BEV 특징 학습을 위한 합성 데이터셋 및 데이터 큐레이션에 관한 연구

심영보*, 성기호
한국전자기술연구원
*youngbo.shim@keti.re.kr, gihosung@keti.re.kr

Synthetic Dataset Construction and Data Curation for BEV Feature Learning

Youngbo Shim*, Giho Sung
Korea Electronics Technology Institute

요 약

최근 자율주행 분야에서 Bird's-Eye-View (BEV) 표현을 활용한 End-to-End (E2E) 자율주행 기술이 주목받고 있다. E2E 자율주행 모델은 인지, 판단, 제어를 통합적으로 처리하며, 대부분의 레이어에서 BEV 특징을 활용하기 때문에 2차원 이미지를 3차원 BEV 공간으로 변환하는 성능이 전체 시스템의 핵심 요소로 작용한다. 그러나 차원 변환 과정에서 발생하는 정보 손실 및 왜곡 문제는 모델 성능 저하의 주요 원인이 되며, 이를 해결하기 위해서는 효과적인 데이터 큐레이션 전략이 필수적이다. 본 논문에서는 BEV 특징 학습의 효율성을 높이기 위해 합성 데이터셋 생성 방법론과 데이터 품질 향상을 위한 큐레이션 기법을 제안한다.

Ⅰ. 서 론

자율주행 기술의 발전과 함께 End-to-End (E2E) 학습 방식이 전통적인 모듈형 파이프라인을 대체하는 새로운 패러다임으로 부상하고 있다. E2E 자율주행 시스템은 센서 입력으로부터 주행 경로 또는 제어까지의 전 과정을 단일 신경망으로 처리하여 복잡한 모듈 간 인터페이스를 제거하고 최적화 효율을 높인다는 장점이 있다 [1, 2]. 특히 Bird's-Eye-View (BEV) 표현은 다중 카메라의 시점을 통합하여 차량 주변 환경을 조감도 형태로 나타내므로, 공간적 관계 파악과 경로 계획에 유리하여 최근 E2E 모델의 핵심 표현 방식으로 자리잡았다 [2, 3].

그러나 BEV 기반 E2E 모델의 성능은 2차원 이미지에서 3차원 BEV 공간으로의 변환 품질에 크게 의존한다. 이 변환 과정은 본질적으로 깊이 정보의 모호성, 폐색(occlusion), 그리고 원근 왜곡 등으로 인해 정보 손실과 부정확한 공간 추정이 발생할 수 있다 [4]. 또한 실제 주행 데이터는 다양한 조명 조건, 기상 상태, 교통 밀도 등에서 불균형하게 수집되어 모델이 특정 시나리오에 편향될 위험이 있다.

이러한 문제를 해결하기 위해 본 논문에서는 두 가지 핵심 방안을 제시한다. 첫째, 시뮬레이션 환경을 활용한 합성 데이터셋 생성 방법론을 통해 다양한 주행 시나리오와 정확한 Ground Truth를 포함하는 학습 데이터를 확보한다. 둘째, BEV 변환 품질에 영향을 미치는 데이터 특성을 분석하여 효과적인 데이터 큐레이션 전략을 제안함으로써 학습 효율성과 모델 성능을 동시에 향상시키고자 한다. 제안하는 방법론은 BEV 특징 학습의 정확도를 높이고, 실제 자율주행 시스템의 안전성과 신뢰성 향상에 기여할 것으로 기대된다.

Ⅱ. 본론

본 연구에서는 자체 개발한 자율주행 시뮬레이터를 활용하여 한국 도로 환경을 디지털 트윈으로 구현하였다. 디지털 트윈 기반 시뮬레이션 환경은 실제 도로의 기하학적 구조, 교통 신호 체계, 그리고 주행 패턴을 정밀하게 재현함으로써 현실성 높은 합성 데이터 생성을 가능하게 한다. 시뮬레이션 환경의 가장 큰 장점은 실제 환경에서 데이터 수집이 어렵거나 위험한 교통 상황을 안전하게 연출할 수 있다는 점이다. 예를 들어, 교차로에서의 복잡한 다중 차량 상호작용, 급정거 상황, 보행자 무단횡단 등 극단적이거나 드물게 발생하는 시나리오를 의도적으로 생성하여 모델의 학습 데이터 다양성을 확보할 수 있다. 이는 실제 주행 데이터 수집 시 발생하는 데이터 불균형 문제를 효과적으로 해결한다.

