Dense 3D Neural Map Reconstruction Only Using a Low-cost LiDAR

Guangming Wang, Ph.D.,[1] Alwyn Mathew, Ph.D.,[1] Mudan Wang, Ph.D.,[1] Ya Wen, Ph.D.,[1] Hamidreza Alavi, Ph.D.,[1] Weiwei Chen, Ph.D.,[2] and Ioannis Brilakis, Ph.D., M.ASCE[1]

[1] Department of Engineering, University of Cambridge

[2] The Bartlett School of Sustainable Construction, University College London

ABSTRACT

Scanning plays a vital role in civil engineering, automation construction, and remote sensing, which is the basis for high-level tasks like Building Information Modeling (BIM) and construction quality control. However, static scanning requires a significant amount of manual labour and time. Mobile scanning has reduced the time and post-processing work compared to traditional static scanning. However, the point cloud density in mobile scanning is often lower than that of static scanning. To address this challenge, we propose a novel approach that only relies on a cheaper LiDAR to enhance scanning resolution. Our approach employs explicit LiDAR odometry for odometry estimation and AI-based neural 3D dense mapping. The approach combines the high-precision registration of current explicit LiDAR odometry with the strategy of jointly optimising global temporal information through neural map representations. As a result, it achieves high accuracy in localisation and increased density in mapping. These case studies on building scenes demonstrate that this method significantly improves the point cloud density of original scans.

INTRODUCTION

3D reconstruction of existing buildings using sensors such as Light Detection And Ranging (LiDAR) is a fundamental step in creating Digital Twins (DTs). This process is essential for asset operation, management, and renovations (Sacks et al., 2018). The static scanner needs to be positioned at many locations where each location is with a long period to obtain dense point cloud reconstruction. Mobile scanners can achieve faster scanning speeds, but the density still needs improvement (Trzeciak et al., 2021). In addition, despite recent progress in mobile scanners using expensive LiDAR (Shan et al.,2020; Xu et al., 2021; Xu et al., 2022), obtaining dense detailed 3D reconstruction with a low-cost LiDAR remains a challenge as low-cost LiDAR have fewer beams and sample sparse 3D points. Recently, some methods have been proposed to use depth estimation (Wang et al., 2020) or depth completion (Trzeciak et al., 2023) to obtain denser 3D point clouds. However, images are susceptible to changes in lighting conditions, and the issue of artifacts in depth estimation and depth completion techniques cannot be overlooked. Therefore, efficiently acquiring dense point cloud data with cost-effective equipment remains an open challenge.

Recently, the development of global neural maps based on Neural Radiance Fields (NeRF) (Mildenhall et al., 2021) has shed new light on addressing the abovementioned challenges. This technology initially involves using a series of images to render images with new perspectives

(Mildenhall et al., 2021). During the process of rendering new perspectives, implicit 3D modelling is essentially taking place. Moreover, thanks to the online training approach, it allows for 3D modelling of scanned scenes without the need for annotated data. Leveraging these advantages, related research started using RGBD cameras (Sucar et al., 2021; Zhu et al., 2022) to construct 3D models of indoor architecture. LiDAR was employed for the global neural maps in 2023 (Isaacson et al., 2023). With a longer sensing distance, LiDAR can be used in harsh conditions, which gives the possibilities for more extensive low-cost civil and architectural 3D mobile scanning.

Based on these latest developments, this paper empirically tests and validates the potential of global neural maps for the low-cost 3D reconstruction of indoor building scenes. It finds that the accuracy of original pose estimation based on global neural maps is relatively low, leading to the loss of 3D structure in indoor environments. Therefore, the paper proposes using a well-established LiDAR odometry algorithm to obtain camera pose estimates. Subsequently, these pose estimates are used to train the global neural map, enabling dense 3D mapping technology to be achieved using only an affordable LiDAR scanner.

