphere. The light source direction vectors $\mathbf{l} = \{\mathbf{l}_1 \cdots \mathbf{l}_K\}$ and their intensities

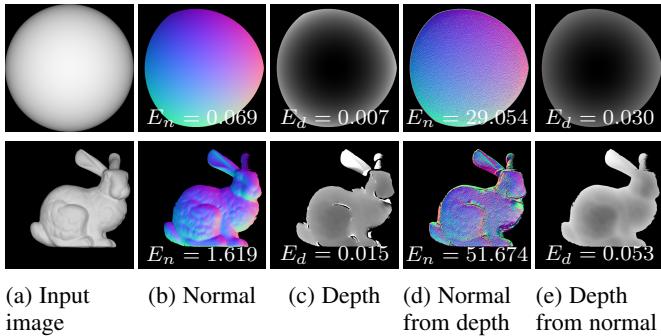


Fig. 4: Quantitative evaluation with synthetic Lambertian objects. (a) Input images synthesized with the base light source. (b)(c) Normal and depth maps estimated by the proposed method. (d)(e) Normal map computed from the estimated depth map and vice versa. Our method achieves higher accuracy for both.

$L = \{L_1 \cdots L_K\}$ can then be estimated by minimizing the L^2 error

$$\arg \min_{\mathbf{l}, L} \sum_{i=1}^n \sum_{j=1}^m \left(k_1 \|\hat{d}(x_j^i) - d(\mathbf{l}, L, x_j^i)\| + k_2 \|1 - \hat{\mathbf{n}}(x_j^i) \cdot \mathbf{n}(\mathbf{l}, L, x_j^i)\| \right), \quad (16)$$

where n is the number of captured images, m is the number of valid pixels, k_1 and k_2 are scalar weights, and d and \mathbf{n} are the estimated depth and normals, respectively.

5 RECOVERY OF NON-LAMBERTIAN SURFACES

For an object with non-Lambertian reflection, we employ the dichromatic model [31] that expresses it as a linear combination of diffuse and specular reflection

$$(\rho_d + (r_s)_i) \mathbf{l}_i^\top \mathbf{n}(x), \quad (17)$$

where ρ_d is the diffuse albedo, and r_s is the specular reflectance. The intensity of the light captured by the camera becomes

$$\begin{aligned} E_i &= (\rho_d + (r_s)_i) \mathbf{l}_i^\top \mathbf{n}(x) L_i \exp[-\hat{\alpha}_i d(x)] \\ &= \boldsymbol{\rho}^\top \mathbf{r}_i \mathbf{l}_i^\top \mathbf{n}(x) L_i \exp[-\hat{\alpha}_i d(x)], \end{aligned} \quad (18)$$

where $\boldsymbol{\rho} = [\rho_d \ 1]$ and $\mathbf{r} = [1 \ (r_s)_i]$.

Let us denote another light source coming from a different direction as

$$E_j = \boldsymbol{\rho}^\top \mathbf{r}_j \mathbf{l}_j^\top \mathbf{n}(x) L_j \exp[-\hat{\alpha}_j d(x)]. \quad (19)$$

By dividing Eq. 19 by Eq. 18, we obtain

$$\begin{aligned} \frac{E_j}{E_i} &= \frac{\boldsymbol{\rho}^\top \mathbf{r}_j \mathbf{l}_j^\top \mathbf{n}(x) L_j \exp[-\hat{\alpha}_j d(x)]}{\boldsymbol{\rho}^\top \mathbf{r}_i \mathbf{l}_i^\top \mathbf{n}(x) L_i \exp[-\hat{\alpha}_i d(x)]}, \\ \frac{\boldsymbol{\rho}^\top \mathbf{r}_j \mathbf{l}_j^\top \mathbf{n}(x)}{\boldsymbol{\rho}^\top \mathbf{r}_i \mathbf{l}_i^\top \mathbf{n}(x)} &= \frac{E_j}{E_i} \exp[(\hat{\alpha}_j - \hat{\alpha}_i) d(x)]. \end{aligned} \quad (20)$$

Given observations under K light sources from different directions, we then have

$$\begin{aligned} \mathbf{P} \frac{1}{\mathbf{l}_1^\top \mathbf{n}(x)} \mathbf{L} \mathbf{n}(x) &= \frac{1}{E_1} \begin{bmatrix} E_2 \exp[(\hat{\alpha}_2 - \hat{\alpha}_1) d(x)] \\ \vdots \\ E_K \exp[(\hat{\alpha}_K - \hat{\alpha}_1) d(x)] \end{bmatrix}, \\ &= \mathbf{g}(d(x)), \end{aligned} \quad (21)$$

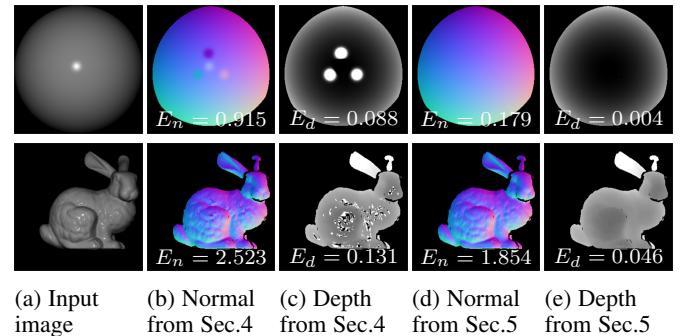


Fig. 5: Quantitative evaluation with synthetic non-Lambertian objects. (a) Input image synthesized with the base light source. (b)(c) Normal and depth maps estimated by the proposed method in Sec. 4. (d)(e) Normal and depth maps estimated by the proposed method in Sec. 5. Note that the object surface is recovered in the common surface region that received sufficient light from all light sources. Our method accurately reconstructs both surface normals and depth values even for the objects with non-Lambertian reflection.

where \mathbf{L} is the light source direction matrix and $\mathbf{P} = \text{diag}\left(\frac{\boldsymbol{\rho}^\top \mathbf{r}_2, \dots, \mathbf{r}_K}{\boldsymbol{\rho}^\top \mathbf{r}_1}\right)$ is a $(K-1) \times (K-1)$ diagonal matrix of the reflection components.

