

시간 정보를 활용한 자율주행 차량에서의 실시간 차량 주변 지도 생성 기술

신홍재, 최세환, 최준원*
한양대학교, *서울대학교

{hjshin, sehwanchoi}@spa.hanyang.ac.kr, *junwchoi@snu.ac.kr

Online HD Map Construction using Temporal Context

Hongjae Shin, Sehwan Choi, Jun Won Choi*
Hanyang Univ., *Seoul National Univ.

요약

본 논문은 실시간 차량 주변 지도 생성을 위한 새로운 시간 융합 기술을 제안하였다. 제안된 네트워크는 지도 요소가 가지는 정적인 특성을 효과적으로 활용하기 위하여 이전 비디오 프레임에서 예측된 벡터화 지도를 활용하는 시간 융합 방법을 사용하였다. 먼저, 연속된 다중 시점 비디오 프레임으로부터 예측한 벡터화 지도를 차량의 움직임 정보를 이용하여 현재 시점의 자차 중심 좌표계에 맞게 변환한다. 변환된 지도를 convolution network 를 이용하여 인코딩하고 현재 시점의 Bird's Eye View (BEV) 특징지도와 결합하여 융합된 BEV 특징지도를 생성한다. 융합된 BEV 특징지도는 BEV 인코더를 통해 강화되어 차량 주변 지도 생성에 활용되며, 기존 방법 대비 높은 성능을 달성하였다. 본 논문에서는 nuScenes 데이터셋을 이용하여 실험을 진행하였다.

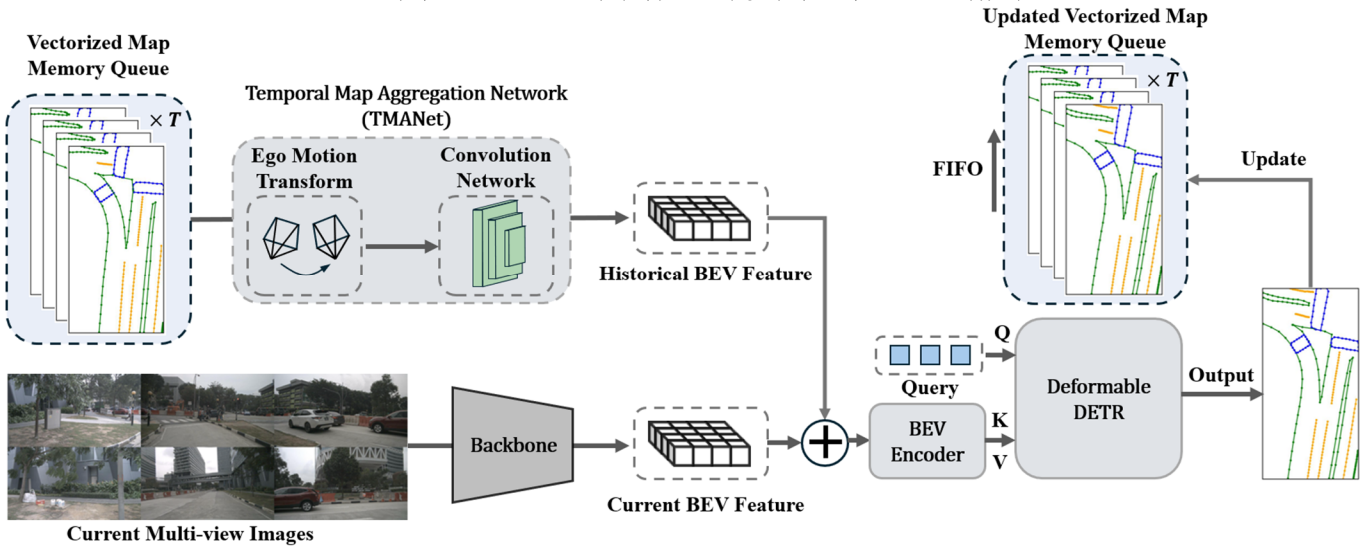


Figure 1. 제안하는 네트워크 전체 구조

I. 서론

본 논문에서는 다중 시점 카메라 기반의 실시간 차량 주변 지도 생성 기술을 위한 효율적인 시간 융합 네트워크를 제안한다. 기존 시간 융합 기술은 과거 정보로부터 동적인 객체의 현재 위치를 정확히 예측하기 어려워 이전 프레임들의 BEV 특징지도를 메모리에 저장하고 현재 시점의 BEV 특징지도와 융합하는 방법을 주로 사용한다. 지도 요소들은 동적인 객체들과 달리 시불변한 절대 좌표를 가지고 있기 때문에, 차량의 움직임만 고려하여도 지도 요소의 현재 위치를 계산할 수 있다. 따라서, 이러한 지도 요소의 정적 특성에 맞는 효과적인 시간 정보 활용 방법이 필요하다. 이를 위해서 과거에 예측된 벡터화 지도를 모델 내의 메모리에

저장하고, 현재 시점의 BEV 특징지도와 융합하여 시간 정보를 활용하도록 모델을 설계하였다. 또한, 여러 프레임의 정보를 활용하기 위해 차량을 기준으로 좌표계를 정렬하기 위한 네트워크를 설계하여 시간 정보를 효과적으로 사용할 수 있게 하였다. 이를 통해 기존 기술 대비 높은 성능의 차량 주변 지도 생성 결과를 얻을 수 있다.

