

자율주행 자동차의 객체 인식 기술 현황

이용구 광주과학기술원 기계공학부 교수

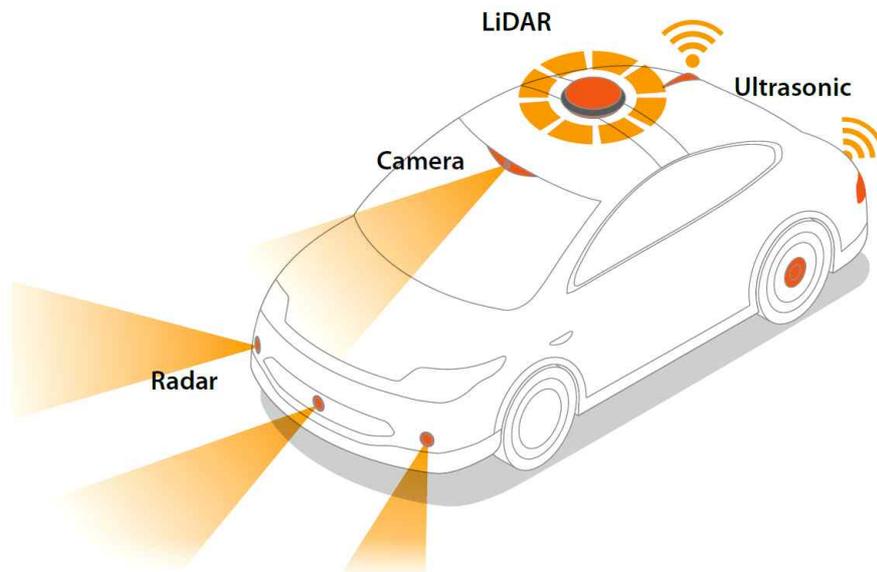
1. 머리말

인공지능의 발달로 자율주행 자동차가 주목을 많이 받는다. 운전할 때는 시각 정보의 처리가 가장 중요하고 사람의 그러한 지능을 컴퓨터로 대체하는 기술이 전통적인 컴퓨터 비전이다. 컴퓨터 비전은 최근 비약적으로 발전한 딥러닝 기술과의 결합으로 정확도가 사람을 대체할 만한 수준에 이르렀다. 시각 정보에서 객체를 탐지하고 이를 토대로 차량의 이동 경로를 계획하고 차를 구동하여 움직이는 기술은 그 시작이 바로 객체 인식 기술에 있다. 자율주행 자동차의 객체 인식 기술과 표준화에 대해 알아보도록 한다.

2. 본문

2.1 자율주행차에서 사용되는 센서

자율주행 자동차의 객체 인식이 타 분야의 객체 인식과 다른 점은 다양하고 이질적인 센서를 동시에 사용한다는 것이다. [그림 1]은 일반적인 자율주행차에서 사용하는 센서들을 보여준다. 동시 사용이 필요한 것은 어느 한 센서의 정보만으로는 만족할 만한 객체 인식이 이루어지지 않기 때문이다. 센서 하나로는 고장에 대처하기 어렵고 아직은 단일 센서의 인식률이 신뢰할 만큼 충분하지 않기 때문에 다양한 센서로 정보를 규합하여 인식률을 높이는 작업이 필요하다.



[그림 1] 자율주행차의 다양한 센서

일반적으로 라이다(LiDAR)는 3차원 정보를 가공해 제공하기 때문에 자율주행차에서 사용하기 매우 쉽다. 레이다(Radar)는 2차원으로 제공되어서 정보의 차원은 낮으나 추돌 위험 상황을 빠르고 신뢰성 있게 파악할 수 있기 때문에 위급하고 돌발적인 상황을 감지하기 위해서 유용하다. 카메라의 정보는 매우 풍부하나 객체들을 인식하려면 여러 단계의 계산을 수행해야 한다. 센서 정보 중 가장 중요한 정보가 시각 정보이다. 이때 정적인 디지털 사진 인식과 동적인 비디오 인식이 같이 이루어져야 한다. 시각 정보는 외부 조명에 의한 빛이 피사체에 반사되어 나오는 정보를 이용한다. 따라서 악천후나 날씨가 흐릴 때 그 정보의 품질이 현저히 떨어질 수 있다. 따라서 자율주행차는 주변 탐지를 추가적으로 하는데 대표적인 두 가지 기술이 바로 라이다 및 레이다이다. 두 방법 다 전자기적인 파동을 외부로 송신하여 물체에 의해 산란된 신호를 수집하는 물리적인 현상을 이용한다. 라이다는 1000nm 이상의 원적외선 대역의 레이저 광원을, 레이다는 10m~200m의 라디오 주파수를 사용한다.

