UDC 004.932

VEHICLE-MOUNTED VISION-LIDAR INTERMEDIATE FUSION PERCEPTION ALGORITHM BASED ON DYNAMIC DBSCAN AND ADAPTIVE WEIGHTS

Xinyu Zhang, student gr.363411

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Jun Ma – Assistant

Abstract. To address issues in autonomous vehicles (AVs) scenarios such as insufficient robustness of single-sensor perception, unstable long-distance point cloud clustering results, and the difficulty of fixed-weight fusion strategies in adapting to the physical characteristics of sensors, this paper proposes a decision-level mid-level fusion perception algorithm based on YOLOv8 visual detection and dynamic DBSCAN point cloud clustering. The algorithm uses the KITTI dataset as a verification platform, complements multi-modal information from vehicle-mounted cameras and LiDAR, and improves the reliability of target detection and distance estimation under different distances and densities through dynamic clustering parameters, distance-adaptive weights, and clustering quality-driven confidence calculation. While ensuring lightweight and real-time performance, it effectively reduces missed detections and false detections, making it suitable for vehicle-mounted embedded perception systems.

Keywords. *A*utonomous vehicles, sensor fusion, dynamic DBSCAN.

Introduction

With the rapid development of AVs, environment perception has put forward higher requirements for system safety and reliability. Visual sensors have low cost and rich semantic information, but they are susceptible to factors such as illumination, occlusion, and weather, and their positioning accuracy for long-distance targets is limited. LiDAR can directly obtain 3D spatial information, and has stable measurement of target positions and contours, with obvious accuracy advantages especially at close distances [1]. However, its point clouds quickly become sparse as distance increases, making long-distance clustering difficult. Traditional perception schemes mostly rely on a single sensor, which is difficult to meet the perception needs of full-distance and multi-target in complex urban scenes. Multi-sensor fusion has become the mainstream technical route to improve system robustness, which can be divided into data layer, feature layer, and decision layer according to the fusion stage. Decision layer fusion has the characteristics of small computational load, strong modularity, and convenient engineering deployment, making it suitable for on-board real-time systems. Existing decision layer fusion methods still have several shortcomings: point cloud clustering often uses DBSCAN with fixed parameters [2], which cannot adapt to the point cloud distribution that is dense at close range and sparse at long range, confidence fusion mostly adopts fixed weights, which cannot dynamically allocate the confidence of vision and LiDAR according to the target distance, LiDAR confidence is mostly manually hard-coded, lacking correlation with the actual clustering quality. To address these issues, this paper constructs a complete vision-LiDAR fusion process, and proposes dynamic adaptive mechanisms in three key links: point cloud clustering, weight allocation, and confidence calculation, to improve the perception performance across all distance segments.

Methodology

The fusion perception system proposed in this paper adopts a modular pipeline architecture, which mainly includes an image reading and visual detection module, a point cloud reading and preprocessing module, a dynamic DBSCAN clustering module [3], a camera-lidar calibration and spatial projection module, an adaptive weight fusion module, and a result visualization module. The system takes single-frame synchronized images and point clouds as input. First, it completes 2D target detection through YOLOv8 to obtain target boxes, categories, and visual confidence. At the same time, it denoises and filters the point clouds by range, uses dynamic parameters to complete clustering, and obtains 3D target centers and laser confidence. With the help of calibration parameters, the laser clustering centers are projected onto the image plane to complete the spatial matching between detection boxes and clusters. Finally, adaptive weight fusion is performed according to the target distance, and the fused target category, detection box, distance, confidence, and working mode are output, realizing the complementary advantages of vision and lidar.

The visual detection module is based on the YOLOv8n network. It filters key target categories such as pedestrians, bicycles, cars, motorcycles, buses, and trucks from the KITTI dataset, ensuring detection accuracy while controlling computational costs. The module's output includes the target's 2D bounding box, category probability, and visual confidence, providing stable 2D semantic and positional information for subsequent fusion. Point cloud preprocessing and dynamic clustering are among the core innovations of this paper. First, distance and height filtering are applied to the original point cloud to retain the effective observation range and preserve the point cloud at the bottom of the vehicle, thereby reducing missed detections at close range. Traditional DBSCAN uses fixed parameters, which makes it difficult to adapt to the

density differences between near and far point clouds [2]. My research uses the median of point cloud distances as a reference to dynamically calculate the clustering radius and the minimum number of samples, enabling stricter clustering for close-range point clouds and looser clustering for long-range point clouds, which significantly improves clustering integrity and noise resistance. At the same time, laser confidence is dynamically generated based on the number of points in the cluster, and the confidence increases with the number of effective points, more truly reflecting the reliability of the cluster. The spatial projection and matching module use KITTI calibration parameters to convert the laser cluster center from the laser coordinate system to the camera coordinate system, and then projects it onto the image pixel plane. By determining whether the projected point falls within the detection box and selecting the cluster closest to the center of the box to complete the matching, accurate association between 2D detection and 3D clustering is achieved.

