

허영진 (주)뉴로메카 CILab(Cobot Intelligence Lab) 팀장

| e-mail : youngjin.heo@neuromeka.com

박종훈 (주)뉴로메카 대표이사

| e-mail : coolcat@neuromeka.com

정완균 포항공과대학교 기계공학과 교수

| e-mail : wkchung@postech.ac.kr

이 글에서는 로봇 비전에 활용 가능한 다양한 딥러닝 추론 기술과 이와 관련된 상용서비스에 대한 현황을 소개하고자 한다.

딥러닝 추론 기술의 중요성

최근 딥러닝 기반의 인공지능 기술이 급속도로 발전함에 따라 로봇, 의료, 금융 등 다양한 산업분야에서 이를 활용하는 사례들이 늘어나고 있다. 산업분야 중 로봇기술에 대해서는 로봇의 위치제어, 자세제어와 같은 로우레벨 제어부터 Learning from Demonstration(LfD)과 같은 로봇의 작업이해 분야까지 딥러닝을 적용하는 많은 연구들이 수행되고 있

다. 특히 로봇비전 분야는 딥러닝 기반의 영상인식 기술의 눈부신 발전에 힘입어 가장 빠르게 발전하고 있다.(그림 1)

로봇기술에 딥러닝을 적용시키기 위해서는 학습모델의 추론정확도 외에 추가적으로 고려해야 할 요소들이 있는데, 이 중 대표적인 것은 학습된 신경망에 대한 실시간 및 고속 추론 기술이다. 기본적으로 머신러닝 및 딥러닝은 학습(training phase)과 추론(inference phase) 두 가지 과정으로 구분된다. 학습은

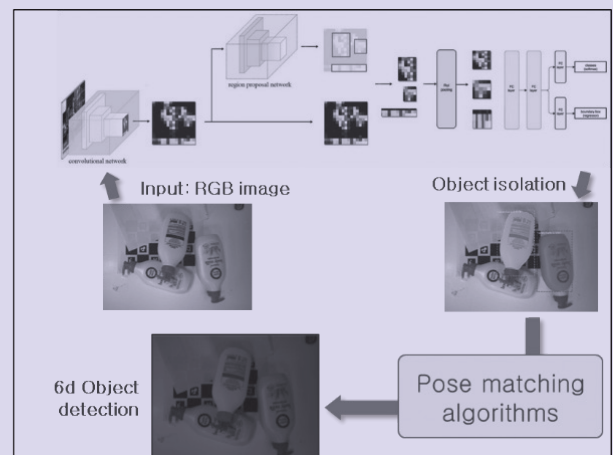
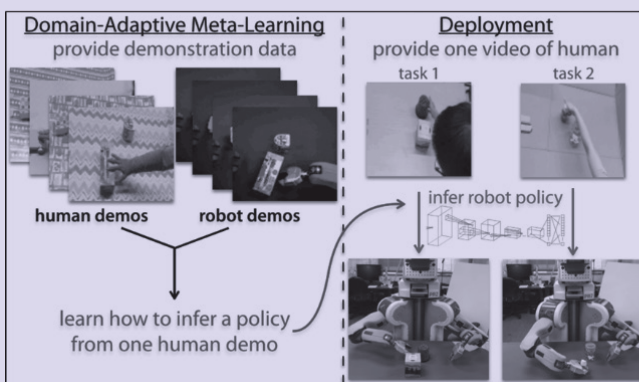








그림 1 Meta-learning을 이용한 LfD 예시(좌), 로봇비전에 대한 딥러닝 적용 예시(우)

Inference Latency 					All Submissions
Objective: Latency required to classify one ImageNet image using a model with a top-5 validation accuracy of 93% or greater.					
Rank	1-example Latency (milliseconds)	Model	Hardware	Framework	
1  Dec 2018	4.2180	ResNet50 Perseus AI Cloud Acceleration team in Alibaba Cloud source	Alibaba Cloud [ecs.gn5i-c8g1.2xlarge]	TensorFlow 1.10.0	
2  Apr 2018	9.9600	ResNet50 Intel(R) Corporation source	Amazon EC2 [c5.18xlarge]	Intel(R) Optimized Caffe	
3  Apr 2018	12.4000	ResNet50 Intel(R) Corporation source	Amazon EC2 [c5.4xlarge]	Intel(R) Optimized Caffe	
4  Apr 2018	17.3800	ResNet50 Intel(R) Corporation source	Amazon EC2 [c5.2xlarge]	Intel(R) Optimized Caffe	
5  Nov 2017	22.2700	ResNet 152 Stanford DAWN source	1 P100 / 30 GB / 8 CPU (Google Compute)	TensorFlow v1.2	







Inference Cost 					All Submissions
Objective: Average cost on public cloud instances to classify 10,000 validation images from ImageNet using an image classification model with a top-5 validation accuracy of 93% or greater.					
Rank	Cost (USD)	Model	Framework	Hardware	
1  Dec 2018	\$0.02	ResNet50 Perseus AI Cloud Acceleration team in Alibaba Cloud source	TensorFlow 1.10.0	Alibaba Cloud [ecs.gn5i-c8g1.2xlarge]	
2  Apr 2018	\$0.02	ResNet50 Intel(R) Corporation source	Intel(R) Optimized Caffe	Amazon EC2 [c5.2xlarge]	
3  Apr 2018	\$0.02	ResNet50 Intel(R) Corporation source	Intel(R) Optimized Caffe	Amazon EC2 [c5.4xlarge]	
4  Nov 2017	\$0.07	ResNet 152 Stanford DAWN source	MXNet 0.11.0	1 K80 / 61 GB / 4 CPU (Amazon EC2 [p2.xlarge])	
5  Nov 2017	\$0.11	ResNet 152 Stanford DAWN source	TensorFlow v1.2	1 P100 / 30 GB / 8 CPU (Google Compute)	

그림 2 DAWNBench 추론지연(좌) 및 추론비용(우), 2019년 2월 14일 기준

많은 데이터를 통해 모델의 손실함수를 최적화해 목표하는 작업에 대한 성능을 달성하기 위한 과정이고, 최종적으로 학습된 모델은 새로이 입력되는 데이터에 대한 추론을 위해 이용된다. 잘 학습된 신경망 모델을 로봇과 융합하여 배치(deployment)할 경우 추론 정확도뿐만 아니라 추론 효율, 추론 속도, 안전성 등에 대한 추론 최적화가 중요해진다.

실시간성이 보장되어야 하는 로봇 작업은 로우레벨 제어 및 신호처리, 고장진단, 충돌감지 등이 있으며, 이러한 작업은 실시간성과 고속추론이 보장되지 않으면 추론 정확도뿐만 아니라 안전성에도 문제가 발생한다. 예를 들어 로봇의 실시간 시계열 데이터를 이용하는 제어, 신호처리 및 고장진단은 실시간성 보장이 되지 않으면 올바른 추론평가 자체를 수행할 수 없으며, 로봇이 사람과 충돌 시 혹은 자율주행 자동차의 비전시스템이 위험상황을 감지하였을 시 추론지연으로 인해 곧바로 대응하지 못하는 경우 안전성에 대한 문제가 발생한다. 로봇비전은 응용분야에 따라 실시간성이 요구되지는 않지만 고속 추론은 여전히 중요하다. 예를 들어 매니플레이터가 픽-앤-플레이스 작업 수행을 위해 딥러닝 기반의 로봇비전을 적용할 경우 비실시간 추론으로도 작업 수행이 가능하지

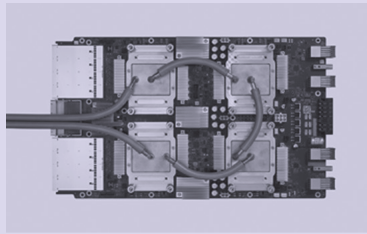
만, 작업시편에 대한 로봇비전의 인식 작업(영상분할 및 물체자세추정)에서 많은 연산시간을 소모하게 되면, 추론 지연시간만큼 전체 로봇 작업의 효율이 떨어지게 되므로 자동화 로봇작업의 효율을 위해서 고속의 추론 연산이 요구된다.

스탠포드의 DAWN 프로젝트에서 발표한 DAWNBench는 딥러닝 학습 및 추론에 있어서 효율에 대한 벤치마크를 제공한다(<https://dawn.cs.stanford.edu/benchmark/>). 기존의 심층신경망과 관련된 벤치마크들이 대부분 정확도(accuracy, mAP 등)에만 집중했던 것과 달리 DAWNBench는 심층학습의 워크로드에 대한 정량화를 통한 연산 효율에 중점을 두었다. DAWNBench에서 경쟁하는 항목은 학습시간, 학습비용, 추론지연, 추론비용이다. 예를 들어 추론지연은 추론비용에 대해서는 고정된 목표성능(영상분류의 경우 ImageNet 검증데이터에 대한 93% 탑-5 정확도) 이상을 가지는 모델에 대한 연산 지연 시간을 msec/image 단위로 나타내며, 추론비용은 1만 개의 검증 데이터를 상용 클라우드 인스턴스를 이용하여 추론연산 할 때 발생하는 평균비용을 나타낸다. 또한 각 경쟁요소에 대한 성과를 달성하기 위해 사용된 다양한 최적화 전략, 모델의 구조, 소프

트웨어 프레임워크, 클라우드, 하드웨어로부터 경쟁항목에 대한 정량화를 통해 딥러닝 워크로드에 대한 래퍼런스를 제공한다.(그림 2)

추론효율 향상 연구 및 시장

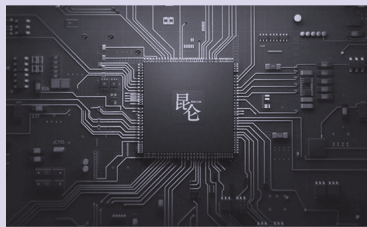
추론효율에 대한 연구는 연산효율을 높이기 위한 알고리즘 개발과 연산속도를 가속시킬 수 있는 하드웨어 개발 두 가지로 나눌 수 있다. 알고리즘의 경우 MIT의 Song Han 교수가 제안한 ‘Deep Compression’이 가장 대표적이며, 가지치기(pruning), 가중치 공유, 양자화, 허프만 코딩 등의 기법을 통해 모델을 압축하여 연산효율을 극대화시키는 방법들을 제안한다. 특히 가지치기는 반복적 재학습 과정을 통해 신경망의 연결노드의 개수를 1/10 수준까지 줄이고서도 신경망의 정확도가 떨어지지 않게 유지시킬 수 있다. 이와 같은 알고리즘 기반의 추론효율 향상만으로도 추론효율을 수십 배 향상시킬 수 있다. 상용화된 추론 전용 플랫폼 중 가장 잘 알려진 것은 NVIDIA에서 제공하는 TensorRT이다. TensorRT는 범용 GPU를 이용하여 추론속도를 극대화시킬 수 있는 CUDA기반의 API를 제공한다. TensorRT는 다양한 딥러닝 프레임워크에 구현된 신경망 레이어에 대한 변환 기능을 제공하지만 빠르게 발전하는 딥러닝 모델들에 비해 지원되는 레이어의 범위는 아직 부족한 편이어서 개발자가 플러그인 레이어 코드를 직접 작성해야 하는 불편함이 있다. 최근 국내에서는 이러한 불편을 해소시키고자 딥러닝 추론을 가속시켜주는 솔루션을 제공하는 스타트업도 등장하여 다양한 기업 및 연구소에 이러한 가속 솔루션을 제공하고 있다.(<http://soynet.io>)



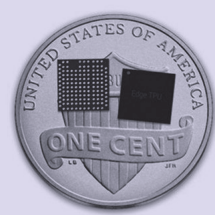
구글 TPU3.0



NVIDIA Tesla T4 GPU



Baidu Kunlun



구글 Edge TPU

그림 3 딥러닝 연산 가속을 위해 개발된 다양한 연산 하드웨어

딥러닝 연산효율 증가를 위한 하드웨어로는 범용적 연산장치인 CPU 또는 GPU가 대표적이지만 이 외에도 FPGA(필드프로그램머블게이트어레이), ASIC(어플리케이션특화직접회로)를 이용한 별도의 특화된 하드웨어를 통한 연산효율 향상 제품들이 출시되고 있다. 딥러닝 가속 ASIC 중 가장 대표적인 것은 구글에서 자체 개발한 TPU(Tensor Processor Unit)이며 최근 구글은 이전 버전보다 8배 빨라지고 100페타플롭의 성능을 낼 수 있는 TPU3.0을 개발했다. 아마존 또한 AWS 인퍼런티아(Inferentia)라고 명명한 ASIC 기반의 추론 칩을 자체 개발하여 추론 성능을 최적화시킨 솔루션 서비스를 출시를 발표했다. GPU의 전통 강자인 NVIDIA는 자사의 튜링 아키텍처에 기반한 튜링 텐서코어를 탑재한 추론 가속 전용 테슬라 T4 GPU를 출시하였다. 이 외에도 바이두의 추론 칩 Kunlun, 구글의 Edge TPU 등은 엣지컴퓨팅과의 통합 솔루션 제공을 위해 개발되었다.(그림3)

클라우드 컴퓨팅과 엣지 컴퓨팅, 어떻게 다를까?

