URSI AP-RASC, Sydney, Australia, 17 – 22 August 2025

High-Precision 3D Point Cloud Reconstruction and Annotation for Multi-Object Scenes

Yanchi Liu (1), Yifei Ma (1), Xiaoya Zhou (1), Xiaoyong Liu (1)(2), Yuqi Lv (2), Hao Du (2), Dan Shi* (1), Canghao Chen* (3)
(1) School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing, 100876, China
(2) The State Radio Monitoring Center Testing Center, Beijing, 100037, China
(3) Southampton Business School, University of Southampton, Southampton, United Kingdom

Abstract

Using three-dimensional (3D) environmental simulation techniques to analyze electromagnetic (EM) environment distribution and predict EM radiation is an effective method for comprehensive and efficient EM management and spectrum optimization. Establishing a 3D model that closely aligns with the real environment is a prerequisite for accurate EM environment prediction and a critical factor influencing the accuracy of simulation results. However, existing modeling methods face challenges such as long modeling times, extensive manual operation, low model accuracy, and poor generalization. This paper proposes a high-precision method for 3D point cloud reconstruction and annotation for multi-object embedded scenes, utilizing a combination of millimeter-wave radar and visual sensors. This approach effectively distinguishes between objects that occupy similar spatial positions but are composed of different materials, making it suitable for multi-object scenarios. The generated 3D model, annotated with material EM parameters, achieves a labeling accuracy of 90.94%, thereby providing high-precision input models for EM simulation and prediction.

1. Introduction

With the widespread adoption of wireless technologies, the exponential increase in frequency equipment has presented significant challenges to EM environment prediction. 3D reconstruction techniques, capable of providing spatial structural information about environments, are vital for analyzing EM distribution and forecasting signal propagation paths. Current EM simulation software primarily employs two methods for 3D modeling: manual drawing and importing 3D maps [1]. However, these methods suffer from low accuracy, hindering rapid and precise outdoor modeling of environments and materials. Additionally, they lack real-time capability and EM parameters of materials, requiring cumbersome manual input process. EM-characterized 3D environment mapping via reconstruction has emerged as a critical solution. While vision and radar-based methods dominate 3D reconstruction, their single-sensor architectures exhibit critical constraints in more complex EM environments.

The advent of the multimodal concept has spurred the increasing application of multi-sensor fusion techniques in 3D reconstruction and simulation. Multi-sensor fusion integrates information from different sensors, leveraging their respective advantages to improve system accuracy. For instance, [2] introduces a multi-dimensional DBSCAN clustering method based on millimeter-wave radar Doppler characteristics, processing radar data effectively. [3] utilizes simultaneous localization and mapping (SLAM) technology to fuse camera and radar data, enabling 3D reconstruction of physical environments to monitor the state of an entire field or individual plants. [4] and [5] propose a high-precision depth estimation method using a deep convolutional neural network (CNN) architecture, which combines sparse 3D LiDAR and dense stereo depth information, achieving higher efficiency and precision in 3D reconstruction. While these methods enable faster and more accurate reconstruction of environments, they do not incorporate material EM parameters and thus cannot be directly applied to EM environment prediction. [6] proposes a method for 3D point cloud segmentation and reconstruction by fusing millimeter-wave radar and visual data, enabling the annotation of material EM parameters based on 3D reconstruction. However, this approach is limited to simple scenes with single objects and cannot handle complex scenarios with multiple targets.

This paper presents an accurate 3D point cloud annotation and reconstruction method for multi-object embedded scenes. First, a DBSCAN algorithm incorporating reflection intensity significantly improves clustering accuracy, reducing Intensity-MAD by 61.80%. Then, the Dual-SIFT method addresses issues of duplicate labels within frames and inconsistent labels across frames, reducing the matching time by 82.60%. Finally, Multiview mode-based label mapping maps 2D semantic labels to 3D point clouds, achieving a labeling accuracy of 90.94%. The multi-target 3D models, annotated with material EM parameters, can be used for EM simulation and prediction, providing high-precision input models for EM environment analysis.

2. Method description

This section introduces the principles of the DBSCAN algorithm based on weighted reflection intensity and dual-SIFT matching, which serve as the foundation for point cloud clustering and semantic segmentation. Additionally, we propose a high-dimensional label mapping approach and outline the process of mapping the 2D segmentation model to the 3D point cloud model with EM parameters.