Similar to Eq. 11, by left-multiplying the inverse matrix of \mathbf{P} , the Moore-Penrose pseudo inverse matrix of \mathbf{L} , and the direction vector of the base light source \mathbf{l}_i to both sides of Eq. 21 we obtain

$$\begin{aligned} 1 &= \mathbf{l}_1^\top \mathbf{L}^+ \mathbf{P}^{-1} \mathbf{g}(d(x)) \\ &= \mathbf{b} \mathbf{P}^{-1} \mathbf{g}(d(x)). \end{aligned} \quad (22)$$

The diffuse reflection component is usually smooth (*i.e.*, its gradient changes slowly), and the specular reflection component is sparse since specularities are observed only under a small number of light source directions. We first assume that the target object is Lambertian and obtain the initial estimates from the proposed method in Sec. 4.2. These initial estimates can then be refined by alternating between estimating depth $d(x)$ and reflectance \mathbf{P} at each pixel until convergence

$$\begin{aligned} \arg \min_{\mathbf{P} \text{ or } d(x)} &\| \mathbf{b} \mathbf{P}^{-1} \mathbf{g}(d(x)) - 1 \|_2^2 \\ &+ k_1 \| \nabla \rho_d(x) \|_2^2 + k_2 \sum_{i=1}^K \| \nabla (r_s)_i(x) \|_1, \end{aligned} \quad (23)$$

$$\text{subject to } 0 \leq \rho_d(x) \leq 1, 0 \leq (r_s)_i(x) \leq 1, 0 < d(x), \forall x \quad (24)$$

where k_1 and k_2 are scalar weights, and $\nabla \rho_d$ and ∇r_s are the gradients of diffuse and specular reflection components, respectively. As with any non-linear optimization, the system converges to a local minimum that may not represent the best possible global solution. We estimate the gradient of Eq. 23 and solve it in an iterative fashion to find a better local minimum (*i.e.*, repeatedly restarting the optimizer from a new position). In the experiments below, 4–5 times of iterations were sufficient for convergence.

Given the depth $d(x)$ and reflectance \mathbf{P} from Eq. 23, the surface normal $\mathbf{n}(x)$ is

$$\mathbf{n}(x) = \frac{\mathbf{L}^+ \mathbf{P}^{-1} \mathbf{g}(d(x))}{\| \mathbf{L}^+ \mathbf{P}^{-1} \mathbf{g}(d(x)) \|}. \quad (25)$$

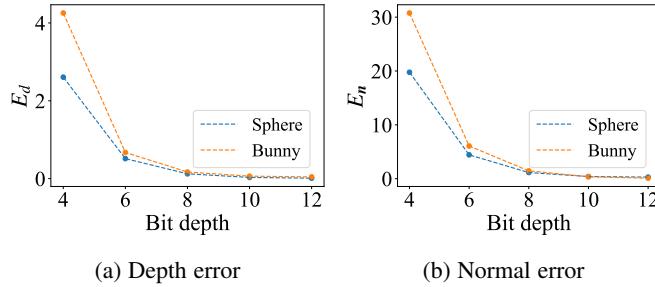


Fig. 6: Estimation errors at different image bit depths. Colors denote different objects.

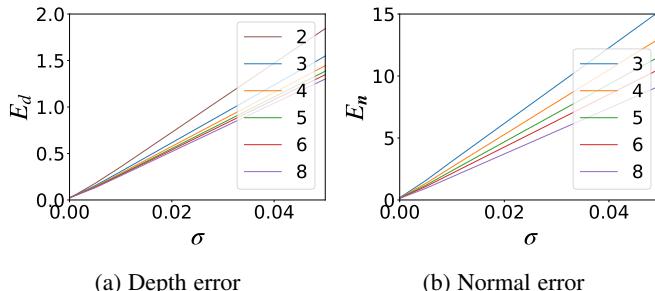


Fig. 7: Estimation errors at different noise levels. Colors denote different numbers of auxiliary light sources.

This derivation shows that we can recover the per-pixel surface normal and depth even for the non-Lambertian surfaces.

6 EXPERIMENTAL RESULTS

We experimentally validate our method using synthetic and near-infrared real light sources in the range of 900nm to 1000nm. We assume that the minimum value that $\mathbf{l}_c^\top \mathbf{l}_i$ can take is $1/\sqrt{2} \simeq 0.71$ when \mathbf{l}_i is tilted 45° from \mathbf{l}_c , while making each \mathbf{l}_i be independent from each other. The absorption coefficient in this wavelength range varies strongly from roughly 5×10^{-3} to 3×10^{-2} .

6.1 Quantitative Evaluation with Synthetic Data

We quantitatively evaluate the reconstruction accuracy and noise robustness of our method using synthetic images rendered with PBRT [32]. We render the images with an orthographic camera facing the target under K light sources, one coaligned with the camera as the base light source and others around it forming a regular $(K - 1)$ -sided polygon as auxiliary light sources. The K absorption coefficients are defined as an arithmetic sequence from 5×10^{-3} to 3×10^{-2} .