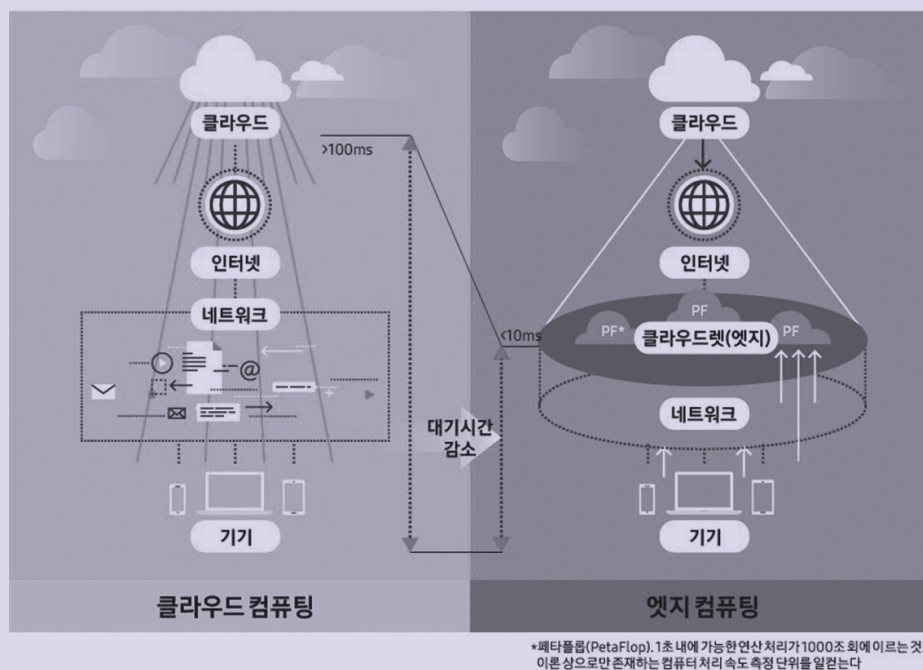


그림 4 클라우드 컴퓨팅과 엣지 컴퓨팅(출처: 삼성 NEWROOM)

클라우드 컴퓨팅 기반의 추론 서비스

딥러닝 응용분야가 다양해짐에 따라 모바일 플랫폼에서의 인공지능 기술 접목 또한 각광받고 있다. 소형 모바일 장치와 사물인터넷 장치에는 고성능 연산 칩을 탑재하는 데에 한계가 있기 때문에 클라우드 컴퓨팅을 이용한 추론 서비스에 대한 수요 및 공급이 증가하고 있다. 클라우드 기반의 연산 플랫폼 중 가장 대표적인 것은 Google Cloud Platform(GCP)이다. GCP는 영상분석, 텍스트분석, 음성분석 등 다양한 클라우드 컴퓨팅 플랫폼을 제공하며 TPU칩을 이용할 수 있는 Cloud TPU 서비스 또한 제공하여 딥러닝 학습과 추론 모두에 이를 이용할 수 있다. 최근 여러 기업에서는 추론기능에 더 큰 중점을 둔 클라우드 서비스를 차례대로 발표하였으며 이들은 모두 대량처리, 저지연성, 저가(high-throughput, low-

latency, cost-effective) 세 가지 요소를 강점으로 내세운다. 아마존은 AWS 인퍼런티아를 자사의 클라우드 서비스 Amazon EC2를 통해 제공하고 있으며, 바이두 또한 Kunlun칩을 이용할 수 있는 클라우드-투-엣지 플랫폼을 발표하였다. NVIDIA는 테슬라 T4 GPU와 TensorRT를 함께 제공하는 GPU 클라우드 플랫폼을 서비스하기 시작했다. 이러한 클라우드 연산 서비스들은 주로 사용시간, 작업량을 기준으로 가격을 책정하기 때문에 사용자는 합리적인 가격으로 추론가속 서비스를 이용할 수 있다. 클라우

우드 컴퓨팅 플랫폼들은 로보틱스, 예지보수, 이상탐지, 머신비전 등에 대한 활용을 주요 타겟으로 삼고 있으며 스마트공장에서의 로봇들의 연결성을 증대시키거나, 스마트시티를 구성하는 데에 큰 기여를 할 것으로 기대된다. 또한 연산 가속 하드웨어 및 소프트웨어에 대한 연구가 활발히 수행되고 저지연 클라우드 컴퓨팅 서비스 시장이 확대됨에 따라 딥러닝의 추론 지연은 더욱 빠르게 줄어들어 로봇의 지능화 범위는 점차 확대될 것이다. 지능화 범위의 확대는 지능로봇의 응용 범위를 확대시켜 로봇 지능 연구와 로봇 시장의 확장을 가속시킬 수 있을 것으로 기대된다.

5G 통신과 엣지 컴퓨팅과의 접목 및 산업용 로봇에서의 응용

클라우드 컴퓨팅 서비스의 경우 딥러닝 연산속도

보다 통신에서 발생하는 지연시간이 더 클 수밖에 없다. 최근 각광받는 엣지 컴퓨팅 기술은 기존 클라우드 컴퓨팅과 대조되는 기술로서 데이터가 생성되는 기기와 가까운 곳(네트워크 끝단, edge)에서 데이터를 처리하는 방식이다. 막대한 양의 데이터를 클라우드를 통해 중앙 데이터센터까지 보내지 않고 end-point에서 데이터를 수집 및 연산하여 데이터 흐름을 가속시킬 수 있고 지연시간을 줄일 수 있다. 이를 통해 저지연(low latency)이 요구되는 다양한 어플리케이션에 응용될 수 있을 것으로 기대된다.

그리고 최근 4차 산업혁명과 관련된 대표적 메가트렌드인 5G 통신 기술은 1msec 수준의 초저지연성의 특성을 가지며 이에 대한 상용화가 코앞으로 다가왔다. 초저지연성을 갖는 5G 통신과 엣지 컴퓨팅의 결합은 클라우드 컴퓨팅의 통신지연 문제를 완전히 해결할 수 있으며 이를 통한 초연결 시대가 도래할 것으로 기대된다. 클라우드 엣지에 딥러닝 추론 가속 하드웨어 및 소프트웨어를 탑재하고 개별 디바이스에서 획득되는 데이터들을 5G를 통해 엣지 컴퓨팅 장치

로 전송하게 되면 딥러닝 연산지연과 통신지연 두 가지를 잡을 수 있는 강력한 시스템이 될 것이다. 이러한 5G 엣지 컴퓨팅 기술의 킬러 어플리케이션으로서 사물인터넷, 로봇틱스, 스마트팩토리, 자율주행기술 등이 가장 많이 언급되고 있다. 특히나 로봇비전에 이러한 통합 시스템을 적용 할 경우, 카메라는 5G 통신모듈을 이용하여 별도의 케이블 연결 없이 영상을 송수신 하고, 카메라 헤더에 내장된 통신모듈만으로 초저지연 딥러닝 연산이 가능한 강력한 로봇비전 시스템이 개발될 수 있을 것으로 기대된다. 이는 빈-피킹 작업과 같이 3차원 비전솔루션과 로봇솔루션이 통합되어야 하는 고가의 자동화 작업솔루션에 대한 비용절감을 가능케 할 것이며, 이를 통해 로봇 및 로봇비전 통합 솔루션에 대한 진입장벽을 낮추어 스마트팩토리 보급을 가속시킬 수 있을 것으로 기대된다. 실제로 5G 통신, 엣지 컴퓨팅, 로봇, 인공지능을 결합한 서비스 개발을 위해 최근 여러 기업들이 협업을 시작하였으며, 빠른 시일 내에 초연결성 추론 기능을 내장한 로봇비전 솔루션이 시장에 등장할 것으로 전망된다.

황 서 연 삼성전자 글로벌기술센터 책임
 노 경 식 삼성전자 글로벌기술센터 마스터

| e-mail: ssy.hwang@samsung.com
 | e-mail: kyung.roh@samsung.com

이 글에서는 산업용 물류이송 로봇의 주행 기술 동향 및 지각(perception) 측면에서의 자율주행 기술, 그리고 향후 과제 및 전망에 대해 소개하고자 한다.

산업용 물류이송 로봇은 주로 공장이나 물류창고 등에서 제품 조립을 위한 자재나, 우편물 또는 택배를 이송 또는 분류하는 목적으로 사용된다. 최근 4차 산업혁명이 대두되면서 산업용 물류이송 로봇에 대한 수요가 빠르게 증가하고 있다. 기존의 공장들도 더욱 지능적인 스마트팩토리 체제로 전향하면서 물류를 이송하는 로봇의 중요성이 날로 커지고 있으며, 고도로 자동화된 물류창고의 경우에는 이미 몇 명의 관리 인력을 제외한 대부분의 인력들이 로봇으로 대체되고 있다.

산업용 물류이송 로봇은 주행 방식에 따라 AGV(Automatic Guided Vehicle)와 좀더 진보된 형태인 AMR(Autonomous Mobile Robot)로 구분할 수 있다. AGV는 주로 지면에 부착되거나 매설된 광학, 자기 유도선 또는 마커를 따라 이동하는 방식을 사용하며, 간편한 설치 및 운용이 가능하여 10년이 넘게 산업 현장에 널리 적용되어 왔다. 이동하고자 하는 경로를 따라 유도선을 설치함으로써 직관적으로 경로를 설정할 수 있지만, 사람, 대차 등에 의해서 부착된 유도선이 쉽게 훼손되거나 매설된 유도선을 옮기기 위해서는 매번 공사가 필요한 단점이 있다.

AMR은 자율주행 방식을 사용하며, 라이다(LiDAR), 비전 센서 등을 기반으로 환경 내에서 현재 위치를 실시간으로 추정하고 주변 상황 분석을 통해

주어진 목적지까지의 최적 경로를 계산하여 이동한다. 장애물 상황에 따라 로봇이 자율적으로 경로를 변경하기 때문에, 주행 중 발생할 수 있는 다양한 돌발 상황에 대해 대처할 수 있는 센서조합 및 문제해결 능력이 구비되어야 장시간 안정적인 주행이 가능하다. 유도 방식에 비해 시스템이 복잡하고 환경에 대한 정보를 모델링해야 하는 등 설치 절차가 까다롭지만, 보다 유연한 운영이 가능하여 점유율이 빠르게 늘어나는 추세에 있다.

AMR은 최근 실내뿐만 아니라 실외환경에까지 영역을 확대해나아가고 있다. 기존에는 건물 내에서 물류를 이송하는 역할을 주로 담당했다면, 이제는 사업장 내 건물 간 이송을 위해 사용해 온 차량들의 역할까지 넘보고 있다. 주행 측면에서는 3차원 라이다 센서 등의 발전에 힘입어 더욱 더 풍부한 환경 정보를 획득하여 실외에서도 안정적인 위치추정을 수행할 수 있게 되었으며, 딥러닝 기반 물체인식 기술의 등장으로 차선, 보행자, 차량 등의 주야간 인식 성능이 크게 향상되어 보다 지능적인 주행 상황 판단이 가능하게 되었다. 실외용 AMR에는 무인자동차와 유사한 주행 기술이 탑재되지만, 일부 제약조건(예, 사업장 내 주행속도 및 영역 제한)하에서 운행되기 때문에 상용화 시점은 상대적으로 더 빠를 것으로 예상된다.

주행 기술 동향

물류창고 자동화를 선도하고 있는 Amazon은 2012년 Kiva Robotics를 인수하여 지속적으로 AGV의 완성도를 높이고 있다. 각 AGV는 물품이 담긴 선반을 통째로 들어올려 목표 지점으로 이동한다. 이전에는 사람이 선반 사이를 이동해야 했기 때문에 별도의 통로 확보가 필요했으나, AGV는 선반 아래로 이동할 수 있으므로 통로를 제거하여 상당한 공간을 절약할 수 있게 되었다. 주행 시에는 격자 형태로 설치된 바닥 마커 정보를 인식하여 로봇의 현재 위치를 파악하며, 중앙 서버에서 Wi-Fi 무선통신을 통해 각 로봇을 제어한다. 상황에 따른 경로 변경은 가능하지만 정해진 인프라(예: 격자 형태의 경로) 내에서만 허용된다. 이와 유사한 주행 방식을 보유한 기업으로는 Grenzebach, Swisslog(KUKA), GreyOrange, Quicktron 등이 있다.

AMR의 개발 및 상용화도 전 세계적으로 활발히 이루어지고 있다. 대부분의 상용화된 실내용 AMR은 2차원 라이다 센서를 기반으로 한 점유격자지도 작성, 최적 경로계획, 장애물 회피 기능이 탑재되어 있다. 주행을 위해 사용하는 지도, 목적지 등의 정보는 물리적으로 확인이 불가능하므로 보다 직관적으로 설치 및 관리를 할 수 있도록 태블릿PC에서도 작업이 가능한 사용자 인터페이스를 제공한다. 해당 인터페이스에서는 지도 작성 및 편집, 주행 방향 및 금지구역 지정 등의 세부적인 설정이 가능하다. Omron은 장애물 감지용 추가 센서 및 천장 조명을 인식하여 위치

추정 성능을 향상시킬 수 있는 카메라 모듈 등을 옵션으로 제공하고 있다. KUKA는 메카넘(mecanum) 휠을 사용하여 전방향으로 이동할 수 있는 주행 플랫폼을 개발하였다. 전방향 이동 능력은 협소한 공간에서의 주행이나 도킹을 위한 모션 생성 등을 용이하게 한다. 이외에도 Mobile Industrial Robots(MiR), Fetch Robotics, Clearpath Robotics 등의 기업들이 유사한 자율주행 기술을 자사의 로봇들에 적용하고 있다.

