Surface Normals and Shape From Water

Satoshi Murai* Meng-Yu Jennifer Kuo*
Ryo Kawahara Shohei Nobuhara Ko Nishino
Kyoto University, Kyoto, Japan
http://vision.ist.i.kyoto-u.ac.jp/

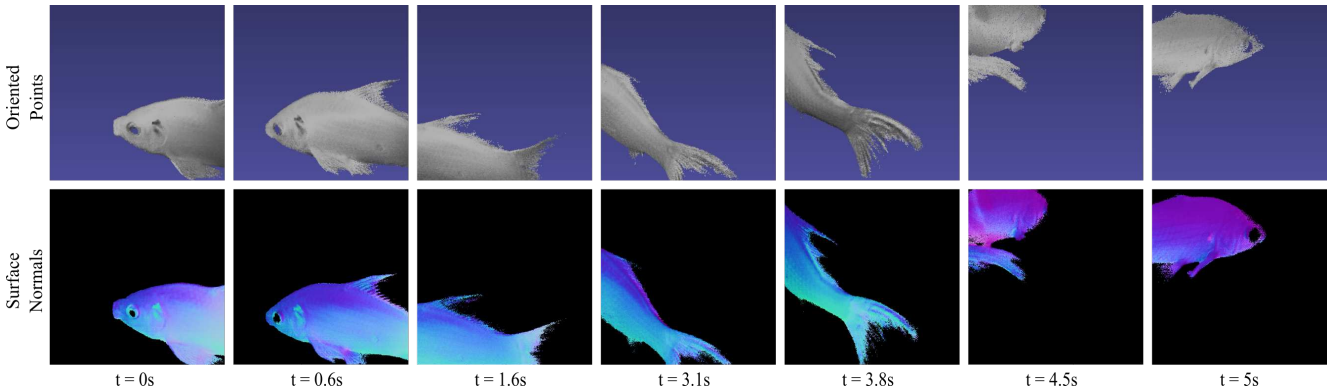


Figure 1: Our method enables video-rate, per-pixel surface normal and depth recovery of dynamic underwater objects without any artificial constraints on one another such as smoothness. The recovered oriented points retain intricate surface details including sharp geometric features that would otherwise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

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. 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.*

*Equal contribution.

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 di-

rectly 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*. [2]. 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. We also 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.

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.

Surface Normals and Shape Reconstruction in Air A representative early work that achieves simultaneous pixel-wise surface normal and shape estimation is Helmholtz stereo [25, 19], which can also handle non-Lambertian surfaces. The method, however, requires interchanged viewing and lighting directions, which precludes single-shot estimation by definition.

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 [7] and PatchMatch stereo [3] 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*. [13] proposed a polarization-based depth enhancement method, and Wu *et al*. [22] and Yu *et al*. [23] 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 [24, 9], structured light [1], and structure-from-motion [14, 11] 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*. [10] 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.

3D Reconstruction in Water Asano *et al*. [2] 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. Most other studies on 3D reconstruction in water such as underwater stereo [15], refractive structure-from-motion [12], underwater photomet-

