# Depth from Defocus Approaches for Video Depth Estimation

Zhengyang Wei

February 17, 2022

## 1 Motivation

Video depth information is important for robotics, autonomous driving, 3D reconstruction, and beyond. Depth information can be acquired by some expensive sensors like LIDAR, stereo cameras. Generating high-quality depth-from-color can inexpensively complement these sensors.

Depth from defocus approaches are used for single image depth estimation and outperform many state-of-art methods. However, defocus blur hasn't been applied to video depth estimation yet. Video depth estimation considers not only the depth of a single frame but also the Spatio-temporal relationship between adjacent frames, which improves the performance of depth estimation. Therefore, we want to explore how to combine the defocus blur cues and the consistency of video frames to achieve a better depth estimation.

## 2 Related Work

### 2.1 Depth from defocus method for MDE

Besides pictorial cues, defocus blur is an important depth cue for monocular depth estimation. Using images with defocus blur in deep learning approaches outperforms the use of all-in-focus images [CSTP+18]. End-to-End coded aperture was used to improve the defocus blur and encode more information [WBC+19]. Ikoma et al. proposed a framework for single RGB image depth estimation, including an occlusion-aware image formation model, a rotationally symmetric phase-coded aperture, and the corresponding preconditioning approach [INM+21].

### 2.2 Video depth estimation

Different from depth estimation for a single image, video depth estimation usually take information between frames into account. Monodepth2 uses minimum reprojection loss to tackle occlusions between frames and auto-masking loss to ignore stationary pixels [GMFB19]. Packnet has symmetrical packing and unpacking blocks and it uses the neighbor frames s temporal context to realize self-supervised scale-aware structure-from-motion [GAP+20]. SC-Depth penalizes the inconsistency of predicted depths of adjacent with frames geometry consistency loss [BZW+21]. M4Depth maintains the Spatio-temporal consistency with time recurrence and motion information [FED21]. Robust CVD jointly optimize camera poses as well as depth deformation in 3D and resolve fine-scale details using a geometry-aware depth filter [KRH21].

## 3 Final Goals

The project will focus on setting up a framework including depth from defocus deep networks and the video consistency models. We will compare the performances of different video depth estimation models modified by the depth from defocus module with chosen datasets and design different loss functions to adapt the modified structure.

# 4 Milestones and Timeline

Week 7:

- Review previous approaches for depth from focus and video depth estimation.
- Try to understand the source code corresponding to the approaches.
- Submit Project Proposal (2/18/2022).

Week 8:

- Choose suitable datasets for the project.
- Set up video depth map estimation framework for selected methods..
- SImplement the algorithms and train models.

Week 9:

- Continue to implement the algorithms and train models.
- Conduct the ablation study.

Week 10

- Analyze the experiment results and write the report.
- Design the poster.

# References


[BZW+21] Jia-Wang Bian, Huangying Zhan, Naiyan Wang, Zhichao Li, Le Zhang, Chunhua Shen, Ming-Ming Cheng, and Ian Reid. Unsupervised scale-consistent depth learning from video. *International Journal of Computer Vision (IJCV)*, 2021.

[CSTP+18] Marcela Carvalho, Bertrand Le Saux, Pauline Trouvé-Peloux, Andrés Almansa, and Frédéric Champagnat. Deep depth from defocus: how can defocus blur improve 3d estimation using dense neural networks?, 2018.

[FED21] Michael Fonder, Damien Ernst, and Marc Van Droogenbroeck. M4depth: A motion-based approach for monocular depth estimation on video sequences. May 2021.

[GAP+20] Vitor Guizilini, Rares Ambrus, Sudeep Pillai, Allan Raventos, and Adrien Gaidon. 3d packing for self-supervised monocular depth estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[GMFB19] Clément Godard, Oisin Mac Aodha, Michael Firman, and Gabriel J. Brostow. Digging into self-supervised monocular depth prediction. October 2019.

[INM+21] 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. *IEEE International Conference on Computational Photography (ICCP)*, 2021.

[KRH21] Johannes Kopf, Xuejian Rong, and Jia-Bin Huang. Robust consistent video depth estimation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[WBC+19] Yicheng Wu, Vivek Boominathan, Huaijin Chen, Aswin Sankaranarayanan, and Ashok Veeraraghavan. Phasecam3d — learning phase masks for passive single view depth estimation. In *2019 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12, 2019.