Structure from Motion of Underwater Scenes Considering Image Degradation and Refraction

Xiaorui Qiao * **Yonghoon Ji** ** **Astushi Yamashita** *
Hajime Asama *

* *The University of Tokyo,*
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
(e-mail: qiao, yamashita, asama@robot.t.u-tokyo.ac.jp).
** *Chuo University,*
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan
(e-mail: ji@mech.chuo-u.ac.jp)

Abstract: Structure from Motion (SfM) can reconstruct three-dimensional (3D) structures using only image sequences. However, SfM applied to the underwater environment is different from the air environment because image formation in underwater environments suffers from two major problems: image degradation and refraction. Consequently, images captured in underwater environments are loss of contrast, hazy and distorted geometrically. To achieve an accurate 3D reconstruction in such underwater environments, image degradation and refraction problems should be solved simultaneously. In this work, we propose a systematic SfM pipeline containing image enhancement and refraction reduced SfM. The image enhancement part improves the image quality for the following steps in the reconstruction part. We confirm the effectiveness of the proposed pipeline using images captured by a submerged robot in an extreme underwater environment, the disaster area of Unit 3 Primary Containment Vessel (PCV) at Fukushima Daiichi Nuclear Power Station. Experimental results show that our proposed method can successfully reconstruct the structures in that extreme underwater environment.

Keywords: Underwater robots, 3D Reconstruction, Structure from Motion, Image Enhancement, Refraction

1. INTRODUCTION

Three-dimensional (3D) reconstruction has been studied intensively for underwater environments (Murino et al. (1998); Brandou et al. (2007); Yang et al. (2013); Wang et al. (2016); Xu et al. (2016)). Structure from Motion (SfM), as a 3D reconstruction technique, is easy to implement because of using only a single camera. Generally, the camera is confined in a waterproof housing in underwater environments. Thus, the light rays are refracted twice before entering to the camera, one is at the water-housing interface, and the other is at the housing-air interface. Images captured from scenes suffer from distortion due to this refraction. Severe distortions would occur if the refractive distortions are not properly addressed.

Moreover, if floating particles and haze exist in the underwater environment, the environment can be extreme for carrying out vision tasks, such as detection, recognition and visual simultaneous localization and mapping (SLAM). Figure 1 depicts a typical underwater scene with image degradation and refraction. Suspended particles are floating at the object's surroundings, and the light ray is refracted twice before arriving at the image sensor. The quality of the captured image is reduced by floating particles and haze. The geometric distortion also occurs due to the refraction. Thus, the image formation in such an environment is quite different from that in an air environment. The image formation suffers from two major problems simultaneously: image degradation and refraction. Consequently, the images captured are loss of contrast, hazy and distorted geometrically. In the conventional SfM (Hartley and Zisserman (2003)), the image sequences are regarded as clear scenes, and the image formation process adopts the perspective camera model. In the extreme underwater environment, if the two major problems mentioned above are not properly addressed, there may serious problems

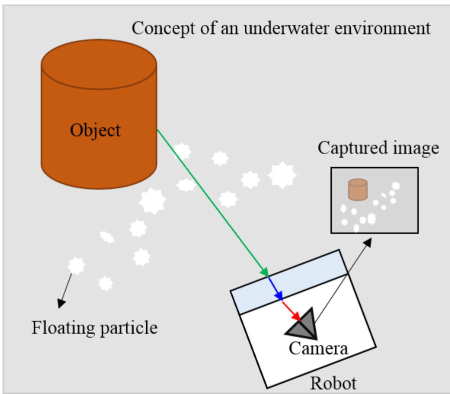

Fig. 1. Concept of an underwater environment.

happen in the 3D reconstruction. One is insufficient feature matches caused by the image degradation. The number of feature matches will affect the camera pose estimation and the number of sparse points in the 3D reconstruction. The other is distortion in 3D reconstruction caused by the refraction. Therefore, to achieve 3D reconstruction in extreme underwater environments, image degradation and refraction problems should be solved simultaneously.

In this work, we propose a novel SfM pipeline to solve the two problems simultaneously for underwater environments with image degradation and refraction. First, the image sequences are enhanced by the enhancement method in our previous work (Qiao et al. (2018)). Meanwhile, the camera system is modeled using ray tracing (Schmitt et al. (1988)) and Snell's law. Then, the camera pose can be estimated based on the modeled system using the enhanced images. Finally, the 3D reconstruction can be obtained by triangulation. Experiments using investigation video at Unit 3 Primary Containment Vessel (PCV) at Fukushima Daiichi Nuclear Power Station show that the proposed pipeline can solve the image degradation and refractive distortion problems simultaneously and reconstruct the object successfully.

