Adaptive Point Cloud Clustering Algorithm for Practical Roadside MmWave Radar Systems

Luyi Zhang*, Jinhang Zhang*, Haixin Shi*, Lu Gao*, Xiaopeng Hu*† and Rui Chen*
*Guangzhou Institute of Technology, Xidian University, Guangzhou 510700, China
†Guangzhou INFOHAND Technology Co., Ltd.
Email: luyizhang@stu.xidian.edu.cn, rchen@xidian.edu.cn

***Abstract*—Millimeter-Wave radar has been widely applied in the field of autonomous driving due to an excellent performance under complex weather conditions. However, in practical roadside scenarios, the challenge of sparse point clouds leading to clustering difficulties and the issue of large vehicle point clouds dispersing, resulting in fragmentation, currently hampers the practical application of radar sensors. We propose an adaptive point cloud clustering algorithm based on DBSCAN. First, we propose an improved DBSCAN clustering algorithm based on distance and speed thresholds, which enhances the differentiation of point clouds between different vehicles, and an adaptive ellipse gate strategy to solve the large vehicle point clouds fragmentation problem. Then, a secondary clustering algorithm based on azimuth is exploited, effectively addressing the issues of large vehicle fragmentation and anomalous speed values. Practical roadside experimental results demonstrate that our proposed algorithm significantly outperforms traditional algorithms, showing considerable potential in practical applications.**

***Index Terms*—Millimeter-Wave radar, roadside, MmWave, point cloud, clustering, DBSCAN, wave gate**

I. Introduction

Millimeter-Wave (MmWave) radar offers unique advantages in challenging weather conditions, low-light environments, and long-range detection, ensuring vehicular safety under diverse meteorological and illumination circumstances. Therefore, MmWave radar, known for its high resolution and compact design, is widely used in vehicles. In recent years, roadside MmWave radar becomes more and more popular because it can boast a broader field of vision, and extended detection range as a complement of automotive radar [1], [2]. However, the point cloud data of roadside MmWave radar exhibit sparse characteristics and is prone to be effected by traffic velocity, detection distance, weather conditions, and other environmental interferences, which brings challenges in point cloud clustering, necessitating specialized algorithms and methodologies for effective processing.

The DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm [3] is a popular clustering method for analyzing MmWave radar data, based on which a number of adaptations aimed at improving its application in traffic environment vehicle clustering have been proposed [4]–[11]. A robust, adaptive radar point cloud clustering algorithm, termed Radar Elliptic Density-Based Spatial Clustering of Applications with Noise (REDBSCAN), was introduced by Zhang et al. [4]. A radar point cloud clustering algorithm with adaptive clustering parameters, named Leaf-DBSCAN, was proposed by Sheikh et al. [5]. The algorithm modifies the clustering strategy to reduce frequent distance queries between target locations formed during clustering, and adapts the clustering parameters according to the spatial distribution of points. ADBSCAN, an improvement upon DBSCAN's handling of uneven LiDAR point cloud distributions with a distance-adaptive clustering radius, was introduced by Jin et al. [6]. An algorithm for dynamic environments such as autonomous driving, which auto-estimates DBSCAN parameters to achieve results comparable to the original without manual adjustments, was created by Mohammad et al. [7]. A method for boundary detection and plane segmentation in 3D point clouds, which enhances DBSCAN for effective plane fitting, was proposed by Chen et al. [9]. Additionally, Campello $et al.$ [10] introduced HDBSCAN, a hierarchical density-based clustering algorithm that simplifies the merging process through a cluster candidate tree, suitable for various applications.

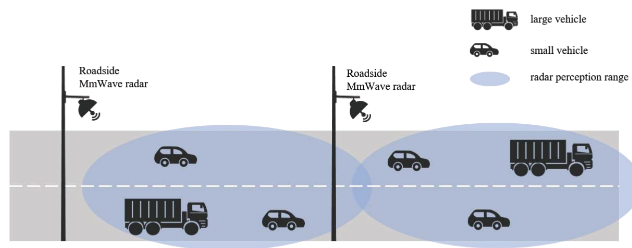

Fig. 1: Scenario of roadside MmWave radar deployment.

However, in the practical context of road environments, clustering point clouds presents significant challenges. To address the issues of large vehicle fragmentation and point cloud sparsity encountered in real-world road scenarios, we propose an adaptive point cloud clustering algorithm based on DBSCAN clustering, utilizing multidimensional information derived from point clouds. Our algorithm employs an improved DBSCAN based on distance and speed, which expands the differentiation of point clouds between different vehicles. We proposes an adaptive ellipse gate strategy for better alignment with vehicle point cloud shapes. In order to improve the accuracy of the clustering and to reduce the probability of fragmentation, we also add an inter-cluster merging strategy after the clustering algorithm.