STATE OF RESEARCH

The existing mobile scanners primarily utilise LiDAR sensors for Simultaneous Localization and Mapping (SLAM). Through point cloud registration techniques, newly acquired frames of point clouds are registered with historical point clouds. The newly registered point clouds are then incorporated into the whole map, gradually expanding the map as scanning progresses. Additionally, due to point cloud distortion during motion, the Inertial Measurement Unit (IMU) sensor is often employed. Representative recent studies in this field include LIO-SAM (Shan et al., 2020), Fast-LIO (Xu et al., 2021), and Fast-LIO2 (Xu et al., 2022). LIO-SAM (Shan et al., 2020) achieves real-time performance by registering selectively chosen new keyframes to a fixed-size set of prior sub-keyframes. However, Fast-moving environments with noise usually lead to rapid degradation of LIO-SAM. Therefore, Fast-LIO (Xu et al., 2021) adopts a tightly-coupled iterated Kalman filter to fuse LiDAR feature points with IMU measurements and proposes a formal back propagation process to compensate for the motion distortion. However, there is usually a heavy computational burden caused by many feature points from LiDAR. Furthermore, Fast-LIO2 (Xu et al., 2022) designs an incremental k-d tree (ikd Tree) data structure that efficiently represents a large dense point cloud map and performs registration directly from the raw point cloud to eliminate feature extraction, making the system applicable to various LiDAR sensors. As the current handheld scanning devices are unable to provide point clouds with sufficient resolution, Trzeciak et al. (2023) utilised deep learning techniques to fuse images and laser scans into dense reconstructions, which are then used in a similar manner to odometry for incremental reconstruction. By utilising RGB images, the density of sparse laser scans can be significantly increased but suffer from artifacts and need additional camera sensors with calibration work. The method based on neural radiance fields allows for constructing a global neural map without the need for cameras (Isaacson et al., 2023). It fits the 3D spatial distribution, utilising LiDAR to densify 3D point cloud reconstruction through AI technology. However, this technology lacks sufficient practical applications. This paper aims to study it.

LiDAR is a commonly used sensor in the current field of surveying. However, methods based on inter-frame registration lack density, and those based on global neural maps are dense but lack registration accuracy. Hence, research and experiments on effectively combining their strengths for optimal real-world applications remain a focus of this paper.

RESEARCH METHODOLOGY

System architecture. As illustrated in Figure 1, our core concept revolves around using Fast-LIO (Xu et al., 2021) as the front end for SLAM to provide a relatively accurate pose estimation. We employ the Neural Radiance Field (NeRF) (Isaacson et al., 2023) for the back-end mapping. As the learned neural radiance field represents a spatially continuous function, the proposed architecture ensures that even with sparse point cloud input, the learned NeRF can also be used to estimate the density of each spatial point, rendering maps with varying density levels.

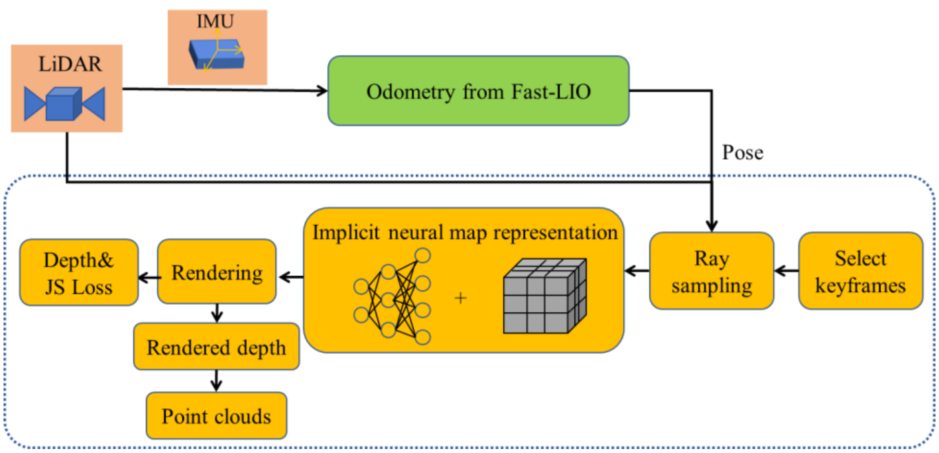


Figure 1. The architecture diagram of the proposed method.

Odometry estimation. Fast-LIO is utilised as the front end to obtain the relative transformation from the last to the current scan. We chose Fast-LIO over algorithms like Nice-SLAM (Zhu et al., 2022) to optimise the global pose variables during training NeRF due to the strong performance of the Fast-LIO algorithm, which conserves GPU resources and maintains real-time capabilities, even on CPUs. The estimated poses from the front end are later refined in the mapping thread.

