Underwater Depth Estimation based on Water Classification using Monocular Image

Conference Paper · November 2020

DOI: 10.1109/LARS/SBR/WRE51543.2020.9307103

3 authors, including:

Paulo Drews-Jr
Universidade Federal do Rio Grande (FURG)
194 PUBLICATIONS **1,655** CITATIONS

Underwater Depth Estimation Based on Water Classification using Monocular Image

Edwilson Silva Vaz Jr.[1] Everson Fagundes de Toledo[1] Paulo L. J. Drews-Jr[1]

***Abstract*— The rapid growth of computational and sensor capacities allows the development of image restoration methods that can be applied to underwater images. Due to its high degree of absorption, water becomes a major challenge for robotic perception applications. A fundamental issue for many underwater robot applications is the requirement of a depth map. One of the challenges to obtaining monocular underwater depth image is the lack of large image sets to validate the method, or even training a learning-based method. For the estimation, some methods have been proposed in the state-of-the-art either based on a physical model and on a deep learning approach. Through the analysis of the strengths and weaknesses of each kind of approach, this work aims to obtain the best depth map by classifying the input color image. For this, the water type of each image is evaluated. The results obtained in this work are promising, showing the capability of the classifier to identify the most suitable for each input image.**

I. Introduction

Nowadays, several techniques have been developed to extract information from images, either through image processing methods or computer vision. Almost all methods are developed for non-participating medium, *i.e.* they assume that the rays of light do not change their path from the scene to the camera sensor. However, the participating medium interferes directly and indirectly in the direction of the light rays, as well as in their intensities. In the underwater robotic field, observing the scene with clearness is crucial for the applications [1].

The water causes distortion and absorption of light before it reaches the camera. The problem becomes relevant since these effects reduce the overall contrast. Moreover, color-shifting is another problem. Both factors have a direct impact on reducing the visibility of the observed scene. This degradation becomes more relevant according to the distance between the camera and the scene. Thus, greater distances impose less visibility of objects in the scene. Therefore, the underwater image carries depth information [2], differently from the non-participating medium image.

Although most computer vision techniques do not consider participating medium, real-world problems in robotics involve media that interfere with acquired images. In general, the algorithms which deal with underwater images restore the input image in a pre-processing step, enabling the application of typical computer vision algorithms for a specific task. However, some works regard the effects of the water, as in the applications of autonomous underwater robots [3], [4], tracking [5], [6], localization [7], [8], inspection [9], [10] and underwater survey [11], [12].

One of the most important challenges in dealing with underwater images is to understand the distance between the objects observed in the scene and the camera [13]. This distance can be translated into a depth map. Each pixel in this map indicates how close or far the scene is to the camera. Some works proposed the creation of datasets that are composed by the RGB images and their corresponding depth map [14], [15]. As the underwater environment does not contain controlled scenes, underwater datasets containing the depth information are rare. Although indoor RGBD datasets enable to simulate the underwater effects [16], many variables must be taken into account in the underwater world, such as turbidity, the color of the color and the depth of captured images. Then, synthetic data does not lead to the best solution in terms of reality.

Once data is limited, learning-based methods can fail for cases where the observed scene has distinctions from the trained dataset. Even generative methods are not able to represent this diversity. In other words, the generalization ability of learning-based methods depends on the training data. On the other hand, model-based methods are can generalize according to the physical model addressed. Although the model-based methods do not estimate optimized results, they are able to generate reliable results to a larger group of water scenes. Some robotics applications depend on the depth estimation, such as obstacle avoidance [4], which the important information is to know which objects are near the robot.

This work focuses on the premise that state-of-the-art approaches in both fields, learning and model-based underwater depth estimation, can be applied in just one system, which understands the input image and guides it to the most accurate depth map inference for the presented task. Our work aims at real underwater images, which provide a wider range of scenarios. Based state-of-the-art methods, we developed a new system which is able to classify which is the best method to generate the depth map from input underwater image. Furthermore, we adapt the model-based state-of-the-art method to underwater restoration [17] of raw image to estimate depth map from compressed images.

