Sea-thru: A Method For Removing Water From Underwater Images

Derya Akkaynak   Tali Treibitz

University of Haifa

derya.akkaynak@gmail.com, ttreibitz@univ.haifa.ac.il

Abstract

Robust recovery of lost colors in underwater images remains a challenging problem. We recently showed that this was partly due to the prevalent use of an atmospheric image formation model for underwater images and proposed a physically accurate model. The revised model showed: 1) the attenuation coefficient of the signal is not uniform across the scene but depends on object range and reflectance, 2) the coefficient governing the increase in backscatter with distance differs from the signal attenuation coefficient. Here, we present the first method that recovers color with our revised model, using RGBD images. The Sea-thru *method estimates backscatter using the dark pixels and their known range information. Then, it uses an estimate of the spatially varying illuminant to obtain the range-dependent attenuation coefficient. Using more than 1,100 images from two optically different water bodies, which we make available, we show that our method with the revised model outperforms those using the atmospheric model. Consistent removal of water will open up large underwater datasets to powerful computer vision and machine learning algorithms, creating exciting opportunities for the future of underwater exploration and conservation.*

1. Introduction

Reconstructing colors in underwater images is a challenging task for which no robust algorithm currently exists. We recently showed that the commonly used image formation model was partly to blame [1], because it was derived for the atmosphere [48] and neglected the strong wavelength dependency of light underwater. We proposed a revised model that showed: 1) direct and backscattered signals are governed by distinct coefficients (the old model assumed them to be the same), and 2) each of these coefficients have dependencies on factors other than the optical properties of the water (the old model ignored them). While the revised model is physically more accurate, it has extra parameters making its application difficult. Here, we present the *Sea-thru* method that outlines how to estimate these parameters for better scene recovery.
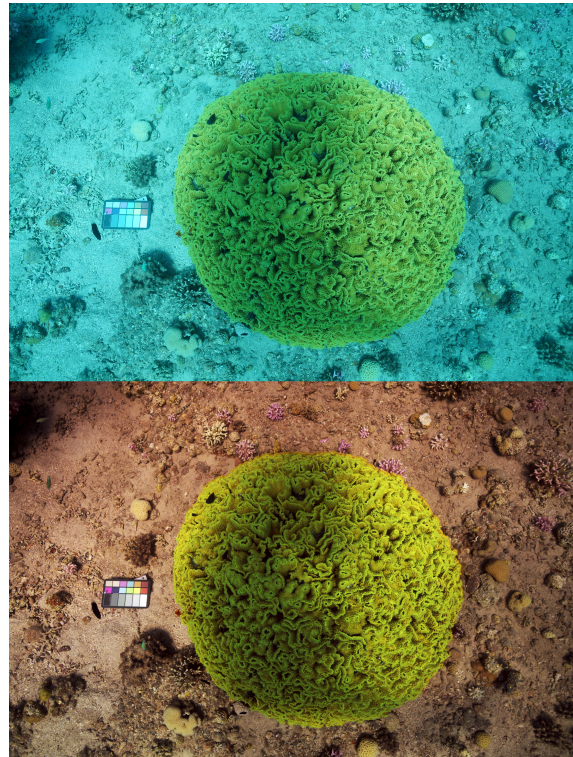

Figure 1. The *Sea-thru* method removes water from underwater images. Best viewed online for color and details.

Large image datasets like ImageNet [20] have been instrumental in igniting the artificial intelligence boom, which fueled many important discoveries in science and industry in the last two decades [39]. The underwater domain, which has no shortage of large image datasets, however, has not benefited from the full power of computer vision and machine learning methods which made these discoveries possible, partly because water masks many computationally valuable features of a scene. An underwater photo is the equivalent of one taken in air, but covered in thick, colored fog, subject to an illuminant whose white point and intensity changes as a function of distance. It is difficult to train learning-based methods for different optical conditions that represent the global ocean, because calibrated underwater datasets are expensive and logistically difficult to
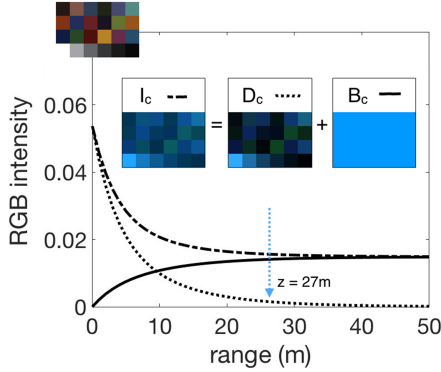
Figure 2. Underwater image formation is governed by an equation of the form $I_c = D_c + B_c$. D_c contains the scene with attenuated colors, and B_c is a degrading signal that strongly depends on the optical properties of the water, and eventually dominates the image (shown here for a gray patch). Insets show relative magnitudes of D_c and B_c for a Macbeth chart imaged at 27m in oceanic water.

acquire. Existing methods that attempt to reverse the degradation due to water are either unstable, too sensitive, or only work for short object ranges. Thus, the analysis of large underwater datasets often requires costly manual effort. On average, a human expert spends over 2 hours identifying and counting fish in a video that is one hour long [59].

The **Sea-thru** method aims to consistently remove water from underwater images, so large datasets can be analyzed with increased efficiency. It works as follows: given an RGBD image, it estimates backscatter in a way inspired from the Dark Channel Prior (DCP) developed for haze [34], but utilizing the known range map. Next, it uses an optimization framework to estimate the range-dependent attenuation coefficient using an illumination map obtained using local space average color [23] as input. We show that the distance-dependent attenuation coefficient can be modeled as a 2-term exponential, which greatly reduces the unknowns in the optimization step. We contribute more than 1,100 images acquired in two optically different water types (available at http://csms.haifa.ac.il/profiles/tTreibitz/datasets/sea_thru/index.html). On these images and another underwater RGBD dataset contributed by [7], we show qualitatively and quantitatively that **Sea-thru**, which is the first to utilize the revised image formation model, outperforms others that use the old model.

2. Related Works

The image formation model for bad weather was developed by Nayar and Narasimhan [48]. It was assumed that the scattering coefficient is constant over the camera sensitivity range in each color channel, resulting in a coefficient per wavelength. This model then became extensively used for bad weather, and later adapted for the underwater envi-

ronment [52]. For scene recovery, these methods required more than one frame of the scene, or extra information such as 3D structure. This model was further simplified to include only one attenuation coefficient, uniform across all color channels. This was done to enable recovery from single images in haze [8, 26, 33, 57] and later used also for underwater recovery [17, 19, 21, 44, 49]. While using the same coefficient for all color channels in underwater scenes is a very crude approximation [1], using a coefficient per channel can yield decent results [9, 13, 52, 56, 58]. Nevertheless, as we further show their accuracy is inherently limited by the model.

