SUCRe: Leveraging Scene Structure for Underwater Color Restoration

Clémentin Boittiaux[1,2,3]   Ricard Marxer[3]   Claire Dune[2]   Aurélien Arnaubec[1]
Maxime Ferrera[1]   Vincent Hugel[2]

[1] Ifremer, Centre Méditerranée, La Seyne-sur-Mer France
[2] Université de Toulon, COSMER, Toulon, France
[3] Université de Toulon, Aix Marseille Univ, CNRS, LIS, Toulon, France

Abstract

Underwater images are altered by the physical characteristics of the medium through which light rays pass before reaching the optical sensor. Scattering and wavelength-dependent absorption significantly modify the captured colors depending on the distance of observed elements to the image plane. In this paper, we aim to recover an image of the scene as if the water had no effect on light propagation. We introduce SUCRe, a new method that exploits the scene's 3D structure for underwater color restoration. By following points in multiple images and tracking their intensities at different distances to the sensor, we constrain the optimization of the parameters in an underwater image formation model and retrieve unattenuated pixel intensities. We conduct extensive quantitative and qualitative analyses of our approach in a variety of scenarios ranging from natural light to deep-sea environments using three underwater datasets acquired from real-world scenarios and one synthetic dataset. We also compare the performance of the proposed approach with that of a wide range of existing state-of-the-art methods. The results demonstrate a consistent benefit of exploiting multiple views across a spectrum of objective metrics. Our code is publicly available at github.com/clementinboittiaux/sucre.

1. Introduction

Images captured underwater are significantly different from those taken above the surface. Water has a large effect on light transport, inducing various changes of appearance in the scene. Particles in suspension propagate light in multiple directions, inducing what is known as scattering. In addition, strong wavelength-dependent absorption greatly modifies the color of light that reaches the sensor. More importantly, these effects are highly dependent on the water conditions and the distance between elements of the scene and the sensor [1, 3, 20, 30, 43]. This source of visual aspect variability presents a challenge for interpreting and
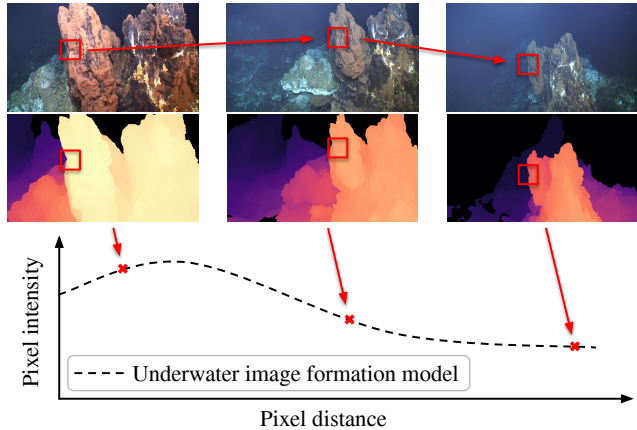

Figure 1: **Multi-view tracking.** We track pixels in multiple images to retrieve their intensities at different distances. We then estimate simultaneously their corrected color and the parameters of an underwater image formation model.

exploiting underwater images and has been long sought to be reduced.

Underwater color restoration aims to remove the effects of the water medium, a strategy that has already been shown to significantly improve the usability of underwater images in many tasks [5, 7, 26]. We can identify two main types of approaches: statistical methods and physics-based models. Statistical methods rely on intrinsic properties of the images to improve their visual aspect. Some methods rely on fusing color and contrast enhanced versions of the original underwater image [4, 5], while others leverage the expressiveness of neural networks by training models using pairs of underwater images and their corresponding references. Such parallel data can be generated synthetically [47], obtained via domain distribution modeling such as GANs [27] or using human preference manual annotations [25]. In contrast, physics-based approaches have focused on explicitly modeling the formation of images under water [1, 3, 20, 30, 32], invert the model and recover the scene appearance as if it

were captured without the effects of the described phenomena [2, 7, 8, 27, 32]. For instance, *Sea-thru* [2] restores colors from a single image and its corresponding distance map by estimating the parameters related to absorption and scattering. However, estimating these parameters from a single image is an ill-posed problem. To cope with this, *Sea-thru* relies on additional assumptions, like a distance-wise adaptation of the dark channel prior [17], to allow independent estimation of the parameters.

We hereby propose exploiting multiple views from the same scene to further constrain the aforementioned ill-posed problem. We rely on Structure-from-Motion (SfM) not only to estimate distance maps, but also to track observed intensities of elements in the scene at different distances based on estimated camera poses. In recent years, SfM has become a common intermediate step of many popular airborne computer vision algorithms [31, 36]. In underwater scenarios, SfM has been proven highly feasible with an appropriate camera setup [12, 16, 22]. Pizarro *et al.* have demonstrated the effectiveness of underwater SfM, even in scenarios with low overlapping images and challenging 3D structures [35]. As such, SfM has also become increasingly popular for use in underwater computer vision algorithms [2, 10, 32, 40, 42]. While it requires additional information about the scene, the proposed approach is well suited for the creation of large-scale datasets with reference images from real-world scenarios, which can be used to train underwater image enhancement neural networks.

Our contributions are the following:
1) We introduce a novel multi-view method that simultaneously estimates the parameters of an underwater image formation model alongside the restored image by tracking points in multiple images to retrieve their intensities at different distances to the scene (see Figure 1).
2) We validate experimentally the developed approach on synthetic and real-world datasets in both natural light and deep-sea scenarios. We perform extensive objective quantitative evaluation on two datasets containing reference ground truth data: synthesized underwater images [47] and color charts captured under water [2]. The applicability and ecological validity of our method is confirmed with qualitative analysis on images from two deep dive surveys: Eiffel Tower [29] and a submarine wreck. Our results demonstrate that leveraging multiple views enables restoring colors that are significantly attenuated in the underwater images due to the distance to the scene. Furthermore, it consistently improves the accuracy of color rendering of elements. Given its improved performance, our approach can be used to produce new reference images for training single-view underwater color restoration methods. To fully appreciate the effectiveness of our underwater color restoration results, we encourage readers to watch the accompanying supplementary videos, which provide compelling visual comparisons.

Variable	Description	Type
I	underwater images	$\mathbb{R}^{N \times H \times W \times C}$
J	restored images	$\mathbb{R}^{N \times H \times W \times C}$
z	distance maps of images	$\mathbb{R}^{N \times H \times W}$
B	veiling light	$\mathbb{R}^{C}$
i	image index	$[1..N]$
c	color channel index	$[1..C]$
p	pixel index	$[1..H \times W]$

Table 1: **Underwater image formation model variables used across the paper.** We use subscripts to index specific images, color channels and pixels, *e.g.*, $I_{i,c,p}$ is the intensity of pixel p of channel c of image i. In some equations the image index i may be omitted for readability. Bold symbols are used for variables encoding spatial information such as images or distances maps. N is the number of images (*i.e.*, views), H and W are the height and width of images and C the number of channels $C = |\{R, G, B\}|$.

2. Related work

A large body of work tackles restoration of single still images without using any additional source of information [7, 8, 13, 18, 26, 38, 46]. Some successful solutions assume the availability of distance maps [2], usually estimated using multiple views.

We focus on making the most out of image sequences, such as those available in underwater video captures, by exploiting SfM results. Bryson *et al.* leverage multi-view observations of texture patches to estimate the parameters of an underwater image formation model and restore the color of the texture map of an underwater 3D model generated by SfM [9, 10]. Nakath *et al.* [32] propose employing a differentiable rendering framework. The authors use the parameters of a previously fitted underwater image formation model to render a textured 3D mesh from multiple views. The error between the acquired and the rendered images is then backpropagated directly into the texture map.