II. Related Works

The monocular depth estimation is an important problem in many applications [4], although most applications focus on underwater image restoration.

[1]All the authors are with the NAUTEC, Centro de Ciencias Computacionais, Universidade Federal do Rio Grande - FURG, RS, Brazil. E-mail: edwilson,eversontoledo,paulodrews@furg.br

He et al. [18] proposed one of the most important dehazing method called Dark Channel Prior (DCP). The method is based in the prior that all regions of a hazy image have at least one channel in a pixel neighborhood which is approximately zero. In [2], the authors propose an adaptation of the *Dark Channel Prior* [18] to underwater images. The Underwater Dark Channel Prior (UDCP) uses only two color channels: green and blue. This is due to the difficulty of understanding the behavior of the red channel when submerged in water since it has strong absorption rate. Other adaptations of DCP are proposed for underwater image restoration, as in [19], [20], [21], [22]. Despite the relevant results, they just estimate the transmission map with a focus on image restoration.

Drews-Jr et al. [23] use the UDCP [2] to estimate an up to scale depth map with the intent to an underwater robot autonomously avoid obstacles. Peng et al. [24] estimate the depth map via image blurriness, which due to the scattering, grows while the distance increases.

The work proposed in [17] is the state-of-the-art in the model-based underwater restoration area. The work proposes a method that takes into account the different types of water, according to Jerlov et al. [25]. In our proposed work, we adapted the work of Berman et al. [17] to estimate depth maps from compressed images, since it is originally proposed to restore raw underwater images.

Some learning-based methods also estimate depth maps. Kuang et al. [26] propose a new network capable of handling underwater omnidirectional images. The idea is to use images in a controlled environment to synthetically generate underwater images and similarly for 360^0 images. Kendall et al. [27] propose a supervised deep learning method that regresses disparity from a pair of stereo images using 3-D convolutions. MegaDepth [28] learns the depth of scenes according to their multi-view photos available on the internet. Although these works show the possibility to estimate depth maps from monocular images even in the air, works for underwater estimation are much more limited due to scarcity of paired datasets.

The state-of-the-art learning-based approach in the underwater depth map estimation area is the method presented by Gupta et al. [29]. This method, called UW-Net, is based on the CycleGan [30] deep network, which has an unsupervised training framework. The method learns to transform images from the underwater domain to the terrestrial RGBD domain with the addition of depth and vice versa, which enables the network to learn the depth from underwater images.

This work proposes a method that can deal with the advantages of both learning and model-based state-of-the-arts. The main idea is to take advantage of what the literature already proposed and direct the input image to the best method to estimate its depth map.

III. Methodology

This work proposes an algorithm that classifies the most appropriate method to generate the depth map for a single input underwater image. Firstly, we present both state-of-the-art methods based on learning and physical model. Furthermore, we present our method, which uses water classification to obtain the depth map.

A. Learning-based method

The state-of-the-art learning-based technique for underwater depth estimation is presented in the work proposed by Gupta et al. [29]. The algorithm consists of an unsupervised network called UW-Net. This work presents a scientific contribution since synthetic images are not used for the networks training. Therefore, our framework does not require paired images. However, terrestrial images are fundamental to the network once the adversarial networks need to have a domain which contains the depth information. Indoor RGBD images are used.

Fig. 1 shows how a terrestrial image (with depth map addition) is translated into an underwater color image. Fig. 2 presents the opposite, with an underwater color image as input and a RGBD indoor image as the output. Each network is composed of two discriminators and two generators.

The adversarial networks can learn how to map images from an input set to a certain domain, having a large enough set of unpaired images. Thus, it's possible to map any set of images from a domain A in a set of images of a domain B.

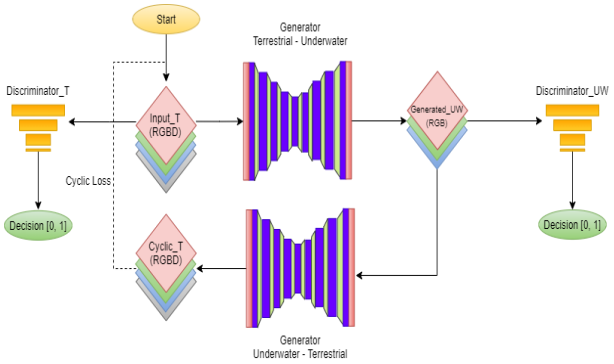

Fig. 1. Cyclic network of translation from terrestrial image to underwater image.

