AR 네비게이션을 위한 딥러닝 기반의 VPS(Visual Positioning System)

□ 정태원, 정계동 / 광운대학교

Ⅰ. 서 론

최근 몇 년 동안 위치 기반 서비스(Location Based Services)에 대한 수요가 증가함에 따라 정확한 위치 정보에 대한 필요성이 높아졌다. 모바일 및 기타 모바일 플랫폼에서 위치를 확인하는 가장 일반적인 방법은 GNSS(Global Navigation Satellite System)이다. 그러나 실내 환경의 경우 GNSS 신호는 장애물에 의해 차단되어 대부분 실외 환경에서만 사용할 수 있다. 실내 위치 인식을 위한 다양한 기술이 제안되었지만 많은 실내 위치 인식 방법은 무선신호를 이용한 Finger Printing 기반 위치 인식 알고리즘 기반을 넘어서지 못하고 있다. 이러한 방법에서는 Wi-Fi RSS(수신 신호 강도) 또는 MFS(자기장 강도)가 수집되어 Finger Printing 데이터베이스의 데이터와 비교된다. Finger Printing 기반 시스템은 구축하기 쉽고 높은 성능을 유지할 수 있지만 신호 패턴은 시스템 환경 변화에 영향을 받기 때문에 성능을 유지하기가 어렵다. 또한 Finger Printing 데이터베이스 구축은 노동 집약적인 많은 시간 투자와 높은 비용이 요구된다. 이런 결함을 극복하기 위해 Optical, RFID(Radio Frequency Identification), Bluetooth Beacons, ZigBee, Pseudo Satellite 등을 포함한 많은 대안이 제안되었지만 복잡한 실내 환경에서는 정확도가 충분하지 않으며, 막대한 비용과 추가 인프라가 필요할 수 있다. 이러한 방법을 극복하기 위해 딥러닝 기반의 VPS(Visual Positioning System)에 대한 연구가 최근에는 활발하게 이루어지고 있다.

Ⅱ. 시각적 위치 결정 시스템

시각적 위치 결정 시스템은 크게 세 가지 범주로 나눌 수 있다. 구조 기반 위치 결정 방법(Structure-based localization methods)과 이미지 기반 위치 결정 방법

(Image-based localization methods) 및 학습 기반 위치 결정 방법(Learning-based localization methods)이다.

구조 기반 위치 결정 방법은 로컬 기능을 활용하여 쿼리 이미지의 기능과 3D 모델의 포인트 클라우드간의 2D 대 3D 일치를 추정하거나 RGB-D 이미지 및 3D 모델을 사용하여 카메라 포즈를 추정한다. 마찬가지로 2D 이미지 기반 위치 결정과 3D 구조 기반 위치 결정을 비교하면 단순한 2D 기반 방법이 가장 낮은 성능의 위치 결정을 하고 3D 기반 방법이 더 복잡한 모델 구성 및 유지 관리를 통해 보다 정확한 포즈를 추정한다. 2D 기반 방법과 간단한 데이터베이스 구축 절차와 정확한 포즈 추정을 모두 갖춘 SfM(local Structure-from-Motion)이 있지만 위치 확인 과정에서 런타임이 길다.