2.1 DBSCAN algorithm based on weighted reflection intensity

The millimeter-wave radar provides spatial coordinate (x, y, z), target velocity v, reflection intensity I, and sampling time t. Velocity-based filtering removes non-zero velocity points to eliminate interference from moving objects. To balance the influence of each feature dimension on the clustering results, zero-mean normalization is applied:

$$z = \frac{x - \mu}{\sigma} \quad (1).$$

where x represents the spatial coordinates x, y, z and the reflection intensity I, μ and σ denote the mean and standard deviation of each feature dimension, respectively.

The Euclidean distance is used to measure the distance between the normalized features of each point cloud:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 + (I_2 - I_1)^2} \quad (2).$$

where x_i, y_i, z_i and I_i represent the target object's horizontal coordinate, vertical coordinate, height, and reflection intensity, respectively.

The traditional DBSCAN algorithm relies solely on spatial features, limiting its ability to distinguish spatially close objects. Therefore, this paper proposes a weighted DBSCAN algorithm incorporating reflection intensity, assigning different weights to each feature dimension:

$$d = \sqrt{w_1(x_2 - x_1)^2 + w_2(y_2 - y_1)^2 + w_3(z_2 - z_1)^2 + w_4(I_2 - I_1)^2} \quad (3).$$

where w_1, w_2, w_3 and w_4 represent the weights for each feature.

This paper uses the Davies-Bouldin index as an inter-cluster evaluation metric:

$$DBI = \frac{1}{N}\sum_{i=1}^{N} max_{j \neq i}\left(\frac{\bar{S_i} + \bar{S_j}}{d_{ij}}\right) \quad (4).$$

where N represents the number of clusters, $\bar{S_i}$ and $\bar{S_j}$ are the average intra-cluster distances between clusters i and j, and d_{ij} represents the distance between the centroids of clusters i and j. The smaller the Davies-Bouldin index, the larger the inter-cluster distance, indicating better clustering performance.

Additionally, we propose Intensity-MAD as an intra-cluster evaluation metric. Intensity-MAD calculates the mean absolute deviation (MAD) using the reflection intensity, and evaluates the similarity of objects within the cluster:

$$Intensity - MAD = \frac{1}{n}\sum_{i=1}^{n} |I_i - \bar{I}| \quad (5).$$

where n represents the number of points in the cluster, I_i represents the reflection intensity of a radar point, and $\bar{I}$ represents the mean reflection intensity of the cluster. The lower the Intensity-MAD, the higher the similarity of objects within the cluster, indicating better clustering performance.

The proposed method achieves an Intensity-MAD value of 0.34, which shows a 61.80% reduction from 0.89 with traditional DBSCAN, demonstrating superior clustering accuracy for multi-dimensional radar data. By integrating reflection intensity, this method enhances the separation of spatially proximate objects, achieving better clustering performance in multi-object embedded scenarios.

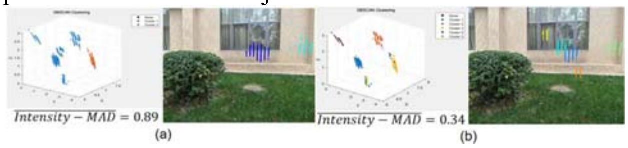

Figure 1. Clustering results and intra-cluster similarity: (a)original DBSCAN algorithm. (b) our method.

2.2 Multi-object segmentation and labeling via radar and dual-SIFT

After completing the clustering preprocessing, the 3D radar point cloud is mapped from the radar coordinate system to the pixel coordinate system. For a radar point (X_w, Y_w, Z_w), the corresponding 2D-pixel point (u, v) in the pixel coordinate system can be obtained:

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \quad (6).$$

where f_x and f_y are the camera focal lengths, c_x and c_y are the principal point coordinates, and R and T are the camera's rotation and translation matrix, respectively.

This paper is based on the SAM model, utilizing radar point clouds after clustering and projection as prompt points to achieve automated multi-object scene segmentation. The reflection intensity is stored as labels in the pixels covered by the segmented mask.

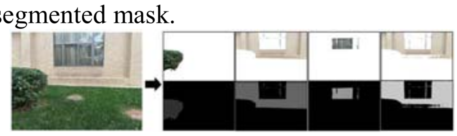

Figure 2. 2D segmentation results from one perspective.

Due to the sparsity of millimeter-wave radar point clouds, multi-frame fusion is employed to increase point density for segmentation. However, the same object may correspond to multiple clusters, and radar labels lack semantic information, causing repeated segmentation.

