Transactions of KSAE, Vol. 34, No. 11, pp.1305-1317 (November, 2026)

Check for updates

Copyright © 2026 KSAE / 252-08
pISSN 1225-6382 / eISSN 2234-0149
DOI http://dx.doi.org/10.7467/KSAE.2026.34.11.1305

Map-View Semantic Segmentation을 통한 레이더 객체 검출 및 추적 성능 향상에 관한 연구

김준녕[1] · 김효민[1] · 유한결[1] · 김준한[1] · 강연식[*2]

국민대학교 자동차 모빌리티 대학원1) · 국민대학교 자동차 공학과2)

A Study on the Improvement of Radar Object Detection and Tracking Performance through Map-View Semantic Segmentation

Junyeong Kim[1] · Hyomin Kim[1] · Hangyeol Yoo[1] · Junhan Kim[1] · Yeonsik Kang[*2]

[1]Department Graduate School of Automobile and Mobility, Kookmin University, Seoul, 02707 Korea
[2]Department of Automotive Engineering, Kookmin University, Seoul, 02707, Korea
(Received 28 April 2026 / Revised 15 June 2026 / Accepted 18 June 2026)

Abstract : This paper presents a camera-radar fusion framework that improves object detection and tracking performance in autonomous driving by reducing radar noise and ghost objects. A cross-view transformer generates map-view semantic segmentation, which the proposed method uses to locate reliable vehicle regions in bird's-eye-view space. After being mapped into this space, radar points are filtered by retaining only those that fall within the predicted regions. DBSCAN groups the selected points, while a Kalman filter estimates object states through nearest neighbor data association and M/N logic-based track management. Results from the nuScenes dataset demonstrate that the proposed method raises detection precision from 67.9 % to 90.1 % in general scenarios and from 45.7 % to 82.7 % in urban environments, while reducing localization error from 1.99 m to 1.33 m.

Key words : Autonomous driving(자율주행), Sensor fusion(센서 융합), Radar denoising(레이더 노이즈 제거), Kalman filter(칼만 필터), Object tracking(객체 추적)

1. Introduction

자율주행 차량의 안정적이고 정확한 주행을 위해 카메라(Camera), 라이다(LiDAR), 레이더(Radar)와 같이 고유의 특성과 장점을 지니고 있는 다양한 센서를 활용한다. 카메라 센서는 물체의 색상, 모양, 텍스처 등 시각적 정보를 세밀하게 포착할 수 있어 신호등, 도로 표지판, 차선, 보행자 인식에 유리하다. 그러나 빛의 영향을 크게 받아 야간이나 악천후 환경에서 인식 성능이 저하되며, 2D 이미지에서 깊이 정보를 직접적으로 얻기 어렵다는 한계가 있다. 이러한 깊이 및 거리 정보의 한계를 보완하기 위해 라이다 센서는 레이저 빔을 활용하여 3차원 포인트 클라우드 형태로 정밀한 거리 및 위치 정보를 제공하지만, 고가의 장비 비용과 악천후 환경에서의 성능 저하 문제가 존재한다. 반면 레이더 센서는 전파를 이용해 주변 물체와의 거리와 상대 속도를 감지하며, 날씨나 조명 조건에 강인하여 악천후 환경에서도 안정적으로 동작한다. 다만 레이더는 반사체의 특성에 따라 실제로 존재하지 않는 고스트(Ghost) 객체를 감지하거나 객체 외곽에 노이즈가 발생하는 문제가 있어, 이를 완화하는 것이 레이더를 신뢰성 있게 활용하기 위한 핵심 과제로 남아 있다.

이러한 단일 센서의 한계를 극복하기 위해 여러 센서를 융합하는 센서 융합(Sensor fusion) 연구가 활발히 진행되고 있다.[1] 그 중에서도 카메라는 풍부한 시각적 정보를 제공하고 레이더는 날씨와 조명에 강인한 거리 및 속도 정보를 제공한다는 점에서 두 센서는 서로의 단점을 보완할 수 있어, 카메라와 레이더를 결합한 연구가 특히 주목받고 있다.[2]

*Corresponding author, E-mail: ykang@kookmin.ac.kr
This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(http://creativecommons.org/licenses/by-nc/3.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium provided the original work is properly cited.

카메라-레이더 센서 융합 방식은 크게 데이터 레벨 융합, 특징 레벨 융합, 결과 레벨 융합으로 분류된다. 데이터 레벨 융합은 원시 데이터를 직접 결합하여 정밀한 분석이 가능하지만 센서 간 캘리브레이션에 민감하다. 특징 레벨 융합은 각 센서에서 추출된 특징을 결합하여 고정밀 분석이 가능하나 정보 손실이 발생한다. 반면 결과 레벨 융합은 각 센서의 출력을 독립적으로 처리한 뒤 결합하므로, 구현이 간단하고 기존 검출·추적 파이프라인에 적용이 용이하다는 장점이 있다. 이러한 융합 방식을 활용하기 위해서는 각 센서 데이터를 공통된 공간에서 표현하고 해석하는 것이 중요하며, 최근에는 이를 위한 방법으로 Map-view Semantic Segmentation 연구가 활발히 이루어지고 있다.[3,4] Map-view는 센서로부터 얻은 데이터를 조감도(Bird's Eye View, BEV) 형식으로 변환하여 객체의 위치와 형태를 직관적으로 파악할 수 있는 방식으로, 자율주행 시스템이 주변 상황을 명확하게 파악하는 데 효과적이다.

따라서 본 논문에서는 결과 융합 방식을 활용한 기반 모델로 Cross-view Transformer(CVT)를 사용하여 Map-view Semantic Segmentation과 레이더 데이터를 융합하여 레이더의 고스트 객체 및 노이즈 문제를 해결해 동적 객체 검출 및 추적(Object detection and tracking) 성능 향상 기법을 제안한다.

레이더 포인트를 Map-view 좌표로 변환한 후 차량 예측 영역 내의 신뢰성 있는 포인트만을 선별하고, DBSCAN 클러스터링 알고리즘으로 객체를 군집화한다. 이후 칼만필터(Kalman filter) 기반 다중 객체 추적 알고리즘과 Nearest Neighbor 기반 데이터 연관 기법 및 M/N Logic의 Vehicle 객체의 Tracking 관리기를 적용하여 안정적인 객체 추적을 구현한다. nuScenes 데이터셋을 활용한 실험을 통해 제안 기법의 성능을 검증하였으며, 본 논문의 주요 기여는 다음과 같다.

첫째, BEV 기반 Semantic segmentation 결과를 활용하여 레이더 측정값을 선별하는 카메라–레이더 융합 프레임워크를 제안한다. 이를 통해 기존 연구에서 충분히 고려되지 않았던 레이더 Ghost object 및 노이즈 문제를 완화한다.

둘째, 레이더 포인트를 BEV 공간으로 투영한 후 차량 영역에 해당하는 포인트만 선택하는 공간 기반 필터링 기법을 제시함으로써, 측정값의 신뢰도를 향상시킨다.

셋째, DBSCAN 기반 클러스터링과 Kalman filter 기반 추적, 그리고 M/N 로직을 결합한 다중 객체 추적 구조를 구성하여, 희소하고 노이즈가 포함된 레이더 환경에서도 안정적인 추적 성능을 확보한다. 또한 개선된 성능을 공개된 nuScenes 데이터 셋을 통해 검증한다.

본 연구는 기존 연구들이 주로 3D 객체 검출 성능 향상 또는 Feature-level fusion에 초점을 둔 것과 달리, 레이더 고스트 포인트가 클러스터링 및 추적 단계에서 오검출과 잘못된 트랙으로 이어지는 현상에 초점을 두고 해당 필터링이 레이더 기반 검출·추적 파이프라인에서 어떤 성능 개선을 가져오는지 분석하였다. 기존 Camera-radar fusion이 학습 기반 Feature-level 융합을 검출기 내부에서 수행하는 것과 달리, 본 연구는 Detector 재학습 없이 BEV Semantic mask로 개별 Radar 측정값을 선별하는 모듈형 Measurement-level 필터링 단계를 검출·추적 파이프라인 앞단에 도입하고 그 효과를 정량적으로 분석한다는 점에서 차별된다.

2. Related Work

자율주행 환경에서 객체 인식 및 추적 성능 향상을 위해 카메라–레이더 센서 융합, 레이더 노이즈 및 고스트 객체 제거, 다중 객체 추적에 관한 연구가 활발히 진행되고 있다. 본 장에서는 각 분야의 대표 연구를 정리하고, 이를 바탕으로 본 연구의 차별성을 기술한다.