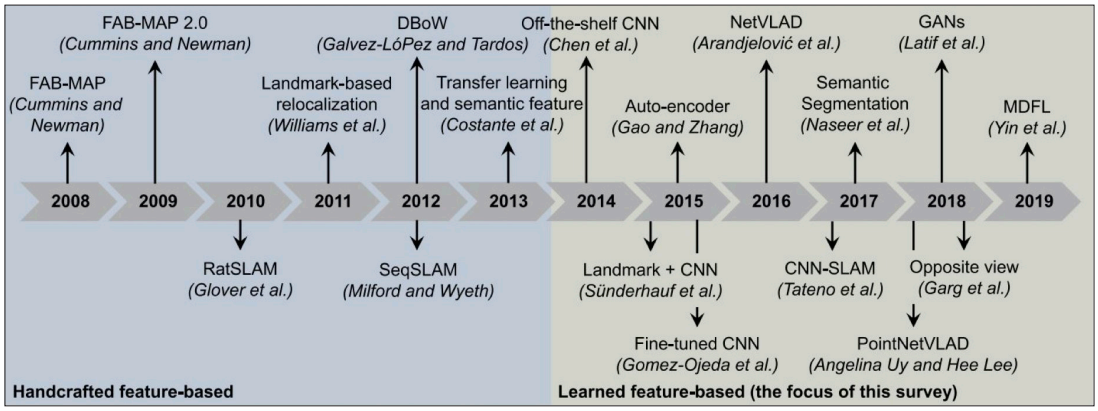
이미지 기반 위치 결정 방법은 공공 지역의 지정된 방대한 이미지 데이터베이스에 의해 추진되었다. 이러한 방법은 쿼리 이미지를 데이터베이스의 이미지와 일치시키는 이미지 검색 기반 전략을 사용한다. 그 후 조회 이미지의 위치는 검색된 참조 이미지의 포즈 정보를 기반으로 계산된다. 소셜 네트워크 및 스트리트 뷰 사진의 빅데이터화로 인해 데이터 기반의 이미지를 참조하는 데 사용할 수 있는 다량의 이미지가 구축되었으며 이미지 검색은 많은 이미지 기반 방법에서 일반적으로 사용되는 대규모 디지털 데이터베이스에서 이미지를 검색하는 시각적 검색 작업이다. 기존의 방법은 로컬 디스크립터 매칭을 기반으로 이미지를 검색하고 정교한 공간 검증을 통해 재정렬한다. 이미지에 대한 콘텐츠 기반 이미지 검색은 가장자리, 색상, 질감 및 모양과 같은 시각적 콘텐츠에 의존한다. 최근에는 이미지 검색을 위해 CNN을 활용하며, 대부분은 사전 훈련 된 네트워크를 특징점 추출기로 사용한다. 또한 일부 작업은 CNN 기능의 기하학적 문제를 해결하고 다양한 크기 및 비율의 이미지를 정확하게 표현할 수도 있다.

학습 기반 위치 결정 방법은 포즈 정보를 사용하여 주어진 이미지에서 모델을 학습시키면 학습된 모델로 장면을 표현할 수 있다. 이러한 학습 기반 위치 결정 방법은 포즈 추정에 대한 일치를 예측하거나 PoseNet, PoseNet2 및 VlocNet과 같은 카메라 포즈를 추정한다. PoseNet은 메트릭 지역화 문제를 해결하기 위해 DCNN을 사용하는 첫 번째 접근 방식이었고, 포즈 불확실성을 해결하기 위해 Bayesian CNN 구현이 사용되었다. 그 후, 장기 단기 메모리(LSTM) 및 대칭 인코더-디코더와 같은 아키텍처를 사용하여 DCNN의 성능을 향상하였다.

Ⅲ. CNN 딥러닝 기반의 VPS

딥러닝은 컴퓨터 비전(CV) 및 로봇 공학을 포함한 다양한 분야에서 주목할 만한 성과를 거두었다. 지난 수년 동안 컴퓨터 비전 및 로봇 공학 커뮤니티의 연구자들은 VPS를 해결하기 위해 딥러닝을 도입하였다. IEEE CVPR(컴퓨터 비전 및 패턴 인식에 관한 국제 회의) 및 ICRA(로봇 공학 및 자동화에 관한 국제 회의)와 같은 국제 회의에서도 딥러닝 기반 VPS에 관한 일련의 워크샵을 개최하여 큰 관심을 보였다. 다수의 논문에서 심층 신경망(DNN), 특히 CNN 기반 VPS 방법의 성능이 기존 방법보다 우수하다는 것이 입증되었다. 〈그림 1〉은 기존의 알고리즘 특징점 기반 방법과 딥러닝 특징점 기반 방법을 포함하여 지난 10여 년간의 VPS 발전 과정이다.