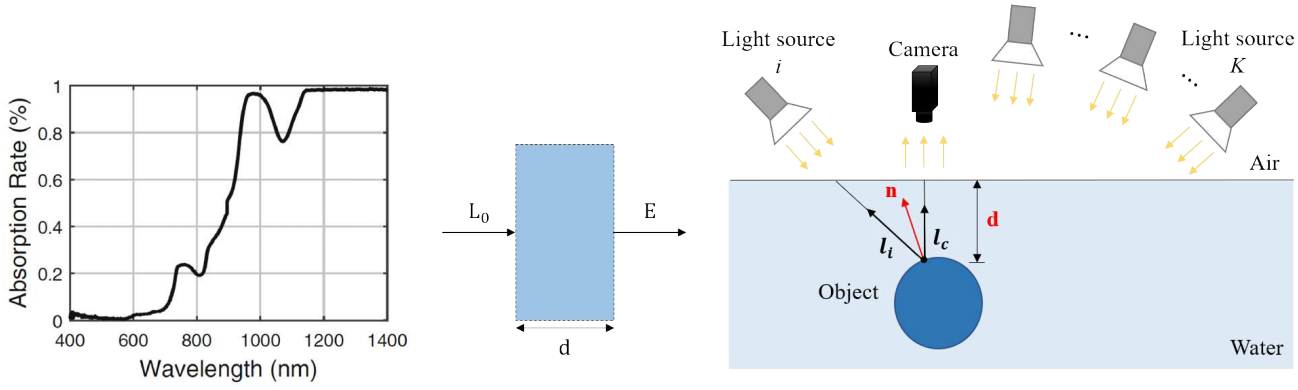


Figure 2: We leverage (a) wavelength-dependent near-infrared light absorption by water (reprinted from [2]) which can be modeled with (b) the Beer-Lambert law that relates light attenuation with the absorption coefficient and light path length. We derive the theory for and implement (c) a multi-wavelength 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.

ric stereo [16, 6], shape from chromatic dispersion [4], and underwater depth from light-field imaging [20] assume absorption by water plays little effect and focus on applying in-air conventional 3D reconstruction methods to underwater imagery. Tsiotsios *et al*. [21] 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 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 almost linearly in the near-infrared range. As depicted in Fig. 2(b), light absorption in water can be accurately modeled with the Beer-Lambert law [18] that expresses transmitted radiance E as that of incident light L_0 which exponentially decays as a function of the distance d it travels

$$E = L_0 \exp\left[-\alpha(\lambda)d\right], \tag{1}$$

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

Asano *et al*. [2] 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 $\boldsymbol{l}_c$ and $\boldsymbol{l}_i$ 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 $\frac{d(x)}{\boldsymbol{l}_c^\top \boldsymbol{l}_i}$ 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

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

where we have defined the *effective absorption coefficient*

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

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*. [2], 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

$$s(\lambda_i) = \rho\,. \tag{4}$$

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

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

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

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

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

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

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

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

Given observations taken under K distinct light sources, we obtain

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

where we have defined the depth vector $\boldsymbol{d}(x) \in \mathbb{R}^{(K-1)\times 1}$ and the light source matrix $\mathbf{L} = \begin{bmatrix} \boldsymbol{l}_2 & \cdots & \boldsymbol{l}_K \end{bmatrix}^\top \in \mathbb{R}^{(K-1)\times 3}$. We refer to $\boldsymbol{l}_1$ as the base light source and all the other light sources (*i.e*., those in the light matrix $\mathbf{L}$) as auxiliary light sources.

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 $\boldsymbol{n}(x)$. Although it does not have a closed-form solution, we can recover a unique depth value $d(x)$ and surface normal $\boldsymbol{n}(x)$ 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:

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

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

$$\frac{\boldsymbol{l}_1^\top \boldsymbol{n}(x)}{\boldsymbol{l}_1^\top \boldsymbol{n}(x)} = \boldsymbol{l}_1^\top \mathbf{L}^+ \boldsymbol{d}(x)\,, \tag{11}$$

$$1 = \boldsymbol{b}\boldsymbol{d}(x)\,. \tag{12}$$

where $\boldsymbol{b} = \boldsymbol{l}_1^\top \mathbf{L}^+ \ (\in \mathbb{R}^{1\times(K-1)})$.

We derive conditions on Eq. 12 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 such as Newton-Raphson, efficiently. Moreover, the optimization can utilize the first and second-order derivatives of $\boldsymbol{b}\boldsymbol{d}(x)$ w.r.t. $d(x)$, thanks to its simple sum-of-exponents form.

Given the depth $d(x)$ from Eq. 12, we can compute the surface normal scaled by $\boldsymbol{l}_1^\top \boldsymbol{n}(x)$ using the right-hand side of Eq. 10. The surface normal $\boldsymbol{n}(x)$ is then given by

$$\boldsymbol{n}(x) = \frac{\mathbf{L}^+ \boldsymbol{d}(x)}{\|\mathbf{L}^+ \boldsymbol{d}(x)\|}\,. \tag{13}$$

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. 13, 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. 12, we observe that the effective absorption coefficients of the auxiliary light sources should all differ from that of the base light source

$$\begin{bmatrix} \hat{\alpha}_2 - \hat{\alpha}_1 & \cdots & \hat{\alpha}_K - \hat{\alpha}_1 \end{bmatrix} \neq \mathbf{0}\,. \tag{14}$$

Recall that the effective absorption coefficient (Eq. 3) is a function of both the absorption coefficient α_i and shading $\boldsymbol{l}_c^\top \boldsymbol{l}_i$. 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. 14.

