VIRTUALLY THROWING BENCHMARKS INTO THE OCEAN FOR DEEP SEA PHOTOGRAMMETRY AND IMAGE PROCESSING EVALUATION

Yifan Song[1,*], Mengkun She[1], Kevin Köser[1]

[1] Oceanic Machine Vision, GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany
{ysong; mshe; kkoeser}@geomar.de

KEY WORDS: Deep Sea Image, Underwater Photogrammetry, Underwater Image Processing, Synthetic Image Dataset, Underwater Image Formation.

ABSTRACT:

Vision in the deep sea is acquiring increasing interest from many fields as the deep seafloor represents the largest surface portion on Earth. Unlike common shallow underwater imaging, deep sea imaging requires artificial lighting to illuminate the scene in perpetual darkness. Deep sea images suffer from degradation caused by scattering, attenuation and effects of artificial light sources and have a very different appearance to images in shallow water or on land. This impairs transferring current vision methods to deep sea applications. Development of adequate algorithms requires some data with ground truth in order to evaluate the methods. However, it is practically impossible to capture a deep sea scene also without water or artificial lighting effects. This situation impairs progress in deep sea vision research, where already synthesized images with ground truth could be a good solution. Most current methods either render a virtual 3D model, or use atmospheric image formation models to convert real world scenes to appear as in shallow water appearance illuminated by sunlight. Currently, there is a lack of image datasets dedicated to deep sea vision evaluation. This paper introduces a pipeline to synthesize deep sea images using existing real world RGB-D benchmarks, and exemplarily generates the deep sea twin datasets for the well known Middlebury stereo benchmarks. They can be used both for testing underwater stereo matching methods and for training and evaluating underwater image processing algorithms. This work aims towards establishing an image benchmark, which is intended particularly for deep sea vision developments.

1. INTRODUCTION

The open ocean, as the largest living space on our planet, covers more than half of Earth's surface with more than 1000m of water. The deep sea is characterized by extremely high pressure and permanent darkness, and remains largely unexplored by humans. Vision systems have been widely applied in ocean exploration missions, which is being increasingly adopted also for deep ocean research. To develop computer vision algorithms for classical applications, benchmarks are often used to evaluate the performance of the methods. Running algorithms on a benchmark with ground truth (GT) allows us to directly compare the performance between different methods. Good vision benchmarks can lead to a boost of the methods' development in the corresponding area, such as ImageNet (Deng et al., 2009) to visual object recognition, the Middlebury datasets (Scharstein et al., 2014) to stereo vision, and KITTI (Geiger et al., 2012) to autonomous driving. Unfortunately, we are lacking such datasets specifically developed for deep sea vision algorithm development. Since the image degradation due to water-based absorption, scattering and light cones are the particular extra challenges for deep sea data, GT data should include also how the images would look without water (as also sought in image restoration algorithms). Simultaneously obtaining a deep sea scene and its corresponding GT appearance without water or lighting effects is practically impossible. The lack of proper deep sea vision datasets with GT is impairing the development of corresponding vision methods (Song et al., 2022).

Synthesized datasets could be a solution to this problem. They must be valid and verifiable, e.g. using physically based techniques to generate realistic deep sea images and utilize existing models or benchmarks to create the GT. There are two types of solutions to simulate the deep sea images. One is converting existing color and depth (RGB-D) images into the deep sea scenario and generate their deep sea twin images (see Fig. 1). Existing benchmarks outside the ocean, e.g. for dense stereo matching, already provide in-air images, as well as the GT depth (or disparity) (Scharstein et al., 2014), to evaluate the results that could be used also to evaluate simulated images with different appearance. However, the GT depth maps of the real world scene often contain missing areas that require additional pre-processing before generating the underwater data.

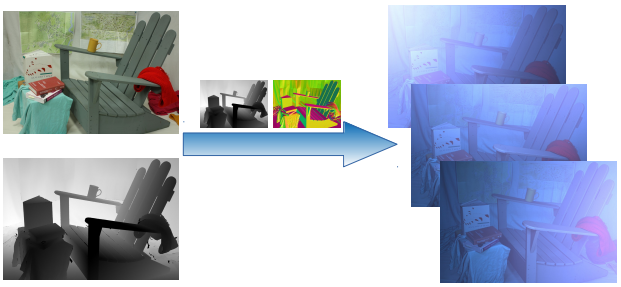

Figure 1. Synthesizing deep sea images with GT for existing real world in-air benchmarks.

The other technical solution is directly rendering deep sea effects from a fine-textured 3D model (Sedlazeck and Koch, 2011, Song et al., 2021a, Zwilgmeyer et al., 2021). Both, the deep sea images, as well as their in-air appearances, can be simulated using the same camera parameters but different lighting

* Corresponding author

and medium configurations, possibly using different rendering pipelines. Simulating from a 3D model using advanced rendering techniques allows to provide perfect pose information and depth images. Nevertheless, they do not show a real world scene, and the rendered images' quality mostly depends on the quality of the 3D model. Existing methods all synthesize deep sea appearance images from virtual scenes. To our best knowledge, there is currently no literature to discuss the synthesis issues of real world scenes in a physically realistic deep sea scenario. To bridge the gaps between real world in-air scenes and deep sea images, we introduce a complete pipeline to synthesize deep sea images with GT from existing in-air benchmarks. Exemplarily, we create deep sea image twin datasets for the Middlebury stereo benchmark that could later become part of a Deep Sea Vision (DSV) dataset, particularly intended for deep sea underwater photogrammetry and image processing evaluation. It consists of high-resolution deep sea stereo image pairs with GT disparities for underwater stereo matching evaluation or for training. Moreover, the in-air/underwater twin color images of the original benchmark can be used as the GT for developing underwater image restoration approaches.