Fig. 4 shows results of Lambertian objects for the minimal imaging configuration of $K = 4$. In addition to the recovered surface normals and depth by our method (a,b), Fig. 4 shows surface normals computed from the estimated depth by numerical differentiation (c), and depth integrated from estimated surface normals [33] (d), each of which corresponds to how surface normals and shape would be estimated using conventional methods such as stereo and photometric stereo, respectively. The errors E_n and E_d represent RMSEs of the estimated normals in degrees and the estimated depth normalized by the object size, respectively. These results clearly demonstrate the fundamental advantage of our simultaneous per-pixel estimation of surface normal and depth.

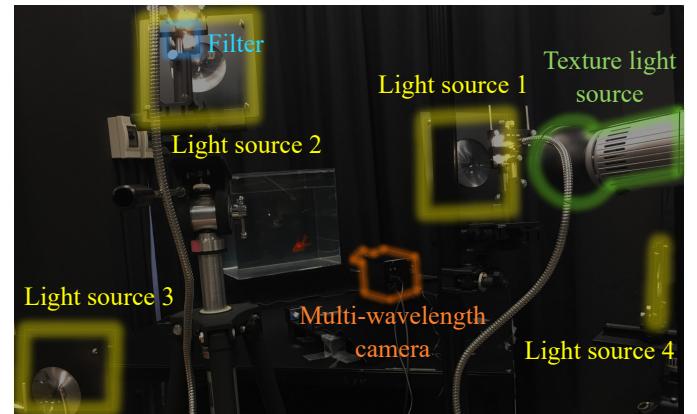


Fig. 8: We implement a video-rate surface normal and shape from water imaging system using four light sources each placed with a Fresnel lens and a near-infrared bandpass filter and a custom-built multi-wavelength camera. A visible spectrum light source is used to also capture texture. This texture light source does not interfere with the near-infrared light sources.

Conventional methods inevitably rely on one estimate to obtain the other, which results in significant reduction in accuracy.

Fig. 5 shows results of non-Lambertian objects observed under $K = 4$ light sources. Even though inter-reflections would affect the effectiveness of the recovery, we can observe that the method described in Sec.5 achieves higher accuracy and enables dense and detailed surface normals and depth recovery (d, e) of objects with diffuse plus specular surfaces.

Fig. 6 plots estimation errors at different bit depths (*i.e.*, the quantization bits of the intensity stored in an image) of the synthetic images. These results show that the recovery accuracy is enhanced pronouncedly by increasing the image depth.

Fig. 7 plots estimation errors at different noise levels. The value σ is the standard deviation of zero-mean Gaussian noise injected to the input intensities ranging in [0 : 1]. The blue to purple curves show results when using 3 to 8 auxiliary light sources, respectively. The brown curve shows results by bispectral shape from water [1], using two light sources of absorption coefficients 5×10^{-3} and 3×10^{-2} coaligned with the camera. These results show that the errors increase linearly with more noise, but that increasing the number of light sources damps the effect.

6.2 Static Object Reconstruction

We implemented the method for off-line surface normal and shape recovery of static underwater objects using off-the-shelf imaging components. The imaging system consists of four light sources each with a Fresnel lens and a monochromatic camera (Grasshopper3 GS3-U3-41C6NIR) equipped with interchangeable bandpass filters. Since the attenuation of light in water varies rapidly in the near-infrared range, we choose the wavelengths within the set that can be purchased as off-the-shelf near-infrared filters to achieve higher reconstruction accuracy. In practice, we use four distinct near-infrared wavelengths of 880nm, 905nm, 925nm, and 950nm. Each light source is placed in different angles w.r.t. the viewing direction, while satisfying the convex cone requirement. We opt for this imaging configuration of four different wavelengths and directions of light sources to make the difference in effective absorption coefficients as large as possible empirically.

