On the Mutual Relation between SLAM and Image Enhancement in Underwater Environments

Monika Roznere
Department of Computer Science
Dartmouth College
Hanover NH, USA
monika.roznere.gr@dartmouth.edu

Alberto Quattrini Li
Department of Computer Science
Dartmouth College
Hanover NH, USA
alberto.quattrini.li@dartmouth.edu

***Abstract*—Typically, underwater model-based image color correction requires depth information to properly enhance raw images, based on heuristic estimates, such as dark prior. Vision-based state estimation systems – such as ORB-SLAM – together with single beam sonar can provide a more accurate estimate for image color correction methods, resulting in a synergistic improvement on state estimation and color correction. This paper discusses the mutual relationship of SLAM and image enhancement in underwater environments for low-cost Remotely Operate Vehicles (ROVs) and Autonomous Underwater Vehicles (AUVs).**

***Index Terms*—underwater SLAM, image color correction**

I. Introduction

From archaeology to biology, underwater exploration is fundamental for preservation and new discoveries, and it will advance with the technological progress of autonomous underwater robotic systems. One of the main challenges includes *visual underwater perception*. In particular, considering low-cost Remotely Operated Vehicles (ROVs) and Autonomous Underwater Vehicles (AUVs), whose sensor configurations are typically composed of Inertial Measurement Unit (IMU), compass, depth sensor, single-beam sonar, and monocular camera.

This paper discusses how underwater perception for Simultaneous Localization and Mapping (SLAM) problem can be more effective and reliable using an inexpensive monocular camera and a single-beam sonar on a low-cost ROV.

State-of-the-art real-time visual SLAM algorithms take raw images as input, extract features from each image, and track them over subsequent frames to then estimate pose and 3D points. While high accuracy have been demonstrated with Inertial Measurement Units (IMU) – typically high-end, in the underwater domain – and stereo cameras, low-cost vehicles are still far from being robust enough to enable autonomous operation.

Indeed, distortions, blurriness, and color degradation are more prevalent features in underwater images compared to images taken above the surface. These optical properties seen in the images are characterized by the dynamic phenomena in the bodies of water, considering the season, the weather, and the quality of marine life [1]. As light propagates underwater, it interacts with suspended particles, causing a proportion of rays of light to be absorbed or scattered. The combined effects

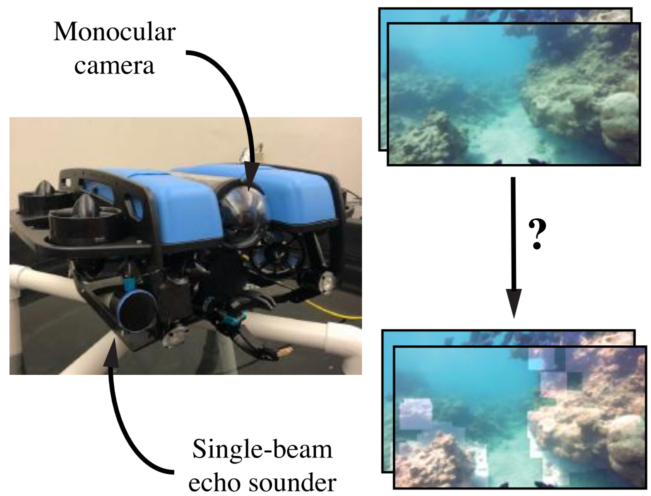


Fig. 1: Given a monocular camera and an echo sounder mounted on a low-cost ROV (BlueROV2 in our lab), how can color correction and a SLAM system be improved?

of absorption and scattering is called *attenuation*. While there has been extensive research in underwater image formation models, they heavily depend on the current scene information, such as distance to the objects (depth), which is difficult to estimate.

Starting from our previous work [2], this paper presents how data from a single beam sonar and a monocular camera can be integrated to mutually improve image feature extraction in SLAM algorithms, while also reducing depth estimation error from color enhancement methods – see Fig. 1 for a depiction of the robot and the problem addressed in this paper. We discuss early results to highlight the feasibility of our proposed approach and the immediate next steps to make low-cost underwater vehicles more autonomous and accessible to the scientific and industrial communities.