II. ROADSIDE MMWAVE RADAR CLUSTERING

A. The Role of Clustering in Roadside Radar System

The roadside MmWave radar system is an important component in achieving intelligent transport, as shown in Figure 1. The clustering algorithm plays a crucial role in the extended point cloud target segmentation and tracking process of MmWave radar data processing. The clustering algorithm precisely segments point cloud clusters of different vehicles by processing MmWave radar detection vehicle point cloud data, so as to accurately distinguish the vehicle data in the detection area and improve the tracking accuracy and system detection capability. Accurately segmenting different vehicle point cloud information on complex traffic roads poses a challenge for the clustering algorithm of roadside MmWave radar.

B. Algorithm framework

Given that the point cloud detected by MmWave radar is sparse and has limited information (including distance, velocity, azimuth and RCS), we analyse the multidimensional feature information and distribution characteristics of the point cloud data, and propose an adaptive point cloud clustering algorithm based on three main modules: clustering based on distance and velocity, adaptive ellipse gate and cluster merging based on azimuth, as shown in Figure 2.

Our algorithm dynamically adjusts the ellipse gate based on point cloud RCS to accommodate the density distribution characteristics of various vehicle point clouds. By integrating distance and speed information, our algorithm segments vehicle target point clouds to achieve optimal clustering results. The algorithm sets a minimum number of points for clustering based on distance to accurately filter out noise points and reduce the rate of missed vehicle targets. At last, we employs cluster merging based on azimuth to address the issues of large vehicle fragmentation and the presence of outliers speed points within clusters. Our algorithm has been tested under real road conditions, demonstrating superior accuracy in clustering results.

III. ADAPTIVE POINT CLOUD CLUSTERING ALGORITHM

A. Clustering Based on Distance and Velocity

The traditional DBSCAN cluster algorithm uses ϵ as the neighborhood distance threshold, however, in real traffic scenarios, speed is also an extremely important feature information used to differentiate between different vehicles. Therefore, we incorporate velocity as an additional dimension threshold and calculate the points that falling into the set

$$\mathbb{N}_\epsilon(x_j) = \{x_i \in \mathbb{D} | \mathrm{dis}(x_i, x_j) \le \epsilon_d, \mathrm{vel}(v_i, v_j) \le \epsilon_v\}, \quad (1)$$

where $\mathbb{D}$ is the set of all points, ϵ_d is the distance threshold, ϵ_v is the velocity threshold, $\mathrm{dis}(x_i, x_j)$ is the distance difference between point i and point j and $\mathrm{vel}(x_i, x_j)$ is the velocity difference between point i and point j.

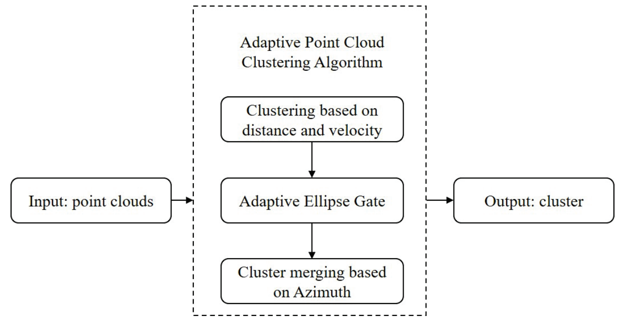

Fig. 2: Framework of roadside MmWave radar clustering algorithm.

Furthermore, when addressing the issue of determining the MinPts required to form clusters at varying distances, selecting the appropriate $\mathrm{MinPts}$is uniquely challenging. The DBSCAN algorithm excels at eliminating noise, making it generally advisable for MinPts to be no fewer than two. However, for roadside MmWave radar detection echoes, the uneven density of the target point clouds, transitioning from dense to sparse from near to far, results in significant characteristic differences in vehicle targets. Adhering strictly to traditional DBSCAN rules may lead to the erroneous classification of target vehicles as noise and their subsequent exclusion, thereby increasing the rate of missed detections, expanding blind spots, and reducing the maximum detection range. Consequently, it is essential to adjust the setting of MinPts required for clustering according to the specific characteristics of different areas as

$$\mathrm{MinPts} = \begin{cases} 2, & \text{if } x < D \\ 1, & \text{if } x > D \end{cases}, \quad (2)$$

where x represents the radial distance of the point, and D is a fixed value. The value of MinPts and D has been selected empirically,and the value of D is generally set as 50m. The parameters of D and MinPts are derived from extensive empirical measurements and engineering adjustments based on radar deployment, encompassing a wide array of environments of roads. The method demonstrated outstanding results in practical roadside MmWave radar clustering experiments.

