

컴퓨팅 계산 오프로딩 위해 ROS 를 사용한 딥러닝 기반의 자율주행카트

한지수, 박지윤, 김채원, 박상수, 김현수

이화여자대학교 컴퓨터공학과

{han_js, jiyoonpark13, cwkim0314}@ewhain.net, {sangsoo.park, hieonn}@ewha.ac.kr

Deep Learning Based Autonomous-Driving Cart Using ROS for Computation Offloading

Jisu Han, Ji-Yoon Park, Chae-won Kim, Sang-soo Park, Hieonn Kim
Dept. of Computer Science and Engineering, Ewha Womans University

요 약

IoT 와 인공지능을 접하려는 시도는 최근 들어서 많은 발전을 보이고 있다. 본 논문은 컴퓨팅 파워가 제한되는 작은 디바이스 IoT 의 한계를 극복하기 위하여 ROS 를 이용하여 복잡한 연산을 무선 통신으로 오프로딩하는 기법을 제안한다. 제안된 자율주행카트 시스템은 카트 이용 고객 개개인을 검출하고 추적하되 컴퓨터 비전 알고리즘과 LiDAR 센서를 이용하며, 음성인식 알고리즘을 적용하여 기계와 인간의 감성공학적 소통이 가능한 융합형 자율주행카트를 구현한다.

1. 서론

시대가 변함에 따라 인공지능 분야의 자율주행 기술이 주목받고 있으며, 작은 디바이스에 자율주행 기술을 접목하는 시도가 증가하고 있다. 본 논문은 인간 생활에 필수적으로 자리 잡고 있는 쇼핑 환경의 개선 뿐만 아니라 감소추이를 보이고 있는 대형 마트의 오프라인 판매에도 기여할 것으로 기대된다.

본 논문에서는 로봇틱스용 운영체제인 ROS[1]를 이용해, CPU, GPU 등을 탑재한 컴퓨터를 쇼핑 카트와 무선으로 연결한다. 그리고 객체인식과 추적에 높은 정확도와 적은 비용으로 고효율의 시스템을 제안한다. YOLOv4(You Only Look Once)[2]와 DeepSORT(Deep Simple Real time Tracker)[3]를 복합적으로 적용한 자율주행카트 ‘미미’는 카트 이용 고객을 높은 정확도로 인식하고 추적한다. 카트의 카메라가 다른 물체나 사람에 의해 가려지는 예외 상황에는 키워드 검출 기법(Keyword Spotting)[4]을 이용해 고객을 빠른 시간 내에 재검출하게 한다. 이를 위해 카트에 탑재된 IoT 장치는 ROS 네트워크로 연결된 컴퓨터 사이에 메시지 교환을 통해 주행한다.

본 논문에서 제안하는 인간과 기계의 음성 소통과 기

계가 사람을 따라 주행하는 등 감성공학시스템의 개발은 유비쿼터스 IoT 를 제공한다. 또한 휠체어나 유모차 사용자도 편리한 쇼핑을 할 수 있도록하여 다양한 사람과 분야에 무한한 확장이 가능할 것으로 예상된다.

2. 관련 연구

컴퓨팅 파워가 작은 IoT 장치에서 객체인식과 추적을 효율적으로 구현하기 위한 관련 연구는 다음과 같다. 실시간 객체 검출법 YOLO는 인간의 시각처럼 한번의 촬영으로 객체를 검출한다. DeepSORT를 객체 검출법과 함께 사용하면 정확한 객체 추적이 가능하다. 음성인식 기술 Keyword Spotting은 특정 단어에 대해 반응하는 기술로 GRU (Gated Recurrent Unit) 신경망을 사용하여 빠르고 정확한 음성인식을 가능하게 한다. ROS (Robot Operating System)는 다양한 로봇의 조작을 범용적으로 가능하게 한 오픈소프트웨어 프레임워크이다. 레이저를 이용하는 LiDAR (Light Detection And Ranging) 센서는 주변환경을 감지하며, SLAM 기술[5]로 주변 환경에 대한 3차원 지도를 생성할 수 있다.

