

Development and validation of a multimodal sonar measurement system for object detection in water bodies

*Sabrina Lennartz, Michael Schneider, Hubert Mantz, Thomas Walter,
University of Applied Sciences Ulm, Prittwitzstraße 10, 89075 Ulm, Germany*

Summary

This paper presents an autonomous, multimodal measurement system designed to detect man-made objects in bodies of water. The aim is the implementation of a system, based on an unmanned surveillance vehicle (USV) equipped with acoustic sensors for detecting underwater objects from various angles. The collected data is then being georeferenced and combined using imaging techniques to create high-resolution 2D and 3D representations. To develop and validate data processing, a rail-based measurement system was set up, enabling precisely georeferenced measurements in a water basin containing defined targets. In both cases, five single-beam sonars arranged in an array, as well as a side-scan sonar, are used. The array provides broad cross-range coverage while maintaining good resolution. Conversely, the side-scan sonar has very high along-track resolution and low angular resolution. Improved data processing is based on three principles: the use of multiple single-beam sonars, the combination of single-beam and side scan sonars and the fusion of data from multiple angles. In test scenarios involving known targets, the system demonstrated significantly improved image quality. Based on these findings, data processing using data collected by the USV can now be tested in a field trial.

Keywords: 3D reconstruction, multimodal sensor technology, multilook, underwater object detection

Introduction

Over 70 % of the Earth's surface is covered by water [1]. Due to its vast extent and depth, as well as factors such as limited visibility caused by light absorption and turbidity, dynamic environmental factors like currents and changing water properties, and general environmental conditions, researching these ecosystems poses a major challenge [2]. A prime example is ALDFG (abandoned, lost, or otherwise discarded fishing gear), which accounts for around 10 % of marine debris [3]. Fish, marine mammals, and birds can become entangled in the nets, and their decomposition leads to the creation of microplastics [4].

Complete, accurate, and comprehensive data are required for the detection and classification of underwater targets. The strengths of both visual and acoustic sensors can be combined using a multimodal sensor fusion system [2]. Better image quality enables users with limited resources or without a scientific background to obtain meaningful results.

This article presents three approaches to improve data processing without requiring expensive hardware. These approaches were tested in a defined setup. For this purpose, targets were placed in a pool, and the sonars

were moved using a rail-based system. This revealed the positive impact of using different sonars and fusing data from different poses. These findings and insights can now be adapted to a USV (unmanned surveillance vehicle)-based system, which introduces additional challenges, such as wave motion, but also offers more possibilities, such as the more flexible incorporation of different viewpoints of the target object. Thus, this work lays the foundation for a framework capable of combining data from various acoustic and non-acoustic sensors, thereby improving underwater imaging.

State of the art

The main difficulty in detecting ALDFGs is that they become entangled in structures such as rocks and wrecks or are pressed to the seabed by their own weight. Distinguishing the nets from the various, often complex, inhomogeneous surfaces is challenging [5]. Current approaches largely rely on artificial intelligence. Regardless of whether classification is performed by humans or artificial intelligence, there is a high false-positive rate. Similar objects, such as sediment waves or cables are often mistaken for net ropes [6].

Higher image quality would lead to better classification. One effective way to achieve this is to combine complementary perspectives and sensor modalities, as demonstrated for the fusion of multiple acoustic systems and for acoustic–optical combinations [2]. How such data is fused best remains an open question. Feature-based techniques, which dominate the field, tend to struggle underwater due to low contrast, turbidity, and the lack of stable visual features. More recent approaches instead exploit navigation information, which requires accurate knowledge of the sensor platform's position and orientation [2, 6]. This work follows the latter line, fusing data from multiple acoustic sensors through precise georeferencing.

Proposed approaches

As outlined above, the main problem lies in the poor quality of much of the collected sonar data. The three approaches presented here do not aim to improve the collection of data but its processing. First, five single-beam sonars are arranged side by side in an array to cover a larger area without sacrificing cross-range resolution. Second, data from a side-scan sonar are added, allowing the complementary beam characteristics of the two sonar types to be fused. Third, targets are observed from multiple angles and combined through suitable algorithms to achieve greater clarity. Table 1 lists the key specifications of the sonars used, the Sounder S500 single-beam sonar by and the Omniscan 450 Side-Scan sonar both from Cerulean Sonar [7].