B. Adaptive Ellipse Gate

The DBSCAN algorithm aggregates data points within a specific density range into clusters. The density range represents a predefined region in space, commonly described as a "beam gate" in the radar domain. Radar systems typically employ circular beam gates, with the observed target as the centre. The determination of whether surrounding data points are within the target's density range is based on assessing whether their distance from the target is less than the preset radius. However, considering the practical scenario of vehicle point clouds, where the shape of vehicles is rectangular, adopting

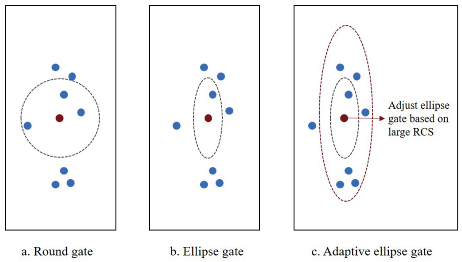

Fig. 3: The effect of different gates on point cloud clustering.

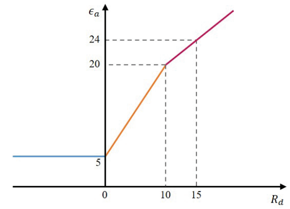

Fig. 4: ε_a varies with the R_d.

an ellipse gate is more appropriate. As shown in Figure 3, adaptive ellipse gate that better conform to the distribution of vehicle-shaped point clouds results in more effective clustering than the original DBSCAN's circular gates. We enhance the ellipse gate's performance by introducing an adaptive ellipse gate, adjusting its major and minor axes based on distance and RCS values. This approach addresses the issue of large vehicle fragmentation and improves clustering outcomes.

The calculation of the ellipse distance between point i and point j can be expressed as

$$\mathrm{dis}(i,j)=\sqrt{\frac{|x_i-x_j|^2}{\varepsilon_a^2}+\frac{|y_i-y_j|^2}{\varepsilon_b^2}}, \tag{3}$$

where in the radar coordinate system, x_i and x_j are the horizontal coordinates of point i and point j, y_i and y_j are the vertical coordinates of point i and point j, and ε_a is the major axis and ε_b is the minor axis to be adjusted by the point cloud's RCS.

In real-world traffic scenarios, large and small vehicles necessitate distinct clustering radius, as employing a uniform radius may result in segmentation issues for larger vehicles. To tackle this challenge, it's pertinent to acknowledge that large and small vehicles exhibit different RCS, thereby warranting an adjustment in clustering radii to accommodate their unique values. Given the variability of RCS detected by MmWave radar in traffic vehicles—attributable to a multitude of factors, such as significant fluctuations within the radar's detection range—a smooth filtering application reveals distinct characteristics in these variation curves. Analyzing a diverse array of vehicles, types, and scenarios enables a rough differentiation of vehicular characteristics. Within a span from 0m to 70m, vehicles fall within the radar's near-field beam, where their RCS precipitously declines; beyond 70m, the targets transition into the radar's far-field beam, exhibiting more stable features. Hence, we statistically analyze the RCS variation curves across various vehicles and scenarios, taking the RCS variation curve of small vehicles as a point of reference for averaging, and employ Fourier series to model the function of small vehicle RCS variation with distance. The Fourier series fitting algorithm [12], a powerful and flexible mathematical tool, is adept at breaking down complex functions into a series of simpler sine and cosine functions. Consequently, we utilize the trigonometric form of the Fourier series for fitting, and have the average RCS reference value R_b as

$$\begin{aligned} R_b =& a_0+a_1\cos(\omega x)+b_1\sin(\omega x)+a_2\cos(2\omega x)+ \\ & b_2\sin(2\omega x)+a_3\cos(3\omega x)+b_3\sin(3\omega x), \end{aligned} \tag{4}$$

where a_0, a_1, a_2, a_3, b_1, b_2, b_3 and ω represent the numerical values obtained following the Fourier fitting process, R_b represents the RCS calculated on the formula. Subsequently, we will compare RCS of the point cloud cluster with the R_b, calculating the difference between the two. The value of R_d =RCS$-R_b$ will serve as the basis for adjusting the radius of the clustering ellipses. If the R_d is large, we surmise that the vehicle is likely a large one, necessitating an increase in ε_a. Given that the distinction between large and small vehicle point cloud data is not markedly evident along the ellipse longitudinal axis, we primarily adjust the clustering radius of the horizontal axis based on the R_d, as illustrated in Figure 4. Based on this scheme, we can calculate the adjusted adaptive ellipse gate. The function for adjusting ε_a is also obtained from empirical measurements and has been extensively tested and fine-tuned on roads.

