

스테레오 비전 및 End-to-End Learning 기반 자율주행 시스템

윤예중¹, 송지환², 변형섭³, 박배성⁴, 김종현⁵
성균관대학교 기계공학부

dbdsb2222@gmail.com, wlghksdl33@g.skku.edu, qusgudtjq123@naver.com, pbs1604@naver.com,
jonghyunkim@skku.edu

An Autonomous Driving System Based on Stereo-Vision and End-to-End Learning

Ye-Joong Yoon¹, Ji-Hwan Song², Hyeong-Seob Byeon³, Bae-Seong Park⁴, Jong-hyun Kim⁵
Dept. of Mechanical Engineering, Sungkyunkwan University

요 약

자율주행 기술에서 스테레오 비전과 End-to-End Driving은 많이 사용되는 기술이며 본 연구에서는 이를 신호등 인식과 주행에 적용하였다. 신호등 인식은 좌우 카메라로부터 적색 원을 인식한 후 스테레오 비전을 통해 신호등과의 거리를 추정한다. 주행 시스템은 End-to-End Learning 기반으로 이루어지며, 출력값인 가변저항을 조향각으로 변환하여 제어할 수 있다. 또한 감마 보정을 통한 데이터 증강을 통해 빛에 대해 민감하지 않게 모델을 학습하였다. 추후 신호등 인식 시 HSV 필터가 빛에 민감한 점과 주행 시 가변저항 값이 일정하지 않은 점이 해결된다면 더욱 안정적인 시스템을 구축할 수 있을 것으로 기대된다.

1. 서론

스테레오 비전 및 End-to-End Driving은 자율주행 분야에서 장애물 인식 및 도로주행에 사용되는 기술이다[1]. 도로 상황에서 신호등의 인식과 거리를 추정하는 것은 매우 중요한 작업이다. 스테레오 카메라는 LiDAR에 비해 가격이 저렴하고, End-to-End Driving은 기존 방식에 비해 더 작은 딥러닝 네트워크를 사용할 수 있다는 장점이 있다[2]. 본 논문에서는 해당 기술을 활용하여 신호등 인식 및 주행에 적용하였다. 이에 두 개의 카메라로부터 신호등을 인식하여 거리를 추정하고 조향각을 예측하여, 시리얼 통신을 통해 제어 명령을 내릴 수 있는 자율주행 시스템을 제안한다.

2. Stereo Vision 기반 신호등 인식

스테레오 비전은 2개의 카메라로 거리(depth)를 추정하는 기술로, 본 논문의 자율주행 시스템에서는 적색 신호등을 인식하여 특정 거리에 도달했을 때 차량이 정지하도록 설계하였다. 신호등과의 거리는 두 카메라 간 특징점 사이의 거리인 시차(disparity)를 통해 계산된다. 본 연구에서는 스테레오 카메라 대신

ABKO사의 APC920W 웹캠 2개를 가로로 붙여 사용하였다. 카메라 내부 파라미터인 베이스라인(baseline, b), 초점 거리(focal length, f), 시차(disparity, d)를 알면 다음과 같은 식으로 거리(depth, z)를 구할 수 있다.

$$z = \frac{b \cdot f}{d}$$

신호등 인식은 Fig. 1과 같은 과정으로 진행된다.

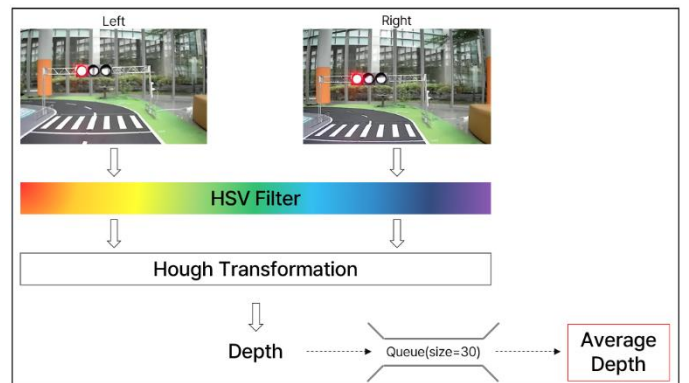


Fig. 1. System Blueprint of Traffic Light Recognition

Table. 1. Accuracy of End-to-End Learning Model

데이터셋	정확도
A2D2	92.16%
직접 수집	87.64%

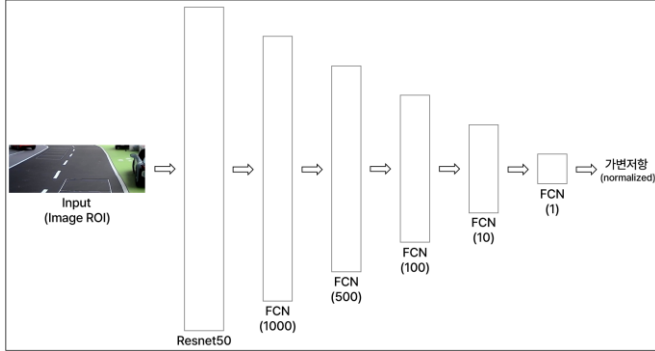


Fig. 2. Architecture of End-to-End Driving Model

우선 HSV 필터를 통해 빨간색 영역을 ROI(Region of Interest)로 지정하고 ROI에 대해 허프 변환을 통해 원을 검출해 신호등으로 인식한다. 좌우 카메라에서 각각 검출된 원의 중심점의 차이를 시차(disparity)로 하여 신호등과의 거리를 추정한다. 또한 안정적인 거리(depth)추정을 위해 사이즈가 30인 큐(queue)를 사용하여 매 프레임마다 구해지는 거리를 인큐(enqueue)하여 30개 거리의 평균값을 최종 거리로 사용하였다.

