향상된 기능을 가진 특정 개체 추적 로봇 설계

김기식*, 이정훈*, 정영빈*, 이승현*, 동홍석*, 황광일*,**
*인천대학교 임베디드시스템공학과, **인천대학교 IoT and Big-data 연구센터
e-mail: wis1906@naver.com

Designing Specific Object Tracking Robots with Enhanced Functionality

Ki-Sik Kim*, Jeong-Hun Lee*, Young-Bin Jeong*, Seung-Hyeon Lee*, Hong-Suk Dong*, Kwang-il Hwang*,**
*Dept. of Embedded System Engineering, Incheon National University
**IoT and Big-data Research Center, Incheon National University

요 약

지능형 로봇 기술은 더 나은 생활을 위한 현대 기술의 집약체이다. 산업, 생활, 정밀 기술 등 다양한 분야에서 응용이 가능한 확장성 넓은 분야이다. 해당 분야의 추적 기술은 LIDAR를 활용하는 방향으로 활발한 연구가 진행 중이다. LIDAR는 사방의 거리를 정확하게 측정할 수 있는 유용한 센서지만, LIDAR만으로는 로봇의 성능을 최대화할 수는 없다. 본 논문은 LIDAR 추적을 연장하여 Vision 기술의 융합에 관련하여 서술한다. Vision 기술의 융합을 통한 향상된 기능을 가지는 추적 로봇 설계 방법을 제안한다.

1. 서론

지능형 로봇 기술은 더 나은 생활을 위한 현대 기술의 집약체이다. 산업, 생활, 정밀 기술 등 다양한 분야에서 응용이 가능한 확장성 넓은 분야이다. 발전을 거듭하면 인간의 대부분 생활에 자연스레 녹아들어 편의를 제공하고, 사람이 할 수 없는 일들을 도맡아 할 수 있다. 이중 '생활' 분야는 다른 분야보다 사람에게 더욱 직접적으로 다가오는 분야이기에, 미래 사회에 꼭 필요한 분야다. 로봇이 생활에 깊이 스며들수록, 여러 종류의 로봇이 개인의 근처에 있게 되고, 로봇의 수가 많아짐에 따라 로봇을 찾기보다는 로봇이 찾아오는 것이 효율적인 상황이 올 것이다.

본 논문은 Robotis 사의 Turtlebot3 burger 모델의 LIDAR 예제에 Vision 기술을 융합함으로써 높은 인식률을 가지는 효율적인 개체 추적 방법을 제안하고, 실험을 통하여 제안 방법의 효율성을 확인한다.

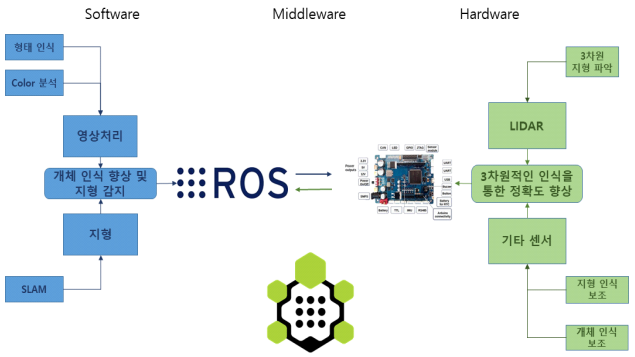
2. 관련 연구

정민우 등[1]은 LIDAR 스캐닝을 기반으로 추출한 거리 및 각도의 데이터를 3D로 재구성하여 개체 인식 능력을 향상하게 하는 방법을 제안하였다. 이종언 등[2]은 LIDAR를 통하여 개체의 이동궤적의 곡률을 계산하여 개체 추적의 성능을 향상하게 하는 방법을 제안하였다. 이처럼 LIDAR를 활용한 개체 추적에 관한 연구는 목적에 따라 다양한 분야에서 진행되고 있다. Robotis 사의 Turtlebot3 burger 모델 로봇 역시도 미들웨어인 ROS(Robot Operating System)를 이용하여 제작한 LIDAR application program 예제는 많이 있고 쉽게 찾을 수 있다. 하지만 LIDAR에 의존하여 로봇을 제어한다는 한계를 가지고 있다. LIDAR는 사방의 거리를 정확하게 측정할 수 있는 유용한 센서지만, LIDAR만으로는 로봇의 성능을 최대화할 수는 없다. Vision 기술을 활용하여 개체의 특성을 파악하여 어떤 개체가 있는지를 인식할 수 있다. '어떤 개체가 있는지를 인식할 수 있다'라는 말은 특정 물체에 대한 행동을 취할 수 있음을 의미한다. 그러므로 로봇에 Vision 기술을 추가함으로써 높은 인식률을 가지는 효율적인 개체 추적이 가능해진다.