Backscatter was previously estimated from single images using DCP [33], some variants of it [17, 19, 21, 44], or other priors [9, 49]. Attenuation coefficients can be measured by ocean optics instruments such as transmissiometers or spectrometers [11]. However, they cannot be used as-is for imaging because of differences in spectral sensitivity and acceptance angle. Plus, these instruments are expensive and cumbersome to deploy. Thus, it is best to estimate the attenuation coefficients directly from images. The most basic method for that is to photograph a calibration target at known distances [60]. In [63], coefficients are taken from the estimated veiling-light, ignoring the illumination color. In [9] the attenuation coefficients per channel were estimated using the grey-world assumption. Others [19, 44, 55] alleviate this problem by using fixed attenuation coefficients measured for just one water type.

Known distances slightly simplify the problem and were used to estimate backscatter together with attenuation by fitting data from multiple images to the image formation model [13, 51, 55]. Deep networks were recently used for reconstructing underwater scenes [43, 53]. Their training, however, relies on purely synthetic data, and thus highly depends on the quality of the simulation models. All the methods so far assume the attenuation coefficients are only properties of the water and are uniform across the scene per color channel, but as we showed in [1, 2], this is an incorrect assumption that leads to errors in reconstruction.

3. Scientific Background

Underwater image formation is governed by:

$$I_c = D_c + B_c \ , \tag{1}$$

where $c = $ R, G, B is the color channel, I_c is the image captured by the camera (with distorted colors), D_c is the *direct signal* which contains the information about the (attenuated) scene, and B_c is the *backscatter*, an additive signal that degrades the image due to light reflected from particles suspended in the water column (Fig. 2). The components D_c and B_c are governed by two distinct coefficients β_c^D and β_c^B, which are wideband (RGB) attenuation and backscatter coefficients, respectively [1, 2].

The expanded form of Eq. 1 is given as [1]:

$$I_c = J_c e^{-\beta_c^D(\mathbf{v}_D)\cdot z} + B_c^\infty \left(1 - e^{-\beta_c^B(\mathbf{v}_B)\cdot z}\right) \ , \quad (2)$$

where z is range (distance) between the camera and the objects in the scene along the line of sight, B_c^∞ is veiling light, and J_c is the unattenuated scene that would be captured at the location of the camera had there been no attenuation along z. Vectors $\mathbf{v}_D = \{z, \rho, E, S_c, \beta\}$ and $\mathbf{v}_B = \{E, S_c, b, \beta\}$ represent the dependencies of the coefficients β_c^D and β_c^B on range z, reflectance ρ, spectrum of ambient light E, spectral response of the camera S_c, and the physical scattering and beam attenuation coefficients of the water body, b and β, all of which are functions of wavelength λ. Previously, it was assumed that $\beta_c^D = \beta_c^B$, and that these coefficients had a single value for a given scene [9], but in [1] we have shown that they are distinct, and furthermore, that they had dependencies on different factors. Eq. 2 is formulated for imaging in the horizontal direction. However, throughout this work, we apply it to scenes captured in different directions with the assumption that the deviations are small. Future work should test the applicability of Eq. 2 to different imaging directions.

The equations connecting RGB coefficients β_c^D and β_c^B to wavelength dependent physical quantities are [1]:

$$\beta_c^D = \ln \left[\frac{\int_{\lambda_1}^{\lambda_2} S_c(\lambda)\rho(\lambda)E(d,\lambda)e^{-\beta(\lambda)z}d\lambda}{\int_{\lambda_1}^{\lambda_2} S_c(\lambda)\rho(\lambda)E(d,\lambda)e^{-\beta(\lambda)(z+\Delta z)}d\lambda} \right] \Big/ \Delta z \ ,$$

$$(3)$$

$$\beta_c^B = -\ln \left[1 - \frac{\int_{\lambda_1}^{\lambda_2} S_c(\lambda)B^\infty(\lambda)(1-e^{-\beta(\lambda)z})d\lambda}{\int_{\lambda_1}^{\lambda_2} B^\infty(\lambda)S_c(\lambda)d\lambda} \right] \Big/ z \ .$$

$$(4)$$

Here, λ_1 and λ_2 are the limits of the visible range (400 and 700nm), E is the spectrum of ambient light at depth d. Light penetrating vertically attenuates based on the *diffuse downwelling attenuation* $K_d(\lambda)$ [47], different than the beam attenuation coefficient $\beta(\lambda)$ [1, 47] which is solely a function of the type, composition, and density of dissolved substances in the ocean [47]. If $E(0, \lambda)$ is light at the sea surface, then $E(d, \lambda)$ at depth d is [2]:

$$E(d, \lambda) = E(0, \lambda)e^{-K_d(\lambda)d} \ . \quad (5)$$

The veiling light B_c^∞ in Eq. 2 is given as:

$$B_c^\infty = \int_{\lambda_1}^{\lambda_2} S_c(\lambda)B^\infty(\lambda)d\lambda \ , \quad (6)$$

where

$$B^\infty(\lambda) = \left[b(\lambda)E(d,\lambda)\right]/\beta(\lambda) \ . \quad (7)$$
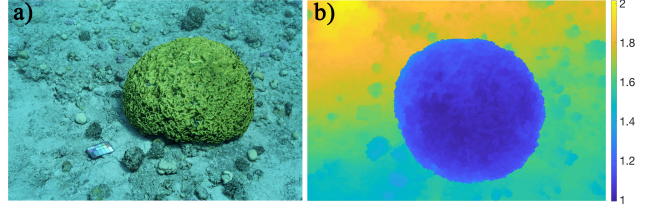

Figure 3. **a)** A 3D model created from 68 photographs using Photoscan Professional (Agisoft LLC). **b)** Range map z (in meters) for the image in Fig. 1 obtained from this model. We placed a color chart on the seafloor to set the scale.

4. The *Sea-thru* Method

Based on Eqs. 2-4, to recover J_c the following parameters need to be known or estimated: optical water type determined by b, β, and K_d; light E_d, distance between the camera and scene along the line-of-sight z, depth at which the photo was taken d, the reflectance of each object in the scene ρ, and the spectral response of the camera S_c. These parameters are rarely, if ever, known at the time an underwater photo is taken. In [1] we showed that β_c^D was most strongly governed by z, and β_c^B was most affected by the optical water type and illumination E. Therefore, we tailor the *Sea-thru* method to tackle these specific dependencies. Since the coefficients vary with imaging angle and exposure [1], we assume that they generally cannot be transferred across images, even those taken sequentially with the same camera, and we estimate the relevant parameters for a given image from that image only.