C. *Cluster Merging Based on Azimuth*

Building upon the initial clustering steps, we have observed instances which becomes challenging for large vehicles to cluster precisely. Furthermore, the radar's detection of velocity outliers within these clusters may have implications in subsequent tracking efforts. It is imperative to incorporate an additional clustering phase to address these issues. The essence of this second phase lies in our belief that clusters encapsulate more information than individual point clouds [13], thus employing distinct data from the first phase. Hence, we utilize azimuth to determine the feasibility of merging two clusters.

Assuming cluster A and cluster B are obtained through clustering, with the shortest Euclidean distance points in the two clusters being points i and j, where $i \in A$ and $j \in B$, the a priori condition for the merger of clusters A and cluster B is

$$d_{\min} < \epsilon_d \wedge \triangle x < \epsilon_x \wedge \triangle y < \epsilon_y. \tag{5}$$

The $d_{\min}$ is the Euclidean distance between point i and point j, and $\triangle x$ and $\triangle y$ represent the differences in the horizontal and vertical distances between point i and point j. ϵ_d, ϵ_x, and ϵ_y are the predefined thresholds for Euclidean distance, lateral distance, and longitudinal distance, respectively. Two clusters that do not meet the above conditions cannot be merged.

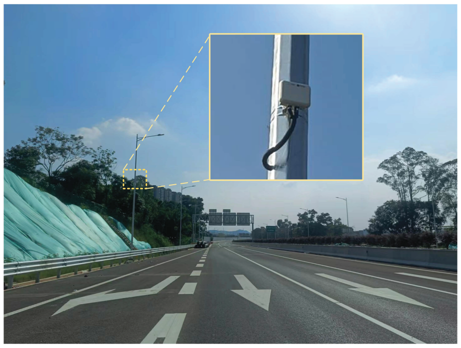

Fig. 5: Experimental scenario of roadside MmWave radar.

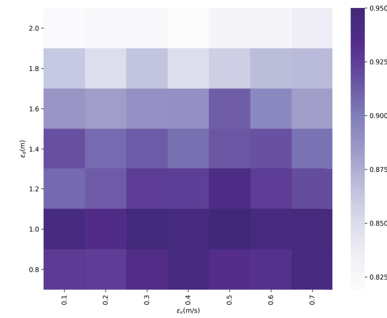

Fig. 6: Heatmap of parameter estimation.

If point i and point j satisfy the above conditions, then comparing the azimuth of point i and point j, if the following condition is met

$$|a_i - a_j| < \epsilon_a, \tag{6}$$

where a_i and a_j is the azimuth of point i and point j, and ϵ_a is the azimuth merge threshold. In this case, clusters A and B are merged and considered as one cluster.

IV. Experiments and Results

This section assesses the algorithm designed and presents the optimized parameter selection. All experiments are conducted based on a real dataset collected from the CongPu Expressway, as shown in Figure 5. The dataset is divided into two non-overlapping segments, each of equal size; one is utilized for parameter estimation, and the other for evaluation. The efficacy of the adaptive point cloud clustering algorithm is evaluated using the V_1-score.

A. Parameter Estimation

In the parameter estimation phase, a comprehensive enumeration search was performed on all thresholds of the adaptive point cloud clustering algorithm to ensure that each combination of thresholds falls within a reasonable preset range. For the initial clustering stage, the velocity threshold, denoted as ϵ_v was explored within a range from $0.1\mathrm{m/s}$ to $0.7\mathrm{m/s}$, and the distance threshold ϵ_d ranged from $0.8\mathrm{m}$ to $2\mathrm{m}$. In the subsequent clustering stage, the azimuth threshold ϵ_a was ranged from $0.5°$ to $2°$, while the lateral and longitudinal distance thresholdsϵ_y, respectively, were determined to be from $1\mathrm{m}$ to $10\mathrm{m}$.