CNN 딥러닝 기반의 VPS는 이미지를 구성하는 특징점을 추출하기 위해 학습된 CNN 모델을 사용한다. 이미지의 유사성을 판단하기 위해 특정 VPS 데이터셋에서 CNN 모델을 미세 조정하거나 새로운 아키텍처로 인식 성능을 향상하였다.

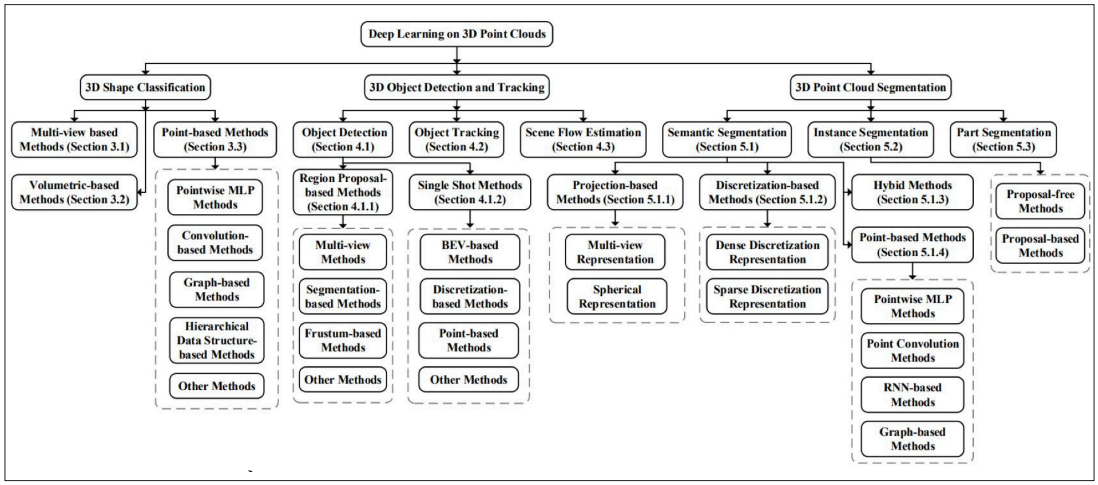


<그림 1> 알고리즘 특징점 기반 방법과 딥러닝 특징점 기반 VPS 발전 과정 (출처: https://doi.org/10.1016/j.patcog.2020.107760)

Ⅳ. 3D 포인트 클라우드 기반의 딥러닝

딥러닝은 다양한 2D 비전 문제를 해결하는 데 성공적으로 사용되었다. 3D 획득 기술의 급속한 발전으로 다양한 유형의 3D 스캐너, LiDAR 및 RGB-D 카메라 등에 의해 수집된 3D 데이터는 풍부한 기하학적 모양 및 스케일 정보로 구성되어 있다. 2D 이미지로 보완된 3D 데이터는 기기의 주변 환경을 더 잘 이해할 수 있는 기회를 제공하며 자율 주행, 로봇 공학, 원격 감지 및 의료 치료 등 다양한 응용 분야에 유용하게 활용되고 있다. 3D 데이터는 일반적으로 깊이 이미지, 포인트 클라우드, 메시 및 그리드를 포함한 다양한 형식으로 표현될 수 있다. 그 중 일반적으로 사용되는 포인트 클라우드는 3D 공간 정보 보존성이 높아 자율 주행 및 로봇 공학관련 응용 프로그램에서 선호되는 데이터 형식이다.

최근 딥러닝 기술은 컴퓨터 비전, 음성 인식, 자연어



<그림 2> A taxonomy of deep learning methods for 3D point clouds (출처: https://github.com/QingyongHu/SoTA-Point-Cloud)

처리와 같은 많은 연구분야에 활용되고 있다. 그러나 3D 포인트 클라우드에 대한 딥러닝 기술은 소규모 데이터 세트, 높은 차원성 및 3D 포인트 클라우드의 구조화되지 않은 특성과 같은 몇 가지 중요한 해결해야 할 과제가 있다.

