

# 라이다 포인트 클라우드 시계열 데이터의 움직임 정보를 활용한 3 차원 물체 검출 기술

고준호, 최준원\*  
한양대학교

jhkoh@spa.hanyang.ac.kr, \*junwchoi@hanyang.ac.kr

## 3D Object Detection using Motion Information from LiDAR Point Cloud Sequences

Junho Koh, Jun Won Choi\*  
Hanyang University

### 요 약

본 논문은 라이다 센서의 시계열 포인트 클라우드 데이터의 시간적 움직임 정보를 활용하여 새로운 3 차원 물체 검출 기술을 제안한다. 우선 라이다 센서에서 취득된 연속된 포인트 클라우드 데이터를 단기 시퀀스와 장기 시퀀스로 나눈다. 단기 시퀀스 내에 존재하는 포인트 정보들 사이의 시간적 분포를 고려하여 복셀 특징지도를 생성한다. 이렇게 얻어진 복셀 특징지도는 백본 네트워크를 통과하여 bird's eye view (BEV) 특징지도를 생성한다. 이 BEV 특징지도는 장기 시퀀스 내의 움직임 정보를 이용하여 현재 BEV 특징지도의 위치에 맞도록 정렬된다. 이렇게 정렬된 BEV 특징지도와 현재 BEV 특징지도는 transformer 에서 활용되는 attention 기법을 시간 영역으로 확장하여 집계되고, 이렇게 집계된 BEV 특징지도는 최종 3 차원 물체 검출 결과를 추출하는데 사용된다. 본 논문에서는 nuScenes 데이터셋을 이용하여 실험을 진행하였다. 각 제안한 방식은 현저한 성능 향상을 보여 기존 방법에 비해 높은 성능을 달성하였다.

### I. 서 론

본 논문에서는 최근 자율 주행 환경에서 활발히 진행되고 있는 라이다 기반 3 차원 물체 검출 기술을 시간 영역으로 확장하여 시간적 변위를 고려한 알고리즘을 소개한다. 실제 라이다 센서에서 취득되는 데이터는 연속적인 시계열 구조를 가지지만, 기존 라이다 기반 3 차원 물체 검출 기술들은 이 연속적인 데이터 내에 존재하는 시간정보를 활용하지 않기 때문에, 성능 향상에 한계가 있었다. 제안하는 알고리즘의 경우, 시간 정보를 단기 시퀀스와 장기 시퀀스에서 각각 활용하여 보다 풍부한 시간 정보, 특히 움직임 정보를 활용하여 기존 기술 대비 높은 성능의 3 차원 물체 검출 결과를 얻을 수 있다.

### II. 본론

#### 2.1) 제안하는 3 차원 물체 검출 기술

제안하는 3 차원 물체 검출 알고리즘은 라이다 포인트 클라우드 시계열 데이터를 단기 시퀀스와 장기 시퀀스로 구성하여 모델의 입력으로 사용한다. 전체 구조는 그림 1 에 첨부하였다.

우선, 단기 시퀀스 내에 존재하는 라이다 포인트 집합을 규칙적인 복셀 구조로 표현하기 위하여, 각 복셀

내에 존재하는 포인트들 사이의 시간적 분포를 고려한 embedding 벡터를 추출한다. 이렇게 추출된 복셀 특징지도는 여러 convolutional layer 를 통과하여 bird's eye view (BEV) 특징지도를 생성하는데 사용된다. BEV 특징지도를 생성할 때, 중간 layer 에서 생성된 multi-scale BEV 특징지도를 합쳐 최종 BEV 특징지도가 여러 크기의 물체를 검출하는데 도움을 주도록 한다.

각 단기 시퀀스에서 추출된 BEV 특징지도는 장기적인 시간적 움직임을 고려하여 융합하기 위하여 우선 현재 BEV 특징지도의 위치에 맞도록 이전 BEV 특징지도를 정렬한다. 이를 위해, 현재 BEV 특징지도와 이전 BEV 특징지도 사이의 움직임 정보를 활용한 Deformable convolution [1] 기술 기반 정렬 알고리즘을 이용하여 이전 특징지도를 정렬한다. 잘 정렬된 이전 특징지도를 현재 특징지도와 최종적으로 융합하기 위하여, deformable DETR 에서 활용된 deformable attention 기법을 시간 영역으로 확장하여 attention 기반 BEV 특징지도 융합을 수행한다.

#### 2.2) 단기 시퀀스 BEV 특징지도 추출

단기 시퀀스 내의 포인트 집합을 이용하여 BEV 특징지도를 추출하기 위해, 우선 규칙적인 복셀 구조로 포인트를 그룹화 한다. 각 복셀 내에 있는 포인트들을 표현하기 위해, 기존에 단순히 포인트 정보 (좌표 값  $x$ ,  $y$ ,  $z$  와 반사율  $r$  로 표현)의 평균을 계산할 뿐만 아니라, 시간적 분포를 함께 고려한다. 시간적 분포를 고려하기