2. RELATED WORK

The first report of filming underwater images dates back to the 19th century (according to (Jaffe, 2014)), and nowadays photogrammetry has been successfully applied to many aspects of shallow underwater research (Bythell et al., 2001, Drap, 2012, Menna et al., 2013). Most of such images are acquired via scuba divers and their operating depth is limited to a few tens of meters. Imaging in the deep sea started much later as it was facing many physical and technological barriers. The first attempt of deep sea imaging was deployed during the second world war (Harvey, 1939). Thanks to the advancement of technologies, nowadays deep sea photogrammetry via robotic platforms becomes applicable and is increasingly being used in deep ocean research (Pizarro and Singh, 2003, Kwasnitschka et al., 2016).

On land, good benchmarks and test data played important roles in the development of vision methods, as improving and validating algorithms requires to continuously evaluate the results and performances. Several datasets with GT were captured under different scenarios for different applications and evaluations. A famous in-door scene 3D vision benchmark is the Middlebury dataset, which contains Stereo (Scharstein et al., 2014), Multi-View Stereo (MVS) (Seitz et al., 2006), and Optical Flow (Baker et al., 2011) datasets. It provides the official page for evaluating the submitted results, which has been widely used in 3D vision and photogrammetry research. Another well known 3D vision benchmark is the ETH3D dataset. It includes MVS (Schops et al., 2017) and simultaneous localization and mapping (SLAM) (Schops et al., 2019) benchmarks. Similar benchmarks exist for dense matching in airborne photogrammetry such as the ISPRS/EuroSDR (Nex et al., 2015) and the Hessigheim 3D (H3D) (Kölle et al., 2021) datasets. The KITTI (Geiger et al., 2012) dataset, is a benchmark currently often used for vision in autonomous driving. It provides various sensor measurements including stereo images in the urban region with GT trajectories, and is utilized for evaluating Visual Odometry and SLAM methods. Besides the KITTI datasets there are several others as well, e.g. the Málaga Urban dataset (Blanco-Claraco et al., 2014) also contributes stereo images and light detection and ranging (LiDAR) measurements in urban scenario for SLAM in autonomous driving. Apart from that, the EuRoC micro aerial vehicle dataset (Burri et al., 2016) provides the images sets with GT poses and a detailed 3D scan of the in-door environment for visual-inertial SLAM. The TUM RGB-D SLAM dataset (Sturm et al., 2012) captured RGB-D images through Microsoft Kinect, with given GT trajectories. Even though various 3D vision datasets are available, they are often still not sufficient for training neural networks with their need for huge amounts of training data and diversity in the data. Here, synthetic rendering of data could help, and many large-scale synthetic datasets with GT exist for different purposes. The MPI Sintel Flow dataset (Butler et al., 2012) derived the optical flow GT from an open source 3D animation *Sintel*. The FlyingThings3D (Dosovitskiy et al., 2015) is another synthetic optical flow dataset which consists of renderings of 3D chairs with different poses and backgrounds. The Monkaa (Mayer et al., 2016) dataset rendered stereo frames with GT from the animated short film *Monkaa*, including optical flow, disparity and disparity change. Many of the above mentioned datasets are later included in the recent Robust Vision Challenge 2020 [1].

Compared to the abundant in-air vision benchmarks, there are only limited underwater vision datasets available to the public for evaluation purposes, due to the inherent difficulties of deploying well-controlled experiments. In underwater robotics, (Mallios et al., 2017) proposed a dataset with imagery collected by an autonomous underwater vehicle (AUV) in underwater caves. (Ferrera et al., 2019) released another real ocean Visual SLAM dataset AQUALOC, captured by a remotely operated vehicle (ROV). In underwater image processing, (Li et al., 2019) collected images from the Internet and presented a real-world underwater image enhancement benchmark (UIEB). The enhanced reference images are manually selected by human inspection among 12 enhanced results. (Akkaynak and Treibitz, 2019, Berman et al., 2020) utilized an underwater stereo camera system and captured in total 57 stereo pairs for underwater restoration evaluation, the reference distances are computed via Structure from motion (SfM). Since water blocks the use of GPS underwater, the real world robotic vision datasets share the same problem: that the GT trajectories and distances are not precise enough for evaluating underwater Visual SLAM. Also, enhanced images used as reference images are still not equal to the medium-free images. As the demand of high-accuracy underwater vision is increasing, the real-world underwater evaluation datasets are by far not sufficient. To our best knowledge, there is still no such dataset available for deep sea vision evaluation yet. Synthesized datasets seem to offer a solution.

To physically simulate underwater images, the physics of light travelling through water, including the attenuation and scattering, has been properly studied (Mobley, 1994). The most frequently used underwater image formation model is inherited from the atmospheric fog model (AFM) (Cozman and Krotkov, 1997), which assumes a global illumination from the water surface. Another famous model is the Jaffe-McGlamery (J-M) model (Jaffe, 1990, McGlamery, 1980), which considers point light sources. It is regularly applied in photometric stereo approaches in participating media. See (Song et al., 2022) for a recent survey on the different models.