1) BEV 특징 학습을 위한 합성 데이터셋 생성

그림 1에서와 같이 본 시뮬레이터는 다양한 센서 모달리티의 합성 데이터를 동시에 생성할 수 있도록 설계되었다. RGB 카메라 이미지를 기본으로 하여 Depth 카메라, Instance 및 Semantic Segmentation, Lidar(Semantic Lidar 포함), 3D Bounding Box Ground Truth, Radar, GNSS, IMU 등의 센서 데이터를 완벽하게 동기화된 형태로 제공한다. 특히 모든 센서 데이터는 정확한 Ground Truth와 함께 생성되므로, 실제 환경에서는 취득하기 어려운 완벽한 레이블 정보를 확보할 수 있다.

BEV 특징 학습의 관점에서 본 합성 데이터셋의 핵심 가치는 카메라 perspective view에서 모든 픽셀의 정확한 metric depth 정보를 제공한다는 점이다. 일반적으로 실제 카메라 이미지에서는 깊이 정보가 명시적으로 제공되지 않으며, Lidar와의 융합을 통해서도 sparse한 깊이 정보만을 얻을 수 있다. 그러나 시뮬레이션 환경에서는 렌더링 과정에서 각 픽셀의 정확한 3차원 좌표가 계산되므로, dense하고 정확한 metric depth 정보를 모든 픽셀에 대해 획득할 수 있다.

이러한 픽셀 단위의 정확한 깊이 정보는 2차원 이미지를 3차원 BEV 공간으로 변환하는 view transformation의 완벽한 supervision 신호로 작용한다. 기존 연구들에서 BEV 변환 학습 시 겪었던 깊이 추정의 불확실성과 이로 인한 공간적 왜곡 문제를 합성 데이터의 정확한 depth ground truth를 통해 해결할 수 있다. 모델은 이러한 완벽한 supervision을 통해

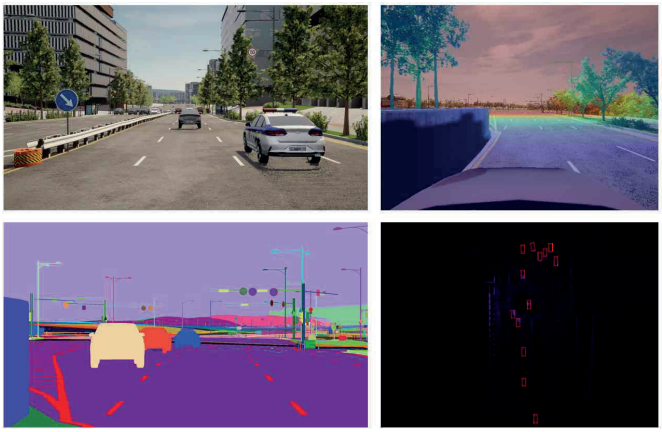

그림 1. 합성데이터 예시: RGB 카메라(좌·상단), Depth 카메라(우·상단), Instance Segmentation(좌·하단), LiDAR & GT(우·하단)

카메라 이미지의 각 픽셀이 BEV 공간의 어느 위치에 매핑되어야 하는지를 정확하게 학습할 수 있으며, 이는 최종적으로 BEV 기반 인지 성능 향상으로 이어진다.

