A Learning Based Depth Estimation Framework for 4D Densely and Sparsely Sampled Light Fields

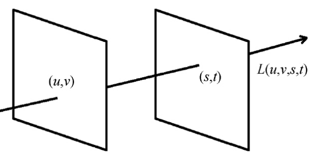
Xiaoran Jiang, Jinglei Shi and Christine Guillemot

INRIA Rennes Bretagne-Atlantique, Rennes, France

DEPTH ESTIMATION EXPLOITING LIGHT FIELDS

Light fields:

- 4D: Intersection with 2 planes **LF(u, v, s, t)**
- 3D scene geometry estimation and reconstruction



Three categories of methods:

- **Based on SAI (Sub-Aperture Images):** Patch-based block matching
- **Based on EPI (Epipolar Plane Images):** Slope is proportional to the disparity value
- **Based on refocused images**

Drawbacks and Challenges

- Most of these methods are designed for dense view sampling.
- Very few methods have been proposed for sparse light fields, including deep learning approaches.
- EPI-based methods only suitable for dense light field.
- Needs of prior knowledge on the disparity range.

DATASETS

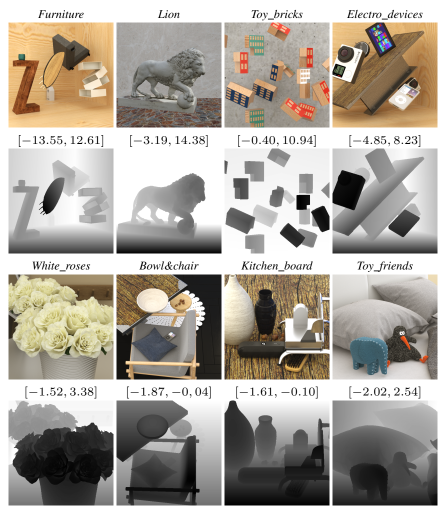
The effectiveness of data-driven algorithms significantly depends on the quality and the quantity of training data.

Existing synthetic datasets

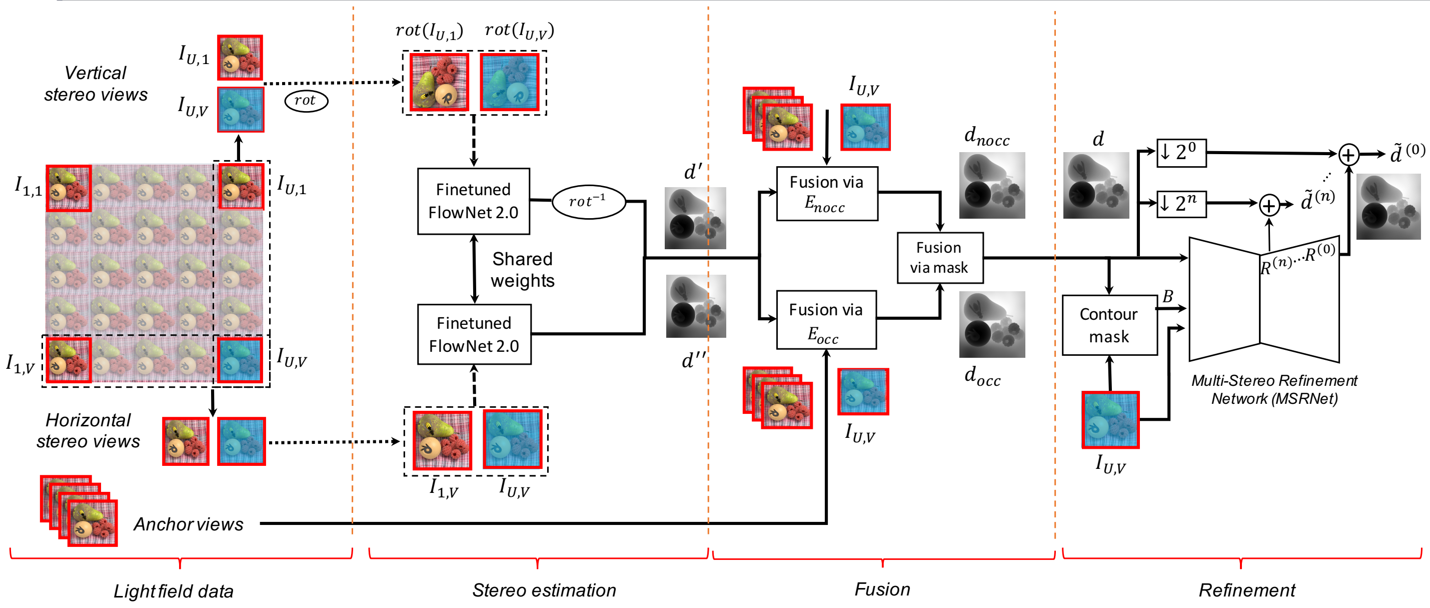
- MIT Light Field Archive (without ground truth disparity/depth values)
- HCI Light Field Dataset (dense light fields only)