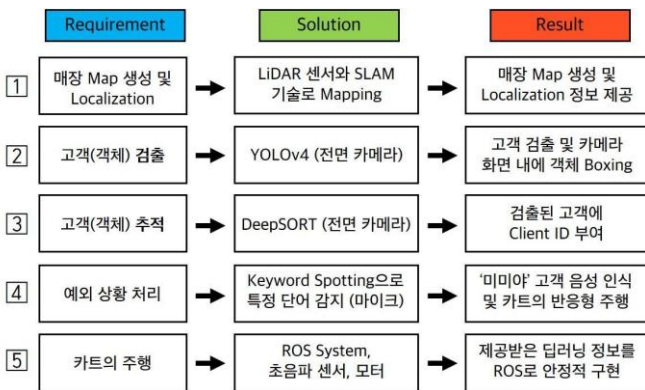
3. 시스템 구성도

본 논문에서 개발한 카트의 하드웨어의 구성요소는 (그림 1)과 같다. 자율주행카트는 마트 전체 3차원 지도를 구성하는 LiDAR 센서와 고객 추적에 사용되는 전면 카메라, 고객의 음성 응답을 처리하기 위한 음성 인식 모델에 필요한 마이크를 탑재하고 있다. 카트 간 혹은 카트와 고객, 카트와 진열대 등과의 충돌 등 긴급 상황 탐지에 필요한 초음파 센서와 카트 구동에 필요한 모터, 이를 운영하기 위한 라즈베리 파이를 탑재한다.



(그림 1) 카트의 구성 요소

카트가 추적하는 특정 고객을 클라이언트라 할 때 (그림 2)는 클라이언트 검출 및 추적 상황에서의 요구 사항과 이용되는 기술을 정리한 도표이다.



(그림 2) 클라이언트 검출/추적 요구사항

카트는 카메라로 수집한 정보를 YOLO 모델을 통해 분석하여 클라이언트를 검출한다. 카트는 클라이언트를 검출한 후, DeepSORT 를 이용하여 검출한 클라이언트를 추적한다.

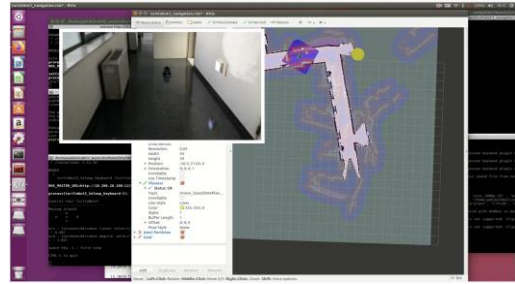
클라이언트가 카메라를 벗어나거나 카트가 다른 물체와 충돌하는 것을 대비하여 음성인식 Keyword Spotting 모델과 초음파 센서를 적용함으로써 정확하고 빠른 동작과 더불어 예외 상황도 처리함으로써 안전성을 확보한다.

4. 설계 및 구현

4.1 LiDAR 센서의 SLAM 기능을 통한 매장 지도 구성

카트는 서비스를 제공하기 전 매장의 위치 정보를 파악해야 한다. 이를 위해, 서비스를 제공하기 전 GMapping 기술을 이용해 매장 지도를 생성한다. 이를 통해 카트는 매장 전체의 지도를 가지게 되고, 카트는 매장에서 현재 위치를 파악할 수 있다. 해당 기술은

ROS 기반의 SLAM 기술을 이용하여 구현하였으며 생성된 지도의 예시는 (그림 3)와 같다.



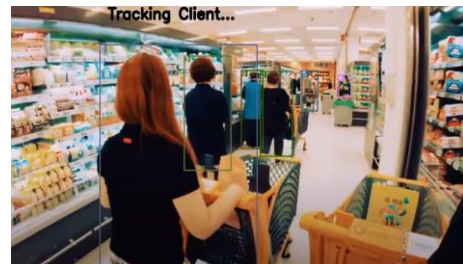
(그림 3) SLAM 을 구현한 GMapping 예

정밀도가 떨어지는 실내 측위 기술에 비해 SLAM 기술을 통한 정확한 구역화 기법으로 매장 안에서의 정확한 위치를 확인할 수 있다

4.2 객체 탐지 기법(Object Detection)에 의한 고객 검출

고객 검출시 카트는 전면 카메라를 통해 객체를 탐색한다. 전면 카메라에서 제공하는 이미지를 이용한 객체 탐지 기술을 통해 고객을 검출하게 된다.

