EPI-NEIGHBORHOOD DISTRIBUTION BASED LIGHT FIELD DEPTH ESTIMATION

Junke Li, Xin Jin, Senior Member, IEEE

Shenzhen Key Lab of Broadband Network and Multimedia,
Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China

ABSTRACT

In this paper, a novel depth estimation algorithm tackling foreground occlusion is proposed based on the neighborhood distribution in the sheared epipolar images (EPIs). First, the EPI is sheared to perform refocusing. Next a series of sheared EPI's neighboring pixels in a local window are selected and the corresponding histogram distributions are analyzed by the proposed novel tensor, Kullback-Leibler Divergence (KLD). Then, depths calculated from vertical and horizontal EPIs' tensors are fused according to the tensors' variation scale for a high quality depth map. Finally, confident depth points are propagated to the whole image by global optimization. Experimental results show that the proposed algorithm achieves better performance relative to state-of-the-art algorithms.

Index Terms—Depth estimation, light field, EPI neighborhood distribution.

1. INTRODUCTION

Light field image [1] records spatial and angular information of light for a three-dimensional scene, with which one can shift view points and refocus images after acquisition. Moreover, depth information can be extracted by using this fruitful information.

The conventional light field depth estimation algorithms can be mainly classified into four categories: depth from epipolar images (EPIs) [2-7]; depth from defocus [8-9]; depth from stereo correspondence [10-13]; depth from combined defocus and stereo correspondence [14-17], among which approaches tackling the inaccuracy caused by occlusions were proposed. S. Zhang *et al.* [6] integrated a spinning parallelogram operator (SPO) to divide the EPI into different regions to handle occlusions. C. Chen *et al.* [7] proposed a bilateral metric on angular patches to measure the probability of occlusions by their similarity to the central pixel. T.C. Wang *et al.* [15] proposed an occlusion-aware depth estimation method using edge detection on central sub-aperture image. J. Chen *et al* [18] detected partially occluded boundary regions (POBR) via super-pixel based regularization to handle occlusion boundaries. H. Zhu *et al.* [16] deduced a multi-occluder model in light field according to the pixel distribution similarity between the spatial and angular spaces. However, they treated occlusion explicitly using the pixel intensity changes in sub-aperture images, which might mistake textures as occlusion boundaries, thus resulting in texture copy.

In this paper, a robust depth estimation method, which is based on the EPI-neighborhood distribution is proposed. First the EPI is sheared to perform refocusing. Once refocused at the right depth, the 2D EPI appears as a perpendicular line since there is no disparity. Then, two-side neighborhood intensity distributions along the EPI textures are analyzed and the two-side distribution difference is modeled as the depth tensor by Kullback-Leibler Divergence (KLD). Confidence is measured to assign different weights to depth tensors from horizontal and vertical EPI slices. Finally, the confident depth is propagated to the whole depth map using a global optimization model. Compared with the state-of-the-art approaches, the proposed method can generate depth maps with richer details and sharper object boundaries, which has been demonstrated by testing on the standard dataset.

The rest of the paper is organized as follows. Section 2 describes the proposed algorithm in detail. Experimental results are shown in Section 3. The conclusions are drawn in Section 4.

2. THE PROPOSED ALGORITHM

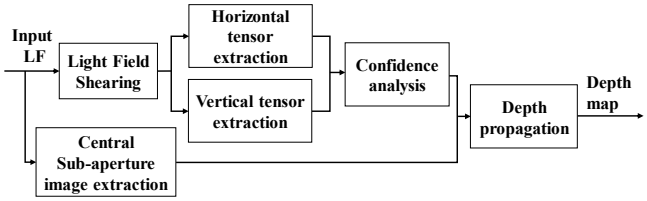


Fig. 1. The framework of the proposed algorithm

The framework of the proposed depth estimation algorithm is shown in Fig. 1. First, *Light field shearing* is to extract sheared EPIs at different depths [1]:

$$L_\alpha(x,y,u,v) = L_F(x+u(1-\frac{1}{\alpha}), y+v(1-\frac{1}{\alpha}), u, v), \quad (1)$$

where L_F is the input light field image; L_α is the refocused light field image; and α is the shear value, which represents the relative depth label. (x, y) and (u, v) represent the spatial coordinates and angular coordinates, respectively. A light field image volume is obtained in this way and each points' corresponding EPI stack can be extracted. For easier conceptual understanding, we use the 2D EPI for illustration.