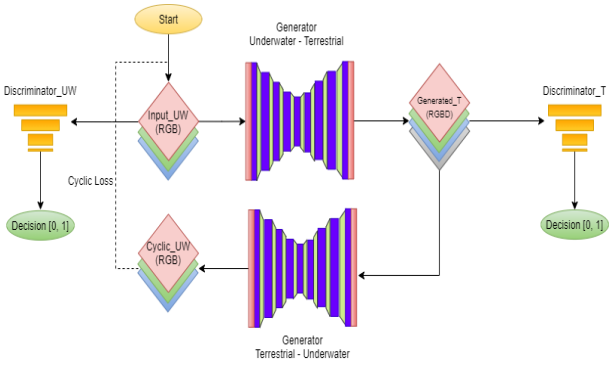

Fig. 2. Cyclic network for translating underwater images to terrestrial images.

The UW-Net [29] is based on the CycleGAN network [30]. This cyclic network can translate an image of a domain A

for a domain B and then, it is possible to translate it back to the domain A and vice-versa. The most important difference between UW-Net and CycleGAN is due to the use of a fourth channel in one of the domains. CycleGAN usually handles RGB images in both domains. On the other hand, UW-Net uses underwater RGB images on one side and terrestrial RGBD images on the other side. We adopted the trained model made available by Gupta et al. [29].

B. Model-based method

Model-based methods have a great capacity for generalization. However, the physical models are a simplified representation of the world. Thus, they are not optimized for specific circumstances and consequently, they cannot achieve such fine results as neural networks. However, model-based methods do not need any training data, only the input image.

The work proposed by Berman et al. [17] is the state-of-the-art on underwater raw image restoration. However, we adapted it to obtain depth maps since the original method estimates the transmission map as an intermediary step in the restoration pipeline. The method considers several water types, according to Jerlov [25]. Each water type leads to a different restoration procedure. The best result is evaluated based on the Gray-World assumption in the restored images.

The transmission is estimated through the simplified image formation model. The authors designed the model as a function of the blue channel, thus being possible to obtain only one transmission map. Such a map is described in Eq. 1, in which t_B is the transmission map in the blue channel, and I_c, J_c, A_c are the input image, the restored image and the illumination coefficient in the color channel $c \in \{R, G, B\}$, respectively. β_{BG} and β_{BR} are the attenuation ratios on blue-green (β_B/β_G) and blue-red (β_B/β_R) channels. Berman et al. [17] uses the similar formulation to the Haze-Lines [31] to predict J and then a first estimation of the transmission.

$$\begin{bmatrix} (I_R(\boldsymbol{x}) - A_R)^{\beta_{BR}} \\ (I_G(\boldsymbol{x}) - A_G)^{\beta_{BG}} \\ (I_B(\boldsymbol{x}) - A_B) \end{bmatrix} = t_B(\boldsymbol{x}) \begin{bmatrix} (J_R(\boldsymbol{x}) - A_R)^{\beta_{BR}} \\ (J_G(\boldsymbol{x}) - A_G)^{\beta_{BG}} \\ (J_B(\boldsymbol{x}) - A_B) \end{bmatrix} \tag{1}$$

The method also calculates the lower-bound transmission based on each pixel, considering that $J_c \geq 0$. The lower-bound transmission is used for the cases where the Mahalanobis distance between two pixels (x) is small compared to the estimated illumination A_c. This occurs in regions that contain only water. The transmission map is calculated based on the initial transmission when the Mahalanobis distance is large compared to A_c. Otherwise, a linear composition between the lower-bound transmission t^{LB} and the initial transmission $\tilde{t}$ is adopted, (Eq. 2).