II. Background

Much work has been done in solving the problem of SLAM, starting from *filtering-based* approaches. The initial research in

SLAM [3] focused on the notion of applying Kalman Filters, and has been extended to Bayesian filtering and PHD filters [4], [5]. These algorithms can depend on data either from motion sensors, video cameras, or imaging sonars [6]. Recently, feature-based SLAM systems has received more attention for its real-time applicability. Examples include ORB-SLAM [7] and OKVIS [8], which are based on keyframes and have been tested using monocular and stereo cameras. They implement different feature detection and descriptor methods to extract image features and find correspondences to the keyframe features. Our method employs ORB-SLAM as it relies solely on camera images – a reliable high-frequency IMU, which is required by OKVIS [8] and VINS [9], is not available on our robot.

In general, fusing data from more sensors can increase the quality of the estimate of the algorithms, e.g., using multi-calibrated cameras [10] and a profiling scanning sonar [11]. As mentioned, low-cost ROVs and AUVs are not equipped with such costly devices and are usually installed with one camera and a single-beam echo sounder. However, monocular SLAM algorithms suffer from scale ambiguity and drift [12]. Additionally, single-beam echo sounders quantify only one depth measurement at a time.

Many SLAM algorithms assume that the collected images used for feature extraction are invariant to a certain degree to changes, e.g., in illumination. While it is possible to use methods for enhancing images before feeding to SLAM algorithms, these types of methods do not typically depend on information from the environment. They use *statistics-based algorithms* to estimate parameters and then enhance the images. These include equalization methods that distribute pixel values in different color spaces [13]–[15] and methods that apply filters to eliminate initial noise and post-processing artifacts [16], [17].

Other color correction methods use the underwater imaging formation model, which explicitly considers both direct and the backscatter signal – the light that directly goes from the object to the camera and the light that scatters due to particles present in the water before reaching to the camera, respectively. These approaches are classified as *physics-based methods*. However, the depth (distance from the camera to the object), which is a fundamental parameter in the model, is infeasible to extract without supplementary sensors. One way to estimate the depth is by using the dark channel prior [18] and implementing dehazing algorithms [19]. Then, such a method can compensate the wavelengths of each RGB channel, according to the underwater imaging formation model. Another method is proposed by Lu *et al.* [20]: a simple prior is calculated from the difference in attenuation among the red color channels, which then can be used to estimate the transmission map and noise filter. Similarly, Carlevaris-Bianco *et al.* [21] exploit the attenuation differences between the RGB channels in water to estimate the depth of the scene, and then use the underwater imaging formation model to correct the image.

Last year, Akkaynak and Treibitz [22] proposed an alteration of the underwater imaging formation model, such that the coefficients for the direct and the backscatter signals are now different from one another and dependent on environmental parameters, like water depth, object distance, and water type. In this paper, we apply this image enhancement method and assume that the required environmental parameters can be received from on-board devices, like single-beam echo sounders, or from extracted features from images processed by SLAM algorithms.

III. Approach

A. Image Color Correction

We apply our recently proposed color correction algorithm [2]. It is based on the new underwater imaging formation model, proposed by Akkaynak and Treibitz [22], that assumes that the attenuation coefficients for backscatter and direct signal are different. Our system automatically derives the two attenuation coefficients necessary to solve the imaging formation model equation, such that images can be color corrected directly on-board an underwater robot.

Given the unattenuated (corrected) image J_c, the imaging range z, the wideband veiling light B_c^∞, and the attenuation coefficients β_c^D and β_c^B for direct signal and backscatter, the raw image taken by the camera can be expressed by the underwater imaging formation model as:

$$I_c = J_c e^{-\beta_c^D(\mathbf{v}_D)z} + B_c^\infty(1 - e^{-\beta_c^B(\mathbf{v}_B)z}) \tag{1}$$

where c represents each of the *RGB* color channels. Note that in the new underwater imaging formation model, the attenuation coefficients depend on $\mathbf{v}_D = [z, \rho, E, S_c, \beta]$ and $\mathbf{v}_B = [E, S_c, b, \beta]$, where ρ is the reflectance spectrum of the object, E is the ambient light, S_c is the spectral response of the camera, and $\beta = a+b$ is the beam attenuation coefficient, with a and b being the beam absorption and scattering coefficient.

