

기술동향 보고서

3D Pedestrian localization

서지영 | 컴퓨터비전 | 21.12.18

Introduction

최근 자율주행차, 로봇기술 등의 발달로 인해 딥러닝에 기반한 객체 인식 기술의 향상으로 보행자 인식에 대한 여러 알고리즘이 제안되고 있다. 보행자 인식 기술의 경우 여러 사고의 위험으로 부터 보행자를 보호하기 위해서도 필요성이 대두되고 있으며 이 경우 여러 환경 안에서 사람의 특징을 잘 찾을 수 있는 알고리즘이 필요하다. 즉 다양한 주야간 환경 속에 존재하는 보행자를 검출해야 한다.

3D Detection

영상기반 객체 검출을 위해 다양한 특징점과 학습 알고리즘이 제안되고 있으며 특히 다수의 보행자가 검출되는 상황에서 보행자의 특징 묘사를 위하여 HOG (Histogram of Oriented Gradients), 경계선, 형태의 맥락, 기울기 값의 매그니튜드와 칼라의 형태 등 다양한 특징들이 활용되고 있다.

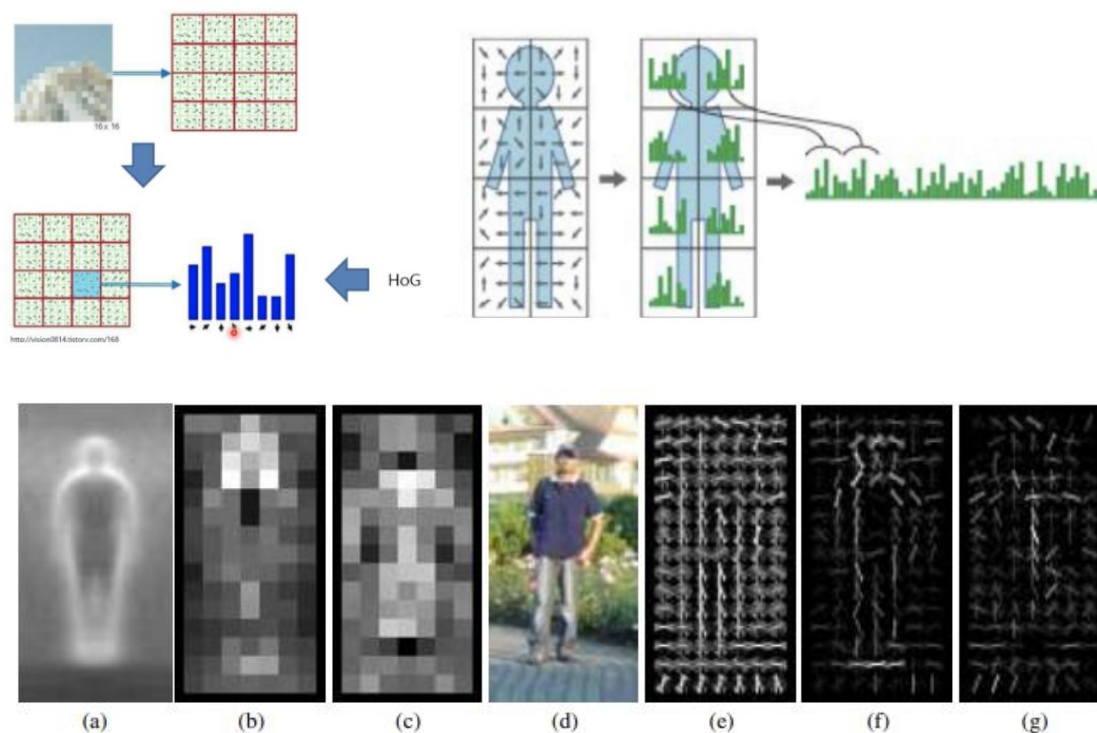


Figure 1 HOG

HOG의 경우 영상을 일정 크기의 셀로 분할하고, 분할한 각 영역에 대해 gradient의 방향값의 히스토그램을 구한다. vectorization 또는 concatenation을 통해 보행자 영상과 보행자가 아닌 영상을 표현한다. 즉 서있는 사람 영상 전체적인 형태에 주목하여

그래디언트를 구하고 그 크기와 방향 성분을 사용하여 사람이 서 있는 형태에 대한 특징 벡터를 추출하는 방식을 사용하였다. 그 후 머신러닝의 일종인 서포트 벡터 머신(SVM : Support Vector Machine)알고리즘을 이용하여 입력 영상에서 보행자 위치를 검출하는 방법을 제안한다.

