Real-time Image Enhancement for Visual-Inertial SLAM in Underwater Scenarios

Team 22: Dongyao Zhang[1], Maithili Shetty[1], Moyan Lyu[1], Quintin Riley Harter[1] and Xinyi Liu[1]

[1]*University of Michigan, Ann Arbor, Michigan*

***Abstract*—Simultaneous localization and mapping (SLAM) in underwater environments pose major challenges due to limited lighting, turbidity, lack of relevant features, imprecision of sensors, presence of marine currents, etc. Therefore, in this project we realize a real-time image enhancement algorithm for visual simultaneous localization and mapping (SLAM) for autonomous underwater vehicles (AUV). Compared to atmospheric environments, underwater environment have issues with low contrast and color distortion. Our method focuses on underwater grayscale image enhancement for a real-world dataset by using a simplified enhancement model. In this project, the visual-inertial SLAM methods such as ORB-SLAM3 and Open Keyframe-based Visual-Inertial Odometry (OKVIS) are used. The results indicate that the enhanced images have better performance than the original images in feature points extraction and overall tracking performance.**

***Keywords*—ORB-SLAM3, OKVIS, image enhancement, underwater scenarios**

I. Introduction

In recent times, there has been an increasing demand for the exploration of underwater environments with the help of a remotely operated vehicle (ROV) or an autonomous underwater vehicle (AUV). These vehicles have become very popular to explore, collect data and reconstruct 3D maps of underwater environments. Simultaneous localization and mapping (SLAM) of these vehicles is challenging due to the unreliability of radio communication and global positioning systems (GPS). In such cases, visual images can provide useful information for the purpose of navigation. However, underwater visual images are strongly dependent on the water medium and are often severely degraded by poor conditions such as turbidity and tidal current, which results in a haze, color distortion, and low contrast.

Sensor fusion can be used to improve performance by merging information from different available sensors. In our project, we make use of tightly-coupled visual-inertial SLAM methods in order to safely localize and navigate the ROV. The combination of visual and inertial information provides robustness to poor textures, occlusions, and motion blur. Feature detection is an important component in VI SLAM or odometry systems [1]. These strategies rely on identifying features and determining their position in the images. VI SLAM methods mainly use features such as ORB [2], FAST [3], BRISK [4], SIFT [3], etc. An increased number of feature correspondences based on feature detection will lead to increased accuracy. However, from our previous discussion, visual information from underwater images is usually degraded which leads to a lower number of feature detections and thereby poor tracking performance. Therefore, a lot of recent works have focused on image enhancement and color correction techniques for underwater images. Reference [5] discusses the state-of-the-art image enhancement and color restoration techniques. In the context of SLAM, our objective is to integrate the existing image enhancement techniques [6] within the visual-inertial pipeline to obtain real-time performance. In particular, we integrate these methods within the ORB-SLAM3 and OKVIS pipeline and analyze the effects of image enhancement.

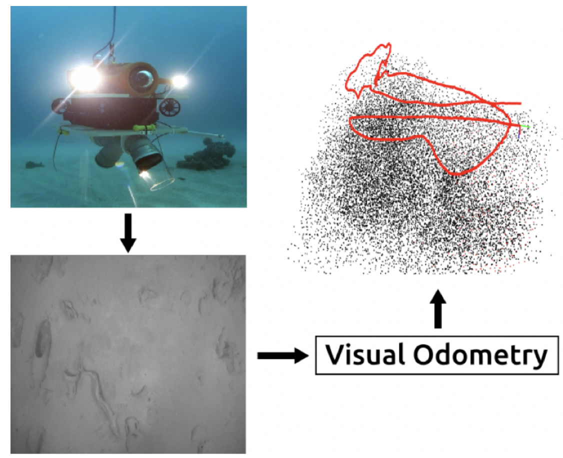


Fig. 1: ROV used with an underwater image of the seabed and the estimated trajectory in red [7]

The remainder of this report is organized as follows. Section II discusses our contributions in greater detail. The information on the dataset we used is provided in Section III. Section IV elaborates on the different SLAM and image enhancement techniques used in this project and their integration. Finally, Section V discusses the results of our project. We end our discussion by going over the concluding remarks and future work in Section VI.

