DSMNet: Deep High-precision 3D Surface Modeling from Sparse Point Cloud Frames

Changjie Qiu, Zhiyong Wang, Xiuhong Lin, Yu Zang, *Member, IEEE*, Cheng Wang, *Senior Member, IEEE*, Weiquan Liu, *Member, IEEE*

Abstract—Existing point cloud modeling datasets primarily express the modeling precision by pose or trajectory precision rather than the point cloud modeling effect itself. Under this demand, we first independently construct a set of LiDAR system with an optical stage, and then we build a HPMB dataset based on the constructed LiDAR system, a High-Precision, Multi-Beam, real-world dataset. Second, we propose an modeling evaluation method based on HPMB for object-level modeling to overcome this limitation. In addition, the existing point cloud modeling methods tend to generate continuous skeletons of the global environment, hence lacking attention to the shape of complex objects. To tackle this challenge, we propose a novel learning-based joint framework, DSMNet, for high-precision 3D surface modeling from sparse point cloud frames. DSMNet comprises density-aware Point Cloud Registration (PCR) and geometry-aware Point Cloud Sampling (PCS) to effectively learn the implicit structure feature of sparse point clouds. Extensive experiments demonstrate that DSMNet outperforms the state-of-the-art methods in PCS and PCR on Multi-View Partial Point Cloud (MVP) database. Furthermore, the experiments on the open source KITTI and our proposed HPMB datasets show that DSMNet can be generalized as a post-processing of Simultaneous Localization And Mapping (SLAM), thereby improving modeling precision in environments with sparse point clouds.

Index Terms—3D surface modeling, modeling evaluation, point cloud registration, point cloud sampling, post-processing of SLAM.

I. INTRODUCTION

OVER the past few decades, most widely used LiDARs have centimeter-level range errors and visible point spacing due to low-beam. How to utilize such sparse point clouds for 3D surface modeling has become essential in various fields, such as autonomous driving, robot navigation, industrial manufacturing, etc. Existing point cloud modeling datasets (KITTI[1], OXFORD[2], etc.) do not provide a unified and complete method to evaluate the modeling effect, whereas primarily use pose precision for modeling evaluation. However, due to the inhomogeneous surface point clouds of sparse point cloud frames, pose-based methods[3] will cause inevitable errors compared with the accurate modeling precision.

Fig. 1. **Our proposed DSMNet performs accurate surface modeling of complex objects (horse in the HPMB dataset), better than BALM[4].**

In this paper, under the above demand, first, we propose HPMB dataset, a High-Precision, Multi-Beam real-world, LiDAR dataset. HPMB consists of numerous low-precision LiDAR-scanned sequences with high-precision position and modeling ground truth. Second, we propose an evaluation criterion focusing on modeling effect. In detail, we utilize a simple similarity method to measure the difference between the low-precision LiDAR modeling results and high-precision LiDAR capture. Following this criterion, we can evaluate scene-level and object-level modeling precision.

Furthermore, existing methods (BALM[4], etc.) demonstrate ineffective in modeling complex object surfaces in sparse point clouds (see Fig. 1). The problem of learning modeling of the environment is customarily known as Simultaneous Localization and Mapping (SLAM). Among these, mapping tasks are often operated as post-processing, such as moving least squares, bundle adjustment and others[5]. Most methods rely on the matching features on the prior inter-frame registration results for modeling correction, which leads to low adaptability in sparse point clouds[5]. In this case, the shape characteristics of the object will become increasingly blurred and continue to disturb the inter-frame pose estimation.

Therefore, preliminary inter-frame pose estimation plays a vital role in this task, which is accomplished with point cloud registration (PCR). There are many effective traditional algorithms in the field of PCR. Besl et al. [6] proposed a groundbreaking iterative optimization method based on point-wise distance. Dellenbach et al.[7] utilizes point-based feature matching. Besides, methods based on deep learning demonstrated excellent results. Li et al.[8] proposed IDAM based on the idea of ICP, and Pan et al.[9] proposed GMCNet, which introduced the popular transformer in the matching framework.