Synthesized images are widely applied in learning based underwater research. Most of them use the AFM to convert RGB-D images to underwater scenario as the training data (Li et al., 2020, Li et al., 2017, Ueda et al., 2019). Meanwhile, some underwater vehicle simulators, such as the Unmanned Underwater Vehicle (UUV) Simulator (Manhães et al., 2016), also apply

[1] `http://www.robustvision.net/index.php`

this model for generating camera outputs due to its simple implementation. However, (Song et al., 2021a) addressed that the AFM is not able to imitate the complex artificial lighting effect in the deep sea scenario, to which the J-M model better suits. (Sedlazeck and Koch, 2011) adapted the J-M model to simulate deep sea underwater images. Later (Song et al., 2021a) extended the model for multi directional light sources and optimized the rendering strategy for faster implementation. Besides physically based image formation models, the ray-tracing techniques such as volumetric rendering, are also used for synthesizing light transportation in media (Bitterli and Jarosz, 2017, Crane et al., 2007, Novák et al., 2018). (Zwilgmeyer et al., 2021) applied the latest Monte-Carlo path-tracing rendering from the rendering engine Blender (Blender Online Community, 2021) to generate the underwater dataset VAROS. In order to evaluate refraction induced by underwater housings (Nakath et al., 2022) have introduced the GEODT toolkit based on blender.

3. APPROACH

This section describes the details of using the real word RGB-D images for synthesizing their deep sea twins. The raw disparity data require careful pre-processing in order to be compatible with the rendering procedure. Firstly, the disparities need to be converted to real world depths for physical rendering. Secondly, the converted depth map requires further refinement to fill the incomplete data. Thirdly, we have to estimate initial normals from the refined depth map and create a mask for depth discontinuity regions, where normals are uncertain. Lastly, the masked normals are smoothed by using a median filter. The complete pre-processing pipeline is illustrated in Fig. 2 and explained in Section 3.1.

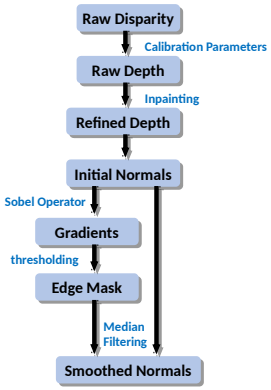

Figure 2. Workflow of pre-preprocessing real world depth

3.1 Scene Preparation and Pre-processing

Stereo benchmarks provide disparity maps for directly evaluating dense image matching performance. The disparity value refers to the pixel coordinate difference between two corresponding points in the stereo image pair. It cannot be directly used for physical model based simulation as this requires the knowledge of the scene depth in the real world scale. According to the Middlebury dataset descriptions, the disparity value d [pixel] can be converted to the real depth Z [m] by the camera calibration parameters:

$$Z = \frac{\mathrm{b} \cdot f}{1000 \cdot (d + d_{\mathrm{offs}})}, \tag{1}$$

where b represents the camera baseline [mm], f the pinhole's focal length [pixels] and d_{offs} is the horizontal difference of the principal points.
Real world raw depth maps usually contain empty values where no depth information has been recorded (See Fig. 3 left). They are often captured by external high resolution devices, e.g. structured light systems (Scharstein and Szeliski, 2003). In this case, the offset between the camera and the infrared emitter can introduce a stereo shadow and a specular object surface can cause missing data. These incomplete depth maps can not be directly used for deep sea rendering as the rendering relies on distance information per pixel and a "hole" would create strong artifacts in the renderings. To avoid such artifacts, the depth maps have to be filled, where the filled values are only used for creating the underwater image appearance, but do not serve as GT in the evaluation. Several approaches have been proposed for filling missing depth values. In this paper, we adopt the inpainting method from (Bertalmio et al., 2001), which is based on the Navier-Stokes equations, for refining depth maps.

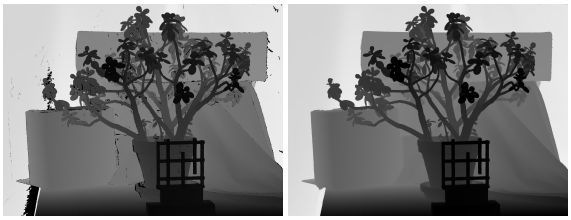

Figure 3. Left: The original raw depth map contains incomplete data. Right: Refined depth map using inpainting.

Besides incomplete depth values, the real world benchmarks often contain complex scene geometry for evaluation purposes. This creates depth discontinuities that have to be carefully handled when computing the surface normals from the depth map with local operators. Here, depth discontinuities can cause wrong normal computation results for rendering, which can lead to obvious dark contours around objects in the image under the standard Lambertian reflection. In order to avoid such artifacts, normals facing away from the camera are median-filtered. Additionally, we adopt also the ambient term from the Phong reflection model (will be discussed in Section 3.2.2) in our image formation model, which essentially resembles the scattered light that is present in the scene. The computation of normal maps from the depth contains the following steps:

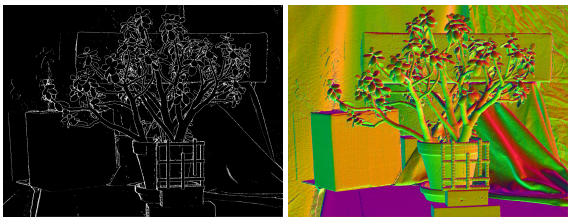

Figure 4. Left: Filtering mask for normal map edge smoothing. Right: Smoothed normal map.

