

# 영상 기반의 3차원 보행자 위치 인식 연구 분석 및 동향

김태주

세종대학교 지능기전공학과

## 1. 서 론

최근 인공지능의 발전으로 자율 주행 로봇에 대한 연구와 관심이 많아졌다. 특히, 자율 주행의 3대 프로세스 중 하나인 인지 기능에서 보행자 검출은 목적지로 도착하기 위한 경로 계획과 안전을 위한 회피 계획에 있어 빠질 수 없는 요소이다. 이러한 이유로 물체 검출은 빠르면서 높은 정확도를 요구하게 되었고, 이와 더불어 많은 연구들이 이뤄지고 있다. 보행자 검출은 다양한 센서(LiDAR, 카메라, Wi-fi...)를 이용하여 이뤄진다. 여러 센서 중 인간의 시각 능력을 가장 많이 닮은 카메라 센서가 대표적인 센서로 사용되고 있다. 본 문에서는 여러 센서 중 카메라 센서를 이용한 방법론에 집중하고자 한다.

영상 기반의 보행자 검출은 2차원 물체 검출로부터 시작하였다. 대표적인 방법론으로 Faster R-CNN[1], SSD[2] 이 있으며, 지속적인 발전으로 우수한 성능과 추론 속도를 갖추게 되었다. 하지만 해당 방법론은 2차원이 아닌 3차원에 존재하는 실제 세계에 적용하기에는 한계가 있다. 2차원 물체 검출은 물체 위치를 영상 좌표계에서의 2차원 경계 상자만으로 예측한다. 즉, 실제 세계에서의 물체의 거리와 실제 스케일은 알아낼 수가 없어, 자율 주행 로봇의 경로 계획과 회피 계획에 적용하기에 적합하지 못하다는 한계가 있다. 이러한 한계로 실제 세계에서의 물체 위치를 검출하기 위해 영상 좌표가 아닌 3차원 좌표를 예측하는 3차원 물체 검출 연구들이 자율 주행 진영의 관심을 받고 있는 현황이다. 3차원 물체 검출은 2개의 좌표축 정보를 추정해야 하는 2차원 경계박스보다 많은 6~8 개의 좌표 혹은 변수를 예측해야만 한다[3]. 또한, 3차원 정보를 활용하기 위해 깊이 정보를 추정하여 3차원 정보를 예측하거나

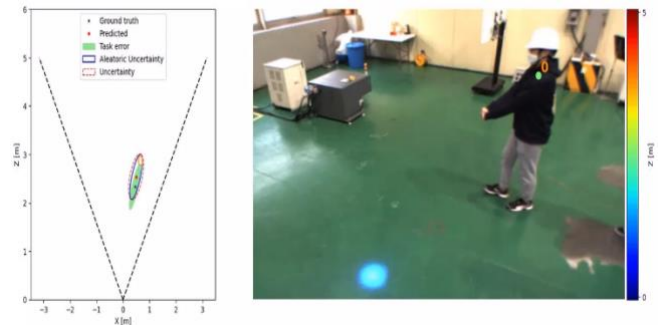


그림 1 3차원 보행자 위치 인식. 왼쪽은 Monoloco[7]를 이용한 정량적 시각화 예시, 오른쪽은 예측에 사용된 영상.

[3-5], 2차원 물체 검출 후, 동일하게 검출된 2차원 경계 박스 간의 3차원 관계를 이용하여 3차원 경계 박스 추정하는 기법[6, 10]들이 존재한다. 이처럼 영상 기반의 3차원 물체 검출은 2차원 영상 정보로부터 3차원 정보를 추정해야 하기에 모델의 계산 복잡도가 높아진다. 이로 인해 2차원 물체 검출에 비해 성능과 보장되는 추론 속도가 낮아지게 되어, 여전히 활발히 연구가 진행 중이며, 연구가 가속화되어야 하는 분야이다.

