Smart solutions for underwater 3D exploration and SLAM – possibilities and limitations

Christian Bräuer-Burchardt [1], Martin Hubold [1], Matthias Heinze [1], Peter Kühmstedt [1], Gunther Notni [1,2]

[1] Fraunhofer Institute for Applied Optics and Precision Engineering, 07745 Jena, Germany – christian.braeuer-burchardt@iof.fraunhofer.de, [martin.hubold, matthias.heinze, peter.kuehmstedt, gunther.notni]@iof.fraunhofer.de
[2] Machine Engineering Faculty, Technical University Ilmenau, Ehrenbergstraße 29, D-98693 Ilmenau, Germany; gunther.notni@tu-ilmenau.de

Keywords: Underwater 3D Scanning and SLAM, Monocular 3D Reconstruction, Multi-Aperture Camera, AI.

Abstract

Monocular 3D exploration of underwater objects is a growing field of application due to powerful new developments of computer technology, hardware components, and algorithms such as AI-supported modules. Using appropriate software tools, video streams of simple cameras may produce complete 3D models. This technique makes it straightforward and convenient to collect image data and produce 3D models of certain objects. In this work we introduce several approaches for further simplification of underwater camera systems for production of 3D models of underwater structures. The principles of monocular 3D reconstruction using image sequences and those of multi-aperture camera technology are described and their use for underwater applications are discussed. The initial experimental results are presented, along with an outlook to approaches of small and lightweight powerful 3D reconstruction systems for (not exclusively) diver use.

1. Introduction

Underwater 3D exploration is used for a variety of applications. Throughout the development of underwater 3D measurement technology, various measurement principles have become established. Whereas rough distance estimations to subsea obstacles are performed using sonar techniques (Guerneve and Pettilot, 2015), energy production structures inspection may require high accuracy 3D surface measurements, obtained by photogrammetry or laser scanning (McLeod et al., 2014). Other applications include the documentation of archaeological heritage sites (Eric et al., 2013), reconstruction of sunken shipwrecks (Balletti et al., 2015), estimation of the mass of certain fish species (Costa et al., 2006), or underwater cave exploration (Campbell, 2018; Massone et al., 2021). The necessity for varying levels of accuracy in different applications require the employment of suitable reconstruction principles. Optical methods such as stereo scanning are increasingly applied in connection with artificial diffuse or structured illumination (Bräuer-Burchardt et al., 2024; Lin et al., 2025).

Due to the ever-increasing challenges in the development of maritime resources and the requirements for energy efficiency and climate protection, the development of a user-friendly, lightweight, and low-budget measurement system for underwater 3D measurements for a wide range of applications would help to offer measurement systems for many users. New developments in camera technology, illumination systems, and innovative algorithms including tools using artificial intelligence (AI) provide new possibilities for construction and realization of low budget, lightweight, and user-friendly systems for underwater 3D exploration.

First examples towards such systems have been presented recently (Gaglianone et al., 2018).

In this work, we endeavor to show approaches to achieve appropriate solutions. We suggest combining monocular video streams, miniature camera techniques, and additional use of AI. We hope to use these to inspire further research topics.

2. New Approaches for Underwater 3D Exploration

2.1 3D Reconstruction Using Monocular Video Streams

In order to achieve a compact design of the measurement system and avoid unnecessary weight, the approach of monocular 3D reconstruction is pursued instead of the classic stereo camera arrangement. In this case, images of the measurement scene are captured sequentially from different positions by the camera and then processed offline in a post-processing step to create a 3D model.

Several approaches to obtain monocular 3D reconstruction in air have been introduced recently. Several authors use motion for 3D reconstruction (Wandt et al., 2014; Kokkinos & Kokkinos, 2021), additional illumination (Wu et al., 2020), or inertial tracking components (Yang et al., 2020). Systems for underwater applications face far greater challenges. An example of an underwater system for deep sea application is presented by Kwasnitschka et al. (2016). This monocular system is used for mapping the seafloor and uses additional information such as velocity of the carrier AUV and acoustic distance estimation for 3D reconstruction.

Monocular underwater exploration provides many advantages over conventional stereo approaches. In addition to a considerable reduction of volume and weight, a compact realization of the complete system is possible. Illumination of the scene has fewer restrictions and no calibration between two cameras to determine the exact orientation between them is necessary. However, there are also some disadvantages. Measurement accuracy is usually lower than that of stereo systems, and when producing maps by SLAM (Simultaneous Localization and Mapping) algorithms, systematic measurement errors may cumulate.