실외 주행이 가능한 이송 플랫폼도 속속 발표되고 있다. Honda는 2018년 소비자 가전 전시회(CES)에서 실내 플랫폼과 더불어 실외 플랫폼인 3E-D18을 발표한 바 있다. 이는 상위 모듈만 교체하면 건설 현장, 구조 및 탐색, 농업 등 다양한 산업에 적용 가능할 것이 특징이다. Volvo도 2018년에 개발 중에 있는 무인 컨테이너 이송 차량인 Vera를 발표하였다. Vera는 클라우드 기반의 서비스로 연계되며, 우선 규모가 큰 사업장 내에서 운행하는 것을 목표로 하고 있다. 향후에는 이와 같은 실외 주행을 위한 플랫폼들에 대한 수



Amazon 물류창고 로봇



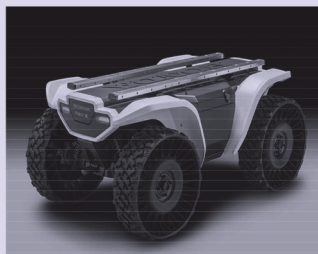
Omron LD



KUKA KMR iiwa



MiR Hook10



Honda 3E-D18



Volvo Vera

그림 1 산업용 물류이송 로봇의 예

요가 점차적으로 증가하여 보다 다양한 모델이 출시 될 것으로 예상된다.

지각(Perception) 측면에서의 자율주행 기술

주행 환경은 보통 정형(structured), 비정형(unstructured)으로 구분할 수 있다. 정형 환경은 명확히 정의되어 있고 변수가 없는 예측 가능한 공간을 의미하며, 완전히 무인화 되어 있는 공장, 물류센터 등이 이에 해당된다. 비정형 환경에는 사람, 조명 변화 등 예측하기 어려운 변수들이 다수 존재하며, 일반 가정환경, 작업자가 공존하는 제조라인 등이 해당된다. 현재 AMR의 적용처는 대체로 비정형 환경에 해당된다. 아직은 로봇과 사람이 공존하는 환경이 대다수이며, 설비 위치 변화 등의 레이아웃 변경이 빈번하게 발생하므로 로봇 주행에 있어서 도전적인 환경이라 할 수 있다. 비정형 환경에서의 안정적인 주행을 위해서는 다양한 센서 정보의 융합을 통한 인지능력의 향상, 각종 예외 상황에 대처할 수 있는 유연한 소프트웨어의 적용이 필요하다.

대부분의 AMR은 사전에 작성된 환경 지도 상에서 로봇의 현재 위치를 실시간으로 추정하고, 목적지까지의 최적 경로를 계획하여 이동하는 방식을 채택하고 있다. 초기에는 로봇의 위치를 추정하기 위해서 환경 내에 2차원 라이다 센서용 반사판, 마커 등의 인공

표식을 설치하고, 각각의 좌표를 지도에 입력하는 방식을 주로 사용하였다. 이 방식은 주행 인프라 구축에 별도의 비용과 시간이 소요되지만, 단시간에 높은 안정성과 정밀도를 갖는 위치추정 시스템을 구성할 수 있었다.

이후에는 로봇을 이동시키면서 획득한 라이다, 비전 센서 정보 등을 기반으로 지도작성과 위치추정을 동시에 수행하는 SLAM(Simultaneous Localization and Mapping) 기술이 AMR에 적용되기 시작했다. 2000년대 후반 iSAM, g2o와 같은 로봇 위치와 표식 등 간의 구속조건을 노드와 엣지로 표현하는 그래프 기반의 SLAM 기술이 더욱 발전하면서 넓은 주행 환경에 대한 정확한 지도를 작성하는 것이 한결 수월해졌다. SLAM 방식은 주행 인프라 구축 비용을 절감할 수 있는 장점이 있지만, 다양한 환경에서 후처리 없이 정확한 지도를 작성하는 부분에서는 높은 수준의 기술을 필요로 한다.

실내 환경은 대체로 평탄한 지면으로 이루어져 있어 로봇의 롤(roll), 피치(pitch) 방향에 대한 자세 변화가 적어 2차원 라이다 센서로 SLAM이 가능하다. 일반적으로는 장애물에 대한 점유격자지도를 작성하는 방식이 사용되고 있으며, 변화가 많이 일어나는 비정형 환경이거나 공간이 너무 넓어 라이다 센서로부터 충분한 정보를 얻기 어려운 경우에는 로봇에 비전 센서를 추가로 장착하여 마커 또는 천장 조명과 같은

인공표식에 대한 인식 정보를 함께 활용하는 사례(예: Omron)도 있다. 실외 환경에는 지면 굴곡 및 다수의 요철이 존재하여 모든 방향으로의 자세 변화가 발생하므로, 안정적인 주행을 위해서 실내에 비해 더 많은 양의 센서 정보가 필요하다. 따라서 실외 환경에서의 SLAM을 위해서는 보통

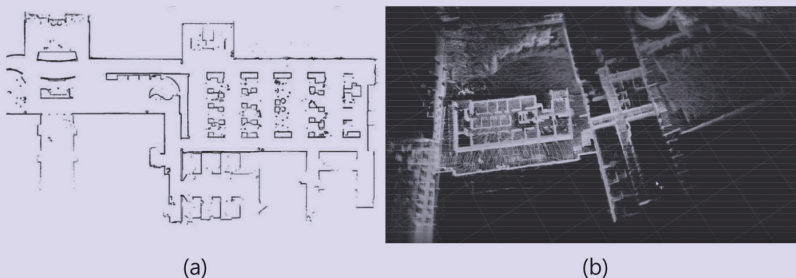


그림 2 SLAM으로 작성한 지도의 예: (a) 2차원 점유 격자지도(Omron, 미국); (b) 3차원 점군 지도(UC Berkeley, 미국)

16채널 이상의 라이다 센서를 기반으로 작성한 3차원 점군(point cloud) 지도를 사용한다.

비정형 환경에서 장기간 동안 신뢰할 수 있는 지도를 유지하기 위해서는 사람, 대차와 같이 이동하는 물체와 책상, 의자 등 잠재적으로 이동 가능성이 있는 물체들을 구분하고, 시간에 따라 변화된 환경을 꾸준히 지도에 반영하는 기술이 필요하다. 현재 사용되고 있는 대부분의 위치추정 방법들은 지도에 의존적이기 때문에 지도와 실제 환경의 일치 여부가 주행 안정성에 큰 영향을 미친다. 경우에 따라서는 벽 앞에 임시로 쌓아둔 자재들이 위치 오차를 증가시켜 의도한 경로로 이동하지 않거나, 위치 오차가 발산하여 복구 불능 상태가 되기도 한다. 이와 같은 문제를 해결하기 위해 각 AMR이 실시간으로 작성한 부분 지도를 별도의 관리 서버로 전송하여 환경 변화를 감지하고, 전체 지도를 지속적으로 갱신하는 시스템을 구축한 사례도 있다.

구축된 지도상에서 최적의 경로를 계산하고 목적지까지 안전하게 이동하기 위해서는 장애물 감지 및 추적 기술도 중요하다. 실내에서 1m/s 내외의 느린 속도로 이동하는 경우에는 장애물이 근접하면 잠시 정지해도 운영에 큰 문제는 없으나, 실외에서 빠른 속도로 이동하는 다른 차량의 움직임을 예측하지 못하면 자칫 큰 사고로 이어질 수 있다. 라이다 센서 정보를 물체 단위로 분할(segmentation)하고, 비전 센서 정보로 사람, 대차, 차량 등 물체의 종류를 구분(classification)하면 보다 세분화 된 이동 모델을 각 장애물에 적용할 수 있다. 각 물체에 알맞는 이동 모델이 설정되면 보다 정확하게 다음 움직임을 예측할 수 있으므로, 교차로 통과 또는 차선 변경 등이 필요할 때 다음 행동을 보다 적절하게 결정할 수 있다.

향후 과제 및 전망

배터리 교체, 정비 등의 이유로 로봇을 임의의 위치

로 이동시킨 후 주행을 재개하려고 하면 전체 지도 내에서 현재 위치를 계산하는 전역위치추정 문제를 풀어야 한다. 고유한 아이디 값을 갖는 인공표식을 사용하거나 GNSS(Global Navigation Satellite System), RTLS(Real-time Locating System), IPS(Indoor Positioning System) 등 별도의 측위 시스템을 사용하는 경우에는 즉시 위치를 결정할 수 있다. 그러나 현재 센서 정보에 위치를 구분할 수 있는 명확한 단서가 없는 경우에는 일정 거리를 이동하면서 정보를 수집하고 후보 지역을 수렴시켜 나아가는 지연(delayed) 방식의 전역위치추정 방법을 주로 사용한다. 지연 방식을 사용하는 경우 비슷한 형태의 생산라인이 늘어서 있는 공장, 선반이 격자 형태로 배치되어 있는 창고 등 반복적인 패턴이 다수 존재하는 환경에서는 센서 정보로부터 위치 결정에 도움이 되는 단서를 찾는 것은 쉬운 일이 아니다. 만약 이로 인해 산업현장에서 주행 정상화에 오랜 시간이 소요된다면 물류가 정체되어 막대한 손해가 발생할 수도 있다. 향후에는 유용한 단서를 제공해줄 수 있는 다양한 시맨틱(semantic) 정보를 지도에 통합하는 기술(예: Imperial College의 SemanticFusion)과 end-to-end 방식의 딥러닝 기술(예: University of Cambridge의 PoseNet) 등이 발전하게 되면 전역위치추정 성공률 향상에 도움이 될 것이다.

실내 또는 실외 환경에서의 안정적인 주행을 위해서 일반적으로 고정밀(high definition) 지도가 사용된다. 특히 3차원 고정밀 지도의 경우 영역 크기가 커질수록 데이터 용량이 기하급수적으로 늘어나게 되며, 관리 및 유지보수 측면에서 높은 비용이 발생하게 된다. 또한 사전에 모델링 된 환경이 아닌 경우에는 주행이 불가능하므로 영역에 제한이 있다. 미래에는 이러한 고정밀 지도에 대한 의존성을 점차적으로 낮추고, 각 공간의 연결을 그래프 구조로 표현하는 토폴로지(topology) 지도 기반의 주행 기술이 발전할 것으로 보인다. MIT에서는 별도의 3차원 지도 정보 없



그림 3 Openstreetmap.org에서 제공하는 토폴로지 지도와 Velodyne HDR-64 라이다 센서 정보를 이용한 주행 및 도로 인식 결과(MIT, 미국)

이 센서 기반의 인지 시스템으로 도로 영역을 인식하고, 목적지까지 자율적으로 주행할 수 있는 MapLite

라는 기술을 발표한 바 있다. 이 분야는 아직 초기 단계이지만, 향후 머신 러닝 등을 기반으로 한 높은 수준의 인지 기술이 뒷받침 된다면 주행 영역의 비약적인 확대 및 비용 감소 측면에서 기술적인 돌파구를 열 수 있을 것이다.

산업용 물류이송 로봇의 관제를 위해서는 보통 별도의 물리적인 서버에서 정보를 수집하여 종합적으로 상황을 판단하고, 각 로봇에 명령을 내리는 중앙집중제어(centralized control) 방식이 선호되어 왔다. 이는 전체 로봇의 실시간 상태 정보를 기반으로 제어가 용이한 이점이 있으나, 로봇 대수 증가에 따른 서버 부하 증가 문제와 서버 오류, 고장 등이 발생하는 경우 소속된 모든 로봇이 영향을 받게 되는 문제를 안고 있다. 따라서 최근에는 Agilox 등의 기업들이 로봇 간 통신 기술을 적용하여 해당 문제를 해결하고자 노력하고 있다. 각 로봇이 독립적으로 최적 경로를 계산하여 이동하고, 잠재적인 장애물 정보를 포함한 주변 상황을 실시간으로 다른 로봇에 공유하면 상황 판단을 도울 수 있다. 또한 특정 설비가 자재 또는 제품 이송을 요청하면, 설비에 인접한 로봇들 간의 합의를 통해 어떤 로봇이 해당 작업을 수행할지를 결정할 수 있다. 향후 5G와 같은 통신 기술이 완전히 보급되어 로봇, 설비 간 연결이 일반화되면 이와 같은 주행 기술에 어떠한 패러다임의 변화가 일어날지 기대된다.

최 성 록 한국전자통신연구원 지능로보틱스연구본부 선임연구원

| e-mail: sunglok@etri.re.kr

이 글에서는 하나의 카메라를 이용해 영상에서 3차원 정보, 즉 공간의 3차원 형태와 카메라의 3차원 위치를 획득하는 기술들에 대해 소개하고자 한다.