앞서 설명한 바와 같이 영상 기반의 3차원 물체 검출은 높은 계산 복잡도로 정형적인 형태를 가진 차량에서도 성능 개선을 위한 노력을 활발하게 진행되고 있으나, 상대적을 비정형적이고 유동적인 움직임을 가지고 있는 3차원 보행자 검출에 대한 관심은 비교적 낮은 편이다. 위와 같은 대표적인 원인으로는 현재 3차원 물체 검출 방법론의 주된 방법이 깊이 정보를 활용한다는 점이다. 보행자인 경우, 관절의 움직임으로 인해 비정형적인 형태를 가지고 있으며, 차량에 비해 스케일이 작아 상대적으로 깊이 정보 추정이 어려워져 3차원 물체 검출도 어려워진다는 문제가 있다.

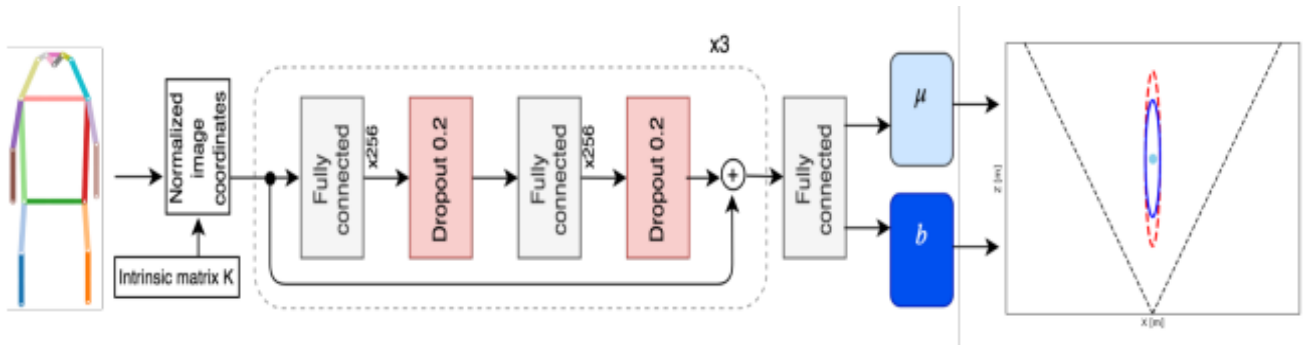


그림 2. Monoloco의 모델 구조

3차원 보행자 검출에 있어 가장 큰 영상 기반의 3차원 물체 검출의 문제점은 높은 계산 복잡도로 인해 빠른 추론 속도를 보장받지 못한다는 점이다. 빠른 추론 속도는 보행자의 안전과 직결되는 문제이기에 반드시 해결되어야 하는 문제이다.

이와 같은 한계를 극복하기 위해 깊이 정보가 아닌 보행자에 집중하고, 모델의 계산 복잡도를 낮춰 추론 속도를 보장하는 3차원 보행자 위치 인식[7]이 등장하게 되었다. 본 문에서는 3차원 보행자 위치 인식에 대해 소개하며, 최근 연구 동향에 대해 분석해보고자 한다.

## 2. 3차원 보행자 위치 인식

### 2.1 Monoloco

본 문에서 다룬 3차원 보행자 위치 인식은 저명한 컴퓨터 비전 학회 ICCV(IEEE International Conference on Computer Vision) 2019년에 발표한 Monoloco[7]에서 처음 제안되어졌다. 앞서 다룬 3차원 물체 검출과는 다르게 보행자의 3차원 위치만 인지함으로써, 모델의 계산 복잡도를 낮춰 모델의 효율성을 높였다. 또한, 비정형적인 보행자 정보를 효율적으로 압축하기 위해 사전 학습된 포즈 추정기를 이용하여 보행자의 관절 키포인트를 이용하여 보행자의 3차원 위치를 추정하였다. 또한, 보행자 위치 추정에 있어 내부/외부적 불확실성을 정의하였다. 단안 영상을 이용한 3차원 위치 추정은 영상을 구성하는 픽셀의 한계로 3차원 위치 정보를 표현에 한계를 가지는 내부적 모호성을 정의한 task error를

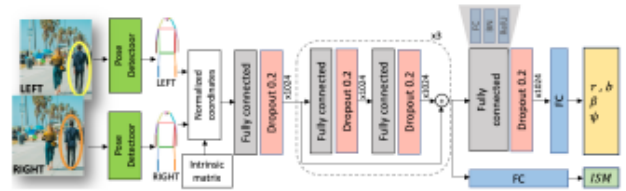


그림 3. MonStereo의 구조

정의하였다. 저자는 이러한 한계를 오히려 학습하여 불확실성을 예측함으로써, 정확도를 향상시켰다. 이를 통해 기존 단안 영상을 이용한 3차원 물체 검출 방법론 보다 높은 성능과 빠른 추론 속도를 선보인다.

