Supplementary Material - Learning Fused Pixel and Feature-based View Reconstructions for Light Fields

Jinglei Shi* Xiaoran Jiang* Christine Guillemot
INRIA Rennes - Bretagne Atlantique, France
{firstname.lastname}@inria.fr

In this document, additional materials are provided to supplement our main paper. In the first section, we provide the link for demo codes. In the second section, we compare the running time of different learning-based methods. In the third section, network structure is depicted. In the fourth section, more visual comparisons of the reconstructed views are given, which is not possible in the main submission due to the page limit.

Note that we also provide video files that show the whole reconstructed light fields both in the case of interpolation and that of extrapolation.

1. Test codes

Our demo code is available in the following github link:
https://github.com/JingleiSHI/FPFR. The complete source code will be available soon.

2. Runtime

We compare the runtime of the three learning-based reference methods[1, 2, 3] against ours. In Table 1, we record the running time for synthesizing a light field of resolution $768 \times 768 \times 7 \times 7$ on a GPU of type Nvidia Tesla V-100 with 32GB memory. (The method [3] is run on CPU Intel i7-7600U 2.80GHz 16GB RAM. We did not succeed in running it on GPU, and in their paper the running time comparison was also done with CPU.) The running time includes model loading and computation of color views.

Methods	DeepBW[1]	LLFF[2]	EPI[3]	FPFR	FPFR*
Runtime	778s	212s	1332s(CPU)	172s	684s

Table 1. Runtime for generating a $768 \times 768 \times 7 \times 7$ light field for learning-based methods.

3. Network structure

The structure details for layers of PixRNet and FeatRNet are depicted in the Table 2.

4. Additional qualitative results

We show in Fig. 1 (for synthetic light fields) and Fig. 2 (for real-world light fields) additional visual comparisons that are not included in the paper due to the page limit.

*equal contribution

PixRNet	***k***	***s***	***in/out***	***input***
conv3Px	3	1	16/64	*concat*($\{\widetilde{L}^{\mathbf{i}}_{\mathbf{t}}, m^{\mathbf{i}}_{\mathbf{t}}, \forall\mathbf{i}\}$)
conv2Px	3	1	64/32	*conv3Px*
conv1Px	3	1	32/16	*conv2Px*
$\hat{L}^{Pix}_{\mathbf{t}}$	3	1	16/3	*conv1Px*
FeatRNet				
conv3_1Ft	3	1	1028/256	*concat*($\{\widetilde{f}^{\mathbf{i},3}_{\mathbf{t}}, m^{\mathbf{i},3}_{\mathbf{t}}, \forall\mathbf{i}\}$)
conv3_2Ft	3	1	256/256	*conv3_1Ft*
$\hat{f}^{3}_{\mathbf{t}}$	3	1	256/256	*conv3_2Ft*
up3	4	2	256/32	$\hat{f}^{3}_{\mathbf{t}}$
conv2_1Ft	3	1	548/128	*concat*($\{\widetilde{f}^{\mathbf{i},2}_{\mathbf{t}}, m^{\mathbf{i},2}_{\mathbf{t}}, \forall\mathbf{i}\}$,*up3*)
conv2_2Ft	3	1	128/128	*conv2_1Ft*
$\hat{f}^{2}_{\mathbf{t}}$	3	1	128/128	*conv2_2Ft*
up2	4	2	128/16	$\hat{f}^{2}_{\mathbf{t}}$
conv1_1Ft	3	1	276/64	*concat*($\{\widetilde{f}^{\mathbf{i},1}_{\mathbf{t}}, m^{\mathbf{i},1}_{\mathbf{t}}, \forall\mathbf{i}\}$,*up2*)
conv1_2Ft	3	1	64/64	*conv1_1Ft*
$\hat{f}^{1}_{\mathbf{t}}$	3	1	64/64	*conv1_2Ft*
conv0_1Ft	3	1	64/32	$\hat{f}^{1}_{\mathbf{t}}$
conv0_2Ft	3	1	32/32	*conv0_1Ft*
$\hat{L}^{Feat}_{\mathbf{t}}$	3	1	32/3	*conv0_2Ft*
Fusion				
$\breve{f}^{Pix}$	3	1	16/8	*conv1Px*
$\breve{f}^{Feat}$	3	1	32/8	*conv0_2Ft*
conv3_Fn	3	1	16/8	*concat*($\breve{f}^{Pix}$,$\breve{f}^{Feat}$)
conv2_Fn	3	1	8/8	*conv3_Fn*
conv1_Fn	3	1	8/4	*conv2_Fn*
M	3	1	4/1	*conv1_Fn*

Table 2. The proposed network architecture. k, s and *in/out* represent the kernel size, the stride and the number of input/output channels, whereas '*up*' and '*concat*' represent upsampling by deconvolution and concatenation.