3. 프로세스 개요

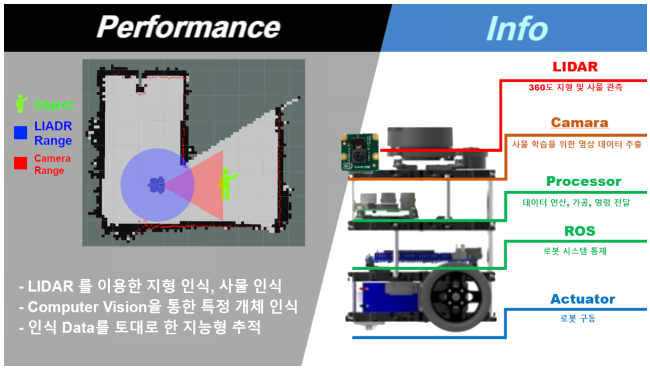
Robotis 사의 Turtlebot3 burger 모델을 활용하여 LIDAR를 사용한 개체 추적 application 예제는 주변의 지형 장애물을 파악하여 2D 도면을 그리는 기술인 SLAM을 활용한다. 이 기술은 2D 지형 도면에 파악된 지형을 실시간으로 재수정하며 더욱 정확한 지형 정보를 얻을 수 있게끔 한다. SLAM은 Vision이 가지는 몇몇 미약한 점을 보완할 수 있다. Vision은 제한된 시야에서의 사물의 정확한 특성을 파악할 수 있지만, 그 사물까지는 거리와 각도, 그 사물까지 도달하기 위한 경로를 정확하게 파악하기는 어렵다. SLAM을 통해 지형을 파악하게 된다면, Vision을 통해 인식한 개체에게 도달하기 위한 경로를 유추할 수 있다.

[본 논문은 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.]



(그림 1) 로봇 구동의 동작 설계도

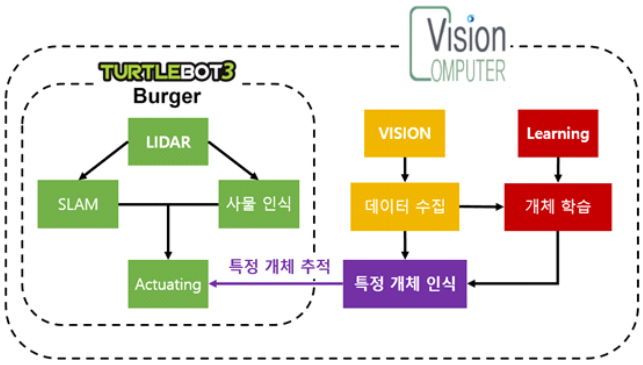
제안하고자 하는 로봇의 설계도는 (그림 1)과 같다. H/W 센서는 가속도 센서 등과 같은 IMU와 Lidar를 이용하여 3차원 적인 공간을 인식하고, 특정 물체를 인식하기 위한 동작을 하여 그에 따른 Feedback 데이터를 프로세서에 전달한다. 센싱값들은 전달된 프로세서에서 상호 보완과 노이즈 제거를 통한 정확도 강화 등의 과정을 거쳐 기체를 지능적으로 움직이기 위한 유용한 정보로 거듭난다. S/W 프로그램은 인식해야 하는 특정 개체의 특징을 분석/학습하여 영상 처리기술에 이용하고 Localization을 위한 SLAM 기술과 융합되어 개체 추적 기능을 한다. Lidar를 통해 추상적으로 잡은 상이 요청한 개체인지 검증하고, 검증 결과를 토대로 Processor에게 진위 여부를 판단할 인자를 반환한다. 실시간 Vision을 통해 끊임없이 상을 추적하고, 학습을 통해 더욱 정확한 정보를 제공할 수 있도록 설계한다. M/W ROS는 S/W 프로세서와 H/W 프로세서(OpenCR) 간의 규격화된 통신을 제공하므로 적극적으로 사용한다. Wireless환경에서도 안정적인 통신을 할 수 있는 설계 방안을 채택하고, 실시간으로 빠르게 들어오는 데이터를 가공하여 즉각적으로 엑추에이팅 할 수 있는 환경을 제공한다. M/W ROS는 S/W 프로세서와 H/W 프로세서(OpenCR) 간의 규격화된 통신을 제공하므로 적극적으로 사용한다. Wireless환경에서도 안정적인 통신을 할 수 있는 설계 방안을 채택하고, 실시간으로 빠르게 들어오는 데이터를 가공하여 즉각적으로 엑추에이팅 할 수 있는 환경을 제공한다.