1. **Compute initial normals**. The initial normal for each pixel is computed from the cross product of two vectors formed from the 3D difference of the neighboring pixels.
2. **Extract filtering mask**. Whenever a normal is facing away from the camera (z-component thresholding) and has a high local variation (thresholding of Sobel operator res-

ult) a potential depth discontinuity is detected and marked in a mask image (see Fig. 4 left).

3. **Normal map smoothness**. A Median filter is applied to the masked region in the initial normal map, which selects either the foreground or the background normal.

3.2 Deep Sea Image Formation

The deep sea lies in complete darkness, where no sunlight penetrates. Artificial light sources are required to provide illumination for camera imaging. Therefore the J-M image formation model is utilized rather than the AFM, as it considers the complete transmission path from point light sources (See Fig. 5). The J-M model describes the underwater image formation as a linear composition of direct signal, forward scattering, and backscatter. Our approach is based on the modified J-M model from (Song et al., 2021a), which additionally supports multi-spotlights for simulating deep sea lighting effects.

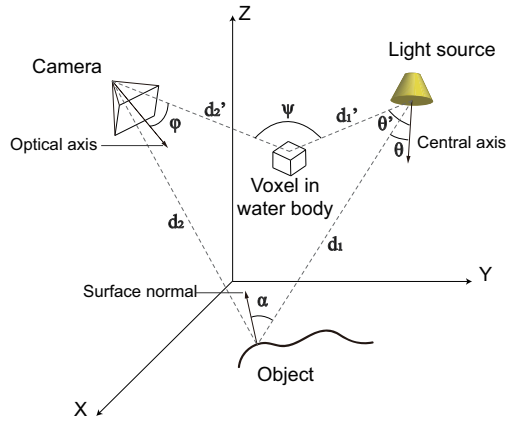

Figure 5. Adapted J-M model with spotlight (Song et al., 2021a).

3.2.1 Attenuation One of the iconic properties in underwater images is their color distortions caused by the water attenuation. The appearance of oceanic water varies depending on the locations, time, and environmental conditions. Attenuation is a wavelength-dependent coefficient. (Jerlov, 1968) measured and categorized waters on Earth into fourteen different spectra, which are also known as Jerlov water types. Later (Akkaynak et al., 2017) utilized Jerlov water types to constrain the space of the oceanic attenuation coefficients for underwater vision applications. The water attenuation with point light sources can be formulated by:

$$E(\lambda, d) = I_\theta(\lambda)\frac{e^{-\eta(\lambda)\cdot d}}{d^2}. \quad (2)$$

where λ = RGB channels
d = light traveling distances from spotlight in [m]
η = water attenuation coefficient in [m^{-1}]
I_θ = spotlight irradiance at angle θ

3.2.2 Phong Reflection Model The Phong reflection model (Phong, 1975) is a well-known model which has been widely applied in computer graphics. It describes the reflected light as a summation of the ambient, the diffuse, and the specular terms. Since underwater specular effects are rare, we omit the specular term in the original Phong reflection model and integrate it into the J-M model. The J-M direct signal $D(\lambda)$ with Lambertian Phong reflection can be expressed by:

$$D(\lambda) = J(\lambda)E(\lambda, d_1)e^{-\eta(\lambda)\cdot d_2}(\cos\alpha + \mathrm{f}_{\mathrm{ambient}}). \quad (3)$$

where J = surface albedo
d_1 = distance from light to object in [m]
d_2 = distance from object to camera in [m]
α = angle between the incident light ray and the surface normal
$\mathrm{f}_{\mathrm{ambient}}$ = ambient factor in the range $(0, 1)$

3.2.3 Scattering Phase Function For deep sea vehicles that cannot avoid the light source being somewhat close to the camera, significant backscatter can be observed in the images, which is characteristic for robotic imaging in the deep ocean (Song et al., 2021a). The scattering effect is caused by the photons interacting with the medium and deviating from their original direction. A common quantity used to describe the scattering is the scattering phase function. It characterizes the angular distribution of scattered light and is often expressed as a 1D function of the angle between the incident light ray and the outgoing ray (which is most relevant in case this is the camera viewing ray) (see Fig. 6). Several phase functions have been proposed for modeling different light scattering effects, such as Rayleigh and Mei phase functions. Oceanographers built special instruments and measured the phase functions, or their unnormalized counterparts, the Volume Scattering Functions (VSFs), in different types of water. A seminal work is by (Petzold, 1972) who measured a wide range of angles for clear, coastal, and turbid ocean water scattering (see Fig. 6). This paper applies the popular Heney-Greenstein (H-G) phase function (Henyey and Greenstein, 1941) to render the J-M backscatter component.

$$p_{HG}(\psi) = \frac{1}{4\pi}\frac{1-g^2}{(1+g^2+2g\cdot\cos\psi)^{3/2}}. \quad (4)$$

where ψ = angle between the incident light ray and the camera viewing ray
g = the asymmetry parameter in the range $(-1, 1)$

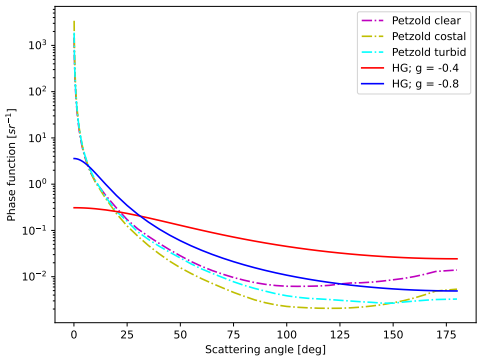

Figure 6. Volume Scattering Phase Function measured by (Petzold, 1972) and H-G Phase Function with different g.