The right hand side of Eq. 12 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 $\boldsymbol{b}$) and $\hat{\alpha}_i - \hat{\alpha}_1$ match. In reality, the light sources cannot be located in the water and they should be all on the same hemisphere, which makes the elements of $\boldsymbol{b}$ 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 $i = 2, \ldots, K$, 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 $\boldsymbol{b}$ are non-negative.

By the definition of pseudo inverse matrix [8], $\boldsymbol{b} = \boldsymbol{l}_1^\top \mathbf{L}^+$ 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, 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 $\boldsymbol{b}$ 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 $\boldsymbol{b} \geq 0$ holds is not trivial. In practice, however, since $\hat{\alpha}_i$ and $\boldsymbol{b}$ 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 $\boldsymbol{b}$. That is, once the calibrated light source directions satisfy the non-negative constraint, the depth estimation by function Eq. 12 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 $\boldsymbol{b}$ 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 $\boldsymbol{l}_i = \boldsymbol{l}_1$, the right hand side of Eq. 9 becomes an all-one vector and the depth estimate becomes

$$d(x) = -\frac{1}{\hat{\alpha}_k - \hat{\alpha}_1} \ln\left(\frac{E_k}{E_1}\right), \tag{15}$$

for every k. This is equivalent to the depth estimate from bispectral shape from water [2]. Eq. 15 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 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

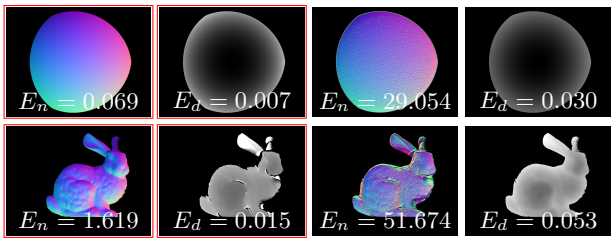
$$\begin{aligned} \Delta d &= -\frac{1}{\hat{\alpha}_k - \hat{\alpha}_1}\left(\ln\frac{E_k + \epsilon}{E_1 + \epsilon} - \ln\frac{E_k}{E_1}\right) \\ &= -\frac{1}{\hat{\alpha}_k - \hat{\alpha}_1}\ln\left(\frac{1 + \frac{\epsilon}{E_k}}{1 + \frac{\epsilon}{E_1}}\right). \end{aligned} \tag{16}$$

Eq. 16 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 are maximized, while the observed radiance due to each corresponding light source is also made as large as possible.

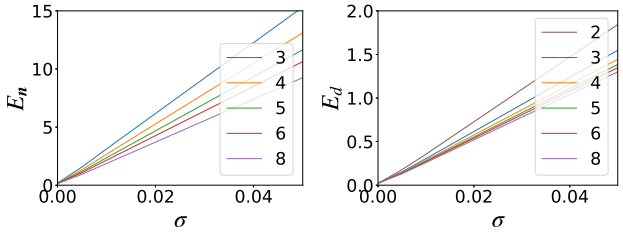
4.5. Calibration

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. We derive a



(a) Normal (b) Depth (c) Normal from depth (d) Depth from normal

Figure 3: Evaluations with synthetic data. (a)(b) Normal and depth maps estimated by the proposed method. (c)(d) Normal map computed from the estimated depth map and vice versa. Our method achieves higher accuracy for both.



(a) Normal error (b) Depth error

Figure 4: Estimation errors at different noise levels. Colors denote different numbers of auxiliary light sources.