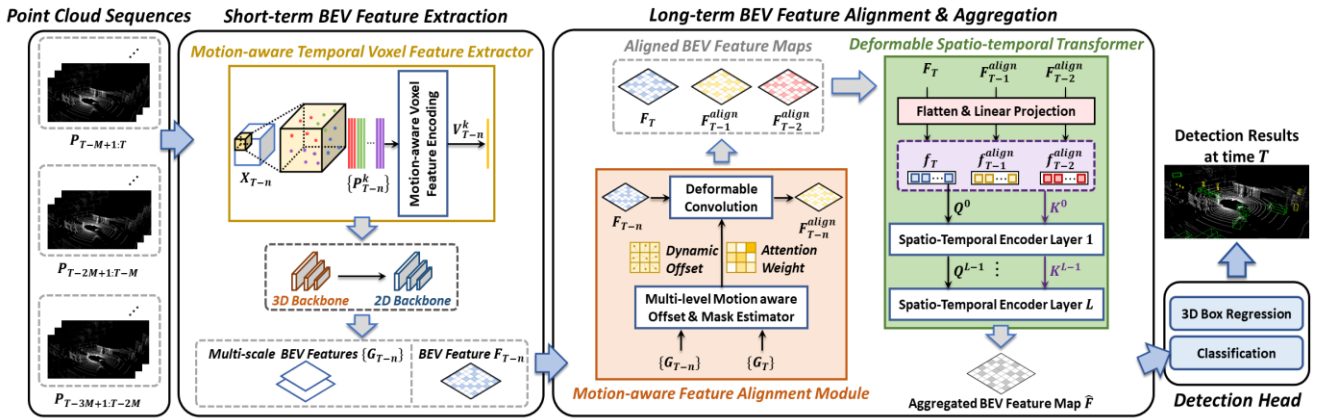


그림 1

위해, 각 시간별 포인트 집합들의 평균을 구하고, 이들 사이의 움직임을 표현하기 위해 MLP 구조를 통과한다. 이렇게 얻은 latent embedding 벡터의 움직임 정보를 극대화하기 위하여 self-attention 기법을 활용하여 최종 포인트 움직임 벡터를 추출한다. 이 움직임 벡터는 기존에 사용된 포인트 정보를 평균으로 표현한 벡터와 concatenation 되어 최종 복셀 표현으로 사용된다.

이렇게 표현된 복셀 특징지도는 BEV 특징지도를 추출하기 위하여 convolution layer 를 통과한다. 이 때, 중간 layer 에서 얻은 여러 크기의 특징지도들을 합하여 더욱 강화된 최종 BEV 특징지도를 추출한다. 이때 사용된 중간 특징지도는 장기 시퀀스에서 각 BEV 특징지도를 정렬하는데 다시 사용된다.

### 2.3) 장기 시퀀스 BEV 특징지도 융합

각 단기 시퀀스에서 얻은 BEV 특징지도는 에고 차량(ego-vehicle)의 움직임과 타 물체의 움직임으로 인해 물체 위치가 계속해서 변한다. 이는 이후에 이루어질 BEV 특징지도의 시공간적 융합을 수행하는데 방해요인이 된다. 따라서, 이러한 변위를 고려하여 이전 BEV 특징지도를 현재 BEV 특징지도로 잘 정렬하는 작업이 필요하다. 우선, 이전 BEV 특징지도와 현재 BEV 특징지도 사이의 변위를 표현하는 움직임 특징지도를 추출한다. 이 움직임 특징지도는 특징지도 정렬에서 자주 활용되는 deformable convolution [1]의 offset 과 attention weight 를 예측하는데 사용되고, 이는 deformable convolution 의 파라미터로 활용되고, 이전 특징지도를 입력으로 deformable convolution 을 연산하여 정렬된 특징지도를 추출한다.

잘 정렬된 특징지도는 transformer 구조에서 활용되는 attention 기법을 통해 융합된다. 특히, 최근에 매우 효율적인 기법으로 알려진 deformable DETR [2]에서 사용되는 deformable attention 방식을 활용하여 정렬된 이전 BEV 특징지도와 현재 BEV 특징지도를 융합한다.

### 2.4) 실험 결과

제안하는 네트워크를 학습 및 성능 평가를 위하여 nuScenes 자율주행 공개 데이터셋을 이용하였다.

제안하는 기법의 성능 평가 결과는 Table 1 과 같이 나타났다. Table 1 을 통해 nuScenes 3 차원 물체 검출 데이터셋에서 기존 시간 정보를 활용하지 않은 대표적인 3 차원 물체 검출 기술인 CenterPoint [3]에 비해 높은 성능을 나타냈다.

Method	NDS (%)	mAP (%)
CenterPoint [3]	66.8	59.6
Our method	70.6	64.8

표 1

## III. 결론

본 논문에서는 라이다 포인트 데이터의 시간적 정보를 활용한 새로운 3 차원 물체 검출 기술을 소개하였다. 단기 시퀀스 내에 존재하는 포인트들의 시간적 분포를 고려하여 강화된 복셀 특징지도를 추출하고, 이는 BEV 특징지도로 사용된다. 이 BEV 특징지도는 장기적인 움직임 정보를 고려하여, 이전 특징지도들을 현재 특징지도의 위치로 정렬시키고, 시공간 attention 기법을 활용하여 집계하여 강화된 BEV 특징지도를 추출한다. 해당 알고리즘은 nuScenes 공개 데이터셋에서 학습과 성능 평가가 이루어졌으며, 타 알고리즘 대비 매우 높은 성능을 보였다.

## ACKNOWLEDGMENT

“이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.2021-0-01314, 자율주행차량 음영지역 데이터 제공을 위한 주행환경 데이터 스티칭 기술개발)”

## 참 고 문 헌

- [1] Dai, J., et al. "Deformable convolutional networks," In Proceedings of the IEEE international conference on computer vision (pp. 764-773).
- [2] Zhu, X., et al. "Deformable DETR: Deformable transformers for end-to-end object detection," In International Conference on Learning Representations.
- [3] Yin, T., et al. "Center-based 3d object detection and tracking," In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11784-11793).