$$t(x) = \alpha(\boldsymbol{x}) \cdot t^{LB}(\boldsymbol{x}) + (1 - \alpha(\boldsymbol{x})) \cdot \tilde{t}(\boldsymbol{x}), \tag{2}$$

where t is the estimated transmission map and $\alpha(\boldsymbol{x})$ is the weight coefficient. Based on the transmission map, we estimated the depth map d of the input underwater image as shown in Eq. 3. We classify the image using the illumination coefficient A_c, detailed in the next section.

$$d = -\frac{log(t)}{\beta_c}. \tag{3}$$

Furthermore, we do not perform any image restoration since our objective is to estimate depth maps. Contrast-enhancement steps proposed by Berman et al. [17] are not performed. They are not reliable for real underwater images acquired by typical underwater cameras whereas due to the embedded compression and pre-processing algorithms. The work of Berman et al. [17] assumes a raw input image, this is not the case of most of the available underwater images. Finally, we use an additional Guided filter [32] step to smooth and enhance the depth map. This step is required due to the limited quality of the estimated depth map. Despite such a limitation, the final algorithm can generalize the depth map estimation for every water type.

C. Classifying the water type

This work also proposes an algorithm that classifies the water type in a simple but effective way. In 1976, Jerlov [25] created a new classification of water types that has ten categories, which were determined through experiments with different waters around the world. Jerlov [25] collected underwater samples from seas and oceans to structure the classification system. The mean attenuation of each water is estimated for each water type in many wavelengths. The wavelengths $600nm$, $525nm$, and $475nm$ represent R, G, B color channels, respectively. Thus, we can simulate the effects of each type of water. Fig. 3 shows the color of each water type in a depth ranging from $1m$ to $20m$. Furthermore, each water type had its attenuation coefficient estimated and is presented as a function of wavelength (λ) in Fig. 4. The classification of Jerlov's water types is subdivided into ocean waters (I, IA, IB, II, III) and coastal waters (1, 3, 5, 7, 9). Using the Jerlov's classification, we can classify an underwater image. We assume the image is naturally lit by a pure white light in the water surface. Thus, we can use the model presented by Roser et al [33] and adapt it to our problem, as shown in Eq. 4.

$$A(wd) = TI_0 e^{-\beta(\lambda)wd}, \tag{4}$$

where T is the transmission coefficient on the water surface, wd is the water depth, i.e. the distance between the scene and the water surface, and I_0 is the atmospheric intensity on the surface. Therefore, considering that the light is white, the Eq. 4 simplifies to Eq. 5:

$$A(wd) = e^{-\beta(\lambda)wd}, \tag{5}$$

where $\beta(\lambda)$ is estimated in [25]by measuring the values directly in the water.

Although used in image restoration methods, such as in [17], the classification of Jerlov's water types in terms of application can be somewhat complex since some water types are too similar, as is the case of types I, IA and IB.

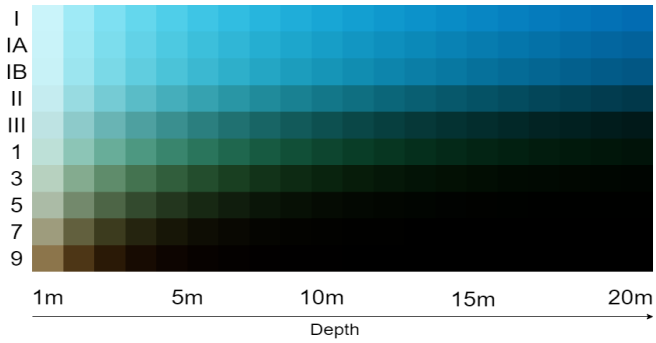

Fig. 3. Classification of Jerlov water types. Representation of the color of each of the 10 water types, being: I, IA, IB, II and III, ocean waters; and 1, 3, 5, 7, 9, coastal waters. Each water type is represented with 20 colored boxes, assuming a distance from $1m$ to $20m$.