The remainder of this paper is organized as follows. In Section 2, the previous work related to the underwater 3D reconstruction is presented. Section 3 presents an overview of the proposed SfM pipeline. Section 4 gives a detail description of the proposed pipeline. In Section 5, experiments using investigation video at Unit 3 PCV at Fukushima Daiichi Nuclear Power Station confirm the effectiveness of the proposed SfM pipeline. Section 6 presents the conclusion.

2. RELATED WORK

For 3D reconstruction in underwater environments, the related studies are mainly classified into two categories.

2.1 Considering Image Degradation

Research on solving the underwater image degradation problem is widely studied to reduce noise (Yamashita et al. (2006); Farhadifard et al. (2017)) and dehazing (Ancuti et al. (2012); Drews Jr et al. (2013)). 3D reconstruction for clear underwater environments usually does not consider the image enhancement as the first step process. However, in turbid underwater environments, image degradation problem should be firstly taken into consideration. Li et al. (2015) presented a method to estimate scene depth and dehazing using a foggy video sequence simultaneously. In this study, the depth cues from stereo matching and fog information reinforce each other and outperform the conventional stereo or dehazing algorithms. Xu et al. (2016) developed an underwater 3D object reconstruction model for the continuous video stream. In this study, image enhancement is taken into consideration as the first step. However, they used the guided filter to acquire clear images. For haze and low contrast underwater images, only the guided filter is unable to achieve satisfactory restoration.

2.2 Considering Refraction

To deal with the refractive problem, the camera systems were modeled by approximate methods, such as focal length adjustment (Ferreira and J. P. Costeira (2005)), lens radial distortion (Pizarro et al. (2003)) and a combination of the two (Lavest et al. (2000)) in early works. However, due to the non-linear camera system, these methods are inadequate to solve the refraction problem.

Refractive Structure from Motion (RSfM), which models the camera system based on ray tracing and Snell's law, has been studied in recent years. Most of studies focus on the flat shape camera housing (Jordt-Sedlazeck and Koch (2013); Shibata et al. (2015, 2018)). However, a waterproof housing can be designed with various shapes, depending on the application, such as cylindrical camera housing equipped on the robot sent to Unit 3 PCV for investigation (Adachi et al. (2018)). Consequently, specific constraints for a flat camera housing, such as the case where all the light rays are on the same plane, will be invalid in a general situation. Thus, the camera system modeling, calibration method and camera pose estimation method should be extended to a more general situation.

Gedge et al. (2011) proposed refractive epipolar geometry to find stereo correspondences for underwater image matching. The corresponding point of a pixel in one image is restricted to a curve, not a line. Jordt-Sedlazeck and Koch (2013) considered the refractive effects of thick flat housing for 3D reconstruction, and virtual camera errors were proposed for the bundle adjustment. Kang et al. (2017) proposed two new concepts, "Ellipse of Refrax" and the "Refractive Depth" of a scene point, and is able to simultaneously perform underwater camera calibration and 3D reconstruction automatically without using any calibration object. However, the camera rotation should be known first.

There are multitude methods can achieve the 3D structures in underwater environments, a miniature single-camera system can be achieved using SfM technique for implementation in a narrow space. To achieve an accurate underwater 3D reconstruction, image degradation and refractive distortion problems should be solved simultaneously. Generally, the current studies focus on only one problem, image degradation or refraction. Image enhancement is essential in turbid water environments because the low image quality affects the feature extraction in the SfM pipeline. Refraction always exists when the camera capturing images in water. The current studies mainly focus on flat refraction. The camera housing can be designed with various shapes depending on the application. Thus, the method considering refraction in general camera housing is acquired. Therefore, underwater SfM is different from the conventional SfM due to the image degradation and refraction. Thus, an approach containing these two factors is required to obtain accurate 3D reconstruction.