사람이 양안을 통해 공간의 3차원 깊이를 인지한다는 것은 일반적인 상식이지만, 양안을 통해 얻은 두 장의 영상은 사람의 3차원 인지의 전부는 아니다. 사실 우리는 한 쪽 눈을 감고 물체를 잡을 수 있고, 벽에 부딪히지 않고 걸을 수 있다. 또 한 장의 영상이나 그림에서 공간이나 물체의 대략적인 3차원 형태도 추측할 수 있다. 이는 단순히 물리적으로 주어진 두 장의 영상 외에 한 장의 영상에서도 그동안 눈과 뇌가 체득한 원근, 물체의 크기와 음영, 상호 가려짐 등의 단서를 통해 물체나 공간에 대한 3차원 인지가 가능하기 때문이다.

이 글에서는 하나의 카메라를 통해 공간이나 물체, 카메라의 3차원 정보를 획득하는 다양한 기술들을 소개하고자 한다. 물론 3차원 거리 데이터를 획득하는 데는 레이저 스캐너(LiDAR), 시간차(Time-of-Flight; ToF) 카메라, 스테레오 카메라 등을 사용할 수 있지만, 이 글에서는 하나의 카메라, 즉 단안 카메라를 이용해 3차원 정보를 획득하는 기법들을 소개한다. 단안 카메라는 스마트폰이나 차량의 블랙박스과 같이 우리 주변에 볼 수 있는 매우 흔한 데이터 획득의 형태이다. 또 카메라는 레이저나 빛의 형태로 에너지를 발산하는 레이저 스캐너와 ToF 카메라와 같은

하나의 카메라에서 얻은 영상을 통해 3차원 인지가 가능하고, 자율차, 로봇, 증강현실 등 다양한 분야에 응용되고 있다

능동형(active) 센서와 달리 수동형(passive) 센서로 간접 문제나 소비전력 관점에서 장점을 갖는다. 또 스테레오 카메라와 달리 매우 작고 경제적으로 제작이 가능하다. 카메라는 3차원 실세계를 2

차원 이미지 평면에 투영(projection)한다. 따라서 일반적인 단안 카메라에서 획득된 영상을 이용해 투영과정 중에 잃어버린 실세계의 3차원 정보를 복원하는 일은 간단하지 않다. 그러나 3차원 정보의 유용성과 필요성 때문에 컴퓨터비전과 머신비전, 로봇공학, 원격탐사(remote sensing), 증강현실 등의 분야에서 영상을 이용한 다양한 3차원 인지 기법들이 연구 및 상용화되고 있다.

한 장의 영상을 이용한 3차원 인지 기법들

한 장의 영상이나 그림에서 3차원 정보를 획득하는 다양한 기법들이 고안되어 왔고, 이러한 기법들은 획득된 3차원 정보의 형태나 밀도에 따라 분류할 수 있다. 우선 영상의 각 픽셀에 카메라로부터의 물체까지 거리 데이터가 기록된 깊이영상(depth image)은 데이터의 밀도가 가장 높은 3차원 인지의 예이다. 반대로 영상에서 관찰된 사람이나 물체의 크기나 위치를

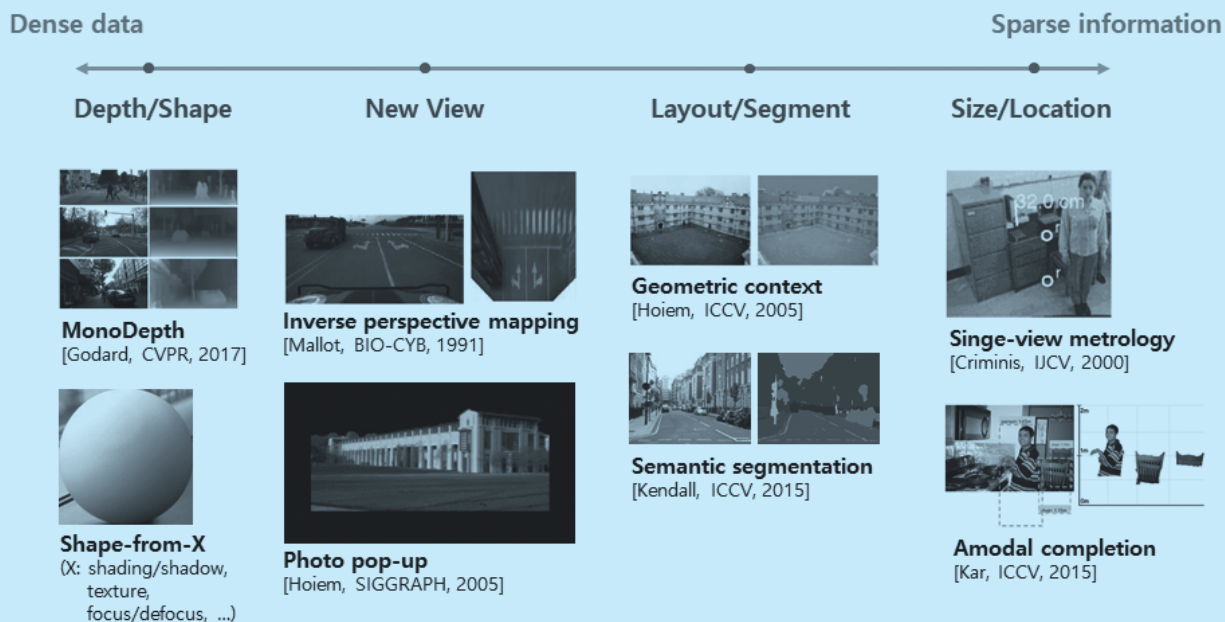


그림 1 한 장의 영상을 이용한 3차원 인지 기술들의 분류(기준: 3차원 정보의 형태)

미터단위로 측정한 값은 가장 밀도가 낮은 3차원 정보의 형태로 볼 수 있다.

3차원 깊이와 모양(Depth and Shape)

컴퓨터비전 분야에서는 shape-from-X라는 이름으로 한 장의 영상에서 물체의 모양(shape)을 알아내는 연구가 오래 전부터 많이 이뤄져왔다. X는 물체의 모양에 대한 단서로서 빛에 의한 물체의 음영(shade)과 그림자(shadow), 물체 표면의 텍스처(texture), 그리고 물체와 카메라의 거리에 따른 초점흐림(defocus)이 대표적인 예이다. 예를 들어 물체 표면에 높낮이 변화가 있다면 외부의 빛에 의해 물체 표면에 밝고 어두운 부분, 즉 음영이 발생한다. 따라서 외부의 빛의 방향을 알고 있거나 추정하게 되면, 발생한 음영을 통해 물체 표면의 높낮이 변화를 추측할 수 있다. 물론 한 장의 영상이 아닌 각각 다른 방향에서 입사되는 조명 환경에서 촬영된 다수의 영상을 통해 더욱 고정밀, 고밀도의 결과를 얻을 수 있다.

또한 한 장의 영상을 통해 3차원 깊이영상을 획득

할 수 있다. 컴퓨터비전 분야에서 2012년 AlexNet의 등장 이후 심층신경망(DNN: Deep Neural Network), 특히 2차원 영상에 사용하기 적합한 컨볼루션신경망(CNN: Convolutional Neural Network)을 이용한 연구들이 많이 이뤄져왔다. CNN을 이용한 3차원 인지의 초기 연구 중 하나가 바로 한 장의 영상에 대응되는 깊이영상을 획득하는 연구였다. 다계층의 CNN을 참값이 있는 대규모 깊이영상 데이터셋으로 교사학습(supervised learning)하여 기존의 연구 결과를 크게 뛰어넘는 결과들을 만들어냈다. MonoDepth는 그 중 가장 인상적인 연구 중 하나이다. 기존 연구의 경우 참값이 있는 대규모의 데이터셋이 필요하다는 점이 큰 걸림돌이었는데, MonoDepth는 참값 대신에 양안 영상을 이용하고, 양안 영상의 상호 일관성(left-right consistency)을 고려한 비교사학습(unsupervised learning)을 수행한다. MonoDepth는 소스코드가 공개되어 있고 학습 과정에 깊이영상의 참값이 불필요하기 때문에 이후 이를 이용한 다양한 응용 연구들이 진행되고 있다.

새로운 시점의 영상(New View)

한 장의 영상을 이용해 새로운 시점의 영상을 합성하는 것도 가능하다. 차량의 1차원 시점의 영상을 탑뷰(top-view)의 도로 영상으로 변환하는 inverse perspective mapping은 고전적인 예이다. 카메라의 자세와 초점거리와 같은 정보가 주어지면 도로가 국부적으로 평평하다는 가정하에 1차원 시점 영상을 탑뷰의 도로 영상으로 재투영할 수 있다. 합성된 탑뷰의 영상은 차선과 같은 도로 위의 물체의 분석을 더욱 용이하게 해준다. Automatic photo pop-up은 영상의 각 영역을 지면, 지면 위의 물체, 그리고 하늘의 세 가지 영역(ground, vertical, sky)으로 구분하고, 영상을 다른 시점에서 재구성한다. Photo Pop-up 기술도 공간의 3차원 모델 생성을 위해 카메라의 자세와 초점거리가 필요한데, 세 가지로 구분된 간략한 3차원 모델을 바탕으로 하기 때문에 영상 내 지평선이나 대략적 고정된 초점거리와 카메라 높이로도 충분히 그럴듯한 데모를 보여준다.

영상의 레이아웃(Scene Layout and Segments)

영상의 각 픽셀의 깊이/거리 정보 대신 각 영역을 그 의미별로 분리하는 기술은 3차원 인지의 기본이 된다. 초창기 연구인 Photo pop-up의 기하학적 문맥(geometric context) 기술은 영상에서 얻을 수 있는 색상과 에지와 같은 여러 가지 영상 특징과 SVM(Support Vector Machine) 분류기를 이용해 영상의 각 부분들을 세 가지 영역(ground, vertical, sky)로 분리한다. 이후 딥러닝의 발전과 더불어 이 분야의 기술들은 의미적 분할(semantic segmentation)이라는 이름으로 많은 연구와 응용이 이뤄지고 있다. 그 중 SegNet은 특징 추출과 분류기가 혼합된 CNN의 인코더와 디코더 구조를 통해 도로, 인도, 차량, 사람, 나무, 건물, 하늘 등 다양한 의미 영역을 분리할 수 있다. 또한 최근의 Mask R-CNN 기술은 물체 검출 기술 분야에 사용되던 기존의 R-CNN 구조를 확

장하여 기존의 물체 검출이나 영역 분할을 융합한 개별 물체별 분할(instance segmentation)이 가능함을 보였다. 즉 기존의 영역 분할이 모여 있는 사람들을 하나의 영역으로 분리한 반면, Mask R-CNN 기술은 해당 영역을 개별 사람별로 구분한다.

물체의 미터단위 위치 및 크기(Metric Location and Size)

한 장의 영상이 주어졌지만, 영상에 보이는 물체의 미터단위의 위치나 크기를 측정하는 것이 가능하다. 전술한 기술들이 깊이영상이나 새로운 시점의 영상, 의미적 분할 결과와 같이 주어진 영상의 크기에 준하는 결과를 주는 반면, 물체의 위치와 크기와 같은 정보는 상대적으로 매우 밀도가 낮은 형태이다. 단일-뷰 계측법(single-view metrology) 기술은 지평선, 소실점, 그리고 평행한 평면들과 같은 기하학적 정보와 제약조건을 이용하여 카메라 정보가 주어지지 않은 상황에서도 미터단위의 거리 측정과 카메라 좌표계를 기준으로 한 물체의 미터단위 위치를 측정할 수 있음을 보인다. 사실 카메라의 자세와 초점거리와 같은 정보가 사전에 주어진 경우, 물체의 미터단위 위치와 크기는 훨씬 간편하게 측정이 가능한데, 이는 여러 가지 측정 및 계측장비에 사용되는 머신비전(machine vision) 기술에서 많이 적용되는 방법이다. Amodal completion 기술은 영상에서 가려짐이나 잘림에 의해 물체가 모두 관찰되지 않은 경우에도 대략적인 미터단위의 크기와 상대적인 거리를 추정하는 흥미로운 기술이다. CNN을 통해 물체 검출을 교사학습할 때 물체의 경계가 되는 상자(bounding box)를 화면에서 잘려지거나 가려진 부분을 포함하여 이를 참값으로 학습시키는 것이 핵심적인 아이디어이다.

여러 장의 영상을 이용한 3차원 인지 기법들

컴퓨터비전 분야에서 여러 장의 영상으로 3차원 모

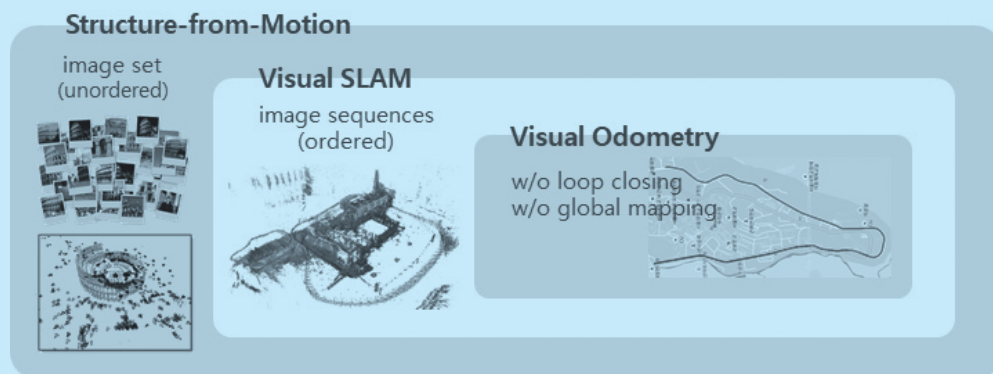


그림 2 여러 장의 영상을 이용한 3차원 인지 기술들의 관계(기준: 입력 영상 특성)

SfM과 visual SLAM, visual odometry는 각각 서로 다른 문제가 아니고, 유사한 3차원 인지를 수행하지만 입력 영상에 대한 특성 및 가정이 다르고, 서로 포함되는 관계를 가진 문제들이다.