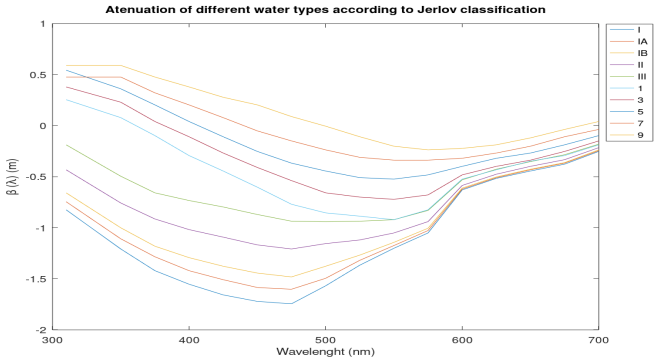

Fig. 4. Attenuation of different water types according to Jerlov's classification depending on the wavelength. RGB colors are represented by the wavelengths $600nm$, $525nm$ and $475nm$, respectively.

Also, a great part of the datasets present in the literature may contain images captured in several water types. These datasets are hybrid, in terms of Jerlov's classification, despite the similarity between the colors of the water. The cases described in [34], [35], [36] are examples of image datasets which vary their classification of water type, according to Jerlov's classification.

The illumination A needs to be estimated. There are many ways to estimate it [4]. We adopted the methodology proposed in [17], in which initially illumination coefficient of the image is estimated. Given an input image, a local histogram equalization is applied. Therefore, an edge map is generated, where smooth and non-textured regions are selected. The color of the largest region of the image that fits the required characteristics is assumed as A.

Based on the illumination estimation, we optimize both unknown variables to minimize the Eq. 6. We optimize water depth wd for each water type, following Jerlov's classification. The optimization is subject to bounds, thus we use the method proposed by Coleman et al. [37]. We limited the water depth to $100m$ due to the assumption of a naturally lit image. The obtained β allows us to know the water type of the input image in a much faster way than the method proposed in [17]. We also obtain the estimated water depth for each input image. This is another result of our optimization method.

$$\min_{\beta,wd} \sum_{c\in R,G,B} A_c - e^{-\beta_c wd}. \tag{6}$$

D. Proposed algorithm for depth estimation

The water classifier allows us to understand each image context. Assuming the previous knowledge of the training set of the UW-Net, we can discover whether or not the state-of-the-art neural network has sufficient knowledge to generate a consistent depth map. In other words, the algorithm understands if the input image presents adequate results when the trained neural network is used. If not, the generated depth map is obtained using the model-based method since the network cannot generate a reliable map.

Assuming a proper knowledge of the water type of the input image, as well as the knowledge of the water types used in the training dataset of the employed network, the proposed method aims to generate the best depth map for each input image, according to the available methods in the framework. There is a method based on a physical model with high generalization capability but limited depth map estimation. Thus, every input image which the water type is previously trained by the learning network, we use the depth estimated by this model. The Algorithm 1 presents the operation steps of the proposed method.

Input: underwater image
Output: depth map
Load the water types used in the training set of the learning-based method ($types[]$);
Classify water type of the input image (t_input);
if $t\_input \in types[]$ **then**
    Generates depth map through learning-based method;
else
    Generates depth map through model-based method;
end
End algorithm

Algorithm 1: Proposed method

The first step consists in loading which water types are known by the learning-based approach. The water types of the trained model are previously classified using the algorithm presented in the last section. Then, the algorithm classifies the water type of the input image. After this step, it is possible to compare if the water type of the input image is previously known or not by the convolutional neural network (CNN).

By comparing the water types learned by the network and the water type of the input image, we define whether using the model-based method or the learning-based method. If the water type of the input image is known by the network, then the most reliable depth map is generated by the learning-based method. If the input water type is unknown by the network, then we choose the model-based method since we can not guarantee a reliable result from the CNN.

