Completion of 3D Point Clouds Derived from 2D Sonar Images Using PCTMA-Net

Philipp Knestel[1,2], Yusheng Wang[1], Nico Reeb[2], Negin Karimi[2], Florian Walter[2,3], Hiroshi Tsuchiya[4], Jun Ota[1], Hajime Asama[1], Qi An[1], Alois Knoll[2] and Atsushi Yamashita[1]

1: The University of Tokyo; 2: Technical University of Munich;

3: Technical University of Nuremberg; 4: Wakachiku Const. Co. Ltd.

I. INTRODUCTION

To accomplish autonomous navigation and the completion of tasks, the ability to accurately map and perceive the environment in three dimensions is crucial, as highlighted by studies such as Maddern's analysis of 3D perception in autonomous systems [1] and O'Mahony's exploration of the role of 3D perception in Robotics [2]. Underwater robotics is no exception. Only the initial conditions differ significantly, including distortion, reduced visibility, acoustic interference, and pressure-related challenges. This prevents a flawless transfer of reconstruction and following completion methods from above-water to underwater environments.

Sonar images are intensity maps that colourise the images depending on the backscattered intensity of an object [3]. Here, speckle noise is a granular disturbance or interference that commonly affects the quality of images acquired by radar and sonar systems. Therefore, one of the primary challenges in this domain is the generation of accurate 3D models from 2D imaging sources. This work focuses on refining and completing incomplete and noisy point clouds that are 3D reconstructed out of 2D sonar images using a method of elevation angle estimation by [4], which generates 3D point clouds out of 2D sonar images by training a model to estimate the elevation angle. Nevertheless, even if this method is very effective, the resulting point clouds still need to be more accurate to provide a useful representation of the environment for autonomous systems.

To achieve effective refinement and completion of point clouds, we employ the PCTMA-Net for dense point clouds, as proposed by [5]. Initially developed for noise-free environments, this trained network is notably capable of completing single-point clouds since even small correlations are sufficient to recognise and complete a shape [5].

Adapting PCTMA-Net to handle the nuances of sonar data, notably its noise and irregularities, forms the crux of our research. This study aims to rigorously evaluate the adaptability and effectiveness of PCTMA-Net in processing sonar-derived point clouds. Specifically, it seeks to answer how well PCTMA-Net can interpolate missing regions, handle varying levels of sonar noise, and enhance the accuracy of 3D reconstructions from sonar data.

II. PROPOSED METHOD

A. Problem Setting

A first output of the elevation-net [4] for real sonar image data can be seen in Fig. 2. The sharp edges of the ground truth cannot be appropriately reconstructed, and the point cloud becomes more noisy as the distance from the sensor increases. In contrast to incomplete but noise-free single point clouds, with which the PCTMA-Net is evaluated, our input will be noisy, with multiple objects in the scene. The Elevation-Net, evaluated using a dataset with 18 objects and the floor deactivated, showed these noisy results. The Hausdorff distance reached a low of 3,8. The f_{score} was 52.6% for less than 1mm and increased to 80.0% for less than 3mm. The Chamfer distance (multiplied by 10^4) is recorded at 134.5. The values are calculated using the reconstructed point cloud and the ground truth to measure the method's

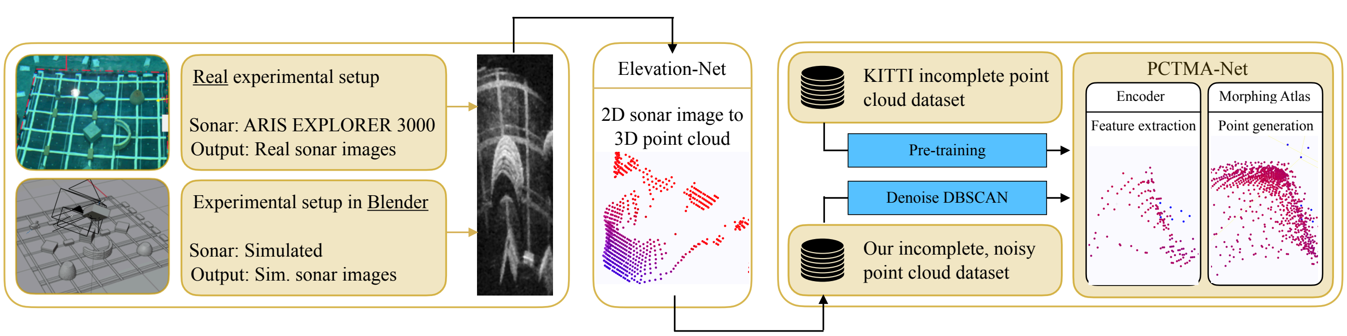

Fig. 1. Pipeline showing dataset generation in Blender, transformation of 2D sonar images into 3D point clouds, denoising of point clouds, and their completion and refinement using PCTMA-Net

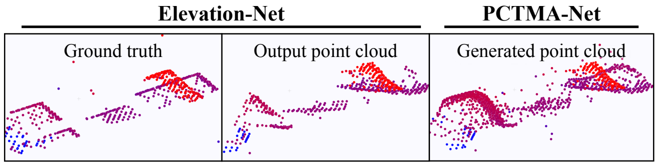

Fig. 2. Comparing the ground truth and the point cloud, transformed by the Elevation-Net, with the generated point cloud by the PCTMA-Model. The scene, captured by the sonar sensor, contains 18 objects but has a deactivated floor.

effectiveness. Both can be seen in Fig. 2. A first test of the PCTMA-Net also archived unsatisfactory results, as shown in Fig. 2. We can define three main problems:
1. The model focuses on only two out of the five objects in the point cloud, limiting completion.
2. The completion itself is not satisfactory since the sharp nodes of the ground truth model are not hit. The noise adds a high level of uncertainty.
3. The available sonar image-generated datasets are too small (3230 samples) in comparison to the 28,000 samples [6] the original PCTMA-Net used to train with [5].

