U-shaped Networks for Shape from Light Field

Stefan Heber[1]
stefan.heber@icg.tugraz.at
Wei Yu[1]
wei.yu@icg.tugraz.at
Thomas Pock[1,2]
pock@icg.tugraz.at

[1] Graz University of Technology
Graz, Austria
[2] Austrian Institute of Technology
Vienna, Austria

Overview This paper presents a novel technique for Shape from Light Field (SfLF), that utilizes deep learning strategies. Our model is based on a fully convolutional network, that involves two symmetric parts, an encoding and a decoding part, leading to a u-shaped network architecture. By leveraging a recently proposed Light Field (LF) dataset, we are able to effectively train our model using supervised training. To process an entire LF we split the LF data into the corresponding Epipolar Plane Image (EPI) representation and predict each EPI separately. This strategy provides good reconstruction results combined with a fast prediction time. In the experimental section we compare our method to the state of the art. The method performs well in terms of depth accuracy, and is able to outperform competing methods in terms of prediction time by a large margin.

Contribution The proposed method is inspired by the method of Heber and Pock [2], that uses a conventional Convolutional Neural Network (CNN) in a sliding window fashion to predict depth information. They showed that CNNs have a large capacity to learn from data to predict the orientation of the lines in the EPIs. However, due to the sliding window approach, their method suffers from considerable high computational costs. Compared to [2] we were able to significantly reduce the computation time by predicting complete EPIs at once using u-shaped networks, cf. Figure 1. Besides drastically reducing the prediction time the proposed network architecture also allows to reduce the errors in homogeneous regions, because the proposed model can overcome the patch-nature of the network proposed in [2]. Our experiments demonstrate that the proposed method is able to predict an entire 4D disparity field within a few seconds, cf. Table 1. Moreover, due to the fact that our network architecture does not include any fully connected layer, our method also allows to process LFs with varying resolutions.

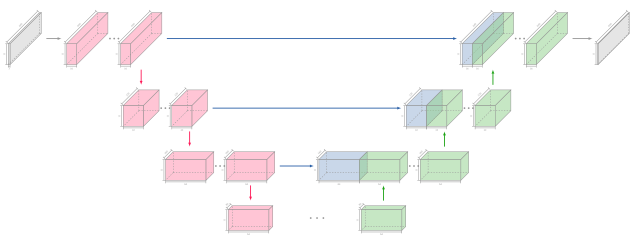

Figure 1: Illustration of the proposed u-shaped network architecture. The encoding and decoding parts of the network are highlighted in purple and green, respectively. The pinhole connections are marked in blue.

	Wanner [5]	Tao [4]	Heber [1]	Jeon [3]	Heber [2] (CNN)	proposed
RMSE	3.91	2.33	2.50	2.49	1.87	0.80
MAE	2.94	1.06	0.79	0.75	1.13	0.35
0.5%	22.00	16.32	8.47	9.64	17.96	7.34
0.2%	35.22	28.48	13.20	16.46	31.61	14.76
Time	3min 18s	23min 4s	4min 44s	2h 12min 30s	35s	2s
GPU	✓	✗	✓	✗	✓	✓

Table 1: Quantitative results for various SfLF methods averaged over 50 synthetic LFs. The table provides the RMSE, MAE, the percentage of pixels with a relative disparity error larger than 0.2% and 0.5%, and the computational time of the method.

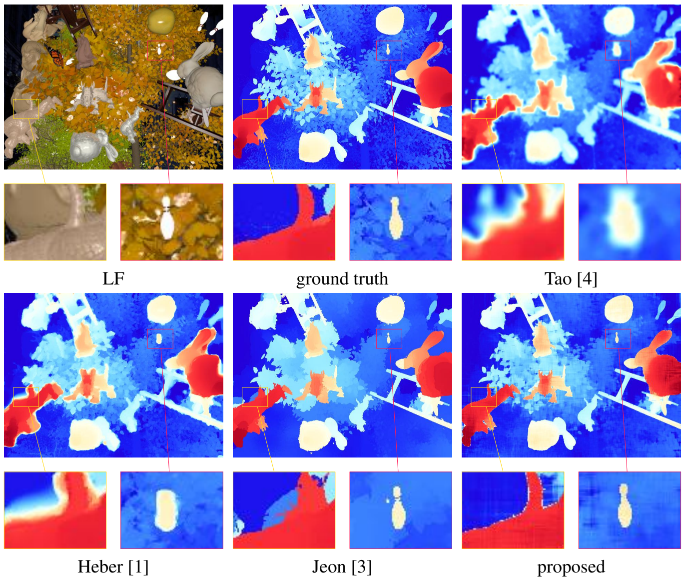

Figure 2: Comparison to state-of-the-art methods on the synthetic POV-Ray dataset. The figure shows the center view of the LF, the color-coded ground truth, the results for three state-of-the-art SfLF methods [1, 3, 4], followed by the result of the proposed method.

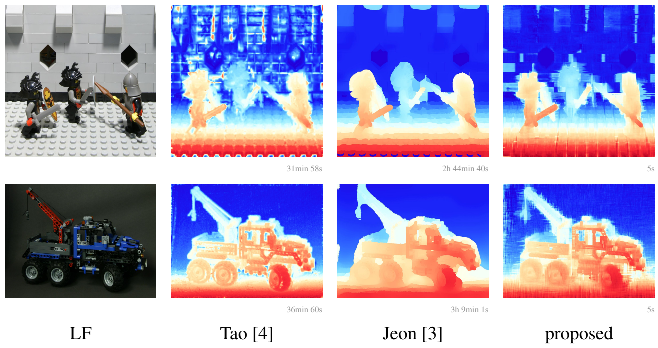

Figure 3: Qualitative comparison for LFs from the SLFA. The figure shows from left to right the center view of the LF, followed by the results for the methods proposed by Tao et al. [4] and Jeon et al. [3]. The results to the right correspond to the proposed method.

[1] Stefan Heber and Thomas Pock. Shape from light field meets robust PCA. In *Proceedings of the 13th European Conference on Computer Vision*, 2014.

[2] Stefan Heber and Thomas Pock. Convolutional networks for shape from light field. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[3] H. G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y. W. Tai, and I. S. Kweon. Accurate depth map estimation from a lenslet light field camera. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1547–1555, June 2015.

[4] Michael W. Tao, Sunil Hadap, Jitendra Malik, and Ravi Ramamoorthi. Depth from combining defocus and correspondence using light-field cameras. In *International Conference on Computer Vision (ICCV)*, December 2013.

[5] S. Wanner and B. Goldluecke. Globally consistent depth labeling of 4D light-fields. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.