KI ABSICHERUNG
Safe AI for Automated Driving

Sensor Fusion for Robust Pedestrian Detection and Human Pose Estimation

Michael Fürst, Rene Schuster,
Didier Stricker

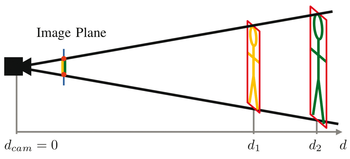
DNN specific Safety Concern: Brittleness of DNNs

Introduction

Sensors have specific strengths which need to be considered for optimal DNN performance.

Depth Ambiguity with Monocular Camera

In a single camera, two pedestrians of different size can appear identical if they are at different distances. This effect makes depth perception for monocular approaches hard.



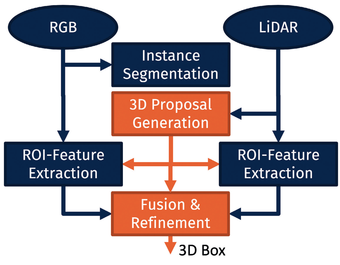
Resolution Limitations of LiDAR

LiDAR has accurate depth perception but lacks resolution. Above 30 m distance pedestrians can have 1-7 measurements, rendering them unrecognizable in the LiDAR.

Specific Strengths of Sensors:
RGB for Recognition: The camera has a high resolution with many pixels/measurements even for distant objects. LiDAR for Localization: LiDAR sensors measure the distance of surface points with centimeter precision, however, only for a few points.

Long Range Pedestrian Detection (LRPD)

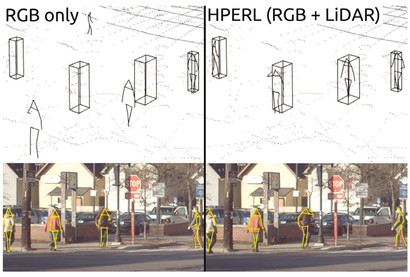
Using the principle above we derived LRPD [1] with strong long-range detection on KITTI [2].



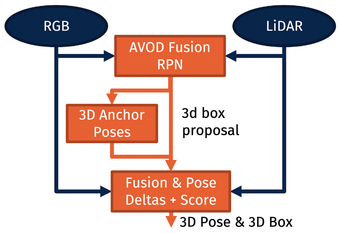
Specifically, the mAP over 30 m distance was significantly improved: PointPillars [3] with LiDAR only has 4.24 mAP, AVOD[4] using fusion achieves 12.55 and our LRPD [1] even 15.63 in the moderate difficulty category of KITTI [2]. The appropriate usage of sensors in the DNN has a significant impact on the performance.

Human Pose Estimation using RGB and LiDAR

RGB and LiDAR for 3D pose estimation remove depth ambiguity and increase precision.



Our HPERL [5] uses the AVOD [4] fusion RPN and the LCR-Net++ [6] head.



In HPERL the center depth error (CDE) decreases from 4.88 to 0.95 meters as well as the error in the orthogonal XY-Plane (XYE).

TABLE I
COMPARISON OF RGB BASELINE VS HPERL ON PEDX. LIDAR SIGNIFICANTLY IMPROVES THE PRECISION OF 3D LOCATION (1/5 CDE, 1/3 XYE). 2D RESULTS IMPROVE SLIGHTLY (MPJPE AND PCKH@0.5).

Model	Type	2D MPJPE	PCKh	CDE	XYE
RGB Base. [ours] (RGB only)	2D	87.76px	65.02%	-	-
	3D	87.66px	65.92%	4.88m	1.44m
HPERL [ours] (RGB + LiDAR)	2D	45.66px	70.08%	-	-
	3D	**45.65px**	**70.22%**	**0.95m**	**0.39m**

Conclusion

By fusing sensor appropriately, the performance can significantly improve compared to using single sensors or other fusion.

References:

[1] M. Fuerst et al. „Lrpd: Long range 3d pedestrian detection leveraging specific strengths of lidar and rgb.“ ITSC, 2020.
[3] A. H. Lang et al. „Pointpillars: Fast encoders for object detection from point clouds.“ CVPR, 2019.
[4] J. Ku et al. „Joint 3d proposal generation and object detection from view aggregation.“ IROS, 2018.
[5] M. Fürst et al. „HPERL: 3d human pose estimation from RGB and lidar.“ ICPR, 2021.
[6] G. Rogez et al. „Lcr-net++: Multi-person 2d and 3d pose detection in natural images.“ T-PAMI, 2019.

For more information contact:
Michael.Fuerst@dfki.de
Didier.Stricker@dfki.de

KI Absicherung is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Climate Action.

Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag

https://www.ki-absicherung-projekt.de @KI_Familie KI Familie

Image sources: Fig. 1: DFKI GmbH; Fig. 2: DFKI GmbH; Fig. 3: DFKI GmbH / 2019 UM & Ford Center for Autonomous Vehicles (FCAV); Fig. 4: DFKI GmbH; Fig. 5: (table) DFKI GmbH