While single image restoration is the most general and challenging formulation of the problem, work on the multi-view setting is of great benefit to produce approximate references for evaluation or target pairs for supervised learning, *e.g.*, in neural network-based approaches [25, 27, 34]. Our work relies on two main prior developments that are presented here as background.

Underwater Image Formation Model (UIFM): A key component of physics-based color restoration methods is their UIFM that describes how the colors of the observed scene are affected by the water medium [1, 3, 20, 30, 32]. While there are many phenomena that impact the qual-
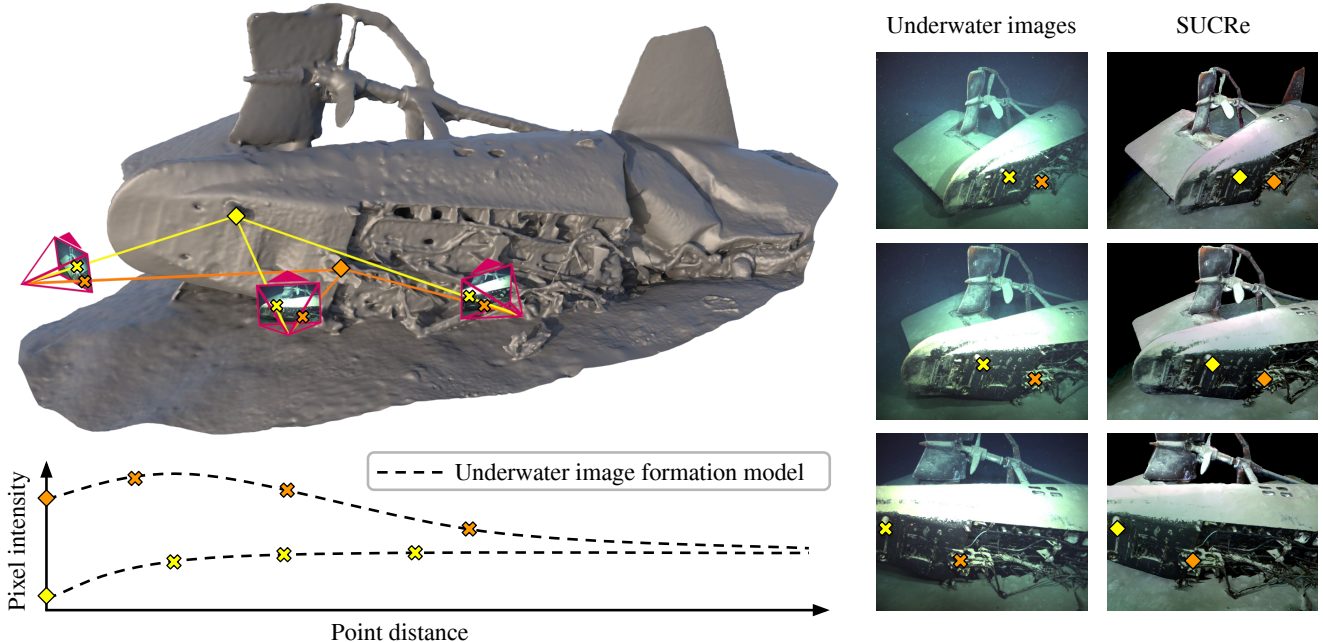
Figure 2: **SUCRe pipeline.** We use camera poses, intrinsics and depth maps resulting from a SfM to pair geometrically pixels between different views. We project pixels from one view to another, enabling us to pair points in low contrast areas. We then simultaneously estimate an UIFM parameters along with the restored image. This figure illustrates our method on a real-world deep-sea dive at a submarine wreck.

ity of underwater images, two of them are predominant: *i)* light collides with floating particles that then act as sources of light, inducing scattering [1]; *ii)* light is attenuated through the water column [3]. Because both phenomena strongly depend on the wavelength [1, 3], UIFM parameters are often wavelength-dependent. In most cases, color restoration is performed for each color channel independently [2, 7, 8, 32, 38, 44].

Many underwater color restoration methods [7, 8, 13, 38] rely on the UIFM introduced in [38] to model backscatter and color attenuation in natural light conditions:

$$\boldsymbol{I}_{c,p} = \boldsymbol{J}_{c,p}e^{-\alpha_c \boldsymbol{z}_p} + B_c(1 - e^{-\alpha_c \boldsymbol{z}_p}), \quad (1)$$

where $\alpha \in \mathbb{R}^C$ is the wavelength-dependent coefficient weighting the distance dependency of color attenuation and backscatter. The other variables are described in Table 1.

Akkaynak *et al.* [1, 3] further revised this model to account for differences between backscatter and attenuation coefficients:

$$\boldsymbol{I}_{c,p} = \boldsymbol{J}_{c,p}e^{-\beta_c \boldsymbol{z}_p} + B_c(1 - e^{-\gamma_c \boldsymbol{z}_p}), \quad (2)$$

where $\beta \in \mathbb{R}^C$ is the color attenuation coefficient and $\gamma \in \mathbb{R}^C$ is the backscatter coefficient.

Sea-thru is a state-of-the-art underwater image color restoration method that relies on images in a raw file format

and their corresponding distance maps [2]. It focuses on inverting the UIFM described by Eq. 2. With the help of the distance information, the problem has $|\boldsymbol{J}_c|$ equations and $|\boldsymbol{J}_c|+|k|$ unknowns, with $k = \{\beta_c, B_c, \gamma_c\}$ the set of UIFM parameters. Given there are more unknowns than observations, the problem is underdetermined and requires additional assumptions to constrain the optimization. For example, an extreme trivial solution can be found with $\boldsymbol{J}_c \rightarrow \boldsymbol{I}_c$, $\beta_c \rightarrow 0$ and $B_c \rightarrow 0$. To tackle this, *Sea-thru* relies on a distance-based alternative to the dark channel prior [17] to retrieve B_c and γ_c, and an illuminant map estimation [15] to retrieve β_c. $\boldsymbol{J}_c$ is then retrieved from Eq. 2 using these parameters:

$$\boldsymbol{J}_{c,p} = \left(\boldsymbol{I}_{c,p} - B_c(1 - e^{-\gamma_c \boldsymbol{z}_p})\right)e^{\beta_c \boldsymbol{z}_p}. \quad (3)$$

3. Method

The proposed approach, named SUCRe, overcomes the limitations of single-view underwater image color restoration methods by leveraging multiple observations of the scene, thus eliminating the need for additional assumptions. Our method takes as input undistorted underwater images together with their corresponding camera poses, intrinsics and depth maps. This information is retrieved using an off-the-shelf SfM pipeline [36]. By pairing pixels in multiple images, we are able to follow the intensity evolution of points at different distances and estimate the parameters

3

of the UIFM and pixel intensities at a hypothetical zero meter distance, implying a lack of disturbance by the water medium (see Figure 2).

SfM pipeline: As a preliminary step to our approach, we outline the procedure to obtain the inputs for SUCRe using SfM. We adopt the pipeline proposed by Sarlin *et al*. for visual localization, which is designed to be robust to significant changes in the environment [36]. NetVLAD [6] pairs similar images between which pixels are matched using SuperPoint [14] features and SuperGlue [37] matcher. The bundle adjustment is performed with COLMAP SfM [39] to recover the camera poses and intrinsic parameters. We then undistort the images using the estimated intrinsic parameters and build a 3D mesh of the scene. Finally, depth maps are obtained by ray-casting the images onto the 3D mesh. This results in depth maps less noisy than those obtained by multi-view stereo methods.