델을 재구성하고 카메라의 위치를 알아내는 기하학 이론과 방법론은 그 태동기부터 지속적으로 연구되어 왔다. 카메라 투영의 수학적 모델을 바탕으로 에피폴라 기하학(Epipolar Geometry), 다중시점 기하학(Multiple-view Geometry)라고 불리는 이론들이 체계화되었고, 이를 바탕으로 한 Structure-from-Motion(SfM), visual SLAM, visual odometry와 같은 세 가지 형태의 문제가 잘 정의되어 자율차, 로봇, 증강현실과 같은 응용과 맞물려 지속적으로 연구/적용되고 있다. 세 가지 문제 유형은 유사한 3차원 인지를 수행하지만, 입력된 영상의 특성에 따라 구분할 수 있다. 우선 SfM에 입력되는 영상은 영상 간에 순서가 없는 영상들의 집합이고, visual SLAM은 보통 하나의 카메라에서 촬영된 비디오, 즉 영상 간에 시간적 관계가 주어진 경우이다. 따라서 visual SLAM은 영상 사이의 시간적 연결 관계를 적극적으로 활용하여 SfM보다 빠르고 정확하게 문제를 해결하는 것이 핵심이 된다. Visual SLAM은 입력된 영상들 중 이전에 촬영된 위치에 가까운 위치(loop)에서 다시 촬영된 영상이 있는 경우를 고려하고, visual odometry는 카메라 궤적의 겹침(loop)을 고려하지 않는 보다 단순화된 문제이다. 따라서 visual SLAM은 visual odometry와 비교하여 재방문(loop closure)을 검출하고 이를 3차원 인지에 고려하는 과정이 추가된다.

Structure-from-Motion

공간의 3차원 구조(structure)와 영상이 촬영된 각기 다른 카메라의 위치(motion)를 동시에 추정하는 SfM 기술은 일반적으로 bundle adjustment라고 불리는 최적화 문제로 접근한다. 카메라의 위치를 알고 있는 경우 공간의 3차원 구조를 추정이 가능하고, 공간 구조를 알고 있는 경우 이를 관찰한 카메라의 위치를 추정할 수 있다. 그러나 이 두 가지가 미지수인 SfM의 문제는 '닭과 계란의 문제(chicken and egg problem)'와 같이 두 미지수가 서로 얽혀있는 문제이다. 일반적으로 이 꼬여있는 문제는 좋은 해를 찾기에 용이한 영상을 하나씩 문제 범위에 포함시키며, 좋은 초기값을 유지하며 점층적(incremental)으로 최적화를 수행하며 해결한다. 이러한 방법 중 가장 유명한 Bundler는 인터넷을 통해 수집한 여러 관광지의 대규모 영상들에서 해당 관광지의 3차원 외형을 복원하는데 적용되어 그 유용성과 확장성(scalability)을 입증하였다. 이후 Visual SFM과 COLMAP은 이 기술을 시간복잡성(time complexity)과 기능성/사용성 측면에서 한 단계 개선하였다.

Visual SLAM

로봇공학에서 많이 연구되던 SLAM(Simultaneous

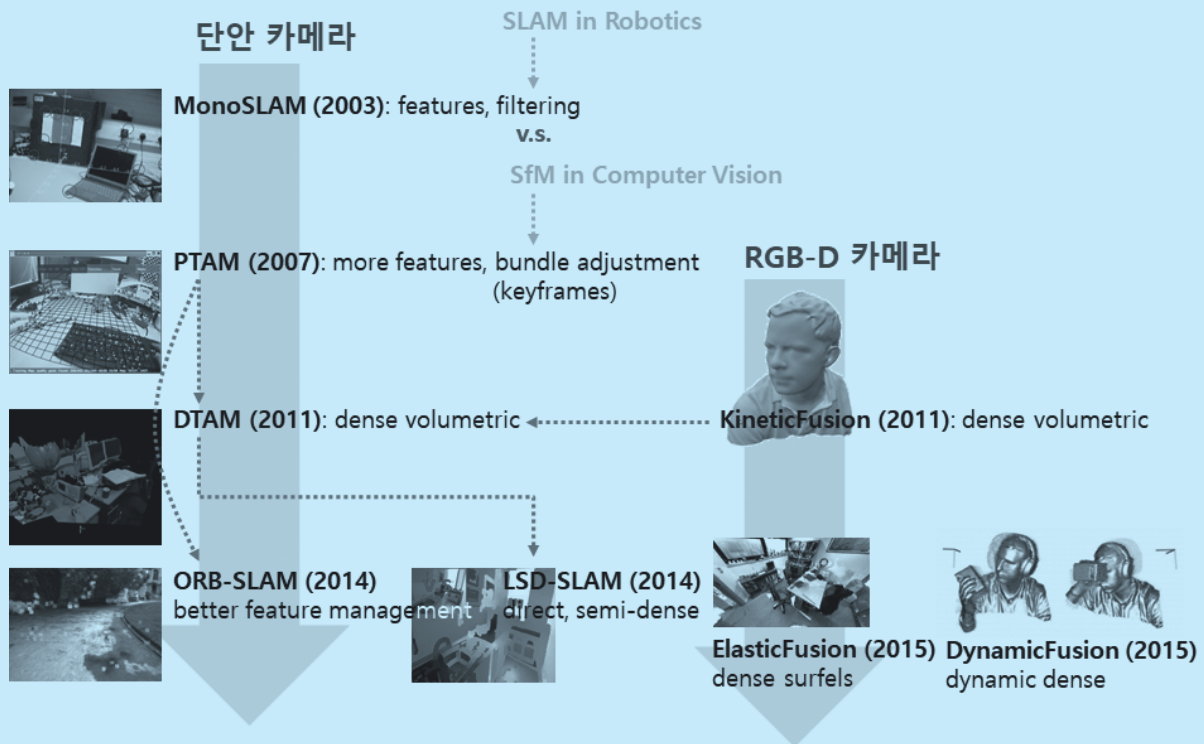


그림 3 Visual SLAM 기술의 계보 및 특징

Localization and Mapping) 문제는 컴퓨터비전의 SfM 문제와 동일하게 카메라의 위치와 공간의 형태를 동시에 추정하는 문제이다. 전술한 바와 같이 visual SLAM은 SfM과 달리 주어진 영상들 사이에 시간적 관계가 있고 이를 효과적으로 이용하는 것이 핵심이다. 따라서 visual SLAM의 초기에는 로봇공학에서 시간적 관계를 가진 데이터를 다루는데 널리 사용되는 칼만필터나 입자필터(particle filter)와 같은 베이지안(Bayesian) 필터링을 통해 SLAM 문제를 해결하였다. MonoSLAM은 이러한 접근의 대표적인 연구로 확장칼만필터(EKF)를 이용하여 수백개의 영상 특징점으로 표현할 수 있는 공간에 적용되었다. 그러나 필터링을 이용한 접근은 높은 시간복잡

Structure-from-Motion, Visual SLAM, Visual Odometry는 모두 공간의 3차원 형태와 카메라의 위치를 추정하는 것을 목적으로 하고, 입력 영상의 특성과 가정이 다른 포함 관계를 갖는 기술이다

도와 마르코프(Markov) 가정과 같은 제약으로 인해 확장성과 성능의 한계가 있었다.

향상된 컴퓨팅 성능에 힘입어 SfM에 사용되던 bundle adjustment의 비선형 최적화 기법이 실시간 visual SLAM에도 적용되었다. PTAM(Parallel Tracking and Mapping)은 위치추정 문제

와 공간추정 문제를 병렬적으로 수행되도록 분리하고, 이를 각각 단시간(short-term) 최적화와 장시간(long-term) 최적화로 접근하였다. 또 마르코프(Markov) 가정에서 벗어나 현재 영상 프레임뿐만 아니라 이전의 키 프레임들을 고려하여 정확도와 안정성도 개선하였다. ORB-SLAM2는 PTAM의 접근을 따르는 가장 정점에 있는 기술로 대형 공간과 장기간

동작, 그리고 RGB-D와 스테레오 등 여러 가지 형태의 카메라를 지원하는 것이 특징이다.

2011년 깊이영상이 센서에서 바로 주어지는 Microsoft의 Kinect의 등장은 많은 연구에 RGB-D 카메라의 사용을 촉발하였고, KinectFusion이나 ElasticFusion, DynamicFusion과 같은 RGB-D SLAM 기술이 탄생하게 하였다. 이에 영향을 받은 DTAM(Dense Tracking and Mapping) 기술은 영상 특징점이 아닌 영상의 대부분의 영역을 고밀도로 추적하고 매핑하는 RGB-D SLAM의 접근(direct method)을 단안 카메라를 적용한 visual SLAM 기술이다. 또 이러한 접근에 기반을 둔 LSD-SLAM 또한 bundle adjustment의 재투영 에러(reprojection error)가 아닌 광도 에러(photometric error)를 최소화하는 방법을 따랐다. 아직까지 영상 특징점을 이용한 기존 방식(feature-based method)과 영상의 대부분을 직접 이용하는 방식(direct method)은 서로의 장단점으로 인해 기술적 우위에 대한 논란에 서있다.

Visual Odometry

Visual odometry는 visual SLAM에서 재방문(loop closure)를 검출하고 이를 추정에 반영하는 부분이 제외된 단순화된 문제이다. 따라서 대부분의 visual SLAM의 접근이 visual odometry 기술에 그

대로 적용되었다. 영상 특징점과 bundle adjustment를 이용한 기존 접근과 영상의 대부분의 영역을 바로 이용하는 최근 접근 모두 visual odometry에서 시도되었다. 그러나 visual SLAM과 달리 재방문을 검출하지 않는 점은 후자의 최근 접근 방법에 유리한 측면이 많기 때문에 후자의 방법을 따르는 visual odometry 기술이 최근 많이 시도되고 있다. SVO(Semi-direct Visual Odometry)는 전자와 후자의 접근의 장점을 융합한 기술이고, DSO(Direct Sparse Odometry)는 후자의 접근을 크게 개선한 최신 기술이다.

하나의 카메라를 이용한 다양한 형태의 3차원 인지 기술들을 살펴보았다. 입력 영상의 개수에 따라 한 장의 영상을 이용한 방법과 여러 장의 영상을 이용한 방법을 살펴보고, 한 장의 영상을 이용하는 방법은 결과가 되는 3차원 정보의 형태에 따라서 살펴보았다. 여러 장의 영상을 이용하는 방법은 입력 영상의 특성 및 가정에 따라 구분하여 살펴보았다. 카메라는 우리 주변에서 쉽게 접할 수 있는 값싸고 많은 정보를 제공하는 센서이고 이를 이용한 3차원 인지는 많은 분야에 큰 파급 효과를 낼 수 있다. 이 글을 통해 이러한 인식과 공감대가 공유되길 기대한다.

김 아 영 한국과학기술원 건설 및 환경공학과 조교수

| e-mail : ayoungk@kaist.ac.kr

이 글에서는 도시 내 자율 주행 로봇이 도시를 인식하고 판단하는 방법인 센싱 기술과 인식 기술에 관한 설명을 다룬다.

자율 주행의 시대를 맞아 도심에서 로봇 기술이 다양하게 적용되고 있다. 로봇은 탑재된 센서를 통해 주변 공간을 인식하고 판단하는데, 여러 가지 기술 중에서도 주변 환경을 파악하여 지도를 만들고 자신의 위치를 파악하는 기술을 SLAM 기술이라고 한다. 이 SLAM 기술은 자율 주행과 임무 수행 등의 작업을 위해서 선행되어야 하는 기술이기에 그 중요도가 크며, 그 핵심에는 로봇 탑재 센서로부터 공간을 인식하는 기술이 존재한다. 그렇다면 로봇은 도심 공간을 어떻게 센싱하고 파악하는 것인지 이 글을 통하여 설명하고자 한다.

이렇듯 변화가 다양한 도시의 공간을 표현하는 것은 로봇의 안정적인 도시 내 주행을 위해 중요하다

로봇의 인식과 재인식

이 과정에서 가장 중요한 점은 로봇의 “인식(perception)”과 “재인식(recognition)”으로 볼 수 있다. 어떤 공간을 처음으로

이동할 때 로봇은 인식을 통해 주변 환경에 대한 지도(map)를 작성한다. 작성하는 지도에는 안타깝게도 불가피한 동적 물체나 순간적인 인식 오류 등이 포함될 수 있다. 이러한 오류를 보정할 수 있는 것이 로봇의 재인식 기술이다. 이러한 재인식 기술은 로봇에 있어 매우 중요한데, 어떤 공간을 재방문하거나 이전에 인식되었던 공간이 다시 인식될 때 그 간극에 존재하



그림 1 도심의 점군 데이터 예(Complex Urban LiDAR Dataset, KAIST, 한국)



그림 2 다양한 라이다 센서(좌로부터 벨로다인 사의 VLP-16, 이베오 사의 LUX HD, 쉐너지 사의 S3)

을 가정하여 예상되는 풍광을 만들어 낼 수 있다면 다양한 상황에 대처할 수 있는 인식이 가능하다.

