Synthetic MBES Dataset for Underwater 3D Object Detection

1 General Information

Dataset title:	Synthetic Multi-Beam Dataset for 3D Underwater Object Detection
Corresponding author:	M. Salman Shaukat, muhammad.shaukat@uni-rostock.de
Date:	August 2025
Type:	3D Point Cloud Data
Language:	English
Rights:	Creative Commons Attribution 4.0 International (CC BY 4.0)
DOI:	10.18453/rosdok_id00005050
Funding:	BMBF, OTC-Digital Twin & Analytics (OTC-DaTA), 03ZU1107FB

2 Abstract

This dataset contains 100 synthetically generated multi-beam echosounder (MBES) 3D point clouds for the development and evaluation of underwater 3D object detection methods. The simulated environments were inspired by the artificial reef at the Digital Ocean Lab (DOL) near Nienhagen, Germany, and consist of procedurally generated seabeds populated with four types of artificial reef structures: small tetrapods, large tetrapods, reef cones, and concrete rings. The dataset contains a total of 5,491 annotated object instances. Each point cloud is accompanied by ground-truth 3D bounding-box annotations specifying the object class, dimensions, position, and orientation. The dataset was generated using a physics-based simulation pipeline and is intended to support research on 3D object detection from sonar point clouds, particularly in scenarios where annotated real-world sonar data are scarce.

3 Reference to other publications describing/evaluating the data set

N/A

4 Objective / Problem Statement

The objective of this dataset is to provide a synthetic multi-beam echosounder (MBES) 3D point cloud dataset for underwater 3D object detection. The dataset supports the development and evaluation of both deep learning–based and training-free detection methods operating directly on raw sonar point cloud data.

5 Dataset Description

The dataset consists of 100 synthetically generated multi-beam echosounder (MBES) sonar point cloud samples, each representing a simulated sonar scan of an underwater scene with artificial reef objects placed on a procedurally generated seabed. The scenes were created using a physics-based simulation pipeline implemented in Blender.

A total of 5,491 objects are included across all scenes. Each point cloud is accompanied by ground-truth 3D object annotations suitable for supervised learning and evaluation.

6 Data Format

Point Cloud Files

The point cloud data is stored in binary format (`.bin`). Each file contains raw, unstructured 3D point coordinates (x, y, z) representing simulated multi-beam sonar returns. The dataset comprises a total of 100 point cloud samples.

Annotations

Each point cloud sample has a corresponding annotation file stored in the `labels/` directory. Annotations follow a KITTI-style text format, where each row describes one object instance. Each row contains the object class label and the parameters of a 3D bounding box, including dimensions, center position, and orientation. This format enables direct compatibility with common 3D object detection frameworks.

7 Archive Structure

```
dataset/
|-- data/        # 100 sonar point clouds (.bin)
|-- labels/      # KITTI-style 3D bounding box annotations (.txt)
‘-- scripts/     # Visualization and environment setup
    |-- viewer.py
    |-- README.md
```

```
‘-- environment.yml
```

Each point cloud file in the `data/` directory has a corresponding annotation file with the same base filename in the `labels/` directory.

8 System Requirements

The dataset can be accessed and visualized on Linux, macOS, or Windows systems using Python 3.10. Visualization is supported via Open3D, with required dependencies specified in the provided `environment.yml` file.

9 References/Bibliography

M. S. Shaukat, Y. Käckenmeister, S. Bader and T. Kirste, "Toward Training-Free Underwater 3-D Object Detection From Sonar Point Clouds: A Comparison of Traditional and Deep Learning Approaches," in IEEE Journal of Oceanic Engineering, vol. 51, no. 3, pp. 2284-2297, July 2026, doi: 10.1109/JOE.2026.3675841.