We fix a horizontal pixel row of constant spatial coordinate y^* and constant angular coordinate v^*, and restrict the light field $L(x, y^*, u, v^*)$ to an $(x\text{-}u)$ slice $E_{y^*, v^*}(x, u)$. The resulting 2D image $E_{y^*, v^*}(x, u)$ is called an $(x\text{-}u)$ EPI, where x and u represent horizontal and vertical axis in EPI, respectively. Then, each sheared EPI with different α value makes up an EPI stack $E_{y^*, v^*, \alpha}(x, u)$. In $E_{y^*, v^*, \alpha}(x, u)$, two-side neighborhood intensity distributions along the EPI textures are analyzed and depth tensor KLD modeling two-side distribution difference is extracted in *tensor extraction*. KLD profile from horizontal and vertical EPI stacks are measured for reliability estimation, which determines weight for each pixel in *confidence analysis*. The definition of the depth tensors, confidences and the related analysis will be described in detail in the following. Meanwhile, *Central sub-aperture image extraction* is applied by picking out the central pixel in each angular patch for calculating the smoothness constraints in the following processing [15]. To ensure each pixel has a reasonable depth value, a *depth propagation* model considering central sub-aperture gradient, smoothness, and second derivative constraints is used to propagate confident depth to the whole depth map.

2.1 Tensor extraction

How to estimate the depth for an object point from its $(x\text{-}u)$ EPI stack is described in this Subsection. The case of $(y\text{-}v)$ EPI stack is similar. Since a correctly focused object presents zero disparity among the viewpoints, which results in no change in angular domain, its corresponding EPI line is perpendicular to the spatial axis. Therefore, α to the perpendicular EPI line is the correct depth label.

Fig. 2 shows how to extract depth tensors and assign a correct depth label α for each pixel. p is an object point to be estimated with its $(x\text{-}u)$ EPI shown at top of Fig. 2 (a). And its sheared EPIs of different α are shown at top of Fig. 2 (c-e) with the corresponding neighborhood distribution histograms shown at the bottom. The neighborhood pixels are in a local rectangular window centered at p, sized $h \times w$. The window is divided into two sides by a hypothetical perpendicular center line (the dotted line in Fig. 2(c-e)) passing through p. Since the correctly focused EPI is also a perpendicular line, the pixel at the left side and the right side of the center line are from different objects and present big differences in intensity, as shown in Fig 2 (d). In this figure, red histogram corresponds to pixel intensity distribution on the left side of the center line and blue histogram corresponds to that right side. It can be seen that the distributions of two sides are quite different. Comparing with the EPIs at the out-of-focused distance, like that shown in the top of Fig. 2 (c) and (e), the EPIs present breaking structures containing oblique lines. Therefore, the two sides divided by the center line might contain same points from different view, thus resulting in similar pixel distributions, shown in bottom of the figures. Therefore, our objective is to find a tensor which is able to describe such EPI-neighborhood difference between the pixels on the two sides.

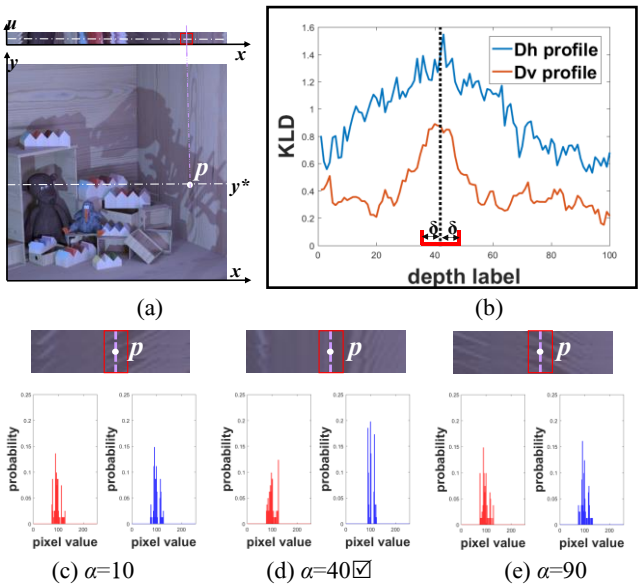


Fig. 2. Tensor for general cases. (a) Central sub-aperture image with an $(x\text{-}u)$ EPI of object point p. (b) The KLD profile for p, with ground truth α=40. (c-e) Sheard EPI at different α (top), and the neighboring pixels' distribution histogram (bottom).