라이다(LiDAR) 센서를 통한 도시 공간 인식

로봇이 사용하는 센서 중 가장 널리 쓰이는 센서는 단연코 카메라일

것이다. 단안 또는 양안 카메라를 통한 위치 추정 및 지도 작성법은 이미 널리 알려져 있으며, 카메라는 다른 센서에 비해 가격이 저렴하여 연구자들에게 널리 사랑받는 센서이다. 하지만 이런 카메라에 치명적인 단점이 존재하는데 지난번 테슬라의 자율 주행 차량 사고에서와 같이 유사 색상에 따른 인식의 착오가 발생하게 되면 자칫 중대한 오류를 일으킬 수 있다.

가격은 카메라보다 고가이지만 최근 로봇 분야에서 각광받는 센서가 바로 라이다(LiDAR) 센서이다. 우리가 흔히 알고 있는 레이더(Radar)센서와 달리 빛(light)을 측정하는 방식의 라이다(LiDAR) 센서는 전파를 사용하는 레이더보다 직진성이 좋은 빛의 특성으로 인하여 주변의 환경 데이터를 풍성하게 제공한다. 3차원 정보가 2차원으로 투영된 이미지와 달리 라이다 센서는 3차원 공간을 3차원 점들로 표현한다. 이를 점군(point cloud)이라고 한다. 점군 데이터는 그 자체로 3차원을 표현한다. 광조건에도 덜 민감하며 낮/밤에 영향을 받지 않고 데이터를 제공한다.

이렇듯 변화가 다양한 도시의 공간을 표현하는 것은 로봇의 안정적인 도시 내 주행을 위해 중요하다. 그렇다면 로봇이 안정적으로 도시 내에서 이동하기 위해서는 어떻게 해야 할까? 첫 번째로 변화에 조금 덜 민감한 센서를 사용하는 방법이 있다. 카메라 영상의 경우 조도에 영향을 많이 받기 때문에 광조건이 크게 변하는 환경이라면 카메라가 아닌 다른 센서를 사용해서 인식을 수행하는 것이다. 또 다른 방법은 로봇이 인식하기 좋은 공간의 특징을 정의해 주는 것이다. 공간을 대표할 수 있는 특정값을 정의하여 한 공간의 지문(fingerprint)처럼 사용하면 공간의 재인식이 조금 더 강건해질 수 있다. 세 번째로는 변화하는 다양한 상황을 모두 다 기억하는 방법이 있다. 어떤 공간에 대해서 봄, 여름, 가을, 겨울, 낮, 밤 등의 다양한 모습을 모두 기억한다면 다양한 상황에서도 인식이 가능하다. 만일 눈 오는 도시를 본 적이 없는 경우에도 눈이 내린 상황

를 가정하여 예상되는 풍광을 만들어 낼 수 있다면 다양한 상황에 대처할 수 있는 인식이 가능하다.

로봇이 사용하는 센서 중 가장 널리 쓰이는 센서는 단연코 카메라일 것이다. 단안 또는 양안 카메라를 통한 위치 추정 및 지도 작성법은 이미 널리 알려져 있으며, 카메라는 다른 센서에 비해 가격이 저렴하여 연구자들에게 널리 사랑받는 센서이다. 하지만 이런 카메라에 치명적인 단점이 존재하는데 지난번 테슬라의 자율 주행 차량 사고에서와 같이 유사 색상에 따른 인식의 착오가 발생하게 되면 자칫 중대한 오류를 일으킬 수 있다.

가격은 카메라보다 고가이지만 최근 로봇 분야에서 각광받는 센서가 바로 라이다(LiDAR) 센서이다. 우리가 흔히 알고 있는 레이더(Radar)센서와 달리 빛(light)을 측정하는 방식의 라이다(LiDAR) 센서는 전파를 사용하는 레이더보다 직진성이 좋은 빛의 특성으로 인하여 주변의 환경 데이터를 풍성하게 제공한다. 3차원 정보가 2차원으로 투영된 이미지와 달리 라이다 센서는 3차원 공간을 3차원 점들로 표현한다. 이를 점군(point cloud)이라고 한다. 점군 데이터는 그 자체로 3차원을 표현한다. 광조건에도 덜 민감하며 낮/밤에 영향을 받지 않고 데이터를 제공한다.

라이다 센서는 쉽게 말해 레이저 포인터와 같다고 생각할 수 있다. 하나의 레이저마다 센서로부터 해당 물체의 표면까지의 거리를 측정하는데, 이 레이저 여러 개를 모아서 부채살처럼 한 줄로 배치시켜 공간을 스캔할 수도 있으며, 부채살처럼 배치한 레이저 다발을 360도 회전시켜서 3차원의 데이터를 얻기도 한다. 이처럼 다양한 라이다 센서의 종류는 상단의 그림과 같다. 다양한 라이다 센서는 주변 환경을 인식하는 데

사용될 뿐 아니라 3차원 지도 제작을 위해서도 사용되는 등 그 활용 범위가 폭넓다.

라이다의 가장 큰 단점인 가격 역시 추후 SSD 기반의 센서가 널리 도입된다면 개선될 예정이니, 향후 라이다 기반의 인식과 자율 주행 기술의 전망도 긍정적이다. 라이다 단독 사용에 국한되지 않고, 카메라와의 센서 융합을 통해 상호 단점을 보완하고 더욱 풍부한 정보를 제공할 수 있는 것도 주목할 만한 부분이다. 더 나아가 라이다 데이터로 기구축된 3차원 점군 데이터상에서 전혀 다른 센서인 카메라 정보의 매칭을 통해 재인식을 수행하기도 한다. 일례로 지도와 같은 데이터베이스 작성에는 조금 더 강건하고 고가인 센서를 사용하고, 그 지도상에서의 주행에는 카메라와 같은 상대적으로 저가의 센서를 사용하는 방식으로 이중 센서 간의 융합을 통해 단일 센서의 성능을 개선한 연구 사례가 있다.(KAIST, 한국)

도시 공간 인코딩

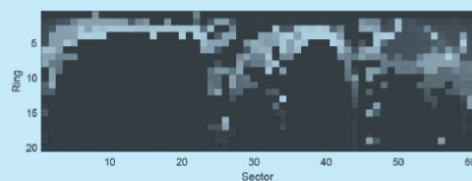
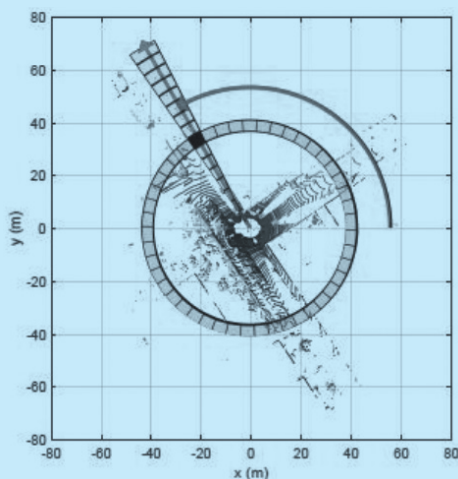
한 공간을 대표하는 어떤 값을 찾아낼 수 있다면 공간 인식과 재인식 능력을 크게 향상시킬 수 있다. 공간에도 그 공간을 대표하는 “지문”과 같은 특징을 규정할 수 있다면 어떤 공간이 방문해 본 곳인지 판단할 때 공간 지문을 비교해 보면 손쉬울 것이다. 사람의 지문처럼 개인의 독창적인 패턴

공간에도 그 공간을 대표하는 “지문”과 같은 특징을 규정할 수 있다면 어떤 공간이 방문해 본 곳인지 판단할 때 공간 지문을 비교해보면 손쉬울 것이다.

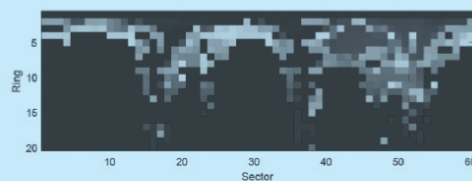
을 정의하려는 노력은 최근 여러 곳에서 진행되어 왔다. 그중 TOMTOM이라는 해외 지도 제작 업체는 Road DNA라는 이름으로 도로 위의 한 지점을 표현한다. Road DNA는 도로의 양옆에 나타나는 건물의 실루엣을 하

나의 이미지처럼 투영하는 방식으로 만들어진다. 이러한 “지문”의 제작을 위해, 앞에서 소개한 라이다와 같은 센서로 제작한 3차원 지도를 활용한다면 조금 다른 방법으로 공간의 특징값을 정의할 수 있다.

도시 공학에서는 한 공간이 얼마나 열려 있는지를 정의하기 위한 노력을 기울여 왔다. 도시 공간의 열림(openness)은 보안, 치안, 조망권 등과 관련하여 연구되어 왔으며 그 측정을 위해서 기존 연구들은 주로 모델이나 항공사진을 기반으로 한 분석을 진행해 왔다. 그러던 것이 최근 3차원 지도가 등장하고 점군 데이터가 등장함에 따라 점군 데이터를 활용한 3차원의 공간 분석이 수행되고 있다. 도시의 공간 분석 기법에 착안하면 한 공간의 특징점을 규정하는데 이러한 공간의 열림의 정도를 사용할 수 있다. 다시 말해 주변 건물의 skyline을 그려서 한 공간의 특징 패턴으로



(a) The query scan context (3280th scan, KITTI00)



(b) The detected scan context (2345th scan, KITTI00)

그림 3 (좌) 도시 공간 함축 패턴 생성 과정 / (우) 다른 시점, 같은 공간에서 취득된 패턴 (KAIST, 한국)



그림 4 (좌) Oxford 연구진이 GAN을 통해 만들어낸 이미지(Porav et. al., 2018), (우) 호주 연구진에 제공한 같은 공간에서 다양하게 취득되는 영상 정보(Chen et al., 2017)

정의한다면, 로봇이 skyline으로 만든 사진을 보고 이곳이 도시 공간에서 어디에 해당하는지를 유추하는 데 사용할 수 있다.

주변 건물의 분포로 정의된 이 skyline의 특징 패턴은 잘 변하지 않는다. 한 지점을 기준으로 주변 환경이 완전히 변하는 것이 아니라면(도시의 특정 구역이 모두 소실되는 경우나 황무지에 도시가 새로 건설되는 경우) 절반 정도의 변화가 발생하여도 나머지 패턴을 사용하여 인식이 가능하다. 더 나아가 단 하루의 데이터에서 제작한 패턴은 1년의 시간이 흘러도, 계절의 변화나 주변 동적 물체의 변화가 있더라도 크게 변하지 않는다. 계절과 시간의 변화에 크게 바뀌지 않는 건물의 skyline을 인코딩하였기 때문이다.

공간의 가상 기억

라이다와 같이 환경 조건에 덜 민감한 센서를 사용하기 어려운 상황이라면 로봇은 도시 공간에서 어떻게 안정적으로 주행을 할 수 있을까? 영상 정보를 활용한 장기적 자율 주행(long-term navigation)이라는 연구 토픽하에 다양한 연구자들의 활발한 연구가 진행 중이다. 특히, 최근 대두된 딥러닝 기법의 도움으로 그간 SLAM 분야의 어려움으로 보였던 환경 변

화에 대한 강건성 확보는 큰 돌파구를 마련하였다.

CNN으로 영상이 포함하는 객체를 분류하는(classification) 문제는 이미 컴퓨터 비전 분야에서 널리 적용되어 그 효과를 입증한 바 있다. 초반에 ImageNet에서 객체를 분류하던 연구자들은 이제 공간 재인식(place recognition)으로 관심을 옮기고 있다. 기존에 자동차, 개, 고양이 등의 물체를 구분하는 것에 중점을 두었던 ImageNet 데이터셋과는 달리 공간을 위한 데이터와 그 데이터에 대한 학습의 중요성을 강조한 호주 연구진의 연구는 딥러닝이 어떻게 다양하게 변화하는 공간을 추상화하고 일반화할 수 있는가에 대한 고민을 담고 있다. 네트워크를 통해 학습이 이루어진 중간 layer에서 특징점을 추출한다든가 공간의 인식을 위한 집중이 어느 곳에서 활발한지를 이용하거나 동적 물체를 예측하는 과정을 통해서 공간 인식 능력을 향상시키려는 시도들이 있었다. 이러한 연구 모두 딥러닝이 다양한 데이터로부터 강건한 일반화 능력을 보여준다는 점에 착안하여 로봇 기술에 적용한 예이다.