Dense multi-view pixel pairing: The first step of our approach is to pair pixels in a dense manner between different views. This is accomplished by projecting pixel coordinates from one view to another using the depth maps as well as the poses of the cameras and their intrinsics parameters. Let x_1 be the homogeneous coordinates of a pixel in image view i_1 with depth $d_1 \in \mathbb{R}$ and homogeneous pose matrix ${}^{w}\boldsymbol{T}_{i_1} \in SE(3)$, which represents the rigid 3D transformation between frame i_1 and frame w such as: $\lambda^{w} = {}^{w}\boldsymbol{T}_{i_1} \odot \lambda^{i_1}$, where $\lambda^{i_1} \in \mathbb{R}^3$, $\lambda^{w} \in \mathbb{R}^3$. Let $\boldsymbol{K} \in \mathbb{R}^{3 \times 3}$ be the intrinsic calibration matrix. The projection of x_1 in image view i_2 with pose ${}^{w}\boldsymbol{T}_{i_2}$ can be obtained by:

$$x_2 = \boldsymbol{K} \, {}^{i_2}\boldsymbol{T}_w \odot \left({}^{w}\boldsymbol{T}_{i_1} \odot \boldsymbol{K}^{-1} d_1 \, x_1 \right). \quad (4)$$

We then back-project x_2 in i_1 view using i_2 depth map:

$$x_1' = \boldsymbol{K} \, {}^{i_1}\boldsymbol{T}_w \odot \left({}^{w}\boldsymbol{T}_{i_2} \odot \boldsymbol{K}^{-1} d_2 \, x_2 \right), \quad (5)$$

where d_2 is the depth of x_2 in image view i_2. The pixels in both images (x_1, x_2) are only paired if x_1' and x_1 land on the same pixel coordinate, *i.e.*, pixels are matched to each other in both directions, from i_1 to i_2 and from i_2 to i_1. This ensures that each pixel has only one match in both images. It also filters out points occluded by the structure. This geometry-based approach allows us to robustly pair pixels in scenarios where feature matching algorithms fail, *e.g.*, in low contrast areas where most image signal has been attenuated like in the top left corner of Figure 3.

Optimization: With multiple observations of the same $\boldsymbol{J}_{c,p}$, our problem becomes well-posed. In SUCRe, we formulate the model described by Eq. 2 in a multi-view setting:

$$\boldsymbol{I}_{i,c,p} = \boldsymbol{J}_{c,p} e^{-\beta_c \boldsymbol{z}_{i,p}} + B_c (1 - e^{-\gamma_c \boldsymbol{z}_{i,p}}). \quad (6)$$
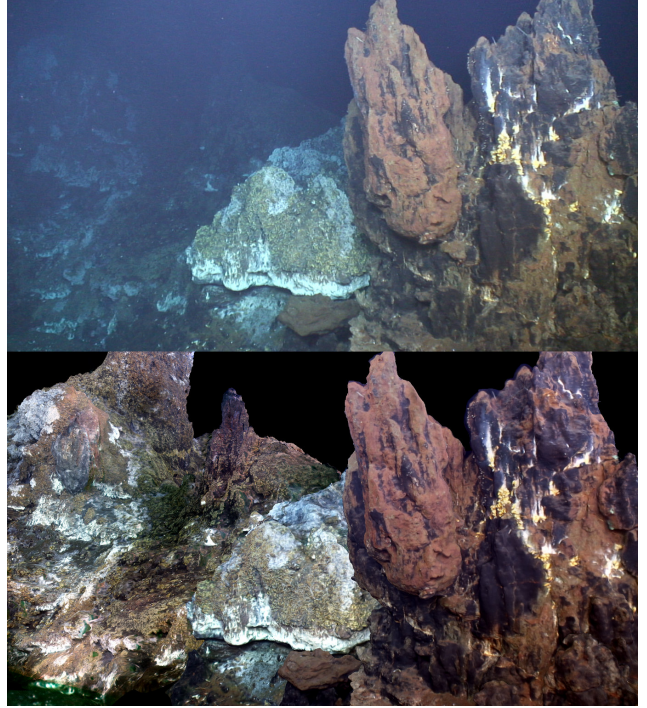

Figure 3: **Applying SUCRe on a deep-sea image** from the Eiffel Tower dataset [29] captured by a ROV equipped with an artificial lighting system, depicts the recovery of colors in low contrast areas (top left of the image). Pixels without depth information are left blank.

Parameters of Eq. 6 are then estimated by fitting the model in a least squares manner (see Appendix A for details):

$$\underset{\boldsymbol{J}_c, B_c, \beta_c, \gamma_c}{\arg\min} \sum_i \sum_p \| \boldsymbol{I}_{i,c,p} - \boldsymbol{J}_{c,p} e^{-\beta_c \boldsymbol{z}_{i,p}} $$
$$ - B_c (1 - e^{-\gamma_c \boldsymbol{z}_{i,p}}) \|^2. \quad (7)$$

Because some pixels in low contrast areas were matched with closer observations (see Figure 2), we are able to retrieve their color despite insufficient information about them on the image being restored (see Figure 3). Pixels with depth that have not been paired in other images contribute to the optimization as single observations.

4. Experiments

This work has been developed in the context of deep-sea applications. The image formation model, used in *Sea-thru* (Eq. 2) and expressed in a multi-view setting within SUCRe (Eq. 6), was conceived for natural light conditions. However, it mainly describes the effects of backscattering and color attenuation, which are phenomena shared by all underwater images regardless of the scenario. In this section, we demonstrate the effectiveness of our approach in both

Method	Varos				Sea-thru D5			
	PSNR ↑	SSIM ↑	UCIQE ↑	UIQM ↑	$\bar{\psi} \downarrow$	$\bar{\psi}$ std ↓	$\Delta E_{00} \downarrow$	ΔE_{00} std ↓
Underwater image	10.71	0.39	0.60	1.40	37.14	3.72	36.93	3.68
Fusion [4]	10.25	0.35	0.51	2.10	29.85	6.38	30.60	6.34
*Sea-thru** [2]	10.15	0.39	0.52	1.88	27.55	3.68	30.64	5.46
Water-Net [25]	11.20	0.38	0.54	1.96	29.12	4.11	31.49	5.89
FUnIE-GAN [19]	11.02	0.35	**0.62**	2.51	32.91	3.63	35.55	5.07
Haze-Lines [7]	9.64	0.36	0.57	2.00	25.80	7.14	28.85	6.89
TACL [28]	10.02	0.36	0.44	**2.52**	29.28	4.27	30.50	4.93
SUCRe (ours)	**12.13**	**0.42**	0.32	1.99	**21.45**	**2.63**	**22.56**	**2.84**

Table 2: **Restoration evaluation.** Our approach shows significant improvements across all full-reference metrics in comparison to other methods. Additionally, its achieves high performance independently of the color chart position, as indicated by the lower standard deviation on *Sea-thru* D5 errors.

natural light and deep-sea environments. Appendix B provides empirical examples that illustrate how the UIFM fits deep-sea images.

4.1. Datasets

First experiments are done in a controlled environment using the Varos synthetic underwater dataset [47]. It provides pairs of underwater and reference images that enable quantitative evaluation using objective measures. Another set of objective metrics are computed on *Sea-thru* D5 dataset [2]. Images of reference color charts captured under natural light in an underwater scene allow for measuring color rendition accuracy in a realistic scenario. Finally, we restore two ecologically valid datasets acquired before the development of the method. This showcases the real-life practical applicability of SUCRe for two sites of interest, *i.e.*, the Eiffel Tower hydrothermal vent [29] and a submarine wreck.