이러한 문제점을 해결하기 위해 특히 지난 몇 년 동안 연구가 활발히 진행중이며 ModelNet, ScanObjectNN, ShapeNet, PartNet, S3DIS, ScanNet, Semantic3D, ApolloCar3D 및 KITTI Vision Benchmark Suite 데이터 세트는 3D 포인트 클라우드에 대한 딥러닝 연구를 촉진시켰다. 또한 3D 형상 분류, 3D 객체 감지 및 추적, 3D 포인트 클라우드 분할 및 처리와 관련된 다양한 문제를 해결하기 위해 점점 더 많은 방법이 제안되고 있다.

Ⅴ. 3D 복셀(Voxel)을 이용한 VPS

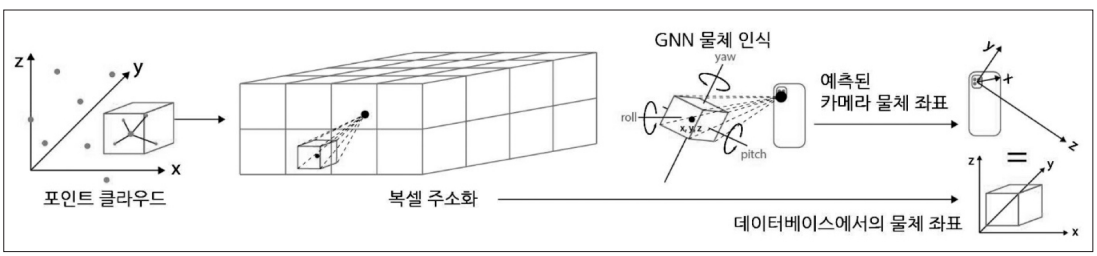
3D 포인트 클라우드를 이용한 딥러닝 기반의 VPS는 LiDAR(Light Detection And Ranging) 센서에서 생성된 포인트 클라우드를 딥러닝 네트워크에서 학습하여 물체의 위치와 포즈를 추정한다. 학습된 네트워크는 물체의 중심 좌표(x, y, z), 물체의 크기(길이, 폭, 높이), 물체의 회전 값(X축 회전-Roll, Y축 회전-Pitch, Z축 회전-Yaw)을 추정한다. 추정된 물체의 위치와 크기 및 회전 각도는 사용자 카메라 센서의 원점 좌표(0, 0, 0)를 기준으로 추정된다.

딥러닝 기반으로 물체를 인식하여 물체의 카메라 좌표계에서의 좌표값과 포즈를 추정한다. 데이터베이스는 공간을 복셀 단위로 주소화하고 데이터베이스 좌표계에서의 물체는 중심 위치 좌표(x, y, z)와 길이, 폭, 높이, 물체의 회전 값으로 구성된다.

딥러닝 네트워크에서 추정된 물체의 크기를 매칭하기 위해선 카메라 좌표계에서의 물체의 중심 좌표를 미리 정의된 데이터베이스 좌표계의 중심 좌표로 이동변환한뒤 다시 네트워크에서 추정된 물체의 회전 값만큼 데이터베이스 좌표계에서 회전하면 사용자 카메라 센서의 원점 좌표가 데이터베이스 좌표계에서의 카메라 위치 좌표로 변환된다.

〈그림 3〉은 카메라 좌표계와 데이터베이스 좌표계를 매칭한 후 물체의 좌표 이동 변환 및 회전 변환이다. 데이터베이스는 복셀 단위로 공간을 분할하게 된다. 분할된 공간에 각각의 주소를 생성하며 이렇게 생성된 복셀 주소는 카메라의 위치를 추정하기 위한 공간의 주소로 사용된다. 카메라의 예측한 물체의 포즈를 회전 변환 행렬로 변환하고 카메라의 원점을 역추적하여 공간의 복셀 주소로 변환한다. 역추적된 데이터베이스의 카메라의 위치는 사용자의 위치를 결정하게 된다.