4.1. Imaging and Range Map Generation

As β_c^D heavily depends on z we require having a range map of the scene, which we obtain using structure-from-motion (SFM), commonly used underwater to measure structural complexity of reefs (e.g., [12, 15, 16, 27, 28, 42]) and in archaeology (e.g., [45, 46]). Our method requires an absolute value for z, whereas SFM provides range only up to scale, so we placed objects of known sizes in the scene (Fig. 3). When imaging from underwater vehicles, their navigation sensors can provide velocity or altitude. An alternative is stereo imaging (e.g., [7, 9, 35, 54]), which requires the use of two synchronized cameras, and a straightforward in-water calibration before imaging survey begins.

4.2. Scene Reconstruction

From Eqs. 1 & 2 we have:

$$J_c = D_c e^{\beta_c^D(z)z} \ , \quad (8)$$

where $D_c = I_c - B_c$. Here we explicitly kept the z dependency of β_c^D as we must account for it, but we will ignore other dependencies. J_c is an image whose colors are only corrected along z, and depending on the imaging geometry, it may need further correction to achieve the colors of
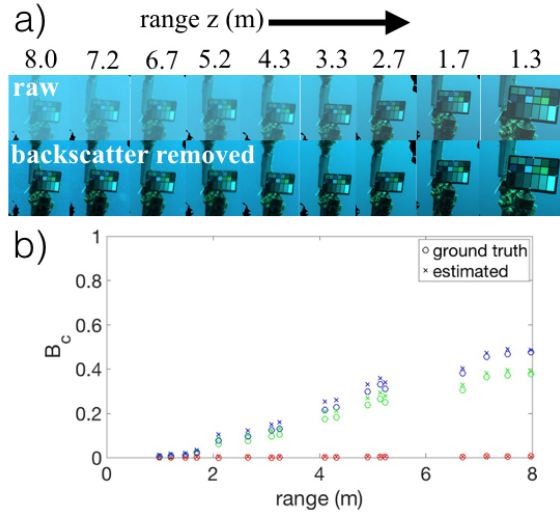
a)
range z (m) →
8.0 7.2 6.7 5.2 4.3 3.3 2.7 1.7 1.3

raw

backscatter removed

b)

Figure 4. A color chart imaged at various ranges. Top row in **a** shows the raw images I_c, and bottom row shows the corresponding D_c, or backscatter-removed images. **b)** B_c calculated for each color channel with the method described here (x's), and the color-chart based backscatter calculation method described in [1] (o's); values are almost identical.

an image that was taken at sea surface. Let J_s denote the image taken at the surface. Then,

$$J_s = J_c/W_c \quad , \qquad (9)$$

where W_c is the white point of the ambient light at the camera (i.e., at depth d), and J_s is J_c globally white balanced.

4.3. Backscatter Estimation

Backscatter increases exponentially with z, and eventually saturates (Fig. 2). Where scene reflectance $\rho_c \to 0$ (all light absorbed), or $E \to 0$ (complete shadow), the captured RGB intensity $I_c \to B_c$. We search the image for very dark or shadowed pixels, and use them to get an initial estimate of backscatter. Our approach is similar to DCP in that it attempts to find the backscattered signal where the D_c is minimum, but it differs in the fundamental way that we utilize a known range map, rather than try to estimate it. Additionally, we search for the darkest RGB triplets rather than finding the darkest pixels independently in each color channel and we do not form a dark channel image. The small number of unconnected pixels our method identifies is sufficient, because we have the corresponding range information, and a physical model of how B_c behaves with z (but see note about imaging angle in Sec. 3).

We estimate backscatter as follows: first we partition the range map into 10 evenly spaced clusters spanning the minimum and maximum values of z. In each range cluster, we search I_c for the RGB triplets in the bottom 1 percentile, which we denote by Ω. Then across the whole image, $\hat{B}_c(\Omega) \approx I_c(\Omega)$ is an overestimate of backscatter, which

we model as:

$$\hat{B}_c = B_c^\infty (1 - e^{-\beta_c^B z}) + J_c' e^{-\beta_c^{D'} z} \quad , \qquad (10)$$

where the expression $J_c' e^{-\beta_c^{D'} z}$ represents a small residual term that behaves like the direct signal. Using non-linear least squares fitting, we estimate parameters B_c^∞, β_c^B, J_c', and $\beta_c^{D'}$ subject to the bounds $[0, 1]$, $[0, 5]$, $[0, 1]$, and $[0, 5]$, respectively. For this step, we ignore the z-dependency of $\beta_c^{D'}$. If information about the camera sensor, water type, etc., is available, the bounds for β_c^D and β_c^B can be further refined using the loci described in [1, 2].

Depending on the scene, the residual can be left out of Eq. 10 if the reflectance of found dark pixels are perfect black; if they are under a shadow; if z is large; or if the water is extremely turbid ($B_c \gg D_c$). In all other cases, the inclusion of the residual term is important. In reef scenes, due to their complex 3D structure, there are often many shadowed pixels which provide direct estimates of backscatter.

Fig. 4 demonstrates the performance of this method in a calibrated experiment. We mounted a chart on a buoy line in blue water (to minimize interreflections from to the seafloor and surface), and photographed as we swam towards it. The veiling effect of backscatter is clearly visible in the far images, decreasing as z between the camera and the chart decreases (Fig. 4a). For each image, we calculated the ground-truth backscatter using the achromatic patches of the chart as described in [1], and also estimated it using the method described here (Fig. 4b). The resulting B_c values are almost identical; no inputs (e.g., water type) other than I_c and z were needed to obtain this result. Note that the black patch of the color chart was not picked up in Ω in any of the images, indicating that it is indeed a just dark gray and much lighter than true black or shadowed pixels.

4.4. Attenuation Coefficient Estimation

4.4.1 β_c^D as a Function of z

We previously showed that β_c^D varies most strongly with range z [1, 2]. Inspecting Eq. 3 suggests that this variation is in the form of an exponential decay. Before we describe how to extract $\beta_c^D(z)$ from images, we formulate the relationship between β_c^D and z.