4. THE DSV DATASET GENERATION

4.1 Stereo Benchmarks and Simulation Platforms

The Middlebury Stereo dataset[2] is one of the most famous vision benchmarks in photogrammetry and computer vision. It currently consists of six subsets created in different years. The dataset includes high-resolution stereo pairs with complex scenes. The corresponding pixel-accurate GT disparities were obtained from high accuracy structured light measurements. In this paper, 23 stereo pairs with GT disparities from the Middlebury Stereo 2014 datasets (10 evaluation/training sets with GT and 13 additional datasets with GT), are selected for deep sea data synthesis. Each stereo pair consists of two views with different illumination and exposure settings, the default left and right view images, together with the camera calibration file and corresponding GT disparities. The depth maps are processed according to the methods described in Section 3.1.

Synthesis of the deep sea images are based on the DeepSeaRenderer from (Song et al., 2021a). It is a rasterization-based rendering tools which is initially used for deep sea robotic vision simulation. It utilizes RGB-D images as inputs and implements a modified J-M image formation model supporting multiple spotlights with angular characteristic. On top of the DeepSeaRenderer, we integrated the Lambertian Phong reflection (Section 3.2.2), in particular ambient light to account for the multiple scattered light in the scene, and the H-G phase function (Section 3.2.3) in the renderer to further improve the rendering results.

4.2 Parameter Settings

Setting physically correct rendering parameters is a challenging task. It often requires professional knowledge and instruments in order to define physically meaningful values. Wrong settings would significantly change the image appearance (see Fig. 7) and lead to unrealistic synthesis results.

Figure 7. Simulated images vary significantly with different settings of lighting conditions (first row), attenuation parameters (second row), and g in H-G phase function (third row).

In our case, the camera intrinsics are obtained from the original datasets, the attenuation parameters refer to (Akkaynak et al., 2017) for Jerlov water type IB. The other radiometric parameters are defined according to (Song et al., 2021b), which provides a detailed explanation of their deep sea image rendering settings referring to real deep ocean images. Details about the parameter settings in this paper are listed in Table 1. In our modified renderer, the default values of ambient factors $f_{ambient}$ are set to 0.2 for all RGB channels and the g parameter of the H-G phase function is set to -0.4. Two scenarios with different lighting configurations are defined for synthesizing deep sea images. One places a spotlight 0.5m away on the right side of the origin in the local stereo camera coordinate system (Setup 1) and another moves the light on top of the origin with 0.5m distance (Setup 2). Both setups define the central axis of the spotlight pointing parallel to the camera viewing direction.

Parameter Name	Values
scale factor	3.5
scale factor bs	1600.0
volumetric max depth	4.0
num volumetric slabs	10
slab sampling method	EQUAL_DISTANCE
white balance	[2.498, 1.0, 1.448]
water attenuation RGB	[0.37, 0.044, 0.035]
light spectrum RGB	[0.25, 0.35, 0.4]
light RID type	1
auto iso	false

Table 1. Parameter settings for the rendering.

4.3 Dataset Structure and Usage

The data structure of the DSV stereo datasets is inspired by the original Middlebury 2014 Stereo datasets. Each DSV stereo pair is saved in a individual folder that shares the same name with the Middlebury datasets. In each scene folder, the refined depth maps (depth{0,1}_rf.exr) for left and right views are saved as 16-bit OpenEXR files. The synthesized deep sea stereo images with lighting Setup 1 (im{0,1}_ds1.png) and Setup 2 (im{0,1}_ds2.png) are exported as the same 8-bit PNG format with the original stereo images. Detail about the dataset structure is shown in Fig. 8.

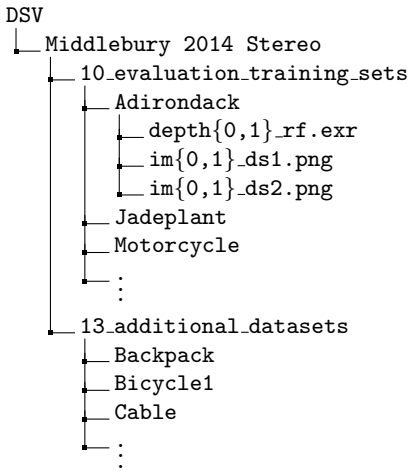

Figure 8. Folder structure of the DSV - Middlebury 2014 Stereo datasets (version 1).

Each synthetic stereo pair can be used for testing corresponding stereo matching methods. The evaluation SDK provided by the Middlebury official web page enables the standardized evaluation of the result, referring to the GT disparity maps. Moreover, the synthetic deep sea images with their original in-air images can be formed into the in-air/underwater pairs, which can be used for training or testing underwater image restoration approaches for a certain underwater approximation setting.