Table 1: Specification of the sonars used [7].

Parameter	Single-Beam	Side-Scan
Frequency [kHz]	500	450
Beam Width [°]	5	0.5
Beam Height [°]	-	50
Resolution [mm]	3	< 1/1200 th of range
Max. Range [m]	100	150

Measurement system USV

The goal of this project is the improved processing of data recorded by an autonomous USV. This is based on a small catamaran to the sensors can be attached. The single-beam sonars transmit sequentially to avoid interference between them. They look downward, while the side-scan sonar is mounted on the hull and pointed slightly downward in the cross-range direction. Figure 1 shows the USV. On either side are two gray GPS antennas. Between them, at regular

intervals, are the single-beam sonars. The one at the very front can be seen in red on the side of the hull. Behind it, underwater, is the side-scan sonar mounted on the hull of the boat. The light gray box in the center of the USV contains the necessary electronics for the sensors.



Figure 1: Image of the USV, equipped with sonar and GPS antennas

Measurement system pool

To enable experiments in a controlled environment, an additional measurement system was set up. A carriage is mounted on a linear axis, to which the various sonars can be attached, as shown in Figure 2.

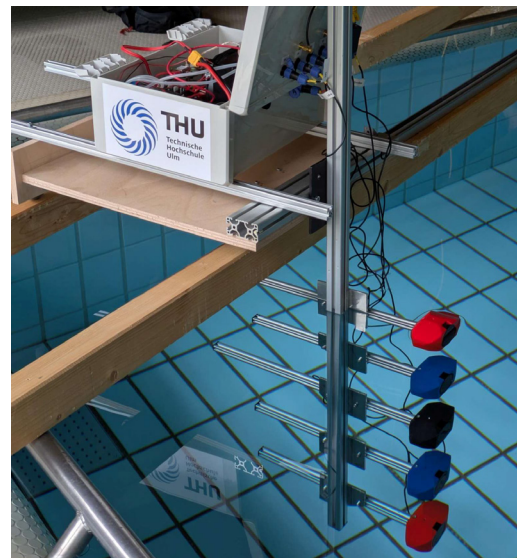


Figure 2: Picture of the measurement system

The pool measured 6.5 by 10 m, with a constant depth of 1.2 m.

Two vertical poles with a diameter of 6mm each (red circle) and two plastic boxes, one of which was covered with a net to simulate the real-life situation of a snagged net, were used as targets.

The experimental setup can be seen in Figure 3. The measuring system was positioned successively in the two corners to the left and right of the capture point.

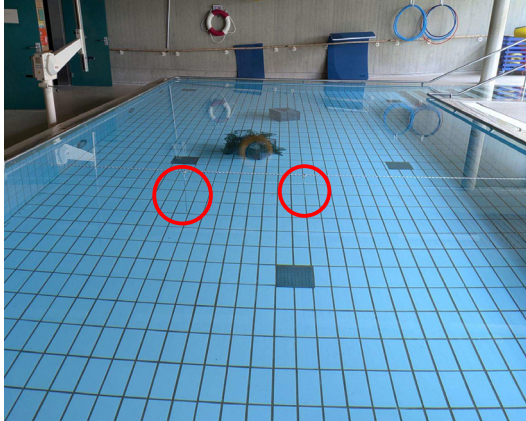


Figure 3: Image of the experimental setup. Front: Leash with two vertical metal rods, Middle: Box with net, Back: Box without net

Before presenting the scans from this experiment, the two coordinate systems used for the axis labels are briefly introduced. The parent coordinate system consists of the X and Y axes, which represent the sides of the pool, and the Z axis, which extends into the pool. Since the rail was positioned at an angle in a corner, a second coordinate system rotated relative to the main coordinate system was established for it. Here, the sonars are moved in the x_s direction, lie underneath each other on the z_s axis, and face in the y_s direction. The coordinate system definitions are illustrated in Figure 4.