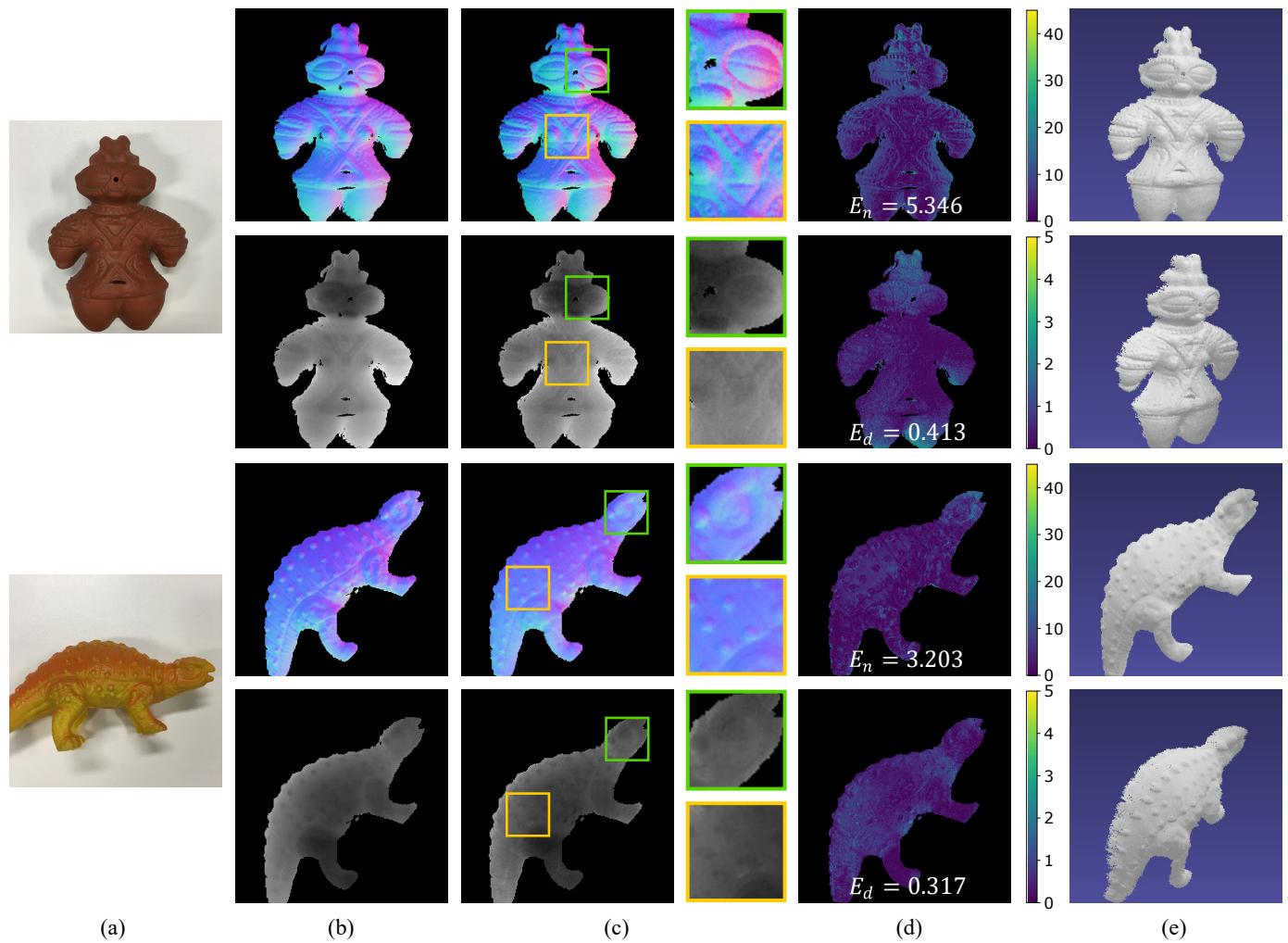


Fig. 9: Quantitative evaluation with static Lambertian objects captured with our multi-wavelength near-infrared imaging system using a regular near-infrared camera with interchangeable band-pass filters. (a) Room-light appearance. (b)(c) Ground truth and recovered surface normals and depth, and (d) error maps obtained after aligning 3D reconstruction results with the ground truth. E_d is the mean absolute error in millimeter and E_n is the mean angular error in degree. (e) Recovered 3D surface as oriented point clouds from two different viewpoints. The results show that our method successfully recovers both surface normals and depth at each pixel and retain geometric details as evident in the zoomed-in insets.

The absorption coefficients, four light source directions, and their intensities need not be known beforehand, which significantly increases flexibility in the imaging setup. Absorption coefficients can be calculated from the Beer-Lambert law by capturing a flat white target in water at a known depth. Both light source directions and intensities are estimated with the calibration method in Sec. 4.5. We eliminate the environment illumination by taking images under ambient light, and subtracting them from the captured images. E_n and E_d before and after calibration were 27.039 and 0.183, and 7.728 and 0.002, respectively. E_n and E_d of another sphere not used in calibration put at a different position were 7.85 and 0.002, which validates the accuracy of calibration.

We quantitatively evaluate the reconstruction accuracy of our method on both Lambertian and non-Lambertian objects with complex surface geometry including sharp bumps and creases as well as discontinuities. The ground truth depth and normals are computed from structured light stereo reconstruction and photometric stereo, respectively, using two calibrated cameras and a projector. We coated non-Lambertian objects with white matte

spray before obtaining the ground truth. Fig. 9 and Fig. 10 show our method successfully estimates surface normals and depth of real objects with varying color and texture as can be observed in the error maps computed after aligning them with the ground truth. Note that the object surface is recovered only in the surface region where the lights are sufficiently irradiated in all images. The mean absolute errors E_d of estimated depth for each static Lambertian object are 0.413mm and 0.317mm, respectively. The mean angular errors E_n are 5.346 degrees and 3.203 degrees, respectively. The mean absolute errors E_d of estimated depth for each static non-Lambertian object are 0.373mm and 0.325mm, respectively. The mean angular errors E_n are 6.438 degrees and 5.182 degrees, respectively. Compared with the object sizes (about 8cm for Lambertian objects and about 6cm for non-Lambertian objects), the relative errors are about 0.5% and 0.6%, respectively. These results quantitatively validate the accuracy of our method for reconstructing accurate per-pixel depth and surface normals of objects with arbitrary material that preserve surface details irrespective of surface discontinuities or abrupt changes in normal

