Pedestrian Environment Model for Automated Driving

Adrian Holzbock, Alexander Tsaregorodtsev, and Vasileios Belagiannis

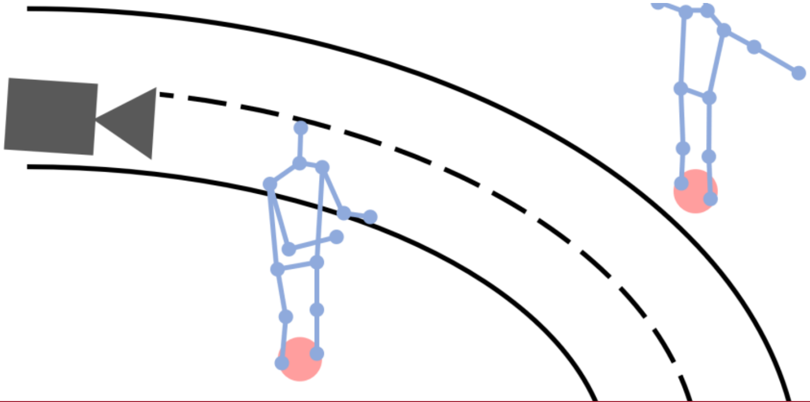
This presentation has been held at the 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), 24-28 September 2023, Bilbao, Bizkaia, Spain.

Citation information of the original publication:

A. Holzbock, A. Tsaregorodtsev and V. Belagiannis, "Pedestrian Environment Model for Automated Driving," 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), Bilbao, Spain, 2023, pp. 534-540, doi: 10.1109/ITSC57777.2023.10422032.

Citation information of the open-access publication:

A. Holzbock, A. Tsaregorodtsev and V. Belagiannis (2023): Pedestrian Environment Model for Automated Driving. Open Access Repositorium der Universität Ulm und Technischen Hochschule Ulm . https://dx.doi.org/10.18725/OPARU-52409



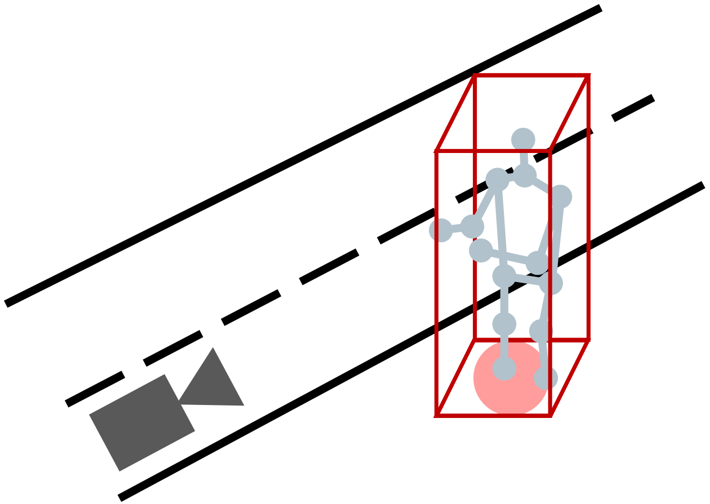
Pedestrian Environment Model for Automated Driving

Adrian Holzbock[1], Alexander Tsaregorodtsev[1], and Vasileios Belagiannis[2]

[1]Institute of Measurement, Control and Microtechnology, Ulm University
[2]Chair of Multimedia Communications and Signal Processing, Friedrich-Alexander-Universität Erlangen-Nürnberg

Motivation

- Automated vehicles have to fully recognize the environment to safely navigate.
- Pedestrians use gestures to communicate with the automated vehicle.
- The special needs of pedestrians are not supported by current environment models (grid maps and object lists).
- Gesture recognition, human behavior understanding, or body pose forecasting can be performed on the pedestrians environment model.