A monocular 3D reconstruction system for underwater application called goSEA3D was developed at our institute. It generates 3D information by triangulation of sequential single images supported by pose information using a tracking unit.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-2-W10-2025-27-2025

The system consists of a color camera, a tracking camera, and a control unit. The power supply is realized by a rechargeable battery which is housed with the control unit separate from the handheld scanning unit. Batteries and control unit are connected to the scanning unit by one cable. The color camera is of type Baumer VCXG-204C with a 20 MPix resolution and a frame rate of 6 fps. The tracking unit collects position, orientation, and velocity of the sensor at approximately 150 Hz. The camera and tracking unit are installed in a housing designed for underwater use (Figure 1). The underwater objects are illuminated by an independent spotlight, which can also be attached to the underwater housing. Size (290 mm length and 230 mm diameter) and weight (about 13.4 kg in air) of the handheld scanning unit have not yet been optimized. The system is constructed for diver use at operating depths up to 20 meters.

Figure 1. Monocular underwater 3D system goSEA3D.

The tracking unit is a low-power, standalone SLAM device – the Intel® RealSense™ Tracking Camera T265. It consists of two fisheye cameras, an inertial measurement unit (IMU), and an image processing unit. The tracking unit enables low-latency estimation of both position and orientation of the goSEA3D scanner. All captured camera images can be reliably oriented in unknown environments without additional markers or external tracking devices, using only positions and orientations.

The process of 3D data generation is split into acquisition with the scanning unit and the subsequent 3D reconstruction on a workstation. At recording, the scanning unit continuously takes up to six camera images per second while the user guides the camera perspective over the object's surface. The goSEA3D system provides feedback regarding brightness and motion speed to avoid images of poor quality. Exposure time may be set to short values below 3 ms to reduce image blur caused by motion during exposure.

The tracking unit estimates its own motion and provides a trajectory of the sensor's temporally changing poses with approximately 150 Hz. The coordinate systems from the color camera and the tracking unit are linked, and the calculated 3D data are stored in a common world coordinate system. Matching pose data and camera images are identified using time stamps. The well-known photogrammetric principle is used for the generation of the 3D model (see, e.g., Luhmann et al., 2006). It is performed in three major steps: alignment, 3D mesh generation, and texture bending.

In the first step, the accurate alignment of the camera poses is determined by bundle block adjustment. In the goSEA3D system this processing is sped up by using the data of the tracking camera. Only images with positions close to each other are selected for image feature matching and, the poses of the tracking camera give an initial solution for the bundle block adjustment.

In the second step, the complete full 3D surface model of the object is reconstructed. Depth maps are calculated using a multi-view stereo method for mesh generation. Distance calculation and averaging is realized using up to 16 neighbouring images. This results in a depth map for each selected image and all depth maps build the full 3D model of the object.

The third step includes the mapping of the texture as a layer on the 3D model. All three steps are realized in a automated workflow using the commercial photogrammetry software "Agisoft Metashape" (Agisoft, 2025). Examples of reconstruction results of underwater object measurements are presented in section 3.

2.2 Ultra Compact Multi-Aperture Camera Technique

Although high-resolution camera systems with high-quality lenses can already be relatively small, the development of unobtrusive cameras, such as those designed to be attached to a diver's head using a forehead mount should prioritize minimal size and weight. Multi-aperture camera (MAC) technology (Hubold et al., 2021) may soon offer alternatives to existing camera systems. These cameras have an array of microlenses instead of a single lens. Each microlens objective images the entire field of view onto tiled areas of the image sensor without the need for an additional lens.

With a defined field of view of the entire system, the focal length of the microlenses may be only a fraction of the focal length of a single-aperture system. This can reduce the overall length and optical system design complexity and significantly increase the depth of field. Figure 2 shows the imaging principle of a MAC and Figure 3 shows the MAC and several components of the system used here.

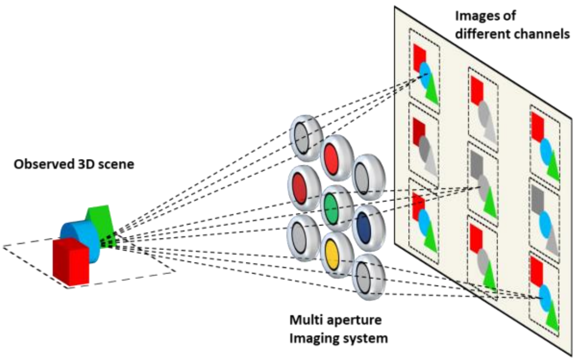

Figure 2. Imaging principle of a multi aperture camera.

