

다중 에이전트 인지를 위한 오토인코더 기반 특징 데이터 압축기법 연구

김영석, 황선욱, 박세웅
서울대학교 전기정보공학부 뉴미디어통신공동연구소

{yskim, swhwang}@netlab.snu.ac.kr, sbahk@snu.ac.kr

A Study on the Autoencoder-based Feature Data Compression Technique for Multi-Agent Perception

Youngseok Kim, Sunwook Hwang, and Saewoong Bahk
Department of Electrical and Computer Engineering, INMC, Seoul National University

요약

본 논문은 다중 에이전트 인지를 위한 특징 데이터 공유 과정에서 통신 오버헤드를 낮추기 위한 압축기법을 다룬다. 특히, 오토인코더 기반의 특징 데이터 압축 모델에 대하여 채널 차원과 공간 차원의 압축이 인지 성능에 미치는 영향을 분석하여 향후 고도화된 압축 기법의 설계에 있어서 참고해야 할 점이 있는지 확인해보고자 한다.

I. 서론

로봇, 드론, 차량 등을 포함하는 자율주행 에이전트에 있어서 주변 환경에 대한 인지과정은 빠르고 정확하게 이루어져야 한다. 그러나 단일 에이전트 내에서의 인지는 정확도 측면에서 여러가지 한계점을 지닌다. 먼저 센서의 센싱 범위가 한정적이기 때문에 먼 거리에 있는 대상은 감지할 수 없으며, 장애물로 인해 시야가 가려지게 되면 사각지대가 생길 수 있다. 또한 예기치 못한 센서의 오작동으로 인하여 정상적인 인지가 불가능해지는 상황이 발생할 수도 있다. 이러한 측면에서 최근 비전 정보 공유를 통한 다중 에이전트 인지 연구가 활발히 이루어지고 있다. [1-7]. 다중 에이전트 인지는 서로 다른 위치에 있는 에이전트들 간의 협력을 통해서 인지 과정의 정확도와 안정성을 크게 높일 수 있을 것으로 기대된다.

본 논문에서는 다중 에이전트 인지를 위한 비전 정보 공유 과정에서의 통신 오버헤드 측면을 다루고자 한다.

II. 본론

- 다중 에이전트 인지 모델의 분류

다중 에이전트 인지 모델은 공유되는 정보의 레벨에 따라 크게 세가지로 분류될 수 있다. 첫째, 입력 데이터 공유는 정보량이 많아서 가장 높은 정확도를 달성할 수 있지만, 데이터 사이즈가 크기 때문에 통신 오버헤드가 증가하게 된다. [1-2]. 둘째, 출력 데이터 공유는 데이터 사이즈가 작아서 통신에 대한 부담이 적지만, 정보 융합시 기대되는 정확도 이득 또한 제한적이다. [3-4]. 셋째, 특징 데이터 공유는 적절한 정보 융합 기법을 적용할 경우 입력 데이터 공유만큼의 정확도 이득을 달성할 수 있으면서도, 압축이 용이하다는 특성이 있어서 통신의 부담을 낮출 수 있다는 장점을 지니고 있다. [5-7]. 최근에는 정확도 이득과 데이터 사이즈 간의 트레이드 오프를

달성할 수 있는 특징 데이터 레벨의 공유에 대한 연구가 활발히 이루어지고 있다.

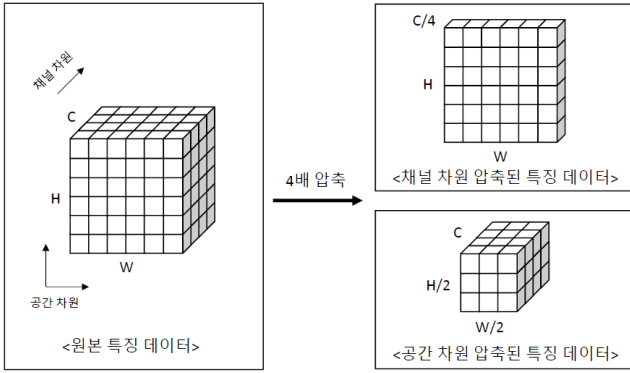
- 특징 데이터 공유 기반 다중 에이전트 인지 모델

특징 데이터 공유 기반 다중 에이전트 인지 모델은 입력 데이터로부터 특징 데이터를 추출하는 단계, 특징 데이터를 압축 및 전송하는 단계, 수신된 특징 데이터를 융합하는 단계, 융합된 특징 데이터를 기반으로 최종 출력을 도출하는 단계로 이루어져 있다. 본 논문에서는 대표적인 특징 데이터 공유 기반 다중 에이전트 인지 모델인 V2X-ViT [7] 을 사용하였다. V2X-ViT 에서는 LiDAR point cloud 데이터를 입력 받아서 3D 물체 감지를 수행하며, 특징 데이터의 융합 과정에서 Vision Transformer 모델을 사용한다.

- 오토인코더 기반 특징 데이터 압축

이미지 데이터에 대한 압축 기법으로서 가장 널리 이용되고 있는 딥러닝 모델은 오토인코더 기반의 모델이다. 오토인코더는 주어진 데이터의 차원을 감소시켰다가 다시 늘리는 방식을 통해서 압축 및 복원을 수행할 수 있다. 다만 채널 차원의 크기가 비교적 작은 이미지 데이터에 비하여 특징 데이터는 큰 차원의 채널을 갖기 때문에 오토인코더를 통해 압축을 할 때, 공간 차원 뿐만 아니라 채널 차원까지도 고려할 수 있다. 실제로 V2X-ViT 에서도 오토인코더를 통한 특징 데이터 압축을 적용하였으며 특히, 3x3 사이즈의 컨볼루션 필터를 사용하여 채널 차원에서 데이터 압축을 시도하였다.