Wideband veiling light B_c^∞ is calculated as follows:

$$B_c^\infty = \frac{1}{k}\int_{\lambda 1}^{\lambda 2} S_c(\lambda)\frac{b_c E_c(d,\lambda)}{\beta_c}d\lambda \tag{2}$$

where k is a scalar directing image exposure and λ is the wavelength. $E(d, \lambda)$, the ambient light at a given wavelength λ, at depth d, is

$$E(d,\lambda) = E_0\frac{K_d(\lambda)}{d} \tag{3}$$

where E_0 is the ambient light at the surface and K_d is the diffuse attenuation coefficient. The coefficients a, b, and K_d depend on the type of water, as defined by Jerlov [23], and can be derived from the current charts.

We would like to point our interested readers to [2] for a descriptive analysis of our observations and assumptions for calculating the wideband veiling light B_c^∞. In this paper, we simplify the calculations by assuming the wideband veiling light to be the average pixel value of the background [15]. This approach is appropriate for diverse water conditions or when prior knowledge is unreliable.

If J_c is known, in cases when a color chart is utilized, β_c^D and β_c^B can be estimated by taking pixel samples from two color patches of the same image. In cases where J_c is not known, the attenuation values can be optimized from previous experiments. Using these assumptions and estimations, the unattenuated image J_c can be solved using Equation (1) given an input image I_c and the corresponding depth values z, characterizing the distances of the objects in the scene from the camera.

B. Depth Extraction

Depth measurements can be extracted either from specific devices, such as echo sounders and imagery sonars, or from tracked feature points processed by SLAM algorithms. In our approach, we assume the true depth measurement is captured by a low-cost single-beam echo sounder and surrounding image features with corresponding depth estimations are handled by ORB-SLAM [7].

Similar to [24], the echo-sounder can be installed and directed intentionally for collecting more informative data. Note that, a typical single-beam echo sounder generates sound waves and returns the distance measurement of the strongest returned response. Particularly, the echo-sounder will be positioned, such that its measurements can be identified by a region in the images taken by the on-board camera. This assumes that the echo sounder and the camera are not moving independently or unknowingly of one another.

Monocular ORB-SLAM [7] is a real-time keyframe-based SLAM system that handles scene features for mapping, tracking, and localization tasks. For our interests, the features are characterized by pixel points assigned in the grayscale-converted image and its corresponding depth estimations. However, depth estimation from monocular SLAM systems is faulty due to self-drift and scale uncertainty.

Depth measurements from the single-beam echo sounder can account for this ambiguity. If features detected by the ORB-SLAM system lie in the region detected by the echo sounder, their depth measurements can then be altered to match the measurements read by the physical sensor. The remaining depth measurements of the features detected by the system can be adjusted according to some ratio calculated in the previous step, when converting estimated values to accurate readings.

C. Design

The camera and echo sounder require to be calibrated when applying the algorithm in a new environment. This is especially important for color correction, as the attenuation values, β_c^D and β_c^B, may be different in location or change over time. One approach is to create a 3D structure that is characterized by known irregularities that will be detected by the ORB-SLAM system and then matched with the echo sounder readings. In addition, the color chart is required for retrieving prior attenuation values. The two procedures can be combined by adding the white and black patches from the color chart to the 3D calibration structure. The camera is also calibrated using Aprilgrid targets [25].

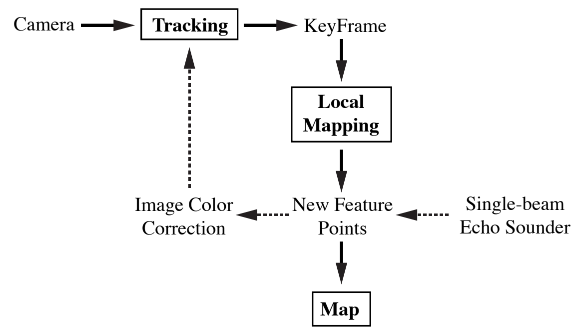


Fig. 2: Overview of the integration of the image color correction method and the single-beam echo sounder in the ORB SLAM system.

The depth readings from the echo sounder are relatively accurate, which can then be used to tune the tracked scene features in ORB-SLAM. The depth reading will continuously be used as a guiding point, not as a direct ground truth. In other words, it is an indicator of appropriate estimations. The resulting feature points and associated depth estimations will be more informative for the image color correction method.