The objective of the parameter estimation process is to enhance the precision and robustness of the clustering algorithm. Accurate parameter estimation enables the algorithm to identify and cluster similar data points more effectively, reducing misclassification and thereby significantly improving the algorithm's accuracy and reliability. Moreover, sensible parameter settings can diminish the computational load of the algorithm, augmenting processing speed and thereby enhancing efficiency. A meticulous parameter estimation process was conducted to thoroughly explore and evaluate the performance of the adaptive point cloud clustering algorithm, with the outcomes displayed in Figure 6.

This figure illustrates the variation in the algorithm's accuracy across different parameter combinations, where the deeper purple regions indicate higher accuracy, and the lighter purple regions denote lower accuracy. Through this process, it was discovered that a parameter combination of a $1\mathrm{m}$ distance threshold and a $0.5\mathrm{m/s}$ velocity threshold exhibited optimal performance, a finding visually corroborated by the deep purple optimum point in the heatmap. The same methodology was applied for the parameter estimation of the second clustering stage, ultimately identifying the optimal parameter combination as a $1°$ azimuth threshold, a $1\mathrm{m}$ lateral distance threshold, and a $5\mathrm{m}$ longitudinal distance threshold.

B. Evaluation Metric

The $\mathrm{V1\text{-}measure}$ is selected as the evaluation metric for our experiments, a criterion designed to assess the effectiveness of clustering, especially in unsupervised learning scenarios where the ground truth is known. Comprised of homogeneity and completeness, it aims to holistically appraise the quality of clustering outcomes. Homogeneity measures whether a cluster contains only a single actual class data point—achieving its maximum when each cluster consists solely of a single class. Completeness evaluates whether all data points of an actual class are assigned to the same cluster, reaching its maximum when all data points of a class are included in a single cluster.

$$\mathcal{H} = 1 - \frac{H(C|K)}{H(C)}. \tag{7}$$

Here, $H(C|K)$ represents the conditional entropy of the true classes given the clustering result, and $H(C)$ is the entropy of the true classes. This ratio illustrates how clustering information reduces the uncertainty of true class information.

$$\mathcal{C} = 1 - \frac{H(K|C)}{H(K)}. \tag{8}$$

Here, $H(K|C)$ is the conditional entropy of the clustering result given the true classes, and $H(K)$ is the entropy of the clustering results. This ratio indicates how true class information reduces the uncertainty of the clustering outcome.

TABLE I: Detection Clustering Experiments along with The Best Determined Parameters and Scores

#	Algorithm	V_1-score	Optimized Parameter Set
1	baseline-DBSCAN	65.18%	$\epsilon_d = 1.0\mathrm{m}$
2	distance and velocity DBSCAN	68.52%	$\epsilon_d = 1.0\mathrm{m}, \epsilon_v = 0.5\mathrm{m/s}$
3	distance and velocity DBSCAN with adaptive gate	70.02%	$\epsilon_d = 1.0\mathrm{m}, \epsilon_v = 0.5\mathrm{m/s}$
4	**adaptive point cloud clustering algorithm**	**71.61%**	$\epsilon_d = 1.0\mathrm{m}, \epsilon_v = 0.5\mathrm{m/s}, \epsilon_x = 1\mathrm{m}, \epsilon_y = 5\mathrm{m}, \epsilon_a = 1°$

$$\mathrm{V}_1\text{-measure} = \frac{2\mathcal{H}\mathcal{C}}{\mathcal{H}+\mathcal{C}}. \tag{9}$$

The V_1-measure, the harmonic mean of Homogeneity and Completeness, is employed to evaluate the quality of clustering outcomes comprehensively. The harmonic mean is chosen because it is susceptible to any of the measures being significantly low. The advantage of the V_1-measure is that it does not necessitate a direct correspondence between clustering results and authentic classes, making it particularly suited for evaluating the clustering effects in unsupervised learning. Additionally, considering both the homogeneity and completeness of the clustering results provides a comprehensive and balanced assessment of clustering quality.