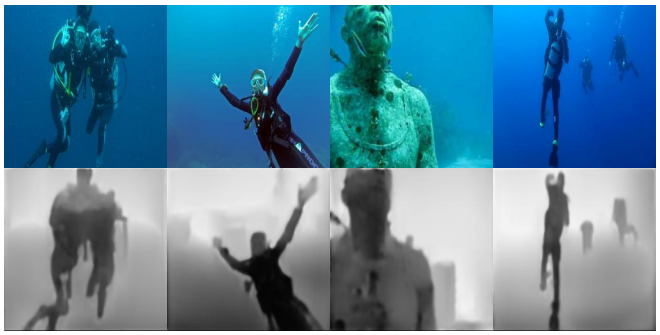

Fig. 5. Depth maps generated for previously known water types by the network. We present the input image, on the top row, and the result generated by the learning-based method, on the bottom row.

Fig. 6. Depth maps generated for unknown water types by the learning-based method. The first row (a) represent the input image. The second one (b), the model-based result, which was chosen to generate the depth map by the algorithm. In the last row (c), the result that would be generated by the learning-based method if it was chosen.

IV. Results

In the proposed algorithm, the first step is to identify which water types the learning network was trained with. The network proposed by Gupta et al. [29] was trained by the authors with a dataset composed massively of I, IA, IB and II water types. Then, for an input captured in one of these water types, the method generates accurate depth maps.

In real-world scenarios, there are some other water types, varying turbidity and color. We evaluated the proposed method using the *Real Underwater Segmentation Dataset* [38]. This dataset is composed of 700 underwater images collected online which have a diversity of water types. Such a variety is crucial, once we need to evaluate different situations and analyze the behavior of the proposed algorithm.

Figures 5 and 6 show qualitative experimental results. Fig. 5 presents many cases where the water type of the input image is known and the results of the CNN are reasonable. In Fig. 6 are depicted the results obtained using the proposed model-based method, where the water types are unknown to the UW-Net. The comparisons in Fig. 6 indicate that the results obtained by the learning-based method are useless for tasks as obstacle avoidance. However, the results achieved by the model-based method are promising.

Despite the obtained results, it is possible to note that the learning-based method is not able to generate accurate results for most of the images. The model-based method does not obtain results as accurate as of the learning-based method when the water type matches those present in the training dataset. However, the model-based method has a stronger ability to generalize than general CNNs.

For quantitative results, two metrics are adopted: Scale-Invariant Mean Squared Error (SI-MSE) proposed in [39] and Pearson correlation coefficient which is defined as $\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y}$. Larger values of the Pearson correlation and lower values of SI-MSE lead to better results. Both metrics require a dataset that contains ground truth depth maps since the output of each method must be compared with a reference. We adopted the dataset proposed by Berman et al. [17]. To the best of our knowledge, it is the only underwater image dataset with a reliable ground truth depth map.

The images from the aforementioned dataset were acquired in the Mediterranean Sea and Red Sea. Both environments are included in water types I and II according to Jerlov's book [25] and our classification algorithm. Thus, the dataset contains images whose domains are known to the pre-trained CNN. Our algorithm chooses considering the all these cases, the learning-based method. Table I presents the quantitative results for Pearson correlation coefficient and the SI-MSE. Note that the learning-based method presents better values for both cases. Our method correctly classifies the water type of the input images and successfully chooses the best method to estimate the depth map.

V. Conclusions

This work presents a depth map estimation using state-of-the-art methods based on learning and physical model. Our approach is based on an algorithm that classifies the water type of the input image and decides the best way for generating its depth map. We proposed a new classification algorithm for each water type of the underwater input images.

Although our work is supported by state-of-the-art methods, we can offer a hybrid solution, where a knowledge of the water type is used to decide the best solution according to the existing data. In this work, the weaknesses and strengths of each method are regarded. We identified the cases in which the UW-Net obtain accurate and poor results. Thus, we deliver the best depth map possible for the input.

Future works will be focused on image fusion. Instead of just release the state-of-the-art result, we fuse the information to obtain a better depth map. In this case, the depth map is partially estimated with each method. Furthermore, other methods from the literature can be fused. We also plan to generate new underwater datasets with depth maps in many water types, this will allow us to evaluate the algorithm in different water conditions.

TABLE I

QUANTITATIVE RESULTS. LARGER VALUES FOR PEARSON CORRELATION (ρ) AND LOWER VALUES FOR SI-MSE ARE BETTER.