Since Kullback-Leibler Divergence (KLD) [19] describes the relative difference between two probability distributions, we choose it as our depth tensor. For each sheared α of $(x\text{-}u)$ EPI, the tensor $KLD_{x,u}(h^1_{x,u} \mid h^2_{x,u})$ is defined as:

$$KLD_{x,u}(h^1_{x,u} \mid h^2_{x,u}) = \sum_i h^1_{x,u}(i) \log \frac{h^1_{x,u}(i)}{h^2_{x,u}(i)}, \quad (2)$$

where $h^1_{x,u}(\alpha, i)$ and $h^2_{x,u}(\alpha, i)$ are the probabilities of intensity i in the separated two sides at depth label α. $KLD_{x,u}(h^2_{x,u} \mid h^1_{x,u})$ is computed in the same way. The horizontal depth tensor D_h of pixel $\mathbf{p}$ ($\mathbf{p} = (x, y)$) is formulated as:

$$D_h(\mathbf{p}) = \frac{1}{2}(KLD_{x,u}(h^1_{x,u} \mid h^2_{x,u}) + KLD_{x,u}(h^2_{x,u} \mid h^1_{x,u})). \quad (3)$$

The vertical depth tensor D_v of pixel $\mathbf{p}$ can be calculated by vertical $(y\text{-}v)$ EPI stack by similar processes as that of $(x\text{-}u)$ EPI stack. It is defined by:

$$D_v(\mathbf{p}) = \frac{1}{2}(KLD_{y,v}(h^1_{y,v} \mid h^2_{y,v}) + KLD_{y,v}(h^2_{y,v} \mid h^1_{y,v})). \quad (4)$$

Fig. 2(b) shows D_h and D_v profiles for p, where both of them reach the peak at the right depth label α=40. We can draw the conclusion that $D_h(\mathbf{p})$ and $D_v(\mathbf{p})$ can reflect the difference adequately. The raw depth map from the horizontal or vertical tensor can be obtained by:

$$\alpha_*(\mathbf{p}) = \arg\max_{\alpha} D_*(\mathbf{p}), \quad (5)$$

where * denotes h or v alternatively for horizontal and vertical cases.

The proposed algorithm is robust to occlusions since the tensor is able to capture the occlusion property from the

intensity distribution histograms, as illustrated in Fig. 3. The object point q is grey-colored on the wall in the central sub-aperture image in Fig. 3 (a), while it is occluded by the blue or the red object in some other sub-apertures. Due to the foreground occlusion, q's grey EPI line is intersected by the blue and the red occlusion line in Fig3. (c) and (e), respectively. Even correctly focused, q's perpendicular EPI line is intersected by the oblique blue line in Fig. 3 (d). Although there are a few foreground pixels with the same intensity value in the two sides of q's EPI line, the difference between the histograms remains at a local maximum since the probability of the grey pixels is still higher and that of the blue pixels is much lower. Fig. 3(b) shows D_h profile for q is able to reach its maximum at the correct depth. D_v might not reflect the difference adequately since the occlusions are mostly in vertical direction. Therefore, it is essential to combine the horizontal and vertical tensors to make best use of them.

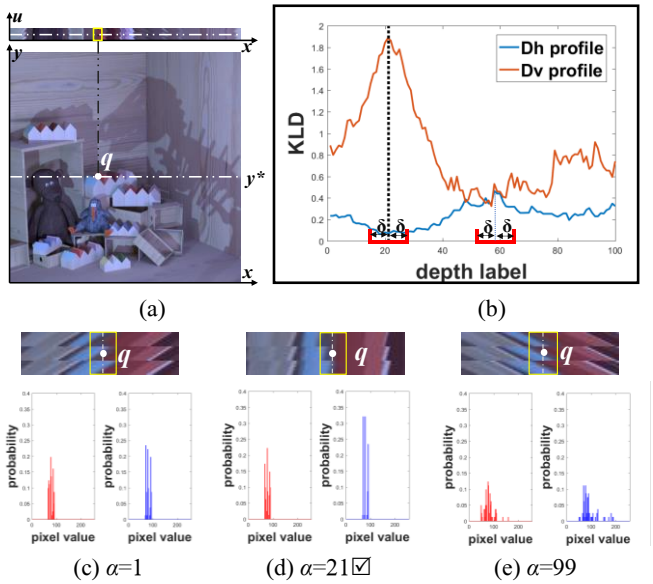


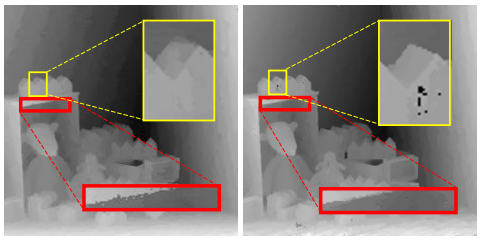
Fig. 3. Tensor for occluded cases. (a) Central sub-aperture image with an (x-u) EPI of object q. (b) The KLD profile for q, with ground truth α=21. (c-e) Sheared EPI at different α (top), and the neighboring pixels' distribution histogram (bottom).