Varos is a synthetic dataset embedding 4,715 images that were rendered with Blender within a simulated underwater setting [47]. The dataset benefits from Blender's ray tracing technology, which enables the simulation of scattering and attenuation effects that are commonly observed in underwater scenes. A notable feature of Varos is its provision of reference images under uniform lighting conditions, which are useful for assessing the accuracy of color restoration techniques. These reference images offer a consistent baseline for comparison and facilitate the measurement of standard metrics.

Sea-thru D5 is composed of 43 raw images captured under natural light conditions, along with their corresponding distance maps obtained through SfM [2]. The scene contains four color charts with known patterns positioned throughout the scene. These charts serve as ground truth for computing metrics used to evaluate the performance of color restoration algorithms.

Eiffel Tower consists of a ROV dive to the Eiffel Tower hydrothermal chimney [29]. Due to the depth of the vent, which is around 1,700 meters, the ROV was equipped with an artificial lighting system since no light from the surface reaches such depths. The dataset comprises 4,875 images that were extracted from the ROV's video feed.

Submarine wreck comprises 4,595 images extracted from a ROV's video feed during a dive to a submarine wreck at a depth of around 1,150 meters. Similar to the Eiffel Tower dataset, the ROV was equipped with an artificial lighting system to illuminate the wreck due to the lack of natural light at such depths.

4.2. Implementation

Methods such as *Sea-thru* [2] have been designed to work on raw image files. However, in most real world applications, the images available have been archived in viewable formats with low dynamic range. Despite such sub-optimal conditions to perform color restoration, our approach is able to tackle this issue. Therefore, all experiments have been run on 8-bits images, which were processed from the raw images of the *Sea-thru* D5 dataset.

*Sea-thru**: Due to the lack of a publicly available implementation, the *Sea-thru* algorithm [2] was re-implemented to enable its comparison with our proposed method. We adapted the implementation to work on 8-bit images, and this revised version is referred to as *Sea-thru*.

(a) 3D model textured with underwater images
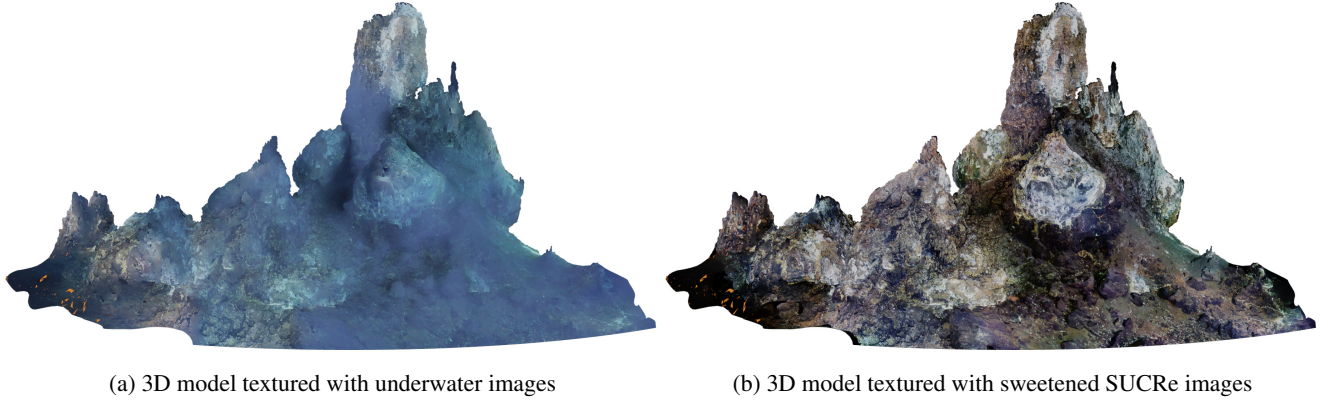

(b) 3D model textured with sweetened SUCRe images

Figure 4: **Texturing** the Eiffel Tower hydrothermal vent 3D model with images restored using our method results in a final model with improved visual quality, including finer details and more accurate colors compared to the original model.
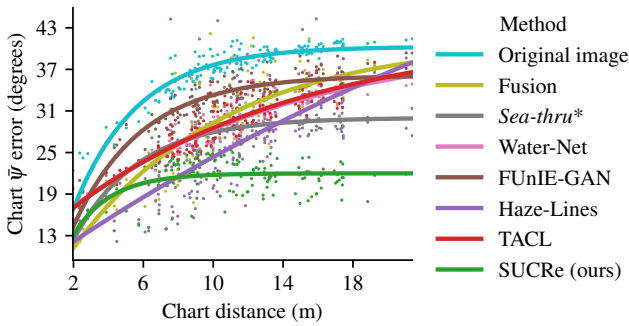

Figure 5: **Chart $\bar{\psi}$ error vs. distance.** Compared to other methods, SUCRe demonstrates lower and more consistent $\bar{\psi}$ errors independently of the distance of the color chart.
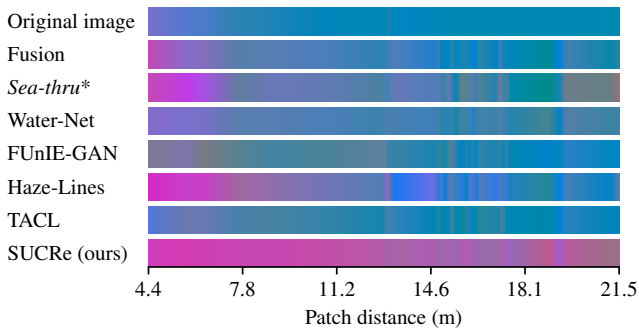
Figure 6: **Hue vs. distance.** Tracking the hue value of the magenta color patch at different distances on *Sea-thru* D5 dataset shows that our method outperforms other approaches in maintaining stable hue values.

Initialization: Solving the system described by Eq. 7 requires only a coarse initialization. In our experiments the following naive initial solution is sufficient: $J = I$ and $\beta = B = \gamma = 0.1$. Nevertheless, we may also use the parameters estimated with *Sea-thru** or other methods.

Optimization: To optimize the many parameters involved in jointly estimating the restored image along with the UIFM parameters, we use gradient descent with an Adam optimizer [23]. Each step of the gradient descent is computed using all matched observations by minimizing the function described by Eq. 7. Specifically, we perform 200 optimization steps with a learning rate of 0.05.

Normalization: Most color restoration methods need to perform white balancing as a final processing step. This is usually done using the Gray World Hypothesis algorithm [11]. Instead, we choose to apply a simple channel-wise histogram stretching.

Processing time: The full pipeline processing time depends on the number of images in the dataset and evolves linearly with the number of observations in SUCRe's optimization. Details are provided in Appendix C.

4.3. Quantitative evaluation

Evaluating the performance of underwater color restoration methods is a challenging task. Ground truth restored colors are generally unavailable for real-world underwater images. Nonetheless, synthetic datasets like Varos or images containing color charts like those included in the *Sea-thru* D5 dataset can be employed to provide reference values. This allows for the use of so called full-reference metrics [25]. We may also employ proposed no-reference metrics to assess other factors like contrast and saturation [33, 45]. However, recent literature has raised doubts about the ability of the latter to accurately measure the correction of physical phenomena such as attenuation and scattering [21, 25].