레이다의 경우 단일 파장대의 광원을 사용하기 때문에 반사되어 나온 신호 처리가 매우 쉽다. 또한 자율주행차가 인식해야 할 물체보다 광원의 파장이 106 이상으로 훨씬 더 작기 때문에 광원이 직진과 반사를 하는 선(ray)으로 간주되어 매우 간단한 물리 현상으로 해석이 가능하다. 또한 레이다는 값이 싸고 안개나 비에 의해 전자기파가 흡수되지 않기 때문에 많은 자율주행차에서 사용하고 있다.

마지막으로 현대 자동차에서 주차 시 인접 차량과 근접하는 경우 경고음을 내는 센서로 많이 사용하는 초음파(Ultrasound) 센서가 있다. 초음파 센서는 기계적인 음파를 발생하고 주변 물체와의 반사를 측정하여 거리를 계산한다. 초음파는 악천후에도 오류 없이 작동하며 심지어 센서가 먼지에 덮히는 악조건에서도 잘 작동한다. 이는 전천후 작동을 기대하는 자동차 특성상 매우 신뢰성 있는 센서이지만 아직 대부분의 자율주행차에서는 사용하지 않고 있다.

2.2 카메라를 이용한 객체 인식 방법

카메라에서 얻는 정보는 카메라에 수직인 방향으로 투영되는 피사체의 가시광선대 빛의 세 가지 파장대에서 나타나는 세기이다. 디지털 카메라의 보급으로 얻어지는 정보는 투영방향에 수직인 두 방향으로 배열을 이룬다. 이러한 다채널 배열정보로부터 객체를 인식하는 방법의 핵심은 객체가 놓여 있는 영역을 잘라내는 것(region selection)과 해당 영역에 있는 객체를 분류(classification)하는 것이다. 계산을 빨리하기 위해 많은 네트워크 아키텍처는 두 가지 작업을 동시에 수행한다. 자율주행은 정지된 영상뿐 아니라 연속으로 이어진 영상에서도 객체를 인식해야 한다. 또한 시간적으로 진행되는 영상에서 각 객체에 아이디를 부여하고 이들을 추적하는 계산을 성공적으로 수행해야 한다. 이러한 추적은 객체 간의 가림(occlusion) 현상으로 매우 어려워진다. 또한 카메라를 부착한 차량이 고속으로 이동을 할 때는 객체를 탐지할 시간이 매우 제한된다.

2.3 라이다와 비디오카메라의 데이터 융합

두 개의 카메라를 사용하면 피사체의 3차원 좌표 값을 얻는다. 자율주행차에서는 객체를 인식할 뿐만 아니라 객체가 멀리 또는 가깝게 있는지를 판별하여 가까운 물체에 더 많은 관심을 가져야 한다. 계산을 효율적으로 하고자 일정거리에서 벗어난 객체는 일반적으로 탐지하지 않

는다. 테슬라와 모빌아이와 같이 라이다를 사용하지 않는 자율주행차 기술에서는 이러한 방식으로 도로상의 객체에 대한 3차원 위치 정보를 얻는다. 라이다를 사용하는 자율주행차의 경우에는 피사체의 거리 정보를 센서에서 얻을 수 있기 때문에 원근 구별이 매우 쉽다.

