

V2X 환경에서 주행 맥락 인지 기반 의미론적 마스킹을 활용한 시맨틱 통신 기법

임영훈, 오성현, 김정곤*

한국공학대학교 전자공학부

dladjwlsrh@tukorea.ac.kr, osh119@tukorea.ac.kr, jgkim@tukorea.ac.kr*

Driving Context-Aware Semantic Masking for Semantic Communications in V2X Environments

Young Hoon Lim, Sung Hyun Oh, Jeong Gon Kim*

Dept. of Electronic Engineering, Tech University of Korea

요약

ITS(Intelligent Transport Systems) 기반 V2X(Vehicle-to-Everything) 환경에서 협력형 자율주행을 위해 카메라 및 인프라 센서 정보의 공유가 요구되지만, 제한된 대역폭과 가변 채널로 인해 지연 및 손실이 발생하여 안전 메시지의 적시 전달이 어렵다. 본 논문은 주행 장면의 맥락을 반영하여 전송 전략을 능동적으로 조절하는 상황 인지 기반 적응형 시맨틱 통신 프레임워크를 제안한다. 제안 프레임워크는 SegFormer 기반 특징 추출을 통해 다중 해상도 의미 특징을 생성하고, 보행자·차량·도로 등 안전 핵심 클래스를 우선순위로 설정하여 Semantic 마스킹을 수행함으로써 중요 영역의 정보 보존을 강화한다. 이후 마스킹된 특징을 CNN(Convolutional Neural Network) 기반 Deep-JSCC(Deep Joint Source-Channel Coding) 구조의 시맨틱 인코더로 저차원 잠재 표현으로 압축하여 채널을 통해 전송하고, 수신단에서 디코더로 복원한다. nuScenes-mini 데이터셋의 전방 카메라 영상을 이용해 AWGN(Additive White Gaussian Noise) 채널에서 SNR(Signal to Noise Ratio) 변화를 적용한 시뮬레이션 결과, 낮은 SNR 환경에서도 우선순위로 지정된 핵심 클래스 영역은 상대적으로 안정적으로 복원되며, 비중요 배경 정보는 의도적으로 열화되는 경향을 확인하였다. 이를 통해 제한된 전송 자원을 안전과 직결된 의미 정보에 집중하여 통신 효율과 신뢰성을 동시에 향상할 수 있음을 보인다.

I. 서론

최근 AI(Artificial Intelligence)와 5G/6G 통신 기술의 발전으로 교통 시스템은 ITS(Intelligent Transport Systems)로 고도화되고 있으며, 차량과 인프라가 실시간으로 정보를 공유하는 협력형 자율주행은 안전성과 교통 효율 향상을 위한 핵심 기술로 자리 잡고 있다[1]. 이러한 환경에서는 차량의 LiDAR, RADAR(Radio Detection And Ranging), 고해상도 카메라 등 센서 데이터와 함께, 신호등 및 RSU(Road Side Unit)에서 생성되는 정보가 V2X(Vehicle-to-Everything) 통신을 통해 공유되어야 한다.

그러나 센서 데이터의 고용량화로 인해 제한된 대역폭에서 모든 정보를 원본 수준으로 전송하는 방식은 통신 지연과 패킷 손실을 야기하며, 이는 안전 관련 메시지의 적시 전달을 어렵게 만든다[2]. 이를 해결하기 위한 대안으로, 수신 목적에 필요한 의미 정보만 전달하는 시맨틱 통신이 주목받고 있다[3].

다만 기존 시맨틱 통신은 주행 환경과 상황 맥락을 충분히 반영하지 못한 채 고정된 의미 추출/전송 정책을 적용하는 경우가 많다. 그 결과, 도심·고속도로 등 환경이 달라져도 동일한 전송 정책이 유지되어 중요도 변화에 적응하지 못하고, 불필요한 전송으로 대역폭 활용 효율이 저하될 수 있다.

이에 본 논문은 주행 상황을 인지하여 전송 전략을 능동적으로 조절하는 상황 인지 기반 적응형 시맨틱 통신 기법을 제안한다. 제안 기법은 주행 맥락에 따라 핵심 의미 정보를 선별하고 네트워크 상태에 맞춰 전송 정책을 최적화함으로써 통신 자원을 절감하는 것을 목표로 한다.

본 논문의 구성은 다음과 같다. 2절에서는 제안 시스템 모델과 실험 결과, 3절에서는 결론 및 향후 연구 방향을 논의한다.