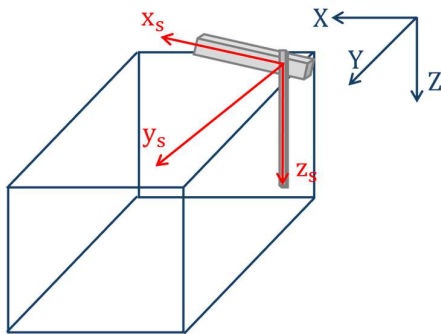


Figure 4: Schematic diagram of the parent coordinates (blue) and the measurement coordinates (red)

First Approach: Sonar Array

To use the array of single-beam sonar correctly, certain constraints must be met, which stem from the fundamental characteristics of sonars. The basic principle of a sonar is to measure the time t between transmitting and receiving acoustic signals. Based on the speed of sound c , which depends on temperature, pressure and salinity, Eq. (1) calculates the range l [1].

$$l = \frac{c \cdot t}{2} \quad (1)$$

The diameter d_v of the circular field of view changes depending on the distance from the sonar. To provide seamless data the fields of view of the sonar in the array must overlap. The diameter d_v of these circular fields of view can be calculated at a distance l by Eq. (2)

$$d_v = 2 \cdot l \cdot \tan\left(\frac{\alpha}{2}\right) \quad (2)$$

It becomes clear that a certain distance from the target is necessary or, conversely, a certain distance d_s between the sonars must not be exceeded, to achieve overlap, as in Figure 5.

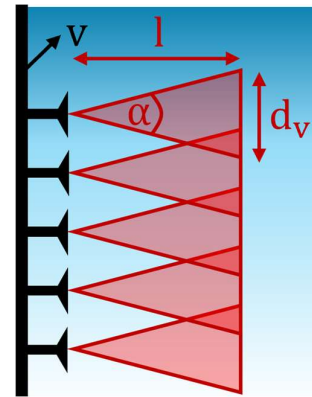


Figure 5: Drawing illustrating how the distance between the sonars, their opening angle, their distance to the target and the overlapping areas are related

Besides the overlap between adjacent sonars, consecutive pings of the same sonar must also overlap. This imposes an upper bound on the travel speed: within the ping interval t , the platform moves a distance $s = v \cdot t$, which must not exceed the footprint diameter d_v . The resulting speed limit is given by Eq. (3). Since d_v grows with the distance l to the target, the admissible travel speed increases accordingly.

$$v < \frac{2 \cdot l \cdot \tan\left(\frac{\alpha}{2}\right)}{t} \quad (3)$$

Under this condition, consecutive scans overlap as shown in Figure 6.

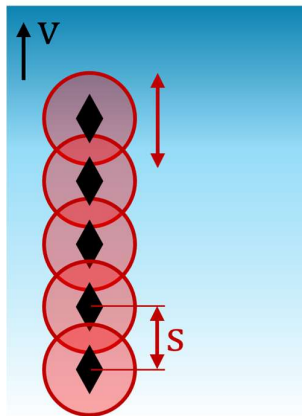


Figure 6: Drawing to illustrate how the measurement speed and the opening angle are related.

This knowledge must be applied to measurement systems. The single-beam sonars have a beam width α of 5° . In the case of the pool measurement, the nearest box - which was the focus of this study - was in a distance of 4.5 m, which would require a sonar distance of at least 39 cm. To have appropriate coverage, 18 cm was chosen.

In the case of the USV measurement vessel (as shown in Figure 1), the boat imposes certain constraints due to its twin hulls. With an even spacing between the five sonars, they can be mounted 22 cm apart from each other. This results in a minimum scan range of 5 m.

The speed of the linear axis can be specified individually. For a target at a distance of 4 m and a ping interval of 0.5 s for each sonar, this results in a maximum measurement speed of 0.78 m/s. In this study, a measurement speed of 0.7 m/s was used.

The USV has a minimum speed of 1 m/s. This results in a ping interval of 0.5 s. Consequently, the minimum distance to the targets is 5.7 m.

Second Approach: Sonar Combination

Effective fusion of different sonars requires their individual characteristics to be considered, most importantly the radiation pattern and the operating frequency.