Additionally, clustering labels for the same object differ across perspectives. The SIFT (Scale-Invariant Feature Transform) algorithm, widely used for image recognition and 3D reconstruction, offers rotation and scale invariance. Leveraging SIFT, this paper proposes dual-SIFT labeling and deduplication to unify, correct, and eliminate redundant object labels across different perspectives.

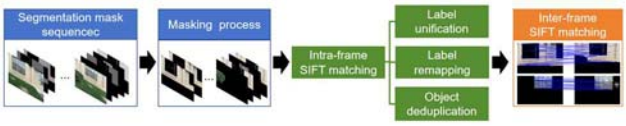

Figure 3. Process of multi-object segmentation and labeling via radar and dual-SIFT.

First, intra-frame SIFT matching is performed to match repeated segmentation images of the same object within the same view, followed by label remapping for consistency. Duplicate objects within each frame are removed by retaining the one with the most feature points. The comparison of multi-target objects before and after intra-frame matching is shown in Table I.

TABLE I

	Obj1	Obj2	Obj3	Obj4
Original label	0	1	2	3
Label after re-projection	0	1	2	1
Label after deduplication	0	1	2	x

Next, inter-frame SIFT matching is performed across consecutive perspectives to correct labels of the same objects. Simultaneously, new objects that appear in later perspectives but not in previous ones are assigned new object labels, enabling multi-target scene handling. The number of objects and matching iterations before and after our dual-SIFT method is shown in Table II.

TABLE II

	Number of objects	Number of matches
Raw data + SFT	26	650
Data + dual-SIFT	16	113

This method addresses the issues of label confusion caused by duplicate objects within the same frame and label mismatches between frames. It also significantly reduces the overall matching computation time, with an 82.60% reduction in total.

2.3 Multiview mode-based label mapping

The masks of all target objects, obtained through 2D segmentation for each view, are aggregated to form the complete mask for the respective perspective. Each pixel in the image is assigned a corrected object label. However, at this stage, all labels correspond to pixel points in the 2D image plane. Consequently, the reconstructed 3D point cloud model does not yet incorporate this labeling information.

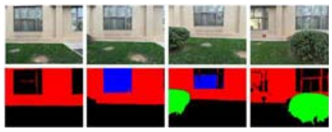

Figure 4. The labeled 2D segmentation results from different perspectives.

This paper proposes the multiview mode-based label mapping method that extracts reflection intensity semantic labels from multi-view images using camera projection transformations and then fuses label to generate labeled 3D point clouds. A 3D cloud point $P(x, y, z)$ is first converted to homogeneous coordinates $P_{hom}(x, y, z, 1)$, and then projected onto the 2D image plane using the camera intrinsic and extrinsic parameters. After projection, the 2D image plane coordinates Q' are normalized.

$$Q = K[R|T]P_{hom} \quad (7).$$

$$Q' = \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} Q_x/Q_z \\ Q_y/Q_z \end{bmatrix} \quad (8).$$

The reflection intensity label l of Q' obtained from the projection is extracted. For the 3D point P, which corresponds to n different views, the label sequence $label(P) = \{label_1, label_2, \dots, label_n\}$ from the 2D points is collected. The label with the highest occurrence from the set of non-background labels is chosen as the label for P:

$$label'(P) = \{label_i | label_i \neq NaN, i = 1,2, \dots, n\} \quad (9).$$

$$l_p = mode(label'(P)) \quad (10).$$

where $label'(P)$ represents the set of 2D point labels visible from n different views, and l_p represents the final label for the 3D point P.

This method enables the automated mapping and labeling of 3D reconstructed point clouds by projecting 2D labels from each view in the 2D pixel space to the 3D point cloud space.

3. Results and discussion

The experimental scenario in this paper is a multi-target embedded scene measuring 12 × 6 × 8 meters, consisting of a teaching building and the lawn in front of it. The objects in the scene include the building's walls, windows embedded in the walls, the lawn in front of the building, and shrubs on the lawn. The materials involved are categorized as buildings, glass, ground, and plants. The scene is scanned using a 3D radar-vision fusion integrated system, constructed with the TI AWR1843 radar and a monocular camera.

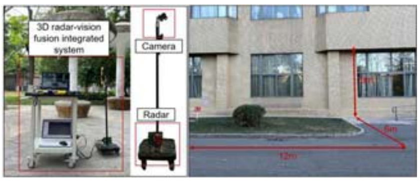

Figure 5. The experimental system and scenario.