Metric	**UDCP [2]**	**Godard et al. [40]**	Model-based Method	Learning-Based	**Ours**
ρ	0.556	0.803	0.669	**0.828**	**0.828**
SI-MSE	0.45	0.311	0.518	**0.211**	**0.211**

ACKNOWLEDGMENT

The authors would like to thank to National Council for Scientific and Technological Development (CNPq) and the Coordination for the Improvement of Higher Education Personnel (CAPES).

REFERENCES

[1] A.-N. Ponce-Hinestroza, L. A. Torres-Méndez, and P. Drews-Jr, "Using a MRF-BP model with color adaptive training for underwater color restoration," in *ICPR*, 2016, pp. 787–792.

[2] P. Drews-Jr, E. do Nascimento, F. Moraes, S. Botelho, and M. Campos, "Transmission estimation in underwater single images," in *IEEE ICCVw*, June 2013.

[3] C. Fabbri, M. J. Islam, and J. Sattar, "Enhancing underwater imagery using generative adversarial networks," in *IEEE ICRA*, 2018, pp. 7159–7165.

[4] P. Drews-Jr, E. Hernández, A. Elfes, E. R. Nascimento, and M. Campos, "Real-time monocular obstacle avoidance using underwater dark channel prior," in *IEEE/RSJ IROS*, 2016, pp. 4672–4677.

[5] P. Drews-Jr, V. Kuhn, and S. Gomes, "Tracking system for underwater inspection using computer vision," in *International Conference on Offshore and Marine Technology: Science and Innovation*, 2012, pp. 27–30.

[6] S. Marini, E. Fanelli, V. Sbragaglia, E. Azzurro, J. D. R. Fernandez, and J. Aguzzi, "Tracking fish abundance by underwater image recognition," *Scientific reports*, vol. 8, no. 1, pp. 1–12, 2018.

[7] A. Concha, P. Drews-Jr, M. Campos, and J. Civera, "Real-time localization and dense mapping in underwater environments from a monocular sequence," in *IEEE OCEANS*, 2015.

[8] H. Chame, P. Drews-Jr, and S. Botelho, "Towards a biologically-inspired model for underwater localization based on sensory-motor coupling," in *2017 Latin American Robotics Symposium (LARS) and 2017 Brazilian Symposium on Robotics (SBR)*, 2017, pp. 1–6.

[9] P. Ridao, M. Carreras, D. Ribas, and R. Garcia, "Visual inspection of hydroelectric dams using an autonomous underwater vehicle," *Journal of Field Robotics*, vol. 27, no. 6, pp. 759–778, 2010.

[10] P. Drews-Jr, V. Kuhn, and S. Gomes, "Controlling a system for underwater visual inspection," *Marine Systems & Ocean Technology*, vol. 8, no. 2, pp. 121–128, 2013.

[11] G. Ferri, A. Munafò, A. Tesei, P. Braca, F. Meyer, K. Pelekanakis, R. Petroccia, J. Alves, C. Strode, and K. LePage, "Cooperative robotic networks for underwater surveillance: an overview," *IET Radar, Sonar & Navigation*, vol. 11, no. 12, pp. 1740–1761, 2017.

[12] J. A. Donaldson, P. Drews-Jr, M. Bradley, D. L. Morgan, R. Baker, and B. C. Ebner, "Countering low visibility in video survey of an estuarine fish assemblage," *Pacific Conservation Biology*, vol. 26, no. 2, pp. 190–200, 2020.

[13] P. Drews-Jr, E. R. Nascimento, M. F. M. Campos, and A. Elfes, "Automatic restoration of underwater monocular sequences of images," in *IEEE/RSJ IROS*, 2015, pp. 1058–1064.

[14] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *ECCV*, 2012, pp. 746–760.

[15] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213–3223.

[16] A. Duarte, F. Codevilla, J. D. O. Gaya, and S. S. C. Botelho, "A dataset to evaluate underwater image restoration methods," in *IEEE OCEANS*, 2016, pp. 1–6.