(그림 2) 로봇 구조와 구현 설명

실험에 사용하고자 하는 로봇의 하드웨어 구조와 구현하고자 하는 기능에 대한 설명은 (그림 2)과 같다. 먼저 로봇의 구조를 살펴보자면, LIDAR는 로봇 근처 360°의 지형 정보를 파악하여 SLAM 지도를 맵핑한다. Camera는 촬영된 사물 프레임을 실시간으로 촬영한다. Processer파트는 LIDAR와 Camera를 활용하여 가공한 개체 추적 데이터를 연산하여 Processeor 하단에 위치하는 OpenCR 보드에 로봇의 액추에이팅 데이터와 센싱 정보를 넘겨주고, 이 데이터를 기반으로 바퀴가 동작한다. 로봇의 종합적인 동작은 실시간 분석된 주변 지형 환경 맵을 인지하여 해당 개체를 효율적으로 추적하게 된다. 최종적인 동작 모습은 (그림 2)의 Performance 부분에서 확인할 수 있다.

4. Vision과 LIDAR 센서의 융합

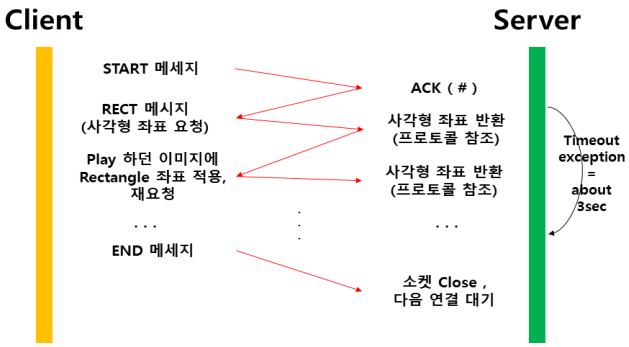


(그림 3) Vision과 Lidar 센서 융합에 관한 다이어그램

Lidar는 넓은 감지 각과 정확한 거리 측정이 가능하여 여러 분야에서 활용되고 있다. Computer Vision 역시 활용 분야가 넓고 발전에 따른 기술 향상이 돋보이는 분야 중 하나다. 하지만, Lidar는 정확한 각도와 거리 측정은 가능하지만, 전방에 어떤 사물이 있는지는 알 수 없다. Computer Vision은 특정 개체 판단이 가능하지만, 사물의 정확한 위치는 알 수 없다. 따라서 이 두 가지를 융합하여 서로의 단점을 보완하겠다는 결론에 도달한다. Computer Vision이 가지는 특정 개체의 인식, Lidar가 가지는 정확한 각도와 거리 인식이 합해진다면 개체의 정확한 위치를 파악하고 대상이 무엇인지 인지할 수 있다.

(그림 3)의 좌측의 Turtlebot3 Burger 로봇에 장착된 Lidar로 지형을 Mapping 하는 SLAM을 이용해 장애물 요소를 미리 인지하고, 그 외 요소들의 위치 정보를 파악하는 프로세스가 진행된다. (그림 3)의 우측의 Vision은 Darknet 이라는 신경망 프레임워크에서 학습된 결과물 중 하나인 YOLOv3를 이용해 추적하고자 하는 사람 개체를 탐지하여 좌표를 반환한다. YYOLOv3는 53개의 Convolutional layer 네트워크를 사용하고 retinaNet 같은 다른 주류의 신경망 네트워크와 비슷한 성능을 가지고 있으면서 최대 3.8배 빠른 성능을 제공하기에 YOLOv3를 이

