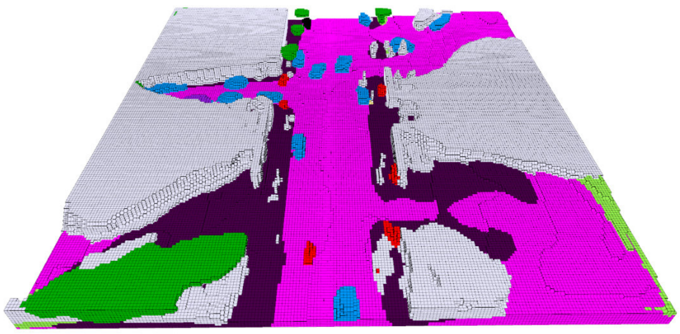
3D Occupancy Prediction from Multi-View 2D Surround Images



Master's Thesis

Institute of Photogrammetry and Geoinformation
Leibniz University Hannover

Samer M. R. Abualhanud
Matriculation Number 10006569

Hanover, 31. March 2024

Examiner	Prof. Dr.-Ing. habil. Christian Heipke
Supervisor	Dr.-Ing. Max Mehltretter
Bosch Supervisor	Eashwara Sudharsan Erahan

Abstract

An accurate geometric 3D representation of the environment is essential in various autonomous driving tasks such as path planning and obstacle avoidance. This thesis delves into 3D perception within autonomous driving, moving away from traditional geometric sensors like LiDAR towards vision-based perception methods that utilize camera sensors for comprehensive geometric and semantic scene understanding. Despite LiDAR's precision in creating 3D maps of the environment, its prohibitive cost and the sparse nature of its point clouds have prompted a shift towards camera-based solutions. These solutions provide rich visual information but face challenges in accurately recovering 3D geometry. Our work focuses on 3D occupancy prediction, reconstructing scene geometry as a voxel grid where each voxel is assigned both an occupancy status and a semantic label. Utilizing color images from multi-view surround cameras, their camera parameters, and the vehicle's ego pose, we learn a model capable of predicting 3D occupancy maps including semantics from unseen images during training. We explore supervising the network with 2D labels to reduce the costs associated with generating and annotating 3D ground truth, while addressing the challenges presented by dynamic objects. We propose a fully self-supervised model that does not depend on any geometric ground truth, demonstrating significantly improved performance over current state-of-the-art self-supervised occupancy prediction models on the Occ3D dataset [Tia+24].