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 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 $\hat{\boldsymbol{n}}(x)$ at any given point x on the sphere. The light source direction vectors $\boldsymbol{l} = \{\boldsymbol{l}_1 \cdots \boldsymbol{l}_K\}$ and their intensities $L = \{L_1 \cdots L_K\}$ can then be estimated by minimizing the $L2$ error

$$\arg\min_{\boldsymbol{l},L} \sum_{i=1}^{n} \sum_{j=1}^{m} \Big(k_1 \|\hat{d}(x_j^i) - d(\boldsymbol{l}, L, x_j^i)\| + k_2 \|1 - \hat{\boldsymbol{n}}(x_j^i) \cdot \boldsymbol{n}(\boldsymbol{l}, L, x_j^i)\| \Big) , \tag{17}$$

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 $\boldsymbol{n}$ are the estimated depth and normals, respectively.

5. Experimental Results

We experimentally validate our method using synthetic and near-infrared real light sources in the range of 900nm to 1000nm. The absorption coefficient in this wavelength range varies dramatically from roughly 5×10^{-3} to 3×10^{-2}.

5.1. Quantitative Evaluation with Synthetic Data

We quantitatively evaluate the reconstruction accuracy and noise robustness of our method using 8-bit synthetic images rendered with PBRT [17]. 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. 3 shows results for the minimal imaging configuration of $K = 4$. In addition to the recovered surface normals and depth by our method (a,b), Fig. 3 shows surface normals computed from the estimated depth by numerical differentiation (c), and depth integrated from estimated surface normals[5] (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. Conventional methods inevitably rely on one estimate to obtain the other, which results in significant reduction in accuracy.

Fig. 4 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 [2], 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 dampens the effect.

5.2. 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 near-infrared bandpass filters. 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 as it gives maximum difference in effective absorption coefficients.

The absorption coefficients, four light source directions, and their intensities need not be known beforehand, which significantly increases flexibility in the imaging setup. Ab-

