ICRA2026 Late Breaking Results Poster presented at
2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Hierarchical Grid-based Sensor Pose Extraction for Demonstration Dataset Generation

Doyu Lim, Chaewon Park, and Soohee Han

Abstract— **High-quality 3D reconstruction of unknown small objects with complex surface details is important in applications such as digital preservation and cultural heritage archiving. In practice, such scanning procedures rely heavily on skilled human experts, but the high cost of expert training and the large number of objects requiring digitization make this process difficult to scale. This motivates the need to construct expert demonstration datasets as a foundation for future automated view planning. However, available scan data often contain only frame-level geometry without per-frame sensor poses. To address this issue, we propose a hierarchical grid-based method for extracting sensor poses from frame-based scan data. The proposed method progressively refines candidate poses through coarse-to-fine grid search and selects poses that effectively observe the target surface. Experimental results show an average coverage of 0.85, demonstrating the practicality of the proposed approach for expert demonstration dataset construction.**

I. INTRODUCTION

High-quality 3D reconstruction of unknown small objects with complex surface details has received increasing attention in fields such as digital preservation and cultural heritage archiving [1]. In particular, cultural heritage objects, such as the mounted dishes shown in Fig. 1, often exhibit complex and unstructured surfaces without reliable prior models, making scan quality highly dependent on sensor viewpoint and orientation. For high-performance sensors with a narrow field of view (FOV) and a limited operating range, view planning is therefore essential for precise surface reconstruction.

In practice, such scanning procedures rely heavily on skilled human experts who determine sensor poses by considering object geometry, occlusion, and fine surface details. However, the high cost of expert training and the large number of cultural heritage objects requiring digitization make this manual process difficult to scale. This motivates the need to formalize expert scanning behavior as structured data, laying the foundation for expert demonstration dataset construction and future automated view planning.

A major challenge is that available scan data often contain only frame-level geometric observations, while the corresponding sensor poses are not explicitly recorded for each frame. To address this problem, we propose a hierarchical grid-based method that extracts sensor poses from frame-based scan data through coarse-to-fine refinement. Experimental results show an average Chamfer distance of 8.1 mm, supporting the practicality of the proposed method for demonstration dataset construction.

Doyu Lim, Chaewon Park, and Soohee Han are with the Department of Convergence IT Engineering, Pohang University of Science and Technology (POSTECH), Pohang 37673, South Korea (e-mail: doyulim@postech.ac.kr; cwonpark@postech.ac.kr; soohee.han@postech.ac.kr).

Fig. 1. Examples of scanned objects used in this study. The dataset consists of mounted dishes with diverse shapes and damage patterns.

II. HIERARCHICAL GRID-BASED SENSOR POSE EXTRACTION

The overall grid-based sensor pose extraction pipeline is illustrated in Fig. 2. The proposed method determines the sensor pose through a hierarchical search procedure consisting of three stages: a global stage, a local refinement stage, and a fine refinement stage. As shown in Fig. 2(a), the search begins with broadly distributed candidate grids and progressively focuses on more promising regions, ultimately yielding the selected grid.

Fig. 2(b) presents the procedure executed at each stage. At each stage, k grid points are placed along each axis of the current search region, yielding a total of k^3 candidate grids. These candidates are then filtered according to the sensor working range to remove infeasible grids. For the remaining candidates, visibility is evaluated using the hidden point removal (HPR) operator [2], and the sensor orientation is subsequently optimized for each grid based on the similarity between the surface expected from the estimated pose and the actually observed frame. Each candidate is then assigned a score based on the expected–observed surface similarity, visibility, and the distance from the position estimated in the previous frame. The Top-N candidates are retained and passed to the next stage for further refinement. By varying this parameter across stages, the proposed pipeline achieves broad exploration in the early stage and increasingly focused refinement in the later stages. Through this iterative process, the proposed method progressively concentrates the search on high-quality candidate regions, avoids exhaustive search over the entire 3D space, and finally determines the selected grid.