Our proposed datasets

- **Sparse Light Field Dataset (SLFD): 53 scenes with disparity range [−20,20]**
- **Dense Light Field Dataset (DLFD): 43 scenes with disparity range [−4,4]**



A SUPERVISED LEARNING BASED DISPARITY ESTIMATION SCHEME



Purpose of the scheme

- Handle **either a dense or a sparse light field** for disparity/depth estimation **based on 4 corner views**.
- **Generate one disparity map for each light field view.**

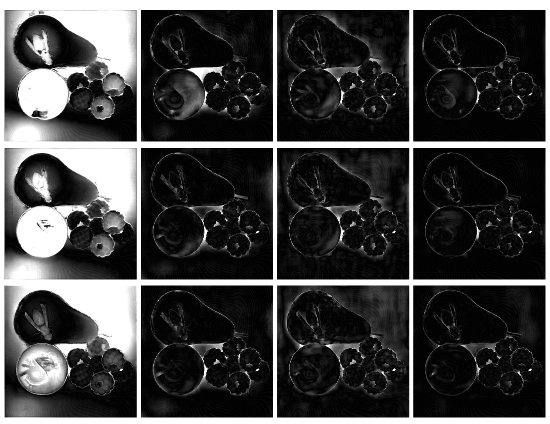
Method

- Deep learning based cascaded framework.
- A pre-trained FlowNet 2.0 is fine-tuned by pairs of stereo images, and the obtained model is used to estimate disparity between pairs of anchor views, arranged horizontally or vertically.
- These coarse estimates are then fused at each anchor viewpoint by exploiting the warping error from other anchor viewpoints.
- Multi-scale residual learning for the refinement of the fused disparity map.

$$\mathcal{L}^{(n)} = \lambda_1 \mathcal{N}(\tilde{d}^{(n)}, d_{GT}^{(n)}) + \lambda_2 \mathcal{G}(\tilde{d}^{(n)}, d_{GT}^{(n)})$$

- The propagation of disparity from anchor viewpoints towards other viewpoints is performed by an occlusion-aware soft 3D reconstruction method [5].

Fine-tuned FlowNet 2.0 for disparity estimation



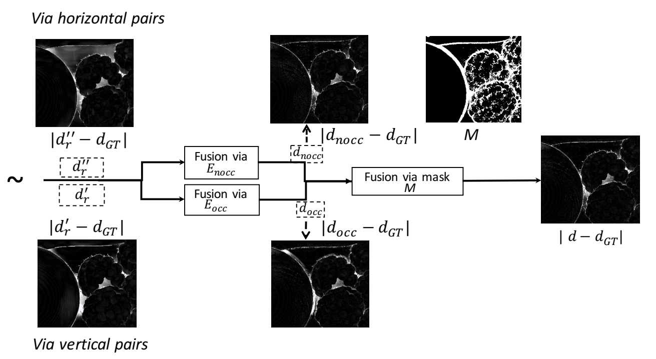
(a) FN2 (b) FN2-ft (c) DispNet-ft (d) FN2-ft-stereo

FN2: the pre-trained FlowNet 2.0 model

FN2-ft: finetuned with no constraint on view positions

DispNet-ft and FN2-ft-stereo: finetuned with stereo light field views only

Fusion based on warping error maps



$E_{\mathrm{nocc}}(d'_r) = \mathrm{mean}_i\Big(\mathcal{E}(I_r, I_r^i(d'_r))\Big)$ for occlusion-free areas

$E_{\mathrm{occ}}(d'_r) = \min_i\Big(\mathcal{E}(I_r, I_r^i(d'_r))\Big)$ for non-overlapped occlusion zones

$M(p) = \begin{cases} 1, & \text{if } \min\Big(E_{\mathrm{nocc}}(d'_r, p), E_{\mathrm{nocc}}(d''_r, p)\Big) > \theta \\ 0, & \text{otherwise} \end{cases}$ approximated binary occlusion mask

Implementation details

- **Data augmentation**
  - Geometrical transformations (e.g. rotation, translation or scaling) that involve data interpolation bring extra errors in the ground truth disparity values.
  - Only chromatic transformation has been applied by changing the hue, saturation, contrast and brightness of training images.
- **Learning details**
  - Initial learning rate set to 0.0001 for the first 500 epochs, then decreased by half every 200 epochs.
  - 2 days of training with a 15G GPU NVIDIA P-100.