(그림 4)는 전면 카메라를 통해 고객을 검출하는 예이다.



(그림 4) 전면 카메라를 통한 고객 검출

해당 기능은 YOLOv4 로 구현하였으며 기존의 영역 제안 방식(Region Proposal)과 분류(Classification)가 순차적으로 이루어지는 2 단계의 탐지 기법과는 달리, YOLO 계열 탐지 기법은 이러한 과정이 동시에 이루어지는 1 단계의 탐지 기법으로 빠른 성능을 나타낸다.

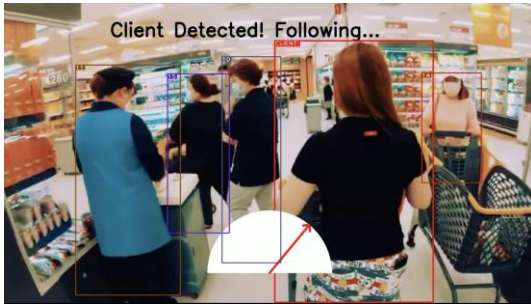
본 논문에서는 YOLOv4 의 선 트레이닝된 모델을 사용하며, 처리 속도를 빠르게 하기 위해 가장 가벼운 YOLOv4s weight 모델을 채택하였다.

4.3 객체 추적 기법(Object Tracking)에 의한 클라이언트 검출 및 추적

마트의 고객들을 다른 물체와 분리하여 검출한 후에는 해당 카트만의 클라이언트를 검출하고 추적하는 과정이 필요하다. 앞선 객체 검출 과정에서 전면 카메라에 의해 검출된 객체 중 가장 높은 빈도로 검출되는 고객을 객체 추적에 적용할 클라이언트로 지정하였다.

저장된 클라이언트 정보를 통해 카트는 검출된 클라이언트를 실시간으로 추적한다. 전면 카메라 화면 속 클라이언트의 좌표와 전면 카메라 화면의 중앙을 기점으로 클라이언트의 위치 정보에 의해 생성된 방향 벡터가 결정된다. 이를 모터에게 전달하여 하드웨어의 이동방향이 결정된다.

(그림 5)은 검출된 클라이언트를 추적하며 진행 방향을 카트가 직접 결정하는 실시간 상황의 예이다.



(그림 5) 검출된 클라이언트를 추적하는 실시간 상황

본 논문에서 적용된 DeepSORT는 위 객체 탐지에서 검출된 객체 정보를 바탕으로 객체들을 추적하는 객체 추적 기법 중 하나이다. 이는 SORT[6] 필터에 딥러닝을 적용한 기술로 영상 전체를 기록해 사물을 분석하는 것이 아닌, 앞뒤의 프레임 분석을 통해 사물을 구별하는 기법이다. 이는 앞뒤 프레임의 광학 흐름을 이용하여 가능해진다. 관찰자와 장면 사이의 상대적인 움직임을 포착하여 추적한다.

본 논문에서는 DeepSORT도 선 트레이닝된 모델을 사용해 빠른 처리 속도를 보인다. 본 모델은 위의 YOLOv4 모델과 연결되어 실행되고, 뒤의 Keyword Spotting 모델과 동시에 실행되어야 한다. 이는 멀티 스레드(Multi-thread) 기법으로 구현한다. 각 모델마다 다른 스레드에서 작동되도록 구현하여 두 모델 다 실시간에서 빠른 처리 속도를 보인다.

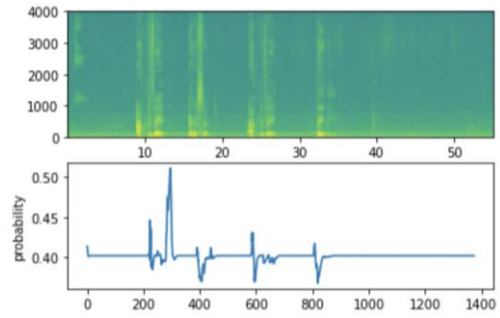
4.4 예외 상황에서의 재검출

정확도 높은 객체 검출과 추적에도 안정성을 위해서는 적은 확률의 예외도 고려되어야 한다. 본 논문에서는 몇가지 예외 상황을 예상하고 해결책을 제시한다.