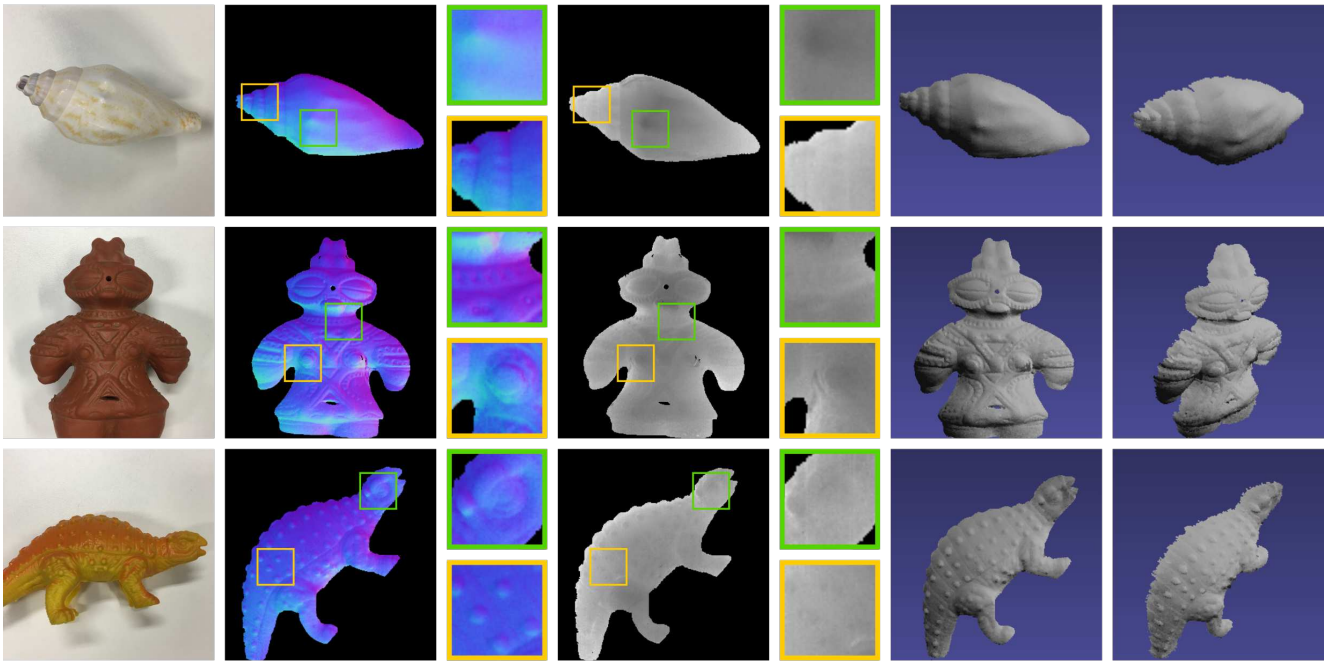


Figure 5: Reconstruction results of static objects captured with our multi-wavelength near-infrared imaging system using a regular near-infrared camera with interchangeable band-pass filters. For each object, from left to right, we show the room-light appearance, recovered surface normals, depth, and 3D surface as oriented point clouds from two different viewpoints. The results show that our method successfully simultaneously recovers both surface normals and depth at each pixel and retain geometric details as shown in the zoomed-in insets.

sorption 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. 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.

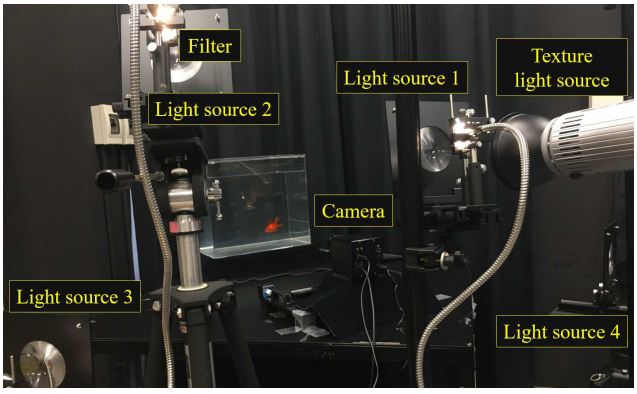


Figure 6: We implement a video-rate surface normal and shape from water imaging system using four light sources each placed with a Fresnel lens and a near-infrared band-pass filter and a custom-built multi-wavelength camera. A visible spectrum light source is used to also capture texture.

We apply our method to real objects with complex surface geometry including sharp bumps and creases as well as discontinuities. Fig. 5 shows the estimated surface normals and depth of real objects with varying color and texture. We can observe that our method is capable of reconstructing accurate per-pixel depth and surface normals that preserve surface details irrespective of surface discontinuities or abrupt changes in normal orientations.

6. Dynamic Object Reconstruction

We also implement our method as a video-rate 3D sensing system. For this, as shown in Fig. 6, we replace the camera with a custom-built 10-bit multi-wavelength camera by EBA Japan that can capture the scene in six different wavelengths each at 14fps. We used 852, 880, 905, 950nm for the near-infrared band-pass filters. 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. 6). 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. Theoreti-