HOG 알고리즘의 경우 보행자 검출을 위한 영상은 기본적으로 1. 보행자 검출을 위한 영상은 기본적으로 64x 128 크기의 영상을 사용합니다.

2. 입력 영상에서 그래디언트를 계산합니다

3D Detection 의 경우 Monocular(단안), Stereo, Vision-based 3D Localization Ambiguity 로 구분할 수 있다. Monocular 의 경우 한대의 카메라를 사용하여 포즈를 구하는 방식이며 Stereo 의 경우 두대의 카메라를 사용해 포즈를 구하는 방식이다. RGB-D SLAM 의 경우 구조광이나 ToF 센서를 이용한 카메라를 사용하는 방식으로 한 프레임의 데이터만으로 실제 metric scale 의 3D structure 를 얻어낼 수 있다.



Figure 2 Qualitative Examples

2D 이미지에서 3D 물체를 감지할 때 위치 파악에서 발생하는 어려운 점을 해결하기 위해 다양한 연구들이 진행되고 있다. 정확한 3D 물체 감지를 위해 고정밀 LIDAR 포인트 클라우드를 활용하고 있지만, 이는 비용적인 문제와 날씨의 영향을 받는 단점이 있다.

SMOKE 는 3D bbox 를 직접 예측하는 방식을 사용한다. 3D 입체 중심 투영 시점으로 3D 경계 상자를 인코딩하고 다른 매개변수를 추가 속성으로 제시한다. MonoDIS 의 L1 loss 를 사용하였으며 이는 다중 손실 함수의 가중 합계를 통해 7DoF 매개 변수를 예측하는 방식과 달리 3D 예측에 대한 기여도에 따라 손실을 측정한다.

Monocular 의 RTM3D 와 유사한 방법을 사용하여 정육면체 8 개의 투영 지점을 회귀시킨 후 수직 가장자리 높이를 사용하여 3D 정육면체를 형성한다. 이는 monocular Pseudo-Lidar 보다 더 좋은 결과가 나타난다.

