자율주행을 위한 3 차원 물체 검출 및 추적 통합 기술

고준호, 김재겸, 최준원*
한양대학교

jhkoh@spa.hanyang.ac.kr, jkkim@spa.hanyang.ac.kr, junwchoi@hanyang.ac.kr

Joint 3D Object Detection and Tracking for Autonomous Driving

Junho Koh, Jaekyum Kim, Jun Won Choi*
Hanyang University

요 약

본 논문은 카메라와 LiDAR 센서를 이용하여 새로운 3 차원 물체 검출 및 추적 통합 기술을 제안하였다. 우선 카메라와 LiDAR 에서 취득된 연속된 이미지들을 이용하여 시공간 특징지도를 추출한다. 추출된 시공간 특징지도는 센서융합 기술로 융합된 시공간 특징지도를 추출하고 이를 Tracklet-aware 모듈을 통해 물체 추적 정보를 이용하여 물체 검출 결과를 얻는다. 이렇게 얻어진 물체 검출 결과를 이용하여 Graph Neural Network (GNN)을 통과시켜 rule 기반의 edge-pruning 과 attention 기반의 edge-gating 을 수행한다. 이를 통해 data association 을 하여 물체 추적 결과를 얻는다. 본 논문에서는 KITTI 와 nuScenes 데이터셋을 이용하여 실험을 진행하였다. 각 제안한 방식은 현저한 성능 향상을 보여 기존 방법에 비해 높은 성능을 달성하였다.

Ⅰ. 서 론

본논문에서는 최근 자율 주행 환경에서 활발히 진행되고 있는 3 차원 물체 검출 연구와 물체 추적 알고리즘을 통합하는 알고리즘을 소개한다. 기존에 물체 검출 및 추적 알고리즘은 각각 목적으로 개발되었다. 하지만 물체 검출과 추적 알고리즘은 결과가 서로 연결이 되기 때문에 동시에 최적화를 통하여 서로 협력하여 높은 성능을 얻을 수 있다. 따라서 3 차원 물체 검출에서 사용한 정보와 3 차원 물체 추적에서 나온 정보를 서로 사용하여 상호 보완하는 알고리즘을 제안하였다.

Ⅱ. 본론

2.1) 제안하는 물체 검출 및 추적 기술

제안하는 3 차원 물체 검출 및 추적 통합 알고리즘은 물체 검출 단계와 물체 추적 단계로 구성되어 있으며, Fig. 1 에 전체 구조를 첨부하였다.

물체 검출 부분은 카메라와 LiDAR 센서를 이용하여 센서융합 기반의 3D-CVF 모델[1]에 시공간 특징지도 융합 알고리즘을 결합하여 사용하였다. 시공간 특징지도를 추출하기 위해서 Spatio-temporal Feature Aggregation Network (SFANet)을 사용하여 각 센서 특징지도 추출 단계에서 시간 축으로 특징지도를 융합하였다. 추출된 각 센서 특징지도는 3D-CVF 에서 사용된 융합 방식을 사용하였으며, 이를 Tracklet aware Detection network(Trk-Det)에서 최종 3 차원 물체 검출 결과를 얻었다. Trk-Det 는 이전 시간까지 누적된 물체 추적 결과를 이용하였다. 특히, 이러한 물체 추적 결과는 classification score 가 정제되는 효과를 가지고 있으며, 이는 더 정확한 3 차원 물체 검출 결과를 얻도록 도와준다.

두 번째로 물체 추적 단계에서 Data association 을 수행하기 위해서 물체 검출 알고리즘에서 추출된 특징지도를 이용한다. 물체 검출 네트워크에서 얻은 특징지도에서 물체 검출 결과 부분을 추출한 후, 제안하는 Spatio-temporal Gated Graph Neural Network (SG-GNN)을 통해 feature aggregation 을 진행한다. 이는 rule 기반 edge-pruning 과 attention 기반 edge gating 기법을 이용하여 GNN 을 구성하였다.

2.2) SFANet

SFANet 은 연속된 데이터를 시간 축으로 융합해주는 네트워크이다. 시간 t 와 시간 t-1 에서의 각 특징지도를 concatenation 하고 1-channel CNN 을 통과시키고 Sigmoid 함수를 적용하여 각 시간별 특징지도에 대한 weight 지도를 만들어준다. 이렇게 만들어준 weight 지도를 해당 특징지도에 곱하고 더해주어 선택적으로 융합된 특징지도를 만들어준다.

2.3) Trk-Det

시간 t-1 까지 얻은 물체 추적 결과를 이용하여 시간 t 에서의 물체 검출 결과 성능을 향상시키는 네트워크를 제안하였다. 시간 t-1 에서 박스 결과와 시간 t 에서의 anchor box 사이의 Intersection of Union (IoU)를 계산하여 이를 특징 지도로 mapping 하는 과정을 RPN 단계와 refinement 단계에서 모두 진행한다. RPN 단계에서는 기존의 특징지도, IoU 특징지도와 1 차적으로 얻은 score 지도를 모두 합치고 CNN 을 통과하여 새로운 score 를 계산해준다. 다음으로 refinement 단계에서도 마찬가지로 이러한 과정을 수행하여 최종 detection score 를 계산해준다.