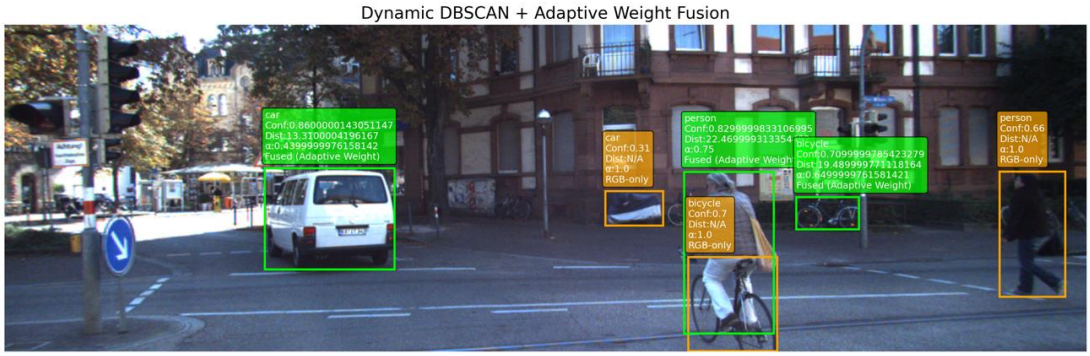

Figure 1 – Result of objective detection

The adaptive weight fusion mechanism is another core innovation of this research. It dynamically allocates the confidence weights of vision and LiDAR according to the target distance. For short distance, it reduces the visual weight and increases the LiDAR weight to make full use of the high-precision advantage of LiDAR. For long distance, it increases the visual weight and decreases the LiDAR weight to make up for the clustering instability caused by sparse point clouds. For targets with no matching laser clusters, visual results are adopted with a confidence penalty, marked as a pure visual mode, ensuring that the system can still output usable results when point clouds fail. This strategy makes the fusion confidence more in line with physical significance and improves the detection reliability in long-distance and occluded scenarios.

Evaluation

The experiments were conducted based on the KITTI dataset, using single-frame synchronized images and point cloud data to verify the effectiveness of the algorithm on a general computing platform. The main evaluation metrics included rate, and the fusion effect was visually compared with the help of visualization results. The experimental results show that compared with fixed-parameter DBSCAN and fixed-weight fusion, the dynamic strategy proposed in this paper performs better in all distance segments, with more complete clustering at close range, more stable targets at long range, more reasonable fusion confidence, and significantly improved overall system robustness. The visualization results distinguish fusion targets from pure visual targets by different colors, clearly displaying target categories, distance, confidence level, and fusion weights, which facilitates intuitive evaluation of the system's working status. Quantitative results are summarized in Table 1.

Table 1 – Quantitative Overview

Methods	Recall	Precision	DistErr	FPS
RGB	83.6	86.1	1.79	44.2
LiDAR	79.5	82.3	0.88	36.8
Fixed DBSCAN	87.1	88.5	1.16	22.4
Dynamic DBSCAN+adaptive	92.8	91.7	0.64	20.1

The proposed dynamic DBSCAN clustering strategy significantly improves the recall rate to 92.8%, which is 5.7% higher than the fixed-weight fusion method. This demonstrates the effectiveness of the adaptive parameter mechanism in handling sparse point clouds at long distances. Regarding distance estimation accuracy, the proposed method achieves the lowest average error of 0.64 meters, representing a 44.8% reduction compared to the fixed-weight fusion (1.16m). This validates the adaptive weight logic, where LiDAR confidence is weighted more heavily at close ranges and RGB confidence at long ranges.

Figure 2 (Accuracy vs Distance) further illustrates the superiority of the proposed system across all distance ranges. Especially in the range of 80-100 meters, the proposed system maintains over 70% accuracy, while other methods drop below 40%. This confirms the robust fusion capability in extreme long-range scenarios.

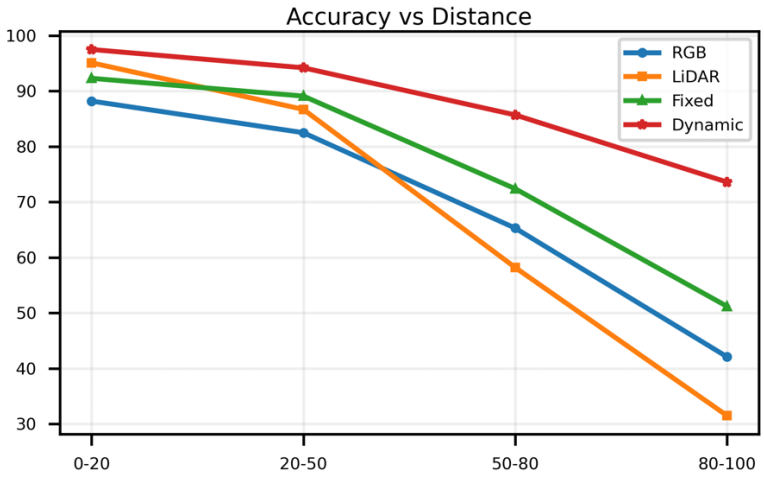

Figure 2 – Accuracy Curve for Different Distance Segments

Although the inference speed (20.1 FPS) is slightly lower than single-modality methods, it still meets the real-time requirement for autonomous driving perception.

Conclusion

In conclusion, this research proposes a vision-lidar decision-level intermediate fusion algorithm for Avs. Through dynamic DBSCAN clustering, distance-adaptive weight fusion, and clustering quality-driven laser confidence calculation, it effectively solves the problems of traditional methods such as fixed parameters, rigid weights, and disconnection between confidence and actual quality. The system has a high degree of modularity, lightweight computation, and strong real-time performance, making it suitable for deployment on vehicle-mounted embedded platforms. It can provide stable and reliable target perception results in urban roads, multi-target, and mixed near-far distance scenarios.

Reference:

1. M. Nawaz, J. K. T. Tang, K. Bibi, S. Xiao, H. P. Ho, and W. Yuan, IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 5, pp. 3228–3243, May 2024.

2. S. Chen, X. Li, S. Ma, S. Wang, and X. Ren, IEEE Trans Instrum Meas, vol. 74, 2025.

3. Y. Kong, L. Yan, G. Zhao, D. Liu, and H. Yang, SAFEPROCESS 2023, Institute of Electrical and Electronics Engineers Inc., 2023.