Low Complexity Networks and Edge Enhancement for Monocular Depth Estimation

A Thesis submitted

in partial fulfillment for the degree of

Doctor of Philosophy

by

Sandip Paul

Department of Avionics
Indian Institute of Space Science and Technology
Thiruvananthapuram, India

July 2024

Abstract

Estimating depth from a 2D image is valuable for robotics, navigation, object recognition, medical diagnosis, and 3D measurements. Mobile platforms often have limitations in size, weight, and power. Cost-effective monocular depth estimation with a single camera and low computational requirements are suited for such applications. Most research on depth recovery lacks work on depth range, dynamic targets, and practical setups. The thesis investigates various depth recovery methods to understand such gaps.

The analysis of experiments using depth from focus (DOD) found that blur detectors can estimate sparse depth maps, but the depth estimate is only accurate for close-range targets due to the shallow depth of field. Additionally, shadows can create false depth and artifacts not previously reported in the literature. This method is not useful for featureless surfaces or moving targets.

A coded aperture using two spectral filters (DFCCA) was developed, resulting in a larger 32-pixel disparity map compared to the 14 pixels reported in the literature. However, the accuracy was again only high for near-targets. The analysis showed that performance is sensitive to spectral leakage, signal-to-noise ratio (SNR), and reduced image resolution. The multi-coded aperture (DMCA) relies on non-overlapping Point Spread Function (PSF) signatures. Here, the depth estimate was found to be accurate for surfaces with dense textures but not suitable for moving targets due to the requirement of two images.

Most depth-recovery methods estimate relative depths. A new calibration method has been developed to recover absolute depth from monocular images. This approach uses a unique blur target with ground truth. The method is not affected by contrast variations, magnification artifacts, or spectral sensitivity. The blur ranges up to a radius of 1.2σ, as demonstrated, but is limited by the inherent camera optic blur.

Modern Deep Neural Networks (DNNs) estimate depth from single images. The key components are network architecture and network training loss functions. Regression loss for robust training is less researched. Here, new loss functions based on edge and Structural Similarity (SSIM) functions were proposed. These improved the log_{10} error and the $\delta 1$ accuracy (85%) emphasizing the training robustness.

Available Deep Learning Networks are computationally intensive for use in mobile platforms. A low-complexity multi-scale network architecture (NDWTN) is designed.

NDWTN has wavelets, attention, dense convolution, residual convolution, batch norm, and efficient activation layers to estimate the full-depth map. This network outperforms previously known DWT-based and UNET++ models in all six performance metrics and provides the best RMSE score among present state-of-the-art models. NDWTN trains faster within 17 minutes per epoch and reaches an accuracy of $> 92\%$ for 10 epochs. These are suitable for conservative mobile systems.