However, the destination of the PCR is only to maximize the coincidence rate of both point clouds, but it do not handle the situation of significant noise and redundant surface points. In this scenario, it is desirable to filter out insignificant points, at which Point Cloud Sampling (PCS) is considered helpful.

Fig. 2. **Overview of DSMNet:** DSMNet alternately processes point cloud frames through the Density-Aware PCR and the Geometry-Aware PCS module. In each alternation, DSMNet predicts the complete and precise point cloud obtained by the current processing, which continuously updates two point-wise weighting maps used for feature fusion.

Fig. 3. **Overview of Density-Aware PCR module:** The module unifies the point-wise density feature through attention kernels of different scales of horizons. Given the source and target point clouds, the registered source point cloud is obtained through the similarity matrix calculated by RANSAC.

Equivalent to PCR, the most commonly used method is the traditional method farthest point sampling[10], which preserves point clouds by iteratively picking the most faraway points. In the field of deep learning, S-Net[11] and SampleNet[12] both adopt the same training method that instructs a separate PCS module by training the downstream task.

In light of this, we propose a learning-based approach, DSMNet, while jointly optimizing the inter-frame pose estimation and the location of points. Instead of using simple surfel or plane methods, we use geometry-aware PCS to enhance the ability to obtain robust geometry features of sparse point clouds. As for feature matching, density-aware PCR achieves overlapping consistency for objects with complex structures. Additionally, we jointly optimize PCR and PCS to enhance feature fusion between modules. We utilize neighbor-focused PCR to transfer local density information to PCS by a point-wise weighting map, making PCS adaptable to uneven density. At the same time, we pass the geometric information as another weighting map of global-focused PCS to PCR, which handles the situation of low overlap and shape confusion. Overall, our main contributions are as follows:

• We independently construct a set of LiDAR system with an optical stage to contribute a high-precision, multi-beam, real-world dataset HPMB, containing over 3,000 points cloud frames with high-precision position and modeling information. Based on HPMB, we propose a unified and adaptable evaluation of the modeling precision method.

• We propose a novel learning-based framework, DSMNet, to jointly optimize geometry-aware PCS and density-aware PCR, which simultaneously learns implicit density and geometry information.

• Extensive experiments demonstrate that DSMNet outperforms previous state-of-the-art methods in PCS and PCR. Besides, DSMNet can be generalized as a post-processing of SLAM, which can precisely model complex objects in sparse point clouds.

II. DSMNET

Given a sequence of point cloud frames $I = \{P_1, \ldots, P_m\}$, where $P = \{p_i \in R^3\}_{i=1}^{n}$. 3D surface modeling from point cloud frames aims to obtain precise and complete modeling of corresponding objects in a unified world coordinate system. The architecture of DSMNet is shown in Fig. 2.

A. Density-Aware Point Cloud Registration

In the case of sparse point clouds, the locations of surface points are affected by LiDAR measurement errors. To address the noise generated by such errors, we follow the construction pattern of variational autoencoder (VAE) for feature extraction. This makes feature distribution close to the normal distribution, reducing the position error caused by normal noise. Moreover, we divide it into a global encoder and a point-wise neighborhood encoder. The details of the feature encoder follow the encoder design of Pointnet++[13].

After obtaining global and point-wise local features, the PCR module adopts a novel local density consistent cross-attention module to combine global and local information (see Fig. 3). Specifically, we use local feature distribution as a query matrix and global feature distribution as a key matrix to query the distribution difference of neighborhoods in global information. Based on this, we use the proportion of the difference to mask the point-wise local features, which filter points with non-uniform densities. Additionally, to accommodate the densities of the corresponding point neighborhoods of the source and target point clouds, we aggregate local features using different scales of horizons. Finally, the relationship matrix M between two point clouds is obtained by RANSAC[14], and a point-wise significance map R oriented to PCR, which is used for cross-module feature fusion, defined as follows:

$$R(p_i) = \sum_{\mathbf{P_i} \in \mathbf{P}} top5(M_{i,1}, M_{i,2}, \ldots, M_{i,n}) \tag{1}$$

where M is the similarity matrix between the source point cloud and the target point cloud. The significance map is the sum of five largest similar weights associated with per point on the similarity matrix, indicating the likelihood that the point belongs to an inline point of another point cloud during the registration process. The loss L_R operated in the registration module is the distance between the predicted and ground truth registered point clouds, formulated as:

$$L_R = \frac{1}{N} \sum_{i}^{N} |(R_{gt}p_i + t_{gt}) - (R_{pred}p_i + t_{pred})| \tag{2}$$

where R is the rotation matrix and t is the translation vector of the source point cloud registered to the target point cloud.

B. Geometry-Aware Point Cloud Sampling

Divergent from the PCR module, PCS utilizes a geometry consistent self-attention module for geometric feature

Fig. 4. **Overview of Geometry-Aware PCS module:** The module explores the implicit shape feature through attention kernels at different resolution scales. Given a point cloud, a complete, precise sampled result is obtained.

Fig. 5. **Overview of HPMB LiDAR System:** The system consists of an optical stage, a milliradian-level precision translation guide, a millimeter-level precision rotary table, and different LiDARs.

TABLE I
COMPARING HPMB WITH EXISTING DATASETS.

Dataset	Translation Precision	Rotation Precision	Modeling Precision
OXFORD[2]	$\geq 1cm$	0.02°	20mm
KITTI[1]	$\geq 1cm$	0.15°	20mm
3DMatch[15]	2cm	0.1°	×
HPMB(Ours)	**0.01cm**	**0.01°**	**5mm**

enhancement. Afterward, an accurate and complete point cloud O with uniform density is obtained by uncomplicated point-wise position regression. Each attention kernel further explores the implicit global shape information of point clouds by fusing point-wise features corresponding to point clouds with different scales of resolution. The details are shown in Fig. 4.

The significance map S of PCS is obtained from the minimum distance calculation between sampled and original point clouds, which is defined as follows:

$$S(p_i) = \min_{\mathbf{p_i} \in \mathcal{P}, \mathbf{o_i} \in \mathcal{O}} \|p_i - o_j\|_2^2 \tag{3}$$

The significance map S is represented as the point-wise importance of the global shape, measured by point-wise minist distance. The loss L_S operated in the sampling module, the traditional Chamfer Distance[12], can be formulated as:

$$L_S(P, O) = \frac{1}{n} \sum_i^n \min_{\mathbf{p} \in \mathbf{P}, \mathbf{o_i} \in \mathbf{O}} \|\mathbf{o}_i - \mathbf{p}\|_2^2 \\ + \frac{1}{n} \sum_i^n \min_{\mathbf{o} \in O, p_i \in P} \|\mathbf{p}_i - \mathbf{o}\|_2^2 \tag{4}$$

C. Cyclic Optimization Based on Point-wise Weighting Map

Our DSMNet is trained end-to-end and optimizes the PCR and PCS modules alternately (see Fig. 2). Initially, we utilize the PCR module to register the point clouds to a unified world coordinate system and the PCS module to sample the registered point clouds for a concise result. In the next iteration, we utilize the sampled point cloud of the previous iteration as the target point cloud for registration.

However, cyclic modeling will lead to the accumulation of errors. In order to effectively suppress this situation, we utilize two point-wise weighting maps for reducing the attention weights of low utility points generated by previous iterations. Precisely, we not only utilize the significance map introduced in Sec. II-A and II-B, but also construct a point-wise neighborhood distribution map D. After computing global and point-wise local feature distributions, M points are randomly sampled from each local feature distribution. Based on these points, the probability belonging to the global feature distribution is calculated as the value of the point-wise distribution map. The specific calculation method is as follows:

$$D(p_i) = \sum_{a_j \in dl(p_i)}^{M} \left(-\frac{(a_j - \mu)^2}{2\sigma^2} - \log(\sigma) - \log(\sqrt{2\pi}) \right) \tag{5}$$

where μ and σ is the mean and variance of the global feature distribution. $dl(\cdot)$ is the samping point set of the local feature distribution. M is the number of sampling points.