Single-beam sonars, usually referred to as down-looking sonars due to their orientation, have a highly focused beam pattern. This results in a high degree of data clarity. However, a drawback is their limited coverage area. Side-scan sonars, on the other hand, have a very wide beam angle in one direction - usually perpendicular to the direction of travel - and a

very narrow beam angle in the direction of travel. This results in very high resolution in the direction of motion, while losing angular information. At the same time, their lateral field of view makes them particularly well-suited for scanning large areas. Combining the two sonar types yields both high depth resolution and precise lateral coverage [1,8].

In addition to correctly aligning the sonars and interpreting their fields of view, an understanding of the measured intensities is important for correctly combining the data.

The sonar equation determines the echo strength (EL) using the source level (SL), the transmission loss (TL), and the target strength (TS), as shown in the Eq. (4).

$$EL = SL - 2 * TL + TS \quad (4)$$

The strength of the echo is thus significantly influenced by the transmitters' power, the frequency (and thus the wavelength), the resulting absorption in the water, and the characteristics of the target [1]. Normalizing the data is therefore necessary to achieve a uniform greyscale distribution when superimposing scans.

The effect of proper normalization was investigated in the pool. The unique feature of the sonars used is that their data is scaled to the same maximum value. However, this results in scans with backgrounds of varying intensity depending on the target intensity for each individual scan. Each sonar must therefore be normalized individually and all sonars jointly against each other. To address this, the median of the measured values was calculated for each individual sonar measurement. This median corresponds to the value of the background, which can be used as a reference value. A scaling factor was then calculated to bring the individual scans onto a uniform scale. Figure 7 shows the improvement achieved. In the figure on the right, the data have been normalized as explained. The improvement in background uniformity across the scan is evident.

The more uniform background is clearly visible. This makes the targets stand out more clearly. At $R_1 = (0.0, 2.7)$ and $R_2 = (1.0, 2.6)$ thin metal rods are located (red circles), at $B_1 = (1.0, 4.5)$ there is a box surrounded by a net (blue circle); and at $B_2 = (1.5, 6.2)$ there is a box without a net (green circle).

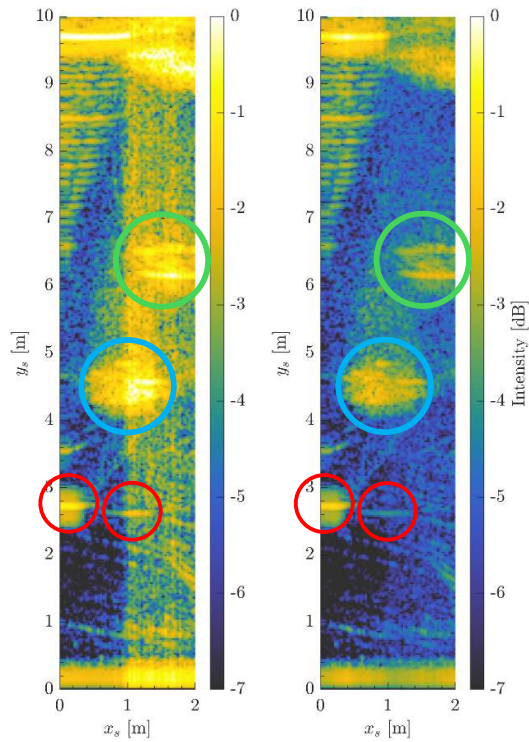


Figure 7: Comparison of the same data. The left one is not further processed while the right image has undergone normalization

In addition to a visual assessment of the improvement in scan quality, background uniformity, the CNR (Contrast to noise ratio) [9] and the SNR (Signal to noise ratio) [10] were calculated as quantitative evaluation metrics. When it comes to background uniformity, a low value is desirable, whereas a high value is advantageous for CNR and SNR [8,10].