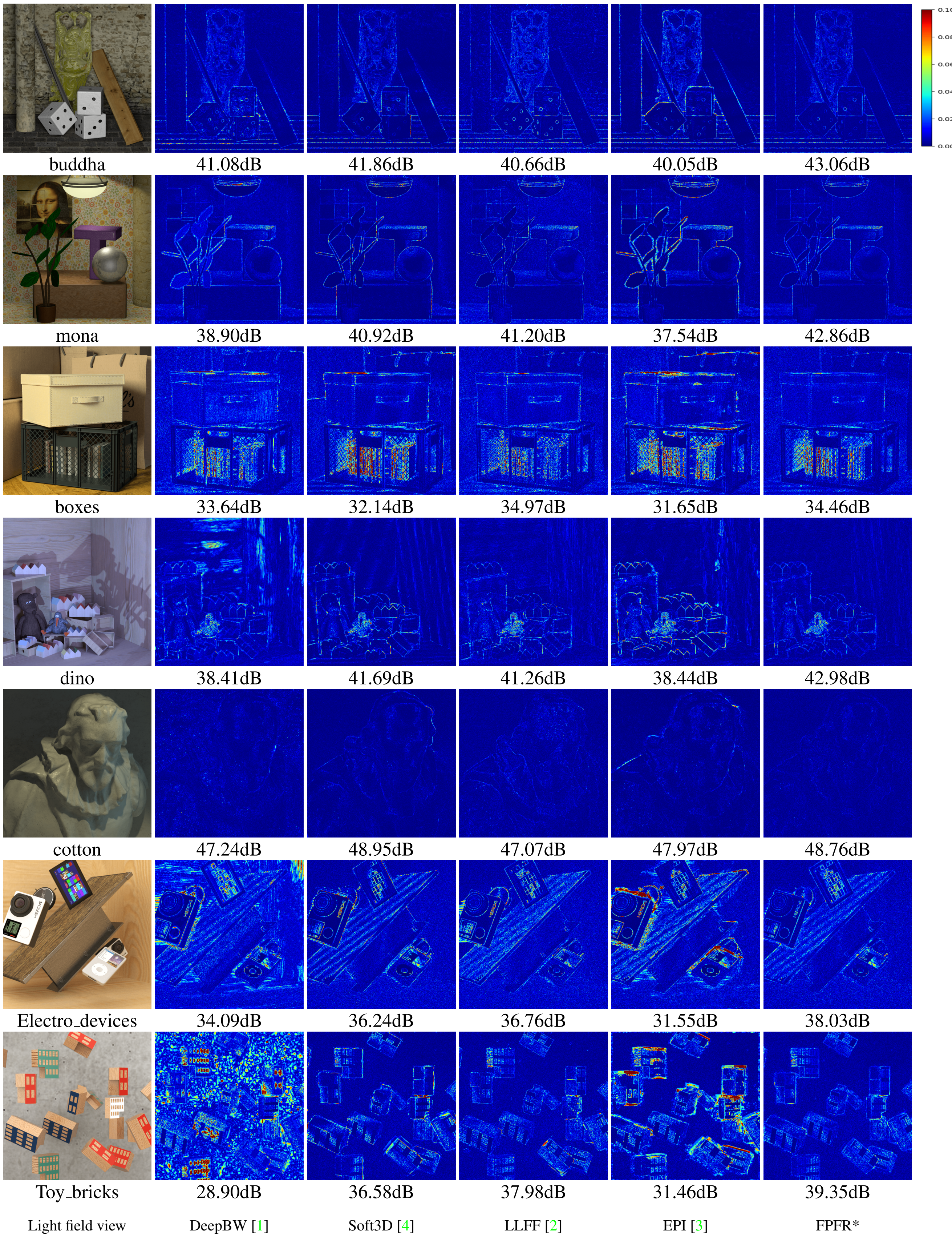

Figure 1. Visual comparison of reconstruction error maps for synthetic light fields.