Fig. 5 shows an experiment from [2] where we mounted a color chart and a Nikon D90 camera on a frame roughly 20cm apart, and lowered this configuration from surface to 30m depth while taking photographs. Backscatter and attenuation between the camera and the chart are both negligible since the distance z between the chart and the camera was small, yielding $I_c \to J_c$. In this setup, the color loss is due to the effective attenuation coefficient acting in the vertical distance d from the sea surface, and is captured in the white point of ambient light W_c at each depth.
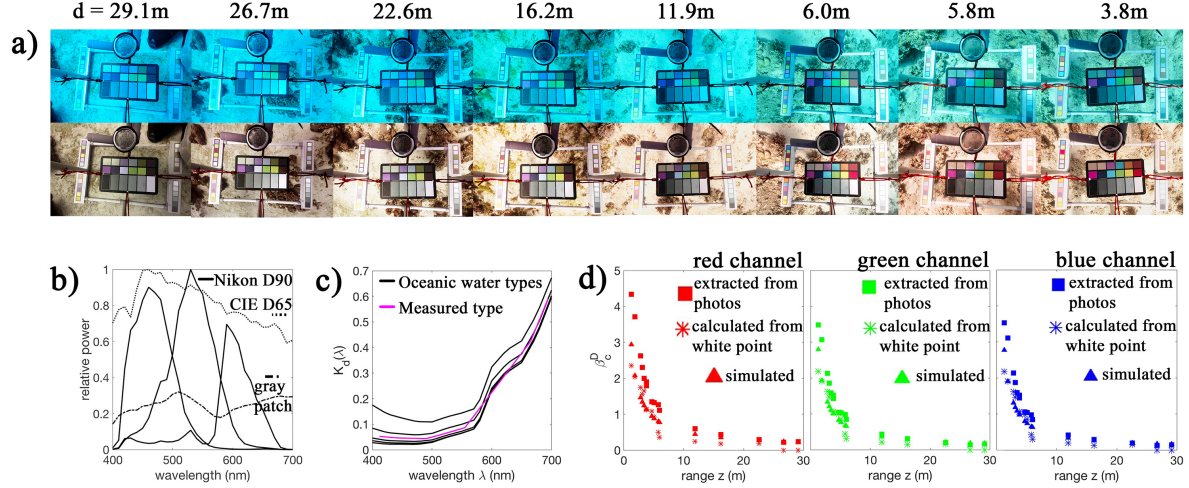
Figure 5. Using a dataset from [2] we calculate $\beta_c^D(z)$ three different ways: extracting it from photos of the same object at two different distances; calculating it using the white point of the ambient light using Eq. 9; and simulating it using Eq. 3. For this dataset, a color chart (DGK Color Tools) and the camera were mounted on a frame very close to each other so that $I_c \rightarrow J_c$. The frame was imaged from surface to 30m depth in the Red Sea using a Nikon D90. **a)** Raw images captured by the camera (top row; not all are shown), and same images after white balancing using the achromatic patch (bottom row). Brightness in each image was manually adjusted for visualization. **b)** For simulating the value of β_c^D, we used the spectral response of Nikon D90 from [38], assumed CIE D65 light at the surface [61], measured the reflectance of the second brightest gray patch (note that it does not reflect uniformly), and for the optical water type, used the diffuse downwelling attenuation coefficient $K_d(\lambda)$ we measured *in situ* (magenta curve in **c**). This curve agrees well with the oceanic water types defined by Jerlov (black curves in **c**) [36, 37]. **d)** β_c^D decays as a 2-term exponential with z as shown by all three methods.

From each image, we calculated the effective β_c^D in the vertical direction two different ways: from pairwise images as described in [2], and by using Eq. 9 with W_c extracted from the intensity of the second (24%) gray patch in the color chart. Additionally, we used Eq. 3 to calculate the theoretical value of β_c^D in that water type using the spectral response of the camera from [38], and the $K_d(\lambda)$ of the water body (which acts in the vertical direction) which we had measured. All three ways of estimating β_c^D in Fig. 5 demonstrate that β_c^D decays with distance, in this case d. Based on the data in Fig. 5 and additional simulations we describe the dependency of β_c^D on any range z using a 2-term exponential in the form of:

$$\beta_c^D(z) = a * \exp(b * z) + c * \exp(d * z) \ . \quad (11)$$

For short ranges, $\beta_c^D(z)$ can also be modeled as a line.

4.4.2 Coarse Estimate of $\beta_c^D(z)$ From an Image

Assuming B_c is successfully removed from image I_c, we can now proceed to estimating $\beta_c^D(z)$ from the direct signal D_c. Note from Eq. 2 that the direct signal is the product of the scene J_c (at the location of the camera) attenuated by $e^{-\beta_c^D(z)z}$. Thus, the recovery of the scene J_c reduces to a problem of the estimation of the illuminant map between the camera and the scene, which varies spatially. Given an estimate of the local illuminant map $\hat{E}_c(z)$, we can obtain

an estimate of $\beta_c^D(z)$ as follows:

$$\hat{\beta}_c^D(z) = -\log \hat{E}_c(z)/z. \quad (12)$$

Estimation of an illuminant locally is a well-studied topic in the field of computational color constancy (e.g., [6, 10, 22, 29, 32, 41]). Several methods, most notably the Retinex model which mimics a human's ability to discount varying illuminations, have been applied on underwater imagery (e.g., [30, 62]), and a recent work showed that there is a direct linear relationship between atmospheric image dehazing and Retinex [31]. If backscatter is properly removed from original images, we can expect many of the multi-illuminant estimation methods to work well on underwater images. Here, we adopt a variant of the local space average color (LSAC) method described in [24], as it utilizes a known range map. This method works as follows: for a given pixel (x, y) in color channel c, local space average color $a_c(x, y)$ is estimated iteratively through updating the equations:

$$a_c'(x, y) = \frac{1}{N_e} \sum_{N_e} a_c(x', y') \quad (13)$$

$$a_c(x, y) = D_c(x, y)p + a_c'(x, y)(1 - p) \ , \quad (14)$$

where the neighborhood N_e is defined as the 4-connected pixels neighboring the pixel at (x, y) which are closer to it than a range threshold $\in$:

$$N_e(x', y') = (x', y') \text{ with } \|z(x, y) - z(x', y')\| \leq \in. \quad (15)$$

Set	Scene	Depth	Angle	B_c	Water Type	# images	Camera	Lens
D1	reef	10m	down	low	clear	559	Sony α7R Mk III	Sony FE 16-35mm f/2.8GM
D2	reef	10m	down	high	clear	318	Sony α7R Mk III	Sony FE 16-35mm f/2.8GM
D3	reef	4m	all	low	clear	68	Sony α7R Mk III	Sony FE 16-35mm f/2.8GM
D4	canyon	4-9m	down	high	turbid	153	Nikon D810	Nikkor 35mm f1.8
D5	reef	5m	forward	med	clear	59	Nikon D810	Nikkor 35mm f1.8

Table 1. Datasets we contribute with SFM-based range maps for each image. Each set contains multiple images with a color chart.