For this purpose, the ranges $0.5 < y_s < 2.0$, $3.0 < y_s < 4.0$, and $5.0 < y_s < 6.0$ were defined as the background, and the position of the box with the net (blue circle) and the rod on the right were defined as targets. Background uniformity is calculated as the ratio of the standard deviation to the mean. The CNR is calculated as the ratio of the difference between the means of the target and the background to the standard deviation of the background [9]. The SNR, on the other hand, is calculated as the ratio of the signal mean to the standard deviation of the background [1].

The significant improvement in both metrics resulting from normalization is shown in Table 2. This is an example of the sonar shown in Figure 7, the other sonars had the same output.

Table 2: Comparison of the background uniformity, the CNR and the SNR of two targets between the original data and the normalized data of the sonar shown in Figure 7.

	Original data	Normalized data
Uniformity	0.35	0.27
CNR _{Rod}	1.01	2.55
CNR _{Box}	2.42	3.24
SNR _{Rod}	3.89	6.98
SNR _{Box}	5.31	6.29

However, when combining different sonar systems, the impact of normalization on the comparability and superimposition of two sonar scans is crucial. Figure 8 shows the superimposition of a single-beam sonar with a side-scan sonar. In the left case, the data is not normalized; in the right case, it is.

It becomes clear that adjusting the pixel values darkens the left side of the scan. This eliminates the false impression of the target on the left, stretching from $Y = 4.5$ until 8.

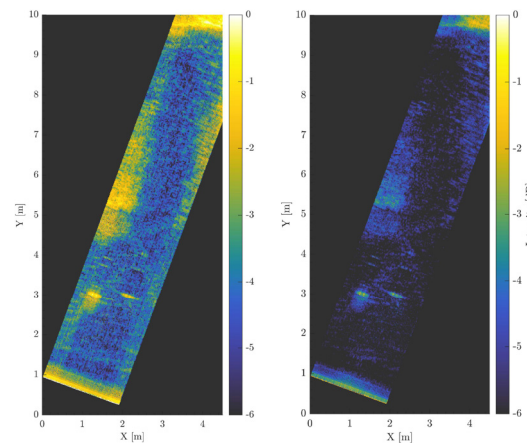


Figure 8: Comparison of the same fused data of one single beam and one side scan sonar. In the left the data is only added to each other, while in the right image had undergone normalization

Third Approach: Multilook

To achieve the third approach - the multilook method - precise georeferencing is necessary. More specifically, at every point in time when a measurement is taken, it must be known where the sonar is located and in which direction it is pointing.

With increasing distance to the target, it is also essential to consider the varying speed of sound in water. While an incorrect assumption of the speed in a single measurement merely results in a shift in the intensities, in a multilook measurement it prevents the correct superposition of the various measurements.

Several models exist for computing the speed of sound in water. Besides simplified approximations, the UNESCO, Del Grosso, and Wilson formulas are established as accurate references [1, 6].

To investigate the effect of using multiple viewing angles, data was collected from two corners of the pool. For correct overlay, each dataset was georeferenced individually and then fused pixelwise, either by multiplying or by adding the corresponding pixel values. In both cases they must be normalized afterwards to adjust the brightness, to prevent overestimation caused by multiple sightings. Figure 9 shows the results of the multiplication on the left and the addition on the right.

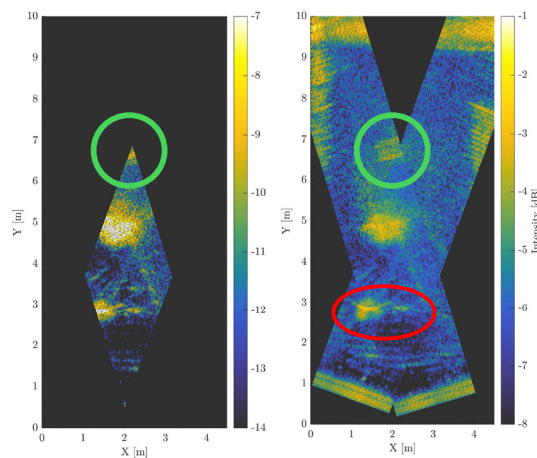


Figure 9: Comparison of multiplication (left) and addition (right) of the pixel values of two poses, measured by one single-beam sonar.