3. End-to-End Learning 기반 주행

End-to-End Driving은 기존 인지-판단-제어 방식을 하나로 합쳐 이미지로부터 바로 조향각을 예측하는 모델이다[2]. 본 논문에서는 차량의 자율주행을 위하여 End-to-End Learning기법을 사용하였으며, 차량의 제어를 위해 정밀 각도가 아닌 가변저항을 모델의 출력으로 학습하였다. Fig. 2는 본 논문에서 제시하는 End-to-End Driving 모델의 아키텍처이다. 왼쪽 카메라의 프레임이 모델의 입력으로 들어가며, ImageNet 데이터셋으로 사전학습된 ResNet50 모델을 통해 특징추출을 수행한다[3]. 특징추출 이후의 FCN(Fully Connected Layer) 부분은 층이 깊어질수록 뉴런의 개수가 작아지도록 하였으며[2], 여러 조합 중 가장 정확도가 높은 조합으로 사용하였다. 각 층 사이에는 배치 정규화 층과 활성화 함수로 LeakyReLU가 적용되었으며, 마지막 층 이후에는 TanH 함수가 적용되었다. 학습 이미지는 하단부 50%만 ROI로 사용하기 위해 상단부를 마스킹 해주었으며, 감마 보정을 적용하여 2배로 데이터 증강을 함으로써 빛에 대해 강건해지게 하였다. 모델의 성능평가는 오픈 데이터셋인 A2D2와 직접 수집한 데이터셋에 대해 정확도를 평가하여 비교하였으며, 그 결과를 Table. 1에 나타내었다.

4. 데이터 수집 및 제어 시스템

End-to-End Driving을 구현하기 위해 두 가지 제어 시스템이 고안되었다. 첫 번째는 학습 데이터 수집을 위한 차량 원격 제어 시스템으로, 키보드 방향키의 입력 시 시리얼 통신을 통해 아두이노로 신호를 전송한다. 아두이노는 입력 받은 방향키에 대응하는 조향 명령을 기어모터에 전달하고, 이 때 기어모터에 측정되는 가변저항 값을 데이터로 저장한다. 동시에 왼쪽 카메라에서 이와 대응하는 프레임을 저장하면 학습에 필요한 데이터셋이 생성된다. 이렇게 수집된 데이터로 학습이 완료된 모델은 입력이미지로부터 가변저항 값을 나옴데, 이 출력값에 맞게 차량을 조향하기 위해 두 번째 제어 시스템이 사용된다. 이 시스템은 앞선 원격 제어 시스템과 동일한 구성을 가지며, End-to-End Driving 모델의 출력값을 아두이노로 전송한다. 아두이노는 이 값을 현재의 조향각과 비교하고, 필요한 조향을 계산하여 기어 모터에 명령을 전달한다. 이러한 일련의 과정을 통해 조향각을 제어할 수 있다.

5. 결론

스테레오 비전 기반 신호등 인식 모델에서는 30개 프레임의 거리 평균값을 사용함으로써 안정적인 거리를 계산할 수 있었다. 하지만 신호등 인식 시 HSV 필터가 빛에 민감하다는 한계점이 있었다. 이는 추후 딥러닝 모델을 적용하여 정확도를 높일 수 있을 것이라 예상된다. 다음으로 End-to-End Learning 기반 주행 모델에서는 원격 제어 시스템으로 수집한 데이터를 바탕으로 학습하여 차선을 잘 주행하였다. 수집한 데이터셋에 대한 정확도는 약 87.64%로, 오픈 데이터셋에 비해 낮은 정확도를 보였다. 이는 데이터 수집 시 가변저항 측정 과정에서 발생하는 노이즈로 인해 값이 일정하지 않은 문제점으로 인한 것이며, 추후 연구를 통해 개선되어야 할 부분이라 생각한다.

ACKNOWLEDGMENT

이 논문은 정부(교육부-산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임(P0022098, 2023 년 미래형자동차 기술융합 혁신인재양성사업)

참고문헌

- [1] Fan, Rui, et al. "Computer stereo vision for autonomous driving." arXiv:2012.03194, 2020
- [2] Bojarski, Mariusz, et al. "End to end learning for self-driving cars." arXiv:1604.07316, 2016
- [3] Tampuu, Ardi, et al. "A survey of end-to-end driving: Architectures and training methods." IEEE Transactions on Neural Networks and Learning Systems 33.4, 2020