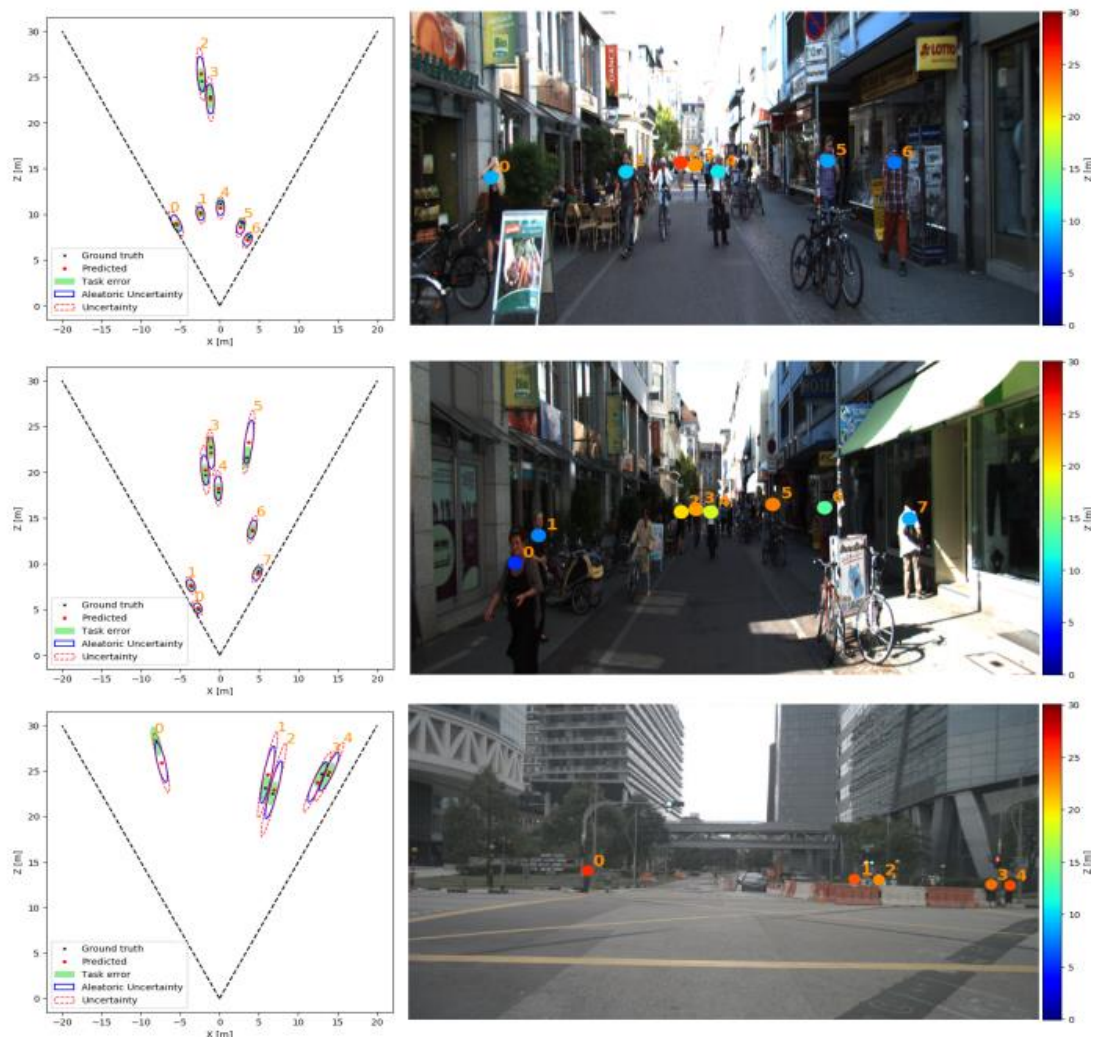


Figure 3 Illustration of results from KITTI

Monoloco 는 단안 RGB 영상으로부터 촬영된 보행자의 3 차원 위치 인식에서의 근본적인 원인을 바라보고 해결하는 알고리즘이다. 해당 알고리즘은 경우 보행자의 위치를 파악하는데 중점을 두었으며 단순히 보행자의 위치를 점으로 보고 추정하는

것이 아니라 Laplace distribution 를 기반으로 하는 손실 함수를 이용하여 신뢰 구간을 예측함으로써 작업의 모호성을 해결하는 방식을 사용한다.

다층 퍼셉트론을 사용하여 불확실성에 대한 예측을 높인 알고리즘이다. 즉 2D 영상을 추출할 시 Lidar 데이터를 기반으로 3D 정보 유추가 가능하다.

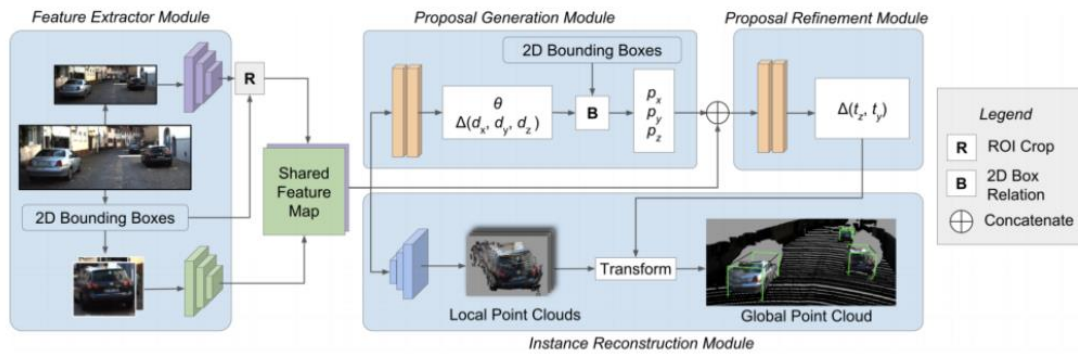


Figure 4 Network Architecture

MonoPSR 은 3D 생성한 다음 동적 객체의 로컬 포인트 클라우드를 재구성한다. 중심 제안 단계에서는 2D 상자 높이와 회귀된 3D 개체 높이를 사용하여 깊이를 추론하고 2D 경계 상자 중심을 예상 깊이의 3D 공간으로 다시 투사하는 방식을 사용한다. proposal sage 의 경우 오차가 1.5m 정도로 낮다. 3D 물체 감지를 위해 MonoGRNet 과 TLNet 의 점을 반영하여 사용하는 방식이다.