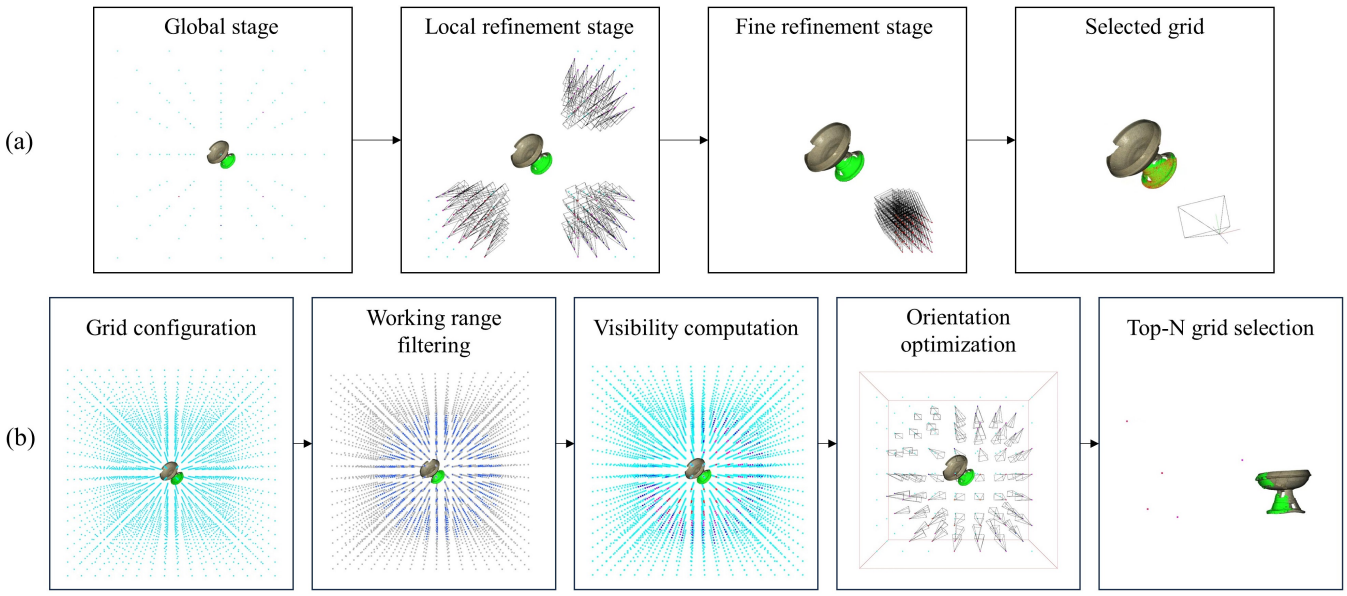

Fig. 2. Overview of the proposed grid-based sensor pose extraction algorithm. (a) The overall pipeline follows a hierarchical coarse-to-fine search strategy consisting of a global stage, a local refinement stage, and a fine refinement stage, which progressively reduce the search space and finally determine the selected grid. (b) At each stage, the same sequence of operations is performed: grid configuration, working range filtering, visibility computation, orientation optimization, and Top-N grid selection. The Top-N grids selected at one stage are passed to the next stage for further refinement.

III. Experiments

A. Expert Demonstration Collection

The proposed dataset consists of 213 demonstrations collected by three human experts, with one demonstration for each object. All objects were scanned using an Artec Space Spider structured-light scanner with a $30^\circ \times 21^\circ$ angular FOV, a 3D resolution of 0.1 mm, and an operating depth of 200–300 mm. The scanned objects are mounted dishes, a type of cultural heritage artifact. As shown in Fig. 1, the dataset includes objects with diverse shapes and conditions, including intact, restored, and fractured examples.

B. Experimental Setup and Parameter Settings

All experiments were conducted on a PC equipped with a 13th Gen Intel(R) Core(TM) i9-13900KF CPU. The algorithm parameters were configured as follows. In the global stage, the parameters were set to $k = 5$ and $N = 3$. For the first frame, the entire search space was explored. For subsequent frames, the search region was restricted to positions within $d = 200$ mm of the pose estimated in the previous frame, in accordance with the physical motion limits of the sensor. Accordingly, the grid spacing in this stage was 40 mm. Orientation optimization was not performed in the global stage. In both the local refinement stage and the fine refinement stage, the parameters were set to $k = 4$ and $N = 1$. The resulting grid spacings were 10 mm and 2.5 mm, respectively. For the selected grid, only orientation optimization was performed.

C. Accuracy

The accuracy of the sensor pose extraction algorithm is evaluated by comparing the point cloud of the object surface expected to be observed from the extracted sensor pose with the point cloud actually observed in the corresponding frame. Table I reports the Chamfer distance (CD) [3], 95th-percentile Hausdorff distance (HD95), Precision, Recall, and F-score. The F-score is computed as the harmonic mean of Precision and Recall.

TABLE I
Quantitative accuracy of the extracted sensor poses

CD (mm)	HD95 (mm)	Precision (%)	Recall (%)	F-score (%)
8.1	24.4	70.5	66.0	68.2

IV. Conclusions

This paper presents a hierarchical grid-based method for extracting sensor poses from frame-based scan data to construct an expert demonstration dataset. The proposed method progressively refines candidate poses and selects those that effectively observe the target surface without requiring explicit per-frame pose information. This makes it possible to convert frame-only scan data into structured expert demonstrations for future learning-based view planning. Future work will use the constructed dataset to support automated view planning through imitation learning.

References

[1] I. D. Lee, J. H. Seo, and B. Yoo, "Autonomous view planning methods for 3D scanning," *Automation in Construction*, vol. 160, no. 105291, 2024.

[2] S. Katz, A. Tal, and R. Basri, "Direct visibility of point sets," *ACM Transactions on Graphics*, vol. 26, no. 3, pp. 24—es, 2007.

[3] T. Wu, L. Pan, J. Zhang, T. Wang, Z. Liu, and D. Lin, "Density-aware chamfer distance as a comprehensive metric for point cloud completion," *arXiv preprint arXiv:2111.12702*, 2021.