

자원이 제한된 자율주행 단말의 실시간 지능형 서비스를 위한 엣지 컴퓨팅 기반 이미지 처리 및 분석 시스템 개발

홍승준, 최인훈, 박수용, 정홍주, 김서현, 허의남*
경희대학교

hongsj1022@khu.ac.kr, inhun321@khu.ac.kr, clapd10@khu.ac.kr,
sub06038@khu.ac.kr, seok02h@khu.ac.kr, *johnhuh@khu.ac.kr

Development of edge computing-based image processing and analysis system for real-time intelligent service by autonomous vehicle with limited resources

Seungjun Hong, Inhun Choi, Suyong Bahk,
Hong-Ju Jeong, Seo-Hyun Kim, Eui-Nam Huh*
Kyung Hee Univ.

요 약

최근 자율주행을 이용한 지능형 서비스의 발전이 큰 주목을 받고 있다. 하지만 자율주행 단말의 제한된 자원으로 인해 이미지 분석을 이용한 다양하고 섬세한 서비스 제공에는 한계를 드러내고 있다. 본 논문은 이런 문제를 해결하고 실시간 지능형 서비스를 제공하기 위해 엣지 컴퓨팅을 활용한 실시간 이미지 처리 및 분석 시스템을 제안하고 평가한다.

I. 서 론

최근 자율주행 분야는 센서와 컴퓨팅 기술의 발전으로 빠르게 진화하고 있다[1]. 그 중 합성곱 신경망(CNN, Convolutional Neural Network)을 이용한 이미지 분석 기술은 주변 상황에 대한 인지와 판단을 담당하며, 자율주행의 큰 축을 담당한다[2].

하지만 자율주행 단말의 제어를 담당하는 보드는 컴퓨팅 자원이 제한적이기 때문에 이미지 분석에 대한 온전한 성능을 낼 수 없다. 이로 인해 생기는 지연시간은 실시간으로 주행을 제어하고 지능형 서비스를 제공함에 있어서 큰 문제를 불러 일으킨다.

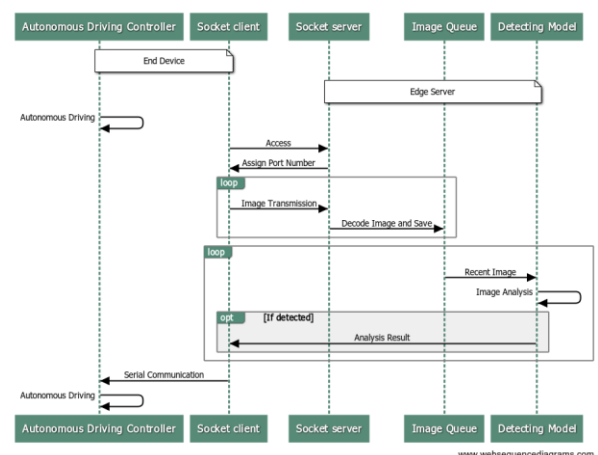
본 논문에서는 이러한 문제를 해결하기 위해 자율주행 단말과 엣지 서버간 통신을 통해 실시간으로 이미지를 처리하고 분석하는 시스템을 구현하고 평가한다.

II. 본론

자율주행 단말과 같은 이동성 단말은 네트워크 제공과 서버와의 통신 방법에 대한 제한이 생긴다. 지속적인 이동으로 고정된 네트워크를 제공할 수 없으며, 주행 경로에 설치된 무선 네트워크를 변경하며 사용할 경우 단말의 IP 주소가 지속적으로 변경되어 서버에서 단말로의 데이터 전송이 불가능하다. 또한, 네트워크가 변경되는 시간에 서비스의 실시간성이 보장되지 못한다.

따라서 본 논문에서는 고정된 엣지 서버의 네트워크 주소를 이용하여 데이터를 주고받는 소켓 통신을 사용하였다. 단말의 네트워크는 스마트폰을 이용한 무선 네트워

크 제공으로 고정되게 하였으며, 단말의 요청에 따라 할당된 서버의 소켓으로 데이터를 전송한다. 이는 단일 자율주행 단말이 아닌 다중 단말을 이용한 지능형 서비스 제공의 경우에도 통신이 가능한 확장성을 지닌다.



[그림 1] 이미지 분석 시퀀스 다이어그램

그림 1은 자율주행 단말과 엣지 서버의 통신을 이용한 실시간 이미지 분석 시나리오를 나타내는 시퀀스 다이어그램이다. STM32 보드 제어를 통해 자율주행차가 주행을 시작하면 단말에서 서버로 소켓 연결 요청을 보낸다. 요청을 받은 서버는 소켓을 할당하고 단말로부터 주행 중 촬영되는 이미지를 해당 소켓으로 전송받는다. 전송받은 이미지는 디코딩 후에 큐에 저장되며 이는 주행이 끝날 때까지 반복된다. 이후 큐에 저장된 최근 이미지를 불러

와 학습된 모델을 사용하여 분석하고 인식 결과가 도출되면 분석을 멈춘다. 인식된 결과는 단말에 전송되고 STM32 보드와의 시리얼 통신을 통해 자율주행차를 제어한다.

이미지를 큐에 저장할 때는 엣지 서버의 디스크 용량을 고려하여 동일한 이름의 파일로 저장하고 학습 모델은 가장 최근 이미지를 분석하게 된다. 이 때, 이미지를 큐에 저장하는 프로세스와 큐에서 이미지를 가져와 분석하는 프로세스는 별도로 동작하기 때문에 두 동작이 겹치는 경우를 대비해 예외처리를 하였다.