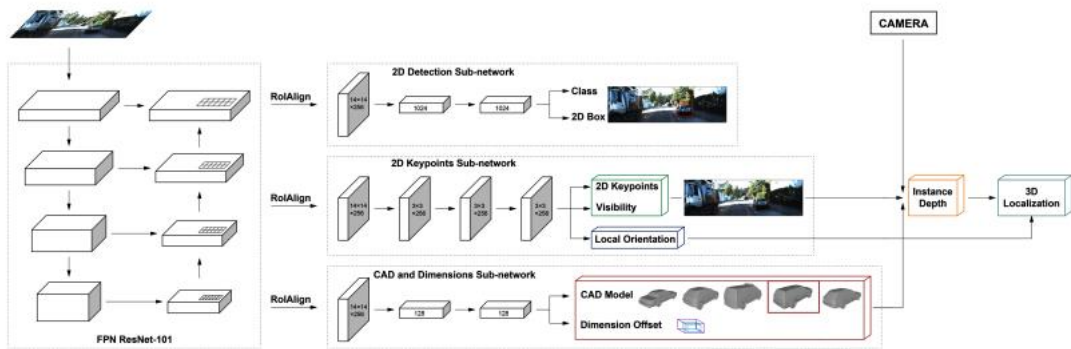


Figure 5 An overview of our monocular 3D detection architecture

MonoGRNet 의 경우 깊이 정보의 손실을 줄이기 위하여 단일 RGB 이미지에서 3D 객체 감지를 위해 새로운 키포인트 기반 접근법을 제안하였다. 이미지에서 2D 키포인트 감지를 중심으로 멀티 브랜치 모델을 구축하고 개념적으로 간단한 기하학적 추론 방법으로 보완하는 알고리즘이다. 전체 3D 포즈와 함께 2D 경계 상자, 키포인트 및 방향과 같은 2D 특성을 동시에 상호 의존적으로 추정하는 방식으로 수행된다.

참고문헌

- [1] 보행자 인식 기술 동향, 신현철, IDEC 기술동향컬럼, 8-9.
- [2] 지능형 자동차를 위한 야간보행자 검출 기술, 김성호, 전자공학학회지, 2017(8), 605-612.
- [3] Monocular 3D Object Detection for Autonomous Driving, Xiaozhi Chen, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, 2147-2156.
- [4] Monoloco: Monocular 3d pedestrian localization and uncertainty estimation, Bertoni, Lorenzo, Sven Kreiss, and Alexandre Alahi, Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.
- [5] Histograms of oriented gradients for human detection, N. Dalal and B. Triggs, In Proc. of CVPR 2005.
- [6] Real-time stereo visual odometry for autonomous ground vehicles, A. Howard, IEEE/RSJ International Conference on Intelligent Robots and Systems, 2008, 3946-3952, doi: 10.1109/IROS.2008.4651147.
- [7] Monocular 3D Object Detection Leveraging Accurate Proposals and Shape Reconstruction, Jason Ku, Alex D. Pon, Steven L. Waslander, Computer Vision and Pattern Recognition, 2019
- [8] Monocular 3D Object Detection via Geometric Reasoning on Keypoints, Ivan Barabanau, Alexey Artemov, Evgeny Burnaev, Vyacheslav Murashkin, Computer Vision and Pattern Recognition, 2019
- [9] SMOKE: Single-Stage Monocular 3D Object Detection via Keypoint Estimation, Zechen Liu, Zizhang Wu, Roland Tóth, Computer Vision and Pattern Recognition, 2020
- [9] https://github.com/xown3197/3D_Pedestrian_Localization_2021ComputerVision