II. Contributions

Our main contributions include:

- Application and performance analysis of the recently developed ORB-SLAM3 method in underwater environments.
- Integration of real-time image enhancement pipeline within ORB-SLAM3 and OKVIS.

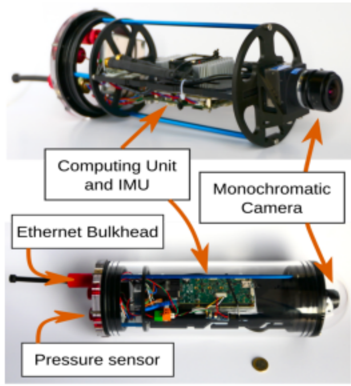


Fig. 2: Acquisition system

- Performance comparison based on absolute translational error (ATE).

III. Dataset overview

For testing the SLAM methods, we used a publicly available dataset called AQUALOC [8]. The data acquisition for this was performed using a ROV equipped with a monocular camera, a low-cost inertial measurement unit, a pressure sensor, a magnetometer and a computing unit all within a single enclosure as seen in Fig. 1. For the purpose of our project, we mainly focused on the camera and Inertial Measurement Units (IMU) information. The camera is equipped with a wide-angle lens and captures images at 20 frames per second with a resolution of 640×512 pixels. The IMU provides measurements at a rate of 200 Hz. It delivers measurements from a 3-axis accelerometer, 3-axis gyroscope and a 3-axis magnetometer.

The camera-IMU setup was calibrated to provide intrinsic parameters such as focal lengths, principal points and distortion coefficients. It also provides the relative position of the camera with respect to the IMU, and the time-delay between the camera and IMU as part of the extrinsic parameter. This calibration was performed using the Kalibr toolbox. It should be noted that the IMU calibration was performed in air to allow for faster motions that are required to correlate the IMU's measurements to the camera's.

The dataset consists of two sets of sequences. The archeological sequence is recorded in the Mediterranean sea, off Corsica's shore at a depth of approximately 270 meters. The second set, i.e., the harbor sequences was recorded at a depth of approximately 3-4 meters over an area of 100 m^2. In our project, we mainly focused on the harbor sequences. In these sequences, the vision was heavily degraded due to light absorption, strong illumination variations, and backscattering. As part of our image enhancement technique, we aim to correct these for better localization and mapping of the ROV.

IV. Technical Background

In this section, we present briefly the technical details relating to ORB-SLAM3 and OKVIS. We also discuss our applied image enhancement technique in greater detail.

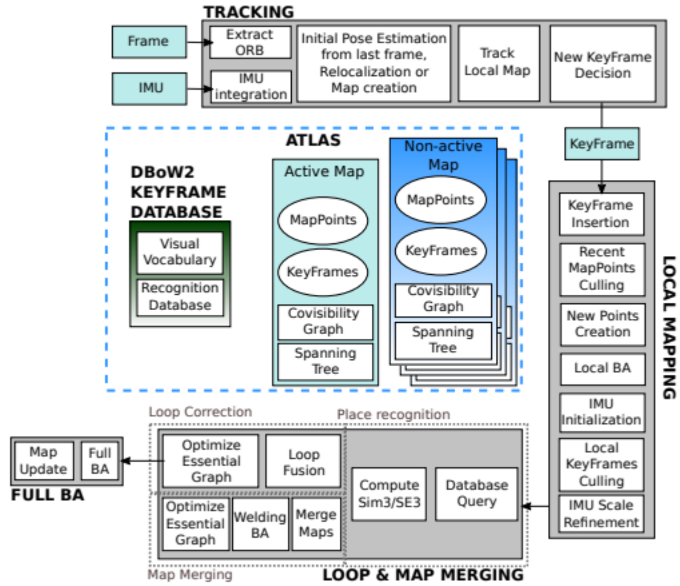


Fig. 3: ORB-SLAM3 system components

A. ORB-SLAM3

ORB-SLAM3 [2] is a tightly-integrated feature-based visual-inertial SLAM system. This is a multi-map and multi-session system capable of operating in pure visual and visual-inertial modes with monocular, stereo, and RGB-D camera sensors. Additionally, ORB-SLAM3 is able to survive long periods of poor visual information owing to its multi-map feature. The main components of the ORB-SLAM3 system are shown in Fig. 3.