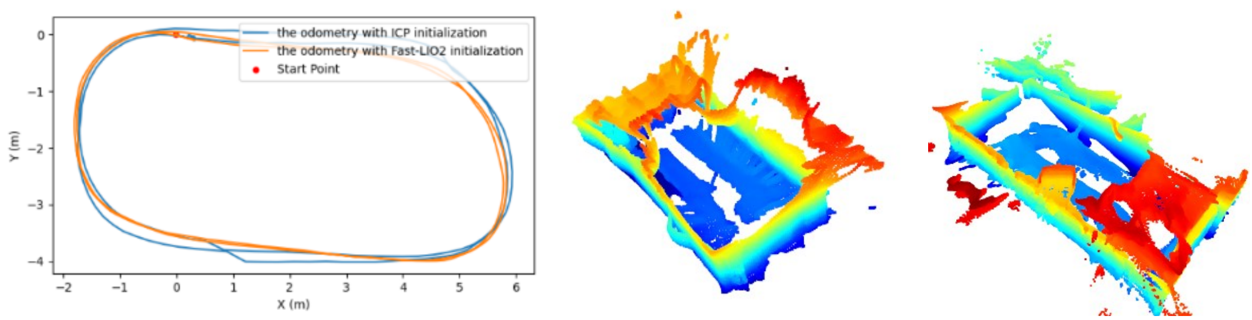
Implicit global neural map representation. We represent the scene using a Multi-layer Perceptron (MLP) and a hierarchical feature grid encoding. During online training, the parameters of the MLP and the grid features are updated to predict the volumetric density for each 3D point. We follow the standard rendering formula to train the network and estimate depth (Isaacson et al., 2023).

Mapping optimisation. The mapping thread of the system receives LiDAR scans and selects a group of keyframes with each new scan. Then, the optimiser updates the network's weights and poses based on the group of keyframes. The optimiser uses the same loss functions, including depth loss and Jensen-Shannon (JS) loss (Isaacson et al., 2023), to optimise the network implicitly representing a global map. Depth loss compares the rendered sparse depth with ground truth sparse depth from LiDAR to learn 3D space density distributions. JS loss is used to constrain the probability distribution in each sampling ray to make the learned 3D space density sharper.

RESULTS

The experiment was conducted on recorded data in a meeting room. A low-cost LiDAR, Velodyne VLP-16, an old version of LiDAR with 16 beams, around $4,000 in 2022, was used to scan the scene. The Velodyne company was acquired by Ouster Company in 2022. Now, the price of the new version of LiDAR, Ouster OS1 with 32 beams is around $8,000, Ouster OS1 with 64 beams is around $12,000, and Ouster OS1-128 beams is around $18,000. Some famous mobile scanners are more expensive: LEICA BLK2GO is priced at around $53,000; the price of FARO Orbis is around $56,000; the cost of ZEB Go is around $33,200.

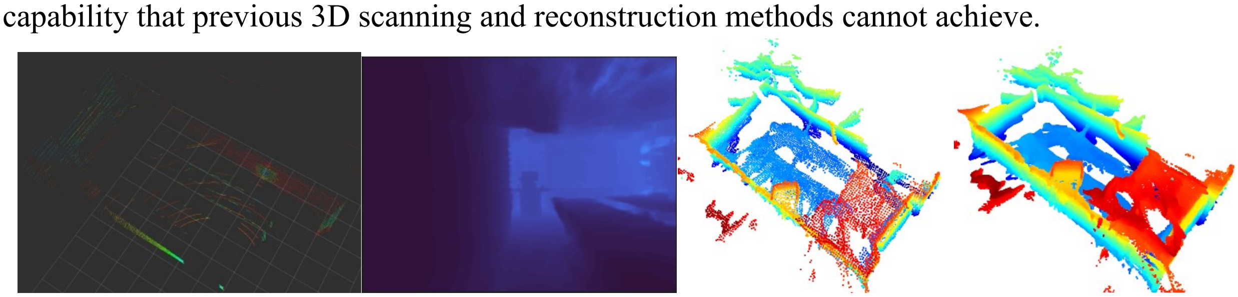
Visualisation of experiment results: A comparison is shown in Figure 2(a) between the poses estimated by Fast-LIO as the front end and those estimated by ICP as the front end. Because the data is recorded by the twice same loop in a meeting room, Fast-LIO, as the front end, provides odometry that better matches the scene. Additionally, we compare the reconstructed maps from the ICP-based pose and Fast-LIO-based pose, as illustrated in Figure 2(b)(c). The Fast-LIO-based front-end method can obtain a denser and more consistent map with real scenes.