II. 본론

2.1) 제안하는 시간 융합 알고리즘

제안하는 실시간 차량 주변 지도 생성을 위한 시간 융합 알고리즘은 이전 비디오 프레임에서 예측한 벡터화 지도를 활용하며, BEV 특징지도를 사용하는 기존의 시간

융합 기술보다 메모리 효율적이게 설계하였다. 이에 대한 전체 구조는 Fig.1 과 같다.

과거 융합 BEV 특징지도를 생성하기 위해서 Temporal Map Aggregation Network (TMANet)을 사용하여 메모리에 저장된 각 프레임에서의 벡터화 지도 예측을 과거 융합 BEV 특징지도로 인코딩하였다. 이후, Lift, Splat, Shoot 추출 방식[1]을 활용하여 현재 시점의 다중 시점 카메라 이미지로부터 현재 시점의 BEV 특징지도를 추출하였다. 과거 융합 BEV 특징지도와 현재 시점의 BEV 특징지도를 더해 융합된 BEV 특징지도를 생성하였고, 이 융합된 BEV 특징지도에 U-Net 네트워크[2]를 BEV encoder 로써 활용하여 강화된 BEV 특징지도를 생성하였다. 이 강화된 BEV 특징지도를 이용하여 벡터화 지도를 예측하고, 이를 메모리에 업데이트한다. 벡터화 지도를 예측하는 네트워크로써 학습 가능한 벡터를 query, 융합 BEV 특징지도를 key, value 로 하는 Deformable DETR 모델[3]을 사용하였다.

2.2) TMANet

과거 시간에서 예측된 벡터화 지도와 현재 시점의 BEV 특징지도의 융합을 위해 예측된 벡터화 지도를 과거 융합 BEV 특징지도로 인코딩하는 과정을 수행한다.

먼저, 시간 $t-1$ 까지의 벡터화 지도 예측을 현재 시점의 자차 BEV 좌표계로 변환한다. 이때, 시불변한 절대 좌표를 갖는 지도 요소의 특징을 고려하여 자차의 움직임 정보만을 활용한다. 변환된 벡터화 지도 예측 중 예측 score가 임계값 이상인 지도 요소들만 선별하여 각 지도 요소 클래스와 각 시간축으로 구분하여 분리한 벡터화 지도 예측을 convolution network 를 통해 BEV 특징지도와 같은 형태로 인코딩하여 과거 융합 BEV 특징지도를 생성한다.

2.3) 실험 결과

제안하는 네트워크를 실험 및 평가하기 위해 공개 자율주행 데이터셋인 nuScenes 데이터셋을 사용하였다.

제안하는 기법의 성능 평가는 Table 1 과 같이 나타났다. 대표적인 단일 프레임을 활용한 실시간 차량 주변 지도 생성 기술인 MapTRv2[4]와 비교하여 더 우수한 성능을 보였다. 대표적인 시간 정보를 활용한 실시간 차량 주변 지도 생성 기술인 StreamMapNet[5]에 비해 높은 성능을 달성하였다.

Table 1. nuScenes 데이터셋을 이용한 실험 결과

Method	AP_{dv} (%)	AP_{ped} (%)	AP_{bou} (%)	mAP (%)
MapTRv2[4]	61.35	60.16	62.63	61.38
StreaMapNet[5]	66.3	61.7	62.1	63.4
Ours	64.01	61.93	65.56	63.83

III. 결론

본 논문에서는 새로운 다중 시점 카메라 기반의 실시간 차량 주변 지도 생성 기술을 위한 효율적인 시간 융합 네트워크를 제안하였다. 이전 시점까지의 벡터화

지도 예측 결과를 활용하여 현재 시점의 BEV 특징지도를 효율적으로 강화하였다. 해당 알고리즘은 nuScenes 데이터셋에서 학습 및 성능 평가를 진행하였으며, 기존 알고리즘 대비 높은 성능 향상을 이루었다.

ACKNOWLEDGMENT

이 논문은 2024 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No. 2020-0-01373, 인공지능대학원 지원(한양대학교))과 2024 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No.2020R1A2C2012146)

참고 문헌

- [1] Philion, Jonah, and Sanja Fidler. "Lift, splat, shoot: Encoding images from arbitrary camera rigs by implicitly unprojecting to 3d." In European Conference on Computer Vision, 2020.
- [2] Ronneberger, O. "U-net: Convolutional networks for biomedical image segmentation." In Medical Image Computing and Computer-Assisted Intervention, 2015.
- [3] Zhu, Xizhou, et al. "Deformable DETR: Deformable Transformers for End-to-End Object Detection." In International Conference on Learning Representations, 2020.
- [4] Liao, Bencheng, et al. "Maptrv2: An end-to-end framework for online vectorized hd map construction." In arXiv:2308.05736, 2023.
- [5] Yuan, Tianyuan, et al. "Streammapnet: Streaming mapping network for vectorized online hd map construction." In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2024.