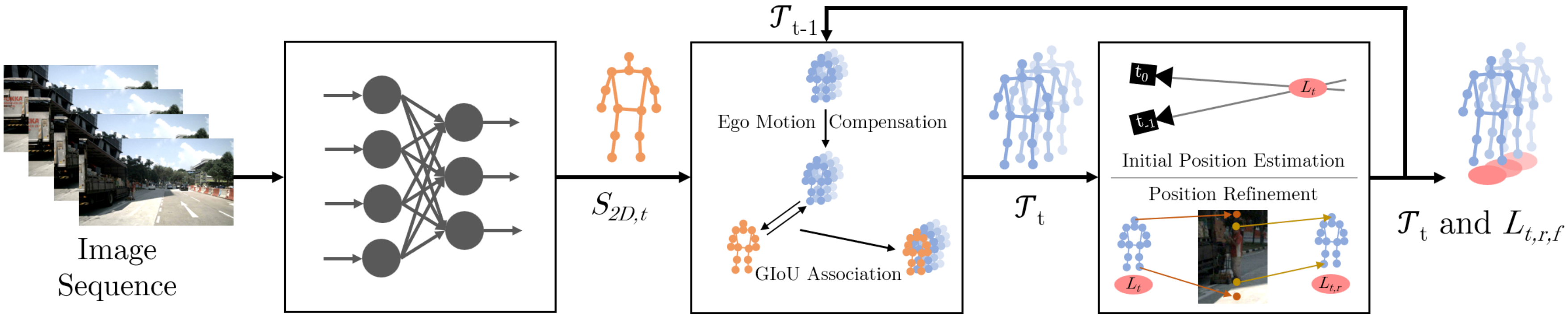
Method Overview

Three steps to generate the pedestrian environment model:

1. Skeleton Extraction
2. Person Tracking
3. World Position Estimation

Input: Image sequence and ego motion data

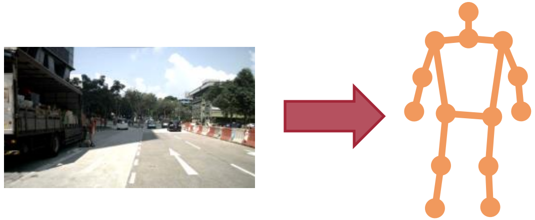
Output: Sequence of poses with position in the world coordinate system



Skeleton Extraction & Person Tracking

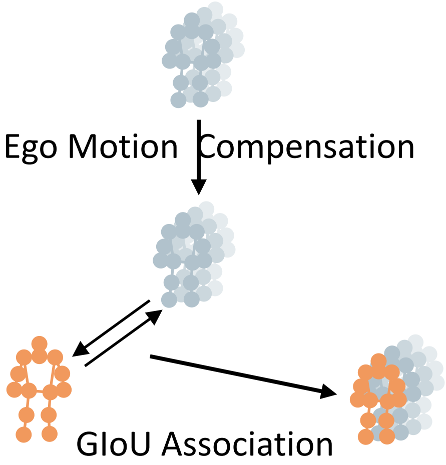
Skeleton Extraction:

- CID[1] human pose estimation network trained on the COCO dataset.
- Pedestrian detection and human pose estimation combined in one network.
- Returns besides the human pose also for every keypoint a detection probability.



Person Tracking:

- Transfer the skeleton data of the last time steps to the current time step with a ego-rotation compensation that uses self-localization data.
- Assign the detected pedestrians from the current time step to them of previous time steps. (Use the Hungarian Algorithm with the negative Generalized Intersection over Union as cost)

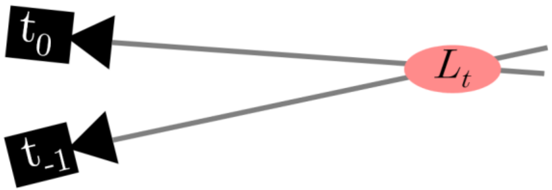


[1] Wang, Dongkai, and Shiliang Zhang. "Contextual instance decoupling for robust multi-person pose estimation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

World Position Estimation

Divided into two steps:

- **Initial Position Estimation:**
  - Use the self-localization data to determine rays for all keypoints.
  - Calculate the intersection of the keypoint rays between the perspectives of two consecutive frames.
- **Position Refinement:**
  - Initial estimation can be erroneous due to pedestrian‘s motion, no ego motion, or inaccuracies in the pose estimation.
  - Refinement of the initial position with re-projection into the image plane and comparison of the pedestrians height.



Experimental Settings

CARLA Dataset [2]:

- Dataset simulated with the CARLA simulator [2]
- Each sequence contains RGB images, ego-motion data, 2D skeletons, and 3D bounding boxes with a frequency of 10Hz.
- Use for the evaluation the generated 2D skeletons.

nuScenes Dataset [3]:

- Evaluate on the nuScenes mini-set-split.
- Camera images with a frame rate of 12 Hz, whereas labels are only available at a frequency of 2Hz.
- Detect the 2D skeletons with the CID pose estimator [1].