[2] https://vision.middlebury.edu/stereo/data/

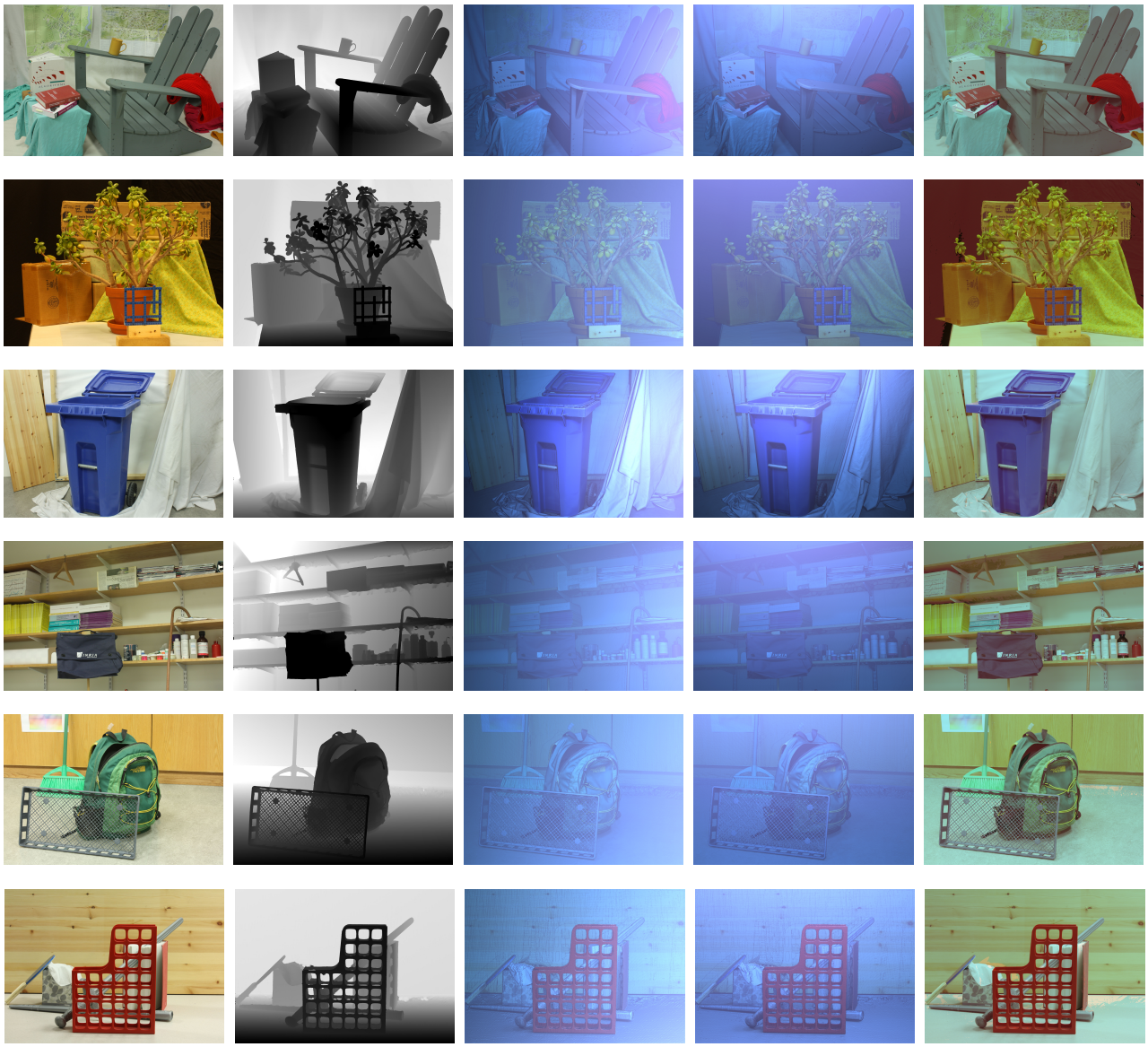

Figure 9. Examples of synthetic deep sea image twins for Middlebury 2014 Stereo datasets. From left to right: original in-air images from Middlebury dataset, Refined depth maps, Synthesized deep sea images with Setup 1 and 2. Synthesized underwater images using the atmospheric fog model (with the same attenuation parameters, background color was set to [110, 137, 212] for RGB). From top to bottom: left view image of Middlebury Adirondack, Jadeplant, Recycle, Shelves, Backpack, and Sword2 dataset.

4.4 Rendering Results

As it is displayed in Fig. 9, example scenes from the Middlebury 2014 datasets were synthesized to approximate aspects of deep sea scenarios with two different lighting setups. As can be seen, the synthesized images resemble some characteristics of deep sea images. The unique scattering pattern and the uneven illumination shading caused by artificial spotlight enable humans to infer the lighting direction from the image. Although the used model is not complete (only single scattering, limited phase function), compared to the simulation results by using the AFM, our approach provides more vision-realistic results.