As ORB-SLAM3 is an indirect SLAM method, low texture environments which are more often found in underwater scenarios account for the main failure case of this algorithm. Therefore, we aim to tackle this with the help of real-time image enhancement. Another challenge in the case of ORB-SLAM3 is the difficulty in IMU initialization in applications with slow motions, or without roll and pitch rotations. This is particularly true in the case of ROVs which leads to frequent tracking failures due to the lack of IMU initialization. Unfortunately, our current work does not cover the plausible solutions to overcome this challenge.

B. OKVIS

Open Keyframe-based Visual Inertial SLAM, or OKVIS, is a simultaneous localization and mapping algorithm that focuses on tightly integrating IMU measurements with camera frames. It also focuses on forming a trajectory primarily based on camera data, using IMU measurements between frames to correct. In addition, OKVIS uses an uncommon paradigm for frame tracking. It stores a "keyframe" that all new frames are compared to calculate the trajectory. When the matching ratio between the keyframe and a new frame drops below a threshold, a new keyframe is chosen. This allows for the body to move slowly or stop and the tracking will still remain stable. This is one reason we selected OKVIS for use in underwater scenarios, where such movement is common. Furthermore, because OKVIS relies so heavily on camera frames we hypothesized that image enhancement would significantly improve tracking.

C. Image Enhancement

Compared to land environments, the knowledge and technologies to analyze underwater scenarios are sparse due to the complexity of the underwater environment. Underwater light propagation is disturbed by suspended solids, dissolved solids, and other substances [9], resulting in issues like low contrast and blurred visual effect, which are not conducive to the extraction of image features. Since the camera is the vital environment sensing sensor to provide visual information for visual-inertial SLAM, the quality of images will directly determine the effect of positioning and mapping.

Among dehazing methods that have been developed to enhance the visibility of images, the most widely used Single Scattering Atmospheric Model (SSAM) models the haze image $I(x)$ as:

$$I(x) = J(x)t(x) + A(1 - t(x)), \tag{1}$$

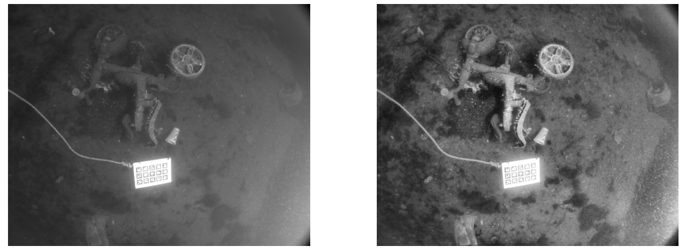
where x is a pixel, J is the haze-free image, A is the global airlight and $t(x) = \exp(-\beta d(x))$ is the transmission which represents color or light attenuation due to the scattering medium, determined by the scene depth $d(x)$ and the attenuation coefficient β.

To enable the model to deal with all degraded phenomena of underwater images, it needs to use light transfer characteristics underwater from particle physics. The unified model for underwater images uses non-uniform artificial light estimation, Point Spread Function (PSF) blur modeling, and random noises from underwater particles and is constructed as [5]:

$$\begin{aligned} I(x) &= [(E_A + E_L(x)) \cdot \rho(x)] * h_{psf}(x) \cdot \exp(-\beta d(x)) \\ &\quad + (1 - \exp(-\beta d(x))) \cdot E_L(x) + N \\ &= J(x) * h_{psf}(x) \cdot \exp(-\beta d(x)) \\ &\quad + (1 - \exp(-\beta d(x)) \cdot E_L(x) + N, \end{aligned} \tag{2}$$

here E_A is the global veiling light bias from the water surface, which is assumed to be constant, and E_L is the light from artificial light source. $J(x)$ is used for expressing reflected radiance for simplification. $\rho(x)$ is the reflectance of the surface. h_{psf} is the PSF kernel. N is the generated noise.