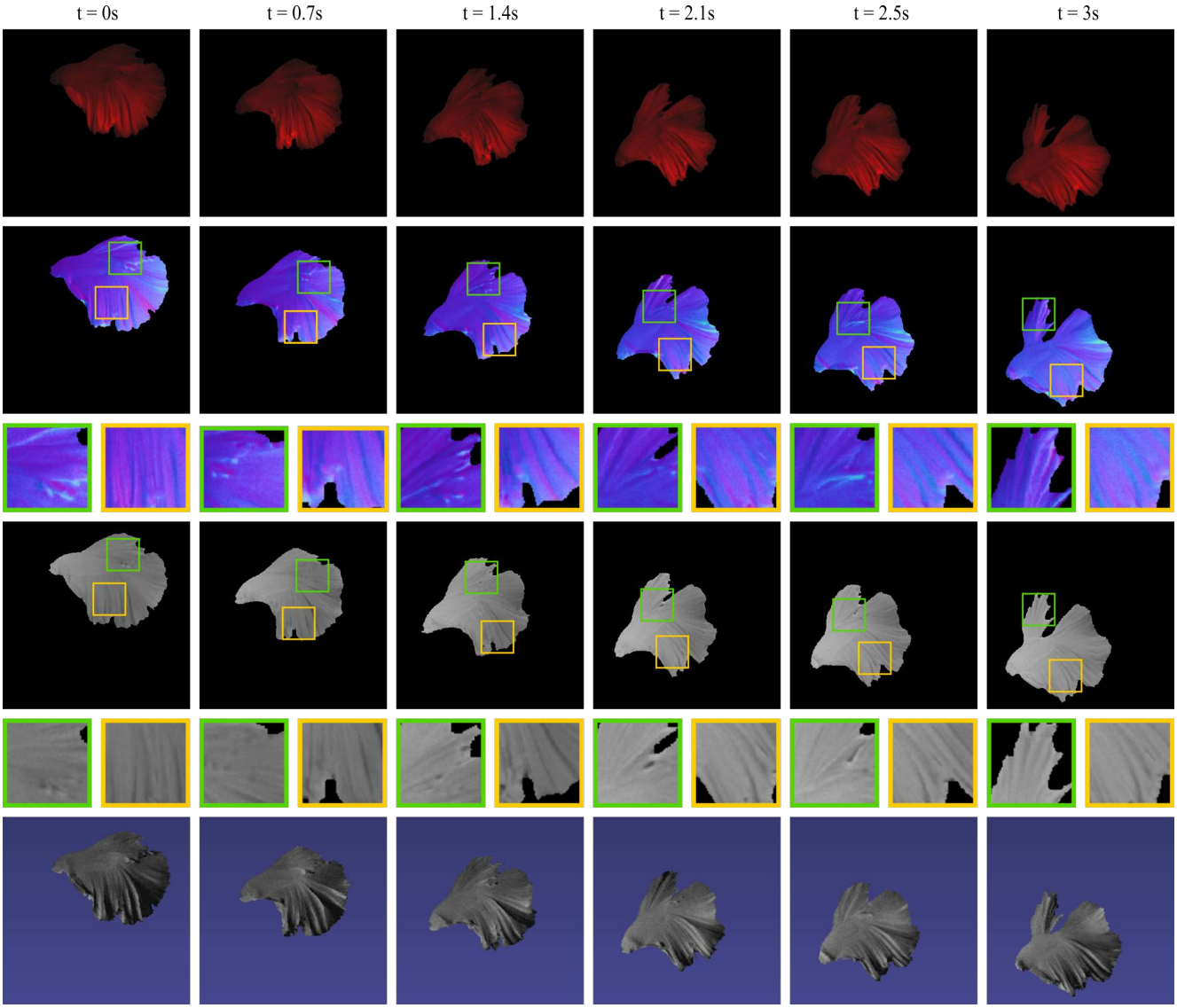


Figure 7: Results of video-rate reconstruction of a swimming Siamese fighting fish using an implementation of our method with a custom-built multi-wavelength camera. From top to bottom, each row shows the texture, surface normal, depth, and shaded oriented points.

cally 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 using estimated optical flow. Fig. 7 shows several frames from the video-rate normal and shape recovery of a swimming fish.

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.

Acknowledgement This work was in part supported by JSPS KAKENHI 15H05918, 17K20143, 18K19815, and 26240023.

References

[1] Daniel. G. Aliaga and Yi Xu. Photogeometric structured light: A self-calibrating and multi-viewpoint framework for accurate 3d modeling. In *Proc. CVPR*, pages 1–8, 2008. 2

[2] Yuta Asano, Yinqiang Zheng, Ko Nishino, and Imari Sato. Shape from water: Bispectral light absorption for depth recovery. In *ECCV*, pages 635–649, 2016. 2, 3, 4, 5, 6

[3] Michael Bleyer, Christoph Rhemann, and Carsten Rother. Patchmatch stereo - stereo matching with slanted support windows. In *Proc. BMVC*, 2011. 2

[4] Xida Chen and Yee-Hong 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*, 26(9):4553–4561, Sep. 2017. 3

[5] Robert T. Frankot and Rama Chellappa. A method for enforcing integrability in shape from shading algorithms. *TPAMI*, 10(4):439–451, 1988. 6

[6] Yuki Fujimura, Masaaki Iiyama, Atsushi Hashimoto, and Michihiko Minoh. Photometric stereo in participating media considering shape-dependent forward scatter. In *Proc. CVPR*, 2018. 3