첫째, 클라이언트가 화면에서 검출되지 않으면 카트는 추적을 멈추고 재검출의 피드백을 진행한 후 다시 추적하면서 자율주행을 진행한다.

둘째, 카트가 마트의 다른 고객과 클라이언트를 혼동하는 경우가 있다. 사람이 서로 이름을 불러 소통하는 것에서 착안하여, 클라이언트가 카트의 이름을 부르고 Keyword Spotting을 통해 카트가 클라이언트를 재검출하고 추적한다.

(그림 6)은 테스트 데이터 중 하나를 적용한 예시이다.



(그림 6) Keyword Spotting 모델 적용 결과

본 논문에서는 잡음이 있는 실제 환경에서의 Keyword Spotting을 위해 세 종류의 데이터를 가공하여 훈련 데이터를 활용했다. 본 논문에서 개발한 자율주행 카트의 이름인 미미, 즉 키워드에 해당하는 퍼지티브 데이터(Positive Data), 키워드 외의 네거티브 데이터(Negative Data), 잡음 역할의 백그라운드 음성 데이터(Background Data)을 준비하여 백그라운드 음성 데이터 속에 퍼지티브 데이터, 네거티브 데이터를 무작위로 삽입하여 잡음이 섞인 훈련데이터 만들고 학습되었다.

해당 모델은 자연어 처리 모델인 LSTM을 단순화한 GRU 신경망을 사용해 적은 데이터를 빠른 속도로 처리하는 효율성을 강조했다. 그 결과 92%의 정확도로 키워드 음성에 의해 카트의 반응을 도출했다.

4.5 하드웨어 구성 및 설계

4.5.1 자율주행 카트

카트의 메인 컴퓨터는 Raspberry-pi4(라즈베리 파이) 모델로 4GB의 램(RAM)을 가진 모델이다. 실시간으로 ROS 메시지를 받아 각 센서들에게 명령을 주기엔 충분하다. 실제 카트에서는 LiDAR 센서를 이용한 거리 탐지가 이루어진다. 라즈베리 파이 모델에서는 HC-SR04 초음파 센서를 이용하여 이를 대신한다.

본 논문은 LiDAR 센서가 장착된 터틀봇 버거(turtlebot burger) 모델을 사용한다. 고객 인식을 하기 위한 카메라와 음성 인식을 위한 마이크를 추가한 터틀봇을 사용한다.

(그림 7)은 실제 상황에 사용된 터틀봇이다.

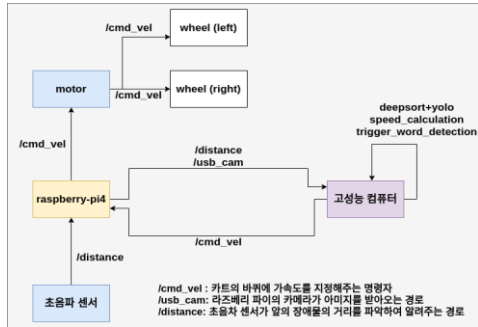


(그림 7) 카트 하드웨어 완성품

4.5.2 ROS 와의 연결

앞서 말한 객체 추적 모델과 Keyword Spotting 모델은 메인 컴퓨터에서 작동된다. 같은 네트워크 내에서 작동하는 시스템으로, ROS 기반의 시스템을 활용하여 네트워크 간의 메시지 전달 방식을 활용한다.

(그림 8)은 카트 내 외부의 통신 토픽(topic) 흐름이다. 높은 컴퓨팅 파워가 요구되는 프로세스를 요구하는 딥러닝 모델들은 컴퓨터 본체에서 작동한다. 라즈베리 파이를 사용하는 카트는 주행 명령과 입력을 받아오는 작업을 수행한다.