The point-wise distribution map, containing features about the neighborhood density, assigns task-irrelevant information for module optimization. The significance map provides task-oriented information and the geometry importance of the local shape in the global shape. Through the cyclic, the map obtained by PCR assists PCS in performing a preliminary sampling based on density. PCS assists PCR in screening geometry outliers. It is worth noting that since there are no corresponding results to calculate in the first iteration, we replace the weighting map with two matrices of all ones.

III. HPMB DATASET

We independently construct a set of LiDAR system with an optical stage, containing a millimeter-level precision translation guide and a milliradian-level precision rotary table to collect data, which is illustrated in Fig. 5.

• **Data Acquisition.** During the capture process, multiple complex structured objects were placed in a static scene at two meters from the LiDAR system. We operate the high-precision translation guide and rotary table to move the low-precision LiDARs to simulate realistic multi-view modeling. After each movement, we record the ground truth of high-precision rotation and translation at the current time node. At the same time, under a unified world coordinate system, we use a millimeter-level precision LiDAR to capture, regarded as the ground truth of high-precision global modeling.

In summary, HPMB provides 3,600 time node, three kinds of LiDARs (Velodyne-16, Hesai-32 and Rigel VZ-1000), and 10 kinds of complex objects (Horse, Cabinet, Sofa, etc.).

• **Data Characteristic.** Table I presents statistics of our HPMB dataset in comparison to other publicly available datasets. Our HPMB dataset has the following advantages: (1) HPMB consists of over 3,000 high-precision optical flow ground truth, while other registration-oriented or SLAM-oriented datasets usually have limited precision. (2) HPMB is comprised of multiple LiDARs with varying beams and precision. Most current datasets lack such high modeling precision.

Fig. 6. **Qualitative results on the MVP-SP and MVP-RG datasets.** The upper part is PCS and the lower part is PCR, which proves DSMNet can handle various challenging conditions (Large Noise, Moving, etc.).

TABLE II
QUANTITATIVE COMPARISON OF PCR IN MVP-RG DATASET.

Method	Rotation Error (E_R)	Translation Error (E_t)
DeepGMR[17]	43.74°	0.353
GMCNet[9]	16.57°	0.174
IDAM[8]	24.35°	0.280
RPMNet[18]	22.20°	0.174
DSMNet(Ours)	**14.17°**	**0.158**

(3) On the basis above two, we can calculate the similarity between the modeling results of low-precision LiDAR and the capture of high-precision LiDAR at any time node, which can be regarded as a unified modeling effectiveness evaluation method. Existing datasets use pose precision to replace the evaluation of modeling precision, which has low precision and inevitable error when the point cloud is sparse.

IV. EXPERIMENTS

In line with previous methods, like VRCNet [16] and GMCNet [9], we evaluate PCR by computing isotropic rotation errors and translation errors: $E_R = \arccos\left(\frac{1}{2}\left(\mathrm{tr}\left(R_{G_T}^{-1} \cdot R_{\mathrm{pred}}\right) - 1\right)\right)$, $E_t = ||t_{GT} - t_{\mathrm{pred}}||_2$. Besides, we evaluate PCS by computing the Chamfer Distance (Eq. (4)), which is also utilized as the evaluation method of modeling precision in the HPMB dataset.

A. Point Cloud Registration and Sampling

MVP [16] is a multi-view partial point cloud dataset, consisting of over 100,000 complete and incomplete point clouds in 16 categories, which renders partial 3D shapes from 26 uniformly distributed camera poses for 3D CAD models.

