TEK5030 Project Report

Bryan Walther, Didrik Samdanger, Carlos Ankour Navarro

May 24, 2023

1 Introduction

Monocular depth estimation in computer vision is the problem of extracting depth information of a scene from single images captured by monocular cameras. This is a challenging problem because the ability to perceive depth is achieved through stereo vision, which adds a constraint to the problem of estimating the depth. In contrast, monocular depth estimation is an ill-posed problem, since we attempt to infer depth information of a scene using only visual cues of a single image. Most modern approaches to monocular depth estimation involves using large datasets with labeled images to train deep learning models, where the ground truth depth information for each sample is included in the training set. A popular model that does this is MiDaS, which is trained on up to 12 different datasets, and predicts the relative depth map for a given image. However, the predicted depth maps only offer depth information up to scale, meaning that we face the issue of scale-ambiguity. Some work has been done on attempting to learn the absolute depth of a scene, a notable example being ZoeDepth. ZoeDepth attempts to learn depth information while also maintaining the metric scale. After testing ZoeDepth however, we find that the accuracy of the model in predicting the true scale is generally poor.

In this project we attempt to tackle the problem of scale-ambiguity in estimated depth-maps by leveraging prior knowledge of the real world geometry of the scene. As a proof of concept, we apply this idea to traffic footage, where the goal is to get good distance estimates from a dash-camera to any object in the scene.

2 Methods

TODO: Explain the approach for getting the correct scale of a depth map by detecting license plates with known real world dimensions.

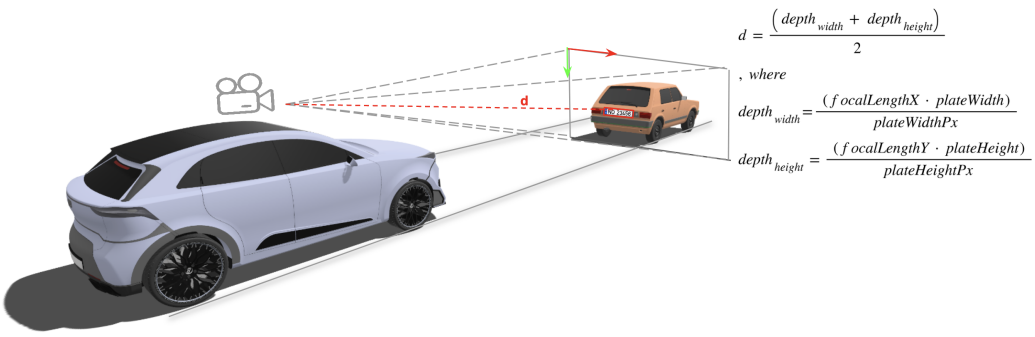

Figure 1: Depth estimation of license plates

3 Results

4 Discussion

5 Conclusion