PSNR and SSIM are full-reference measures of image similarity that are particularly useful when entire ground truth restored images are available. UCIQE [45] and
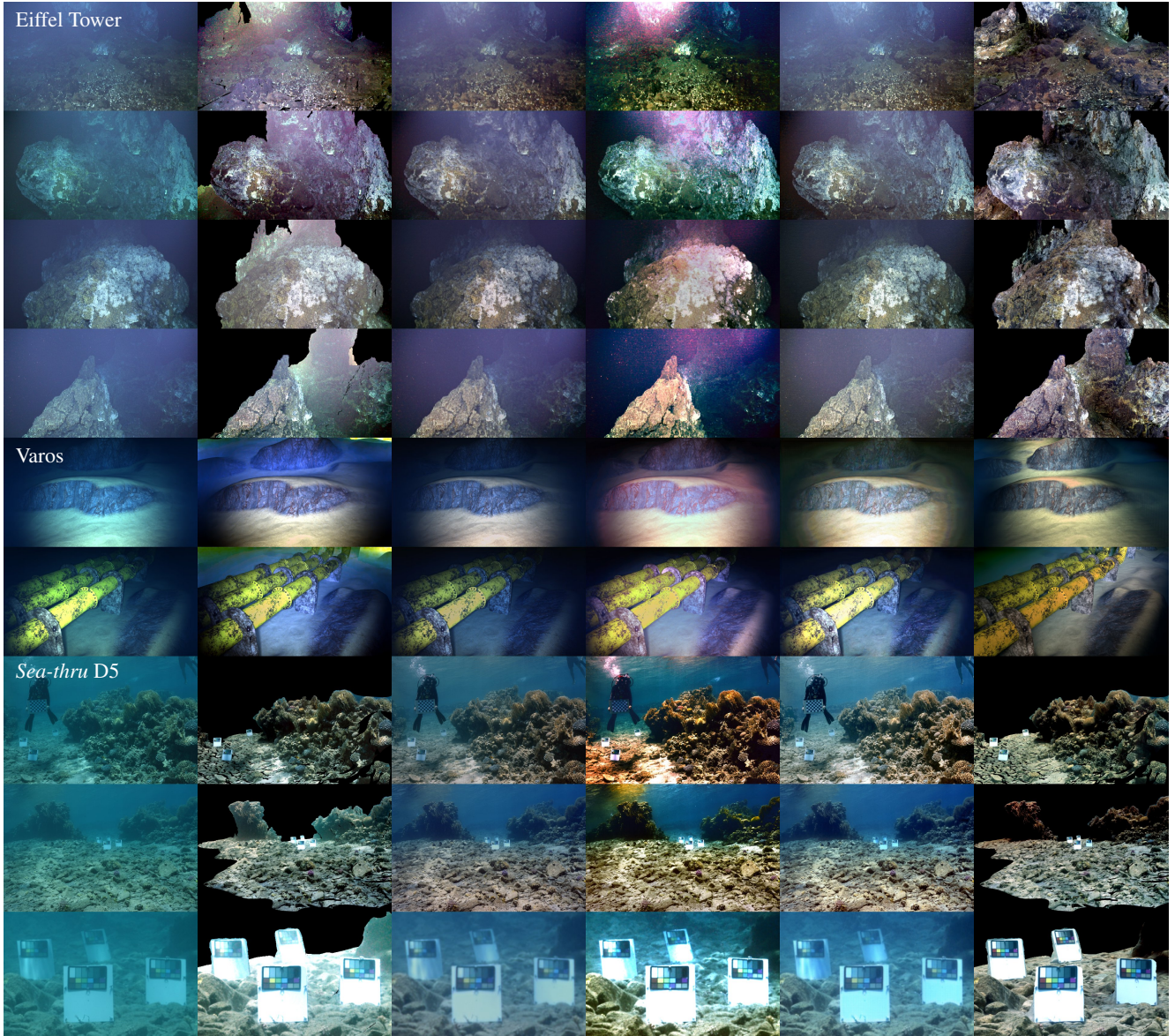
Figure 7: **Visual inspection.** Restoration results of different methods in diverse underwater scenarios. The proposed approach successfully recovers colors of distant elements where other methods struggle. Zooming in on the images is encouraged. As mentioned in Figure 3, pixels without depth information are rendered black for SUCRe and *Sea-thru**.

UIQM [33] are commonly used no-reference metrics for evaluating the visual quality of restored underwater images. The CIEDE2000 (ΔE_{00}) formula was developed by the International Commission on Illumination to evaluate color differences [41] and is commonly used in underwater color restoration [5, 24]. We hereby compute it between the restored and expected color patches of the *Sea-thru* D5 dataset. The $\bar{\psi}$ error was introduced by Berman *et al.* [7] and is designed specifically for images with color charts of

known colors distributed throughout the scene. For a given color chart in an image, the $\bar{\psi}$ error is defined as the average angle in RGB space between grayscale patches and a pure gray color. We redefine the error to take into account all twelve color patches in the DKK color calibration charts used in the *Sea-thru* D5 dataset:

$$\bar{\psi} = \frac{1}{12} \sum_{p \in P} \cos^{-1} \left(\frac{\boldsymbol{I}_p \cdot \boldsymbol{E}_p}{\|\boldsymbol{I}_p\| \cdot \|\boldsymbol{E}_p\|} \right), \qquad (8)$$

Parameters estimation		PSNR	SSIM
Single-view	Multi-view		
J, β, B, γ	—	10.15	0.39
β, B, γ	J	11.32	0.42
—	J, β, B, γ	12.13	0.42

Table 3: **Ablation study on Varos.** We show the benefits of leveraging multi-view observations for the estimation of the UIFM parameters and the restored image.

where P is a set containing pixel indices of the twelve color patches in the given color chart and E_p denotes the expected RGB values of the color patch with pixel index p.

Table 2 reports PSNR, SSIM, UCIQE [45] and UIQM [33] metrics on the Varos dataset, as well as $\bar{\psi}$ error in degrees [7] and CIEDE2000 (ΔE_{00}) color difference [41] on the *Sea-thru* D5 dataset. Results show that the proposed approach significantly outperforms other methods on every full-reference metric. Similarly to previous works [21, 25], we find that there is little correlation between PSNR/SSIM objective criteria and UCIQE/UIQM metrics. Furthermore, the low standard deviations on *Sea-thru* D5 errors show that SUCRe retains a high performance across color chart positions, a unique quality of our method that is highlighted in Figures 5 and 6.

4.4. Qualitative evaluation

As can be seen on Figures 2, 3 and 7, our method successfully restores the colors of far away elements (see last row of Figure 7). The use of multiple overlapping views overcomes the low intensities and partial loss of color information due to 8-bit quantization. In other words, our method fully exploits the virtual dynamic range augmentation offered by multiple observations of the color charts in different images. Additionally this ensures color consistency across elements of the scene at different distances from the sensor. Figure 4 illustrates that, when texturing a 3D mesh, restoring underwater images using our approach leads to significant improvements, including finer details as well as plausible and coherent colors.

4.5. Additional insights

Ablation study: Table 3 investigates the impact of using multiple views when estimating different parts of the complete model. The first row shows *Sea-thru** results, where both the UIFM parameters (β, B and γ) and the restored image J are estimated using a single image. In the second row, the UIFM parameters are fixed to those obtained with *Sea-thru**, yet the restored image is obtained using multi-view observations to minimize the same error as SUCRe (Eq. 7).

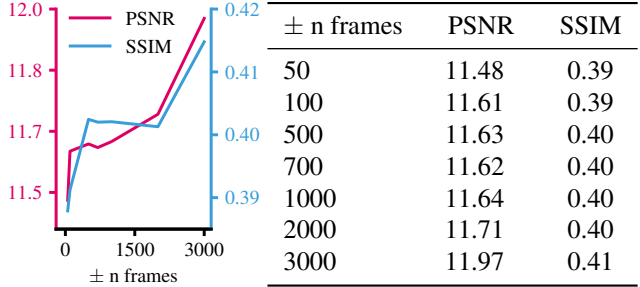| ± n frames	PSNR	SSIM
50	11.48	0.39
100	11.61	0.39
500	11.63	0.40
700	11.62	0.40
1000	11.64	0.40
2000	11.71	0.40
3000	11.97	0.41

Table 4: **Degradation study on Varos.** We evaluate the impact of the number of sequential input frames on the performance of our approach. As the Varos dataset has a recording rate of 10fps, using ± 50 frames as input is equivalent to using 100 frames, or 10 seconds of video.