Following the construction method of the MVP dataset[16], we construct a challenging multi-view registration dataset (MVP-RG) for PCR. In summary, MVP-RG dataset consists of 7,600 partial point cloud frame pairs from sixteen categories. We compared our DSMNet with existing superior methods (see Table. II). Experiments demonstrate that our DSMNet has a favorable effect on the MVP-RG dataset, which represents our PCR module exhibiting excellent performance on challenging, low-overlap partial point clouds (see Fig. 6).

Fig. 7. **Qualitative modeling results on the KITTI dataset.** DSMNet can be used as a SLAM post-processing algorithm to achieve high-precision surface modeling of complex objects (Cars, Pillars, etc.)

TABLE III
QUANTITATIVE COMPARISON OF PCS IN MVP-SP DATASET.

Method	Modeling Precision (Chamfer Distance F-Score@1%)
FPS[10]	0.507
SampleNet[12]	0.5921
Local SampleNet	0.5421
TASL[19]	0.6410
S-Net[11]	0.3954
DSMNet(Ours)	**0.6610**

Unlike the point cloud completion (PCC) introduced in the original MVP dataset, we construct a challenging sampling dataset MVP-SP based on the data of the PCC. Specifically, for each object we randomly select several frames from the corresponding set of multi-view partial point cloud frames. On this basis, we stack all the selected frames, which is regarded as the input of PCS. Subsequently, the complete and concise point cloud provided by the PCC is used as the ground truth. In summary, our MVP-SP dataset consists of 104,000 point cloud frames of 16 categories. Experiments demonstrates that our DSMNet outperforms SOTA methods in the MVP-SP dataset, showing the superiority of our PCS module in the case of non-uniform densities and incomplete point clouds (see Fig. 6 and Table III).

B. Post-processing of Simultaneous Localization and Mapping

The KITTI odometry dataset[1] consists of 22 independent sequences captured during driving over various road environments. Sequences 00-10 (23,201 scans) are provided with ground truth poses obtained from the IMU/GPS readings.

Rather than comparing the precision of trajectory in the KITTI dataset, we focus more on the effect of modeling. Specifically, we use several classic LiDAR odometry methods of SLAM as preprocessing. Based on these prior poses, we tested our DSMNet and several post-processing algorithms for fine modeling. To achieve an intuitive comparison, we selected a special and complex object, car, from the KITTI dataset for visual comparison (see Fig. 7). As the visualization shows, our framework also obtains excellent modeling precision on complex structural objects in the case of high noise or poor

Fig. 8. Qualitative modeling results on the HPMB dataset by different methods. DSMNet can guarantee the integrity of the shape of the complex object (horse) in the case of sparse point clouds.

TABLE IV
QUANTITATIVE COMPARISON OF MODELING EFFECT OF VELODYNE-16 LiDAR ON HPMB DATASET.

PreProcessing	Method	Modeling Precision (Chamfer Distance F-Score@1%)
ICP[6]	MLS[20]	0.017
	BALM[4]	0.024
	DSMNet(Ours)	**0.321**
GICP[21]	MLS	0.103
	BALM	0.142
	DSMNet(Ours)	**0.341**
CTICP[7]	MLS	0.079
	BALM	0.046
	DSMNet(Ours)	**0.411**

prior results. While in the environment of good prior and high signal-to-noise ratio point clouds, our algorithm achieves a similar effect comparable to these algorithms.

However, due to the lack of methods to evaluate the effect of modeling in the KITTI dataset, we can only achieve qualitative experiments. As introduced in Sec. III, our HPMB dataset utilizes a unified modeling effect evaluation method to realize quantitative experiments at the object level. Additionally, we also achieve qualitative experiments like the KITTI dataset.

The visual results shed light on the more lower precision of LiDAR, the more worse modeling quality due to sparsity and cumulative errors (see Fig. 8). In this particular scenario, experiments demonstrate that our DSMNet effectively alleviates this situation and guarantees highly precise point cloud modeling results for complex object. For a more intuitive comparison effect, quantitative experiments demonstrate that DSMNet outperforms SOTA methods of modeling precision in our HPMB dataset (see Table IV).