Here, the initial value of $a(x, y)$ is taken as zero for all pixels, since after a large number of iterations the starting value will be insignificant. The parameter p describes the local area of support over which the average is computed and depends on the size of the image; large p means that local space average color will be computed for a small neighborhood. Then, the local illuminant map is found as $\hat{E}_c = fa_c$, where f is a factor based on geometry scaling all color channels equally and can be found based on the scene viewed. We use $f = 2$ following [23] for a perpendicular orientation between the camera and the scene.

4.4.3 Refined Estimate of $\beta_c^D(z)$

Next, we make use of the known range map z and refine the estimate of $\beta_c^D(z)$ found from Eqs. 12-15 to obey the given z. We rewrite Eq. 12 as:

$$\hat{z} = -\log \hat{E}_c / \beta_c^D(z) \quad , \qquad (16)$$

and minimize:

$$\min_{\beta_c^D(z)} \|z - \hat{z}\|, \qquad (17)$$

where $\beta_c^D(z)$ is defined in the form of Eq. 11 with parameters a, b, c, d. The lower and upper bounds for these parameters to obtain a decaying exponential will be $[0, -\infty, 0, -\infty]$, and $[\infty, 0, \infty, 0]$, respectively; but can be narrowed using the coarse estimate obtained from Eq. 12. Using the refined estimate of $\beta_c^D(z)$, we recover J_c using Eq. 8. In J_c, spatial variation of ambient light has already been corrected, so all that remains is the estimation of the global white point W_c. This can be done using statistical or learning based methods (see [32] for a survey); here for the scenes that contain a sufficiently diverse set of colors, we adopt the simple and fast Gray World Hypothesis [14], and for monochromatic scenes (such as in our dataset **D4** we introduce next), we use a spatial-domain method from [18] that does not rely on color information.

4.5. Photofinishing

We use the camera pipeline manipulation platform described in [40] to convert **_Sea-thru_** outputs to a standard color space, inserting them into the pipeline before Step 6 and specifying an identity matrix for white balance.

5. Datasets

We contribute five underwater RGBD datasets (Table 1). All were acquired under natural illumination, in raw format, with constant exposure settings for a given set, and contain multiple images with color charts.

6. Results: Validation and Error Analysis

We use our dataset from Table 1 and the stereo RGBD dataset from [7] to test the following scenarios:
S1. Simple contrast stretch.
S2. Former model with an incorrect estimate of B_c. Specifically, we use DCP [34], which typically overestimates B_c in underwater scenes. We used the built-in *imreducehaze* function in Matlab.
S3. Former model, with a correct estimate of B_c (i.e., correct B_c^∞ and β_c^B), and assuming $\beta_c = \beta_c^D = \beta_c^B$.
S4. Revised model, with a correct estimate of B_c, and J_c obtained as $J_c = D_c / \hat{E}_c$ without explicitly computing $\beta_c^D(z)$.
S5. **_Sea-thru_**, which uses the revised model where $\beta_c^B \neq \beta_c^D$, and $\beta_c^D = \beta_c^D(z)$ from Eq. 11.
Since **_Sea-thru_** is the first algorithm to use the revised underwater image formation model and has the advantage of having a range map, we do not test its performance against single image color reconstruction methods that also try to estimate the range/transmission. After a meticulous survey of these methods, authors in [7] found that DCP-based ones [21, 50] were not able to consistently correct colors, and others [3, 4, 5, 25] were designed to enhance images rather than achieve physically accurate corrections. The proposed method in [7] does aim to recover physically accurate colors (using the former model), but only works for horizontal imaging with sufficiently large distances in the scene, making it unsuitable for many of our images.

We present raw images, range maps, and the corresponding **S1-S5** results on **D1-5** in Fig. 6, and on the stereo database of [7] in Fig. 7. For evaluation, we used RGB angular error $\bar{\psi}$ between the six grayscale patches of each chart and a pure gray color, averaged per chart

$$\bar{\psi} = (1/6) \cos^{-1} \left[I_c / (\sqrt{3} \|I_c\|) \right] \quad , \qquad (18)$$

as also done by [7]. Lower $\bar{\psi}$ value indicates better correction (though see exceptions below); errors (in degrees) are
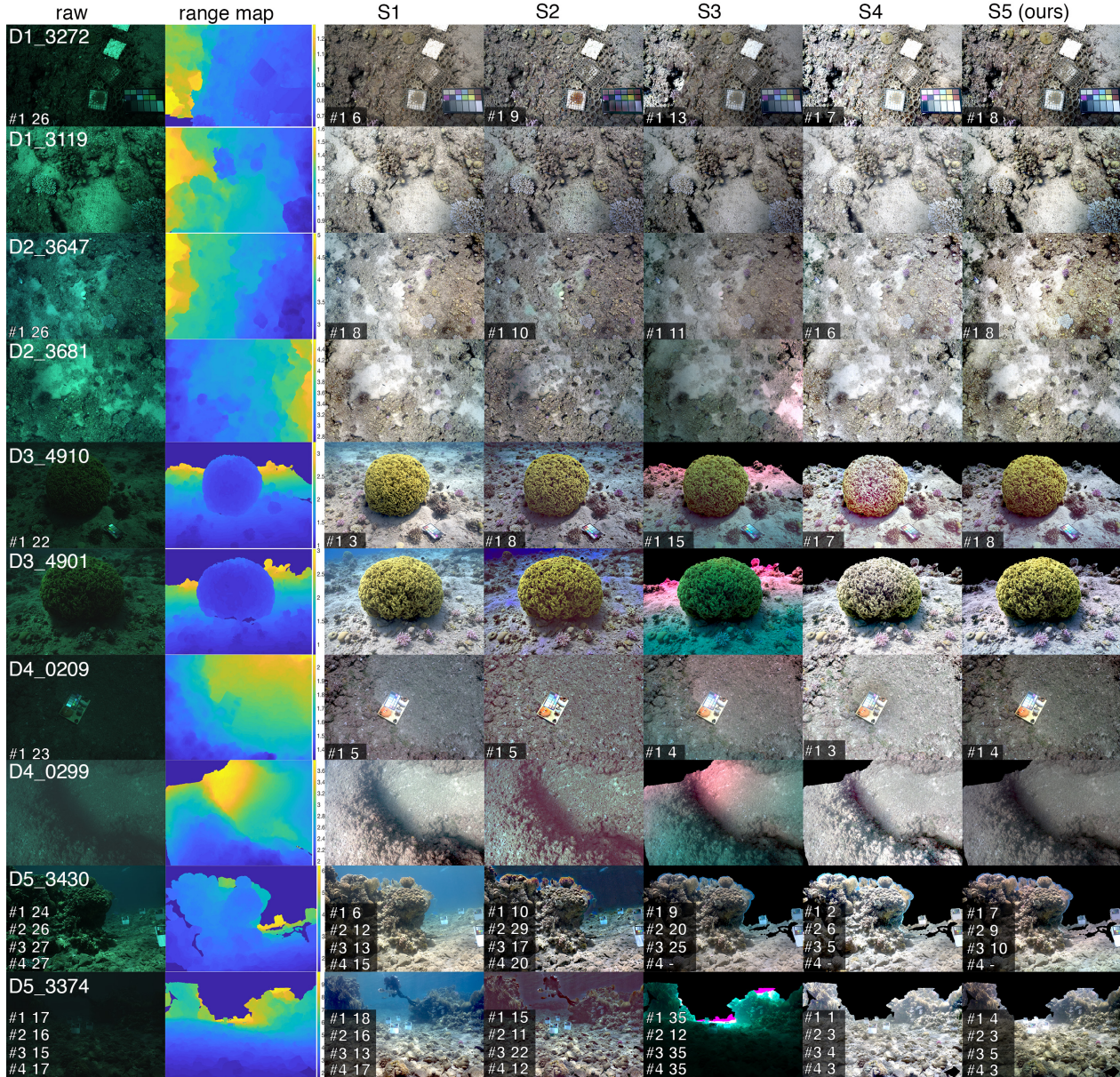
Figure 6. Results on **D1-5** (Table 1). For each chart and method, $\bar{\psi}$ rounded to the nearest integer is given in inset; Chart #1 is closest to the camera. Average errors for the dataset are: **raw:** 20.57, **S1:** 12.49, **S2:** 14.38, **S3:** 21.77, **S4:** 4.13, **S5:** (*Sea-thru*) 6.33.