라이다를 사용할 경우 원시데이터를 사용하기보다는 피사체의 표면을 점군(Point cloud)으로 표현한 가공데이터를 활용한다. 여기서 점은 라이다 센서 좌표계를 원점으로 한 직교 좌표계의 좌표 값 x, y, z 가 된다. 점군에서 객체를 인식할 때는 대부분 PointNet[1]이라는 점군에 특화된 변환 전처리기를 입력단에 둔, 다층 퍼셉트론(Multilayer perceptron)으로 구성된 네트워크를 사용한다. PointNet의 전처리기는 점군이 순서에 무관하고, 점들은 멀고 가까운 거리 구분이 되어야 하고, 회전이나 병진 같은 변환에 무관해야 한다는 특징을 이용한다. PointNet과 그것에 계층구조를 부여한 PointNet++[2]를 사용하면 객체의 인식뿐만 아니라 주변 배경의 분할(Segmentation)도 가능하다. 라이다의 단점은 얻는 점군이 매우 부족하여 센서로부터 일정 거리가 되면 매우 작은 수의 점군만을 얻어 객체인식이 어렵다는 데 있다. 이를 보완하고자 라이다(LiDAR)의 점군을 중심으로 카메라에서 얻은 컬러 이미지를 부착(Augment)하여 측정되지 않는 영역의 데이터를 보충한다. 나아가서 다채널 색상 정보를 객체 인식에 활용하는 연구들도 있다[3]. 이러한 경우에 다른 양상의 센서 정보들을 융합해야 한다. 여러 센서 간의 데이터를 융합하는 것은 센서의 특성뿐 아니라 일부 센서가 원활한 작동을 하지 않을 경우까지 포괄해야 할 것이다.

2.4 자율주행차 데이터셋

현재 다수의 자율주행차 데이터셋이 공개되어 있다. 자율주행차 데이터셋은 차량에 부착된 다수의 센서가 동시에 수집을 하였기에 멀티모달(Multi-modal)의 특성을 가진다. 최소한 카메라 외에 라이다의 점군을 동시에 제공하며 최근에 나온 데이터셋은 레이더 측정도 같이 제공한다. 자율주행차 데이터셋은 멀티모달 특성상 촬영의 단위가 되는 장면(scene)의 숫자는 대체로 1k를 넘지 않으며 객체를 구분하고자 다루는 클래스의 숫자도 10개 내외이다[4]. 공통적으로 다루는 클래스는 자동차, 보행자 그리고 자전거이다. 자율주행 데이터셋에서는 인식한 객체가 3차원 상에서 이루어져 있기 때문에 2차원 박스가 아닌 3차원 박스를 사용하는 것이 일반적이다.

2.5 자율주행 데이터셋의 표준화

자율주행 데이터셋은 각 모드의 특색에 맞는 데이터를 기본적으로 포함해야 한다. 가령 카메라 데이터일 경우 RGB(Red, Green, Blue) 이미지가 기본적으로 있어야 한다. 그밖에 부록 정보로 사용된 카메라 렌즈, 카메라 장착 위치, 촬영 시간, GNSS(Global Navigation Satellite System) 정보, 기후 등의 정보가 있으면 네트워크 정확도를 더 향상시킬 수 있다. 탐지된 객체의 클래스와 바운딩 박스(Bounding box)도 있어야 한다. GIST[5]에서는 국내외에 이러한 정보를 표현하는 표준이 없기 때문에 이에 대한 국내표준을 제안한 바 있다.

국내의 경우 위와 같은 표준이 제안된 반면 국외에는 아직 이러한 표준을 제정하려는 움직임이 알려진 바 없다. 그 이유는 미국의 경우 산업 표준이 산업계에서의 자발적인 제안을 토대로 만들어지는 경우가 일반적인 경우이기 때문이다. 아직 자율주행 자동차는 산업계의 공통된 합의가 만들어지지 않았고 그러한 이유 때문에 표준이 만들어진 바가 없다고 보는 게 옳다.

딥러닝을 위한 데이터셋을 만들 때 위와 같은 표준화된 분류체계와 더불어 바운딩 박스를 실제로 적용하는 규칙에 대한 규범도 따로 만들어야 한다.