Based on the proposed model, the full enhancement would consist of denoising, artificial light estimation, plane induced dehazing, deblurring, and contrast enhancement. However, to process temporal images in real-time, we require a faster image processing method. Since the consecutive image pairs have a large overlap and small viewpoint changes, we adopt a simplified image processing method, which is based on unsupervised image enhancement. It consists of two parts, detail enhancement and contrast enhancement, which solve the problem of background light with noise and contrast degradation respectively.



(a) Unenhanced Image (b) Enhanced Image

Fig. 4: Effect of image enhancement

We first apply the dual guided image filtering method (GIF) to compute the detail layer $\Delta\mathbf{e}$ from foreground image (I_f) and background image (I_b) as

$$\Delta\mathbf{e} = I_f - I_b = \mathtt{GIF}_{k_f,\sigma_f}(I) - \mathtt{GIF}_{k_b,\sigma_b}(I), \tag{3}$$

where kernel sizes $(k_f = 2, k_b = 20)$ and smoothing degrees $(\sigma_f = 0.1^2, \sigma_b = 0.1^2)$ were chosen.

The detail enhanced image is then calculated as

$$I_{\mathtt{enhanced}} = I_b + \alpha \cdot \Delta\mathbf{e}, \tag{4}$$

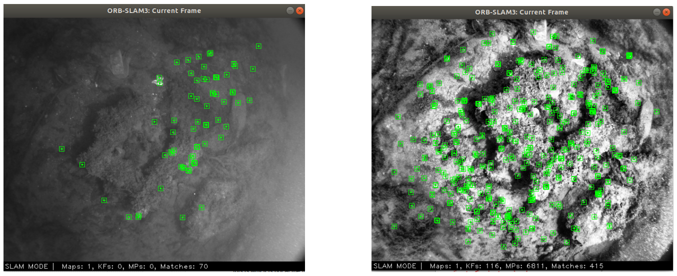
where α is the amplification factor that controls the level of details. In our project, we chose $\alpha = 5$. Then CLAHE is applied for contrast enhancement. Fig. 4 depicts the visible effect of this enhancement on a single underwater image.

V. Results

This section discusses the performance results of ORB-SLAM3 and OKVIS with and without image enhancement. The discussion on results is split into two parts. Firstly, the tracking performance with respect to the features matches is discussed for both the algorithms. Following this, we evaluate the trajectories based on RMS ATE and also discuss their CPU loads. The image enhancement code was developed in C++ for real-time integration with the SLAM systems. For the purpose of our project, we used the first harbor sequence from the AQUALOC dataset. This was chosen as it seemed to be representative of the general underwater visual effects such as turbidity and back-scattering. All our experiments have been run on an Intel Core i7-10750H CPU, at 2.6 GHz and 16 GB RAM. Our publicly available code can be found at: https://github.com/Maithilishetty/Mobile_Robotics_Team22. Our final presentation video can be accessed at: https://youtu.be/INbl_esfQ-Q.

A. Tracking Performance

In this section, we discuss the effect of image enhancement on tracking performance. Both OKVIS and ORB-SLAM3 rely on feature recognition and matching to calculate the estimated change in state from camera frames. Therefore, increased feature matching results in reduced failures and better tracking in both algorithms. The increase in the number of matches as a result of image enhancement for selected keyframes can be analyzed in Fig. 5 for ORB-SLAM3.

