

한국지형에서의 카메라 기반 실시간 3차원 객체인식 기술 연구

한지은, 고석현, 임예원, 이성진

동서울대학교 전자공학과

haneul2069@du.ac.kr, sungjinlee@du.ac.kr

A Study on the Camera based 3D Object Detection in Korea

Jieun Han, Sukhyun Ko, Yewon Lim, Sungjin Lee

Department of Electronic Engineering, Dong Seoul University

요약

자율주행 기술의 실현을 위해 필수적인 3차원 객체 인식 기술에는 카메라 기반 방식과 Lidar 기반 방식의 두 가지 접근법이 있다. 자율주행기술은 안전성이 중요시되기 때문에 빛 반사, 어두움, 착시 현상 등의 영향으로 상황에 따라 불안정할 수 있는 카메라 기반의 3D Object Detection 기술보다는 Lidar 기반 방식이나 그 둘의 Multi-Sensor Fusion 기술이 주로 다루어지고 있다. 본 논문에서는 다양한 환경에서 카메라 기반의 3D Object Detection 기술의 성능을 측정해보고 그 가능성을 분석해 보았다. 특히, 한국 지형과 여러 주행 환경 (도심, 교외, 고속도로) 에 따라서도 안정적으로 잘 동작할 수 있는지를 평가하기 위해 자체 데이터 셋을 구축하여 여러 상황에 따른 실험을 해 보고 각각의 시나리오에서 좀 더 성능을 높이기 위한 전략들을 도출하였다.

I. 서론

최근 딥러닝 기술의 발전으로 자율주행 기술은 운전 보조 기능을 넘어서 완전 자율주행 단계에 이르고 있다. 그런 완전 자율주행을 위해서는 주변 상황에 대한 3차원 인식과 이를 통한 주행가능 경로 파악이 필수적이어서, 관련된 3D Object Detection 기술에 산업과 학계가 주목하고 있다 [1][2].

특히 이런 3D Object Detection 기술에는 카메라 기반 방식과 Lidar기반 방식의 두 가지 접근법이 있다. 우선 카메라 기반 방식은 상대적으로 데이터 양이 적고 해당 객체들의 특징에 대한 표현력이 우수하여 좋은 환경에서는 높은 정확도에 실시간 처리가 가능하다. 하지만 다양한 광학적 어려움, 즉 빛 반사, 어두움, 착시 현상 등의 영향으로 상황에 따라 불안정한 성능을 낼 수 있는 게 사실이다. 반면 Lidar기반 방식은 해당 객체들의 3D 위치를 Point Cloud Data로 상대적으로 정확하게 측정할 수 있지만 해당 데이터의 양이 많고 표현력(Intensity)이 상대적으로 부족하여 Lidar만으로는 객체의 class를 정확하게 분별해 내기 힘들어서 이를 보강하기 위해 카메라와 함께 사용하는 Multi-Modal Sensor Fusion 방식을 주로 사용한다 [1][2]. 하지만 실시간으로 처리되어야 하는 자율주행 상황에서는 두 입력신호 각각에 대한 3D Object Detection 적용 및 Multi-Modal 신호처리를 위한 정합과정을 위한 느린 동작 속도에 대한 위험성이 존재한다.

그러므로 실시간 처리 및 정확도 모두를 확보할 수 있도록 본 논문에서는 카메라 기반 3D Object Detection 기술에 대해 집중하여 해당 기술의 불리한 상황에서도 정확도를 높일 수 있는 방법에 대해 연구하였다. 특히 지역적으로 훈련 데이터와 다른 상황들, 즉 한국 지형과 여러 주행 환경 (도심, 교외, 고속도로), 시간대 (낮, 밤) 에서 어느 정도의 성능 열화가 발생하는지 분석하였고 이런 환경 분석을 위해 자체 데이터 셋을 구축하여 시나리오 별 성능 향상을 위한 기법들을 제시하였다.