3. OVERVIEW OF THE PROPOSED SFM PIPELINE

As shown in Fig. 2, the process is the proposed SfM pipeline considering image degradation and refraction. The blue parts are different compared with the conventional SfM pipeline. Based on the two major problems

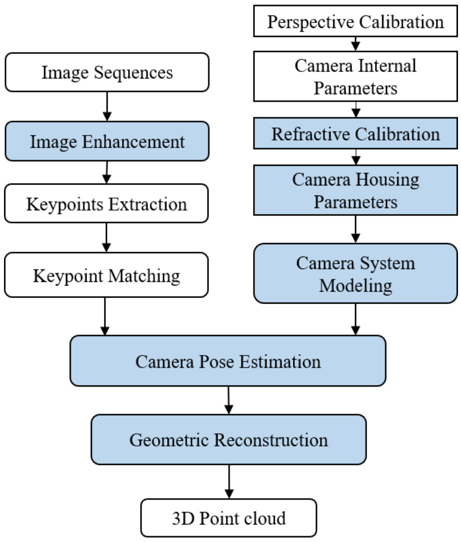

Fig. 2. Proposed SfM pipeline.

existing in underwater environments, the different parts are summarized into two aspects. First, the image enhancement part is added to the pipeline to solve the image degradation problem and improve the image quality in underwater environments. The objective of this part to achieve "clear" images for feature detection and feature matching. Second, refraction is considered in the 3D reconstruction part. The corresponding steps, such as calibration, pose estimation, and geometric reconstruction, use the modeled camera system to eliminate the refractive effects on these steps. In the refractive calibration part, the parameters between a camera and its housing are estimated. These parameters can be known parameters for the camera system modeling. Then, based on the modeled camera system, the camera poses can be estimated for the following 3D reconstruction.

4. DETAIL OF THE PROPOSED SFM PIPELINE

4.1 Underwater Image Enhancement

In the underwater environments, occlusion noises, such as floating particle and water bubble, and haze exist. The main process of image enhancement is as depicted in Fig. 3. There are two major steps for image enhancement. One is occlusion noise removal based on multi-frame method, and

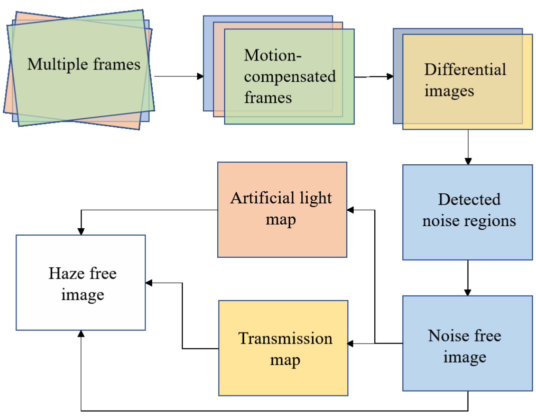

Fig. 3. Image enhancement flowchart.

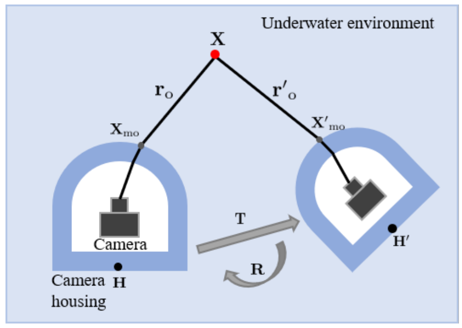

Fig. 4. Conceptual image of RSfM.

the next step is the haze removal based on the single frame estimation.

Occlusion Noise Removal Multiple frames can be obtained from the video sequences. There is slight motion between the consecutive frames due to the moving of the underwater vehicle. The motion compensation is implemented to achieve the aligned frames. Thus, the differential images can be acquired by subtraction and the setting threshold. The noise regions can be detected based on chromatic properties and Combined Temporal Discontinuity Field (CTDF) (Qiao et al. (2018)). The detected noise regions can be restored by background pixels.

Haze Removal The noise-free image is the input of the dehazing process. The artificial light map is estimated based on maximum reflectance prior (Zhang et al. (2017)), and transmission map can be estimated based on the red channel. Then, the haze can be removed based on the underwater light model (Jaffe (1990)).

4.2 RSfM

Refraction is considered in the calibration, camera modeling and camera pose estimation steps.