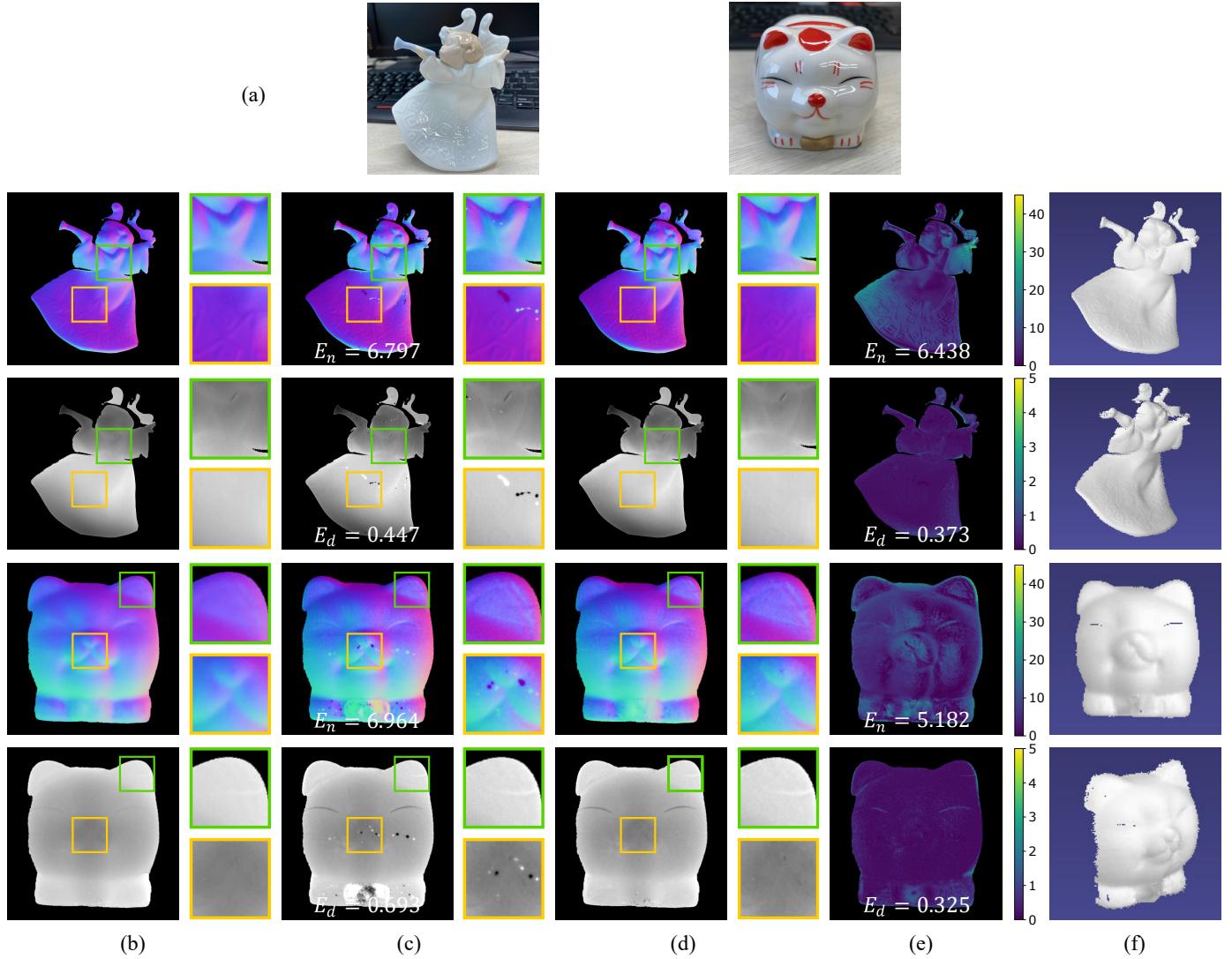


Fig. 10: Quantitative evaluation with static non-Lambertian objects. (a) Room-light appearance. (b) Ground truth and (c)(d) Recovered surface normals and depth from methods described in Sec. 4 and Sec. 5. (e) Error maps obtained after aligning 3D reconstruction results (d) with the ground truth. E_d is the mean absolute error in millimeter and E_n is the mean angular error in degree. (f) Recovered 3D surface as oriented point clouds from two different viewpoints. The results show that our method can accurately reconstruct per-pixel surface normals and depth even for objects with non-Lambertian reflectance.

orientations.

6.3 Dynamic Object Reconstruction

We also implement our method as a video-rate 3D sensing system. For this, as shown in Fig. 8, we replace the camera with a custom-built 10-bit multi-wavelength camera by EBA Japan that relies on time slicing using a spinning filter wheel and can capture the scene in six different wavelengths each at 14fps. We used off-the-shelf near-infrared band-pass filters of 852, 880, 905, and 950nm. We used two additional filters in the green and blue wavelength ranges, which combined with the 852nm provide regular RGB color information of the scene. To capture this color information, we added a regular light source (as shown in Fig. 8). Note that this additional light source does not interfere with the near-infrared light sources. The fact that we can capture texture information simultaneously with the surface normals and shape is another advantage of our method. Theoretically speaking, this implementation is not a one-shot imaging system, as the six images are captured consecutively

not at once. In practice, the time difference between the different wavelengths was small enough for the dynamic objects we used in the experiments. If necessary, we can align the observations by linearly interpolating the drift-time relationship in different wavelength captures using the estimated optical flow of RGB images. We conducted our experiments in sufficiently clear water in which scattering is negligible. In the case of real-world water (*e.g.*, river, lake, and ocean), the visibility of the captured images also degrades by the scattering induced by suspended particles. In visible light, scattering is the dominating extinction process, while in near-infrared light, absorption is dominant. If necessary, we can explicitly account for scattering in water by adopting models introduced by Tsitsios *et al.* [34], Sun *et al.* [35], and Murez *et al.* [21] similar to how Kuo *et al.* have demonstrated [16]. Fig. 11 and Fig. 12 show several frames from the video-rate normal and shape recovery of dynamic objects in water.