2.1 카메라–레이더 센서 융합 연구

카메라와 레이더는 각각 시각적 의미 정보와 거리, 속도 정보에 강점을 가지므로, 두 센서의 상호보완성을 활용한 융합 연구가 자율주행 인지 분야에서 활발히 이루어지고 있다. 특히 최근의 Camera-radar fusion 연구는 두 센서 정보를 공통된 Bird's-Eye-View(BEV) 공간에서 정렬하고 융합함으로써 3D 객체 검출 성능과 환경 강건성을 향상시키는 방향으로 발전해왔다. 또한 CenterFusion과 ClusterFusion은[5,6] 카메라가 검출한 객체를 기준으로 Frustum 또는 Cluster 단위로 Radar point를 연관시켜 검출기 내부의 Feature를 보강하는 학습 기반 Fusion인 반면, 본 연구는 Detector 재학습 없이 BEV semantic mask로 개별 Radar 측정값의 유효성을 판정하여 Ghost/Noise point를 Clustering 및 Tracking 이전 단계에서 제거하는 모듈형 Measurement-level filtering이라는 점에서 설계 철학이 구분된다.

먼저, Lin 등[2]은 RCBEVDet를 통해 카메라와 레이더를 BEV 공간에서 정렬 및 융합하여 3D 객체 검출 성능을 향상시켰다. 특히 Attention 기반 구조를 통해 두 센서의 공간적 불일치를 완화하였다.

Kim 등[7]은 CRN을 제안하여 카메라와 레이더 데이터를 BEV 공간에서 융합하고, Multi-modal deformable attention을 통해 멀티모달 특징을 통합하였다.

다양한 환경 조건에서 강건한 3D 객체 검출을 위한 레이더–카메라 융합 프레임워크를 제안하였으며, 악천후

및 저조도 환경에서 카메라 기반 인식의 한계를 레이더 정보로 보완하였다.

이들 연구는 카메라–레이더 융합이 객체 검출 성능과 강건성을 향상시킬 수 있음을 보여주었다. 그러나 대부분의 연구는 검출 성능 향상 자체에 초점을 두고 있어, 레이더의 다중 경로 반사로 인해 발생하는 고스트 객체 및 노이즈를 직접적으로 제거하는 접근은 상대적으로 제한적이다.[8] 본 연구는 기존 BEV 기반 Camera-radar fusion 연구와 달리, 새로운 3D detector 구조를 설계하는 것보다 BEV semantic 정보를 Radar measurement filtering 기준으로 활용하여 Ghost point가 후속 Clustering 및 Tracking 단계로 전달되는 것을 줄이는 데 초점을 둔다.

2.2 레이더 노이즈 및 고스트 객체 문제

레이더 센서는 다중 경로 반사와 희소한 포인트 분포 특성으로 인해 실제로 존재하지 않는 고스트 객체나 객체 외곽 노이즈가 발생할 수 있으며, 이러한 문제는 복잡한 도심 환경에서 더욱 두드러진다.[9,10]

예를 들어, Zhou 등[11]은 4D 이미징 레이더의 가능성을 제시하면서도, 단일 레이더만으로는 정밀한 객체 인식에 한계가 있음을 지적하고 카메라와의 융합 필요성을 강조하였다.

또한 Ouaknine 등[12]은 CARRADA 데이터셋 기반의 레이더 시맨틱 분할 기법을 통해 노이즈가 포함된 레이더 포인트 클라우드에서도 객체를 인식할 수 있는 가능성을 제시하였다.

이와 같은 연구들은 레이더 노이즈 및 고스트 문제의 중요성을 보여주고 있으나, 카메라 기반의 Semantic BEV 정보를 활용하여 레이더 포인트를 직접 선별하고 고스트 객체를 제거하는 방법은 충분히 다루어지지 않았다. 특히 기존 연구들은 레이더 자체의 표현 학습 또는 Sensor Feature fusion에 초점을 두는 경우가 많으며, 카메라 기반 BEV semantic mask를 이용해 개별 Radar point의 유효성을 판정하고, 이를 후속 Clustering 및 Tracking 입력으로 사용하는 구조는 충분히 다루어지지 않았다.

2.3 다중 객체 추적 연구

다중 객체 추적은 자율주행 시스템에서 동적 객체의 위치와 궤적을 연속적으로 추정하는 핵심 기술이다. 특히 레이더 기반 환경에서는 포인트 밀도가 낮고 노이즈가 빈번하므로, 상태 예측과 데이터 연관, 트랙 관리가 함께 고려되어야 한다.

관련 연구로, Weng 등[13]은 AB3DMOT를 통해 3D Kalman filter 기반의 실시간 다중 객체 추적 기법을 제안하였으며, 비교적 단순한 구조로도 높은 추적 성능을 달성하였다.

또한, Baek 등[14]은 레이더–카메라 융합 필터링 기반의 검출 및 추적 시스템을 제안하여, 도심 자율주행 환경에서 강건한 추적 성능을 확보하였다.

이들 연구의 경우 추적 단계에서 Kalman filter 기반 예측과 데이터 연관의 유효성을 보여주지만, 입력 측정값 자체에 레이더 고스트 및 노이즈가 포함된 경우 추적 성능이 저하될 수 있다. 따라서 측정값의 신뢰도 향상과 추적 알고리즘의 결합이 함께 고려될 필요가 있다.

종합하면, 기존 연구들은 카메라–레이더 융합, 레이더 노이즈 문제, 다중 객체 추적의 각 측면에서 의미 있는 성과를 보였으나, BEV semantic segmentation을 이용하여 레이더 포인트를 선별적으로 필터링하고, 이를 추적 단계와 통합하여 활용하는 결과 수준 융합 방식의 연구는 상대적으로 부족하다.

이에 본 연구에서는 Cross-view Transformer 기반 Map-view Semantic Segmentation 결과를 활용하여 레이더 포인트를 선별하고, 이를 Kalman filter 기반 다중 객체 추적과 결합함으로써 고스트 객체 제거와 추적 안정성 향상을 동시에 달성하고자 한다. 특히 추적 알고리즘 자체의 복잡도를 높이는 방향보다, BEV semantic segmentation 기반 Filtering을 통해 추적 입력 측정값의 False positive를 줄이고, 이후 Kalman filter 기반 Tracking 구조와 결합하는 방향을 선택하였다.

3. Method

3.1 Model Architecture

본 논문에서 제안하는 시스템은 카메라 기반 Map-view Semantic Segmentation과 레이더 데이터를 융합하여 고스트 객체 및 노이즈를 제거하고 안정적인 객체 인식 및 추적을 수행하는 파이프라인으로 구성된다. 전체 시스템은 크게 Map-view Semantic Segmentation, 레이더 데이터 전처리 및 필터링, 레이더 클러스터링, 다중 객체 추적의 네 가지 모듈로 구성된다. Fig. 1에서 다중 카메라 이미지를

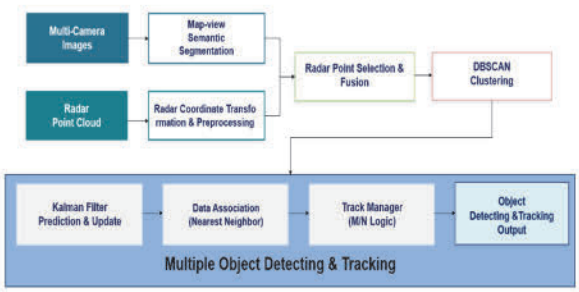


Fig. 1 Overall system architecture for camera-radar sensor fusion-based object detection and tracking

입력으로 받아 기반 모델인 Cross-view Transformer 기반 Map-view Semantic Segmentation을 수행하여 차량 예측 영역을 생성한다. 이후 레이더 데이터를 Map-view 좌표계로 변환하고, 차량 예측 영역 내에 위치한 신뢰성 있는 레이더 포인트만을 선별하여 고스트 객체 및 노이즈를 제거한다.