(a) (b) (c)

Figure 2: (a) the poses from ICP and Fast-LIO, (b) the dense map with ICP as the front end, and (c) the dense map with Fast-LIO as the front end.

Figure 3(a) shows the sparsity of the raw LiDAR scan. Figure 3(b) shows that the proposed method can estimate dense depth even with sparse point clouds from the cheap LiDAR. Furthermore, one significant advantage of using NeRF as the back end is the ability to construct point cloud maps with varying density levels once the scene is well-trained, as illustrated in Figures 3(c) and 3(d). By adjusting the resolution of the redarning, we can render point clouds with different density levels. This is attributed to the global fitting capability of the neural map, a capability that previous 3D scanning and reconstruction methods cannot achieve.



(a) (b) (c) (d)

Figure 3: (a) one sparse scan, (b) the rendered dense depth map, (c) the point cloud map with a resolution of 0.1, and (d) the point cloud map with a resolution of 0.01.

CONCLUSIONS

We applied the AI-based neural map representation to reconstruct 3D building scenes and proposed a combination of the explicit LiDAR odometry with the global neural map representation. Finally, dense 3D reconstruction of building scenes is achieved using only low-cost LiDAR. This technology holds the potential to reduce the construction costs of digital twins significantly. In the future, we will incorporate detailed metrics and statistical analysis on quantitative evaluation regarding its accuracy and 3D reconstruction fidelity and provide a more balanced view of the trade-offs between cost savings and the achievable level of detail in the reconstructed models.

ACKNOWLEDGEMENT

The authors acknowledge the funding received from the European Union's Horizon Europe research and innovation program under grant agreement No. 101079961 (AEGIR project), No. 955269 (OMICRON project), and No. 958398 (Bim2Twin project).

REFERENCES

Isaacson, S. et al., 2023. LONER: LiDAR Only Neural Representations for Real-Time SLAM. IEEE Robotics and Automation Letters.

Mildenhall, B. et al., 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM, 65(1), 99-106.

Sacks, R. et al., 2018. BIM Handbook: A Guide to Building Information Modeling for Owners, Designers, Engineers, Contractors, and Facility Managers, John Wiley & Sons.

Shan, T. et al., 2020. Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping. In 2020 IEEE International Conference on Intelligent Robots and Systems, pp. 5135-5142.

Sucar, E. et al., 2021. iMAP: Implicit mapping and positioning in real-time. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6229-6238.

Trzeciak, M. P., & Brilakis, I. (2021, July). Comparison of accuracy and density of static and mobile laser scanners. In EC3 Conference 2021 (Vol. 2, pp. 197-203). ETH.

Trzeciak, M. et al., 2023. Dense 3D Reconstruction of Building Scenes by AI-Based Camera–Lidar Fusion and Odometry. Journal of Computing in Civil Engineering, 37(4), 04023010.

Wang, G. et al., 2020. Unsupervised learning of depth, optical flow and pose with occlusion from 3d geometry. IEEE Transactions on Intelligent Transportation Systems, 23(1), 308-320.

Xu, W., & Zhang, F., 2021. Fast-lio: A fast, robust lidar-inertial odometry package by tightly-coupled iterated kalman filter. IEEE Robotics and Automation Letters, 6(2), 3317-3324.

Xu, W., Cai, Y., He, D., Lin, J., & Zhang, F., 2022. Fast-lio2: Fast direct lidar-inertial odometry. IEEE Transactions on Robotics, 38(4), 2053-2073.

Zhu, Z. et al., 2022. Nice-slam: Neural implicit scalable encoding for slam. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 12786-12796).