The advantages of the MAC technique compared to commonly used camera techniques include the possibility to realize several channels with varying properties on one chip. The features may be different spectral channels, different foci, or fields of view. Additionally, stereo camera properties may be realized in order to obtain 3D measurements for close object distances. However, the focal length is very small (smaller than dimension of the chip).

The developed MAC has nine channels with different spectral bandpass filters as shown by Figure 4. However, the filters can be changed according to a specific application. For underwater use, wavelengths in the blue or green range will be preferred. The total number of pixels is 3870 x 2940, i.e. 1290 x 980 per

https://doi.org/10.5194/isprs-archives-XLVIII-2-W10-2025-27-2025

channel, and the size of one pixel is 2.74 µm. The dimensions are about 30 x 30 x 30 mm³, and the weight is 44 g.

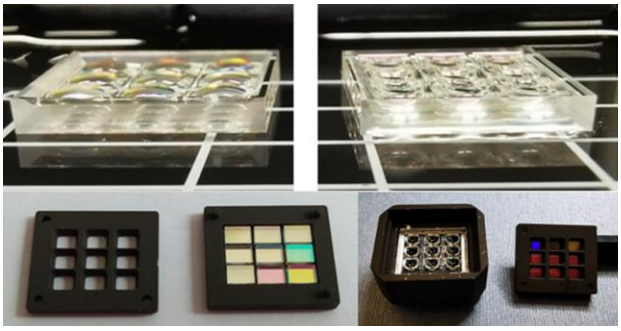

Figure 3. Components of the realized MAC: 3x3 hot embossed glass lens array (convex side – upper left, concave side - upper right), black sun vizor and glued filters (lower left), assembly status (lower right).

Figure 4. Realized MAC system.

In order to evaluate the applicability for underwater measurements, initial experiments have been performed using an glass aquarium and different measurement objects placed in the water, whereas the MAC was placed outside the aquarium. More details are given in section 3.

2.3 Underwater 3D Reconstruction Supported by AI

Artificial intelligence supports technological developments in many areas and leads to improved performance parameters. For instance, in the 3D reconstruction of underwater scenes, monocular depth estimation using AI can be advantageous, e.g., by filling in unobserved surface elements with high-probability values.

Recently, many researchers proposed new algorithmic methods to achieve monocular underwater 3D reconstruction. Roznere and Li proposed a combination of a monocular camera and a single beam echosounder to obtain depth information of underwater scenes (Roznere and Li, 2020). Varghese et al. propose a self-supervised deep learning network (Varghese et al., 2023). They achieved high-accuracy depth maps with very short processing times, enabling calculations at 55 fps. Ebner et al. use sparse priors for metrically scaled depth estimation (Ebner et al., 2023) proposing applications for underwater robots. Zhang et al. propose a method for generation of photorealistic underwater images as input for a depth estimation network (Zhang et al., 2023). The technique introduced by Wang et al. is based on a physical-guided transformer (Wang et al., 2024).

Additionally, several software packages for application to single images (Bochkovskii et al., 2025) or video streams are available (Chen et al., 2025). Monocular depth estimation can even generate complete 3D models if scale accuracy is not required. In section 3, examples of 3D reconstructions using monocular depth estimation are given.

2.4 Advantages and Disadvantages of the new Approaches

The advantages of monocular 3D computation include greater compactness, reduced weight and lower costs for the camera system. The disadvantages include offline processing, limited accuracy, and the loss of scale without additional scale representation in the images. However, in connection with AI tools, considerable improvements of monocular techniques may be expected for underwater applications.

The advantages of the multi-aperture technique include the potential to realize very flat cameras and the ability to place different spectral channels on a single camera chip. This may, e.g., help to amplify certain wavelengths which improve underwater visibility. Additionally, multiple channels allow true stereo conditions on one chip, although with limited accuracy potential due to the short base length. The disadvantages are lower spatial resolution and accuracy and the restricted visibility under water connected with the need to prolong exposure times. However, the development of this technology may soon make enormous progress in terms of its performance.