Refractive Calibration As shown in Fig. 4, a camera and a housing are called "local camera system" in this study. The center of the camera system is set at $\mathbf{H}$. Estimation of the relative pose of the camera to the camera housing is called refractive calibration. After the refractive calibration, the local camera system can be modeled accurately. The back-projection (from a image pixel to a 3D point) can be represented based on ray tracing (Schmitt et al. (1988)) and Snell's law. A checkerboard pattern is used as the method in Zhang (2000). When the checkerboard is captured by the camera system. A corner point on the checkerboard, projecting to the image sensor, corresponds a image pixel. If we project the image pixel back to the corner point, there is an error between a back-projected point and the corresponding ground truth point on the checkerboard. Minimization of the back-projection error is used to estimate the unknown parameters.

Camera System Modeling After obtaining the parameters of the local camera system, the camera system can be modeled by ray tracing (Schmitt et al. (1988)) and Snell's law. As shown in Fig. 4, for each feature point, there is a corresponding outer ray. Thus, the corresponding outer rays can be used for the relative camera pose estimation.

Camera Pose Estimation The camera pose estimation is based on the modeled camera system. A camera system with a waterproof housing is shown in Fig. 4. The origin of local camera coordinate system $\mathbf{H}$ is set on camera housing center. After the camera housing calibration, the relative pose of camera to camera housing is estimated, and the outer intersected point and outer ray can be obtained by ray tracing (Schmitt et al. (1988)) and Snell's law. The camera system undergoes a rotation $\mathbf{R}$ and a translation $\mathbf{T}$. The first camera local camera system is assumed to be the world coordinate system origin. In first local camera system, the outer line can be denoted by

$$\mathbf{L} = (\mathbf{d}, \mathbf{m}) = (\mathbf{r}_{\mathrm{o}}, \mathbf{X}_{\mathrm{mo}} \times \mathbf{r}_{\mathrm{o}}), \tag{1}$$

where $\mathbf{d}$ and $\mathbf{m}$ is the direction and moment of the plücker line, respectively.

Similarly, in the second local camera system, the outer line can be denoted by

$$\mathbf{L}' = (\mathbf{d}', \mathbf{m}') = (\mathbf{r}'_{\mathrm{o}}, \mathbf{X}'_{\mathrm{mo}} \times \mathbf{r}'_{\mathrm{o}}). \tag{2}$$

In order to determine the relative camera pose of the second camera, the plücker vectors of the outer line in the second coordinate system is transformed to the world coordinate system:

$$\mathbf{d}'^{\mathrm{w}} = \mathbf{R}\mathbf{r}'_{\mathrm{o}}, \tag{3}$$

and

$$\begin{aligned} \mathbf{m}'^{\mathrm{w}} &= (\mathbf{R}\mathbf{X}'_{\mathrm{mo}} + \mathbf{T}) \times (\mathbf{R}\mathbf{r}'_{\mathrm{o}}) \\ &= \mathbf{R}\mathbf{m}' + [\mathbf{T}]_{\times} \mathbf{R}\mathbf{r}'_{\mathrm{o}}. \end{aligned} \tag{4}$$

Thus, the outer line $\mathbf{L}'$ in the world coordinate system can be written as:

$$\mathbf{L}'^{\mathrm{w}} = (\mathbf{R}\mathbf{r}'_{\mathrm{o}}, \mathbf{R}\mathbf{m}' + [\mathbf{T}]_{\times} \mathbf{R}\mathbf{r}'_{\mathrm{o}}). \tag{5}$$

Plücker lines $\mathbf{L}$ and $\mathbf{L}'^{\mathrm{W}}$ intersect at point $\mathbf{X}$. The intersection both lines is determined by

$$\mathbf{d}^{\top}\mathbf{m}'^{\mathrm{w}} + \mathbf{m}^{\top}\mathbf{d}'^{\mathrm{w}} = 0. \tag{6}$$

The following equation can be obtained according to (4), (5), and (6).

$$\begin{aligned} &\mathbf{r}_{\mathrm{o}}{}^{\top} \left(\mathbf{R}\mathbf{m}' + [\mathbf{T}]_{\times} \mathbf{R}\mathbf{r}'_{\mathrm{o}}\right) + \mathbf{m}^{\top}\mathbf{R}\mathbf{r}'_{\mathrm{o}} \\ &= \begin{pmatrix} \mathbf{r}_{\mathrm{o}} \\ \mathbf{m} \end{pmatrix}^{\top} \underbrace{\begin{pmatrix} [\mathbf{T}]_{\times}\mathbf{R} & \mathbf{R} \\ \mathbf{R} & \mathbf{0}_{3\times 3} \end{pmatrix}}_{\mathbf{E}_G} \begin{pmatrix} \mathbf{r}'_{\mathrm{o}} \\ \mathbf{m}' \end{pmatrix} \\ &=0, \end{aligned} \tag{7}$$

where $[\mathbf{T}]_{\times}$ is the skew-symmetric matrix of $\mathbf{T}$, $\mathbf{E}_G$ is the generalized essential matrix. The equation is called the generalized epipolar constraint (Pless (2003)). Therefore, we can calculate $\mathbf{T}$ and $\mathbf{R}$ by (7).