The diagram in Fig. 2 displays our overall implementation that integrates with the ORB-SLAM system. The depth measurements from the single-beam echo sounder will be integrated in the local mapping when new scene feature points are created. Afterwards the image color correction method can apply the feature points and enhance the regions in the image that are tracked. The corrected images can then be incorporated back into the tracking step, as a substitute of the previous raw image frame, which can be described as a form of back propagation. As a possible example, each color corrected frame with its corresponding adjusted feature points will be fed back as a frame or key-frame reference for the incoming image frames.

IV. Preliminary Results

In this section, we show the fundamental results that will be used as building blocks for the final system. First, we will show the results of the image enhancement method applied on images of a color chart over different depths and distances. Then, the same image enhancement method is applied to images processed by the ORB-SLAM system. In the following section, we will discuss the plans of integrating measurements from an echo sounder into the image enhancement and SLAM processes. All experiments and data collection were performed at different locations in the Caribbean Sea off the coast of Barbados.

A. Color Chart Verification

To test the application of our image enhancement method, we deployed the BlueROV2 and used its installed Sony

depth	3.26 m	6.06 m	8.98 m	12.25 m	15.11 m
Raw					
Corrected					

TABLE I: Image enhancement over depth. First row: Raw images taken by the BlueROV2 in the Caribbean Sea. Second row: Our image color correction method [2].

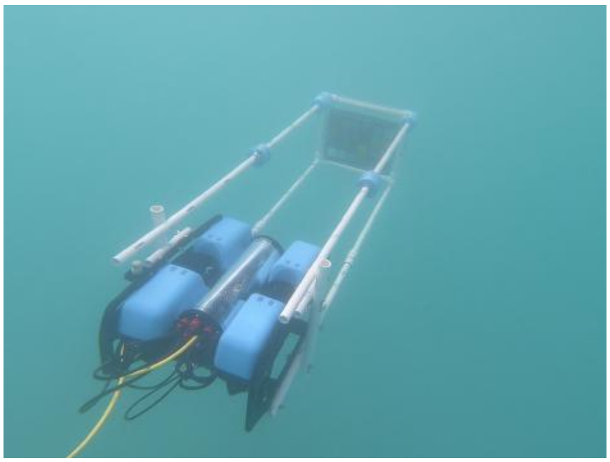


Fig. 3: BlueROV2 with color chart attached.

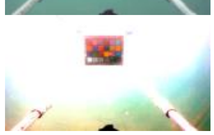
	0.33 m	**0.98 m**
Raw		
Corrected		

TABLE II: Image enhancement over viewing distance. Images taken at a depth of around $2\,\mathrm{m}$. Left column: color chart $0.33\,\mathrm{m}$ away. Right column: color chart $0.98\,\mathrm{m}$ away. First row: Raw images. Second row: Our image color correction method [2].

IMX322LQJ-C camera to collect images as it descended down to $25\,\mathrm{m}$. This camera has a resolution of 5 MP, a horizontal field of view of $80°$, and a vertical field of view of $64°$[1]. An X-Rite ColorChecker was attached to the robot at a set distance of $0.33\,\mathrm{m}$ from the camera. Table I displays the images that were taken by the camera at recorded depths of $3.26\,\mathrm{m}$, $6.06\,\mathrm{m}$, $8.98\,\mathrm{m}$, $12.25\,\mathrm{m}$, and $15.11\,\mathrm{m}$, as well as the results of applying our image color correction method. The attenuation values were estimated using the white and black color patches for each frame, while the wideband veiling light was calculated using Equation (2).

We also set the color chart at different distances from the BlueROV2 camera. Table II displays the images of the color chart set at two distances, $0.33\,\mathrm{m}$ and $0.98\,\mathrm{m}$. Both images were taken at a depth close to $2\,\mathrm{m}$. Fig. 3 illustrates the experimental setup when the color chart is set at 0.98 m away from the camera. In this case, as water conditions were poor, we assumed the veiling light to be the background color.

Comparison results with other image correction methods will not be described here, but we point the interested reader to [2] for such supplementary details. The visual results indicate the strengths of the image formation model, as well as the reliance on environment factors such as depth and viewing distance.

B. Depth Estimation using SLAM

Without the knowledge of the distances between the camera and the scene, the underwater image formation model becomes infeasible. The image depth can be estimated by feeding images through the ORB-SLAM system [7]. By slightly modifying the Monocular implementation to retrieve the estimated distances of the tracked feature points, we can integrate the depth estimations in the image formation model.

The method color corrects the pixels in the different patches, enhancing the color and increasing the contrast of the background and sand from the reef and fish. Table III displays the experimental image results. For further examination, please look at our previous work in [2].