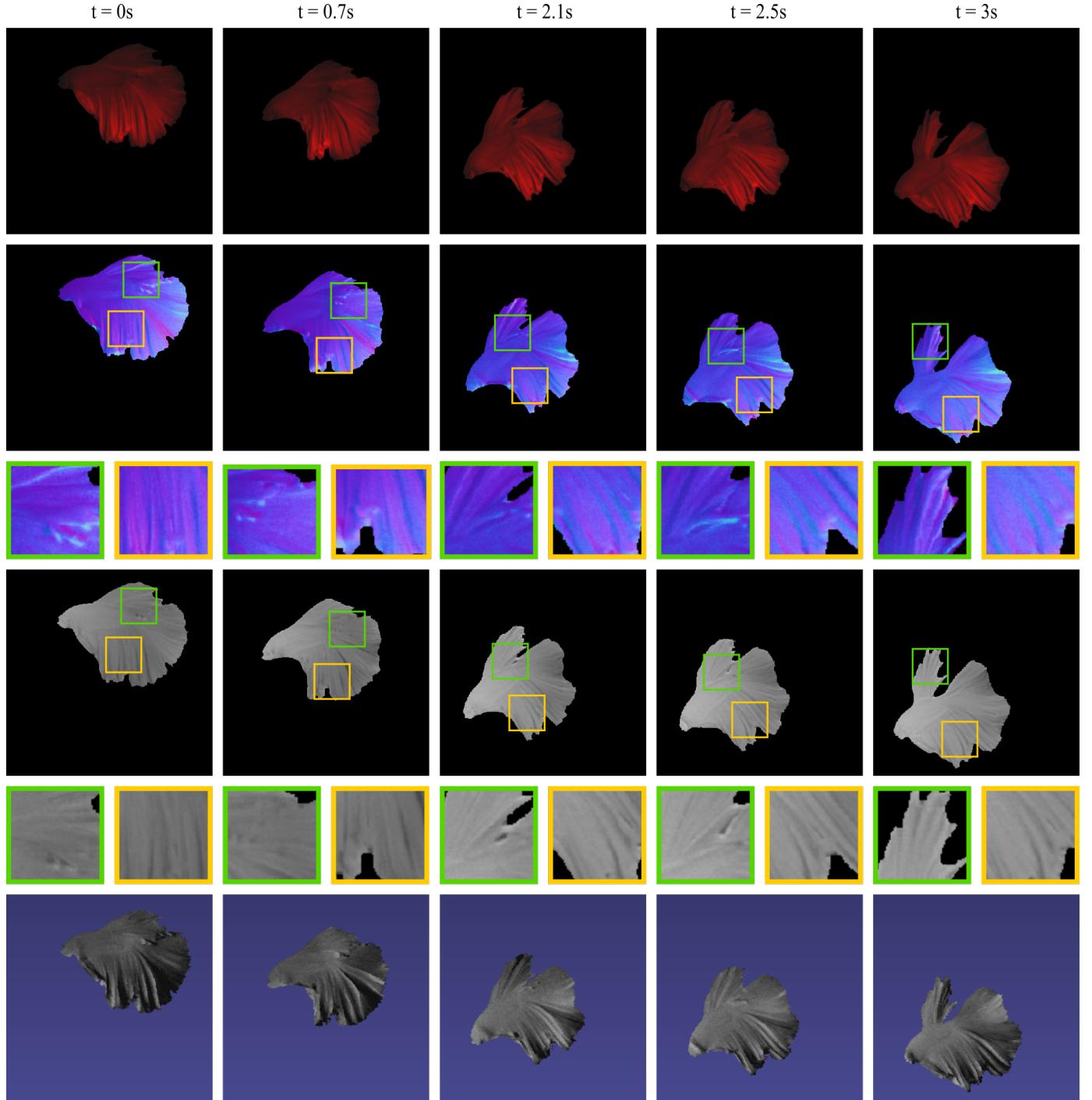


Fig. 11: Results of video-rate reconstruction of a swimming Siamese fighting fish using an implementation of our method with a custom-built multi-wavelength camera. From top to bottom, each row shows the texture, surface normal, depth, and shaded oriented points.

7 CONCLUSION

In this paper, we introduced a novel method for simultaneous recovery of surface normals and depth of objects in water. We derive the near-infrared multi-wavelength imaging principle based on the idea of leveraging light absorption along different underwater light paths associated with surface shading. Experimental results show that our method can reconstruct accurate pixel-wise depth and surface normal of complex dynamic surfaces with challenging geometric features. We believe surface normals and shape from

water would be a viable option for 3D sensing, especially as it can directly measure dynamic 3D surfaces as textured oriented points in real-time.

ACKNOWLEDGMENTS

This work was in part supported by JSPS KAKENHI 15H05918, 17K20143, 18K19815, 26240023, and 20H05951, and JST grant number JPMJCR20G7.