Geometric Reconstruction After obtaining the estimated poses and outer rays, the 3D reconstruction can be achieved by triangulation methods (Hartley and Sturm (1997)).

5. EXPERIMENTS

Experiments using real images were implemented to confirm the effectiveness of the proposed method. In the experiments, the images captured by the submersible robot sent to Unit 3 PCV at Fukushima Daiichi Nuclear Power Station were used. The inside environment of Unit 3 PCV is extreme because the water is filled with floating particles and turbid. We use the proposed SfM pipeline to reconstruct a typical scene in this underwater environment as shown in Fig. 5

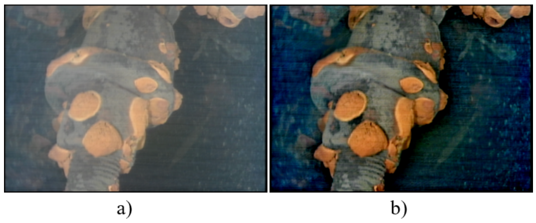

Fig. 5. Enhanced result: a) the original image, b) the enhanced result.

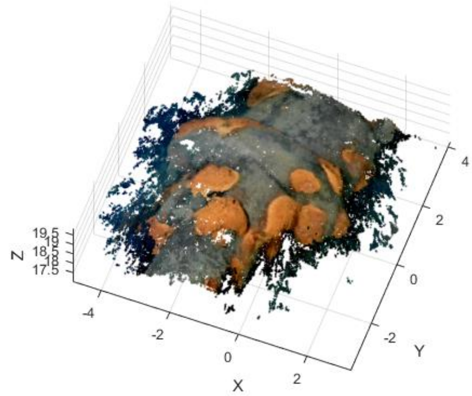

Fig. 6. Reconstructed object in the Unit 3 PCV at Fukushima Daiichi Nuclear Power Station.

5.1 Image Enhancement

Figure. 5 shows an example of the enhanced result by proposed method. The enhanced image has more contrast, less noise and less haze compared with the original image. The image quality is improved obviously.

5.2 Two-view Reconstruction

Two enhanced frames are selected for the 3D reconstruction. The dense optical flow (Farnebäck (2003)) method is applied to achieve each pixel's correspondence for two consecutive frames. We remove the outliers in each local patch. Thus, the dense reconstruction can be obtained. As shown in Fig. 6, the object in the scene can be successfully reconstructed using the proposed method.

6. CONCLUSION

We herein proposed a novel SfM pipeline for underwater environments to address image degradation and refraction problems. To solve the two problems simultaneously, the image enhancement and refractive SfM methods are proposed for the 3D reconstruction. We confirmed the effectiveness of the proposed SfM pipeline using the image data from an extreme underwater environment, the inside of

Unit 3 PCV at Fukushima Daiichi Nuclear Power Station. The results show that the proposed pipeline can reconstruct the underwater scene successfully with enhanced images. It also shows the effectiveness of the proposed method in underwater environments with image degradation and refraction.

ACKNOWLEDGEMENTS

The authors would like to thank IRID and TOSHIBA for supporting this research. This work in part was supported by JSPS KAKENHI Grant Number JP18H03309.

REFERENCES

Adachi, H., Okada, S., and Matsuzaki, K. (2018). Investigation robots for inside primary containment vessel of fukushima daiichi nuclear power station. *Journal of the Robotics Society of Japan*, 36(6), 395–398.

Ancuti, C., Ancuti, C.O., Haber, T., and Bekaert, P. (2012). Enhancing underwater images and videos by fusion. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 81–88.

Brandou, V., Allais, A.G., Perrier, M., Malis, E., Rives, P., Sarrazin, J., and Sarradin, P.M. (2007). 3d reconstruction of natural underwater scenes using the stereovision system iris. In *Proceedings of OCEANS 2007 - Europe*, 1–6.