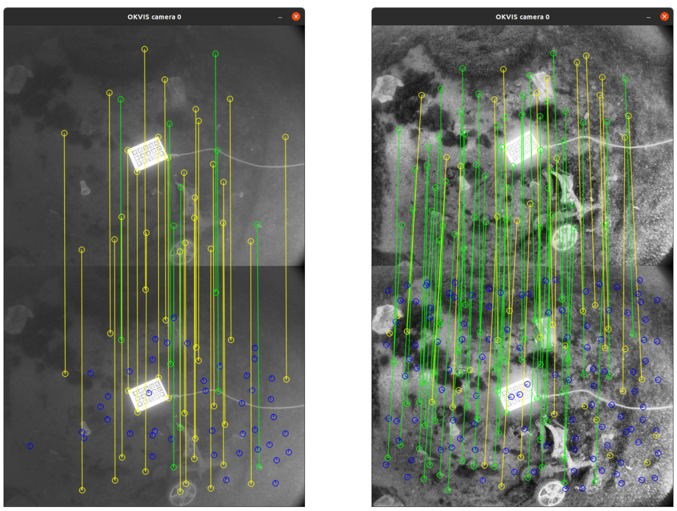


(a) Matches before enhancement

(b) Matches after enhancement

Fig. 5: Effect of image enhancement on feature detection for ORB-SLAM3

A similar observation can be drawn in the case of OKVIS as seen in Fig. 6.



(a) Matches before enhancement

(b) Matches after enhancement

Fig. 6: Effect of image enhancement on feature detection for OKVIS

Figures 8 and 7 show the match rate for all the frames of the harbor sequence. The number of matches found in both ORB-SLAM3 and OKVIS significantly increased when the versions with image enhancement were run. Furthermore, OKVIS no longer ever fell below the minimum matching threshold (shown in blue in the figures). This often occurred in the unenhanced case which led to a complete tracking failure. After enhancement, this was no longer an issue. It should be noted that the main tracking failure in the case of ORB-SLAM3 was due to the lack of IMU initialization in both the enhanced and unenhanced versions.

B. Evaluation Metrics

In this section, we provide a comparison of the SLAM trajectory estimation task with and without image enhancement. The global consistency can be evaluated by comparing the absolute distances between the estimated and ground

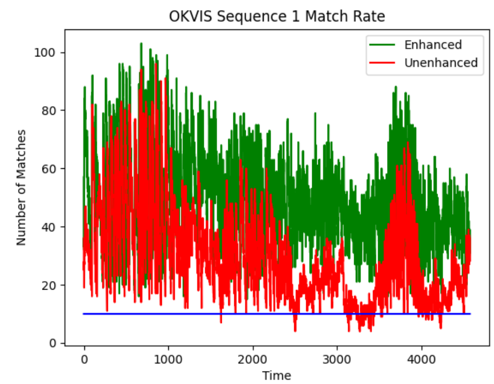


Fig. 7: OKVIS match rate result

truth trajectory. As both trajectories are specified in arbitrary coordinate frames, we use Sim(3) Umeyama's method [10] to align the trajectories with respect to the ground truth, as shown in Figures 9 and 10, then the trajectory estimate performance is evaluated based on Absolute Translational Error (ATE).

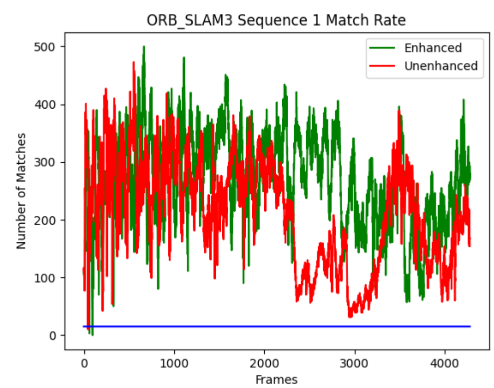


Fig. 8: ORB-SLAM3 match rate result

Method	Mean	RMSE	STD
Unenhanced OKVIS	1.782807	2.011744	0.932048
Enhanced OKVIS	0.727308	0.820046	0.378812
Unenhanced ORB-SLAM3	1.867629	2.190443	1.144554
Enhanced ORB-SLAM3	1.873155	2.091043	0.929383

TABLE I: Comparison between unenhanced and enhanced methods

The estimation visualization and the error statistics from Table. I show that the image enhancement technique can generally improve the trajectory estimation performance. Especially in the case of OKVIS, the estimated trajectory from enhanced image yields significantly lower ATE error.

