COSE474-2024F: Final Project Proposal
Depth Estimation from Monocular Image

Sanghwa Song

1. Introduction

Perceiving the surrounding environment is a crucial challenge in autonomous driving and drone navigation. While a variety of sensors are used for this purpose, devices like LiDAR are often expensive. This motivates the use of single RGB cameras to estimate depth, offering a more cost-effective solution.

The aim is to develop a reliable system that estimate accurate depth information based on RGB image. This approach can lower expenses of using other specialized sensors. RGB-based depth estimation can sometimes be used with other sensors, enhancing overall performance and improving the robustness of autonomous systems.

2. Problem definition & challenges

Given a single RGB image $I \in R^{W \times H \times 3}$, the task is to estimate a depth map $D \in R^{H \times W}$. Our goal is to find a function $f : I \to D$ that maps input image to depth map.

Monocular Depth Estimation is challenging because there's no parallax information like in stereo vision, leading to ambiguity in depth prediction. The model must rely on less reliable visual cues like lighting, shading, and object size, which can vary across scenes. Additionally, it requires learning priors about objects to make accurate predictions, and multiple valid depth maps can correspond to the same image, making it difficult to ensure a unique and correct solution. The model should overcome the ambiguity and estimate a depth map that is plausible to human expectations.

3. Related Works

Monodepth (Godard et al., 2017) introduced an unsupervised approach to monocular depth estimation using left-right image consistency, allowing models to learn depth without ground-truth depth maps.

DORN (Fu et al., 2018) introduced a novel ordinal regression approach for depth estimation, framing the task as an ordering problem rather than a traditional regression task.

MiDaS (Ranftl et al., 2020) uses models pretrained on large-scale image datasets and then fine-tunes them for depth estimation tasks across multiple datasets. It achieves robust performance across various datasets and environments.

In addition to the above research, there have been a lot of studies conducted with remarkable progress in the field. However, I believe exploring research in other areas of 3D computer vision, such as NeRF and mesh generation, will provide us with additional insights and inspire new ideas.

4. Datasets

The NYU-Depth V2 is a large dataset from indoor scenes. It consists of RGB images paired with depth maps captured from the Microsoft Kinect.

The KITTI dataset is a popular dataset focused on outdoor scenes. It contains RGB images and corresponding depth maps captured by LiDAR sensors mounted on a car.

5. State-of-the-art methods and baselines

The model was selected considering both benchmark performance and influence based on the number of GitHub stars. Marigold (Ke et al., 2024) uses diffusion model designed for monocular depth estimation. It leverages rich knowledge modern generative image models have. Depth Pro (Bochkovskii et al., 2024) uses ViT encoders on patches extracted at multiple scales. It fuses the patch predictions into a single high-resolution depth map.

As a baseline we can use one of the models mentioned in the related work. I will use Monodepth as a baseline due to its simplicity and its significance as an early contribution to depth estimation using deep learning approach.

6. Schedule

Current - 10/31 Investigate further into various approaches

11/01 - 11/14 Design methodology and architecture

11/15 - 11/30 Conduct experiments

12/01 - Deadline Write a report

References

Bochkovskii, A., Delaunoy, A., Germain, H., Santos, M., Zhou, Y., Richter, S. R., and Koltun, V. Depth pro: Sharp monocular metric depth in less than a second, 2024. URL https://arxiv.org/abs/2410.02073.

Fu, H., Gong, M., Wang, C., Batmanghelich, K., and Tao, D. Deep ordinal regression network for monocular depth estimation, 2018. URL https://arxiv.org/abs/1806.02446.

Godard, C., Aodha, O. M., and Brostow, G. J. Unsupervised monocular depth estimation with left-right consistency, 2017. URL https://arxiv.org/abs/1609.03677.

Ke, B., Obukhov, A., Huang, S., Metzger, N., Daudt, R. C., and Schindler, K. Repurposing diffusion-based image generators for monocular depth estimation, 2024. URL https://arxiv.org/abs/2312.02145.

Ranftl, R., Lasinger, K., Hafner, D., Schindler, K., and Koltun, V. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer, 2020. URL https://arxiv.org/abs/1907.01341.