V. Discussion and Future Steps

The applied image enhancement method can yield outputs of better visual quality or with higher color contrast that can be used for improving SLAM systems performance. Currently, our proposed method provides clearer and more accurate color contrast.

In the experiments described above, the depth measurements from the echo sounder were not integrated into the system. In the future, depth estimations will be adjusted using the Blue Robotics Ping Sonar Altimeter and Echosounder, which can be installed on the BlueROV2. This echo sounder has a frequency of $115\,\mathrm{kHz}$ and a range between $0.5\,\mathrm{m}$ and $30\,\mathrm{m}$, with a

[1]Further technical details can be found at https://www.bluerobotics.com/store/sensors-sonars-cameras/cameras/cam-usb-low-light-r1/

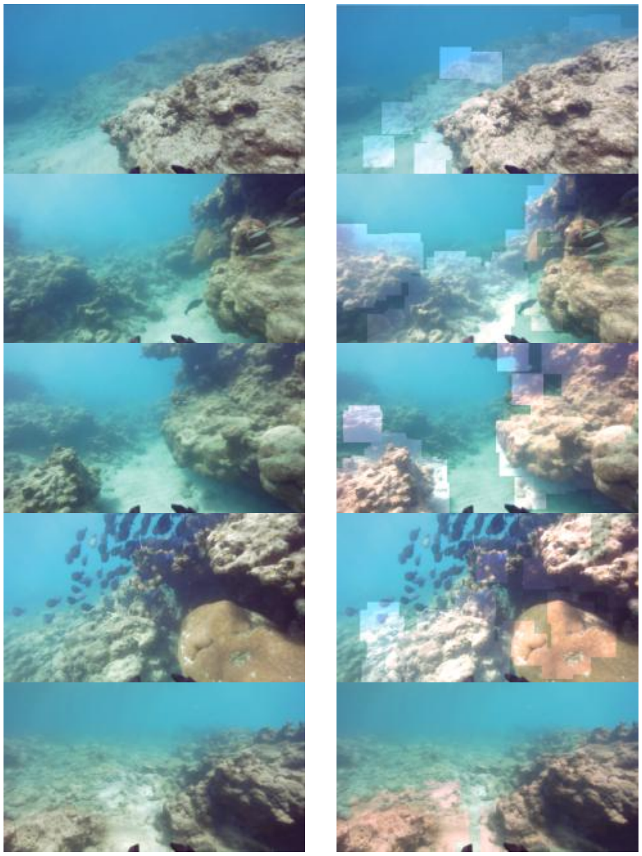


TABLE III: Image enhancement after ORB-SLAM [7] processing. Left: Raw. Right: Our image enhancement method [2], where patches in the image are color corrected according to the tracked features.

resolution of 0.5%[2]. It also provides confidence estimations of its readings, which is helpful when deciding if the readings should be integrated into the method at each frame.

It will be interesting to quantify the improvement in the state estimate with the enhanced images. We plan to collect images over different underwater areas where the visibility condition is different and apply different color correction techniques. One possible suggestion for collecting quantitative results is to compare the number of feature points detected in each frame between corrected and uncorrected images. We would also look into the localization accuracy when implementing such a system.

Note, the depth measurement is incorporated in the color correction method after feature extraction and then the color corrected image can be fed back to ORB-SLAM. A pre-corrected image can be provided to ORB-SLAM using the initial depth measurement before feeding to the SLAM system.

Currently, the depth measurement provides a single point around which patches can be extracted to and then color corrected in the image, as shown in the preliminary results. Our immediate future work will look at ways to apply that measurement to a neighborhood of points describing the same object and its estimate pose in the world. Finally, the system would integrate the color corrected images back into the SLAM system, hypothetically in the the tracking step as a form of back propagation. We would like to test the system in real world environments using two different types of robots, the BlueROV2, which was used in the preliminary results, and an Autonomous Surface Vehicle (ASV), custom made in our lab, coined as the Catabot.

[2]Further technical details can be found at https://www.bluerobotics.com/store/sensors-sonars-cameras/sonar/ping-sonar-r2-rp/

VI. Conclusion

We have discussed a new design of mutually improving a SLAM algorithm and image enhancement method by integrating low-cost sensors, a monocular camera and a single-beam echo sounder. This paper presents initial experiments to show the feasibility of this new design and discussion for future steps. More broadly, effectively using inexpensive sensors mounted on low-cost ROVs and AUVs will augment their autonomy increasing their applicability in many fields.