선별된 레이더 포인트는 DBSCAN 클러스터링 알고리즘을[15] 통해 객체 단위로 군집화되며, 각 클러스터의 중심 좌표를 객체의 위치 정보로 활용한다. 마지막으로 칼만필터 기반 다중 객체 추적 알고리즘을 적용하여[16] 객체의 위치를 지속적으로 예측하고 갱신하며, Data Association nearest neighbor 알고리즘과 M/N Logic 기반 트랙 관리기를 통해 안정적인 객체 추적을 수행한다. 각 모듈은 레이더 기반 검출 및 추적의 단계별 실패 요인에 순차적으로 대응하도록 결합된다. Map-view Semantic Segmentation 기반 필터링은 측정값 수준에서 발생하는 고스트 및 노이즈를 제거하고, DBSCAN은 선별된 희소 레이더 포인트를 객체 단위 측정값으로 변환하여 군집 오류를 줄이며, Kalman filter 기반 추적과 M/N Logic은 측정 불확실성과 일시적 오검출이 잘못된 Track으로 이어지는 것을 억제한다. 이러한 네 모듈은 순차적으로 연결되어 단일 처리 흐름을 구성한다. 다중 카메라 영상은 Cross-view Transformer를 거쳐 BEV 공간의 차량 시맨틱 마스크로 변환되고, 이와 병렬적으로 5개의 레이더에서 수집된 포인트는 각 센서의 캘리브레이션 파라미터를 통해 자차 중심의 통합 좌표계로 정렬된 뒤 동일한 BEV 좌표계로 투영된다. 두 모달리티는 검출기 내부가 아니라 측정값 단계에서 결합되며, 시맨틱 마스크는 각 레이더 포인트의 유효성을 판정하는 공간적 기준으로만 사용된다. 이 과정에서 차량 예측 영역에 포함되지 않는 레이더 포인트는 차량 Class를 기준으로 한 선별 과정에서 제거되어 다중 경로 반사로 인한 고스트 및 노이즈를 완화하며, 선별된 포인트만이 DBSCAN의 입력으로 전달된다. 클러스터링으로 산출된 각 클러스터의 중심 좌표는 객체 단위 측정값으로 변환되어 칼만 필터의 갱신 단계에 입력되며, Nearest Neighbor 데이터 연관과 M/N Logic 트랙 관리기가 프레임 간 측정값을 연결하여 연속적인 객체 궤적을 생성한다. 또한 결과 수준 융합 방식을 채택함으로써 카메라와 레이더가 서로 다른 시점과 표현 공간을 가지더라도 BEV 좌표계라는 공통 기준에서 정합되며, 한 센서의 일시적 성능 저하가 전체 인식 결과에 직접 전파되는 것을 완화한다. 이와 같이 각 모듈은 명확히 구분된 입력과 출력을 가지므로 개별 모듈을 독립적으로 교체하거나 개선할 수 있으며, Segmentation 모델이나 추적 알고리즘을 변경하더라도 전체 파이프라인 구조를 유지할 수 있어 확장성과 유지보수성이 높다.

Map-view Semantic Segmentation은 여러 카메라 뷰에서 수집한 데이터를 조감도(Bird's Eye View, BEV) 형식의 시맨틱 맵으로 변환하는 기법으로, 자율주행 차량이 주행 경로와 주변 객체의 위치를 직관적으로 파악할 수 있도록 한다. 본 연구에서는 다중 카메라 뷰 간의 정보를 융합할 수 있는 Cross-view Transformer 모델을 적용하였다.[17]

Fig. 2에서 Cross-view Transformer는 Attention 메커니즘을 기반으로 각 카메라의 내부 파라미터(K)와 외부 파라미터(R,t)를 활용하여 Camera-aware Positional Embedding을 생성하고, 이를 통해 다양한 카메라 시점 간의 공간적 관계를 학습한다. EfficientNet-B4 인코더를 통해 각 카메라 이미지로부터 다중 스케일 특징을 추출하고, Cross-view Attention 레이어를 통해 이미지 특징과 위치 임베딩을 Map-view 좌표에 맞게 결합한다. 이후 디코더를 통해 고해상도의 시맨틱 분할 마스크를 생성하며, 이 과정에서 별도의 깊이 추정 없이도 객체의 위치와 형태를 Map-view 상에서 정밀하게 표현 가능하다.

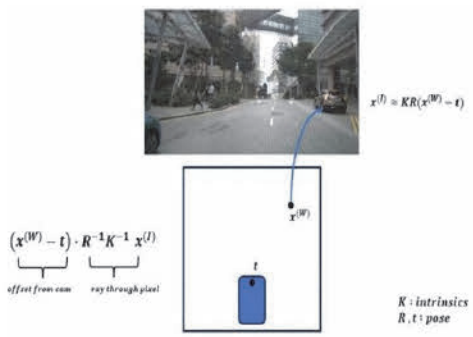


Fig. 2 Geometric relationship between image coordinates and world coordinates in Cross-view Transformer Base Model

최종적으로 생성된 차량 예측 영역은 이후 레이더 포인트 필터링 단계에서 신뢰성 있는 레이더 데이터 선별을 위한 기준으로 활용된다.

3.2 레이더 데이터 전처리 및 필터링

레이더 데이터를 Map-view Semantic Segmentation 결과와 융합하기 위해 레이더 좌표계를 Map-view 이미지 좌표계로 변환하는 전처리 과정이 필요하다.

nuScenes 데이터셋의 레이더는 전방, 전방 좌우측, 후방 좌우측에 총 5개가 장착되어 360도 전방위를 커버하며, 각 레이더 센서의 Rotation 및 Translation 캘리브레이션 파라미터를 적용하여 모든 레이더 포인트를 자차 중심 기준의 통합 좌표계로 변환한다. 각 Radar point는 nuScenes의 Sensor calibration parameter를 이용하여 자차

중심 기준의 Ego/Body frame으로 변환하였으며, 이후 Map-view Semantic Segmentation 결과와 동일한 BEV image frame의 Pixel 좌표 ((u, v))로 변환하여 Filtering을 수행하였다. 본 연구에서 Ego/Body frame은 자차 중심을 원점으로 하는 좌표계이며, BEV image frame은 이미지 좌측 상단을 원점으로 하고 오른쪽 방향을 (u)축, 아래쪽 방향을 (v)축으로 정의하였다. 본 연구에서 레이더 포인트와 BEV Semantic Segmentation 결과는 동일한 자차 중심 좌표계에서 표현되므로, Filtering 단계에서는 전역 자차 자세 보상이 필요하지 않으며 식 (1), (2)는 이 조건에서 적용된다.

Fig. 3에서 확인하듯이 변환된 레이더 좌표 (x, y)는 식 (1), (2)를 통해 Map-view 이미지 좌표 (u, v)로 변환된다. 자차 중심을 기준으로 하는 레이더 좌표계와 이미지 좌측 상단을 원점으로 하는 Map-view 이미지 좌표계 간의 구조적 차이를 보정하기 위해 좌표 변환 수식을 적용하였으며, W와 H는 각각 이미지의 폭과 높이를 나타낸다. 레이더 고스트 및 노이즈 포인트는 다중 경로 반사나 주변 정적 반사체로 인해 실제 차량이 존재하지 않는 위치에 발생한다. 반면 카메라 기반 Map-view Semantic Segmentation은 다중 경로 반사의 영향을 받지 않으므로, 예측된 차량 영역은 실제 차량이 존재할 가능성이 높은 영역을 나타낸다. 따라서 각 레이더 포인트가 차량 예측 영역 내부에 위치하는지를 기준으로 유효성을 판정하면, 카메라가 차량을 관측하지 못한 영역에 찍힌 고스트·노이즈 포인트가 선별 과정에서 제거된다.

$$u = \frac{(x+50)W}{100} \tag{1}$$

$$v = (50 - y)\frac{H}{100} \tag{2}$$

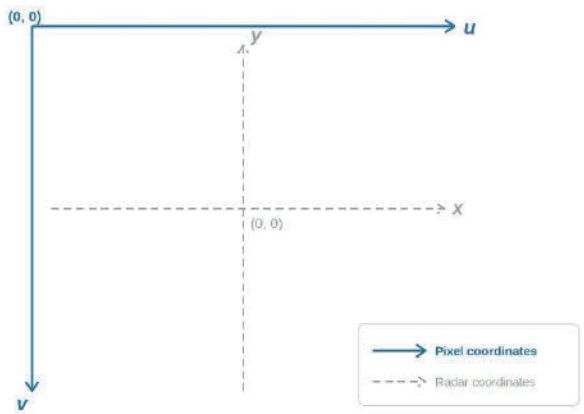


Fig. 3 Comparison of map-view pixel coordinate system (solid, origin at top-left) and radar coordinate system (dashed, origin at vehicle center), illustrating the structural difference requiring coordinate transformation

좌표 변환 후 Fig. 4에서 초록색 CVT 예측 차량 영역 내에 위치한 레이더 포인트만을 선별하였다. Fig. 5에 나타난 바와 같이, 실제 객체가 존재하지 않는 영역에서 발생하는 레이더 고스트 객체 및 노이즈를 제안한 방법을 통해 제거하였다. 본 선별은 차량 Class의 Semantic mask를 기준으로 수행되므로, 차량 영역 외부에 존재하는 비차량 동적 객체가 일부 제거될 수 있다는 점은 본 기법의 한계로 남는다. 각 Sample의 레이더 포인트는 nuScenes에서 제공하는 누적 5 sweep을 사용하였으며, 누적 과정에서 각 Sweep은 현재 Timestamp 기준으로 Ego-motion이 보상되어 통합되나, 빠르게 이동하는 객체에서는 잔여 오차가 남을 수 있다.