복셀 레이블링 기반의 VPS는 데이터베이스화된 월드 좌표계와 딥러닝으로 추정된 카메라 좌표계의 동일 물



<그림 3> 카메라 좌표계와 월드 좌표계 통일 후 물체의 좌표 이동 변환 및 회전 변환

체를 매칭한 후 카메라 원점의 월드 좌표계로의 변환하는 시스템이며 세 단계로 이루어진다. 첫 번째는 데이터베이스의 물체의 박스와 딥러닝 객체 인식된 물체의 박스 생성이다. 두 번째 단계는 동일한 물체의 박스를 데이터베이스의 좌표 공간으로 매칭하여 두 박스의 회전 행렬을 계산한다. 세 번째 단계는 계산된 회전 행렬로 카메라 원점을 회전 변환하여 카메라 원점을 변환시키며 어떤 복셀에 속하는지 계산되는 VPS이다.

VI. 결 론

증강현실에서 기본적이고 핵심적인 기술은 물체 인식과 카메라의 방향 그리고 기기의 연속된 위치 추적인 VPS 기술이다. 딥러닝 기반 VPS는 GPS가 충분하지 않은 경우 학습 기반으로 사용자에게 맞춤형 증강현실 콘텐츠를 제공하기 위한 것이다. 모바일 기기의 지속적인 성능 향상과 네트워크 속도의 향상에도 여전히 증강현실을 실질적으로 활용한 제품이나 서비스는 부족하며 모바일 사용자에게 주목받지 못했다. 증강현실은 가상현실(Virtual Reality, VR)과 달리 현실공간을 활용하기 때문에 실감적인 가상세계를 위한 3차원 모델링과 렌더링의 요소를 줄인 반면, 사용자의 위치와 주변의 환경 변화에 대응하는 가상의 정보나 콘텐츠를 실시간으로 제공하므로 기술적 어려움을 해결해야 한다. 최근에는 구글, 애플, 페이스북, 마이크로소프트, 삼성, 네이버 등 대기업을 중심으로 투자 및 연구가 활발하며, 다양한 기기와의 연동과 폭넓은 응용 분야 개발을 통해 생활 전반에 걸쳐 증강현실을 향유할 수 있을 것으로 기대된다.

참 고 문 헌

[1] Walter C. S. S. Simões, Guido S. Machado, André M. A. Sales, Mateus M. de Lucena, Nasser Jazdi, Vicente F. de Lucena, Jr, "Review of Technologies and Techniques for Indoor Navigation Systems for the Visually Impaired," Sensors 2020

[2] Xiwu Zhanga, Lei Wang b, Yan Su, "Visual place recognition: A survey from deep learning perspective" Pattern Recognition, Volume 113, May 2021

[3] Ramon F. Brena, Juan Pablo García-Vázquez, Carlos E. Galván-Tejada, David Muñoz-Rodriguez, Cesar Vargas-Rosales, James Fangmeyer Jr., "Evolution of Indoor Positioning Technologies: A Survey," Hindawi Journal of Sensors Volume p.21, 2017

[4] Anup S, Abhinav Goel, Suresh Padmanabhan, "Visual Positioning System for Automated Indoor/Outdoor Navigation," Proc. of the 2017 IEEE Region 10 Conference(TENCON), Malaysia, November 5-8, 2017

[5] Weijing Shi and Ragunathan (Raj) Rajkumar, "Point-GNN: Graph Neural Network for 3D Object Detection in a Point Cloud," Computer Vision and Pattern Recognition (cs.CV), arXiv:2003.01251v1 [cs.CV] 2 Mar 2020

[6] Yulan Guo, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, Mohammed Bennamoun, "Deep Learning for 3D Point Clouds: A Survey," arXiv:1912.12033v2 [cs.CV] 23 Jun 2020

필 자 소 개

정 태 원

- 2020년 2월 : 광운대학교 스마트융합대학원 정보시스템학과 석사 졸업
- 2020년 3월 ~ 현재 : 광운대학교 일반대학원 실감융합콘텐츠학과 박사과정
- 주관심분야 : 컴퓨터 비전, 증강현실, 모바일 시스템, 인공지능

정 계 동

- 광운대학교 컴퓨터과학과 이학 박사
- 현재 : 광운대학교 인제니움학부대학 교수
- 주관심분야 : 웹 서비스, 지오펜싱 서비스, 증강현실, 인공지능