Drews Jr, P., do Nascimento, E., Moraes, F., Botelho, S., and Campos, M. (2013). Transmission estimation in underwater single images. In *2013 IEEE International Conference on Computer Vision Workshops*, 825–830.

Farhadifard, F., Radolko, M., and von Lukas, U.F. (2017). Single image marine snow removal based on a supervised median filtering scheme. In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4: VISAPP*, 280–287.

Farnebäck, G. (2003). Two-frame motion estimation based on polynomial expansion. In *Scandinavian conference on Image analysis*, 363–370.

Ferreira, R. and J. P. Costeira, J.S. (2005). Stereo reconstruction of a submerged scene. In *Proceedings of the 2005 Iberian Conference on Pattern Recognition and Image Analysis*, 102–109.

Gedge, J., Gong, M., and Yang, Y. (2011). Refractive epipolar geometry for underwater stereo matching. In *Proceedings of 2011 Canadian Conference on Computer and Robot Vision*, 146–152.

Hartley, R. and Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge university press.

Hartley, R.I. and Sturm, P. (1997). Triangulation. *Computer vision and image understanding*, 68(2), 146–157.

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

Jordt-Sedlazeck, A. and Koch, R. (2013). Refractive structure-from-motion on underwater images. In *Proceedings of the 2013 IEEE International Conference on Computer Vision*, 57–64.

Kang, L., Wu, L., Wei, Y., Lao, S., and Yang, Y. (2017). Two-view underwater 3d reconstruction for cameras with unknown poses under flat refractive interfaces. *Pattern Recognition*, 69, 251–269.

Lavest, J.M., Rives, G., and Lapresté, J.T. (2000). Underwater camera calibration. In *Proceedings of the 2000 European Conference Computer Vision*, 654–668.

Li, Z., Tan, P., Tan, R.T., Zou, D., Zhou, S.Z., and Cheong, L.F. (2015). Simultaneous video defogging and stereo reconstruction. In *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*, 4988–4997.

Murino, V., Trucco, A., and Regazzoni, C.S. (1998). A probabilistic approach to the coupled reconstruction and restoration of underwater acoustic images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1), 9–22.

Pizarro, O., Eustice, R.M., and Singh, H. (2003). Relative pose estimation for instrumented, calibrated imaging platforms. In *Proceedings of the 2003 Digital Image Computing Techniques and Applications*, 601–612.

Pless, R. (2003). Using many cameras as one. In *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, II–587–93.

Qiao, X., Ji, Y., Yamashita, A., and Asama, H. (2018). Visibility enhancement for underwater robots based on an improved underwater light model. *Journal of Robotics and Mechatronics*, 30, 781–790.

Schmitt, A., Müller, H., and Leister, W. (1988). Ray tracing algorithms–theory and practice. In *Theoretical Foundations of Computer Graphics and CAD*, 997–1030.

Shibata, A., Fujii, H., Yamashita, A., and Asama, H. (2015). Absolute scale structure from motion using a refractive plate. In *Proceedings of the IEEE/SICE International Symposium on System Integration*, 540–545.

Shibata, A., Okumura, Y., Fujii, H., Yamashita, A., and Asama, H. (2018). Refraction-based bundle adjustment for scale reconstructable structure from motion. *Journal of Robotics and Mechatronics*, 30, 660–670.

Wang, Y., Negahdaripour, S., and Aykin, M.D. (2016). Calibration and 3d reconstruction of underwater objects with non-single-view projection model by structured light stereo imaging. *Applied Optics*, 55(24), 6564–6575.

Xu, X., Che, R., Nian, R., He, B., Chen, M., and Lendasse, A. (2016). Underwater 3d object reconstruction with multiple views in video stream via structure from motion. In *Proceedings of OCEANS 2016 - Shanghai*, 1–5.

Yamashita, A., Kato, S., and Kaneko, T. (2006). Robust sensing against bubble noises in aquatic environments with a stereo vision system. In *Proceedings 2006 IEEE International Conference on Robotics and Automation*, 928–933.

Yang, Y., Zheng, B., Zheng, H., Wang, Z., Wu, G., and Wang, J. (2013). 3d reconstruction for underwater laser line scanning. In *2013 MTS/IEEE OCEANS - Bergen*, 1–3.

Zhang, J., Cao, Y., Fang, S., Kang, Y., and Chen, C.W. (2017). Fast haze removal for nighttime image using maximum reflectance prior. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 7016–7024.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), 1330–1334.