이를 통해 레이더 데이터의 신뢰성을 높이고 이후 클러스터링 및 추적 단계의 정확도를 향상시킬 수 있다. 특히 영역 밖 측정값을 입력 단계에서 제거해 오검출의 후속 전파를 완화한다. 또한 Segmentation 결과만 사용하는 모듈형 전처리여서 기존 파이프라인 앞단에 그대로 결

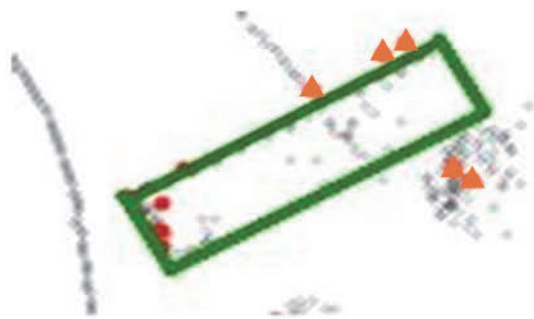


Fig. 4 Radar point selection result based on Map-view Semantic Segmentation, where radar points located inside the CVT-predicted vehicle region (green) are retained as valid measurements (red), while points outside the region (gray) are rejected as ghost or noise

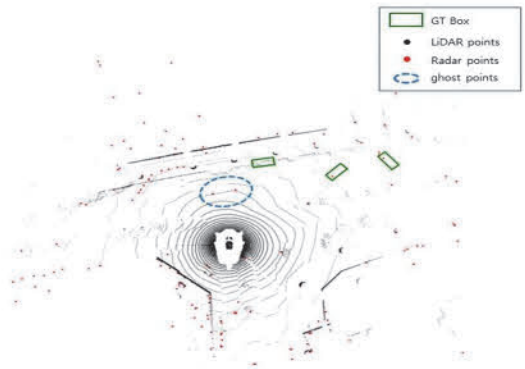


Fig. 5 Example of radar ghost points (blue dashed circle) detected outside ground truth bounding boxes (green), where radar points appear in regions with no actual objects confirmed by LiDAR points (black)

합할 수 있다. 별도의 재학습 없이 가능하다. 이는 카메라의 의미 정보가 레이더의 한계를 보완하는 결과 융합의 이점을 보여준다.

3.3 레이더 데이터 클러스터링

Fig. 6에서 노란색 포인트로 필터링된 레이더 포인트를 객체 단위로 군집화하기 위해 DBSCAN(Density-Based Spatial Clustering of Applications with Noise) 알고리즘을 적용하였다. DBSCAN은 사전에 클러스터 개수를 정의하지 않아도 되며, 일부 노이즈가 포함된 데이터 분포에서도 효과적으로 클러스터를 형성할 수 있어 레이더 데이터와 같이 클러스터 수를 예측하기 어렵고 노이즈가 빈번한 환경에 적합하다.

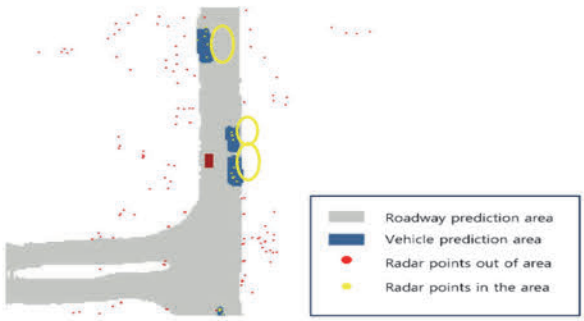


Fig. 6 Radar point clustering results on map-view semantic segmentation output, where yellow points indicate valid radar points within the vehicle prediction area (blue), red points represent noise or ghost radar points outside the prediction area, and gray region denotes the roadway prediction area

이때 차량 예측 영역 내부의 포인트만 군집 대상이 되고 영역 밖 포인트는 제외되므로, 차량 Class를 기준으로 고스트·노이즈가 완화된 클러스터가 형성된다.

DBSCAN의 군집 형성 개념은 Fig. 7과 같다. DBSCAN은 두 가지 주요 파라미터인 Eps와 minPts를 사용한다. Eps는 두 포인트가 동일한 클러스터에 속하기 위한 최대 거리이며, minPts는 클러스터로 확정되기 위한 최소 포인트 수이다. nuScenes 레이더의 경우 각도 해상도와 거리 해상도가 낮아 객체에 포인트가 적게 생성되는 특성이 있어 minPts= 3으로 설정하였다. 또한 Fig. 8과 Fig. 9에서 확인할 수 있는 하나의 객체에서 클러스터가 두 개로 분할되는 문제와 두 개의 객체가 하나의 클러스터로 병합되는 문제를 최소화하기 위해 도로 폭과 차체 크기를 고려하여 Eps=4.5 (m)로 설정하였다.

만약 Eps가 작게 설정될 경우 동일 객체에 속하는 레이더 포인트가 서로 다른 클러스터로 분리되어 Over-segmentation이 발생할 수 있으며, 반대로 Eps가 크게 설정될 경우 인접한 두 객체가 하나의 클러스터로 병합되어 Under-segmentation이 발생할 수 있다. 따라서 본 연구에서는 Eps=4.5에서 객체 분할과 병합 문제가 가장 균형 있게 억제되었으며, Precision과 평균 거리 오차 측면에서도 안정적인 결과를 보였다.

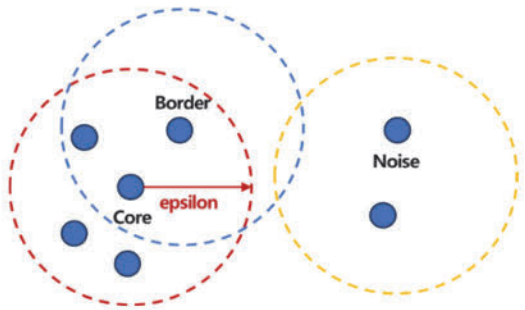


Fig. 7 DBSCAN clustering concept illustrating core points surrounded by the epsilon neighborhood (red dashed circle), where points within the epsilon radius are grouped into the same cluster, and border points (yellow dashed circle) are included in the cluster if they fall within the epsilon range of a core point

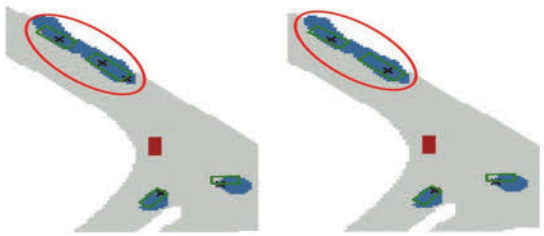


Fig. 8 Comparison of the over-segmentation problem at eps = 4.0 m and 4.5 m, where radar points belonging to a single vehicle are distinguished by color and marker; the points are split into two clusters at eps = 4.0 m and merged into one at eps = 4.5 m

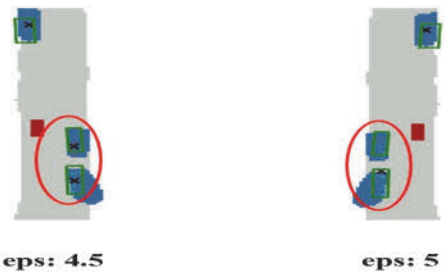


Fig. 9 Comparison of the under-segmentation problem at eps = 4.5 m and 5.0 m, where two adjacent vehicles are distinguished by color and marker; they are separated into distinct clusters at eps = 4.5 m and merged into one at eps = 5.0 m

본 DBSCAN clustering은 BEV image의 Pixel 좌표가 아니라 Ego/Body frame의 2차원 Metric 좌표를 기준으로 수행하였으며, 따라서 Eps의 단위는 Pixel이 아닌 Meter(m)이다. DBSCAN 알고리즘은 특정 포인트 p의 ϵ-이웃을 식 (3)과 같이 정의한다.

$$N_\epsilon(p) = \{ q \in D \mid dist(p,q) \le \epsilon \} \quad (1)$$

여기서 D는 전체 포인트 집합이며, $dist(p,q)$는 두 포인트 간의 유클리드 거리를 나타낸다. 포인트 p가 코어 포인트(Core point)로 분류되기 위한 조건은 식 (4)와 같다.

$$| N_\epsilon(p) | \ge minPts \quad (4)$$

즉, p의 ϵ-이웃 내에 포함된 포인트 수가 $minPts$ 이상일 때 해당 포인트는 코어 포인트로 확정되며, 코어 포인트에 직접 밀도 도달 가능한 포인트들을 동일 클러스터로 군집화 한다.

클러스터링 결과로 생성된 각 클러스터의 중심 좌표를 해당 객체의 위치 정보로 활용하여 이후 다중 객체 추적 단계에 전달한다.

Table 5는 Eps 변화에 따른 Precision과 평균 거리 오차를 비교한 결과로, Eps=4.5 m 선정의 정량적 근거를 제시한다. Table 5에서 확인할 수 있듯이 Eps가 작을수록 Over-segmentation으로 동일 객체가 여러 클러스터로 분리되어 Precision과 위치 정확도가 저하되고, Eps가 클수록 Under-segmentation으로 인접 객체가 하나의 클러스터로 병합되어 성능이 저하되며, Eps=4.5 m에서 Precision과 평균 거리 오차가 가장 균형 잡힌 결과를 보였다.