During the deep sea image synthesis for the Middlebury datasets, we observed some limitations of our method: once the raw disparity contains large areas of missing data, the inpainting algorithms may not able to recover the complex scene accurately. However, these areas do not contain GT for evaluation, which will not contribute to the evaluation metrics. We also noticed

Figure 10. Slight noise in the depth map getting obvious under the spotlight shading. From left to right: In-air color image, Deep sea synthesis, and Normal map.

that the original Middlebury GT disparities contain some artifacts, potentially where inside the shadow area of the structured light. This is also reflected in the official GT sample standard deviations. These issues are not noticeable immediately in the disparities, but are getting obvious when performing shading

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-V-4-2022-353-2022 | © Author(s) 2022. CC BY 4.0 License.

with spotlights (see Fig. 10). We keep this effect in our synthetic results as it reflects the "true" geometry information of the GT depth, which can be interpreted as "millimeters depth scratches" on the object surface.

5. CONCLUSIONS

In this paper, we present a pipeline to generate synthetic images from existing real world in-air RGB-D benchmarks, especially for the deep sea scenario that currently lacks benchmark datasets. It is able to generate more vision-realistic deep sea images with different lighting configurations. The synthesized images share the same GT disparities with the original benchmarks, which can be directly used for evaluating underwater stereo matching algorithms using official platforms and metrics. Additionally, the corresponding in-air/underwater color images can be utilized as references for training or evaluating underwater image restoration methods, though the model applied does not yet cover all deep-sea aspects. Using a physical-model-based rasterization renderer, we demonstrated the creation of more realistic deep sea twins for the Middlebury 2014 Stereo datasets under two different lighting setups. The data will be collected in the DSV dataset, and made available to the public.

ACKNOWLEDGEMENTS

This publication has been funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) Projektnummer 396311425, through the Emmy Noether Programme.

REFERENCES

Akkaynak, D., Treibitz, T., 2019. Sea-thru: A method for removing water from underwater images. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1682–1691.

Akkaynak, D., Treibitz, T., Shlesinger, T., Loya, Y., Tamir, R., Iluz, D., 2017. What is the space of attenuation coefficients in underwater computer vision? *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 568–577.

Baker, S., Scharstein, D., Lewis, J., Roth, S., Black, M. J., Szeliski, R., 2011. A database and evaluation methodology for optical flow. *International journal of computer vision*, 92(1), 1–31.

Berman, D., Levy, D., Avidan, S., Treibitz, T., 2020. Underwater single image color restoration using haze-lines and a new quantitative dataset. *IEEE transactions on pattern analysis and machine intelligence*.

Bertalmio, M., Bertozzi, A. L., Sapiro, G., 2001. Navier-stokes, fluid dynamics, and image and video inpainting. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 1, IEEE, I–I.

Bitterli, B., Jarosz, W., 2017. Beyond points and beams: Higher-dimensional photon samples for volumetric light transport. *ACM Transactions on Graphics (TOG)*, 36(4), 1–12.

Blanco-Claraco, J.-L., Moreno-Duenas, F.-A., González-Jiménez, J., 2014. The Málaga urban dataset: High-rate stereo and LiDAR in a realistic urban scenario. *The International Journal of Robotics Research*, 33(2), 207–214.

Blender Online Community, 2021. Blender - a 3d modelling and rendering package.

Burri, M., Nikolic, J., Gohl, P., Schneider, T., Rehder, J., Omari, S., Achtelik, M. W., Siegwart, R., 2016. The EuRoC micro aerial vehicle datasets. *The International Journal of Robotics Research*, 35(10), 1157–1163.

Butler, D. J., Wulff, J., Stanley, G. B., Black, M. J., 2012. A naturalistic open source movie for optical flow evaluation. A. Fitzgibbon et al. (Eds.) (ed.), *European Conf. on Computer Vision (ECCV)*, Part IV, LNCS 7577, Springer-Verlag, 611–625.

Bythell, J., Pan, P., Lee, J., 2001. Three-dimensional morphometric measurements of reef corals using underwater photogrammetry techniques. *Coral reefs*, 20(3), 193–199.

Cozman, F., Krotkov, E., 1997. Depth from scattering. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, 801–806.

Crane, K., Llamas, I., Tariq, S., 2007. Real-time simulation and rendering of 3d fluids. *GPU gems*, 3(1).

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. *2009 IEEE conference on computer vision and pattern recognition*, Ieee, 248–255.

Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., Van Der Smagt, P., Cremers, D., Brox, T., 2015. Flownet: Learning optical flow with convolutional networks. *Proceedings of the IEEE international conference on computer vision*, 2758–2766.

Drap, P., 2012. Underwater photogrammetry for archaeology. *Special applications of photogrammetry*, 114.

Ferrera, M., Creuze, V., Moras, J., Trouvé-Peloux, P., 2019. AQUALOC: An underwater dataset for visual–inertial–pressure localization. *The International Journal of Robotics Research*, 38(14), 1549–1559.

Geiger, A., Lenz, P., Urtasun, R., 2012. Are we ready for autonomous driving? the kitti vision benchmark suite. *Conference on Computer Vision and Pattern Recognition (CVPR)*.

Harvey, E. N., 1939. Deep sea photography. *Science*, 90(2330), 187.

Henyey, L. G., Greenstein, J. L., 1941. Diffuse radiation in the galaxy. *The Astrophysical Journal*, 93, 70–83.

Jaffe, J. S., 1990. Computer modeling and the design of optimal underwater imaging systems. *IEEE Journal of Oceanic Engineering*, 15(2), 101–111.

Jaffe, J. S., 2014. Underwater optical imaging: the past, the present, and the prospects. *IEEE Journal of Oceanic Engineering*, 40(3), 683–700.

