Benchmark Dataset and Method for Depth Estimation from Light Field Images

(Supplementary material)

Note: Figures 1-2 show results on the rest of the proposed real light field dataset whereas Figures 3-5 show all the samples of the real dataset itself. Figure 6 shows results of DispNet [5].

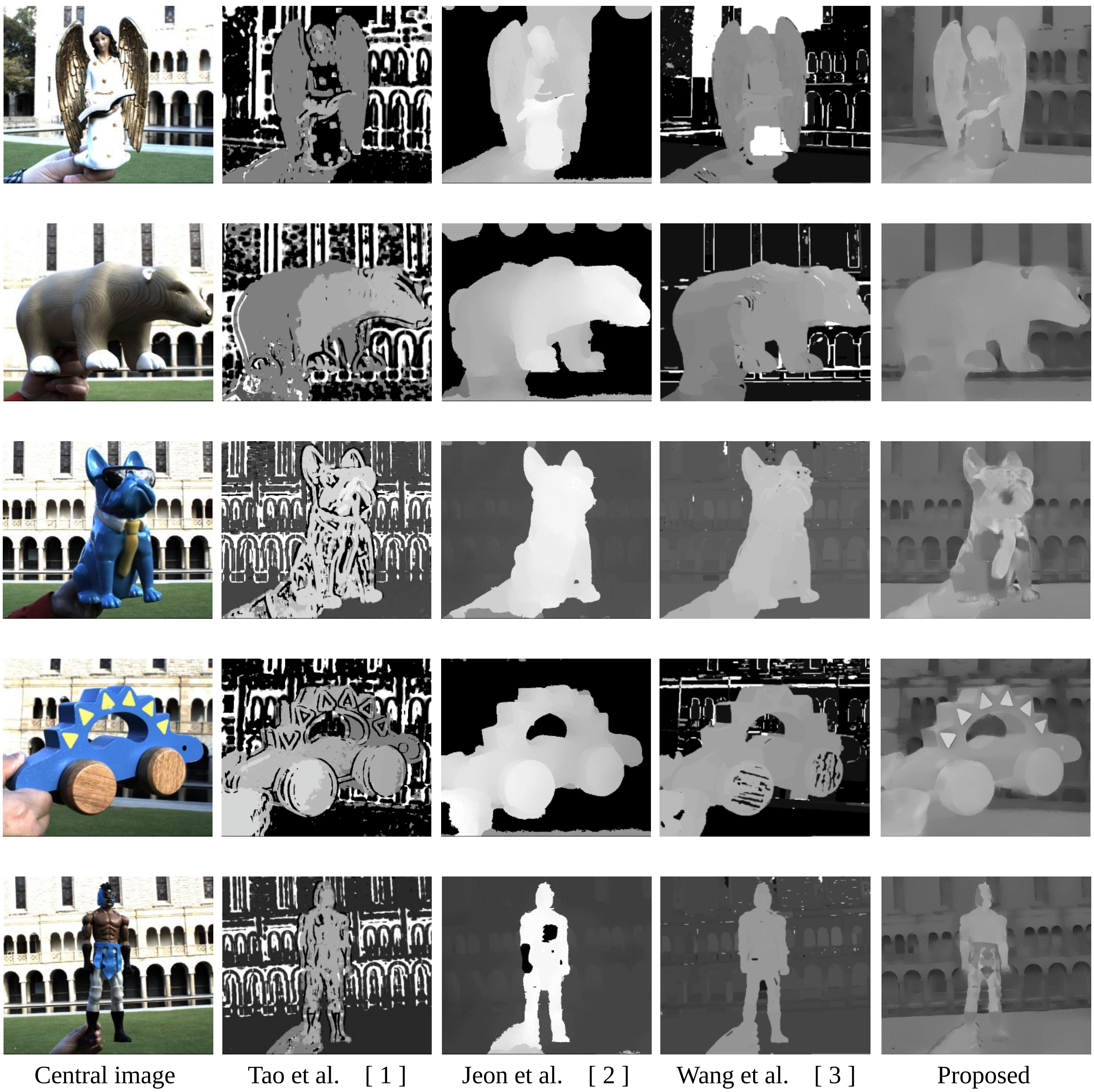
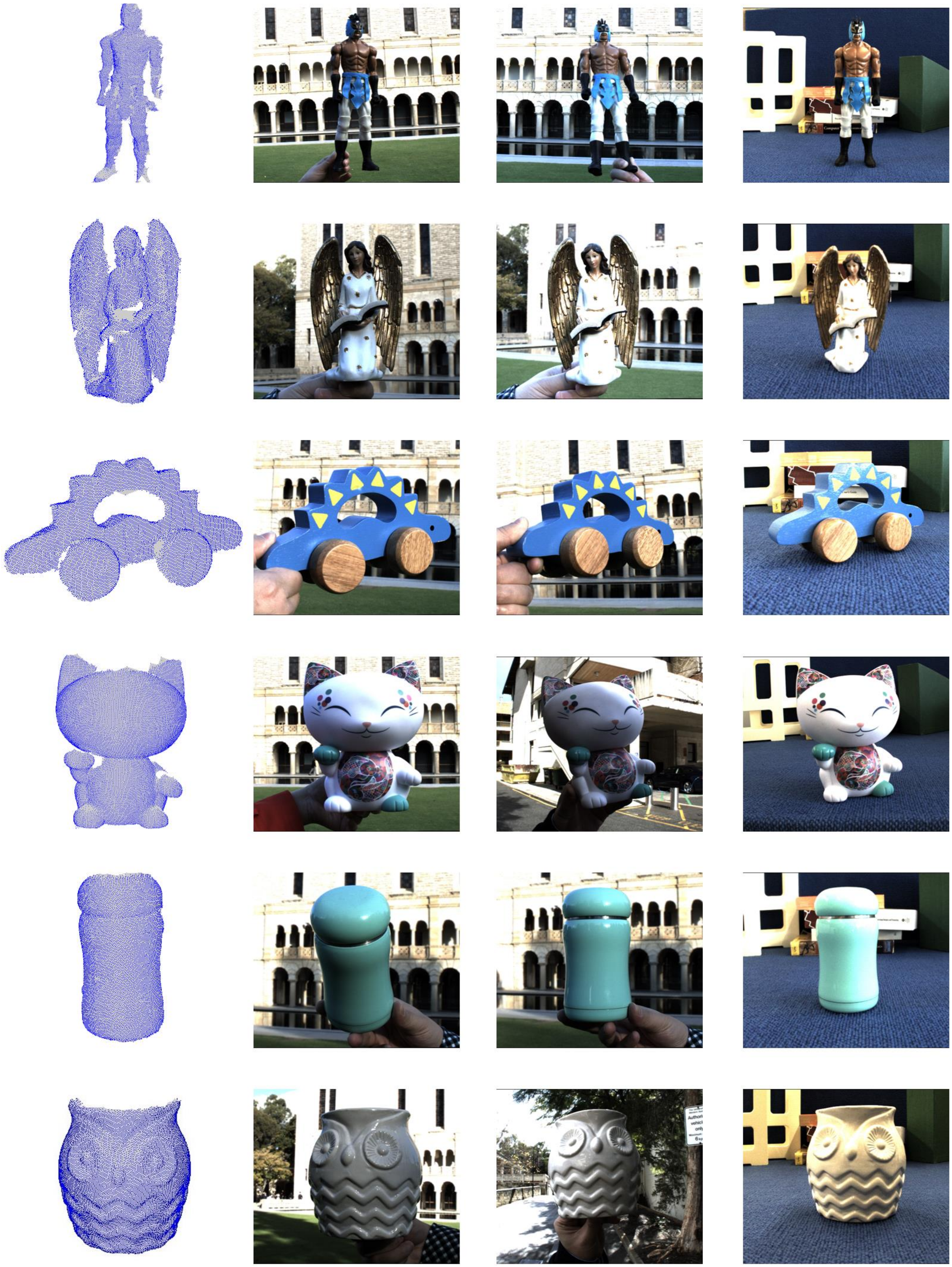