II. 3D Object Detection Model

본 논문에서는 3D Object Detection 모델 중 Fast Detection 성격에 있는 RTM3D 모델 [3]에 기반하여 성능 분석을 수행하였다. 해당 모델은 Fast Detection을 위해 입력이미지로 1장만을 사용한다. 그림 1과 같이 1장의

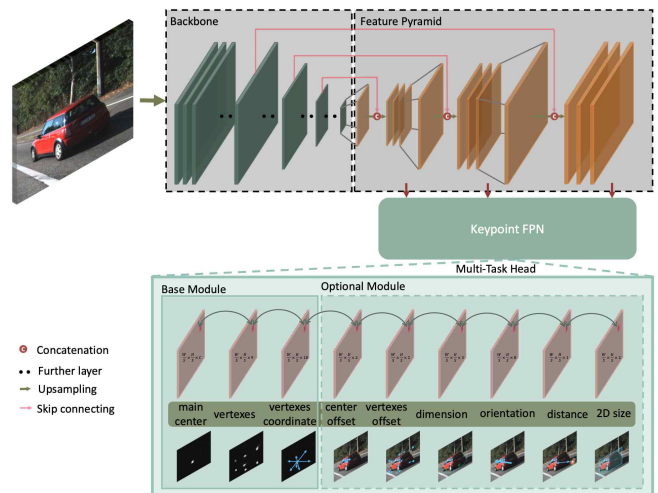


그림 1. RTM3D의 동작 구조도 [3]

이미지를 입력받아 FPN (Feature Pyramid Network) 과 Multi-Task Head를 통해 9개의 Keypoint (main center, vertexes, vertexes coordinate, center offset, vertexes offset, dimension, orientation, distance, 2D size)에 대한 Heatmap을 얻어낸다. 결과적으로 3D Bounding Box B를 위한 Rotation R, Position T, Dimension D를 얻어낸다. 즉, 기존 2D Object Detection에서 많이 사용하는 FPN 방식에 2D Keypoint Feature들을 얻어내고 이를 3D Bounding Box에 Projection 함으로서 성능 도출 시간을 단축시켰다고 볼 수 있겠다.

III. 데이터 셋 구축

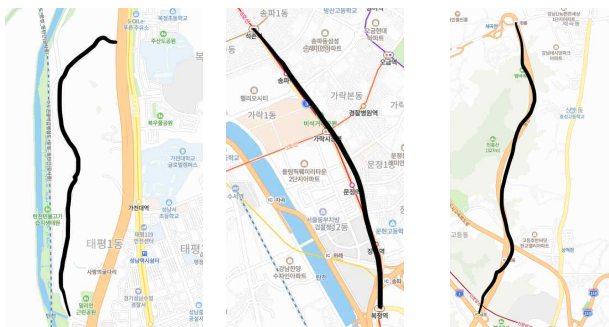
3D Object Detection을 위한 기본 훈련 데이터 셋으로 KITTI [4] 를 사용하였다. 하지만 본 데이터 셋은 미국 지형의 낮 환경 도로 주행에서 만들어진 것이기에 한국 지형과 여러 주행환경 및 밤 도로 환경에 적용하기에는 성능 열화가 발생할 수 있다. 그러므로 본 논문에서는 한국의 도로환경에서 다양한 주행 시나리오와 낮, 밤 시간대를 고려하여 주행 데이터 셋을 구축하였다. 그림 2에서는 데이터 셋 구축을 위해 사용한 기기들의 스펙을 나타내었다.



그림 2. 데이터 셋 구축을 위해 사용된 기기들

위 구축된 기기들 중 주행을 위해서 직접적으로 영향을 미치는 왼쪽 카메라 (운전자 기준)에서 수집되는 영상 이미지를 통해 데이터 셋을 구축하였다. 카메라는 FHD (1920x1080) 30 FPS 126° 화각을 지닌 WebCam을 사용하여 영상을 획득하였다. 경로는 동서울대학교를 중심으로 3가지 코스 즉, 교외 지형, 도심 지형, 고속도로 지형에 대해 낮 (12시30분~1시30분) 과 밤 (8시~9시) 시간대 각각에 대해 데이터를 획득하였다.