3.4 다중 객체 추적

본 연구에서는 클러스터링을 통해 추출된 객체 위치 정보를 기반으로 칼만 필터(Kalman filter)를 적용하여 다중 객체 추적(Multi-Object Tracking, MOT)을 수행하였다. 레이더 센서는 포인트 밀도가 낮고 노이즈가 빈번하게 발생하기 때문에, 안정적인 추적을 위해서는 예측(Prediction)과 갱신(Update)을 반복하는 확률적 추정 기법이 필수적이다. Kalman filter와 Nearest Neighbor association은 시간적으로 연속된 객체 상태를 추정하고 프레임 간 측정값을 연관하기 위해 사용되며, M/N Logic은 일시적 오검출이나 미검출에 의해 잘못된 Track이 생성·유지되는 것을 줄이기 위해 사용되었다. 본 연구의 추적 모델은 위치와 속도로 구성된 상태 벡터 X(k)=[x, y, v_x, v_y]^T를 사용하는 2차원 등속(Constant velocity) 동역학 모델로 구성하였으며, 시간 간격 Δt는 nuScenes의 Sample(키프레임) 간격인 0.5 s(2 Hz)로 설정하였다. 측정값은 DBSCAN 클러스터링으로 산출된 각 클러스터 중심의 2차원 위치(x, y)로 구성하였으며, 레이더의 Radial velocity는 상태 갱신에 직접 사용하지 않았고 별도의 전역 Ego-motion 보상은 적용하지 않았다. 다만 선형 칼만 필터는 등속 직선 운동 가정에 기반하므로, 곡선 주행이나 교차로 회전과 같이 비선형 운동이 두드러지는 상황에서는 추적 성능에 한계가 존재할 수 있다. 상태 벡터는 식 (5)와 같이 정의된다.

$$X(k) = \left[x(k)\ y(k)\ v_{x(k)} v_{y(k)} \right]^T \quad (5)$$

상태 방정식(State equation)과 측정 방정식(Measurement equation)에서 $A(k)$는 상태전이 행렬, $H(k)$는 측정 행렬이며, $w(k)$, $v(k)$는 각각 프로세스 노이즈와 측정 노이즈를 의미한다. 본 연구에서는 일반적인 선형 칼만 필터 구조를 사용하였으며, 전체 흐름은 Table 1과 같다.

Table 1 Kalman filter-based multi-object tracking flow

Kalman Filter-based Multi-Object Tracking (MOT)
Input X(k|k), P(k|k), z(+1), A(k), B(k), H(k), Q(k), R(k), u(k)
Output X(k + 1|k + 1), P(k + 1|k + 1)
[State Equation]
X(k + 1) = A(k)X(k) + B(k)u(k) + w(k)
[Measurement Equation]
z(k + 1) = H(k + 1)X(k + 1) + v(k + 1)
Step 1: Prediction (based on State Equation)
X(k + 1|k) = A(k) X^(k|k) + B(k) u(k)
P(k + 1|k) = A(k) P(k|k) A(k)^T + Q(k)
Step 2: Innovation (based on Measurement Equation)
ẑ(k + 1) = H(k + 1) X(k + 1|k)
ṽ(k + 1) = z(k + 1) − ẑ(k + 1)
Step 3: Innovation Covariance
S(k + 1) = H(k + 1) P(k + 1|k) H(k + 1)^T + R(k + 1)
Step 4: Kalman Gain
K(k + 1) = P(k + 1|k) H(k + 1)^T S(k + 1)^{−1}
Step 5: Update
X(k + 1|k + 1) = X(k + 1|k) + K(k + 1) ṽ(k + 1)
$P(k+1\|k+1) =$ P(k + 1|k) − K(k + 1) H(k + 1) P(k + 1|k)

예측 단계에서는 이전 상태 추정값을 기반으로 다음 시점의 상태를 예측한다. 이후 레이더로부터 새로운 측정값 $z(k+1)$이 입력되면, 예측값과의 차이를 측정 잔차(Innovation)로 정의한다. 이를 바탕으로 칼만 이득(Kalman gain)을 계산하고, 새로운 측정값을 반영하여 상

Table 2 Tracking performance evaluation

Method	Tracking	True positive	False positive	FN	IDSW	Precision	Recall	F1-score	MOTA
Map-view +Radar	Without tracking	81	23	25	12	0.778	0.764	0.771	0.434
	Tracking	91	19	15	4	0.827	0.858	0.842	0.642

Table 3 Comparison of true positive, false positive, and precision (normal scenario)

	True positive	False positive	FN	IDSW	Precision	Recall	F1-score	TP/(TP+FP+FN)	MOTA
Only radar	93	44	35	6	0.679	0.727	0.702	0.541	0.336
Ours	118	13	10	2	0.901	0.922	0.911	0.837	0.805

Table 4 Comparison of true positive, false positive, and precision (city center driving scenario)

	True positive	False positive	FN	IDSW	Precision	Recall	F1-score	TP/(TP+FP+FN)	MOTA
Only radar	75	89	31	12	0.457	0.708	0.556	0.385	−0.245
Ours	91	19	15	4	0.827	0.858	0.842	0.728	0.642

Table 5 Precision and average distance error according to DBSCAN eps variation

Eps (m)	Precision	Mean distance error
3.5	0.81	1.78
4.0	0.87	1.49
4.5	0.901	1.33
5.0	0.84	1.61

태 추정값과 공분산을 갱신한다. 갱신된 상태 추정값과 공분산은 다음 시점의 예측 단계에 다시 사용되며, 이러한 과정을 반복함으로써 레이더의 노이즈 및 일시적 미검출(Missed detection) 상황에서도 안정적인 객체 추적을 수행한다. 특히 측정값이 일시적으로 누락된 구간에서는 예측값만으로 상태를 유지하여 트랙의 연속성을 확보하고, 측정값이 다시 입력되면 갱신을 통해 누적된 추정 오차를 보정한다.

데이터 연관 기법으로는 Nearest Neighbor(NN) 알고리즘을 적용하였다. NN 알고리즘에서 예측된 트랙 위치 $\hat{z}_j$와 측정값 z_i사이의 유클리드 거리는 식 (6)과 같이 정의된다.

$$d(z_i, \hat{z}_j) = \| z_i - \hat{z}_j \|_2 \tag{6}$$

측정값이 트랙에 연관되기 위해서는 식 (7)의 게이트 조건을 만족해야 한다.

$$d(z_i, \hat{z}_j) \leq \gamma \tag{7}$$

여기서 γ는 게이트 임계값으로, 게이트 내에 존재하는 측정값 집합 Z_{gate}중 식 (8)을 만족하는 측정값 z^*를 최종적으로 해당 트랙에 연관시킨다.

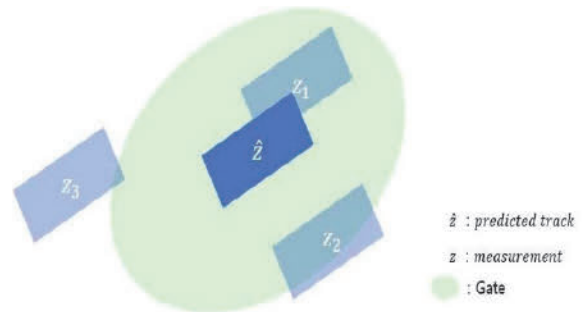


Fig. 10 Illustration of Nearest Neighbor data association, where the gate (green ellipse) defines the valid search region around the predicted track ($\hat{z}$), and the nearest measurement within the gate (z_1 or z_2) is selected for track update, while measurements outside the gate (z_3) are rejected

$$z^* = \arg \min_{z_i \in Z_{gate}} d\left(z_i, \hat{z}_j\right) \tag{8}$$

NN 알고리즘은 칼만필터를 통해 예측된 주변 차량 객체 위치를 기준으로 Fig. 10에서 확인하듯이 게이트(Gate)를 설정하고, 게이트 내에 위치한 측정값 중 유클리드 거리가 가장 가까운 측정값을 해당 트랙에 연관시키는 방식이다. Fig. 10에서 게이트 내에 z_1, z_2 가 존재함을 확인

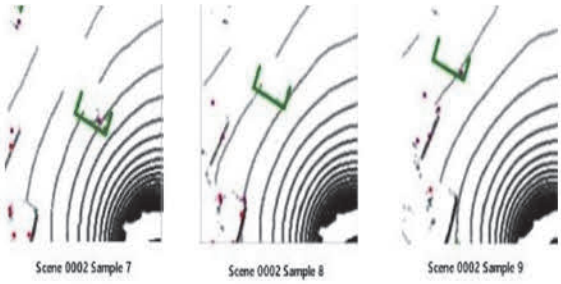


Fig. 11 Example of temporary missed detection across consecutive frames (Scene 0002, Samples 7–9), where the ground truth bounding box (green) shows a vehicle object that intermittently disappears in Sample 8 due to radar ghost points, demonstrating the necessity of the M/N Logic-based track manager

하였다. 결과 z_2는 게이트 내에 존재하며 NN기법을 통해 유클리드 거리가 더 가까운 z_1이 최종적으로 선택된다. z_3의 측정값은 경우에는 게이트 외부에 존재하기 때문에 제외하였다. 이를 통해 급격한 이동이나 방향 전환이 발생하더라도 잘못된 연관을 방지하고 일관된 객체 궤적을 유지한다.