Jerlov, N., 1968. Irradiance optical classification. *Optical Oceanography*, 118–120.

Kwasnitschka, T., Köser, K., Sticklus, J., Rothenbeck, M., Weiß, T., Wenzlaff, E., Schoening, T., Triebe, L., Steinführer, A., Devey, C. et al., 2016. DeepSurveyCam—a deep ocean optical mapping system. *Sensors*, 16(2), 164.

Kölle, M., Laupheimer, D., Schmohl, S., Haala, N., Rottensteiner, F., Wegner, J. D., Ledoux, H., 2021. The Hessigheim 3D (H3D) benchmark on semantic segmentation of high-resolution 3D point clouds and textured meshes from UAV LiDAR and Multi-View-Stereo. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 1, 11.

Li, C., Anwar, S., Porikli, F., 2020. Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognition*, 98, 107038.

Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., Tao, D., 2019. An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, 29, 4376–4389.

Li, J., Skinner, K. A., Eustice, R. M., Johnson-Roberson, M., 2017. WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robotics and Automation letters*, 3(1), 387–394.

Mallios, A., Vidal, E., Campos, R., Carreras, M., 2017. Underwater caves sonar data set. *The International Journal of Robotics Research*, 36(12), 1247–1251.

Manhães, M. M. M., Scherer, S. A., Voss, M., Douat, L. R., Rauschenbach, T., 2016. UUV simulator: A gazebo-based package for underwater intervention and multi-robot simulation. *OCEANS 2016 MTS/IEEE Monterey*, IEEE.

Mayer, N., Ilg, E., Hausser, P., Fischer, P., Cremers, D., Dosovitskiy, A., Brox, T., 2016. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4040–4048.

McGlamery, B., 1980. A computer model for underwater camera systems. *Ocean Optics VI*, 208, International Society for Optics and Photonics, 221–231.

Menna, F., Nocerino, E., Troisi, S., Remondino, F., 2013. A photogrammetric approach to survey floating and semi-submerged objects. *Videometrics, Range Imaging, and Applications XII; and Automated Visual Inspection*, 8791, International Society for Optics and Photonics, 87910H.

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

Nakath, D., She, M., Song, Y., Köser, K., 2022. An Optical Digital Twin for Underwater Photogrammetry. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 90, 69–81.

Nex, F., Remondino, F., Gerke, M., Przybilla, H.-J., Bäumker, M., Zurhorst, A., 2015. ISPRS BENCHMARK FOR MULTI-PLATFORM PHOTOGRAMMETRY. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 2.

Novák, J., Georgiev, I., Hanika, J., Jarosz, W., 2018. Monte carlo methods for volumetric light transport simulation. *Computer Graphics Forum*, 37number 2, Wiley Online Library, 551–576.

Petzold, T. J., 1972. Volume scattering functions for selected ocean waters. Technical report, Scripps Institution of Oceanography La Jolla Ca Visibility Lab.

Phong, B. T., 1975. Illumination for computer generated pictures. *Communications of the ACM*, 18(6), 311–317.

Pizarro, O., Singh, H., 2003. Toward large-area mosaicing for underwater scientific applications. *IEEE journal of oceanic engineering*, 28(4), 651–672.

Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nešić, N., Wang, X., Westling, P., 2014. High-resolution stereo datasets with subpixel-accurate ground truth. *German conference on pattern recognition*, Springer, 31–42.

Scharstein, D., Szeliski, R., 2003. High-accuracy stereo depth maps using structured light. *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, 1, IEEE, I–I.

Schops, T., Sattler, T., Pollefeys, M., 2019. Bad slam: Bundle adjusted direct rgb-d slam. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 134–144.

Schops, T., Schonberger, J. L., Galliani, S., Sattler, T., Schindler, K., Pollefeys, M., Geiger, A., 2017. A multi-view stereo benchmark with high-resolution images and multi-camera videos. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3260–3269.

Sedlazeck, A., Koch, R., 2011. Simulating deep sea underwater images using physical models for light attenuation, scattering, and refraction.

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R., 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, 1, IEEE, 519–528.

Song, Y., Nakath, D., She, M., Elibol, F., Köser, K., 2021a. Deep sea robotic imaging simulator. *Pattern Recognition. ICPR International Workshops and Challenges. ICPR 2021*, Springer, 375–389.

Song, Y., Nakath, D., She, M., Köser, K., 2022. Optical Imaging and Image Restoration Techniques for Deep Ocean Mapping: A Comprehensive Survey. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*.

Song, Y., Sticklus, J., Nakath, D., Wenzlaff, E., Koch, R., Köser, K., 2021b. Optimization of multi-led setups for underwater robotic vision systems. *Pattern Recognition. ICPR International Workshops and Challenges. ICPR 2021*, Springer, 390–397.

Sturm, J., Engelhard, N., Endres, F., Burgard, W., Cremers, D., 2012. A benchmark for the evaluation of rgb-d slam systems. *Proc. of the International Conference on Intelligent Robot Systems (IROS)*.

Ueda, T., Yamada, K., Tanaka, Y., 2019. Underwater image synthesis from rgb-d images and its application to deep underwater image restoration. *2019 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2115–2119.

Zwilgmeyer, P. G. O., Yip, M., Teigen, A. L., Mester, R., Stahl, A., 2021. The varos synthetic underwater data set: Towards realistic multi-sensor underwater data with ground truth. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 3722–3730.