Learned Optics for Imaging and Depth Estimation

Stanford EE367/CS448I Final Project Proposal

Bin (Claire) Zhang
Department of Electrical Engineering
Stanford University
zhangbin@stanford.edu

Megan Zhang
Department of Electrical Engineering
Stanford University
lzhang8@stanford.edu

1 Project description

Introduction. Over the years, there have been numerous attempts to solve the monocular depth estimation (MDE) problem. Learning scene depth from 2D RGB images is crucial to the development of many modern technologies, such as autonomous driving [1] and augmented reality [2]. Conventional approaches involve learning pictorial depth cues from training images [3] or exploiting the depth from defocus (DfD) approach [4] that often requires coded apertures to capture depth information.

Some of the most popular approaches design and deploy an effective neural network architecture. Most end-to-end (E2E) MDE approaches, however, only rely on pictorial depth cues, instead of defocus blur and spatial patterns indicated by occlusions boundaries, which are the proven more effective depth cues [5].

Goal. For this final project, we aim to explore a newly proposed nonlinear occulsion-aware optical images formation model [6] that advances the previous state-of-the-art in MDE. More specifically, we want to study how different point spread function (PSF) modeling techniques in the model helps with the downstream MDE task.

Related Work.

1.1 Learned Optics for MDE

[6] proposed an E2E training paradigm to jointly optimize the optics parameters and the neural network for depth map prediction. Specifically, the PSF of the imaging system is derived using variations in the surface height of the diffractice optical element (DOE), and is thus depth-dependent. This framework is shown to have superior depth estimation quality among E2E computational imaging techniques.

1.2 PSF Modeling with Zernike Polynomials

Zernike polynomials are mathematical bases defined in a circular support, and commonly used in optical imaging to characterize the effect of lens systems.

Methods(Task). We would like to study how different PSF modeling strategies could lead to different depth estimation results within the framework of [6]. Concretely, we will substitute the depth-dependent 3D PSF created by a lens profile with weighted sum of Zernike polynomials. The Zernike coefficients, which can be seen as the weight parameters here, will be jointly optimized in this E2E pipeline. We will analyze the effect of different PSFs on the depth estimation task, and whether they are aided by depth from defocus or occlusion awareness.

Data. As suggested by [6], we will mainly use the FlyingThings3D subset of SceneFlow dataset [7], which provides depth maps aligned with RGB input images. We may additionally use the DualPixels dataset [8].

Baselines. Our model with the Zernike-based PSF modeling will be compared with the following methods: (1) **Ground Truth (GT)**, (2) **All in Focus:** Results obtained from applying the depth estimation CNN employed in [6] to GT images, (3) **DfD:** Results obtained from applying the depth estimation CNN employed in [6] to sensor images with the conventional defocus blur technique, and (4) **Ikoma et al. [6]:** Results obtained from applying the entire E2E nonlinear occulsion-aware optical images formation model [6] to 2D RGB images. We will either test on the publicly available models or use the previously published scores.

Evaluation. The evaluation metrics used in this project will be similar to how the comparisons are done in [6]. For the estimated GT RGB images, we will use MAE, PSNR. Note that the DfD approach does not produce such images. The evaluation metrics we will use to evaluate the depth maps are MAE, RMSE, $\log_1 0$, $\delta < 1.25$, $\delta < 1.25^2$, and $\delta < 1.25^3$.

Milestones. The project timeline is as follows:

- 2/18 (Fri) Project proposal
- Run baseline models
- Set up code for PSF generation and training with Zernike polynomials
- Model evaluation and analysis
- 3/9 (Wed) Project poster or video presentation
- 3/11 (Fri) Project report and code

References

[1] N. Metni, T. Hamel, and F. Derkx. Visual tracking control of aerial robotic systems with adaptive depth estimation. In *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 6078–6084, 2005.

[2] W. Woo, Wonwoo Lee, and Nohyoung Park. Depth-assisted real-time 3d object detection for augmented reality. 2011.

[3] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. *CoRR*, abs/1406.2283, 2014.

[4] Pauline Trouvé, Frédéric Champagnat, Guy Le Besnerais, Jacques Sabater, Thierry Avignon, and Jérôme Idier. Passive depth estimation using chromatic aberration and a depth from defocus approach. *Applied optics*, 52 29:7152–64, 2013.

[5] Marina Zannoli, Gordon D. Love, Rahul Narain, and Martin S. Banks. Blur and the perception of depth at occlusions. *Journal of vision*, 16 6:17, 2016.

[6] Hayato Ikoma, Cindy M. Nguyen, Christopher A. Metzler, Yifan Peng, and Gordon Wetzstein. Depth from defocus with learned optics for imaging and occlusion-aware depth estimation. In *2021 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12, 2021.

[7] N. Mayer, E. Ilg, P. Häusser, P. Fischer, D. Cremers, A. Dosovitskiy, and T. Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. arXiv:1512.02134.

[8] Rahul Garg, Neal Wadhwa, Sameer Ansari, and Jonathan T. Barron. Learning single camera depth estimation using dual-pixels. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.