트랙 관리는 M/N Logic을 기반으로 수행된다. M/N Logic은 특정 객체가 N번의 연속된 측정 중 M번 이상 유효하게 인식될 경우 해당 트랙을 생성하거나 유지하며, 조건을 만족하지 못하는 경우 트랙을 제거하는 방식이다. 처음 객체가 인식되면 예비 트랙으로 분류한 후 M/N Logic 조건을 만족할 경우 확정 트랙으로 승격한다. 이를 통해 Fig. 11에서 확인할 수 있는 레이더 데이터의 일시적인 주변 객체 미인식(Missed detection) 및 오인식(False detection) 문제를 보완하고 안정적인 객체 추적 성능을 확보하였다. Kalman Filter 예측, Nearest Neighbor 연관, M/N Logic 트랙 관리가 결합되어 희소하고 노이즈가 많은 레이더 환경에서도 객체 궤적을 안정적으로 유지하며, 앞 단계의 필터링 군집화 결과를 연속 프레임에서 일관되게 추정한다. 특히 일시적 미검출 구간에서도 트랙 연속성을 안정적으로 유지한다.

4. Results

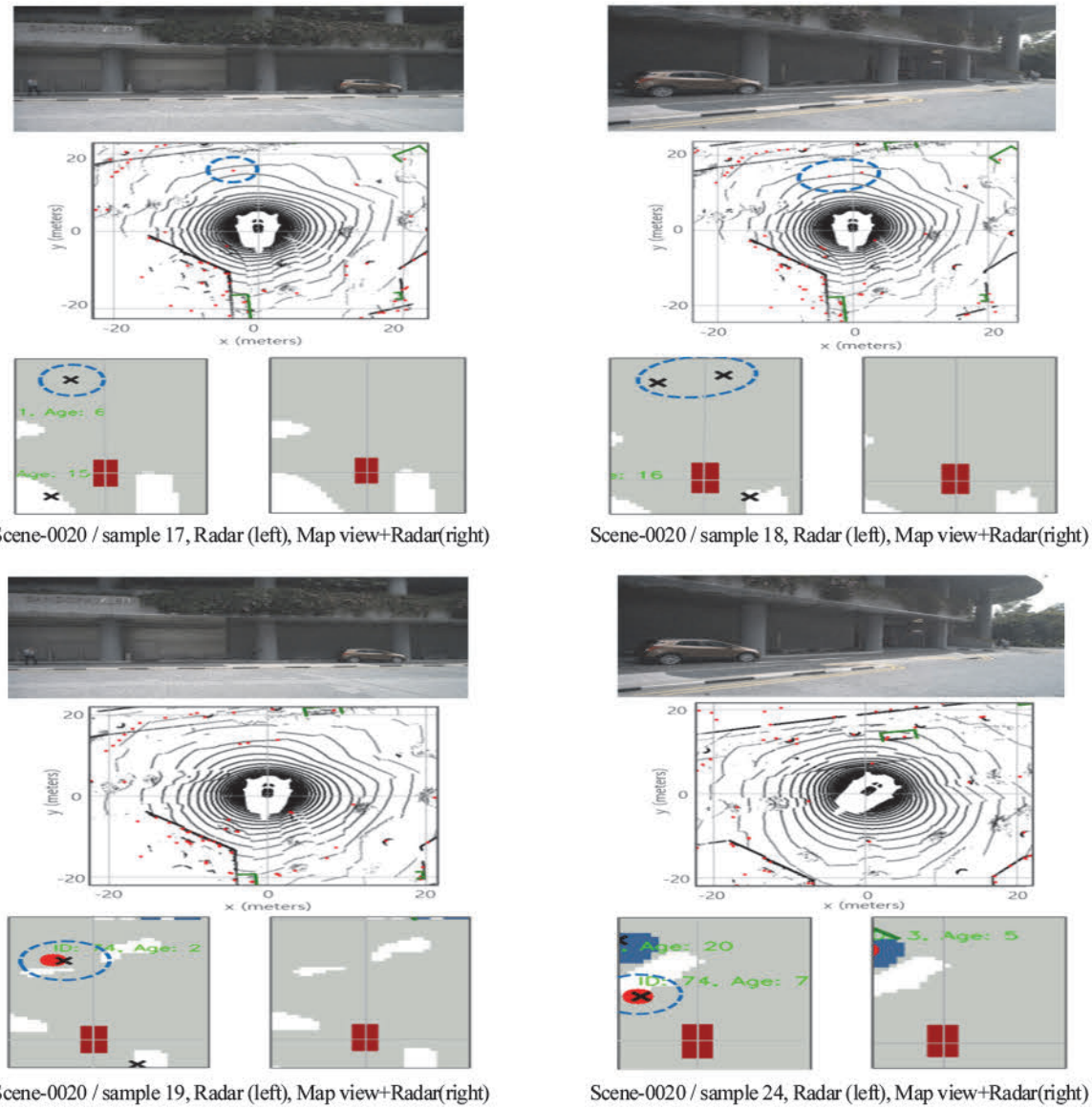
본 연구에서는 제안한 Map-view Semantic Segmentation 기반 카메라–레이더 데이터 융합 기법의 성능을 검증하기 위해 자율주행 연구에서 널리 활용되는 nuScenes 데이터셋[18]을 사용하였다. nuScenes 데이터셋은 다중 카메라와 레이더 센서로부터 수집된 동기화된 데이터를 제공하며, 다양한 도로 환경과 주행 상황을 포함하고 있다. 실험은 일반 주행 시나리오와 도심지 주행 시나리오를 대상으로 수행하여 환경 복잡도에 따른 인식 성능을 비교하였다. 선정 기준은 주변 구조물 및 정적 반사체로 인한 Ghost radar point 발생 여부와, 차량 객체의 밀도, 직선 주행 및 선회와 같은 도로 구조의 다양성, Radar-only 방식에서 False positive가 발생하여 제안한 Semantic filtering의 효과를 확인할 수 있는지 여부이다. 이를 바탕으로 일반 주행, 도심지 주행, 선회 주행 특성을 포함하는 Scene을 평가 대상으로 선정하였다.

평가에는 일반 주행, 도심지 주행, 선회 주행 특성을 대표하는 총 9개 Scene을 선정하였다. 일반 주행 Scene (Scene-0002, 0003, 0012)은 직선 위주의 도로 구조와 상대적으로 낮은 객체 밀도(평균 약 6대)를, 도심지 주행 Scene(Scene-0019, 0020, 0021)은 건물·정적 반사체가 많아 Ghost radar point와 False positive가 빈번하고 객체 밀도가 높은(평균 약 9대) 환경을, 선회 주행 Scene(Scene-0103, 0276, 0553)은 교차로 및 곡선 구간을 포함하여 차량 방향 변화가 큰 환경(평균 약 7대)을 포함한다. 각 Scene은 Radar-only 방식에서 Ghost/Noise point와 False positive가 관찰되어 제안 기법의 효과를 확인할 수 있는지를 기준으로 선정하였다. 각 scene은 40개의 Sample로 구성되었다.

본 평가는 nuScenes validation split에서 위 9개 Scene을 대상으로 수행하였다. 각 Sample의 레이더 포인트는 누적 5 sweep을 사용하였으며, nuScenes의 Sensor calibration parameter로 Ego vehicle 기준 좌표계로 변환한 뒤 Map-view semantic segmentation 결과와 동일한 BEV 좌표계로 변환하여 Filtering을 수행하였다. 객체 Annotation은 nuScenes의 Ground truth 3D bounding box를 기준으로 하였으며, 평가 대상은 Vehicle class로 제한하였다. 본 실험은 nuScenes validation set의 일부 일반 주행, 도심지 주행, 선회 주행을 대표하는 시나리오를 중심으로 한 평가를 진행하였다.