모든 환경 변화에 강건한 특징점(descriptor)을 학습을 통해 구하려는 시도와는 조금 다른 접근 방식도 최근 소개되고 있다. GAN이나 style transfer의 연구를 공간 인식에 적용하려는 시도가 그것이다. 이러한

연구에서는 학습을 통해 GAN이나 style 변환을 적용하여 실제로 본 적이 없는 풍광도 만들어 낼 수 있는 것이다. Oxford 대학의 Paul Newman 그룹에서는 이러한 변화에 강건한 인식 기법에 대해서 활발히 연구를 수행하고 있다. Oxford는 RoboCar 데이터셋(Oxford대학, 영국)이라는 대규모 데이터셋을 기반으로 장기간의 환경 변화를 학습하고 경험하지 못한 풍광을 만들어내는 데 성공했다.

딥러닝을 통해 로봇이 보게 될 도시

딥러닝은 그간 로봇이 공간 인식 문제에서 해결하지 못했던 난제들에 고무적인 답안을 주고 있다. 딥러닝 기술이 발전하고 로봇 분야에 널리 적용되면서 앞으로는 영상뿐 아니라 로봇이 주로 사용하는 다양한 센서에 딥러닝이 적용될 것이다.

도시를 인식하기 위한 문제는 도시라는 공간을 어

떻게 학습하는지로 전환되고, 향후에는 지도의 형태도 학습이 완료된 네트워크로 존재하게 되지 않을까? 어쩌면 미래에 로봇이 도시를 인식할 때 “서울시 지도”가 아닌 학습된 “서울시 공간네트워크”를 사용하여 하는 날이 올지도 모른다.

감사의 글

위 글에서 소개된 연구 내용은 네이버 랩스 사의 LiDAR 및 Camera 센서를 사용한 3차원 SLAM 맵상의 Localization 기법 연구 과제와 산업통상자원부 산업기술혁신사업의 지원(과제 번호:10051867)을 받아 수행되었습니다.

이 덕 진 군산대학교 기계융합시스템공학부 부교수

| e-mail: deokjee@kunsan.ac.kr

이 글에서는 드론의 자율비행 및 충돌회피를 위한 기법 중 기존의 규칙기반(Rule Based) 제어기법(Control Laws)을 적용하는 대신 센서로 획득된 정보 및 데이터를 직접 이용하여 스스로 인식 및 판단을 내리고 모션을 제어할 수 있는 심층학습 및 심층강화학습 응용에 대해 소개하고자 한다.

기존의 드론 및 무인비행체의 자율비행은 기본적으로 사전에 미리 정해진 임무수행을 위하여 경로계획(path planning), 센싱(sensing), 측위(localization), 그리고 법칙(rules) 기반 충돌회피 및 제어(control)를 수행하게 된다. 경로계획의 경우 사전 임무수행에 적합한 전역경로(global path)를 계획하고, 센싱 정보를 이용하여 위치를 계산하며 필요시 지도(map)를 작성하여 상황에 맞게 지역경로(local path)를 생성한다. 또한 장애물을 감지할 경우 미리 정해진 규칙에 따라서 충돌회피 명령을 수행하게 되며, 제어기(controller)를 통하여 생성된 경로를 추종할 수 있도록 가속, 감속 및 조향 등 동작제어(motion control)를 수행하게 된다. 동적장애물(moving obstacle)과의 충돌회피 및 예측하지 못했던 주위 환경의 변화로 생성되는 불확실성(uncertainties)에 대처하기 위한 강건한 해법을 제시가 필요하며, 이러한 복잡한 문제를 해결하기 위해서 새로운 심층학습(deep learning) 및 심층강화학습(deep reinforcement learning) 기반 인공지능 기술을 소개하고자 한다.

인공지능 기반 환경 인식 및 자율비행 기술 현황

무인비행체(unmanned vehicles)의 진보된 충돌

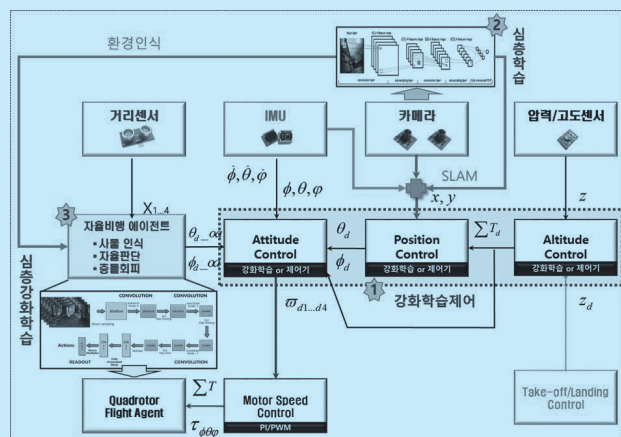


그림 1 드론 자율비행을 위한 인공지능 기반 환경인식, 충돌회피 및 모션제어 구성도

회피 및 자율비행(autonomous flight)을 위한 기법은 크게 ① 강화학습 기반 자세안정화 및 제어, ② 심층학습 기반 환경 인식, ③ 심층강화학습 기반 충돌회피 및 모션제어로 세 가지 방향으로 적용되어 왔다. (그림 1)

첫 번째, 강화학습 기반 무인비행체의 제어 및 자세안정화 기법은 기존의 제어기가 물리적 법칙(physical laws)을 적용한 수학적 모델을 필요로 하는 반면, 축적된 비행체의 모션 및 궤적 데이터를 기반으로 수학적으로 제시하기 어려운 모델을 간접적으로 찾아내고 이를 바탕으로 직접 제어하는 방식이다. 본 강화학

습기반 제어 기법은 스탠포드 대학의 Andrew Ng 교수 연구팀이 최초로 성공적인 결과를 제시하였다. 본 기법에서는 파일럿 조종사가 원격으로 헬리콥터를 조종하고 획득된 데이터를 지도 학습(supervised learning) 기법을 이용 무인 헬기의 복잡한 동역학적 모델을 찾아내고 강화학습 기반 제어를 통해서 자세를 안정화하고 자율비행이 가능한 기법을 제시하였다(그림 2). 이러한 강화학습 기반 직접제어 방식은 환경인식시스템을 이용한 경로생성 이후 강화학습을 이용한 제어알고리즘을 개별적으로 설계해야하는 방식이다.



그림 2 지도학습 기반의 모델 추정과 강화학습 제어를 통한 헬리콥터 자율비행 (Stanford University, 2007)



그림 3 End-to-End 방식 심층학습 기법을 이용한 드론의 숲속 경로추종 및 자율비행(스위스 대학, 2015)

두 번째 기법은 개별적 환경인식, 경로 생성 및 제어 설계과정을 하나로 통합한 end-to-end 방식의 심층학습(deep learning) 기반 자율비행 기법이다. 이 기법은 엔비디아(Nvidia) 사에서 세 개의 전방 카메라 영상 데이터와 심층학습(deep learning) 기법을 적용하여 자동차의 자율주행에 성공적으로 적용된 기법으로, 이후 end-to-end 방식 심층학습을 이용한 드론의 자율비행(autonomous navigation)에 적용되었다. 스위스 Dalle Molle 인공지능연구소(IDSIA)와 취리히(Zurich) 대학이 공동으로 end-to-end 심층학습을 이용하여, 드론이 숲속 산책로(forest trail)를 스스로 인지하고 산책길을 따라서 자율로 비행하는 실험 결과를 제시하였다(그림 3). 본 기법에서는 심층학습 기법을 학습을 위한 데이터 획득을 위하여 헬멧에 세 대의 전방 카메라를 좌, 우, 센터 방향으로 머리에 장착하고 숲속 길을 걸어가면서 카메라에서 들어오는 영상과 그 순간의 숲속 길의 방향각도(heading angle)를 함께 데이터 셋을 저장하여 활용하였다. 획득된 데이터 셋을 활용하여 합성곱신경망(Convolutional Neural Networks) 학습시키고 학습된 심층학습 모델을 이용하여 카메라 영

상에서 들어오는 데이터를 판단하여 드론이 숲속 길을 따라가도록 모션을 제어하는 기법이다. 이러한 end-to-end 방식은 인지부터 제어 명령까지 동시에 학습하고 모션을 제어할 수 있는 효율적 구조를 가지고 있어 많은 관심을 불러일으켰지만, 심층학습 모델을 훈련시키기 위하여 직접 관련 데이터를 수집해서 지도학습(supervised learning)을 해야 하기 때문에, 다양한 시나리오가 발생할 수 있는 환경에 있는 무인 이동체의 경우 모든 상황을 대처하기에는 어려운 한계를 보여줄 수 있다.

심층강화학습 기법 소개

End-to-end 심층학습 기반 모션제어 방식은 지도 학습된 데이터를 필요로 한다. 따라서 지도학습 되지 않은 시나리오의 경우 대처가 되지 않는 경우가 발생할 수 있는데, 이러한 단점을 심층강화학습(deep reinforcement learning)기법을 통하여 극복할 수 있다. 심층강화학습 기법은 2016년 구글 딥마인드(Google DeepMind)에서 소개된 인공지능기법으로,

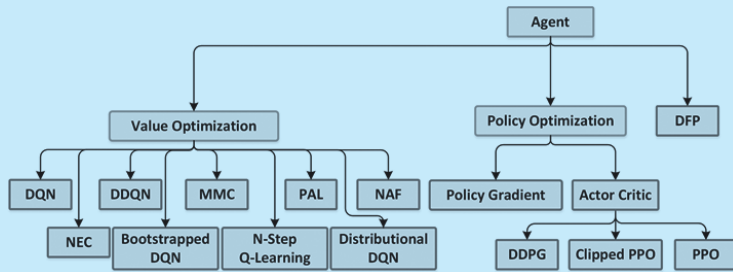


그림 4 심층강화학습 기법의 두 가지 분류 : (좌)가치(Value) 기반 최적화; (우)정책(Policy) 기반 최적화 기법(인텔, Coach 강화학습 라이브러리)

바둑게임에서 이세돌 프로를 이기는 결과를 보여주며 전 세계적으로 커다란 반향을 일으켰으며 과거 2014년 네이처(Nature) 저널에서도 소개된 적 있다.

심층강화학습기법은 학습에 이용되는 에이전트(Agent)의 숫자에 따라서 단일에이전트(single-agent)와 멀티에이전트(multi-agent) 기법으로 분류된다. 단일에이전트 기법은 획득하고자하는 목적이 분명한 시나리오에 적용되며, 멀티에이전트 기법은 획득하고자 하는 목적이 하나 이상이고, 목적 달성이 복잡한 시나리오에 적용하기에 적합하다. 대표적인 멀티에이전트 알고리즘은 계층적 강화학습(hierarchical reinforcement learning)이 널리 사용되고 있다. 또한, 심층강화학습은 적용되는 상태공간(state space)에 따라, 이산상태공간(discrete state space)과 연속상태공간(continuous state space)으로 분류할 수 있다(그림 4). 이산상태공간에 적용되는 기법은 가치(value)기반 최적화(optimization)기법을 기반으로 설계되며, 연속상태공간에 주로 사용되는 기법은 정책(policy)을 최적화하는 방식으로 도출된다.

심층강화학습 이용한 드론 충돌회피 및 자율비행 기술

앞서 소개했듯이, 드론의 자율비행을 위해 적용된 기법들은 주로 end-to-end 기반 심층학습을 이용한

기법들이 주로 사용되었다. 이러한 end-to-end 방식은 지도학습 형태를 사용하기 때문에 심층학습 에이전트를 훈련시키기 위해서 필요한 학습 데이터를 구축하는 데 많은 시간과 노력이 요구되며, 임무수행 중간에 동적장애물(moving obstacle)과 같은 학습에 활용되지 않은 환경에 대처하는 데 한계를 보여준다. 이러한 단점을 극복할 수 있는 방법은 다양한 환경을 시뮬레이션 할 수 있는 학습시스

템을 기반으로 풍부한 학습데이터로부터 스스로 학습할 수 있는 기법을 적용하는 것이다. 이 글에서는 가치 최적화 기법의 가장 대표적인 DQN(Deep Q-Networks) 기법을 이용한 드론의 충돌회피 및 자율비행에 대한 응용결과를 소개하고자 한다.

심층강화학습을 드론에 적용하기 위한 모델을 설계하고, 설계된 모델을 이용 학습(learning) 이후 학습된 알고리즘을 Nvidia TX2 GPU 모듈에 탑재하여 실험을 수행하였다. 첫 번째, 심층강화학습 기법 중 DQN 모델을 설계하였으며, ① 센싱 및 데이터변화, ② 심층합성곱신경망 기반 모션값 출력, ③ 보상을 통한 피드백, 크게 3단계를 걸쳐 드론의 모션을 제어한다. 인식을 위해 라이다(Hokuyo Lidar) 거리센서와 RGB-D 깊이(depth) 카메라(ZED) 센서를 사용하였다(그림 4). 라이다는 장거리 장애물 통산 30m 이내 거리 감지를 위하여 주로 사용되며, RGB-D 깊이(depth) 센서는 5m 이내 거리 근거리 장애물 탐색 및 인식을 위하여 사용된다. 두 번째, 센싱된 데이터 정보는 심층강화학습에 맞게 변환되어 이미지 형태로 전환된 후 합성곱신경망(convolutional neural networks)의 입력값으로 사용되며, 합성곱신경망의 출력값은 드론 모션 제어를 위한 기준값인 Q-데이터 형태로 도출된다. 마지막으로 이후 드론의 행동제어 결과값인 모션에 대한 보상(reward)값이 계산되며 강화학습의 Q-값 업데이트에 이용된다. 학습된 모델 값을 드론에 탑재하여 실제 실험을 진행한다. 드론은

이미 학습된 모델값을 이용하여 드론으로 입력되는 센싱값을 기반으로 학습된 최적의 모션값을 설정하고 방향을 제어한다. 장애물 감지시 보상값의 변화를 통해서 장애물의 유무를 스스로 판단하여, 이러한 경우 기동을 통해서 충돌회피 기능을 수행할 수 있다.