[7] Yasutaka Furukawa and Jean Ponce. Accurate, dense, and robust multi-view stereopsis. In *Proc. CVPR*, pages 1–8, 2007. 2

[8] Gene H. Golub and Charles F. Van Loan. *Matrix Computations (3rd Ed.)*, pages 256–258. Johns Hopkins University Press, Baltimore, MD, USA, 1996. 5

[9] Sk Mohammadul Haque, Avishek Chatterjee, and Venu Madhav Govindu. High quality photometric reconstruction using a depth camera. In *Proc. CVPR*, pages 2283–2290, 2014. 2

[10] Carlos Hernández, George Vogiatzis, Gabriel J. Brostow, Bjorn Stenger, and Roberto Cipolla. Non-rigid photometric stereo with colored lights. In *Proc. ICCV*, 2007. 2

[11] Tomoaki Higo, Yasuyuki Matsushita, Neel Joshi, and Katsushi Ikeuchi. A hand-held photometric stereo camera for 3-d modeling. In *Proc. ICCV*, pages 1234–1241, 2009. 2

[12] Anne Jordt-Sedlazeck and Reinhard Koch. Refractive structure-from-motion on underwater images. In *Proc. ICCV*, pages 57–64, 2013. 2

[13] Achuta Kadambi, Vage Taamazyan, Boxin Shi, and Ramesh Raskar. Polarized 3d: High-quality depth sensing with polarization cues. In *Proc. ICCV*, pages 3370–3378, 2015. 2

[14] Jongwoo Lim, Jeffrey Ho, Ming-Hsuan Yang, and David Kriegman. Passive photometric stereo from motion. In *Proc. ICCV*, volume 2, pages 1635–1642, 2005. 2

[15] Nigel J. W. Morris and Kiriakos N. Kutulakos. Dynamic refraction stereo. *TPAMI*, 33(8):1518–1531, Aug 2011. 2

[16] Zak Murez, Tali Treibitz, Ravi Ramamoorthi, and David Kriegman. Photometric stereo in a scattering medium. In *Proc. ICCV*, 2015. 3

[17] Matt Pharr, Wenzel Jakob, and Greg Humphreys. *Physically based rendering: From theory to implementation*. Morgan Kaufmann, 2016. 6

[18] Erik Reinhard, Erum Arif Khan, Ahmet Oguz Akyz, and Garrett M. Johnson. *Color Imaging: Fundamentals and Applications*. A. K. Peters, Ltd., Natick, MA, USA, 2008. 3

[19] Nadejda Roubtsova and Jean-Yves Guillemaut. Colour helmholtz stereopsis for reconstruction of complex dynamic scenes. In *Proc. 3DV*, volume 1, pages 251–258, 2014. 2

[20] Jiandong Tian, Zachary Murez, Tong Cui, Zhen Zhang, David Kriegman, and Ravi Ramamoorthi. Depth and image restoration from light field in a scattering medium. In *Proc. ICCV*, 2017. 3

[21] Chourmouzios Tsiotsios, Andrew J. Davison, and Tae-Kyun Kim. Near-lighting photometric stereo for unknown scene distance and medium attenuation. *Image and Vision Computing*, 57:44 – 57, 2017. 3

[22] Chenglei Wu, Michael Zollhöfer, Matthias Nießner, Marc Stamminger, Shahram Izadi, and Christian Theobalt. Real-time shading-based refinement for consumer depth cameras. *ACM Trans. Graph.*, 33(6):200:1–200:10, Nov. 2014. 2

[23] Lap-Fai Yu, Sai-Kit Yeung, Yu-Wing Tai, and Stephen Lin. Shading-based shape refinement of rgb-d images. In *Proc. CVPR*, pages 1415–1422, 2013. 2

[24] Qing Zhang, Mao Ye, Ruigang Yang, Yasuyuki Matsushita, Bennett Wilburn, and Huimin Yu. Edge-preserving photometric stereo via depth fusion. In *Proc. CVPR*, pages 2472–2479, 2012. 2

[25] Todd E. Zickler, Peter N. Belhumeur, and David J. Kriegman. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *IJCV*, 49(2-3):215–227, Sept. 2002. 2