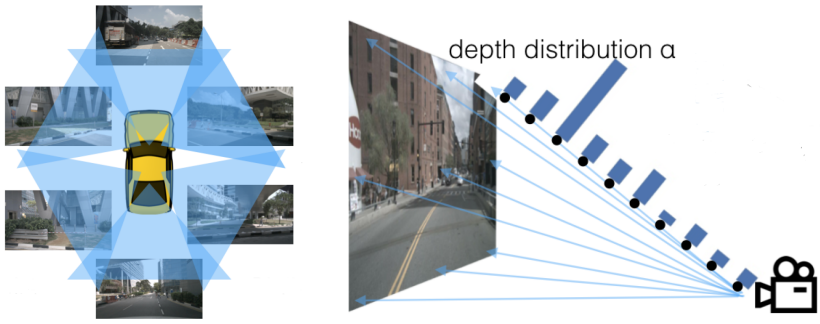
Karlsruhe Institute of Technology

Master Thesis



Analysis and improvement of view transformations for automated driving

3D perception of traffic participants is a crucial task for automated driving. Multiple neural network architectures have been proposed in the last decade to tackle this task. Recently, the transformation of camera images into the birds-eye-view (BEV) is a popular design element of such architectures, which is part of most state of the art methods. Since this transformation requires the depth information for every pixel in the image, most networks introduce a learnable depth estimation. One pioneering work of this approach is Lift-Splat-Shoot [2].

In [1], Harley et al. show that the transformations using such an explicit depth estimation are inferior to other methods in performance and computational efficiency, suggesting the depth estimation has no benefit. However, only methods with a sparse or no depth supervision have been considered in this study, leaving the question whether depth in general or sparse depth is not beneficial for BEV transformation.
The proposed master thesis aims to investigate whether depth improves the BEV transformation, and subsequently the 3D perception, and if so, which depth properties are important.

The proposed thesis consists of the following parts:

+ Literature research about BEV-camera-based 3D object detection networks
+ Literature research about automotive datasets with dense ground truth depth maps, e.g SHIFT [3]
+ Implementation of a BEV-camera-based 3D object detection network utilizing the dense ground truth depth maps
+ Investigation of the influencing properties of the depth information for the 3D perception task by ablation of the network.

I am happy to answer any questions if you are interested. Feel free to ask for an appointment!

[1] Adam W. Harley, Zhaoyuan Fang, Jie Li, Rares Ambrus, and Katerina Fragkiadaki. Simple-BEV: What Really Matters for Multi-Sensor BEV Perception? Technical report, September 2022. arXiv:2206.07959 [cs] type: article.

[2] Jonah Philion and Sanja Fidler. Lift, Splat, Shoot: Encoding Images From Arbitrary Camera Rigs by Implicitly Unprojecting to 3D. Technical report, August 2020. arXiv:2008.05711 [cs] type: article.

[3] Tao Sun, Mattia Segu, Janis Postels, Yuxuan Wang, Luc Van Gool, Bernt Schiele, Federico Tombari, and Fisher Yu. SHIFT: A Synthetic Driving Dataset for Continuous Multi-Task Domain Adaptation. Technical report, June 2022. arXiv:2206.08367 [cs] type: article.

Institute of Measurement and Control Systems (MRT)
Prof. Dr.-Ing. Christoph Stiller

Advisor:
Thorsten Herd, M.Sc.

Programming language(s)[1]:

C++	advanced
Python	advanced

System, Framework(s):
Linux, PyTorch

Required skills:
- Prior knowledge of deep learning for CV
- Work on your own

Language(s):
German, English

For more information please contact:

Thorsten Herd

Email:
thorsten.herd@partner.kit.edu

Or directly send in your application including your current grades as well as our questionnaire!

---

[1] **skill levels:**

beginner	< 500 lines of code (LOC)
advanced	500 – 5000 LOC
proficient	> 5000 LOC