기존에 수행된 심층강화학습 기반 로봇의 모션제어 및 자율비행 연구에서는 주로 단일센서, 카메라 또는 라이다 센서를 사용한 심층강화학습 기술적용과 충돌회피 및 자율비행을 위한 연구가 진행되어 왔다. 이러한 기법은 불확실한 환경에서 제한적인 자율비행 성능을 보여줄 수 있다. 단일센서 기반 심층강화학습의 주된 이유는 심층강화학습에 사용되는 인공신경망(artificial neural networks)의 센서 입력값이 특정한 이미지 데이터 형태를 가져야하기 때문에, 인공신경망에 적합한 형태로 여러 개의 센서를 융합하는 기법은 복잡하고 어려운 문제로 인식되었다. 이 글에서는 이러한 단일센서의 단점을 극복하고 여러가지 센서값을 융합하여 인공신경망에 적합하게 변화된 다중센서 융합데이터 기반 심층강화학습 기법을 적용하여, 우수한 성능을 보여줄 수 있도록 제안하고 있다. 라이다 센서 데이터를 가공 변화하여 (4,90) 프레임 사이즈의 이미지 형태로 변환하고, RGB-D 깊이 센서로부터 획득된 이미지 데이터 (255,255) 이미지를 Binary 형태로 변화한 후 (4,90) 이미지로 변환 최종적으로 (4,90,4) 형태의 새로운 융합이미지를 생성하게 된다. 이렇게 생성된 융합이미지는 심층합성곱신경망의 입력값으로 사용되며, 강화학습을 통하여 드론의 모션을 제어하는 명령을 생성하게 된다. 그림 6의 (좌)는 위에서 언급한 센서융합 기반 융합데이터 형성과정을 보여주며, (우)는 합성곱신경망의 구조를 보여주고 있다. 최종적으로 드론의 모션제어를 위한 명령값을 출력으로 생성하고 있음을 자세히 보

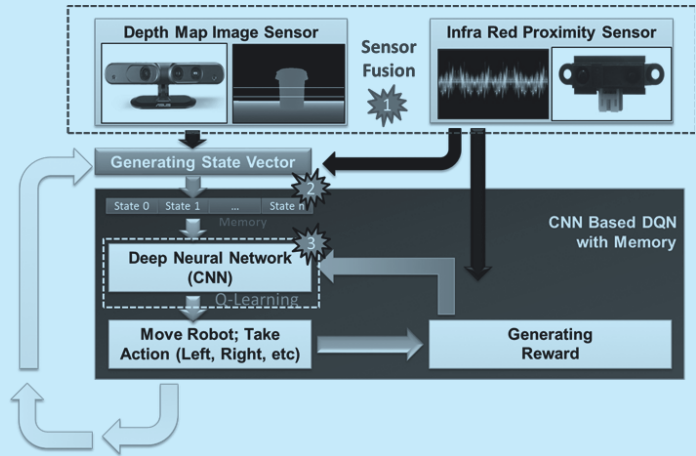


그림 5 메모리 기반 심층강화학습(M-DQN) 구성도

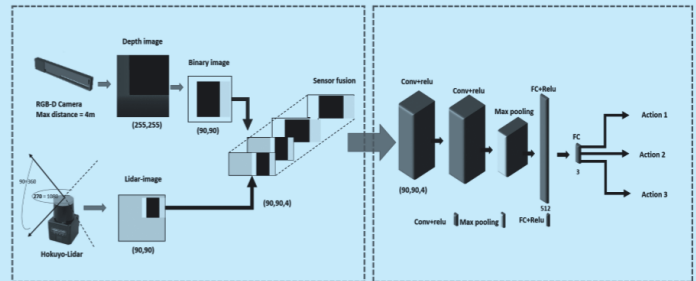
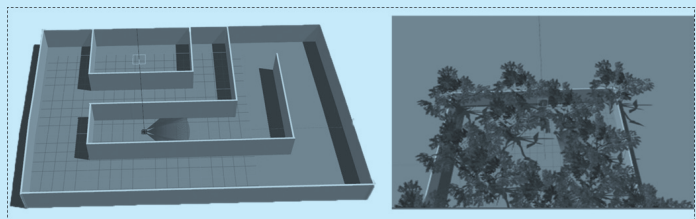


그림 6 (좌)심층강화학습 입력데이터를 위한 센서융합 기법, (우)합성곱 신경망 기반 심층뉴럴네트워크 구조



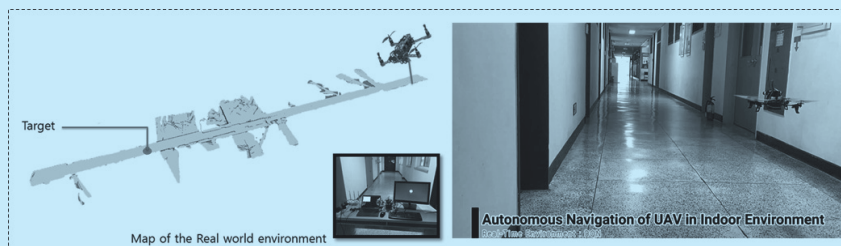
(좌)실내자율비행 학습

(우)실내 충돌회피 및 자율비행 학습

그림 7 메모리 기반 심층강화학습(M-DQN) 기법 기반 실내 충돌회피 및 자율비행 학습과정

여주고 있다.

실제 심층강화학습 기법의 학습을 위하여 학습데이터의 생성이 중요하며, 다양한 환경을 구성하고 이에 따른 데이터를 구축하는 것이 심층강화학습 성능에 중요한 역할을 하게 된다. 실제 실험을 통하여 학습용 데이터를 확보하는 것이 가장 적절하고 실제 실



(좌) Lidar 센서를 이용한 실내 Mapping 지도 (우) 실내 비행 실험 결과

그림 8 메모리 기반 심층강화학습(M-DQN) 기반 충돌회피 및 자율비행 실험 결과

험과 유사한 환경을 구성할 수 있다는 장점을 가지고 있으나, 실험을 통해서 데이터를 확보하는 것은 많은 시간과 비용을 요구하는 문제점을 가지고 있다. 따라서 최근에는 실제 실험 환경과 유사한 가상의 시뮬레이션 환경을 구성하고 다양한 시나리오를 만들어서 학습용 데이터를 생성하고 생성된 데이터를 이용하여 심층강화학습을 훈련시키는 방법이 널리 사용되고 있다. 드론 비행 데이터 생성을 위해 사용되는 여러 가지 시뮬레이터가 존재하는데, 대표적으로 마이크로소프트사의 AirSim이 실제와 유사한 환경을 구성할 수 있으며 선명하다는 장점으로 최근 많은 관심을 받고 있다. 또 다른 학습용 시뮬레이터로 관심을 받고 있는 것은 Gazebo 시뮬레이터이다. Gazebo 시뮬레이터는 ROS(Robot Operating System)와 연동을 통해서 드론 및 다양한 로봇의 모션을 제어하고 시뮬레이션할 수 있다는 커다란 장점을 가지고 있다. Gazebo 학습을 위한 시뮬레이션 결과는 그림 7에서 잘 보여주고 있다. 실내자율비행을 위하여 먼저 실내 시뮬레이션 환경을 구성하고, 파이썬(Python) 프로그래밍 기법을 이용하여 DQN 심층강화학습 알고리즘을 구현하고 ROS와의 연동을 통해서 Gazebo를 이용하여 다양한 실내실험 학습을 구현할 수 있다. 또한 장애물 충돌회피를 위하여 실내에 장애물을 추가하여 충돌회피를 위한 학습도 구현할 수 있다. 최종적으로 드론의 실내충돌회피 및 자율비행을 위한 강화학습 모델의 학습이 끝나게 되면, 학습된 모델을 엔비디아(Nvidia)

GPU 기반의 TX2모듈을 이용하여 드론에 탑재 후 비행실험을 수행할 수 있다. 실내자율비행 성능 검증 결과는 그림 8에서 보여주고 있으며, 실험결과를 지도제작(mapping)을 통한 비행궤적을 통해서 충돌회피 및 자율비행이 성공적으로 수행되었음을 알 수 있다. GPS 측위 정보 활용 없이 라이다와 깊이센서를

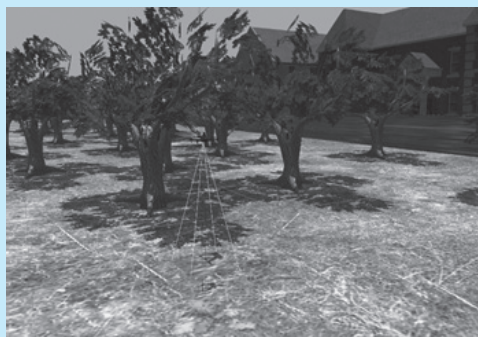
이용하여 실내에서 충돌회피 및 자율비행을 수행할 수 있다는 중요한 결과를 잘 보여주고 있어 향후 실에서의 다양한 응용이 가능할 것으로 보인다.

드론의 외부 활용에서도 GPS 위치 정보를 획득할 수 없는 환경들이 많이 존재하게 하는데, 향후 드론의 활용 범위를 확대하기 위해서는 실외 GPS 측위 정보 없이 드론이 충돌회피 및 자율비행에 대한 요구도가 매우 높을 것으로 판단된다. 이러한 수요를 충족하기 위해서 인공지능을 이용한 인식 및 판단이 가능한 드론 개발이 필요하게 된다. 앞에서 드론의 실내자율비행에 대한 심층강화학습 응용에 대해서 보여주었듯이 실외에서의 자율비행을 성능을 검증하기 위한 작업들이 필요하다. 먼저 실외 시뮬레이션 환경(그림 9(좌))을 나무를 활용하여 구성하고 Gazebo를 이용하여 다양한 실외 환경에 대한 학습데이터를 생성하고 이를 이용한 DQN 모델의 학습을 수행할 수 있다. 학습된 심층강화학습 모델을 드론의 TX2 모듈에 탑재하고 실외 비행실험을 수행할 수 있게 된다. 실외 충돌회피 및 자율비행은 실내 자율비행에 비해 보다 향상된 성능을 요구하게 되는데 무엇보다도 나무와 같은 숲을 충돌 없이 자율적으로 비행하는 성능구현이 매우 필수적인 요소로 작용하게 된다. 실제 실험에서 보여 지듯이 심층강화학습을 이용한 숲에서의 성공적인 자율비행 실험 결과를 보여주고 있다(그림 9(우)). 본 실외 충돌회피 및 자율비행에 사용된 센서는 호쿠유(Hokuyo) 라이다(Lidar)와 ZED 센서이며, 각

두 센서의 데이터를 융합한 융합 데이터가 심층강화학습 알고리즘의 입력값으로 사용되며 심층강화학습을 통해서 충돌회피와 자율비행을 위한 모션제어 판단 명령을 스스로 내리게 된다. 이렇게 인공지능을 이용한 나무와 같은 장애물의 자율충돌회피 기능이 구현이 되면 드론의 응용 범위가 매우 넓어질 수 있음은 명확해진다.

결론 및 향후 전망

본문에서는 드론의 자율비행 및 충돌회피를 위한 기법중 기존의 규칙기반(rule based) 제어기법(control laws)을 적용하는 대신 다중 센서로 획득된 데이터를 이용하여 스스로 인식 및 판단을 내리고 드론의 모션을 제어할 수 있는 심층강화학습 기법에 대해 소개하였다. 최근 다양한 심층강화학습기법들이 개발되고 있으나 아직은 시뮬레이션에 기반 학습용 알고리즘 개발이 주가 되고 있는 상황으로, 본문에서는 실제 심층강화학습 알고리즘 구현 및 비행 실험을 통해서 성능을 검증하였다는데 다른 중요한 의미를 부여해주고 있다. 개발된 심층강화학습 알고리즘의 탑재 및 구현을 위하여 엔비디아 모듈을 활용하였으



(좌)메모리 기반 심층강화학습(M-DQN) 기반 실외 충돌회피 및 자율 비행 실험 결과



(우)실외 충돌회피 및 자율비행 실험 결과

그림 9 심층강화학습을 이용한 자율비행실험 결과

며, 드론에 탑재를 통해서 GPS 위치 정보 획득이 가능하지 않은 실내 및 실외 실험 환경을 구성하여 성능을 검증하였다. 본문에서 제안한 DQN 기반 심층강화학습을 이용한 실내 및 나무와 같은 장애물의 자율충돌회피 기능이 구현을 통해서 드론의 응용 범위가 매우 넓어질 수 있게 드론을 포함한 스마트 모빌리티 시장창출 잠재력을 충분히 발휘할 수 있을 것으로 기대된다.

감사의 글

본 연구는 미래창조과학부의 재원으로 한국연구재단 무인이동체 미래선도핵심기술개발사업단의 지원을 받아 수행되었음.(No 2016M1B3A1A01937245)