HPERL: 3D Human Pose Estimation from RGB and LiDAR

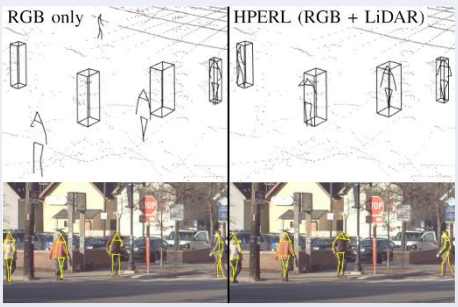
Michael Fürst[1], Shriya T. P. Gupta[2], René Schuster[1], Oliver Wasenmüller[1], Didier Stricker[1]

[1] DFKI – German Research Center for Artificial Intelligence
[2] Birla Institute of Technology and Science (BITS) Pilani

Submission-ID: 2020

Position

- Human Pose Estimation in the Wild has great opportunities in animation, action recognition, intention recognition and prediction for autonomous driving
- Current State-of-the-art focuses on RGB lacking global precision
- LiDAR sensors provide centimeter precision
- Human Pose Estimation using both RGB and LiDAR provides high precision and enables new use cases



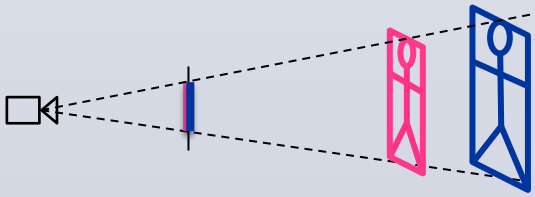
3D Pose Estimation

Task:
- Prediction of the joints of a human in 3D Space
- A point $\boldsymbol{x} = (x, y, z)^T$ describes the position of the joint
- Most approaches only predict relative to the position of the hip of the person. In contrast to them we predict the 3D positions relative to the observer.

→ Opportunity to use for autonomous vehicles and more

Methodologies:
- **Bottom Up:** Predict the position of the joints in heatmaps and then construct skeletons for the persons from the joints.
- **Top Down:** First find the persons in the scene, then predict the joint positions per person.
  - LCR-Net[3,4] with its Faster-RCNN[1]-like architecture was most influential to our work
  - First an RPN predicts 2D Boxes, then a refinement predicts deltas to 3D pose anchors

Limitations of RGB:
- **Depth Ambiguity**: In an RGB image, a large person (blue) further away than a smaller person (pink) can occupy the same pixel space. Given only the image a correct depth estimation proves difficult.



Sensor Fusion RGB and LiDAR in 3D Detection

Importance of LiDAR
- RGB-only approaches lack behind LiDAR + RGB approaches
- Best Detectors on KITTI are mostly LiDAR + RGB Fusion approaches
- LRPD [5] shows characteristics of fusion schemas for pedestrians

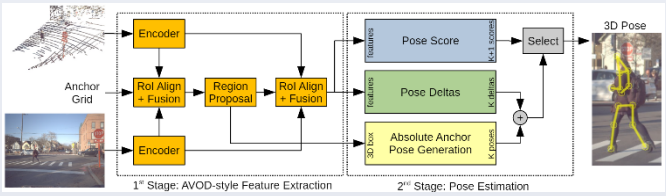
Reference Fusion Schema
- AVOD [2] is a Faster-RCNN[1]-like two stage approach
- First ROI crop using the anchors, fuse and predict region proposals
- Secondly ROI crop using proposals, fuse and predict the final boxes
- This simple schema can be adapted to other tasks as well

HPERL Architecture

- Two Stage Approach (Faster-RCNN[1]-like)

1st Stage:
- Fusion of LiDAR and RGB
  ROI align operation, then averaged or concatenated
- 3D Proposal Generation

2nd Stage:
- **Pose Refinement:**
  Predict deltas to add onto the absolute anchor poses generated
- **Anchor Scoring:**
  Estimate scores for each absolute anchor pose



Sensor Fusion in Human Pose Estimation

- Improve pose estimation by using LiDAR and RGB

Advantages:
- **Localization:**
  Using LiDAR enables high localization precision
- **Occlusion:**
  Combination of LiDAR and RGB helps in occlusion cases, as the LiDAR with the elevated position suffers less from occlusion
- **Precise Pose:**
  Using multiple sensor modalities allows for the precision in pose estimation of RGB and precise 3D understanding with LiDAR

Evidence:
- PedX Pose Estimation Dataset (largest best: PCKh, smallest: else)

	Method	Type	2D MPJPE	PCKh	Center Dist. Err.	XY Error
RGB	RGB-Baseline *(tuned LCR-Net[4])*	2D	87.76px	65.02%	-	-
		3D	87.66px	65.92%	4.88m	1.44m
Fusion	HPERL [ours]	2D	45.66px	70.08%	-	-
		3D	**45.65px**	**70.22%**	**0.95m**	**0.39m**

References

1. Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE International Conference on Computer Vision, 2015.
2. J. Ku, M. Mozifian, J. Lee, A. Harakeh, and S. L. Waslander, "Joint 3d proposal generation and object detection from view aggregation," in IEEE/RSJ International Conference on Intelligent Robots and Systems, 2018.
3. G. Rogez, P. Weinzaepfel, and C. Schmid, "Lcr-net: Localization-classification-regression for human pose," in IEEE Conference on Computer Vision and Pattern Recognition, 2017.
4. G. Rogez, P. Weinzaepfel, and C. Schmid, "Lcr-net++: Multi-person 2d and 3d pose detection in natural images,"in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019.
5. M. Fuerst, O. Wasenmueller, and D. Stricker, "Lrpd: Long range 3d pedestrian detection leveraging specific strengths of lidar and rgb," in IEEE International Conference on Intelligent Transportation Systems, 2020.

Contact Information

Michael Fürst
michael.fuerst@dfki.de
http://av.dfki.de/members/fuerst/

René Schuster
rene.schuster@dfki.de
http://av.dfki.de/members/schuster/