Method	Process Rate	Max Frame Rate	CPU Load
Unenhanced OKVIS	16.31	61.31	-
Enhanced OKVIS	31.03	32.22	1.902x
Unenhanced ORB-SLAM3	55.07	18.15	-
Enhanced ORB-SLAM3	69.21	14.44	1.257x

TABLE II: Image enhancement CPU load

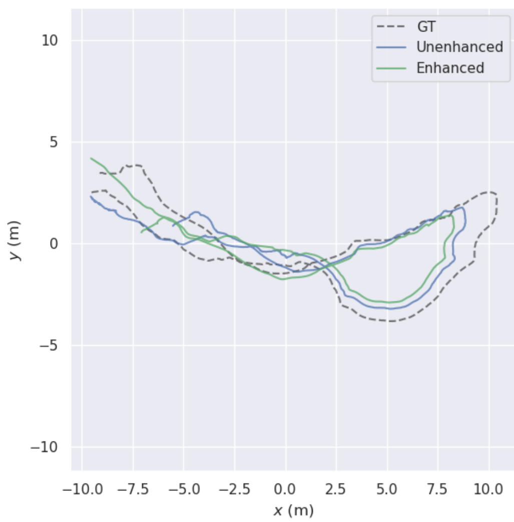


Fig. 9: ORB-SLAM3 trajectory estimation result

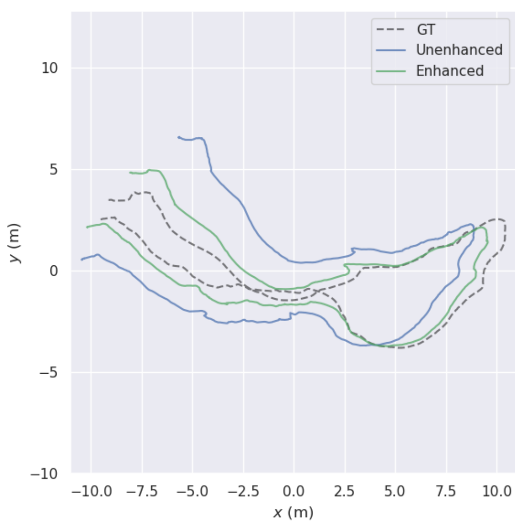


Fig. 10: OKVIS trajectory estimation result

An important factor to consider when implementing SLAM methods is the practicality of implementation on actual hardware. We ran several tests to compute the effect of our image enhancement algorithm on CPU load, this can be seen in Table. II. On our hardware, it increased the length of each frame calculation by about 14ms. As a result, the maximum camera frame rate supported for the enhanced version is 32.2 Hz for OKVIS and 14.44 Hz for ORB-SLAM3. Both of these frequencies are high enough that these algorithms remain suitable for the majority of real-time applications.

VI. Conclusion and Future Work

In this project, we propose a real-time image enhancement for visual-inertial SLAM, with an emphasis on underwater scenarios. The results show that the enhanced version performs better in both ORB-SLAM3 and OKVIS. By comparing the performance of original images in the experiment, the results confirm the main intuition that incorporating image enhancement effectively leads to higher performance in underwater visual-inertial SLAM due to better visual effect, feature matching, and tracking performance. Also, compared to the original images, real-time image enhancement significantly reduced full tracking failures. This shows that visual-inertial SLAM with enhanced vision has promising potential for the safe navigation of ROVs. The limitation of this project is that ORB-SLAM3 sometimes loses tracking between similar runs for both cases due to the lack of IMU initialization. The focus of the future work can be on reducing this tracking failure and on improving tracking results through a few of the methods discussed below:

A. Sensor Integration

Fusing acoustic sensors is an effective approach for robust visual-inertial SLAM in underwater environments. Due to poor visibility caused by suspended particles in water and a lack of light, many state-of-the-art approaches rely on acoustic sensing instead of vision for underwater navigation. Extension of the state-of-the-art ORB-SLAM3 SLAM system using motion priors calculated via acoustic odometry can result in improved performance. Acoustic information from a DVL, a magnetometer and an altimeter/depth sensor can be fused within the SLAM algorithms [11]. Our work will focus on a tightly-integrated Visual-Inertial-Pressure system for integrating depth measurements within the OKVIS framework.

