Intensity-SPAD Sensor Fusion Depth Estimation

Zhanghao Sun
Department of Electrical Engineering
Stanford University
zhsun@stanford.edu

1 Introduction

Depth estimation is critical in lots of promising applications such as auto-vehicle navigation, geometric detection and robotics. Intensive researches have been done on both time-of-flight(ToF) hardware and corresponding signal processing algorithms[1-5]. Recently, an end-to-end neural network is demonstrated to out-perform other approaches in the task of single photon avalanche detector(SPAD) denoising [1]. In this work, 3D SPAD data is processed with a multi-scale CNN with hints provided by 2D intensity image of the same scene. Inspired by it, we propose a different network structure that better fuses the 3D-2D sensor data, which results in a large improvement in depth estimation robustness and accuracy.

2 Related Work

2.1 SPAD denoising

SPAD denoising has long been a challenging and open problem because of the extremely low signal-noise ratio(SNR) and ambient light interference. Most previous works are based on Poisson denoising algorithms [4,5], while the neural network approaches is only explored by David et al. in [1]. In this work, We'll design a stronger neural network to further improve its performance.

2.2 Monocular depth estimation

Human visual systems have the ability of monocular depth estimation (based on textures, occlusion boundaries and so on). Similarly, monocular depth estimation networks only take a single RGB image as input and generate complicated features such as occlusion boundaries, surface normals and 2D depth maps []. Based on these researches, we believe that the information in intensity image is by far fully employed in [1].

2.3 Depth-RGB sensor fusion

Recently, depth-RGB sensor fusion networks is receiving more and more attention in a large variety of fields including depth inpainting [2] and depth super-resolution [6]. Among them, layer-by-layer fusion within a multi-scale network [2,3] remains to be the best architecture. However, all of these networks take a clean 2D depth map as input (though they might be very sparse or have large holes) and thus the sensor fusion is between 2D data. However, a 3D-2D fusion is needed for SPAD raw data denoising, which is much more difficult.

EE367: Computational Imaging and Display, Winter, 2019, Stanford University, CA.

3 Approaches

We noticed that there are basically two aspects that requires improvements in David's previous work [1].

Firstly, they utilized the KL divergence between network predicted SPAD detection rate and ground truth as the only loss function. However, KL loss is insufficient when the two input probability distributions have no overlap. It won't generate a gradient that scales with the distance between these two distributions, which is obviously a drawback in back propagation. Therefore, we'll try out other loss functions such as Wasserstein loss [6], Ordinal regerssion loss [7] and so on to give a more sensitive gradient feedback.

Secondly, in the previous work, intensity image is only served as a regularization. On the other hand, it is already demonstrated that fairly accurate depth maps can be obtained from a single 2D image. Therefore, we'll implement a layer-by-layer fusion architecture all through the multi-scale network [2,3]. Moreover, considering the memory constraints, we'll experiment 3D encoders to compress the huge SPAD data input and make the network deeper.

4 Milestones and Timeline

Week 6-7: Implement and experiment different loss functions

Week 8-10: Implement sensor fusion network

5 References

References

[1] D. Lindell: Single-photon 3D imaging with deep sensor fusion. ACM Trans. Graph. 37 (2018).

[2] Y. Zhang: DeepLiDAR: Deep Surface Normal Guided Depth Prediction for Outdoor Scene from Sparse LiDAR Data and Single Color Image. arXiv:1812.00488 (2018).

[3] F. Ma: Self-supervised sparse- to-dense: Self-supervised depth completion from lidar and monoc- ular camera," arXiv:1807.00275, (2018).

[4] J. Rapp: A few photons among many: Unmixing signal and noise for photon-e cient active imaging. IEEE Trans. Computat. Imaging 3 (2017).

[5] D. Shin: Photon-efficient computational 3D and reflectivity imaging with single-photon detectors. IEEE Trans. Comput. Imaging 1 (2015).

[6] C. Frogner: Learning with a Wasserstein loss. Advances in Neural Information Processing Systems, (2015).

[7] H. Fu: Deep ordinal regression network for monocular depth estimation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2018).