2) Visibility 기반 데이터 큐레이션

BEV 기반 E2E 모델의 학습 과정은 일반적으로 두 단계로 구성된다. 먼저 view transformation 단계에서 task head에서 3D bounding box, segmentation 등 ground truth(GT)를 활용하여 각 task head와 BEV 특징을 동시에 학습한다. 문제는 이러한 GT 중 일부가 카메라의 perspective view에서 볼 때 다른 객체나 장애물에 의해 완전히 가려져 visibility가 0에 가까운 경우에도 학습 데이터로 포함된다는 점이다. 예를 들어, 여러 대의 차량이 밀집된 상황에서 뒤쪽에 위치한 차량은 카메라 이미지에서 전혀 관찰되지 않지만, 3D 공간에서는 존재하므로 GT로 생성된다. 이러한 보이지 않는 객체의 GT를 학습에 활용할 경우, 모델은 이미지에 존재하지 않는 정보를 예측하도록 강제되어 학습이 불안정해지고 성능이 저하된다.

이 문제를 해결하기 위해 본 연구에서는 instance segmentation과 metric depth를 결합하여 각 객체의 visibility를 정량적으로 계산하는 방법을 제안한다. 특정 객체 i에 대한 visibility V_i는 다음과 같이 정의된다;

$$V_i = \frac{A_{visible}}{A_{total}} = \frac{\sum_{p \in P_i} 1(d_p \in B_i)}{|P_i|} \quad (1)$$

여기서 P_i는 객체 i의 instance segmenatsion mask에 해당하는 픽셀 집합, d_p는 픽셀 p의 metric depth 값, B_i는 객체 i의 3D bounding box 내부 깊이 범위 $[d_{\min}, d_{\max}]$이고, $A_{visible}$은 실제로 보이는 픽셀 영역, A_{total}은 instance mask의 전체 픽셀 영역을 나타낸다.

Instance segmentation 정보를 활용하여 각 객체에 해당하는 픽셀들을 식별하고, 해당 픽셀들의 metric depth 값이 객체의 3D bounding box 깊이 범위 내에 존재하는지 확인한다. 만약 특정 픽셀의 depth 값이 bounding box 범위를 벗어나면, 해당 픽셀 위치에서 객체가 다른 물체에 의해 가려진 것으로 판단된다.

Ⅲ. 결론

본 논문에서는 BEV 기반 End-to-End 자율주행 모델의 성능 향상을 위한 합성 데이터셋 생성 방법론과 데이터 큐레이션 전략을 제안하였다. 자체 개발한 시뮬레이터를 통해 한국 도로 환경의 디지털 트윈을 구축하고, 실제 환경에서 수집하기 어려운 다양한 교통 상황을 안전하게 연출할 수 있음을 보였다.

특히 BEV 특징 학습의 핵심인 view transformation을 위해 모든 픽셀에 대한 정확한 metric depth 정보를 제공함으로써, 2D 이미지에서 3D BEV 공간으로의 변환 성능을 향상시킬 수 있는 완벽한 supervision 데이터를 확보하였다. 또한 instance segmentation과 depth 정보를 결합한 visibility 계산을 통해 카메라 시점에서 가려진 객체를 효과적으로 필터링하는 데이터 큐레이션 방법을 제시하였다. 이를 통해 학습 데이터의 품질을 높이고 모델이 실제로 관찰 가능한 정보에 집중하도록 함으로써 학습 안정성과 성능을 동시에 개선할 수 있었다.

제안한 방법론은 실제 데이터 수집의 한계를 극복하고, BEV 기반 자율주행 모델의 학습 효율성을 높이는 데 기여할 것으로 기대된다. 향후 연구에서는 실제 주행 데이터와 합성 데이터의 효과적인 혼합 학습 전략과 다양한 기상 및 조명 조건에서의 합성 데이터 생성 방법에 대한 연구를 진행할 예정이다.

ACKNOWLEDGMENT

이 논문은 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. RS-2025-02221243, 3-Tier 연계형 자율주행 소프트웨어 및 데이터 통합 검증용 클라우드 기반 평가 모델·프로세스 개발)

참 고 문 헌

[1] Bojarski, M., et al., "End to End Learning for Self-Driving Cars," arXiv:1604.07316, Apr. 2016.

[2] Hu, Y., et al., “Planning-oriented autonomous driving.”, In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 17853-17862, 2023.

[3] Philion, J., and Fidler, S., "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.

[4] Song, Z., et al., "Synthetic Datasets for Autonomous Driving: A Survey," arXiv:2304.12205, Feb. 2024.