객체를 인식하기 위한 데이터셋은 꼭 이상적인 사진만 있지는 않다. 객체 인식기는 실제 상황에서 작동해야 하는데 완전히 이해되는 영상이 아닐 경우가 있기 때문이다. [그림2]는 가려진 상황에서의 라벨링 예시이다. 위와 같이 데이터셋에서는 완벽하지 않은 경우라도 객체의 바운딩 박스를 만들어주는 것이 좋은 성능의 검출자를 만드는 데 기여한다. 또한 일부 사진에서는 [그림3]과 같이 가려짐, 빛 반사, 흐림 같은 예외적 상황이 발생할 수 있다. 이러한 경우의 라벨링에 대해서는 가능한 포괄적인 클래스를 표기하고 식별이 어려운 이유 등을 표기하는 것도 좋은 검출자를 만드는 데 기여할 것이다.



[그림 2] 외부 환경에 의해 가려진 객체 주석 표기 예시



[그림 3] 식별 불가 주석 표기 예시 표기 예시

2.6 자율주행 시나리오의 표준화

자율주행은 객체 검출에서 나아가서 객체들이 시간에 따라 어떻게 이동하느냐도 매우 중요한 관심사이다. 최근의 많은 데이터셋은 수십 초 정도의 장면을 단위로 다수의 센서 측정값들을 제공한다. 이러한 장면들이 필요한 교통 상황을 모두 포괄하는지는 자율주행차의 안전성을 위해서 매우 중요하다. 이에 대한 산업계의 요구에 따라 장면을 기술할 수 있는 설명 언어 (Description language)들이 발표되고 있다[6]. 현재 국내에도 이러한 연구들이 이루어지고 있으나 컴퓨터가 이해할 수 있는 형식(Computer readable form)이 아니라 자연어로 기술하고 있다. 장면을 표현하는 방법으로 실제 상황에서는 재현하기 어려운 상황을 가상현실기법으로 만들어내는 도구도 있다. 게임기술에도 사용되는 실사 그래픽 환경과 물리기반으로 다양한 센서의 측정값을 계산해서 데이터셋을 만들어내 부족한 데이터셋을 보충하는 방법을 취한다[7].

3. 맺음말

자율주행 분야의 객체 인식은 다른 분야와 마찬가지로 예외적인 경우까지 포함해 레이블된 데이터셋을 필요로 한다. 레이블에 대한 방식은 산업계에서 아직 충분한 합의가 이루어지지 않아 기초 표준만 나와 있다. 자율주행 분야에서 사용하는 센서는 기본적인 카메라 외에 레이저로 거리 측정을 하는 라이다, 긴급상황을 위한 레이다, 주차 시 활용하는 초음파 등의 매우 다양한 센서가 있다. 이러한 다수의 센서를 융합하여 객체를 판단하는 융합 기술들이 객체 인식의 중요한 부분을 차지한다.

[참고문헌]

- [1] C. R. Qi, H. Su, M. Kaichun and L. J. Guibas, 'PointNet: Deep learning on point sets for 3D classification and segmentation,' 2017 IEEE Conference on computer vision and pattern recognition (CVPR), 2017, pp. 77-85
- [2] C. R. Qi, L. Yi, H. Su, L. J. Guibas, 'Pointnet++: Deep hierarchical feature learning on point sets in a metric space,' Advances in neural information processing systems, 2017, pp. 5099-5108
- [3] S. Vora, A. H. Lang, B. Helou and O. Beijbom, 'PointPainting: Sequential fusion for 3D object detection,' 2020 IEEE/CVF Conference on computer vision and pattern recognition (CVPR), 2020, pp. 4603-4611
- [4] H. Caesar et al., 'nuScenes: A multimodal dataset for autonomous driving,' 2020 IEEE/CVF Conference on computer vision and pattern recognition (CVPR), 2020, pp. 11618- 11628
- [5] '자율주행 자동차의 객체 인식기술에 필요한 도로상 데이터의 객체 분류 체계,' 정보통신단체표준(국문표준), TTA.KO-10.120, 정보통신기술협회 2019.12.11
- [6] 'Measurable scenario description language,' Reference version 20.07, www.foretellix.com, 2020
- [7] cognata.com, 2020

※ 출처: TTA 저널 제191호

(코로나 이슈로 각 표준화기구의 표준화회의가 연기·취소됨에 따라 TTA 저널로 대체합니다)