The last row shows SUCRe results, where all parameters are estimated in a multi-view setting. Results show that SSIM improves mainly by the recovery of low contrast areas when J is estimated using multiple views. While PSNR values indicate that using multi-view observations for estimating both the UIFM parameters and the restored image is required to obtain the peak performance exhibited by our approach.

Degradation study: As the underwater vehicle explores the seabed, it captures more images and acquires additional information that was not available in previous images. Table 4 shows that our approach is able to consistently make use of this information to improve restoration results, yielding higher PSNR and SSIM values when presented with more candidate matching images.

5. Conclusion

In this paper, we have introduced SUCRe, an underwater image color restoration approach that makes use of multi-view observations to simultaneously estimate the parameters of an underwater image formation model along with the restored image. We have demonstrated that the use of overlapping views allows resolving the colors of pixels that are barely visible in the target original image. It also leads to more accurate and consistent color renditions of elements at different distances from the sensor. Results on deep-sea and natural light datasets show that color restoration has much to benefit from multi-view observations of the same pixels. Furthermore the results of the proposed approach could be used as targets to train single-view deep-learning methods. We believe this method is a step forward towards helping us view the world under the surface in the same way we see the one above.

Appendix

The following pages provide additional details on the experiments conducted in the main paper. Appendix A presents the residual analysis which justifies the use of a least squares estimator. Appendix B illustrates how the underwater image formation model used in the paper fits deep-sea images. Appendix C shows the linear trend in processing time with respect to number of observations of the reference implementation of our approach. Appendix D describes the accompanying videos. Appendix E provides additional plots similar to Figure 6, tracking the hue values of yellow and orange color patches at different distances.

A. Least squares

Assuming that the errors between our model and measured individual per-channel pixel intensities follow a zero-centered Normal distribution, the least squares estimator leads to the maximum likelihood solution. This is the technique used in SUCRe to obtain parameters of Eq. 6. Figures 10 to 12 serve to analyze the fitting residuals on a real-world deep-sea image (Figure 3). The deviation from normality of the residuals remains small, which supports the choice of model and estimator.

B. Application to deep-sea images

Figure 8 shows how the underwater image formation model described by Eq. 6 fits a deep-sea image (Figure 3). The plot illustrates that the estimated model accurately captures the evolution of observed pixel intensities with distance. This suggests that the main components described by the model, *i.e.*, backscattering and color attenuation, adequately capture the main phenomena affecting pixel intensity changes in our specific application.

C. Processing time

The processing time of our approach can be divided into two main components: *i)* pairing pixels between the image to be restored and every other images; *ii)* the optimization procedure described by Eq. 7. The pairing step depends on the size of the images and the number of candidate matching images in the dataset. Figure 9 illustrates that the optimization step evolves linearly with the number of matched observations. To restore an image from a dataset comprising 4,875 images with a resolution of 1920 by 1080 pixels, our approach takes about 50 seconds to compute pixel pairs and 1 minute and 40 seconds for the optimization procedure using 100 millions matched pixel observations, for a total of 2 minutes and 30 seconds processing time. All computations were performed with 32 CPU threads and a RTX A5000 GPU.
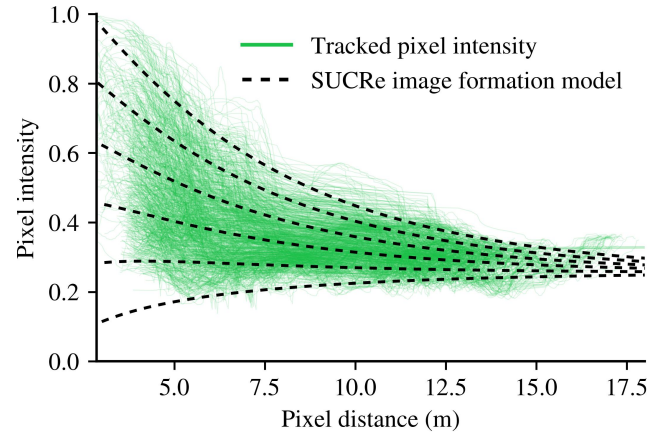

Figure 8: **Fitting the underwater image formation model on a deep-sea image.** Each green curve represents one point observed in multiple images at different distances (curves have been smoothed for visualization purposes). To illustrate how the underwater image formation model used in SUCRe fits these intensities, the black dotted lines show how different initial pixel intensities evolve with distance according to the estimated model.
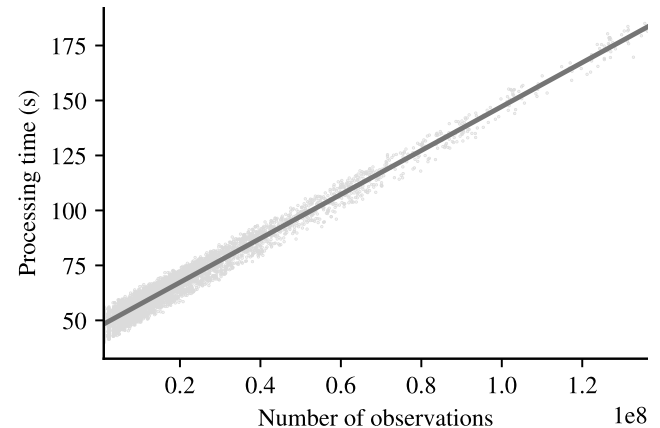

Figure 9: **Processing time vs. number of observations.** Scatter plot illustrating the relation between the number of observations in an image and the processing time of restoring it. Each point corresponds to an image of the Eiffel Tower dataset. The number of observations is the total number of pixels that have been paired to another pixel in the processed image. The processing time of the optimization procedure follows a very consistent linear increase with the number of observations. The intercept of approximately 50 seconds is due to the pairing step that computes pixel pairs for all 4,875 images in the dataset.

(a) Residual errors on red channel   (b) Residual errors on green channel   (c) Residual errors on blue channel
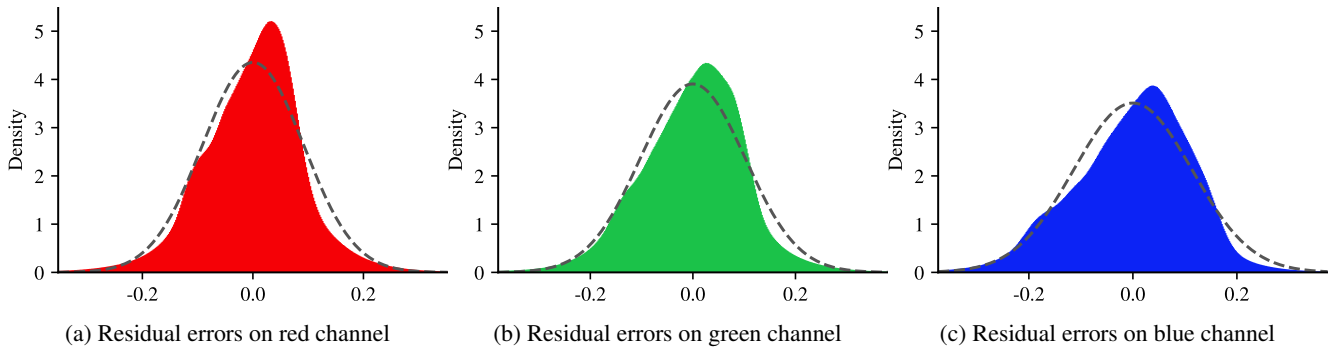
Figure 10: **Residual errors distributions.** Histograms of residual errors on a deep-sea image (Figure 3) illustrate that their distribution is close to a Normal.