### 2.2 MonStereo

MonStereo[8]는 저명한 국제 학회 ICRA 2021(IEEE Robotics and Automation Society)에 출판된 방법론으로 기존의 스테레오의 고질적인 문제로 꼽히는 long tail에 집중하였다. Long tail은 스테레오간 유사한 정보간 매칭에 실패한 정보들이 깊이 정보에 반영될 시, 매우 큰 값을 가지는 문제에 해당한다. 저자는 이러한 문제를 스테레오 구성의 카메라의 시점 차이에서 발생한다고 주장한다. 이로 인해 스테레오 구성에서의 내부적 모호성으로 정의하였고, 저자는 이점을 해결하기 위해서 매칭이 없는 경우에는 단안 기법을 이용하여 예측하자 아주 직관적인 해결책을 제안한다.

먼저 양안 영상으로부터 사전 학습된 포즈 추정기를 이용하여 보행자의 관절 키포인트를 추출한다. 이후, 왼쪽 카메라를 기준으로 오른쪽 카메라로부터 추출된 보행자의 관절 키포인트와 거리를 계산하여 결합한다. 그림 3과 같이 매칭쌍 유무를 예측하는 Instance-based

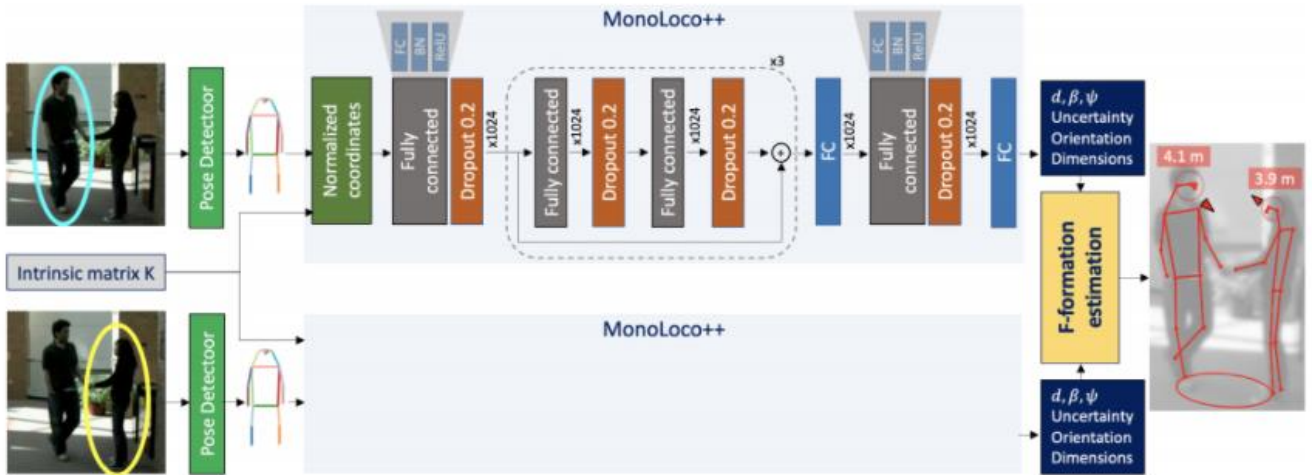


그림4. MonoLoco++의 모델 구조

Stereo Matching 모듈로 판별을 하여 단안 혹은 스테레오를 이용한 예측을 판별함으로써 long tail 문제를 극복할 수 있도록 하였다. 또한 데이터 셋 자체의 키 분포도에 한정된다는 단점을 극복하기 위해 가상으로 만들어진 다양한 관절 키포인트를 추가하는 Knowledge Injection을 제안함으로써 데이터 셋의 내부적 한계를 극복하였다.