2.4) SG-GNN

물체 추적은 물체 검출 결과를 기반으로 하여 물체 검출 네트워크에서 얻은 특징지도에서 RoI align 을

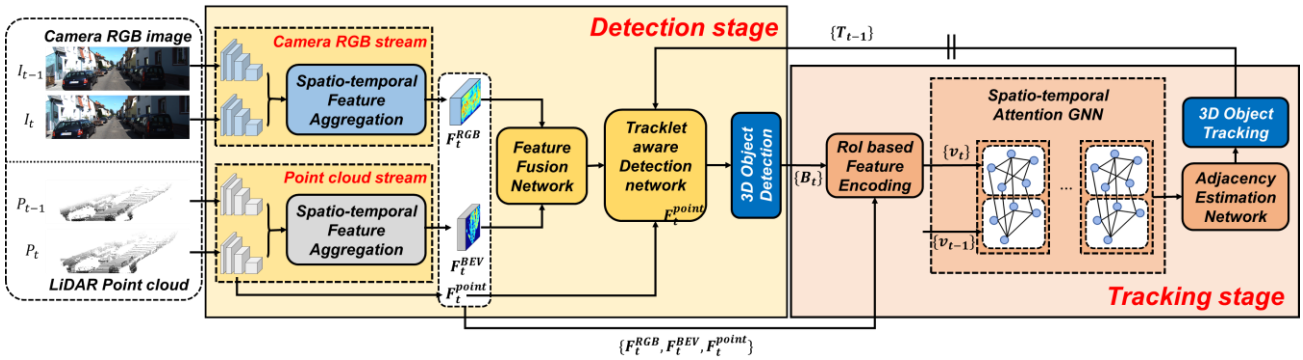

Figure 1. 제안하는 네트워크 전체 구조

진행하여 얻은 특징지도를 SG-GNN 의 입력으로 사용하였다. SG-GNN 은 rule 기반의 edge-pruning 방법과 attention 기반의 edge-gating 방법을 이용하여 GNN 을 수행하였다. rule 기반 edge-pruning 기법은 시간 t 에서의 물체와 시간 t-1 에서의 물체를 일정 거리 이하에 있는 물체들끼리 연결해주었다. Attention 기반 edge-gating 기법은 물체를 연결할 때 softmax 함수를 이용하여 연결 기여도를 판단하고 연결해주었다.

2.5) 실험

제안하는 네트워크를 실험을 하기 위해서 KITTI 물체 추적 데이터셋과 nuScenes 데이터셋을 이용하여 실험을 진행하였다.

제안하는 기법의 성능 평가는 아래 Table 1, Table 2 와 같이 나타났다. Table 1 을 통해 KITTI 물체 추적 데이터셋에서 기존의 다른 방식보다 더 높은 성능을 보이는 것을 확인할 수 있었다. Table 2 에서는 제안하는 기법이 nuScenes 데이터셋에서 baseline 대비 검출 및 추적에서 더 높은 성능을 나타내는 것을 확인하였다.

Table 1. KITTI 데이터셋을 이용한 실험 결과

Method	sAMOTA (%)	AMOTA (%)	AMOTP (%)	MOTA (%)	MOTP (%)
AB3DMOT [2]	93.28	45.43	77.41	86.24	78.43
GNN3DMOT [3]	93.68	45.27	78.10	84.70	79.03
Method [4]	92.37	44.96	76.83	84.49	78.32
Ours	96.38	48.83	81.48	91.40	82.21

Table 2. nuScenes 데이터셋을 이용한 실험한 결과

Method	mAP (%)	sAMOTA (%)	AMOTA (%)	AMOTP (%)
Baseline	49.15	39.48	8.99	24.17
Ours	52.89	45.60	11.43	27.69

Ⅳ. 결론

본논문에서는 새로운 3 차원 물체 검출 및 추적 통합 기술을 제안하였다. 연속된 시계열 데이터를 SFANet 으로 시간 축으로 융합하고 센서 융합을 진행하여 시공간 특징지도를 추출하였다. 또한 이전 시간까지의 물체 추적 결과를 이용하여 물체 검출 score 를 정제하여 3 차원 물체 검출 결과를 얻었다. 또한 SG-GNN 으로 계산 량을 줄이고 효율적으로 특징지도를 연결하여 Data association 을 효율적으로 하였다. 실험은 KITTI 와 nuScenes 에서 진행했으며, 기존 방법들 보다 더 높은 성능을 보였다.

ACKNOWLEDGMENT

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터육성지원사업의 연구결과로 수행되었음" (IITP-2021-2017-0-01637)

참 고 문 헌

[1] J. H. Yoo, et al, “3d-cvf: generating joint camera and lidar features using cross-view spatial features fusion for 3D object detection,” in ECCV 2020.

[2] X. Weng, et al, “3D multi-object tracking: a baseline and new evaluation metrics,” in arXiv preprint arXiv:1907.03961, 2020.

[3] X. Weng, et al, “GNN3DMOT: graph neural network for 3d multi-object tracking with 2d-3d multi-feature learning,” in CVPR, 2020.

[4] X. Weng, et al, “Joint 3d tracking and forecasting with graph neural network and diversity sampling,” in arXiv preprint arXiv:2003.07847, 2020.