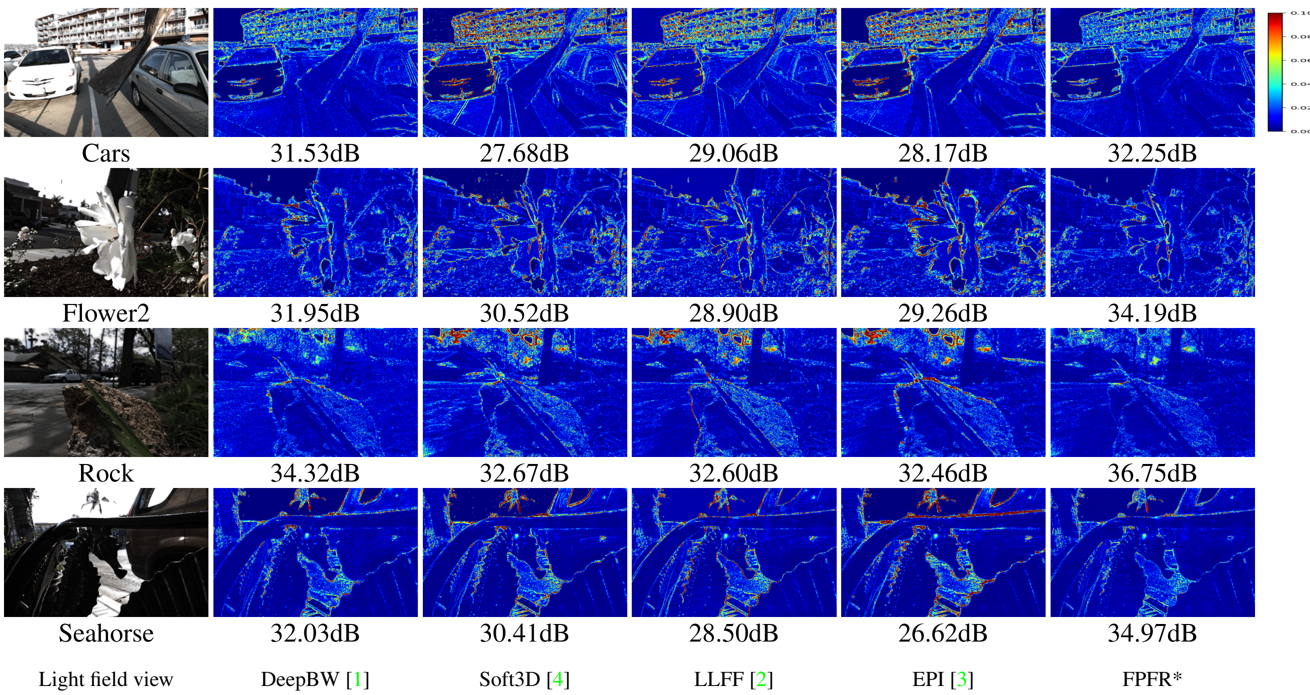

Figure 2. Visual comparison of reconstruction error maps for real light fields.

References

[1] Nima Khademi Kalantari, Ting-Chun Wang, and Ravi Ramamoorthi. Learning-based view synthesis for light field cameras. *ACM Trans. on Graphics*, 35(6):193:1–193:10, 2016. 1, 3, 4

[2] Ben Mildenhall, Pratul P. Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Trans. on Graphics*, 2019. 1, 3, 4

[3] Gaochang Wu, Yebin Liu, Qionghai Dai, and Tianyou Chai. Learning sheared epi structure for light field reconstruction. *IEEE Trans. Image Process.*, 28(7):3261–3273, 2019. 1, 3, 4

[4] Eric Penner and Li Zhang. Soft 3D reconstruction for view synthesis. *ACM Trans. on Graphics*, 36(6):235:1–235:11, 2017. 3, 4