The advantages of 3D reconstruction supported by AI include the filling of 3D measurement values leading to enforced completeness of incomplete accurate 3D measurements and SLAM algorithms. Disadvantages are the lower accuracy and longer calculation times which thus far prevent real-time applications. Typically, complete 3D models can only be obtained offline. However, recent developments promise the generation of complete 3D models, e.g., SLAM mappings of seafloor environments in real-time (Varghese et al., 2023; Bochkovskii et al., 2025). Fundamental improvements in image quality and processing speeds can be expected soon.

3. Experimental Evaluation

3.1 Experiments with the goSEA3D system

The monocular system called goSEA3D has been successfully implemented for clearwater measurements in a water basin. A crane was used for positioning the goSEA3D scanner and certain tools for fixing the measurement objects, e.g. an anchor chain (see Figure 5) hanging on a rope.

Figure 5. Photograph of the goSEA3D scanner under water hanging on a crane with measurement object anchor chain (left), 3D measurement result by mesh representation (middle), and 3D model with mapped texture (right).

Figure 6 exemplifies the 3D measurement of a clay figure, and Figure 7 shows the result of the measurement of a clay pot.

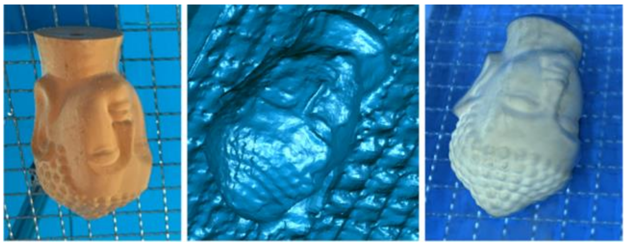

Figure 6. Photograph of the measurement object clay figure (left), 3D measurement result mesh by representation (middle), and 3D model with mapped texture (right).

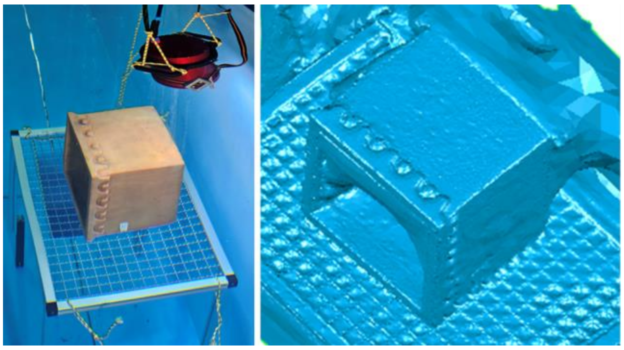

Figure 7. Photograph of the measurement object clay pot (left) and 3D measurement result by representation (right).

Experiments have also been performed in murky lake water. Unfortunately, the results have not met expectations (see the example of Figure 8). Although the anchor chain (same object as used in clearwater) could be measured, the result is unfortunately not to scale.

Figure 8. Photograph of the anchor chain in murky water (left) and 3D measurement result with mapped texture (right).

3.2 Experiments with Multi Aperture Camera System

As part of a funded project, our institute is developing a miniaturized underwater camera for assessing biodiversity based on qualitative and quantitative plankton measurements. An initial hardware development based on multi-aperture technology was experimentally investigated and tested in initial laboratory experiments with encapsulated plankton samples. A glass aquarium with dimensions of 1500 mm (length) x 500 mm (width) x 600 mm (height) was used for this purpose, in which the plankton samples were placed.

The MAC was placed outside an aquarium recording a scene and an additional LED illumination source was placed on the opposite side outside the aquarium (Figure 9). First experiments were performed concerning the sharpness of the MAC mapping depending on the measurement distance. A specific line target (see Figure 10) for spatial resolution determination was used for evaluation. The measurement objects were encapsulated plankton samples (see Figure 11). Additionally, certain plastic toy figures and the clay figure (Figure 12) were recorded fur further experiments according to 3D reconstruction.

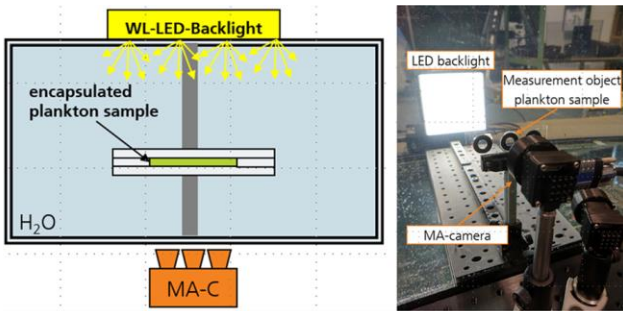

Figure 9. Scheme of the experimental arrangement using the MAC outside the aquarium, view from above (left). Photograph of the setup, view through the side windows of the aquarium (right).