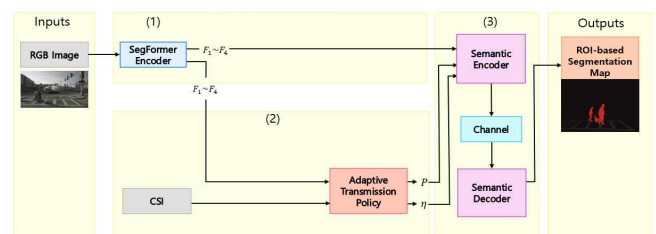
II. 본론

본 절에서는 제안하는 시스템 구조를 설명한다. 제안하는 시스템의 전체 구조는 [그림 1]과 같으며, (1) 상황 인지 모듈(Context Awareness Module), (2) 적응형 전송 정책 모듈(Adaptive Transmission Policy), 그리고 (3) 시맨틱 인코더/디코더로 구성된다.

2.1 상황 인지 모듈 (Context Awareness Module)

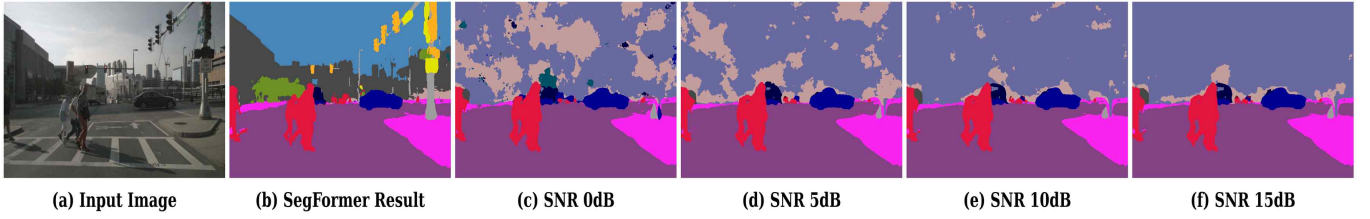
상황 인지 모듈은 차량의 전방 카메라 영상으로부터 주행 장면을 구성하는 객체와 배경의 의미 정보를 추출하고, 전역적 문맥(Global Context)을 반영한 특징을 생성하는 역할을 한다. 본 연구에서는 다양한 해상도의 특징을 효율적으로 추출하기 위해 SegFormer를 백본 네트워크로 채택하였다[4]. 입력 RGB 영상을 $I \in R^{H \times W \times 3}$ 라고 할 때, I 는 SegFormer의 계층적 Transformer Encoder를 통과하며, 네 개의 stage에서 다중 해상도 특징 맵 F_i 가 순차적으로 추출된다. 각 특징 맵은 수식 1과 같이 정의된다.

$$F_i \in R^{C_i \times \frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}}}, \quad i = 1, 2, 3, 4 \quad (1)$$



[그림 1] 전체 시스템 구성도

* : 교신저자



[그림 2] SNR 변화에 따른 세그멘테이션 결과 비교. (a) 입력 영상, (b) 정답(GT), (c)-(f) SNR 0, 5, 10, 15dB에서 복원 결과

여기서 F_i 는 SegFormer의 계층적 인코더의 i 번째 stage에서 출력되는 특징 맵을 의미하며, $R(\cdot)$ 는 F_i 가 실수 값을 갖는 텐서임을 의미한다. C_i 는 i 번째 stage의 채널 수로서 추출되는 특징의 표현 차원을 의미하며, stage가 깊어질수록 증가한다. 또한 H 와 W 는 입력 영상의 높이와 너비를 의미한다. 즉, 인코더의 stage i 가 증가할수록 특징 맵의 공간 해상도는 단계적으로 축소되고, 채널 차원은 증가함으로써 세밀한 지역 정보로부터 전역 문맥 정보까지를 포괄하는 다중 scale 표현을 형성한다.

2.2 적응형 전송 정책 모듈(Adaptive Transmission Policy)

적응형 전송 정책 모듈은 상황 인지 모듈에서 추출된 다중 스케일 특징과 현재 통신 채널 상태를 입력으로 최적의 전송 전략을 수립한다. 본 모듈은 주행 안전 중요도에 따라 클래스별 우선순위를 보행자, 차량, 도로 순으로 사전 정의하며, 이를 바탕으로 입력된 특징 맵에서 의미론적 마스킹을 수행하여 영역별 중요도를 산출한다. 이와 동시에, 현재 채널 상태를 분석하여 의미 표현을 어느 정도까지 세밀하게 보존하여 전송할지를 결정하는 적응형 제어 정보도 산출한다. 도출된 정보는 시맨틱 인코더로 전달되어 압축 및 복원 과정을 수행한다.