(a) Normal plot on red channel   (b) Normal plot on green channel   (c) Normal plot on blue channel
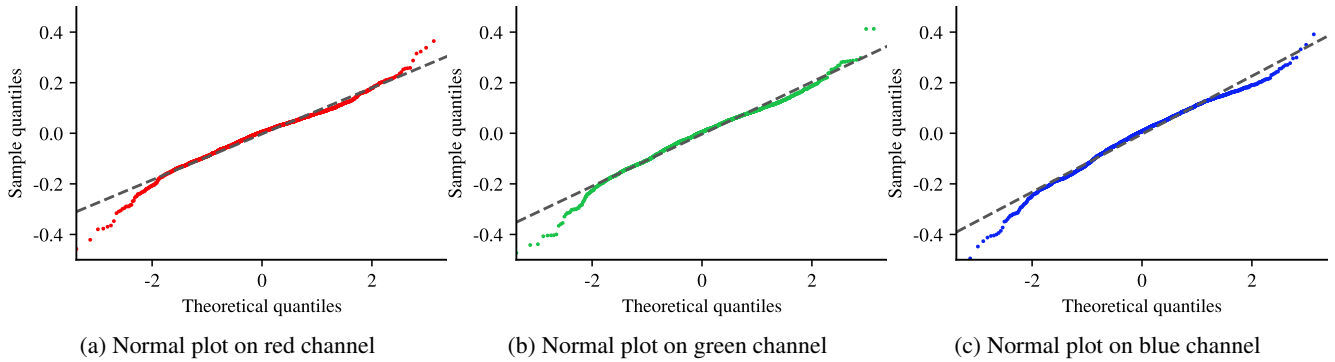
Figure 11: **Normal probability plots.** The x-axis refers to the expected quantiles of a normal distribution, while the y-axis corresponds to the actual quantiles of the residuals. Points following a straight line suggest that the data follows a Normal distribution. Plots on all three channels indicate that the residuals are close to being normally distributed.

(a) Residuals vs. fits on red channel   (b) Residuals vs. fits on green channel   (c) Residuals vs. fits on blue channel
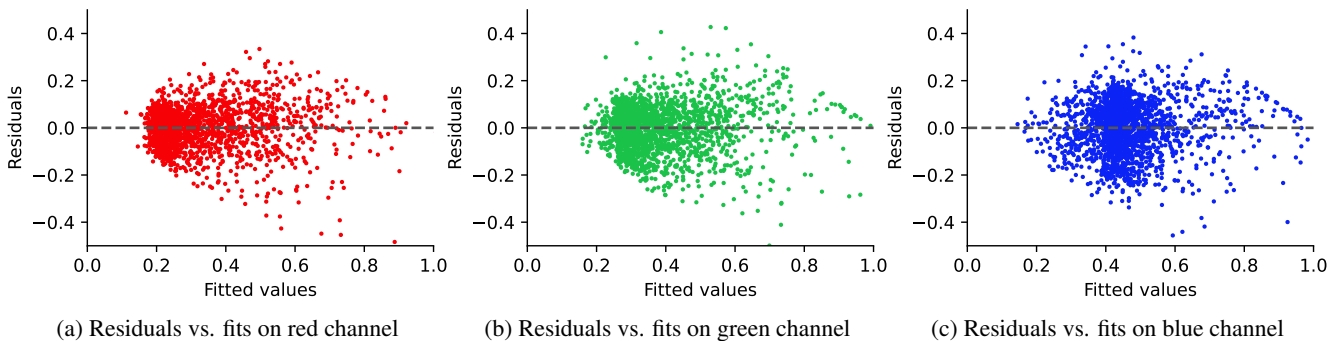
Figure 12: **Residuals vs. fits.** These plots are used to identify patterns in the residuals that may indicate problems with the model, such as non-linearity or heteroscedasticity. A well-fitting model should have residuals randomly scattered around zero independently of the fitted value, which is the case on all three channels.

D. Videos

The accompanying videos show that our approach yields consistent colors independently of the distance between the camera and the scene. Moreover, the videos show that Haze-Lines [7] is not well-suited for underwater color restoration of video feeds because it relies on a per-frame estimation of Jerlov water types [43]. Consecutive changes in the water category selection leads to flickering images in the video. While Water-Net performs a consistent restoration across frames in the video, backscattering and color attenuation remain highly visible and low contrast areas cannot be resolved without exploiting multiple views of the scene.
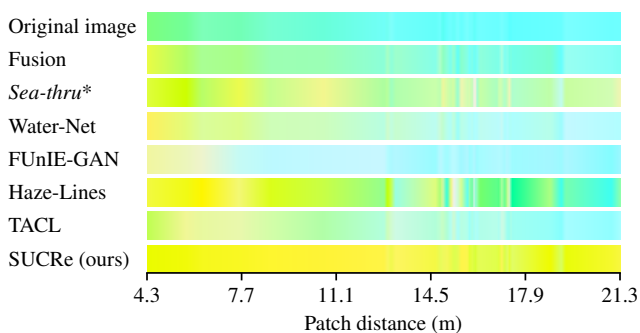
E. Tracking hue values

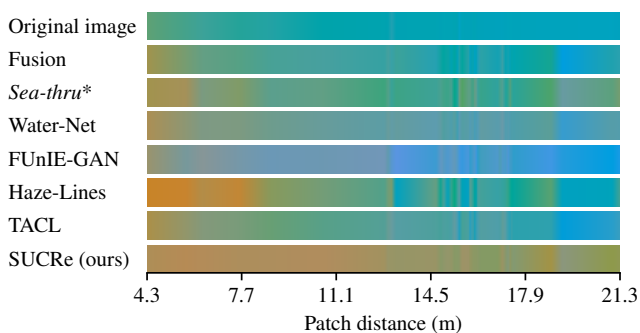Figure 13: **Hue vs. distance** on yellow color patch.

Figure 14: **Hue vs. distance** on orange color patch.

Similar to Figure 6, Figures 13 and 14 show the restored hue values of two other color patches whose appearance is greatly affected under water.

References

[1] Derya Akkaynak and Tali Treibitz. A revised underwater image formation model. In *CVPR*, pages 6723–6732, June 2018. 1, 2, 3

[2] Derya Akkaynak and Tali Treibitz. Sea-Thru: A method for removing water from underwater images. In *CVPR*, pages 1682–1691, June 2019. 2, 3, 5, 7

[3] Derya Akkaynak, Tali Treibitz, Tom Shlesinger, Yossi Loya, Raz Tamir, and David Iluz. What is the space of attenuation coefficients in underwater computer vision? In *CVPR*, pages 4931–4940, July 2017. 1, 2, 3

[4] Cosmin Ancuti, Codruta Orniana Ancuti, Tom Haber, and Philippe Bekaert. Enhancing underwater images and videos by fusion. In *CVPR*, pages 81–88, 2012. 1, 5

[5] Codruta O. Ancuti, Cosmin Ancuti, Christophe De Vleeschouwer, and Rafael Garcia. Locally adaptive color correction for underwater image dehazing and matching. In *CVPR Workshops*, pages 997–1005, July 2017. 1, 7

[6] Relja Arandjelovic, Petr Gronat, Akihiko Torii, Tomas Pajdla, and Josef Sivic. Netvlad: Cnn architecture for weakly supervised place recognition. In *CVPR*, pages 5297–5307, June 2016. 4

[7] Dana Berman, Deborah Levy, Shai Avidan, and Tali Treibitz. Underwater single image color restoration using haze-lines and a new quantitative dataset. *TPAMI*, 43(8):2822–2837, 2021. 1, 2, 3, 5, 7, 8, 11