용하기로 결정하였다[3]. YOLOv3 는 로봇에 탑재된 라즈베리 파이의 프로세서에서 동작하기엔 매우 무거운 프로그램이기 때문에 별도의 서버 컴퓨터에서 동작한다. 로봇에서 S. R. Casner가 제안한 프로토콜인 RTP(Realtime Trasport Protocol) 프로토콜을 통해 영상 정보를 송신하면 서버 컴퓨터에서는 그 영상 프레임에 사람이 있을 것으로 예측된 사각형 좌표정보를 TCP(Transfer Control Protocol) 프로토콜을 통해 반환한다. RTP란 IP 네트워크를 통해 오디오와 비디오를 전달하기 위한 표준화된 패킷 포맷을 정의하는 프로토콜이다[4]. 로봇은 반환된 좌표정보와 센서의 정보를 종합적으로 고려하여 로봇이 움직일 방향을 정하고, 그 방향으로 이동하기 위해 OpenCR 보드에 명령을 내려 바퀴 모터를 제어한다.



(그림 5) 해당 작품 TCP 통신 프로토콜

(그림 5)는 TCP 프로토콜을 이용해 로봇과 서버 컴퓨터가 통신하기 위한 자체적인 프로토콜과 통신 과정의 그림이다.

(a)

(b)

(그림 6) 로봇에 달린 카메라의 시점과 3인칭 시점

(그림 6)은 테스트의 초반부와 후반부 장면이다. (a)와 (b) 두 그림의 좌측은 로봇을 찍은 3인칭 시점, 오른쪽 위는 영상처리 서버 시점, 오른쪽 아래는 로봇에 장착된 카메라에서 본 시점이다. 그림의 장소에서 테스트를 진행하였을 때 인식된 사람의 좌표와 라이더 센서 정보를 토대로 개체 추적을 하는 모습이다. (a)의 로봇 시점에서 두 상을 잡았으나 중앙의 상을 놓치지 않고 추적하는 것을 볼 수 있다. 로봇은 상의 추적 상태를 (a)에서 (b)로 실시간 이동하며 테이블 등의 장애물을 부드럽게 피해 회전하여 개체를 추적한다.

5. 실험 결과

로봇의 프로세서는 Raspberry Pi3 Model B+ 에 장착된 ARM 프로세서이고, 서버 컴퓨터의 CPU와 GPU는 각각 Intel i7-8750H, Nvidia GTX 1060 6GB VRAM 환경으로 구성된다.

평가 항목은 영상 정보를 송신하고, 해당 영상 내 예측된 개체 좌표가 돌아오는 데까지 걸린 시간, 추적하던 개체가 사라졌을 때 판단하는지에 대한 평가, 추적하고자 하는 대상을 올바르게 따라가는지에 대한 평가로 구성된다.

5.1 Round trip 평가

실험 환경은 실내이며, 실험 데이터는 500회(프레임)의 측정을 통해 평균값과 최솟값, 최댓값을 분석한 자료이다.

<표 1> 영상 송신 후 결과 반환까지의 round trip

	mean	min	max
같은 공유기 내 Round trip (ms)	980	930	1040
LAN 직접 연결 Round trip (ms)	900	760	1000

로봇과 서버 컴퓨터를 같은 라우터(공유기) 내의 인터넷 환경에 두고 영상을 전송, 결과를 반환받았을 때 평균 980ms가 소요되었다. LAN 케이블을 직접 연결한 환경에선 평균적으로 900ms가 소요되었다. 전반적으로 LAN 케

이블을 직접 연결한 환경이 좋았지만, Wireless 환경에서도 준수한 성능을 보이는 것을 알 수 있다. 또한, 두 환경 모두 최솟값과 최댓값의 편차가 크지 않은 것으로 보아 안정적인 데이터 전송 및 반환이 이루어진 것을 확인할 수 있다.