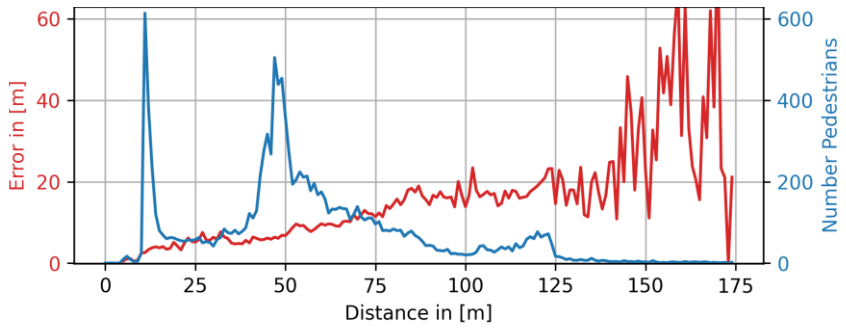
Evaluation Metrics:

- e_{abs}: absolute distance error in meters between the vehicle and the pedestrian.
- e_{rel}: relative distance error regarding to the distance between the vehicle and the pedestrian.

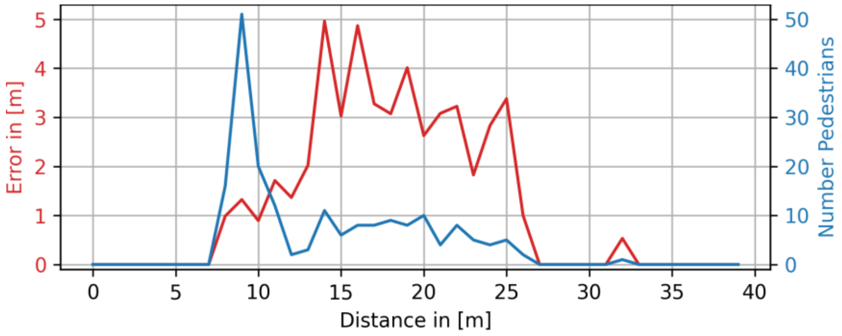
[1] Wang, Dongkai, and Shiliang Zhang. "Contextual instance decoupling for robust multi-person pose estimation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
[2] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, “CARLA: An open urban driving simulator,” in Proceedings of the 1st Annual Conference on Robot Learning, 2017, pp. 1–16.
[3] H. Caesar, et al., “nuscenes: A multimodal dataset for autonomous driving,” arXiv preprint arXiv:1903.11027, 2019.

Evaluation

Data	Absolute Error e_{abs} in [m]	Relative Error e_{rel} in [%]	Number Pedestrians
CARLA all	9.123	16.85	12771
nuScenes all	2.524	15.66	194



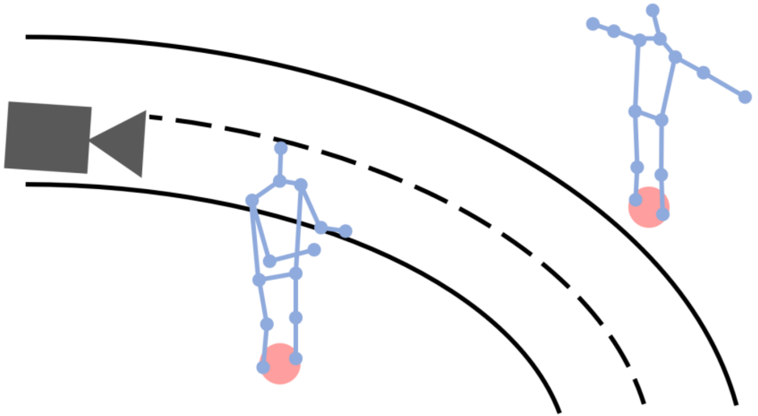
CARLA Dataset



nuScenes Dataset

Conclusion

- Our pedestrian environment model contains besides the humans position in world coordinates also skeletal data which are important for further pedestrian depending tasks.
- The pedestrian environment model is generated only with RGB images from a monocular camera and self-localization data.
- The approach is evaluated on a dataset generated with the CARLA simulator and the nuScenes mini-set-split.



Thank you for your attention!

Adrian Holzbock

Institute of Measurement, Control and Microtechnology

Ulm University

adrian.holzbock@uni-ulm.de

We gratefully acknowledge the funding of this work:

EVENTS

Funded by the European Union

This project has received funding under grant agreement No 101069614. It is funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.