[17] D. Berman, D. Levy, S. Avidan, and T. Treibitz, "Underwater single image color restoration using haze-lines and a new quantitative dataset," *IEEE TPAMI*, 2020.

[18] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE TPAMI*, vol. 33, no. 12, pp. 2341–2353, 2010.

[19] N. Carlevaris-Bianco, A. Mohan, and R. M. Eustice, "Initial results in underwater single image dehazing," in *IEEE OCEANS*, 2010, pp. 1–8.

[20] A. Galdran, D. Pardo, A. Picón, and A. Alvarez-Gila, "Automatic red-channel underwater image restoration," *Journal of Visual Communication and Image Representation*, vol. 26, pp. 132–145, 2015.

[21] J. Y. Chiang and Y.-C. Chen, "Underwater image enhancement by wavelength compensation and dehazing," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1756–1769, 2011.

[22] H. Lu, Y. Li, L. Zhang, and S. Serikawa, "Contrast enhancement for images in turbid water," *JOSA A*, vol. 32, no. 5, pp. 886–893, 2015.

[23] P. Drews-Jr, E. Nascimento, S. Botelho, and M. Campos, "Underwater depth estimation and image restoration based on single images," *IEEE Computer Graphics and Applications*, vol. 36, no. 2, pp. 24–35, 2016.

[24] Y. Peng, X. Zhao, and P. C. Cosman, "Single underwater image enhancement using depth estimation based on blurriness," in *2015 IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 4952–4956.

[25] N. G. Jerlov, *Marine optics*. Elsevier, 1976, vol. 14.

[26] H. Kuang, Q. Xu, and S. Schwertfeger, "Depth estimation on underwater omni-directional images using a deep neural network," *arXiv preprint arXiv:1905.09441*, 2019.

[27] A. Kendall, H. Martirosyan, S. Dasgupta, P. Henry, R. Kennedy, A. Bachrach, and A. Bry, "End-to-end learning of geometry and context for deep stereo regression," in *IEEE ICCV*, 2017, pp. 66–75.

[28] Z. Li and N. Snavely, "Megadepth: Learning single-view depth prediction from internet photos," in *IEEE CVPR*, 2018, pp. 2041–2050.

[29] H. Gupta and K. Mitra, "Unsupervised single image underwater depth estimation," *arXiv preprint arXiv:1905.10595*, 2019.

[30] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *IEEE ICCV*, 2017, pp. 2223–2232.

[31] D. Berman, S. Avidan *et al.*, "Non-local image dehazing," in *IEEE CVPR*, 2016, pp. 1674–1682.

[32] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE TPAMI*, vol. 35, no. 6, pp. 1397–1409, 2012.

[33] M. Roser, M. Dunbabin, and A. Geiger, "Simultaneous underwater visibility assessment, enhancement and improved stereo," in *IEEE ICRA*, 2014, pp. 3840–3847.

[34] M. J. Islam, C. Edge, Y. Xiao, P. Luo, M. Mehtaz, C. Morse, S. S. Enan, and J. Sattar, "Semantic segmentation of underwater imagery: Dataset and benchmark," *arXiv preprint arXiv:2004.01241*, 2020.

[35] A. Gomez Chavez, A. Ranieri, D. Chiarella, E. Zereik, A. Babić, and A. Birk, "CADDY underwater stereo-vision dataset for human–robot interaction (HRI) in the context of diver activities," *Journal of Marine Science and Engineering*, vol. 7, no. 1, p. 16, 2019.

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

[37] T. Coleman and Y. Li, "An interior trust region approach for nonlinear minimization subject to bounds," *SIAM Journal on Optimization*, vol. 6, no. 2, pp. 418–445, 1996.

[38] I. Souza, "Segmentação de imagens subaquáticas baseada em aprendizado profundo (in portuguese)," Undergraduate Thesis, Universidade Federal do Rio Grande - FURG, 2017.

[39] D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network," in *Advances in neural information processing systems*, 2014, pp. 2366–2374.

[40] C. Godard, O. Mac Aodha, and G. J. Brostow, "Unsupervised monocular depth estimation with left-right consistency," in *IEEE CVPR*, 2017, pp. 270–279.