5.2 개체 인식에 대한 평가

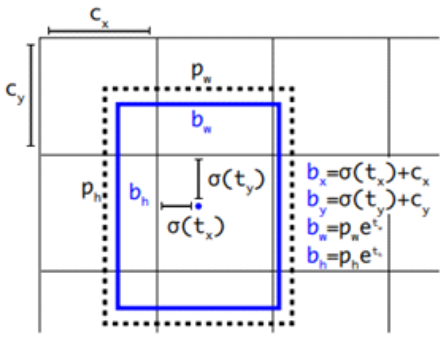
개체 인식에 대한 평가 기준으로 YOLOv3 논문에 기재된 mAP를 사용하였다.

$$recall = \frac{detected\ TRUE}{total\ number\ of\ existing\ TRUE}$$

$$precision = \frac{TRUE\ detections}{whole\ detections\ of\ an\ algorithm}$$

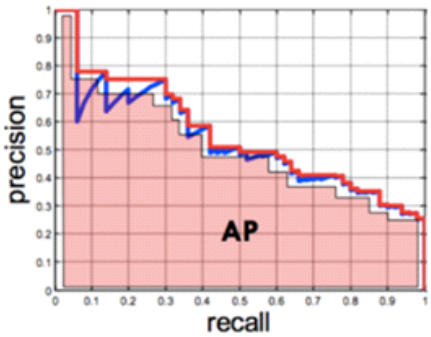
(그림 7) 검출률과 정확도 도출 수식

검출률(재현율)이라고 할 수 있는 용어 recall은 검출하고자 하는 클래스의 물체들을 얼마나 빠뜨리지 않고 잘 잡아내는지를 나타낸다. 정확도로도 표현되는 용어 precision은 검출된 결과가 얼마나 정확한지 (검출된 결과들 중 Ground Truth와 일치하는 결과가 얼마나 포함되는지) 나타낸다.



(그림 8) IOU를 설명하기 위한 그림

프로그램이 예측한 예측 바운딩 박스 Bp와 Ground Truth 바운딩 박스 Bgt에 대해 평가 척도인 IOU (Intersection Over Union) 를 사용하여 '두 박스가 얼마나 겹치나'를 평가하게 된다.



(그림 9) AP를 설명하기 위한 그래프

AP란 precision-recall 그래프의 면적을 뜻하고, AP50, AP75 등의 표현은 IOU의 임계값이 0.5, 0.75일 때의 AP 값이다. 최종적으로 자주 쓰이는 평가 척도인 mAP는 여러 클래스(개체 종류)를 검출해낼 때 각 클래스에 대한 AP의 평균을 뜻한다.

서버 컴퓨터에서는 YOLOv3를 사용하였고, 구성 환경 사양의 한계로 인식률 평가는 Joseph이 제안한 YOLO 평가 지표를 참조하여 진행할 수 있는데, 그 내용에 따르면 YOLOv3-416을 사용하는 환경으로 구성한 평가 방법이 제안되어 있다[5]. 인식률은 mAP-50 기준 55.3으로 준수한 성능을 보인다.

6. 결론

본 논문에서는 Robotis 사의 Turtlebot3 burger 모델을 활용하여 LIDAR와 Vision 기술의 융합을 통한 개체 추적 기능 향상 방법을 제안하였다. 실험 결과, 안정적인 Round trip과 함께 준수한 인식 성능 확인할 수 있었다. 향후 프로세서 등의 하드웨어 업그레이드를 통한 연산 능력 향상으로 실시간 처리 및 응답성이 강화된다면 추적의 퍼포먼스도 더욱 향상될 것으로 예상한다. 제안한 추적 방법을 통해 산업, 생활 편의 등 복합적인 분야로 활용 방안을 넓힐 수 있을 것이다.

참고문헌

[1] Minwoo Jung, Sang Hoon Lee, "Object Perception Algorithm based on LiDAR for Autonomous Vehicle", Conference of The Institute of Electronics and Information Engineers, pp.639-642, 2017.

[2] Jong Eon Lee, Young Chol Kim, "Estimation of the Preceding Vehicle Trajectory from an Autonomous Vehicle using 2D LIDAR", INFORMATION AND CONTROL SYMPOSIUM , pp.196-197, 2014.

[3] Joseph Redmon, Ali Farhadi, "YOLOv3: An Incremental Improvement", University of Washington, pp.1-6

[4] Stephen Lewis Casner, "RTP: A Transport Protocol for Real-Time Applications", University of Southern California, pp.6-8, 2003.

[5] Joseph Redmon, Ali Farhadi, 1