Figure 10. Recorded target for spatial resolution determination at three different distances.

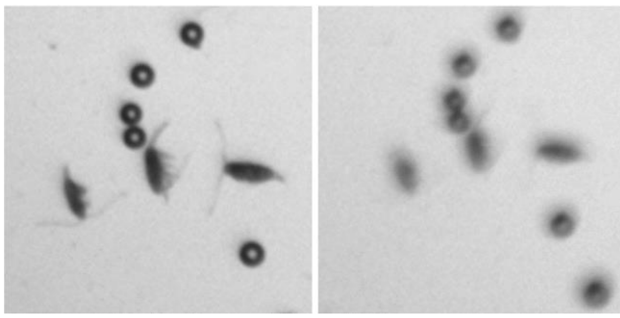

Figure 11. Recorded plankton sample at two different distances of 50 mm and 60 mm showing the small range of sharpness.

Compared to a conventional camera, the required exposure time of the MAC at recording of the image shown by Figure 12 was about ten times longer.

3.3 Experiments Using Monocular Depth Estimation

Monocular depth estimation using AI has been applied to several underwater images and video streams. Figure 13 shows an example of an underwater image, the estimated depth map, and the reconstructed 3D scene without scale using the "Depth-Pro" software (Bochkovskii et al. 2025). The scene recorded with the multi-aperture camera and shown by Figure 12 could not yet successfully be reconstructed by "Depth-Pro". This is

https://doi.org/10.5194/isprs-archives-XLVIII-2-W10-2025-27-2025

probably due to the unusual scene content and thus the lack of training using similar scenes in the software.

Figure 12. MAC outside the aquarium (left) and recorded test scene in the aquarium by the blue channel of the MAC (right).

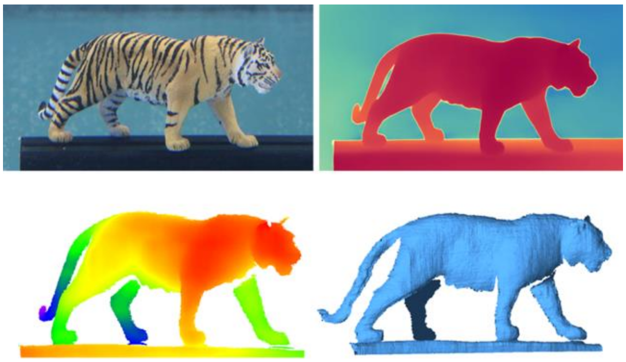

Figure 13. Original image (top left), depth image (top right), 3D results, false colors (bottom left), mesh (bottom right) using AI.

Figure 14 shows the application of the “Depth-Pro” algorithm to one single image of the water basin experiments. The 3D result obtained from “Depth-Pro” applied to an image of the blue channel of the MAC is shown by Figure 15.

An example of an AI supported reconstruction using an image taken in the Baltic Sea during previous experiments (Bräuer-Burchardt et al., 2024) is shown by Figure 16. Due to the inaccurate depth estimation, the structure of the ornamental decoration cannot be recognized in the 3D mesh image.

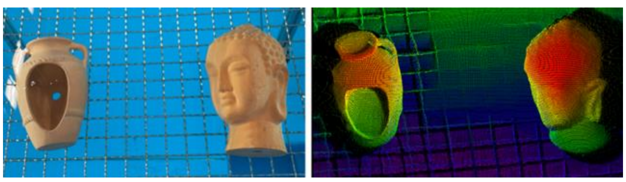

Figure 14. Photograph of measurement objects (left) and 3D result of a single image using AI (right).