V. CONCLUSION

In this letter, we independently construct a set of LiDAR system with an optical stage to contribute a high-precision, multi-beam, real-world dataset HPMB. Based on HPMB, we propose a method to evaluate modeling precision, which fills the gap that the existing datasets cannot evaluate modeling quality. Besides, we propose a novel learning-based joint framework, DSMNet. DSMNet jointly optimizes density-aware PCR and geometry-aware PCS for modeling complex objects of sparse point cloud. Experiments show that DSMNet achieves the best results in PCS and PCR, and can be generalized as SLAM post-processing to solve the difficult challenge of high-precision modeling in complex environments.

REFERENCES

[1] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 3354–3361.

[2] D. Barnes, M. Gadd, P. Murcutt, P. Newman, and I. Posner, "The oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 6433–6438.

[3] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: a versatile and accurate monocular slam system," *IEEE transactions on robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[4] Z. Liu and F. Zhang, "Balm: Bundle adjustment for lidar mapping," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3184–3191, 2021.

[5] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment—a modern synthesis," in *Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms Corfu, Greece, September 21–22, 1999 Proceedings*. Springer, 2000, pp. 298–372.

[6] P. J. Besl and N. D. McKay, "Method for registration of 3-d shapes," in *Sensor fusion IV: control paradigms and data structures*, vol. 1611. Spie, 1992, pp. 586–606.

[7] P. Dellenbach, J.-E. Deschaud, B. Jacquet, and F. Goulette, "Ct-icp: Real-time elastic lidar odometry with loop closure," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 5580–5586.

[8] J. Li, C. Zhang, Z. Xu, H. Zhou, and C. Zhang, "Iterative distance-aware similarity matrix convolution with mutual-supervised point elimination for efficient point cloud registration," in *European conference on computer vision*. Springer, 2020, pp. 378–394.

[9] L. Pan, Z. Cai, and Z. Liu, "Robust partial-to-partial point cloud registration in a full range," *arXiv preprint arXiv:2111.15606*, 2021.

[10] C. Moenning and N. A. Dodgson, "Intrinsic point cloud simplification," *Proc. 14th GrahiCon*, vol. 14, p. 23, 2004.

[11] O. Dovrat, I. Lang, and S. Avidan, "Learning to sample," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2760–2769.

[12] I. Lang, A. Manor, and S. Avidan, "Samplenet: Differentiable point cloud sampling," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7578–7588.

[13] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," *Advances in neural information processing systems*, vol. 30, 2017.

[14] O. Chum, J. Matas, and J. Kittler, "Locally optimized ransac," in *Pattern Recognition: 25th DAGM Symposium, Magdeburg, Germany, September 10-12, 2003. Proceedings 25*. Springer, 2003, pp. 236–243.

[15] A. Zeng, S. Song, M. Nießner, M. Fisher, J. Xiao, and T. Funkhouser, "3dmatch: Learning local geometric descriptors from rgb-d reconstructions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1802–1811.

[16] L. Pan, X. Chen, Z. Cai, J. Zhang, H. Zhao, S. Yi, and Z. Liu, "Variational relational point completion network," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 8524–8533.

[17] W. Yuan, B. Eckart, K. Kim, V. Jampani, D. Fox, and J. Kautz, "Deepgmr: Learning latent gaussian mixture models for registration," in *European conference on computer vision*. Springer, 2020, pp. 733–750.

[18] Z. J. Yew and G. H. Lee, "Rpm-net: Robust point matching using learned features," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11824–11833.

[19] Y. Lin, L. Chen, H. Huang, C. Ma, X. Han, and S. Cui, "Task-aware sampling layer for point-wise analysis," *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[20] S.-L. Liu, H.-X. Guo, H. Pan, P.-S. Wang, X. Tong, and Y. Liu, "Deep implicit moving least-squares functions for 3d reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 1788–1797.

[21] A. Segal, D. Haehnel, and S. Thrun, "Generalized-icp." in *Robotics: science and systems*, vol. 2, no. 4. Seattle, WA, 2009, p. 435.