EXPERIMENTAL RESULTS

Table 1. Quality evaluation of the estimated disparity maps on center view for dense light fields

Light fields	MSE*100					BadPix(0.01)					BadPix(0.03)					Q25				
	[1]	[2]	[3]	[4]	Ours	[1]	[2]	[3]	[4]	Ours	[1]	[2]	[3]	[4]	Ours	[1]	[2]	[3]	[4]	Ours
StillLife	2.02	1.72	2.56	1.16	**1.14**	81.2	76.2	**71.3**	74.4	71.5	51.0	32.1	25.0	37.1	**24.5**	1.36	1.02	0.87	**0.86**	0.88
Buddha	1.13	0.97	0.82	**0.40**	0.46	57.7	41.2	34.9	51.3	**25.8**	24.4	14.8	12.3	13.4	**6.6**	0.51	0.34	0.31	0.52	**0.28**
MonasRoom	0.76	0.58	0.53	0.56	**0.38**	46.0	42.5	38.6	45.5	**25.2**	22.1	17.8	18.6	17.8	**11.4**	0.38	0.34	0.33	0.35	**0.24**
Butterfly	4.79	0.74	1.84	0.70	**0.54**	82.5	78.9	70.8	82.4	**62.9**	49.1	48.5	36.0	50.8	**28.7**	1.47	1.22	0.85	1.28	**0.66**
Boxes	14.15	**8.23**	12.71	10.05	12.48	72.7	62.3	65.8	83.6	**60.5**	45.5	**28.1**	37.7	57.1	32.8	0.89	0.62	0.68	1.54	**0.55**
Cotton	9.98	1.44	1.18	1.23	**0.67**	60.5	41.7	42.6	72.1	**29.6**	23.3	11.1	10.7	33.7	**8.0**	0.59	0.36	0.42	0.89	**0.25**
Dino	1.23	**0.29**	0.88	0.53	0.50	76.6	57.5	49.1	80.9	**35.9**	48.4	17.9	20.0	48.0	**12.6**	1.08	0.55	1.32	0.42	**0.29**
Sideboard	4.16	**0.92**	10.31	1.31	1.60	67.8	64.3	61.7	79.8	**48.8**	39.3	31.0	37.5	46.4	**23.2**	0.74	0.66	1.26	0.51	**0.37**
Average	4.78	**1.86**	3.85	1.99	2.22	68.1	58.1	54.4	71.2	**45.0**	37.9	25.2	24.7	38.0	**18.5**	0.88	0.64	0.62	0.80	**0.44**

Table 2. Quality evaluation on center view for sparse light fields

Light fields	MSE			BadPix(0.1)			Q25		
	[3]	[4]	Ours	[3]	[4]	Ours	[3]	[4]	Ours
Furniture	1.94	**0.38**	0.78	41.3	61.3	**22.0**	2.52	6.17	**1.10**
Lion	0.87	**0.08**	0.15	59.5	21.4	**8.0**	4.47	2.51	**0.61**
Toy_bricks	1.10	**0.18**	0.44	44.6	36.0	**16.6**	3.61	2.72	**0.94**
Electro_devices	0.63	**0.18**	0.23	43.4	55.5	**24.5**	2.71	4.93	**1.35**
Average	1.14	**0.21**	0.40	47.2	43.6	**17.8**	3.33	4.08	**1.00**

Our algorithm can be also **naturally integrated into a light field view synthesis pipeline**, since it is able to infer disparity information for a view that the color information is unknown.

[1] 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 Int. Conf. on Computer Vision and Pattern Recognition, 2015.

[2] S. Zhang, H. Sheng, C. Li, J. Zhang, and Z. Xiong, "Robust depth estimation for light field via spinning parallelogram operator," J. Computer Vision and Image Understanding, 2016.

[3] X. Jiang, M. Le Pendu, and C. Guillemot, "Depth estimation with occlusion handling from a sparse set of light field views," in IEEE Int. Conf. on Image Processing, 2018.

[4] C.T. Huang, "Empirical bayesian light-field stereo matching by robust pseudo random field modeling," IEEE Trans. on Pattern Analysis and Machine Intelligence, 2018.

[5] E. Penner and L. Zhang, "Soft 3d reconstruction for view synthesis," *ACM Trans. on Graphics (Proc. SIGGRAPH Asia)*, 2017.

ERC advanced grant CLIM