listed in the insets of Figs. 6 & 7 per color chart for scenes that had them, rounded to the nearest integer.

In all cases, the simple contrast stretch **S1**, which is global, works well when scene distances are more or less uniform. The DCP method (**S2**) often overestimates backscatter (which can improve visibility), and generally distorts and halucinates colors. For example what should be uniformly colored sand appears green and purple in both datasets. In D1_3272, the gray patches of the color chart in **S2** have visible purple artifacts, yet their $\bar{\psi}$ error is lower than that of **S5**, suggesting that $\bar{\psi}$ is not an optimal metric for quantifying color reconstruction error. In **S3-S5**, the

correct amount of B_c is subtracted. In **S3** attenuation is corrected with a constant β_c^D, as had been done by methods using the former model. When there is large variation in range (e.g., Fig. 7), the failure of the constant β_c^D assumption is most evident, and this is also where **S5** has the biggest advantage (though **S3** also fails on scenes with short ranges, e.g., **D3&4**). Range maps often tend to be least accurate furthest from the camera, which also adds to the difficulty of reconstructing colors at far ranges. **S4** sometimes yields lower errors on the color cards than **S5**. This makes sense as it is easier to calculate the illuminant on the cards; however **S5** results are better on the complete scenes. **S4** can be
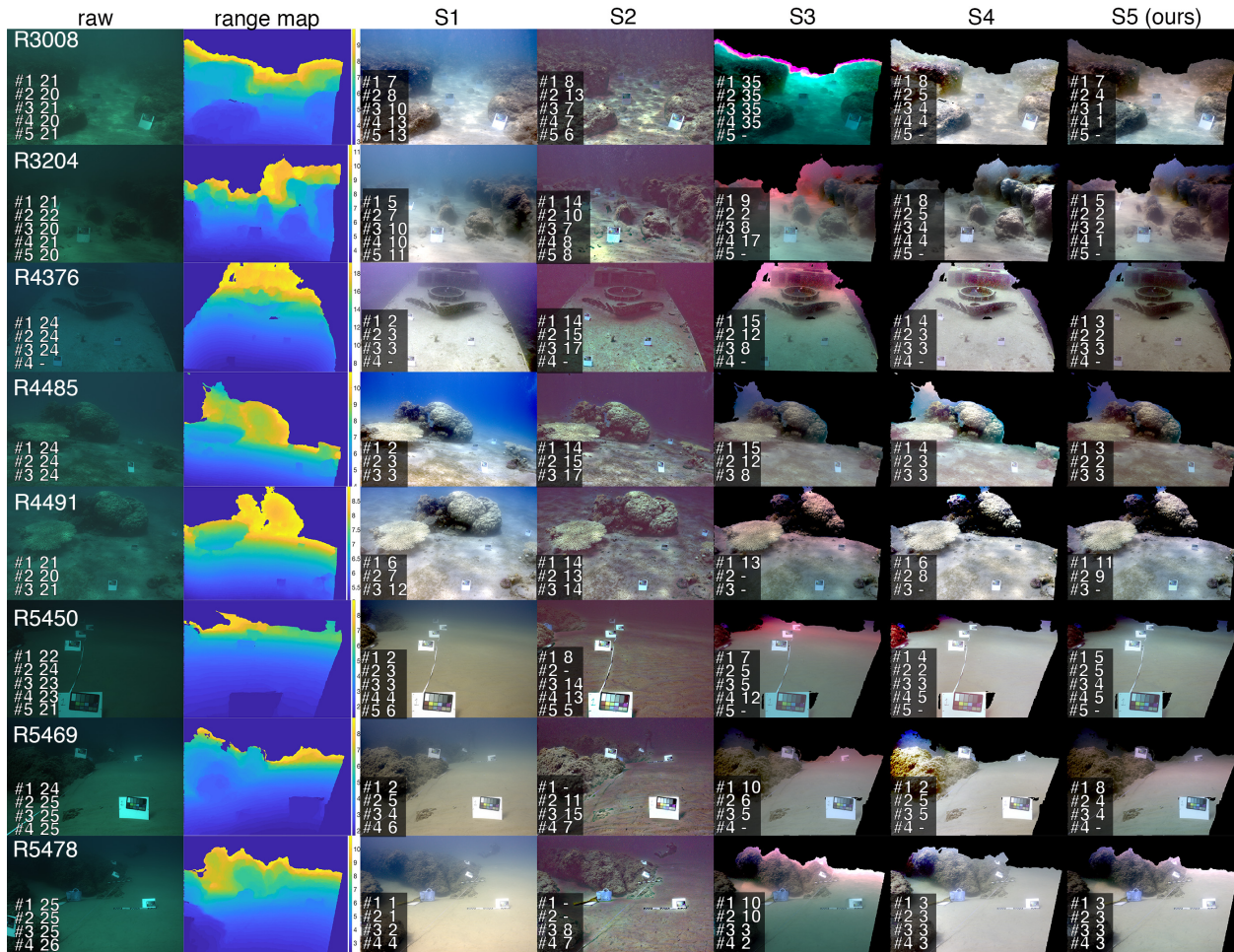
Figure 7. Results on the stereo dataset of [7]. Their range maps were further processed to remove spurious pixels. $\bar{\psi}$ rounded to the nearest integer is given in inset; chart #1 is closest to the camera. **S1** and **S2** do not utilize range maps; for others, lack of $\bar{\psi}$ values indicate missing range information. The average errors across all images are: **raw:** 22.28, **S1:** 6.83, **S2:** 10.03, **S3:** 12.04, **S4:** 4.46, **S5:** (*Sea-thru*) 4.94.

used for a first-order correction that is better than previous methods.