4.1 Precision 성능 비교 결과

본 절에서는 레이더 단독 사용 방식과 제안한 Map-view Semantic Segmentation 기반 카메라–레이더 융합 기법의 객체 인식 성능을 Precision 지표를 통해 비교 분석하였다. Precision은 False positive 감소 효과를 정량적으로 평가하기 위한 지표로, 고스트 객체 및 노이즈로 인한 오인식 개선 정도를 확인하는 데 활용하였다. 일반 주행 시나리오에서는 상대적으로 객체 밀도와 환경 복잡도가 낮아 레이더 단독 사용 시에도 일정 수준의 인식 성능을 보였으나, 고스트 객체와 노이즈로 인해 False positive가 발생하였다. 레이더 단독 사용 시 Precision은 67.9 %로 나타났으며, Map-view와 레이더 데이터를 융합한 결과 False positive가 감소하고 True positive가 증가하여 Precision이 90.1 %로 향상되었다.



Scene-0020 / sample 17, Radar (left), Map view+Radar(right)

Scene-0020 / sample 18, Radar (left), Map view+Radar(right)

Scene-0020 / sample 19, Radar (left), Map view+Radar(right)

Scene-0020 / sample 24, Radar (left), Map view+Radar(right)

Fig. 12 Comparison of radar-only (left) and Map-view+Radar fusion (right) results across consecutive frames (Scene-0020, Samples 17, 18, 19, and 24) in an urban driving scenario, demonstrating effective ghost object removal through camera-radar fusion, where ghost detections appearing in radar-only results (blue dashed circles) are successfully eliminated in the fusion results

Fig. 12는 도심 주행에서 레이더 단독과 융합 결과를 비교한 것으로, 레이더 단독에서 나타난 고스트 검출이 융합 후 제거됨을 보여준다. 도심지 주행 시나리오에서는 건물, 나무, 차량 등 다양한 반사체로 인해 고스트 객체와 노이즈가 빈번하게 발생하여 레이더 단독 사용 시 Precision이 45.7 %로 크게 저하되었다. 반면, Map-view semantic segmentation 결과를 활용하여 유효한 레이더 포인트만을 선별한 경우 False positive가 대폭 감소하였으며, Precision이 82.7 %로 크게 개선되었다. 이는 복잡한 도심 환경에서 제안한 융합 기법이 레이더의 한계를 보완하고 객체 인식의 신뢰성을 향상시킨다는 것을 보여준다. Precision 외에 Recall, F1-score, 종합 검출 지표 TP/ (TP+FP+FN), 추적 지표 IDSW와 MOTA를 함께 산출하였다. 일반 주행 시나리오에서 제안 기법은 Recall이 0.727에서 0.922로, F1-score가 0.702에서 0.911로, TP/(TP+FP+FN)이 0.541에서 0.837로 향상되었으며, IDSW가 6에서 2로 감소하여 MOTA가 0.336에서 0.805로 개선되었다. 도심지 주행 시나리오에서도 Recall이 0.708에서 0.858로, F1-score가 0.556에서 0.842로, TP/(TP+FP+FN)이 0.385에서 0.728로 향상되었으며, IDSW가 12에서 4로 감소하여 MOTA가

–0.245에서 0.642로 개선되었다. 여기서 MOTA는 1–(FN+FP+IDSW)/GT로 계산하였으며, Ground truth 객체 수는 일반 주행 128개, 도심지 주행 106개이다. 도심지 레이더 단독 방식의 MOTA가 음수로 나타난 것은 False positive와 Missed detection의 합이 전체 Ground truth 수를 초과하였기 때문이며, 제안 기법 적용 시 두 오류가 모두 감소하여 양의 값으로 회복되었다. 제안 기법은 차량 예측 영역 외부의 레이더 포인트를 제거함에도 Recall이 함께 향상되었는데, 이는 영역 기반 필터링으로 False positive가 감소하는 동시에 Kalman filter 기반 추적과 M/N Logic 트랙 관리기가 일시적 미검출 구간의 객체를 연속 프레임에서 복원하였기 때문이며, 따라서 본 기법은 Precision–Recall trade-off의 한쪽을 희생시키지 않고 두 지표를 동시에 개선하는 경향을 보였다.

4.2 거리 오차 분석

Map-view semantic segmentation 기반 카메라–레이더 융합 기법의 객체 위치 추정 정확도를 평가하기 위해 거리 오차(Distance error)를 분석하였다. 거리 오차는 nuScenes Ground Truth Bounding Box 중심 좌표와 추적 결과 중심 좌표 간의 유클리드 거리를 기준으로 계산하였으며, 주행 상황에 따른 성능 차이를 확인하기 위해 직진 주행 시나리오와 선회 주행 시나리오로 구분하여 평가하였다.

직진 주행 시나리오에서 레이더 단독 사용 시 평균 거리 오차는 1.99 m로 나타났으나, Map-view와 레이더 데이터를 융합한 결과 평균 거리 오차가 1.33 m로 감소하였다. 특히 횡방향 오차가 크게 감소하였으며, 이는 레이더 데이터가 객체의 전체 형상보다는 제한된 방향으로 분포하는 특성에 기인한 것으로 분석된다. 이는 차량 영역 밖에 분포하던 고스트 노이즈 포인트가 제거되어 클러스터 중심이 실제 객체 중심에 가깝게 형성되기 때문으로 분석된다. 레이더 단독 방식에서는 객체 주변에 산재한 반사 포인트가 클러스터 중심을 한쪽으로 치우치게 하여 거리 오차를 키우는 경향이 있으나, Map-view 기반 선별을 적용하면 이러한 이상 포인트의 영향이 줄어든다. 특히 레이더는 각도 해상도의 한계로 횡방향 측정 불확실성이 크게 나타나는데, 영역 기반 필터링이 경계 밖 포인트를 제거하면서 횡방향 분산이 효과적으로 감소하였다. 다만 융합에 따른 거리오차 개선 폭은 주행환경의 복잡도에 따라 달라질 수 있다. 또한 종방향 오차는 거리 측정이 비교적 안정적이어서 횡방향에 비해 거선 폭이 작았으며, 이는 레이더가 반경 방향거리에는 강인하지만 방위각 추정에는 취약한 특성과 일치한다.

Fig. 13에서 노이즈가 많은 도로 공사 환경에서도 레이더 단독 사용 시 평균 거리 오차가 2.82 m로 크게 나타난

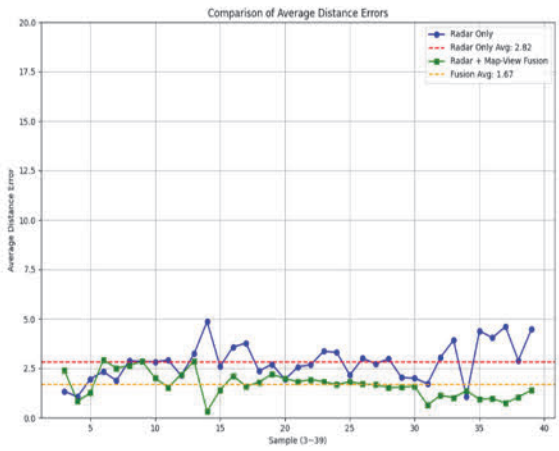


Fig. 13 Comparison of average distance errors between radar-only (blue, avg: 2.82 m) and Radar-Map-View fusion (green, avg: 1.48 m) across samples in a noisy straight-driving scenario, demonstrating that the proposed fusion method consistently reduces distance error throughout all frames

반면, 제안 기법을 적용한 경우 1.48 m로 감소하여 Map-view 기반 융합이 노이즈 환경에서도 위치 추정 정확도를 효과적으로 향상시킴을 확인하였다. 선회 주행 시나리오에서는 총 3개의 Scene에 대해 거리 오차를 비교 분석하였다. 모든 Scene에서 레이더 단독 사용 대비 Map-view와의 융합 결과 평균 거리 오차가 감소하였으며, 특히 횡방향 오차가 일관되게 개선되는 경향을 보였다. 이는 선회 주행과 같이 차량의 방향과 위치 변화가 큰 상황에서도 Map-view 기반 시맨틱 정보가 레이더 데이터의 불확실성을 보완하여 안정적인 객체 위치 추정을 가능하게 함을 의미한다.

4.3 추적 성능평가

추적 성능 평가는 도심지 주행 시나리오를 대상으로 추적 알고리즘 적용 전후를 비교하였다. IDSW는 동일한 Ground truth 객체에 연관된 Track ID가 연속 프레임 사이에서 변경된 횟수로 산출하였으며, M/N Logic 적용 전후의 동일 Scene에 대해 동일한 연관 기준으로 계산하였다. 추적을 적용하지 않은 경우 True positive 81, False positive 23, FN 25, IDSW 12로 Precision 0.778, Recall 0.764, F1-score 0.771, MOTA 0.434를 보였다. 반면 Kalman Filter 기반 상태 예측과 Nearest Neighbor 데이터 연관, M/N Logic 트랙 관리기를 적용한 경우 True positive가 91로 증가하고 False positive가 19, FN이 15, IDSW가 4로 감소하여 Precision 0.827, Recall 0.858, F1-score 0.842, MOTA 0.642로 향상되었다. 특히 FN과 IDSW의 동반 감소는 M/N Logic 기반 트랙 관리기가 레이더 환경에서 빈번한 Missed

detection과 잘못된 Track 생성을 완화함을 정량적으로 뒷받침한다.