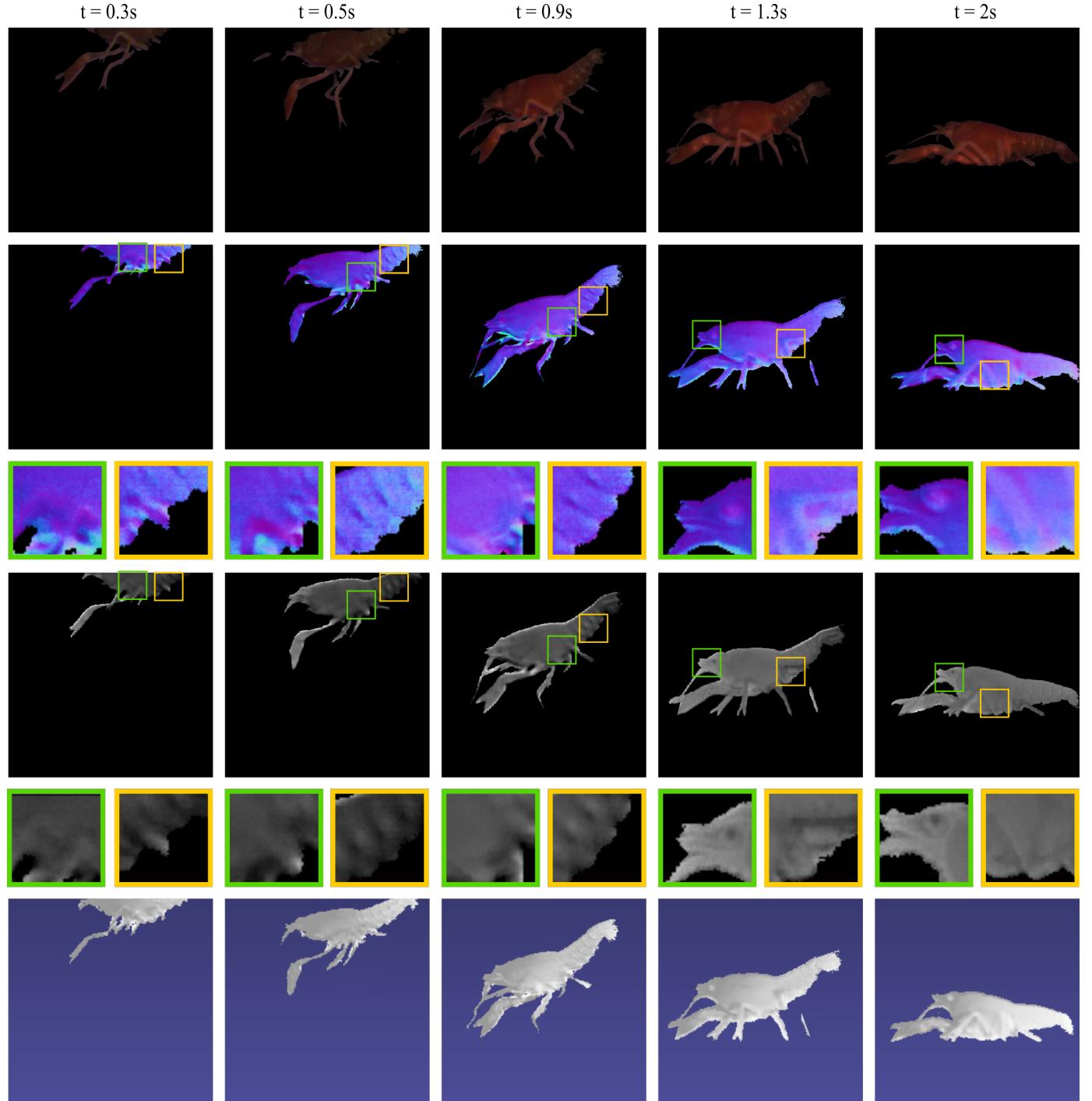


Fig. 12: Results of video-rate reconstruction of a swimming crayfish. From top to bottom, each row shows the texture, surface normal, depth, and shaded oriented points.

REFERENCES

- [1] Y. Asano, Y. Zheng, K. Nishino, and I. Sato, "Shape from water: Bispectral light absorption for depth recovery," in *ECCV*, 2016, pp. 635–649.
- [2] T. E. Zickler, P. N. Belhumeur, and D. J. Kriegman, "Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction," *IJCV*, vol. 49, no. 2-3, pp. 215–227, Sep. 2002.
- [3] N. Roubtsova and J.-Y. Guillemaut, "Colour helmholtz stereopsis for reconstruction of complex dynamic scenes," in *Proc. 3DV*, vol. 1, 2014, pp. 251–258.
- [4] M. Zhou, Y. Ding, Y. Ji, S. S. Young, J. Yu, and J. Ye, "Shape and reflectance reconstruction using concentric multi-spectral light field," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 7, pp. 1594–1605, 2020.
- [5] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multi-view stereopsis," in *Proc. CVPR*, 2007, pp. 1–8.
- [6] M. Bleyer, C. Rhemann, and C. Rother, "Patchmatch stereo - stereo matching with slanted support windows," in *Proc. BMVC*, 2011.
- [7] A. Kadambi, V. Taamazyan, B. Shi, and R. Raskar, "Polarized 3d: High-quality depth sensing with polarization cues," in *Proc. ICCV*, 2015, pp. 3370–3378.
- [8] C. Wu, M. Zollhöfer, M. Nießner, M. Stamminger, S. Izadi, and C. Theobalt, "Real-time shading-based refinement for consumer depth cameras," *ACM Trans. Graph.*, vol. 33, no. 6, pp. 200:1–200:10, Nov. 2014.