### 2.3 MonoLoco++

MonoLoco++은 저명한 저널 중 하나인 T-TIS 2021(IEEE Transactions on Intelligent Transportation Systems)에 출판된 방법론이며, 거리두기에 주 관심을 두었다. Monoloco를 이용하여 모든 보행자의 3차원 위치 정보를 추론하되, 저자는 사회적 거리두기를 고려하기 위해 행동 과학에 포함되는 Social Interactions에 연구된 F-formations의 특성을 고려하여 추정된 보행자 간의 관계와 거리 정보를 유추하도록 하였다. 이를 통해 monoloco++는 실제 보행자 간의 실제 소통이 이뤄지는지에 대한 예측도 가능해진다.

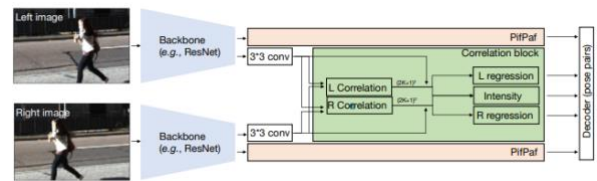


그림 5. PSF의 모델 구조

### 2.4 MonoPair

PSF[12]는 저명한 컴퓨터 비전 학회인 ICRA 2020(IEEE Robotics and Automation Society)에 출판된 방법론이다. MonStereo가 사전 학습된 포즈 추정기를 사용함으로써 의존적인 점으로 개선하기 위해 스테레오 구성 추정된 키포인트 간의 상관 관계를 고려하는 새로운 Part Spatial Field를 제안하였다.

## 3. 결 론

3차원 보행자 위치 인식은 3차원 물체 검출과 달리 효율성을 고려하여 발전된 방법론이며, 추후 자율 주행 로봇에서의 효율성을 생각한다면 가장 현실적인 해결책이라고 생각되어진다. 이후에도 지속적인 연구를 통해 개선되어지며, 3차원 물체 검출에 뒤쳐지지 않는 성능을 선보이길 기대한다.

## 참 고 문 헌

- [1] Ren, Shaoqing, et al. "Faster R-CNN: towards real-time object detection with region proposal networks." *IEEE transactions on pattern analysis and machine intelligence* 39.6 (2016): 1137-1149.
- [2] Liu, Wei, et al. "Ssd: Single shot multibox detector." *European conference on computer vision*. Springer, Cham, 2016.
- [3] Wang, Yan, et al. "Pseudo-lidar from visual depth estimation: Bridging the gap in 3d object detection for autonomous driving." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.
- [4] Qian, Rui, et al. "End-to-end pseudo-lidar for image-based 3d object detection." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.
- [5] DSGN: Deep Stereo Geometry Network for 3D Object Detection
- [6] Li, Peiliang, Xiaozhi Chen, and Shaojie Shen. "Stereo r-cnn based 3d object detection for autonomous driving." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.
- [7] Bertoni, Lorenzo, Sven Kreiss, and Alexandre Alahi. "Monoloco: Monocular 3d pedestrian localization and uncertainty estimation." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019.
- [8] Bertoni, Lorenzo, et al. "Monstereo: When monocular and stereo meet at the tail of 3d human localization." *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021.
- [9] Bertoni, Lorenzo, Sven Kreiss, and Alexandre Alahi. "Perceiving Humans: from Monocular 3D Localization to Social Distancing." *IEEE Transactions on Intelligent Transportation Systems* (2021).
- [10] Liu, Yuxuan, Lujia Wang, and Ming Liu. "YOLOStereo3D: A Step Back to 2D for Efficient Stereo 3D Detection." *arXiv preprint arXiv:2103.09422* (2021).
- [11] Chen, Yongjian, et al. "Monopair: Monocular 3d object detection using pairwise spatial relationships." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.
- [12] Deng, Wenlong, et al. "Joint Human Pose Estimation and Stereo 3D Localization." *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020.