[8] Dana Berman, Tali Treibitz, and Shai Avidan. Color restoration of underwater images. In *BMVC*, pages 44.1–44.12. BMVA Press, September 2017. 2, 3

[9] Mitch Bryson, Matthew Johnson-Roberson, Oscar Pizarro, and Stefan Williams. Colour-consistent structure-from-motion models using underwater imagery. In *RSS*. Robotics: Science and Systems Foundation, July 2012. 2

[10] Mitch Bryson, Matthew Johnson-Roberson, Oscar Pizarro, and Stefan B. Williams. True color correction of autonomous underwater vehicle imagery. *JFR*, 33(6):853–874, October 2015. 2

[11] G. Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310(1):1–26, 1980. 6

[12] Ricard Campos, Rafael Garcia, Pierre Alliez, and Mariette Yvinec. A surface reconstruction method for in-detail underwater 3d optical mapping. *IJRR*, 34(1):64–89, 2015. 2

[13] John Y. Chiang and Ying-Ching Chen. Underwater image enhancement by wavelength compensation and dehazing. *TIP*, 21(4):1756–1769, 2012. 2, 3

[14] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *CVPR Workshops*, pages 224–236, June 2018. 4

[15] Marc Ebner and Johannes Hansen. Depth map color constancy. *Bio-Algorithms and Med-Systems*, 9(4):167–177, 2013. 3

[16] Maxime Ferrera, Vincent Creuze, Julien Moras, and Pauline Trouvé-Peloux. AQUALOC: An underwater dataset for visual–inertial–pressure localization. *IJRR*, 38(14):1549–1559, 2019. 2

[17] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *TPAMI*, 33(12):2341–2353, 2010. 2, 3

[18] Guojia Hou, Jingming Li, Guodong Wang, Huan Yang, Baoxiang Huang, and Zhenkuan Pan. A novel dark chan-

nel prior guided variational framework for underwater image restoration. *Journal of Visual Communication and Image Representation*, 66:102732, 2020. 2

[19] Md Jahidul Islam, Youya Xia, and Junaed Sattar. Fast underwater image enhancement for improved visual perception. *RA-L*, 5(2):3227–3234, 2020. 5

[20] Jules Jaffe. Computer modeling and the design of optimal underwater imaging systems. *IEEE Journal of Oceanic Engineering*, 15(2):101–111, 1990. 1, 2

[21] Qiuping Jiang, Yuese Gu, Chongyi Li, Runmin Cong, and Feng Shao. Underwater image enhancement quality evaluation: Benchmark dataset and objective metric. *TCSVT*, 32(9):5959–5974, 2022. 6, 8

[22] Matthew Johnson-Roberson, Oscar Pizarro, Stefan B. Williams, and Ian Mahon. Generation and visualization of large-scale three-dimensional reconstructions from underwater robotic surveys. *JFR*, 27(1):21–51, 2010. 2

[23] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR (Poster)*, 2015. 6

[24] Chongyi Li, Saeed Anwar, Junhui Hou, Runmin Cong, Chunle Guo, and Wenqi Ren. Underwater image enhancement via medium transmission-guided multi-color space embedding. *TIP*, 30:4985–5000, 2021. 7

[25] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *TIP*, 29:4376–4389, 2020. 1, 2, 5, 6, 7, 8

[26] Chongyi Li, Jichang Quo, Yanwei Pang, Shanji Chen, and Jian Wang. Single underwater image restoration by blue-green channels dehazing and red channel correction. In *ICASSP*, pages 1731–1735, 2016. 1, 2

[27] Jie Li, Katherine A. Skinner, Ryan M. Eustice, and Matthew Johnson-Roberson. Watergan: Unsupervised generative network to enable real-time color correction of monocular underwater images. *RA-L*, 3(1):387–394, 2018. 1, 2

[28] Risheng Liu, Zhiying Jiang, Shuzhou Yang, and Xin Fan. Twin adversarial contrastive learning for underwater image enhancement and beyond. *TIP*, 31:4922–4936, 2022. 5, 7

[29] Marjolaine Matabos and Aurelien Arnaubec. Eiffel tower hydrothermal chimney (lucky srike hydrothermal field, mid atlantic ridge): 3D scene and imagery, 2015. 2, 4, 5

[30] B. L. McGlamery. A Computer Model For Underwater Camera Systems. In Seibert Quimby Duntley, editor, *Ocean Optics VI*, volume 208, pages 221–231. International Society for Optics and Photonics, SPIE, 1980. 1, 2

[31] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020. 2

[32] David Nakath, Mengkun She, Yifan Song, and Kevin Köser. In-situ joint light and medium estimation for underwater color restoration. In *ICCV Workshops*, pages 3724–3733, 2021. 1, 2, 3

[33] Karen Panetta, Chen Gao, and Sos Agaian. Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, 41(3):541–551, 2016. 6, 7, 8

[34] Arpit Pipara, Urvi Oza, and Srimanta Mandal. Underwater image color correction using ensemble colorization network.

In *ICCV Workshops*, pages 2011–2020, October 2021. 2

[35] Oscar Pizarro, Ryan Michael Eustice, and Hanumant Singh. Large area 3-d reconstructions from underwater optical surveys. *IEEE Journal of Oceanic Engineering*, 34(2):150–169, 2009. 2

[36] Paul-Edouard Sarlin, Cesar Cadena, Roland Siegwart, and Marcin Dymczyk. From coarse to fine: Robust hierarchical localization at large scale. In *CVPR*, pages 12716–12725, June 2019. 2, 3, 4

[37] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperGlue: Learning feature matching with graph neural networks. In *CVPR*, pages 4938–4947, June 2020. 4

[38] Yoav Schechner and Nir Karpel. Recovery of underwater visibility and structure by polarization analysis. *IEEE Journal of Oceanic Engineering*, 30(3):570–587, 2005. 2, 3

[39] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *CVPR*, pages 4104–4113, June 2016. 4

[40] Advaith Venkatramanan Sethuraman, Manikandasriram Srinivasan Ramanagopal, and Katherine A. Skinner. Waternerf: Neural radiance fields for underwater scenes, 2022. 2

[41] Gaurav Sharma, Wencheng Wu, and Edul N. Dalal. The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application*, 30(1):21–30, 2005. 7, 8

[42] Katherine A. Skinner, Eduardo Iscar, and Matthew Johnson-Roberson. Automatic color correction for 3d reconstruction of underwater scenes. In *ICRA*, pages 5140–5147, 2017. 2

[43] Michael G. Solonenko and Curtis D. Mobley. Inherent optical properties of jerlov water types. *Applied Optics*, 54(17):5392–5401, June 2015. 1, 11

[44] Yifan Song, David Nakath, Mengkun She, and Kevin Köser. Optical imaging and image restoration techniques for deep ocean mapping: A comprehensive survey. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 90(3):243–267, April 2022. 3

[45] Miao Yang and Arcot Sowmya. An underwater color image quality evaluation metric. *TIP*, 24(12):6062–6071, 2015. 6, 8

[46] Jingchun Zhou, Tongyu Yang, Wenqi Ren, Dan Zhang, and Weishi Zhang. Underwater image restoration via depth map and illumination estimation based on a single image. *Optics Express*, 29(19):29864–29886, September 2021. 2

[47] Peder Georg Olofsson Zwilgmeyer, Mauhing Yip, Andreas Langeland Teigen, Rudolf Mester, and Annette Stahl. The varos synthetic underwater data set: Towards realistic multi-sensor underwater data with ground truth. In *ICCV Workshops*, pages 3722–3730, October 2021. 1, 2, 5