- [9] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, and S. Lin, "Shading-based shape refinement of rgb-d images," in *Proc. CVPR*, 2013, pp. 1415–1422.
- [10] Q. Zhang, M. Ye, R. Yang, Y. Matsushita, B. Wilburn, and H. Yu, "Edge-preserving photometric stereo via depth fusion," in *Proc. CVPR*, 2012, pp. 2472–2479.
- [11] S. M. Haque, A. Chatterjee, and V. M. Govindu, "High quality photometric reconstruction using a depth camera," in *Proc. CVPR*, 2014, pp. 2283–2290.
- [12] D. G. Aliaga and Y. Xu, "Photogeometric structured light: A self-calibrating and multi-viewpoint framework for accurate 3d modeling," in *Proc. CVPR*, 2008, pp. 1–8.
- [13] J. Lim, J. Ho, M.-H. Yang, and D. Kriegman, "Passive photometric stereo from motion," in *Proc. ICCV*, vol. 2, 2005, pp. 1635–1642.
- [14] T. Higo, Y. Matsushita, N. Joshi, and K. Ikeuchi, "A hand-held photometric stereo camera for 3-d modeling," in *Proc. ICCV*, 2009, pp. 1234–1241.
- [15] C. Hernández, G. Vogiatzis, G. J. Brostow, B. Stenger, and R. Cipolla, "Non-rigid photometric stereo with colored lights," in *Proc. ICCV*, 2007.
- [16] M.-Y. J. Kuo, R. Kawahara, S. Nobuhara, and K. Nishino, "Non-rigid shape from water," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021.
- [17] M. B. Hullin, M. Fuchs, I. Ihrke, H.-P. Seidel, and H. P. Lensch, "Fluorescent immersion range scanning," *ACM Trans. Graph.*, vol. 27, no. 3, p. 87, 2008.
- [18] B. Trifonov, D. Bradley, and W. Heidrich, "Tomographic reconstruction of transparent objects," in *Proc. of Eurographics Conf. on Rendering Techniques*, 2006, pp. 51–60.
- [19] N. J. W. Morris and K. N. Kutulakos, "Dynamic refraction stereo," *TPAMI*, vol. 33, no. 8, pp. 1518–1531, Aug 2011.
- [20] A. Jordt-Sedlazeck and R. Koch, "Refractive structure-from-motion on underwater images," in *Proc. ICCV*, 2013, pp. 57–64.
- [21] Z. Murez, T. Treibitz, R. Ramamoorthi, and D. Kriegman, "Photometric stereo in a scattering medium," in *Proc. ICCV*, 2015.
- [22] Y. Fujimura, M. Iiyama, A. Hashimoto, and M. Minoh, "Photometric stereo in participating media considering shape-dependent forward scatter," in *Proc. CVPR*, 2018.
- [23] X. Chen and Y.-H. Yang, "A closed-form solution to single underwater camera calibration using triple wavelength dispersion and its application to single camera 3d reconstruction," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4553–4561, Sep. 2017.
- [24] S. Ishihara, Y. Asano, Y. Zheng, and I. Sato, "Underwater scene recovery using wavelength-dependent refraction of light," in *2020 International Conference on 3D Vision (3DV)*. IEEE, 2020, pp. 32–40.
- [25] J. Tian, Z. Murez, T. Cui, Z. Zhang, D. Kriegman, and R. Ramamoorthi, "Depth and image restoration from light field in a scattering medium," in *Proc. ICCV*, 2017.
- [26] Y. Qian, M. Gong, and Y.-H. Yang, "Stereo-based 3d reconstruction of dynamic fluid surfaces by global optimization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1269–1278.
- [27] Y. Qian, Y. Zheng, M. Gong, and Y.-H. Yang, "Simultaneous 3d reconstruction for water surface and underwater scene," in *The European Conference on Computer Vision (ECCV)*, September 2018.
- [28] C. Tsotsios, A. J. Davison, and T.-K. Kim, "Near-lighting photometric stereo for unknown scene distance and medium attenuation," *Image and Vision Computing*, vol. 57, pp. 44 – 57, 2017.
- [29] E. Reinhard, E. A. Khan, A. O. Akyz, and G. M. Johnson, *Color Imaging: Fundamentals and Applications*. Natick, MA, USA: A. K. Peters, Ltd., 2008.
- [30] G. H. Golub and C. F. Van Loan, *Matrix Computations (3rd Ed.)*. Baltimore, MD, USA: Johns Hopkins University Press, 1996, pp. 256–258.
- [31] S. A. Shafer, "Using color to separate reflection components," *Color Research & Application*, vol. 10, no. 4, pp. 210–218, 1985.
- [32] M. Pharr, W. Jakob, and G. Humphreys, *Physically based rendering: From theory to implementation*. Morgan Kaufmann, 2016.
- [33] R. T. Frankot and R. Chellappa, "A method for enforcing integrability in shape from shading algorithms," *TPAMI*, vol. 10, no. 4, pp. 439–451, 1988.
- [34] C. Tsotsios, M. E. Angelopoulou, T. K. Kim, and A. J. Davison, "Backscatter compensated photometric stereo with 3 sources," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2259–2266, 2014.
- [35] B. Sun, R. Ramamoorthi, S. G. Narasimhan, and S. K. Nayar, "A practical analytic single scattering model for real time rendering," *ACM Transactions on Graphics (TOG)*, vol. 24, no. 3, pp. 1040–1049, 2005.



Meng-Yu Jennifer Kuo received her Ph.D. from Computer Vision Lab at Kyoto University, Japan, in 2021. She received the B.S. degree from the Department of Computer Science and Information Engineering, National Chung Cheng University, Taiwan, in 2014. She joined the Graduate School of Informatics at Kyoto University in 2015 and received the M.S. degree in 2017. Her primary research interests include: computer vision, 3D sensing, computational photography, and physics-based vision.



Satoshi Murai received his B.Sc. in Engineering and M.Sc. in Informatics from Kyoto University, Japan, in 2017 and 2019, respectively. He is currently working at SERVICE & SECURITY CORPORATION, Japan.



Ryo Kawahara is currently an assistant professor in the Department of Artificial Intelligence at Kyushu Institute of Technology. He received the B.S. degree in Electrical and Electronic Engineering in 2013, and received the M.S. and the Ph.D. degree in Computer Vision from Kyoto University in 2015 and 2019, respectively. His research interests are in computer vision, computational photography, and physics-based vision.



Shohei Nobuhara received his B.Sc. in Engineering, M.Sc. and Ph.D. in Informatics from Kyoto University, Japan, in 2000, 2002, and 2005 respectively. Since 2019, he has been an associate professor at Kyoto University. His research interest includes human shape and motion modelling, physics-based vision, and multiple-view geometry. He is a member of IEEE, IPSJ, IEICE.



Ko Nishino is a Professor in the Department of Intelligence Science and Technology at Kyoto University. He received his B.E. and M.E. in Information and Communication Engineering in 1997 and 1999, respectively, and PhD in Computer Science in 2002, all from University of Tokyo. Before joining Kyoto University in 2018, he was a Professor in the Department of Computer Science at Drexel University. His primary research interests lie in computer vision and machine learning including appearance modeling and material recognition, human behavior analysis, and computational photography. He received the NSF CAREER award in 2008.