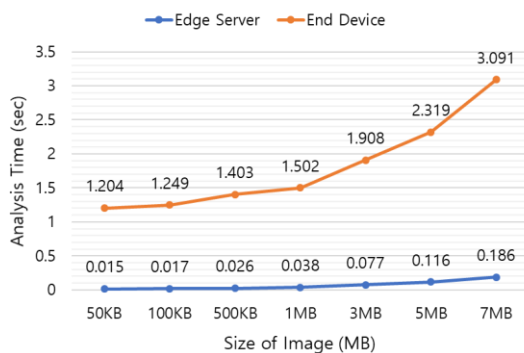
■ 이미지 분석시간 비교 및 평가

본 논문에서는 자율주행 단말로 Raspberry Pi 4B 4GB 모델을 사용하였으며, 엣지 서버는 Ubuntu 18.04.2 LTS Server 를 설치하였다. 그림 2 는 자율주행 단말과 엣지 서버의 상세 스펙을 나타낸다.

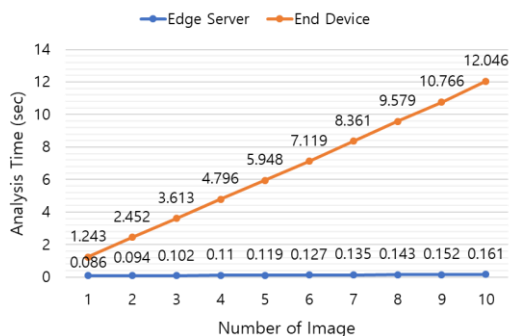
	End Device	Edge Server
CPU	ARM Cortex-A72 1.4GHz	Intel(R) Core(TM) i5-8400 CPU @ 2.80GHz
GPU	Broadcom VideoCore IV	NVIDIA GeForce GTX 1050Ti
RAM	4GB	16GB
DISK	128GB	120GB
OS	Raspbian GNU/Linux 10 (buster)	Ubuntu 18.04.2 LTS Server

[그림 2] 단말과 서버의 상세 스펙

이미지 분석에 사용된 학습 모델은 yolo-v3 tiny 모델을 사용하여 학습시킨 캔 음료 인식 모델이며[3], 단말에서 서버로 이미지를 전송하는 데 걸린 시간은 이미지당 약 0.02 초가 소요되었다.



(A) 이미지 크기에 따른 분석 시간 비교



(B) 이미지 개수에 따른 누적 분석 시간 비교

[그림 3] 자율주행 단말과 엣지 서버의 이미지 분석시간 비교 그래프

그림 3 의 (A)는 픽셀을 제외한 모든 조건이 동일한 크기 별 이미지를 10 회 분석 후 평균시간을 측정한 그래프이다. 이미지의 크기가 증가함에 따라 평균 분석시간 또한 증가함을 관찰할 수 있으며, 엣지 서버에 비해 단말의 분석시간이 더 큰 증가율을 보인 결과, 7MB 를 기준으로 엣지 서버가 단말에 비해 약 94% 빠른 것을 알 수 있다. 이는 더 높은 정확도의 분석과 서비스 제공을 위해 이미지의 해상도를 높일 경우 더욱 긴 지연시간의 발생으로 이어진다.

그림 3 의 (B)는 50KB 크기의 이미지를 기준으로 누적 분석시간을 나타내며, 분석량과 비례하여 소요시간 또한 증가함을 알 수 있다. 이는 10 장의 이미지를 기준으로 단말이 약 98% 빠른 성능을 보이며, 결국 단말의 처리량이 많아질수록 지연시간으로 인한 지능형 서비스를 제공에 치명적일 것으로 평가된다.

III. 결론 및 향후 연구 방향

본 논문에서는 자원이 제한된 단말에서 이미지를 직접 분석하지 않고 엣지 서버로 전송하여 처리 및 분석 후 결과를 수신하는 시스템을 구현 및 평가하였다.

이는 통신 시간과 이미지 분석시간을 모두 고려하여 비교하더라도 엣지 컴퓨팅을 이용한 분석이 실시간성 보장에 더 효과적이라는 것을 알 수 있다. 이는 본 논문에서 사용한 yolo-v3 tiny 모델과 같이 경량화된 학습 모델보다 신경망의 레이어가 추가될 경우에는 더 큰 효과를 기대할 수 있다.

이 시스템은 향후 시각 장애인을 위한 길 안내 서비스 개발에 사용될 것이며 길 안내 서비스에 적합한 모델을 생성하고 실시간으로 분석 후 자율주행차를 제어하는 시스템으로 개발할 예정이다. 또한 사용자의 증가로 다중 자율주행 단말에 지능형 서비스를 제공해야하는 경우를 고려하여 서비스를 확장해 나갈 예정이다.

ACKNOWLEDGMENT

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2017-0-00294, (대학 ICT 기초연구실) 서비스 이동 지원을 위한 분산형 클라우드 핵심원천기술 연구)

참 고 문 헌

- [1] Vasu Singh, Siva Kumar Sastry Hari, Timothy Tsai, Mandar Pitale; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2021, pp. 122-128
- [2] 이정민. "딥러닝을 이용한 영상 기반 자율 주행 자동차 시스템." 국내석사학위논문 세명대학교 대학원, 2018. 충청북도
- [3] In-Hun Choi, "Robot Control System with Deep Learning based Beverage Image Recognition", 1st Korea Artificial Intelligenc Conference, pp. 94-95, 2020.