2.3 시맨틱 인코더 및 디코더 (Semantic Encoder and Decoder)

본 시스템의 시맨틱 인코더와 디코더는 다중 스케일 특징 맵의 효율적인 압축 및 복원을 위해 CNN(Convolutional Neural Network) 기반의 Deep-JSCC(Deep Joint Source-Channel Coding) 아키텍처로 설계되었다. 송신단의 인코더는 정책 모듈을 통해 중요도 마스킹이 적용된 특징 맵을 입력받아, 컨볼루션(Convolution) 레이어, 배치 정규화(Batch Normalization), PReLU(Parametric ReLU) 활성화 함수를 거쳐 채널 효율적인 저차원 잠재 표현(Latent Representation)으로 압축한다. 이후 송신 전력 제약을 준수하기 위해 전력 정규화를 수행한 뒤, 이를 아날로그 심볼 형태로 변조하여 무선 채널로 전송한다. 수신단의 디코더는 잡음이 혼입된 수신 신호로부터 역컨볼루션 연산을 통해 원본 특징 맵을 복원하는 역과정을 수행한다. 본 연구에서는 의미론적 마스킹의 효용성 검증에 주안점을 두어 고정 압축률을 적용하였으나, 향후 CSI(Channel State Information) 피드백을 반영한 동적 압축률 최적화 기술로 확장할 예정이다.

2.4 시뮬레이션 및 결과

제안 시스템의 성능 검증을 위해 nuScenes-mini 데이터셋의 전방 카메라(CAM_FRONT) 이미지를 사용하였으며, SegFormer-B0를 특징 추출기로 활용하였다. 학습 파라미터는 Epoch 100, Batch Size 2, Learning Rate 0.0001로 설정하였고, 무선 채널은 AWGN(Additive White Gaussian Noise) 환경을 가정하였다. 결과는 [그림 2]에서 확인할 수 있다. 주목할 점은 채널 상태의 변동과 관계없이 우선순위로 지정된 핵심 클래스들이 일관되게 안정적인 복원 성능을 유지한다는 것이다. SNR 15dB의 양호한 환경은 물론, SNR 5dB 및 0dB의 극심한 잡음 환경에서도 정책 모듈에 의해 마스킹된 보행자, 차량, 도로 영역은 형태적 왜곡 없이 명확

하게 복원되었다.

반면, 마스킹에 포함되지 않은 배경 정보는 SNR이 낮아짐에 따라 형체가 깨지거나 파편화되는 양상을 보인다. 이러한 배경의 열화는 시스템의 오류가 아니라 배경에 할당된 전송 에너지를 마스킹된 객체 영역에 전부 집중시킨 결과이다. 이는 제한된 전송 전력을 핵심 객체 복원에 집중시킨 의도된 결과로, 제안 기법이 열악한 통신 환경에서도 안전 필수 정보를 효과적으로 보존함을 입증한다.

III. 결론

본 논문에서는 대역폭이 제한적이고 통신 상태가 가변적인 V2X 환경에서 주행 안전성을 확보하기 위해, 상황 인지 기반의 적응형 시맨틱 통신 프레임워크를 제안하였다. 제안하는 시스템은 SegFormer를 백본으로 하여 주행 장면의 맥락을 파악하고, 보행자, 차량, 도로와 같은 핵심 클래스에 대해 시맨틱 마스킹을 수행함으로써 전송 우선순위를 부여하였다. 시뮬레이션 결과 제안된 시스템은 극심한 잡음 환경에서도 중요 객체 영역을 형태적 왜곡 없이 복원하여 시맨틱 통신의 효용성을 입증하였다. 이를 통해 한정된 전송 에너지를 안전과 직결된 객체에 집중시킴으로써 통신 효율과 신뢰성을 동시에 달성할 수 있음을 시사한다. 향후 연구에서는 본 논문에서 검증한 프로토타입을 기반으로, 강화학습을 도입하여 채널 상태와 주행 상황에 맞춰 압축률과 전송 정책을 스스로 최적화하는 방향으로 고도화할 예정이다.

ACKNOWLEDGMENT

이 논문은 2025년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임 (RS-2024-00415938, 2024년 산업혁신인재성장지원사업)

참 고 문 헌

- [1] ETSI, "Automotive Intelligent Transport Systems"
- [2] Nguyen, L. X., Raha, A. D., Aung, P. S., Niyato, D., Han, Z., and Hong, C. S. "A Contemporary Survey on Semantic Communications: Theory of Mind, Generative AI, and Deep Joint Source-Channel Coding," arXiv preprint arXiv:2502.16468 [cs.CV], pp. 1-30, February 2025.
- [3] J. M. Gimenez-Guzman, I. Leyva-Mayorga, and P. Popovski, "Semantic V2X Communications for Image Transmission in 6G Systems," IEEE Network, vol. 38, no. 6, pp. 48 - 54, Nov./Dec. 2024.
- [4] E. Xie et al., "SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers," Adv. Neural Inf. Process. Syst., vol. 34, pp. 12077 - 12090, Dec. 2021.