After applying the proposed segmentation and labeling method, along with multiview mode-based label mapping, the labeling accuracy of the 3D point cloud reached 90.94%. Subsequently, surface reconstruction of the point cloud was performed using Poisson triangulation, and the reconstructed model was imported into the EM simulation platform for ray tracing calculations. The Poisson-reconstructed model had dimensions of approximately 12m × 6m × 8m, containing 59,927 vertices and 119,698 triangles, along with four types of EM parameters representing different materials in the environment. The transmitter and receiver were deployed at (15.00, -5.00, 7.10) and (-5.00, -5.00, 10.28) respectively, utilizing a half-wave dipole antenna configuration. The radiation source was calibrated to 20 W with omnidirectional radiation patterns ($\theta = 0°, \varphi = 0°$) aligned to the global coordinate system. The simulation result is shown in Figure 7.

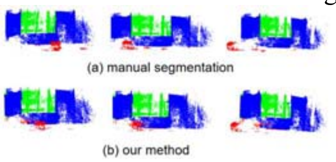

Figure 6. Comparison with the manually segmented results.

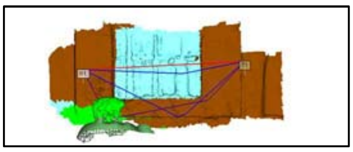

Figure 7. EM environments simulation results.

The 3D reconstruction model with EM parameters enables fast and high-precision ray tracing calculations. The tracing results show that the radio waves emitted by the transmitter bypassed obstacles and reached the receiver through four different paths: one direct path and three reflected paths. The results in the figure further indicate that in areas without obstructions, the radio waves reached the receiver via the direct path, while in areas with obstacles, the waves experienced reflection, forming multiple reflected paths. The results demonstrate that the proposed method can accurately reconstruct real scenarios, enabling precise predictions of radio wave propagation.

4. Conclusion

This paper presents an improved DBSCAN utilizing millimeter-wave radar point cloud reflection intensity, combined with radar prompt segmentation, dual-SIFT label deduplication, and multiview mode-based label mapping for automated 3D point cloud annotation in multi-object scenes. The enhanced DBSCAN reduced Intensity-MAD by 61.80%. Then, the radar and dual-SIFT method resolve intra-frame duplicates and inter-frame inconsistencies, reducing matching time by 82.60%. Finally, the multiview mode-based label mapping method automates 2D to 3D semantic label mapping. Experiments demonstrate 90.94% annotation accuracy in a scene measuring 12×6×8 meters. Furthermore, the Poisson reconstruction 3D model with EM parameters of materials was imported into an EM simulation platform. Simulation results indicate that the final model effectively supports radio wave propagation prediction, providing rapid and accurate input for EM environment analysis in complex multi-object scenarios.

5. Acknowledgements

This study was supported by Key Laboratory of Radio Spectrum Testing Technology, Ministry of Industry and Information Technology.

References

[1] I. Alawneh, J. Barowski and I. Rolfes, "Measuring the Permittivity of Dielectric Materials by Using 140 GHz FMCW Radar Sensor," 2019 13th European Conference on Antennas and Propagation (EuCAP), Krakow, Poland, 2019, pp. 1-4.

[2] H. Li, C. Fu, G. Xi, Q. Wang and H. Peng, "Multi-Dimensional DBSCAN Clustering Approach for Millimeter Wave Radar Point Cloud Data," 2024 International Conference on Microwave and Millimeter Wave Technology (ICMMT), Bejing, China, 2024, pp. 1-3.

[3] L. Meyer, J. Gedschold, T. E. Wegner, G. Del Galdo and A. Kalisz, "Enhancement of Vision-Based 3D Reconstruction Systems Using Radar for Smart Farming," 2022 IEEE Workshop on Metrology for Agriculture and Forestry (MetroAgriFor), Perugia, Italy, 2022, pp. 155-159.

[4] K. Park, S. Kim and K. Sohn, "High-Precision Depth Estimation with the 3D LiDAR and Stereo Fusion," 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, 2018, pp. 2156-2163.

[5] K. Park, S. Kim and K. Sohn, "High-Precision Depth Estimation Using Uncalibrated LiDAR and Stereo Fusion," in IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 1, pp. 321-335, Jan. 2020.

[6] F. Yang, L. Zhan, H. Wei, L. Chengyu and D. Shi, "Mmwave radar and vision fusion for semantic 3D reconstruction," 2023 IEEE 7th International Symposium on Electromagnetic Compatibility (ISEMC), Hangzhou, China, 2023, pp. 1-4.