5. Conclusion

본 연구에서는 자율주행 환경에서 레이더 단독 사용 시 발생할 수 있는 Ghost object 및 Noise point 문제를 완화하기 위해, Map-view semantic segmentation 기반 카메라–레이더 융합 기법을 제안하였다. 제안 기법은 Cross-view transformer 기반 Map-view semantic segmentation 결과를 Semantic prior로 활용하여 차량 예측 영역 내에 위치한 레이더 포인트만을 선별하고, 이를 DBSCAN 클러스터링 및 Kalman Filter 기반 다중 객체 추적 구조와 결합함으로써 Radar 기반 객체 검출 및 추적 입력의 신뢰도를 향상시키는 것을 목표로 한다.

실험 결과, 제안한 Map-view 기반 Radar filtering 기법은 레이더 단독 방식과 비교하여 Ghost point 및 False positive를 효과적으로 감소시키는 경향을 보였다. 일반 주행 시나리오에서는 Precision이 67.9 %에서 90.1 %로 향상되었으며, 도심지 주행 시나리오에서는 45.7 %에서 82.7 %로 개선되었다. 또한 거리 오차 분석에서도 Map-view semantic segmentation 결과를 활용한 경우 레이더 단독 사용 대비 평균 거리 오차가 감소하는 것을 확인하였다. 이는 카메라 기반 Semantic prior가 레이더 측정값 중 실제 차량 영역과 관련성이 낮은 Point를 제거함으로써, 후속 클러스터링 및 추적 단계에서 오검출을 줄이는 데 기여할 수 있음을 보여준다.

다만 본 연구는 제안한 Filtering-oriented fusion 구조의 가능성을 확인하는 데 초점을 둔 만큼, 향후 보다 다양한 조건에서의 추가 검증이 필요하다. 따라서 향후 연구에서는 야간, 우천, 안개 등의 환경 조건을 포함한 평가를 수행하고, Segmentation confidence 기반 Adaptive filtering 또는 Radar-only fallback 구조를 도입하여 카메라 기반 Semantic prior에 대한 의존성을 완화할 필요가 있다.

또한 평가 Scene을 확대하고, Recall, F1-score, Missed detection rate, Track recall 등 다양한 지표를 함께 적용하여 Precision–Recall trade-off와 추적 안정성을 종합적으로 분석할 계획이다. 더불어 레이더 Radial velocity와 Ego-motion 보상을 포함한 비선형 추적 모델(EKF/UKF 등)로 확장하고, Vehicle class뿐만 아니라 보행자, 이륜차, 자전거 등 다양한 동적 객체로 적용 대상을 넓힘으로써 보다 강건한 자율주행 인지 시스템으로 발전시킬 수 있을 것으로 기대된다. 또한 RCBEVDet, CRN 등 SOTA 3D object detection 모델과 CenterFusion, ClusterFusion 등 Camera-radar fusion 방법과의 동일 조건 정량 비교를 수행하고, nuScenes validation split 전체 또는 Mini split 수준으로 평가 범위를 확장하며 scene별 표준편차 및 신뢰구간을 함께 제시할 계획이다. 더불어 거리 임계값 기반 누적 추적 지표인 AMOTA와 AMOTP를 추가로 산출하여 추적 성능을 보다 종합적으로 검증할 계획이다.

한편 제안 파이프라인은 카메라 기반 Cross-view transformer의 Map-view semantic segmentation 결과를 레이더 포인트 필터링의 기준으로 사용하므로, 조명 부족, 우천, 안개 등으로 Segmentation 결과가 누락되거나 부정확해질 경우 유효한 레이더 포인트까지 제거되어 레이더의 환경 강건성이 시스템 수준에서 제한될 수 있다. 또한 차량 예측 영역을 기준으로 선별하므로 차량 영역 외부에 위치한 보행자, 이륜차, 자전거 등 다른 동적 객체의 레이더 포인트가 함께 제거될 수 있으며, 본 연구는 Vehicle class를 중심으로 Ghost/Noise point 제거 효과를 분석하였다는 한계가 있다. 본 연구는 차량 Class를 대상으로 Semantic 영역 기반 선별을 수행하였으며, 향후 보행자, 이륜차, 자전거 등 다양한 동적 객체의 Semantic label을 함께 활용하는 다중 Class 기반 선별 구조로 확장할 계획이다.

후 기

This research was supported by Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE) and Korea Planning & Evaluation Institute of Industrial Technology(KEIT) in 2026(RS-2024-00445826).

References

1) Z. Wang, Y. Wu and Q. Niu, “Multi-Sensor Fusion in Automated Driving: A Survey,” IEEE Access, Vol.8, pp.2847-2868, 2020.
2) Z. Lin, Z. Liu, Z. Xia, X. Wang, Y. Wang, S. Qi, Y. Dong, N. Dong, L. Zhang and C. Zhu, “RCBEVDet: Radar-Camera Fusion in Bird’s Eye View for 3D Object Detection,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.14928-14937, 2024.
3) Z. Liu, H. Tang, A. Amini, X. Yang, H. Mao, D. Rus and S. Han, “BEVFusion: Multi-Task Multi-Sensor Fusion with Unified Bird's-Eye View Representation,” Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pp.2774-2781, 2023.
4) J. Philion and S. Fidler, “Lift, Splat, Shoot: Encoding Images from Arbitrary Camera Rigs by Implicitly Unprojecting to 3D,” Proceedings of the European Conference on Computer Vision (ECCV), pp.194-210,

2020.

5) R. Nabati and H. Qi, "CenterFusion: Center-Based Radar and Camera Fusion for 3D Object Detection," Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp.1527-1536, 2021.
6) I. T. Kurniawan and B. R. Trilaksono, "ClusterFusion: Leveraging Radar Spatial Features for Radar-Camera 3D Object Detection in Autonomous Vehicles," IEEE Access, Vol.11, pp.121511-121528, 2023.
7) Y. Kim, J. Shin, S. Kim, I. Lee, J. W. Choi and D. Kum, "CRN: Camera Radar Net for Accurate, Robust, Efficient 3D Perception," Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp.17615-17626, 2023.
8) S. Yao, R. Guan, X. Huang, Z. Li, X. Sha, Y. Yue, E. G. Lim, H. Seo, K. L. Man, X. Zhu and Y. Yue, "Radar-Camera Fusion for Object Detection and Semantic Segmentation in Autonomous Driving: A Comprehensive Review," IEEE Transactions on Intelligent Vehicles, Vol.9, No.1, pp.2094-2128, 2024.
9) Z. Wei, F. Zhang, S. Chang, Y. Liu, H. Wu and Z. Feng, "MmWave Radar and Vision Fusion for Object Detection in Autonomous Driving: A Review," Sensors, Vol.22, No.7, Paper No.2542, 2022.
10) A. Srivastav and S. Mandal, "Radars for Autonomous Driving: A Review of Deep Learning Methods and Challenges," IEEE Access, Vol.11, pp.97147-97168, 2023.
11) Y. Zhou, L. Liu, H. Zhao, M. López-Benítez, L. Yu and Y. Yue, "Towards Deep Radar Perception for Autonomous Driving: Datasets, Methods, and Challenges," Sensors, Vol.22, No.11, Paper No.4208, 2022.
12) A. Ouaknine, A. Newson, J. Rebut, F. Tupin and P. Pérez, "CARRADA Dataset: Camera and Automotive Radar with Range-Angle-Doppler Annotations," 2020 25th International Conference on Pattern Recognition (ICPR), pp.5068-5075, 2021.
13) X. Weng, J. Wang, D. Held and K. Kitani, "3D Multi-Object Tracking: A Baseline and New Evaluation Metrics," Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp.10359-10366, 2020.
14) S. Baek, J. Kim and K. Yi, "Robust Tracking and Detection Based on Radar Camera Fusion Filtering in Urban Autonomous Driving," Intelligent Service Robotics, Vol.17, No.6, pp.1125-1141, 2024.
15) M. Ester, H. P. Kriegel, J. Sander and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD), pp.226-231, 1996.
16) H. Chiu, C. Wang, M. Chen and S. F. Smith, "Probabilistic 3D Multi-Object Cooperative Tracking for Autonomous Driving via Differentiable Multi-Sensor Kalman Filter," Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pp.18458-18464, 2024.
17) B. Zhou and P. Krähenbühl, "Cross-View Transformers for Real-Time Map-View Semantic Segmentation," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.13760-13769, 2022.
18) H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan and O. Beijbom, "nuScenes: A Multimodal Dataset for Autonomous Driving," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.11621-11631, 2020.