4. Summary, Discussion, and Outlook

Various approaches for generating underwater 3D models of different measurement objects or entire underwater scenes, e.g., for generating seafloor maps using SLAM, were presented and discussed. The advantages and disadvantages of monocular 3D reconstruction, camera miniaturization using multi-aperture techniques, and AI-based applications for monocular depth estimation for underwater images were discussed, and examples of experimental investigations for all three approaches were presented. The use of a single camera instead of a stereo pair, and thus monocular reconstruction for the task of 3D model generation, is primarily motivated by the reduction of volume and weight in underwater applications, as well as improved manageability due to the compactness of a single camera compared to a stereo pair. This generally results in disadvantages in measurement accuracy and processing speed. However, for certain underwater applications, this need not be a disadvantage, especially when real-time generation of the 3D model is not needed, and the required measurement accuracy is not in the millimeter range. This can be the case, for example, for underwater cave explorations, where compactness, lightness, and handling of the equipment are important.

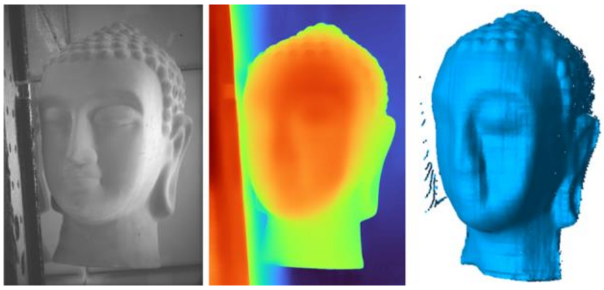

Figure 15. Photograph taken by MAC (left), depth image (middle) and 3D result (right).

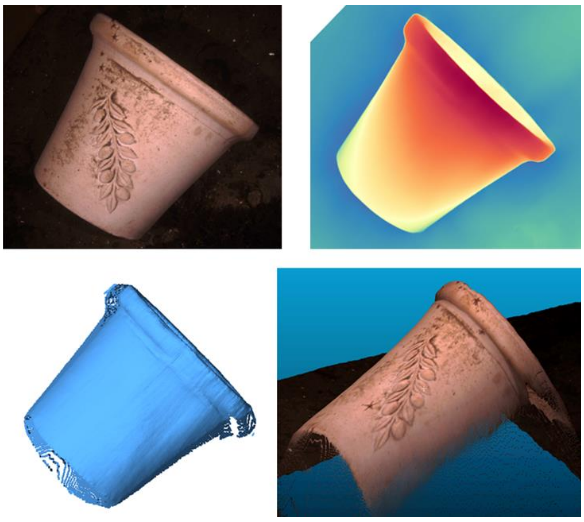

Figure 16. Original image (top left), depth image using AI (top right) and 3D result, mesh representation (bottom left) and with mapped color information (bottom right).

As described, various measurement principles and tools are available for the implementation of such systems (support through tracking tools, special lighting systems, or AI-based algorithms). Depending on the specific application, its constraints, and the available budget, a wide variety of monocular optical underwater systems for 3D reconstruction are conceivable.

The use of multi-aperture technology for underwater cameras is certainly a future scenario that still poses the most challenges for the technical development of corresponding 3D systems.

However, this research field also offers great application potential, so future work in this field will create the conditions for the implementation of 3D measurement systems based on multi-aperture technology and open new areas of application.

Future work will be directed to several objectives. First, the monocular underwater 3D system goSEA3D should be improved in terms of size, weight, and performance. Subsequently, new experimental investigations are intended to quantify the measurement accuracy and ensure the robustness of the measurements in comparison to high-precision stereo systems.

Second, the application and evaluation of multi-aperture techniques for underwater 3D measurements will be further explored. This includes the construction of cameras with different matrices of microlenses to investigate different spectral wavelength sensitivities, real stereo approximation, or different depths of focus of the several channels to increase the potential range of measurement distances.

Finally, the applicability of monocular depth estimation using underwater images will be further tested and evaluated. In this regard, new developments of commercially available software solutions should be continuously considered.

The proposed systems are all still in the experimental evaluation stage, and their integration into a cost-effective, high-performance tool for underwater exploration is still a long way off. However, this is intended to provide inspiration for further research. If the new approaches can be combined, this can result in a compact and efficient tool for SLAM applications.

Acknowledgements

This work was supported by the German Federal Ministry of Education and Research under grant label 03RU1U15 3B.

References

Agisoft Metashape, 2025. https://www.agisoft.com.

Balletti, C., Beltrane, C., Costa, E., Guerr, F., Vernier, P., 2015. Underwater photogrammetry and 3D reconstruction of marble cargos shipwrecks. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *XL-5/W5*, 7–13.

