Light Field Depth Map Estimation

Sarah Xu

EE 367 - Project Proposal
Winter 2021

Motivation

Depth estimation has been and continues to be an important topic in computer vision, it provides critical geometric information about the scene that is being captured, and it has many applications in robotics, autonomous driving, gaming, etc. Light field models the light ray as a 5D plenoptic function of position, angle, wavelength, and time dimensions [1]. It was simplified to a 4D function as radiance is constant along a ray in empty space [2]. Compared to conventional 2D images, the light field contains both spatial and angular information and multiple cues about the scene that can be used for refocusing and estimating the depth map.

With the recent advances in commercializing the light field micro-lens array devices, depth map estimation algorithms have been one of the focused research areas. Since current light field devices capture each pair of sub-aperture images with a very narrow baseline, the disparity of the images and the spatial resolution are restricted, making the accuracy of depth recovery limited. Depth map estimation is also very challenging with non-Lambertian surfaces, i.e. specular materials and with occlusion. In this project, we want to explore various cues, algorithms, and methods that leverage the structure and correlations in spatial, angular, (and temporal) dimensions to produce a depth map estimation from the light field data.

Related Work

There has been a plethora of research on depth map estimation from light field, and they can be subdivided into three categories: 1) sub-aperture image matching; 2) EPI-based methods; 3) Learning-based methods [3]. For the sub-aperture image matching, following Ng et al. [4], Tao et al. used defocus cue, correspondence cue and shading cue to estimate the depth of the scene [5 - 6]. Furthermore, many research works have been done for estimating the depth in scattering medium, occlusion, noise and for glossy materials. [7 - 10] Heber et al. also proposed a method that shears the light field view, and considered each warped image as a row in a matrix that is low rank [11]. In the EPI-based methods, the EPIs are generated from the light field images, and the slopes of the lines are used to measure the depth of different objects [3, 12].

As to the learning-based methods, Heber et al. proposed a CNN-based method that learns an end-to-end mapping and predicts depth information for light field data, and later improved from

5-layer CNN to a U-Net based method [13, 14]. However, the CNN-based method is restricted by the amount of training light field data with accurate ground truth depth map labels, which are difficult to obtain. Recent research tries to address this limitation by data augmentation, training an attention module, and unsupervised or zero-shot feature CNN that exploit the correlations between spatial, angular, and temporal dimensions [15 - 17].

Project Overview

The project will focus on setting up an end-to-end light field depth map estimation pipeline. The pipeline will take in the light field data as input, and outputs the depth map of the light field. We want to explore a CNN-based method to extract the underlying correlations between different views of the light field, since CNN is powerful in learning the relations between data structures. To bypass the limitation of scarce training data and to allow the algorithm to be more generalized to light field data captured using different devices, we want to impose prior assumptions and constraints, and employ a zero-shot or unsupervised learning-based method to estimate the light field depth map.

Milestones

Week 6: 2/15/2021 - 2/21/2021

- Literature review:
    - Review previous methods and approaches.
    - Explore other potential methods we can use to solve the problem.
- Find available public datasets for the project (real world and rendered) [18 - 20].
- Submit Project Proposal (2/17/2021).

Week 7: 2/22/2021 - 2/28/2021

- Data Collection and preprocessing (if needed).
- Setting up depth map estimation pipeline structure and (cloud) infrastructure.
- Establish baseline.

Week 8: 3/1/2021 - 3/7/2021

- Implementation and comparison of different algorithms. Start training.

Week 9: 3/8/2021 - 3/14/2021

- Analyze preliminary results. Revise approach and conduct ablative study if necessary.

Week 10: 3/15/2021 - 3/19/2021

- Analyze and summarize the results.
- Submit Report and code. (3/19/2021, 11:59 pm)

References

[1] Gershun, A. (1939). The light field. *Journal of Mathematics and Physics*, *18*(1-4), 51-151.

[2] Levoy, M., & Hanrahan, P. (1996, August). Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (pp. 31-42).

[3] Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., & Hanrahan, P. (2005). *Light field photography with a hand-held plenoptic camera* (Doctoral dissertation, Stanford University).

[4] Wu, G., Masia, B., Jarabo, A., Zhang, Y., Wang, L., Dai, Q., ... & Liu, Y. (2017). Light field image processing: An overview. *IEEE Journal of Selected Topics in Signal Processing*, *11*(7), 926-954.

[5] Tao, M. W., Hadap, S., Malik, J., & Ramamoorthi, R. (2013). Depth from combining defocus and correspondence using light-field cameras. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 673-680).

[6] Tao, M. W., Srinivasan, P. P., Malik, J., Rusinkiewicz, S., & Ramamoorthi, R. (2015). Depth from shading, defocus, and correspondence using light-field angular coherence. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1940-1948).

[7] Tian, J., Murez, Z., Cui, T., Zhang, Z., Kriegman, D., & Ramamoorthi, R. (2017). Depth and image restoration from light field in a scattering medium. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2401-2410).

[8] Williem, W., & Park, I. K. (2016). Robust light field depth estimation for noisy scene with occlusion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4396-4404).

[9] Lin, H., Chen, C., Kang, S. B., & Yu, J. (2015). Depth recovery from light field using focal stack symmetry. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3451-3459).

[10] Tao, M. W., Su, J. C., Wang, T. C., Malik, J., & Ramamoorthi, R. (2015). Depth estimation and specular removal for glossy surfaces using point and line consistency with light-field cameras. *IEEE transactions on pattern analysis and machine intelligence*, *38*(6), 1155-1169.

[11] Heber, S., & Pock, T. (2014, September). Shape from light field meets robust PCA. In *European Conference on Computer Vision* (pp. 751-767). Springer, Cham.

[12] Bolles, R. C., Baker, H. H., & Marimont, D. H. (1987). Epipolar-plane image analysis: An approach to determining structure from motion. *International journal of computer vision*, *1*(1), 7-55.

[13] Heber, S., & Pock, T. (2016). Convolutional networks for shape from light field. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3746-3754).

[14] Heber, S., Yu, W., & Pock, T. (2016, September). U-shaped Networks for Shape from Light Field. In *BMVC* (Vol. 3, p. 5).

[15] Shin, C., Jeon, H. G., Yoon, Y., Kweon, I. S., & Kim, S. J. (2018). Epinet: A fully-convolutional neural network using epipolar geometry for depth from light field images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4748-4757).

[16] Tsai, Y. J., Liu, Y. L., Ouhyoung, M., & Chuang, Y. Y. (2020, April). Attention-based view selection networks for light-field disparity estimation. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 07, pp. 12095-12103).

[17] Peng, J., Xiong, Z., Wang, Y., Zhang, Y., & Liu, D. (2020). Zero-shot depth estimation from light field using a convolutional neural network. *IEEE Transactions on Computational Imaging*, *6*, 682-696.

[18] The (New) Stanford Light Field Archive: http://graphics.stanford.edu/data/LF/lfs.html

[19] Honauer, K., Johannsen, O., Kondermann, D., & Goldluecke, B. (2016, November). A dataset and evaluation methodology for depth estimation on 4d light fields. In *Asian Conference on Computer Vision* (pp. 19-34). Springer, Cham.

[20] Alperovich, A., & Goldluecke, B. (2016, November). A variational model for intrinsic light field decomposition. In *Asian Conference on Computer Vision* (pp. 66-82). Springer, Cham.