C. Experiment Analysis

To evaluate the proposed algorithm's performance, four experiments are designed, with an overview provided in Table I. The original DBSCAN algorithm is selected as the baseline-DBSCAN experiment (#1). Following the improvements discussed in Section 3, three additional experiments are sequentially designed. Firstly, an enhanced DBSCAN algorithm based on distance and velocity thresholds (#2) is introduced, aimed at augmenting the algorithm's capability to accurately differentiate between moving targets and static backgrounds under varying speed and distance conditions. Subsequently, optimization through the introduction of adaptive ellipse gate (#3) aimed at addressing the issue of large vehicle fragmentation. According to the data in the table, the experimental results obtained by the two alternative algorithms of distance measurement exhibited slight advantages over the baseline DBSCAN algorithm. Building on #3, the study further incorporated a secondary clustering algorithm based on azimuth angles, culminating in the adaptive point cloud clustering algorithm (#4). The experimental outcomes indicate that this two-stage clustering algorithm significantly surpasses the baseline algorithm, achieving a V_1-score of 71.61%.Experimental result shows that compared with the baseline-DBSCAN algorithm, the algorithm we proposed can effectively improve the accuracy of clustering, and each innovative step can also bring better clustering results.

V. Conclusion

We propose an adaptive point cloud clustering algorithm for improving the accuracy of extended target point cloud clustering in roadside traffic sensing environments. The DBSCAN algorithm improved by distance and velocity and adaptive ellipse gate strategy according to vehicle shape, significantly addresses large vehicle fragmentation and increasing clustering accuracy. Cluster merging based on azimuth further resolves large vehicle fragmentation and velocity outlier issues. Experimental and pratical roadside measurement results demonstrate our algorithm's superiority to traditional DBSCAN, and each innovative step can leads to improved clustering outcomes in roadside MmWave sensing.

VI. Acknowledgement

This work was supported by Guangdong Natural Science Fund for Distinguished Young Scholar under Grant 2023B1515020079

References

[1] R. Chen, S. Sun, Y. Liu, X. Hu, Y. Hui, and N. Cheng, "Proactive effects of C-V2X-based vehicle-infrastructure cooperation on the stability of heterogeneous traffic flow," *IEEE Internet of Things Journal*, vol. 11, no. 5, pp. 9184–9197, 2024.

[2] R. Chen, J. Ning, Y. Lei, Y. Hui, and N. Cheng, "Mixed traffic flow state detection: A connected vehicles-assisted roadside radar and video data fusion scheme," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 4, pp. 360–371, 2023.

[3] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the Second lnternational Conference on Knowledge Discovery and Data Mining*, vol. 96, no. 34, 1996, pp. 226–231.

[4] R. Zhang and S. Cao, "Robust and adaptive radar elliptical density-based spatial clustering and labeling for mmwave radar point cloud data," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2019, pp. 919–924.

[5] S. F. Qureshi, H. A. Gonzalez, C. Liu, B. Vogginger, M. Stolba, S. Scholze, S. Höppner, and C. Mayr, "Efficient DBSCAN implementation in a multi-core DSP for FMCW radars," in *2022 IEEE Radar Conference (RadarConf22)*. IEEE, 2022, pp. 1–6.

[6] X. Jin, H. Yang, X. He, G. Liu, Z. Yan, and Q. Wang, "Robust lidar-based vehicle detection for on-road autonomous driving," *Remote Sensing*, vol. 15, no. 12, p. 3160, 2023.

[7] M. El Yabroudi, K. Awedat, R. C. Chabaan, O. Abudayyeh, and I. Abdel-Qader, "Adaptive DBSCAN lidar point cloud clustering for autonomous driving applications," in *2022 IEEE International Conference on Electro Information Technology (eIT)*. IEEE, 2022, pp. 221–224.

[8] C. Li, F. Gao, X. Han, and B. Zhang, "A new density-based clustering method considering spatial distribution of lidar point cloud for object detection of autonomous driving," *Electronics*, vol. 10, no. 16, p. 2005, 2021.

[9] H. Chen, M. Liang, W. Liu, W. Wang, and P. X. Liu, "An approach to boundary detection for 3d point clouds based on dbscan clustering," *Pattern Recognition*, vol. 124, p. 108431, 2022.

[10] R. J. Campello, D. Moulavi, and J. Sander, "Density-based clustering based on hierarchical density estimates," in *Pacific-Asia conference on knowledge discovery and data mining*. Springer, 2013, pp. 160–172.

[11] C. Malzer and M. Baum, "Constraint-based hierarchical cluster selection in automotive radar data," *Sensors*, vol. 21, no. 10, p. 3410, 2021.

[12] C. Lanczos and J. Boyd, *Discourse on Fourier series*. SIAM, 2016.

[13] N. Scheiner, N. Appenrodt, J. Dickmann, and B. Sick, "A multi-stage clustering framework for automotive radar data," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2019, pp. 2060–2067.