(그림 8) 카트 통신 topic 흐름도

라즈베리 파이가 전면 카메라를 통해 실시간 이미지를 가져오기 위해 비동기식 웹 서버 (Asynchronous web server)를 이용한다. ROS 메시지로 이미지를 가져올 경우 OpenCV 사용으로 변환이 필요하지만, 웹 서버를 이용하면 불필요한 변환을 줄인다. 각 센서에서 받아온 속도와 급정지 파라미터를 이용해, 이를 추가한 새로운 ROS 메시지 타입을 정의해 속도를 조정한다.

4.5.3 카트의 주행방법

카트는 전면 카메라에서 받아온 이미지를 바탕으로 딥러닝 처리 후 탐색 대상의 방향 벡터를 계산하여 이동 방향을 결정한다. 카트가 클라이언트를 잃어버렸을 경우, 팬틸트 (pan tilt) 기능을 사용해 360도 클라이언트 재탐색을 수행한다. 클라이언트가 탐색된다면 전면 카메라가 향해있는 방향으로 이동하여 주행을 재개한다.

이와 더불어 카트가 클라이언트 추적 도중에 다른 물체와 충돌하는 예외상황이 있다. 이를 위해 카트는 초음파 센서를 통해 실시간으로 카트와 다른 물체의 사이 거리를 감지하도록 구현하였다.

5. 결론

본 논문은 ROS 네트워크를 통해 고성능 딥러닝 모델이 연결 가능한 고효율 시스템을 제안한다. Object Tracking 모델은 정확도 92%를 보이며 일반적인 상황에서 매우 우수하게 작동됨을 볼 수 있었다. Keyword Spotting 모델은 86%의 정확도를 보이며 일상생활에서 적용 가능성을 보인다.

현재 IoT 서비스들은 대개 컴퓨팅 용량의 한계 때문에 고성능 퍼포먼스를 기대하기 어렵다. 본 논문에서는 ROS를 이용하여 복잡한 연산을 무선통신으로 오프로딩함으로써 자율주행카트의 한계를 극복하였다. 더불어 음성인식 반응 등의 고객친화적 기능을 추가하여 기존에 없던 시스템을 제안하였다. 이는 대형마트의 주 소비자층인 30-50대 성인 뿐만 아니라 어린이, 사회적 약자에게도 친숙하고 편리한 카트가 된다. 기존에 사용되던 간단한 알고리즘을 넘어선 고성능 알고리즘을 통해 더 정확한 인물 추적이 가능하다. 이는 향후에 대형마트 이외의 사용처에서도 유용하게 사용될 것이다.

범용적으로 사용되는 자율주행카트가 되기 위해서는 몇가지 보완할 사항들이 있다. 첫번째는 RGBD 카메라의 탑재이다. 현재의 라이다 센서로는 앞의 장애물이 사람인지 벽인지 구분하지 못한다. RGBD 카메라를 통해 렌더링된 앞의 사물이 무엇인지 구분하게 할 것이다. 두번째는 음성인식 기능의 다양화이다. 현재는 카트가 고객을 잃었을 경우에 재탐색하기 위한 트리거로서만 사용되지만 향후에는 더 많은 기능이 추가 가능할 것으로 예상된다.

참고문헌

- [1]M. Quigley, B. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, and A. Ng, "ROS: an open-source Robot Operating System," in ICRA Workshop on Open Source Software, Vol.3, 2009.
- [2]Bochkovskiy, Alexey, Chien-Yao Wang, and Hong Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." *arXiv preprint arXiv:2004.10934* (2020).
- [3]WOJKE, Nicolai; BEWLEY, Alex; PAULUS, Dietrich. Simple online and realtime tracking with a deep association metric. In: *2017 IEEE international conference on image processing (ICIP)*. IEEE, 2017. p. 3645-3649.
- [4] 임연수, 서덕진, 박정식, 정유철, 한국어 음성 명령어 인식을 위한 자동데이터 구축. 한국 컴퓨터 정보학회 논문지, 24(12), 17-24. (2019).
- [5]Momotaz Begum, George K.I. Mann, Raymond G.Gosine, "An Evolutionary SLAM Algorithm for Mobile Robots," Proc. of IEEE/RSJ Conference on Intelligent Robots and Systems, pp.4066-4071, 2006
- [6]WOJKE, Nicolai; BEWLEY, Alex. Deep cosine metric learning for person re-identification. In: *2018 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 2018. p. 748-756.