Bochkovskii, A., Delaunoy, A., Germain, H., Santos, M., Zhou, Y., Richter, S.R., Koltun, V., 2025. Depth pro: sharp monocular metric depth in less than a second. Proc ILCR 2025, arXiv:2410.02073v2 [cs.CV].

Bräuer-Burchardt, C., Munkelt, C., Bleier, M., Baumann, A., Heinze, M., Gebhart, I., Kühmstedt, P., Notni, G., 2024. Deepwater 3D measurements with a novel sensor system. Appl. Sci. 2024, 14(2), 557; https://doi.org/10.3390/app14020557.

Campbell, P. B., 2018. An introduction to archaeology in underwater caves. Highfield Press. 2018; pp. 5-26.

Chen, S., Guo, H., Zhu, S., Zhang, F., Huang, Z., Feng, J., Kang, B., 2025. Video depth anything: consistent depth estimation for super-long videos. arXiv:2501.12375v2 [cs.CV].

Costa, C., Loy, A., Cataudella, S., Davis, D., Scardi, M., 2006. Extracting fish size using dual underwater cameras. Aquac. Eng. 2006, 35, 218–227.

Ebner, L., Billings, G., Williams, S., 2023. Metrically scaled monocular depth estimation through sparse priors for underwater robots. arXiv:2310.16750v1 [cs.CV].

Eric, M., Kovacic, R., Berginc, G., Pugelj, M., Stopinsek, Z., Solina, F., 2013. The impact of the latest 3D technologies on the documentation of underwater heritage sites. In Proceedings of the IEEE Digital Heritage International Congress 2013, Marseille, France, 2013; Volume 2, pp. 281–288.

Gaglianone, G., Crognale, J., Esposito, C., 2018. Investigating submerged morphologies by means of the low-budget “GeoDive” method (high resolution for detailed 3D reconstruction and related measurements). Acta IMEKO 2018, 7, 50–59.

Guerneve, T., Pettilot, Y., 2015. Underwater 3D Reconstruction Using BlueView Imaging Sonar. *OCEANS 2015 - Genova*, Genova, Italy, 2015, pp. 1-7, doi: 10.1109/OCEANS-Genova.2015.7271575.

Hubold, M., Montag, E., Berlich, R., Brunner, R., Brünning, R., 2021. Multi-aperture system approach for snapshot multispectral imaging applications. Optics Express, Vol 29, No. 5, 7361-7377.

Kwasnitschka, T., Köser, K., Sticklus, J., Rothenbeck, M., Weiß, T., Wenzlaff, E., Schoening, T., Triebe, L., Steinführer, A., Devey, C., 2016. DeepSurveyCam—A Deep Ocean Optical Mapping System. Sensors 2016, 16, 164.

Lin, H., Zhang, H., Huo, J., Li, J., Zhang, H., Li, Y., 2025. High-precision 3D reconstruction of underwater concrete using integrated line structured light and stereo vision. Automation in Construction 169 (2025) 105883.

Luhmann, T., Robson, S., Kyle, S., Harley, I., 2006. Close range photogrammetry. Wiley Whittles Publishing.

McLeod, D., Jacobson, J., Hardy, M., Embry, C., 2014. Autonomous inspection using an underwater 3D LiDAR. In An Ocean in Common, Proceedings of the 2013 OCEANS, San Diego, CA, USA, 23–27 September 2013.

Massone, Q., Druon, S., Triboulet, J. A., 2024. A novel 3D reconstruction sensor using a diving lamp and a camera for underwater cave exploration. Sensors **2024**, 24, 4024. https://doi.org/10.3390/s24124024.

Roznere, M. and Li, A.Q., 2020. Underwater monocular image depth estimation using single-beam echosounder. IEEE, doi:10.1109/IROS45743.2020.9341105.

Varghese, N., Kumar, A., Rajagopalan, A.N., 2023. Self-supervised monocular underwater depth recovery, image restoration, and a real-sea video dataset. IEEE Xplore, 12248-58.

Wang, C., Xu, H., Jiang, G., Yu, M., Luo, T., Chen, Y., 2024. Underwater monocular depth estimation based on physical-guided transformer. IEEE Transactions of Geoscience and Remote Sensing, Vol 62, 4204216.

Zhang, F., You, S., Li, Y., Fu, Y., 2023. Atlantis: enabling underwater depth estimation with stable diffusion. arXiv:2312.12471v1 [cs.CV].