ACKNOWLEDGMENT

We would like to thank the reviewers for their valuable suggestions. This work is supported in part by the Dartmouth Burke Research Initiation Award.

References

[1] J. S. Jaffe, "Computer modeling and the design of optimal underwater imaging systems," *Journal of Oceanic Engineering*, vol. 15, 1990.

[2] M. Roznere and A. Quattrini Li, "Real-time model-based image color correction for underwater robots," *arXiv preprint arXiv:1904.06437*, 2019.

[3] R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," in *Machine Intelligence and Pattern Recognition*, 1986, pp. 1–26.

[4] R. P. S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," in *IEEE Trans. Aerosp. Electron. Syst.*, 2003, pp. 1152–1178.

[5] C. S. Lee, D. E. Clark, and J. Salvi, "SLAM with dynamic targets via single-cluster PHD filtering," *Journal of Selected Topics in Signal Processing*, vol. 7, no. 3, pp. 543–552, 2013.

[6] D. Ribas, P. Ridao, J. Neira, and J. D. Tardós, "SLAM using an imaging sonar for partially structured underwater environments," in *Proc. IROS*, 2006, pp. 5040–5045.

[7] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: a versatile and accurate monocular SLAM system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, 2015.

[8] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual–inertial odometry using nonlinear optimization," *The International Journal of Robotics Research*, vol. 34, no. 3, pp. 314–334, 2015.

[9] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[10] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras," *IEEE Trans. Robot.*, vol. 33, no. 5, pp. 1255–1262, 2017.

[11] S. Rahman, A. Quattrini Li, and I. Rekleitis, "SVIn2: Sonar visual-inertial SLAM with loop closure for underwater navigation," *CoRR*, vol. abs/1810.03200, 2018. [Online]. Available: http://arxiv.org/abs/1810.03200

[12] H. Strasdat, J. M. M. Montiel, and A. J. Davison, "Scale drift-aware large scale monocular slam," in *Proc. RSS*, 2010, pp. 73–80.

[13] G. Buchsbaum, "A spatial processor model for object colour perception," *Journal of the Franklin Institute*, vol. 310, no. 1, pp. 1–26, 1980.

[14] M. Chambah, D. Semani, A. Renouf, P. Courtellemont, and A. Rizzi, “Underwater color constancy: enhancement of automatic live fish recognition,” in *Color Imaging IX: Processing, Hardcopy, and Applications*, vol. 5293. International Society for Optics and Photonics, 2003, pp. 157–169.

[15] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, and B. Wang, “Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior,” in *IEEE Trans. Image Process.* IEEE, 2016, pp. 5664–5677.

[16] R. Garcia, T. Nicosevici, and X. Cufí, “On the way to solve lighting problems in underwater imaging,” in *Proc. OCEANS*, vol. 2. IEEE, 2002, pp. 1018–1024.

[17] S. Serikawa and H. Lu, “Underwater image dehazing using joint trilateral filter,” *Computers and Electrical Engineering*, vol. 40, no. 1, pp. 41–50, 2014.

[18] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, 2011.

[19] J. Y. Chiang and Y.-C. Chen, “Underwater image enhancement by wavelength compensation and dehazing,” *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1756–1769, 2012.

[20] H. Lu, Y. Li, and S. Serikawa, “Underwater image enhancement using guided trigonometric bilateral filter and fast automatic color correction,” in *Proc. ICIP*. IEEE, 2013, pp. 3412–3416.

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

[22] D. Akkaynak and T. Treibitz, “A revised underwater image formation model,” in *Proc. CVPR*, 2018, pp. 6723–6732.

[23] M. G. Solonenko and C. D. Mobley, “Inherent optical properties of Jerlov water types,” *Applied Optics*, vol. 54, pp. 5392–5401, 2015.

[24] P. A. Plonski, J. V. Hook, C. Peng, N. Noori, and V. Isler, “Navigation around an unknown obstacle for autonomous surface vehicles using a forward-facing sonar,” in *Proc. SSRR*. IEEE, 2015, pp. 1–7.

[25] E. Olson, “Apriltag: A robust and flexible visual fiducial system,” in *Proc. ICRA*, 2011, pp. 3400–3407.