Returning to Figure 7, the metal rods appear as hyperbolic shapes typical of point-like scatterers in sonar data. The same rods can be seen in the right picture at $R_1 = (1.5, 2.8)$ and $R_2 = (2.2, 2.8)$, marked with a red circle. For R_1

one can see how the two hyperbolic shapes overlay to a cross, marking the real position of the rod and proving the correct fusion. Both ways of fusion do have advantages and disadvantages. Multiplication yields sharper targets, but any target visible from only one pose is lost — such as the upper box at $B_2 = (2, 6.8)$, marked in green.

This time, the improvement in data quality achieved by the multilook method was quantitatively verified in addition to the visual assessment. For this purpose, the two metal rods were defined as targets, and the CNR and SNR were determined again relative to them. This was performed for the individual sonars as well as for the combination of two directions. The results are shown in Table 3, again representative of the sonar depicted in Figure 8.

Table 3: Comparison of the SNR and CNR of the two rods between the data from a single sonar and the combination of them

Data set	SNR	CNR
Right corner single-beam	2.43	0.45
Left corner single-beam	3.60	1.41
Side-Scan	4.50	1.15
Both corners single-beam	6.52	2.12
All data	9.71	2.73

To combine these three approaches, array of multiple single beam sonars, combination of single-beam sonar and a side-scan sonar and the view out of multiple poses, the total data was combined. After georeferencing and normalization the 3D plot in Figure 10 was created. To make structures recognizable, a threshold was introduced to define which values should be rendered transparent. This makes the targets visible without completely losing the sense of space. Clearly visible are the two rods at $R_1 = (1.5, 2.8)$ and $R_2 = (2.2, 2.8)$ and the two boxes at $B_1 = (2.0, 5.0)$ and $B_2 = (2.0, 6.8)$. Thanks to the additional information about the length of the Z-axis, objects can also be distinguished from artifacts more effectively than is the case in a 2D representation.

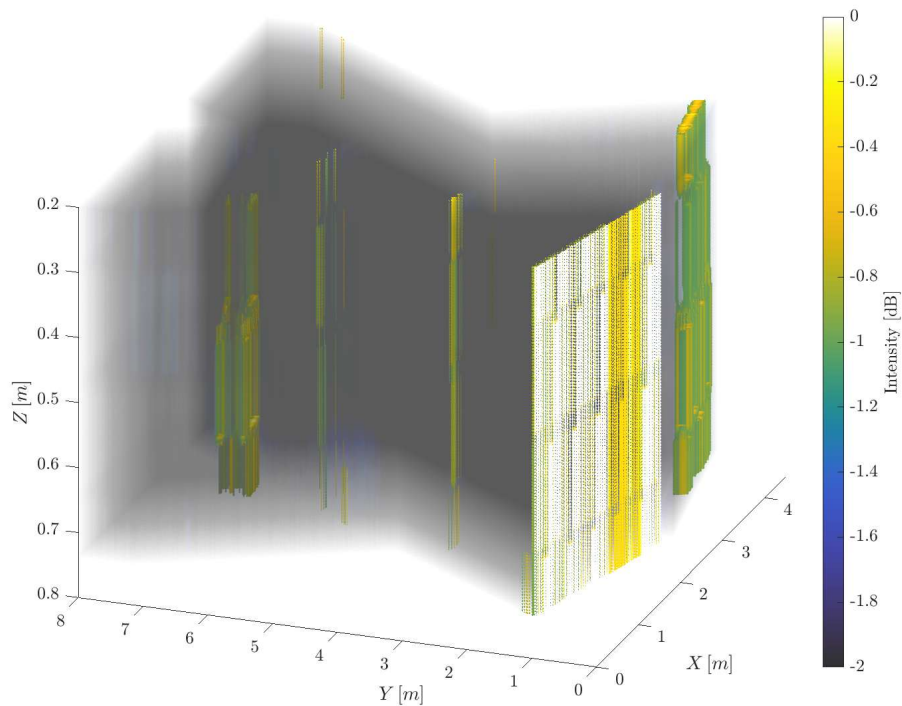


Figure 10: A total 3D view of the pool including the data from the different single beam sonar as well as from the side scan sonar from two different views.