또한 기존의 데이터 셋은 교통상황, 도로 상황에 대한 구분이 없어서 새로운 데이터 셋을 구축하였다. 다양한 상황에 따른 정확도를 검증하기 위해 다음 그림 3과 같은 코스에 따라 그림 2의 기기들을 이용하여 실험 데이터 셋을 수집하였다.



(a) Urban Area (b) Downtown Area (c) Highway Area

그림 3. 도로 유형에 따른 3가지 주행 시나리오

IV. 실험

실험을 위해 다음과 같은 Evaluation Set을 정하여 진행하였다. 우선 기본 KITTI 데이터 셋에 대해

- KITTI dataset : 기본 KITTI 데이터 셋에서 훈련을 진행을 정하였다. 또한 그림 3에서 보듯이 다음의 3가지 상황에 따라 코스를

결정하였다.

- Urban Area Driving dataset: 교외지역으로 교통이 한적하고 간단한 총 2차선 도로 주행 데이터 셋
- Downtown Area driving dataset: 도심지역으로 교통이 복잡한 총 8차선 도로 주행 데이터 셋
- Highway Area driving dataset: 주행 속도가 빠른 고속도로로 주행 데이터 셋

해당 데이터 셋에 대해 IoU 0.5 이상을 Positive의 기준으로 삼았다.

Train	Test			
	KITTI	UA	DA	HA
KITTI	54	51	45	50
UA	51	54	52	53
DA	53	54	54	54
HA	52	53	51	53

Table 1. 4가지 주행시나리오에 대해 정확도 (Average Precision) 실험

위 실험결과에서 보듯이 차량의 밀도가 높은 한국 도심지역 (Downtown)에서의 성능이 일반적으로 열화되어 있는 것을 확인할 수 있다. 하지만, train 셋에 관련 한국 도심지역 데이터를 더 넣어서 훈련 시키면 그 정확도 손실은 다소 복구할 수 있는 것을 알 수 있다. 그 이유로 미국에서 주로 주행되는 차량의 모양과 한국에서 주로 주행되는 차량의 모양이 꽤 달라서 (예, 한국의 시내 버스) 한국 고유의 차량들은 잘 인식이 안되기 때문이다. 하지만, 이런 데이터를 추가로 훈련 데이터에 넣어서 훈련시키면 그 정확도 손실이 복구된다는 것을 알 수 있다.

반면, 교외지역 (Urban) 이나 고속도로 지역 (Highway) 과 같은 곳에는 차량의 밀도가 다소 적고 Kitti 와 비슷한 승용차 위주의 차량들이 많아서 원본 데이터 (Kitti)로 훈련 시킨 것과 거의 비슷한 정확도를 나타내는 것을 알 수 있다.

IV. 결론

본 논문에서는 미국지형에서 훈련된 KITTI 데이터 셋으로 한국지형에서 3D Object Detection 적용을 위한 시나리오 별 성능평가를 수행하였다. 실험결과와 주행되는 차량의 종류, 밀도가 성능평가에 영향을 준다는 것을 알 수 있었고 이를 극복하기 위해서는 추가적인 데이터 셋 확보와 훈련이 중요한 전략이라는 것을 보였다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기본연구사업임(No. NRF-2019R1F1A1062878)

참 고 문 헌

- [1] Rui Qian, Xin Lai, Xirong Li, "3D Object Detection for Autonomous Driving: A Survey"
- [2] Edardo Arnold, etc, "A Survey on 3D Object Detection Methods for Autonomous Driving Applications"
- [3] Peixuan Li, Huaici Zhao, Pengfei Liu, Feidao Cao, "RTM3D: Real-time Monocular 3D Detection from Object Keypoints for Autonomous Driving"
- [4] <http://www.cvlibs.net/datasets/kitti/>