본 논문에서는 이를 보다 확장하여 채널 차원의 압축과 공간 차원의 압축이 인지 성능에 미치는 영향을 분석해보고자 하였다. 다음 [그림 1]은 특징 데이터에 대한 공간 차원 압축과 채널 차원 압축에 대한 개념을 보여주고 있다. 먼저 채널 차원의 압축에 대해서는 공간 차원의 정보를 완전히 배제하기 위하여 1x1 사이즈의 컨볼루션 필터를 사용하였다. 반면 공간 차원의 압축에 대해서는 인코더에서 3x3 사이즈의 컨볼루션 필터에 스트라이드를 2 로 세팅하여 차원을 절반씩 줄일 수 있도록 하였고, 디코더에서 차원을 다시 늘릴 때에는 전치 컨볼루션만



[그림 1] 특징 데이터에 대한 채널 차원 압축과 공간 차원 압축.

Average Precision	압축률				
	비압축	x 4	x 16	x 64	x 256
채널 차원 (3x3) [7]	67.2	67.6	66.2	66.1	59.3
채널 차원 (1x1)		72.8	71.4	64.0	38.4
공간 차원 (전치 컨볼루션)		66.2	48.8	15.0	0.6
공간 차원 (업샘플링&컨볼루션)		59.8	34.8	2.3	0.5

[표 1] 오토인코더 별 압축률에 따른 3D 물체 감지 성능.

사용하는 방식과 업샘플링 및 컨볼루션을 함께 사용하는 방식을 적용해보았다. 다음 [표 1]은 오토인코더의 종류에 따른 3D 물체 감지 성능을 보여준다.

실험 결과, 보편적으로 압축률이 증가할수록 인지 성능이 감소하는 경향성을 확인할 수 있었다. 채널 차원의 압축과 공간 차원의 압축을 비교해보면 같은 압축률일 때, 채널 차원의 압축이 공간 차원의 압축에 비해 훨씬 더 높은 인지 성능을 보여주었다. 이는 공간 차원의 압축은 해상도를 낮춰서 물체의 형태가 일그러지는 효과가 발생할 수 있는데 비해서, 채널 차원의 압축은 해상도를 유지할 수 있기 때문인 것으로 보인다. 특징 데이터는 입력 데이터에 비해서 채널 차원이 훨씬 크다는 특성이 있기 때문에 이러한 점을 고려하여 채널 차원을 위주로 압축 모델을 설계해야 할 것이다.

채널 차원의 압축에 대해서 압축률이 낮을 때에는 1x1 사이즈의 필터를 사용한 경우가 더 높은 성능을 보였는데, 이는 채널 차원의 정보만을 활용한 압축이 보다 강한 정규화 효과를 만들어냈기 때문으로 해석할 수 있다. 3x3 사이즈의 필터를 사용한 경우는 공간 차원에서의 정보를 추가로 활용하여 높은 압축률에서도 안정적인 성능을 보였다.

공간 차원의 압축에 대해서는 전치 컨볼루션을 이용한 경우가 업샘플링 및 컨볼루션을 함께 이용한 경우보다 항상 더 높은 성능을 보였다. 이는 업샘플링 계층에서 물체의 형태가 일그러지는 현상이 추가로 발생하기 때문으로 보인다.

III. 결론

본 논문에서는 다중 에이전트 인지 시나리오에서 오토인코더 기반 특징 데이터 압축 기법의 종류에 따른 인지 성능의 변화를 확인하였다. 공간 차원의 압축의 경우, 해상도가 변하면서 물체의 형태가 보존되기 어렵기 때문에 인지 성능의 하락을 불러오기 쉽다. 반면 채널 차원의 압축의 경우, 해상도를 유지한 채 압축이 이루어지기 때문에 상대적으로 높은 인지 성능을 유지할 수 있다. 향후 고도화된 특징 데이터 압축 모델을 설계하는 데 있어서 이러한 특성을 참고할 수 있을 것이다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터육성지원사업의 연구결과로 수행되었음" (IITP-2024-2021-0-02048). 이 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2022R1A5A1027646).

참고 문헌

- [1] H. Qiu et al., "AVR: Augmented Vehicular Reality," in *Proc. ACM MobiSys*, 2018.
- [2] X. Zhang et al., "Emp: Edge-assisted multi-vehicle perception," in *Proc. ACM MobiCom*, 2021.
- [3] L. Liu and M. Gruteser, "EdgeSharing: Edge assisted real-time localization and object sharing in urban streets," in *Proc. IEEE INFOCOM*, 2021.
- [4] S. Shi et al., "VIPS: real-time perception fusion for infrastructure-assisted autonomous driving," in *Proc. ACM MobiCom*, 2022.
- [5] Y.-C. Liu et al., "Who2com: Collaborative perception via learnable handshake communication," in *Proc. IEEE ICRA*, 2020.
- [6] Y. Zhou et al., "Multi-robot collaborative perception with graph neural networks," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 2289-2296, 2022.
- [7] R. Xu et al., "V2x-ViT: Vehicle-to-everything cooperative perception with vision transformer," in *Proc. ECCV*, 2022.