Conclusion

This paper presented a three-stage processing pipeline for improving target clarity in sonar scans, validated in a pool experiment with a rail-based measurement system. First, several single-beam sonars were mounted one below the other, orthogonally to the direction of motion. This made it possible to cover a larger area without losing resolution. In the second stage, data from a side-scan sonar was additionally utilized. To ensure consistent use of the various sonars and account for varying conditions within a single scan, the data was normalized using the median of each scan. This resulted in more accurate target detection, which can be measured by improved background uniformity, CNR and SNR. The different fusion techniques based on multiplication and addition showed different positive features. While multiplication led to greater focus, addition provided a better overall picture. The third stage, data from two viewing poses were fused, which allowed targets to be localized more precisely and reduced artifacts even with this minimal number of perspectives. The CNR and SNR metrics improved significantly again. Combining all three stages

produced a consistent 3D representation of the pool in which every target could be identified.

Outlook

As outlined in the introduction, the long-term goal is an autonomous multimodal system capable of collecting data from real-world environments. On a USV, sonar positioning is affected by wave motion and GPS inaccuracies, in contrast to the well-defined geometry of the rail system. The use of GPS for coordinating acquisition and the presence of wave motion present additional challenges. Future work will focus on applying the findings to this system.

A second direction concerns the experimental setting itself. The pool environment does not reproduce the conditions encountered in the field such as varying bathymetry, sediment backscatter, currents, and biological clutter. Future experiments in natural water bodies will therefore be carried out to assess the robustness of the approach and to gain deeper insights into the detection and classification of ALDFGs.

Acknowledgment

The authors would like to thank the Friedrich-von-Bodelschwingh school for providing their pool for the measurements which were used in this paper.

References

- [1] X. Lurton, *An Introduction to Underwater Acoustics, Principles and Applications*; ISBN 3-540-42967-0, Springer Verlag, 2002
- [2] F.F.R. Merveille, B. Jia, Z. Xu, B. Fred, *Advancements in Sensor Fusion for Underwater SLAM: A Review on Enhanced Navigation and Environmental Perception*, *Sensors* 2024, 24; doi: <https://doi.org/10.3390/s24237490>
- [3] M. Morfin, S. Méhault, L. Miquerol, Q. C. Dufau, D. Kopp, Using citizen science to inventory and map abandoned, lost, or otherwise discarded fishing gear (ALDFG), *Marine Pollution Bulletin*, Volume 220, 2025; <https://doi.org/10.1016/j.marpolbul.2025.118418>
- [4] C. Wilcox, G. Heathcote, J. Goldberg, R. Gunn, D. Peel, B. D. Hardesty, *Understanding the sources and effects of abandoned, lost, and discarded fishing gear on marine turtles in northern Australia*, *Conservation Biology* Volume 29, Issue 1 pp. 198-206; doi: <https://doi.org/10.1111/cobi.12355>
- [5] M. Schumacher, M. Götz, G. Dederer, *Nutzung künstlicher Intelligenz zur Aufspürung von Geisternetzen in Küstengewässern*, *Journal of Applied Hydrography*, 03/2024; doi: 10.23784/HN127-04
- [6] F. J. Millero, T. Kubinski, *Speed of sound in sea water as a function of temperature and salinity at 1 atm*, *Journal of the Acoustical Society of America* 57, 312-319 (1975); doi: <https://doi.org/10.1121/1.380462>
- [7] <https://ceruleansonar.com/> (last visited: 15.04.2026)
- [8] H. Liu, X. Ye, H. Zhou, H. Huang, *A Mapping Method Fusing Forward-Looking Sonar and Side-Scan Sonar*, *Journal of Marine Science and Engineering* 13, no. 1: 166; <https://doi.org/10.3390/jmse13010166>
- [9] K. Hong-Gi, S. Jungmin, K. Soo Mee, *Underwater Optical-Sonar Image Fusion Systems*, *Sensors* 2022, 2, 8445; doi: <https://doi.org/10.3390/s22218445>
- [10] N. Z. Kolev, *Sonar Systems*; ISBN 978-953-307-345-3