7. Conclusions

Taking its strength from an image formation model derived for the ocean, *Sea-thru* offers a glimpse into the underwater world without skewed colors. It highlights that β_c^D and β_c^B are distinct, and the z-dependency of β_c^D cannot be ignored. We focused on recovering the z-dependency as it is the most prominent, but in the future plan to also recover the ρ-dependency to improve correction. As recovering these intricate dependencies is extremely challenging, deep nets should perform better than the estimation methods we used. Since ground-truth cannot be attained for this environment their training has to be conducted with carefully designed simulations based on the correct image formation models. Careful simulations can also help with another challenge that arose in this work; the evaluation of results. The dataset published in [7] was acquired with significant effort to place

multiple color charts in the scene, but evaluation limited to the charts does not always tell the full story.

Sea-thru is a significant step towards opening up large underwater datasets to powerful computer vision and machine learning algorithms, and will help boost underwater research at a time when our oceans are increasing stress from pollution, overfishing, and climate change.

8. Acknowledgments

References

[1] D. Akkaynak and T. Treibitz. A revised underwater image formation model. In *Proc. IEEE CVPR*, 2018. 1, 2, 3, 4

[2] D. Akkaynak, T. Treibitz, T. Shlesinger, R. Tamir, Y. Loya, and D. Iluz. What is the space of attenuation coefficients in underwater computer vision? In *Proc. IEEE CVPR*, 2017. 2, 3, 4, 5

[3] C. Ancuti, C. O. Ancuti, C. De Vleeschouwer, R. Garcia, and A. C. Bovik. Multi-scale underwater descattering. In *Proc. IEEE Int. Conf. on Pattern Recognition (ICPR)*, pages 4202–4207, 2016. 6

[4] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert. Color balance and fusion for underwater image enhancement. *IEEE Trans. Image Processing*, 27(1):379–393, 2018. 6

[5] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, L. Neumann, and R. Garcia. Color transfer for underwater dehazing and depth estimation. In *Proc. IEEE Int. Conf. Image Processing (ICIP)*, pages 695–699, 2017. 6

[6] S. Beigpour, C. Riess, J. Van De Weijer, and E. Angelopoulou. Multi-illuminant estimation with conditional random fields. *IEEE Trans. Image Processing*, 23(1):83–96, 2014. 5

[7] D. Berman, D. Levy, S. Avidan, and T. Treibitz. Underwater single image color restoration using haze-lines and a new quantitative dataset, 2018. 2, 3, 6, 8

[8] D. Berman, T. Treibitz, and S. Avidan. Non-local image dehazing. In *Proc. IEEE CVPR*, 2016. 2

[9] D. Berman, T. Treibitz, and S. Avidan. Diving into haze-lines: Color restoration of underwater images. In *Proc. British Machine Vision Conference (BMVC)*, 2017. 2, 3

[10] M. Bleier, C. Riess, S. Beigpour, E. Eibenberger, E. Angelopoulou, T. Tröger, and A. Kaup. Color constancy and non-uniform illumination: Can existing algorithms work? In *Proc. IEEE ICCVWorkshops*, pages 774–781, 2011. 5

[11] D. L. Bongiorno, M. Bryson, and S. B. Williams. Dynamic spectral-based underwater colour correction. In *Proc. MTS/IEEE OCEANS*, 2013. 2

[12] M. Bryson, R. Ferrari, W. Figueira, O. Pizarro, J. Madin, S. Williams, and M. Byrne. Characterization of measurement errors using structure-from-motion and photogrammetry to measure marine habitat structural complexity. *Ecology and Evolution*, 7(15):5669–5681, 2017. 3

[13] M. Bryson, M. Johnson-Roberson, O. Pizarro, and S. B. Williams. True color correction of autonomous underwater vehicle imagery. *J. Field Robotics*, 2015. 2

[14] G. Buchsbaum. A spatial processor model for object colour perception. *J. the Franklin institute*, 310(1):1–26, 1980. 6

[15] J. Burns, D. Delparte, R. Gates, and M. Takabayashi. Integrating structure-from-motion photogrammetry with geospatial software as a novel technique for quantifying 3d ecological characteristics of coral reefs. *PeerJ*, 3:e1077, 2015. 3

[16] J. Burns, D. Delparte, L. Kapono, M. Belt, R. Gates, and M. Takabayashi. Assessing the impact of acute disturbances on the structure and composition of a coral community using innovative 3D reconstruction techniques. *Methods in Oceanography*, 15:49–59, 2016. 3

[17] N. Carlevaris-Bianco, A. Mohan, and R. M. Eustice. Initial results in underwater single image dehazing. In *Proc. MTS/IEEE OCEANS*, 2010. 2

[18] D. Cheng, D. K. Prasad, and M. S. Brown. Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution. *JOSA A*, 31(5):1049–1058, 2014. 6

[19] J. Y. Chiang and Y.-C. Chen. Underwater image enhancement by wavelength compensation and dehazing. *IEEE Trans. Image Processing*, 21(4):1756–1769, 2012. 2

[20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009. 1

[21] P. Drews, E. Nascimento, F. Moraes, S. Botelho, and M. Campos. Transmission estimation in underwater single images. In *Proc. IEEE ICCV Underwater Vision Workshop*, pages 825–830, 2013. 2, 6

[22] M. Ebner. *Color constancy*, volume 6. John Wiley & Sons, 2007. 5

[23] M. Ebner. Color constancy based on local space average color. *Machine Vision and Applications*, 20(5):283–301, 2009. 2, 6

[24] M. Ebner and J. Hansen. Depth map color constancy. *Bio-Algorithms and Med-Systems*, 9(4):167–177, 2013. 5

[25] S. Emberton, L. Chittka, and A. Cavallaro. Underwater image and video dehazing with pure haze region segmentation. *Computer Vision and Image Understanding*, 168:145–156, 2018. 6

[26] R. Fattal. Dehazing using color-lines. *ACM Trans. on Graphics (TOG)*, 34(1):13, 2014. 2

[27] R. Ferrari, D. McKinnon, H. He, R. N. Smith, P. Corke, M. González-Rivero, P. J. Mumby, and B. Upcroft. Quantifying multiscale habitat structural complexity: a cost-effective framework for underwater 3D modelling. *Remote Sensing*, 8(2):113, 2016. 3