B. Methodology

To reduce the noise at objects' edges and due to general outliers, density-based spatial clustering of applications with noise (DBSCAN) is used (see Fig. 1) [7]. In addition to the existing two datasets (with and without floor, 18 objects), seven datasets are added. Six datasets with three to five objects, with and without floor and one with real captured data [4]. In total, nine datasets with various objects and, therefore, noise levels will have 14,535 samples. The target is to test the PCTMA-Net under these changing conditions and fine-tune the network to handle a larger number of objects and different levels of noise using two distinct methods:
The first is to pre-train with Completion3D from the ShapeNet dataset (28,000 samples) [6] and afterwards train with all nine datasets. The Completion3D contains incomplete but noise-free samples of single object point clouds. The idea is to use these pre-trained weights to train the noisy data of our datasets.
The second is to address the challenges of noise and multiple objects in point clouds. Several modifications of the hyperparameter are tested and compared with the pre-trained model's results. To enhance the model's capability for complex data, we increase the encoder layers and the number of attention heads, improving focus on relevant features. Expanding the layer dimension is crucial to managing complexity. With more intricate point clouds, more decoder points are needed for detailed completions. More encoder channels will broaden the model's data-handling capacity. Shifting the loss calculation from Chamfer distance to Earth Movers distance (EMD) is advised, as EMD offers greater sensitivity to delicate structures. Lastly, adding more primitives will enable the model to represent diverse and complex shapes accurately.

III. EXPERIMENT

The datasets are generated as [4] described it. Blender is used to generate the synthetic datasets (Fig. 1). The number of objects in the Blender scene changes between datasets. Afterwards, the Elevation-Net generates the point clouds with the method [4] stated (Fig. 1). The PCTMA-Net is then trained with all nine datasets to ensure an evaluation of the influence of noise and the number of objects. In order to establish the comparability of the method, the Chamfer distance and of the Elevation-Net is compared with those of the PCTMA-Net. All results are measured after 300 epochs, multiplied by 10^4, and the mesh grid is set to 0.05. Since the noise points in the point clouds are not deleted, the f_{score} is not changing too much. In a comparative evaluation, the pre-trained weights as the starting point of PCTMA-Net demonstrated significant enhancement in point cloud completion. A Chamfer distance of 33.8 and an f_{score} of 64.58% under 1mm, surpassing Elevation-Net's Chamfer distance of 134.5 and an f_{score} of 52.6%. In PCTMA-Net with tuned hyperparameter, the Chamfer distance was 31.2 and f_{score} was 70.56%, showing a marked improvement even over the pre-trained weights.

IV. CONCLUSION

Our study highlights the PCTMA-Net's adaptability in completing point clouds with diverse noise levels and object counts. Key enhancements in its architecture and training, including pre-training on noise-free datasets and adjustments for complexity, have notably boosted its performance. Results from varied datasets confirm PCTMA-Net's improved handling of complex sonar data, significantly advancing 3D reconstruction for underwater autonomous navigation. However, while the model refines point clouds effectively, eliminating noise is an aspect that requires further development.

REFERENCES

[1] Will Maddern, Geoffrey Pascoe, Chris Linegar, and Paul Newman. 1 year, 1000 km: The oxford robotcar dataset. The International Journal of Robotics Research, 36(1):3–15, 2017.
[2] Niall O'Mahony, Sean Campbell, Anderson Carvalho, Lenka Krpalkova, Daniel Riordan, and Joseph Walsh. 3D vision for precision dairy farming. IFAC-PapersOnLine, 52(30):312–317, 2019.
[3] Joseph W Goodman. Speckle phenomena in optics: theory and applications. Roberts and Company Publishers, 2007.
[4] Yusheng Wang, Yonghoon Ji, Chujie Wu, Hiroshi Tsuchiya, Hajime Asama, and Atsushi Yamashita. Motion Degeneracy in Self-supervised Learning of Elevation Angle Estimation for 2D Forward-Looking Sonar. In 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 6133–6140, 2023.
[5] Jianjie Lin, Markus Rickert, Alexander Perzylo, and Alois Knoll. Pctma-net: Point cloud transformer with morphing atlas-based point generation network for dense point cloud completion. In 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 5657–5663, 2021.
[6] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, and others. Shapenet: An information-rich 3d model repository. arXiv preprint arXiv:1512.03012, 2015.
[7] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, and others. A density-based algorithm for discovering clusters in large spatial databases with noise. In kdd, volume 96, number 34, pages 226–231, 1996.