Fig.1. Evaluation on the proposed real dataset. The first column is central light field image and the following columns are the outputs from [1], [2], [3] and our proposed method respectively.

Fig.2. Evaluation on the proposed real dataset. The first column is central light field image and the following columns are the outputs from [1], [2], [3] and our proposed method respectively.

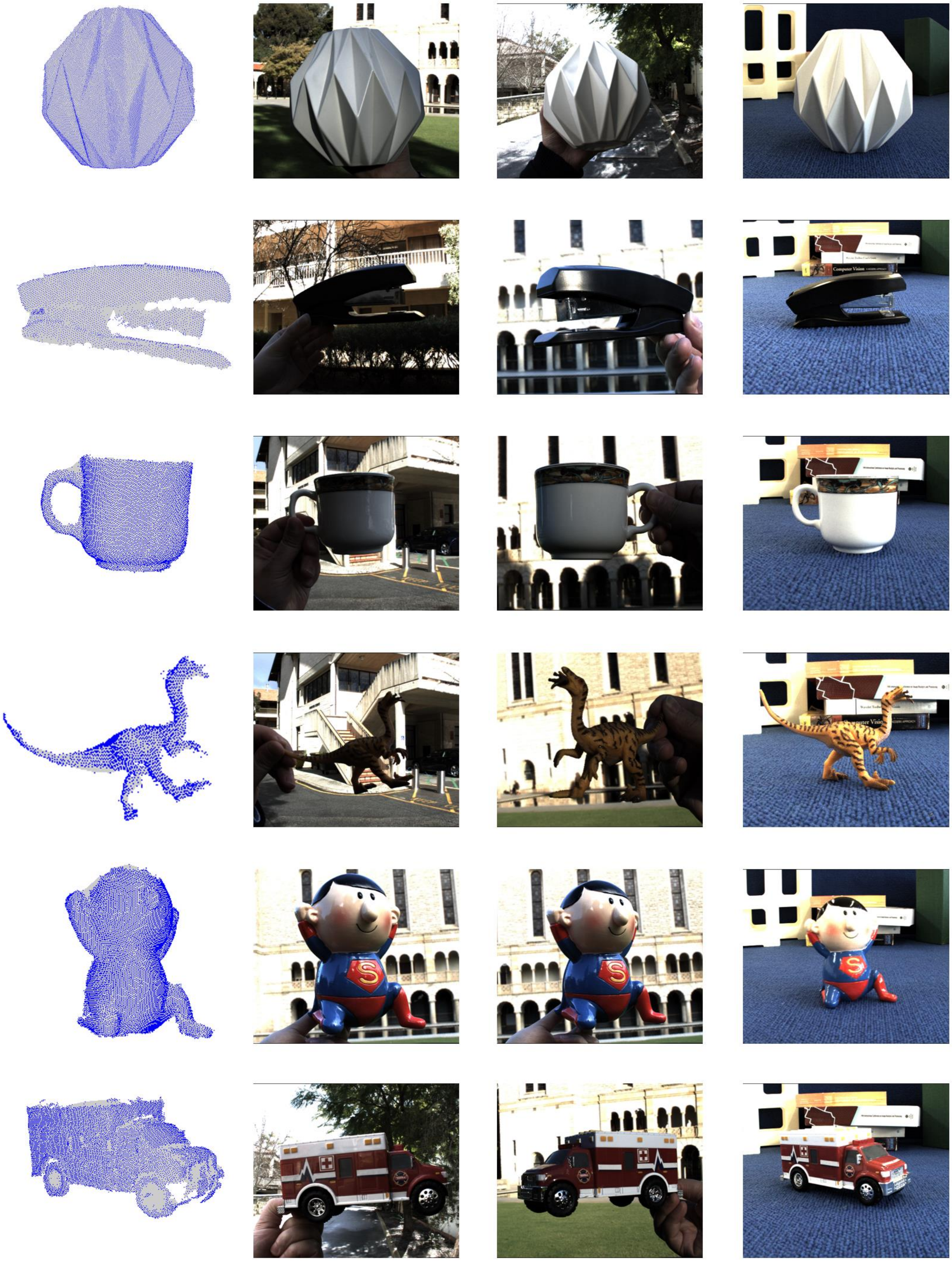
(a) (b) (c) (d)

Fig 3. The proposed real dataset. (a) 3D pointcloud; (b) and (c): center view of light field images captured outdoors; (d) center view of light field images captured indoors.



(a) (b) (c) (d)

Fig.4. The proposed real dataset. (a) 3D scan pointcloud; (b) and (c): center view of light field images captured outdoors; (d) center view of light field images captured indoors.



(a) (b) (c) (d)

Fig.5. The proposed real dataset. (a) 3D pointcloud; (b) and (c): center view of light field images captured outdoors; (d) center view of light field images captured indoors.

Table.1. Comparison of the estimated depths from different networks on the test images of the synthetic HCI dataset. For the second network, we added two pooling layers, one after each convolution layer which reduced the network to just two convolution layers (7x7 and 5x5 kernels). We trained this new network on exactly the same training data as we trained our original network on. We can see that our proposed network without pooling achieves better results.

	Town	Pillows	Medieval2	Kitchen	Dot
Heber and Pock's net output [4]					
RMSE	0.6917	0.8205	0.158	0.7846	0.5805
MAE	0.6121	0.7396	0.067	0.6668	0.5761
PSNR	51.3324	49.8492	64.1577	50.2378	52.8548
Two stream net with pooling layers					
RMSE	0.5873	0.671	0.1509	0.6083	0.569
MAE	0.5014	0.5817	0.072	0.5148	0.5536
PSNR	54.1271	52.8368	70.9842	53.898	53.2669
Proposed net output					
RMSE	**0.2589**	**0.165**	**0.1432**	**0.2739**	**0.2809**
MAE	**0.1649**	**0.1019**	**0.0506**	**0.1713**	**0.1761**
PSNR	**59.8682**	**63.7811**	**65.0119**	**59.379**	**59.1598**

Fig.6. Comparison of the depth maps produced by DispNet [5] (second row) and our method (third row) on the five test images of the HCI dataset.

References

1. M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, “Depth from combining defocus and correspondence using light-field cameras,” in Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 673–680.
2. H.-G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y.-W. Tai, and I. So Kweon, “Accurate depth map estimation from a lenslet light field camera,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1547–1555.
3. T.-C. Wang, A. A. Efros, and R. Ramamoorthi, “Occlusion-aware depth estimation using light-field cameras,” in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3487–3495.
4. S. Heber and T. Pock, “Convolutional networks for shape from light field,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3746–3754.
5. N. Mayer, E. Ilg, P. Hausser, 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 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp.4040–4048.