[28] W. Figueira, R. Ferrari, E. Weatherby, A. Porter, S. Hawes, and M. Byrne. Accuracy and precision of habitat structural complexity metrics derived from underwater photogrammetry. *Remote Sensing*, 7(12):16883–16900, 2015. 3

[29] G. D. Finlayson, B. V. Funt, and K. Barnard. Color constancy under varying illumination. In *Proc. IEEE ICCV*, pages 720–725, 1995. 5

[30] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X.-P. Zhang, and X. Ding. A retinex-based enhancing approach for single underwater image. In *Proc. IEEE Int. Conf. Image Processing (ICIP)*, pages 4572–4576, 2014. 5

[31] A. Galdran, A. Alvarez-Gila, A. Bria, J. Vazquez-Corral, and M. Bertalmío. On the duality between retinex and image dehazing. In *Proc. IEEE CVPR*, pages 8212–8221, 2018. 5

[32] A. Gijsenij, T. Gevers, J. Van De Weijer, et al. Computational color constancy: Survey and experiments. *IEEE Trans. Image Processing*, 20(9):2475–2489, 2011. 5, 6

[33] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. In *Proc. IEEE CVPR*, 2009. 2

[34] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *Trans. IEEE PAMI*, 33(12):2341–2353, 2011. 2, 6

[35] J. Henderson, O. Pizarro, M. Johnson-Roberson, and I. Mahon. Mapping submerged archaeological sites using stereovision photogrammetry. *Int. J. Nautical Archaeology*, 42(2):243–256, 2013. 3

[36] N. Jerlov. *Irradiance Optical Classification*. Elsevier, 1968. 5

[37] N. G. Jerlov. *Marine optics*, volume 14. Elsevier, 1976. 5

[38] J. Jiang, D. Liu, J. Gu, and S. Susstrunk. What is the space of spectral sensitivity functions for digital color cameras? In *IEEE Workshop Applications of Computer Vision (WACV)*, pages 168–179, 2013. 5

[39] M. I. Jordan and T. M. Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015. 1

[40] H. C. Karaimer and M. S. Brown. A software platform for manipulating the camera imaging pipeline. In *Proc. ECCV*, pages 429–444. Springer, 2016. 6

[41] E. H. Land and J. J. McCann. Lightness and retinex theory. *J. Optical Society of America*, 61(1):1–11, 1971. 5

[42] J. X. Leon, C. M. Roelfsema, M. I. Saunders, and S. R. Phinn. Measuring coral reef terrain roughness using structure-from-motionclose-range photogrammetry. *Geomorphology*, 242:21–28, 2015. 3

[43] Y. Li, H. Lu, J. Li, X. Li, Y. Li, and S. Serikawa. Underwater image de-scattering and classification by deep neural network. *Computers & Electrical Engineering*, 54:68–77, 2016. 2

[44] H. Lu, Y. Li, L. Zhang, and S. Serikawa. Contrast enhancement for images in turbid water. *JOSA A*, 32(5):886–893, 2015. 2

[45] J. McCarthy and J. Benjamin. Multi-image photogrammetry for underwater archaeological site recording: an accessible, diver-based approach. *J. Maritime Archaeology*, 9(1):95–114, 2014. 3

[46] J. Mertes, T. Thomsen, and J. Gulley. Evaluation of structure from motion software to create 3d models of late nineteenth century great lakes shipwrecks using archived diver-acquired video surveys. *J. Maritime Archaeology*, 9(2):173–189, 2014. 3

[47] C. D. Mobley. *Light and water: radiative transfer in natural waters*. Academic press, 1994. 3

[48] S. K. Nayar and S. G. Narasimhan. Vision in bad weather. 1, 2

[49] Y.-T. Peng, X. Zhao, and P. C. Cosman. Single underwater image enhancement using depth estimation based on blurriness. In *Proc. IEEE ICIP*, pages 4952–4956, 2015. 2

[50] Y.-T. Peng, X. Zhao, and P. C. Cosman. Single underwater image enhancement using depth estimation based on blurriness. In *Proc. IEEE Int. Conf. Image Processing (ICIP)*, pages 4952–4956, 2015. 6

[51] M. Roser, M. Dunbabin, and A. Geiger. Simultaneous underwater visibility assessment, enhancement and improved stereo. In *IEEE Conf. Robotics and Automation*, pages 3840–3847, 2014. 2

[52] Y. Y. Schechner and N. Karpel. Recovery of underwater visibility and structure by polarization analysis. *IEEE J. Oceanic Engineering*, 30(3):570–587, 2005. 2

[53] Y.-S. Shin, Y. Cho, G. Pandey, and A. Kim. Estimation of ambient light and transmission map with common convolutional architecture. In *Proc. MTS/IEEE OCEANS*, 2016. 2

[54] M. Shortis and E. H. D. Abdo. A review of underwater stereo-image measurement for marine biology and ecology applications. In *Oceanography and marine biology*, pages 269–304. CRC Press, 2016. 3

[55] K. A. Skinner, E. Iscar, and M. Johnson-Roberson. Automatic color correction for 3D reconstruction of underwater scenes. In *Proc. IEEE International Conference Robotics and Automation (ICRA)*, pages 5140–5147, 2017. 2

[56] O. Spier, T. Treibitz, and G. Gilboa. *In Situ* target-less calibration of turbid media. In *Proc. IEEE ICCP*, 2017. 2

[57] R. T. Tan. Visibility in bad weather from a single image. In *Proc. IEEE CVPR*, 2008. 2

[58] T. Treibitz and Y. Y. Schechner. Active polarization descattering. *IEEE Trans. PAMI*, 31(3):385–399, 2009. 2

[59] K. Williams, C. Rooper, and J. Harms. Report of the national marine fisheries service automated image processing workshop. *NOAA Technical Memorandum NMFS-F/SPO-121*, 2012. 2

[60] G. Winters, R. Holzman, A. Blekhman, S. Beer, and Y. Loya. Photographic assessment of coral chlorophyll contents: implications for ecophysiological studies and coral monitoring. *J. Exp. Marine Biology and Ecology*, 380(1):25–35, 2009. 2

[61] G. Wyszecki and W. S. Stiles. *Color science*, volume 8. Wiley New York, 1982. 5

[62] S. Zhang, T. Wang, J. Dong, and H. Yu. Underwater image enhancement via extended multi-scale retinex. *Neurocomputing*, 245:1–9, 2017. 5

[63] X. Zhao, T. Jin, and S. Qu. Deriving inherent optical properties from background color and underwater image enhancement. *Ocean Engineering*, 94:163–172, 2015. 2