B. Image Enhancement Improvements

We tested two alternative strategies, namely gamma correction [12] and unsharp masking [13] in addition to our current implementation. These strategies however failed to produce good results and have not been included in the report for the sake of brevity. In the future, other recently developed image enhancement and color restoration techniques can be implemented and could be compared with our current baseline. A modified visual-inertial SLAM using Generative Adversarial Networks (GANs) to enhance the quality of underwater images seems promising to get a better navigation and localization accuracy. However, it should be noted that there always exists a trade-off between the degree of enhancement and the cost of computation for obtaining a real-time performance.

C. Semantic-based SLAM

One of the key issues of these systems is to recognize the objects in the underwater environment. The future work may integrate a method to provide a semantic mapping using acoustic images acquired by forward-looking sonar (FLS) [14]. The method represents the environment using Gaussian probability density functions. We may create a semantic map of the scene and evaluate the method in a real dataset acquired by an underwater vehicle performing autonomous navigation and mapping tasks.

VII. Acknowledgement

We would like to thank the instructional team of EECS 568 for delivering this incredible course and for all their support during our project.

References

[1] C. Chen, H. Zhu, M. Li, and S. You, "A review of visual-inertial simultaneous localization and mapping from filtering-based and optimization-based perspectives," *Robotics*, vol. 7, no. 3, 2018.

[2] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. M. Montiel, and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.

[3] J. Hartmann, J. H. Klüssendorff, and E. Maehle, "A comparison of feature descriptors for visual slam," in *2013 European Conference on Mobile Robots*, pp. 56–61, 2013.

[4] 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.

[5] R. Schettini and S. Corchs, "Underwater image processing: State of the art of restoration and image enhancement methods," *Journal of Advanced Signal Processing*, vol. 746052 (2010), 2010.

[6] Y. Cho and A. Kim, "Visibility enhancement for underwater visual slam based on underwater light scattering model," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 710–717, 2017.

[7] F. Maxime, M. Julien, T.-P. Pauline, and C. Vincent, "Real-time monocular visual odometry for turbid and dynamic underwater environments," *Journal of Marine Science and Application*, 2018.

[8] M. Ferrera, V. Creuze, J. Moras, and P. Trouvé-Peloux, "Aqualoc: An underwater dataset for visual–inertial–pressure localization," *The International Journal of Robotics Research*, vol. 38, no. 14, pp. 1549–1559, 2019.

[9] L. Yan and L. Bo, "Underwater image bidirectional matching for localization based on sift," *Journal of Marine Science and Application*, vol. 13, no. 2, pp. 225–229,, 2014.

[10] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, pp. 376–380, 1991.

[11] E. Vargas, R. Scona, J. S. Willners, T. Luczynski, Y. Cao, S. Wang, and Y. R. Petillot, "Robust underwater visual slam fusing acoustic sensing," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2140–2146, 2021.

[12] S. Rahman, M. Rahman, and M. Abdullah-Al-Wadud, "An adaptive gamma correction for image enhancement," *Journal on Image and Video Processing*, 2016.

[13] L. Levi, "Unsharp masking and related image enhancement techniques," *Computer Graphics and Image Processing*, vol. 3, no. 2, pp. 163–177, 1974.

[14] P. N. M. Machado, P. Drews and S. Botelho, "Semantic mapping on underwater environment using sonar data," in *2016 XIII Latin American Robotics Symposium and IV Brazilian Robotics Symposium (LARS/SBR)*, pp. 245–250, 2016.