2.2 Confidence analysis and depth propagation

Depth maps from D_h and D_v are shown in Fig 4. (a) and (b), respectively. There is a vertical-dominant structure in yellow rectangle and horizontal-dominant structure in red rectangle. It can be seen that D_h performs better in vertical structures and D_v is robust in horizontal ones. Therefore, confidence analysis gives a measure of how reliable the depth estimate is to take advantage of both horizontal and vertical tensors. Points p and q are taken into consideration, whose tensor profiles are shown in Fig. 2(b) and Fig. 3(b), respectively. Profiles D_h and D_v for p, as well as D_h for q, are all presenting drastic changes around the peak, which indicate the right depth. However, D_v for q is a relatively flat curve with little change around its peak, which indicates a wrong depth. A number of experimental results prove that *KLD* curves with drastic change around its peak will show higher confidence, and vice versa. We investigate variance around the highest point of the tensor profile and define it as the confidence c_* :

$$c_*(\mathbf{p}) = \mathrm{var}(D_{*e}(\mathbf{p}))\big|e \in H(\alpha(\mathbf{p})), H(\alpha(\mathbf{p})) = [\alpha_*(\mathbf{p}) - \delta, \alpha_*(\mathbf{p}) + \delta], \quad (6)$$

where * denotes h or v alternatively; and var(·) is the operation to calculate variance. H(α($\mathbf{p}$)) denotes the neighborhood of the raw depth label along the depth axis. δ is the size of the neighborhood. The confidence measure produces higher values at peaks with drastic changes and lower otherwise as shown in Fig. 3(b).



(a) (b)

Fig. 4. Depth maps from (a)D_h and (b) D_v, respectively.

In order to propagate confident depth values to improve the overall quality of the depth map, we use a multi-label optimization model for propagation. It is given by

$$\alpha_{final} = \arg\min_{\alpha} \sum_{\mathbf{p}} c_h(\alpha(\mathbf{p}) - \alpha_h(\mathbf{p})) + c_v(\alpha(\mathbf{p}) - \alpha_v(\mathbf{p})) + \lambda_1 \cdot (g_\alpha(\mathbf{p}) - g_{I_c}(\mathbf{p})) + \lambda_2 \cdot (\partial^2 \alpha(\mathbf{p}) / \partial \mathbf{p}^2), \quad (7)$$

$\alpha_h(\mathbf{p})$ and $\alpha_v(\mathbf{p})$ denote the raw depths from horizontal and vertical tensors, respectively. $g_\alpha(\mathbf{p})$ and $g_{I_c}(\mathbf{p})$ represent the gradient of depth map and central sub-aperture image at pixel $\mathbf{p}$, respectively. λ_1 and λ_2 control the constraint for edge preserving and smoothness of refined depth map, respectively. Minimizing (7) will give us a final depth map α_{final} with richer details and sharper object boundaries.

3. EXPERIMAENTAL RESULTS AND ANALYSIS

The performance of the proposed method is evaluated by the images of HCI 4D datasets [20].We compare our work with four state-of-the-art methods *LF*[11], *LF_OCC*[15], *EPI*1[5], and *OMG*[16]. Parameters of the algorithms are set to the default. α ranges from 0 to 100, h is 9, w is 5, δ is 2, λ_1 is 2, λ_2 is 0.5. All results are changed into disparity maps for comparison according to [20].

Some of the evaluation results are shown in Fig. 5. *LF* used phase shift theorem for sub-pixel accuracy but it loses depth transitions like the fattening boundary of scene *Cotton*. *LF_OCC* assumes only one occluder and fails at multi-occlusion cases like the grid structure of scene *Boxes*. *EPI*1 is noisy and of high foreground fattening at trivial structures like lamp in *Sideboard*. *OMG* formed occlusion model for small disparity scenes, thus resulting in errors in cases with large disparities like the halos in scene *Dino* and *Boxes*. Since the proposed algorithm analyzes tensors from pixel distribution histogram without explicitly modeling the

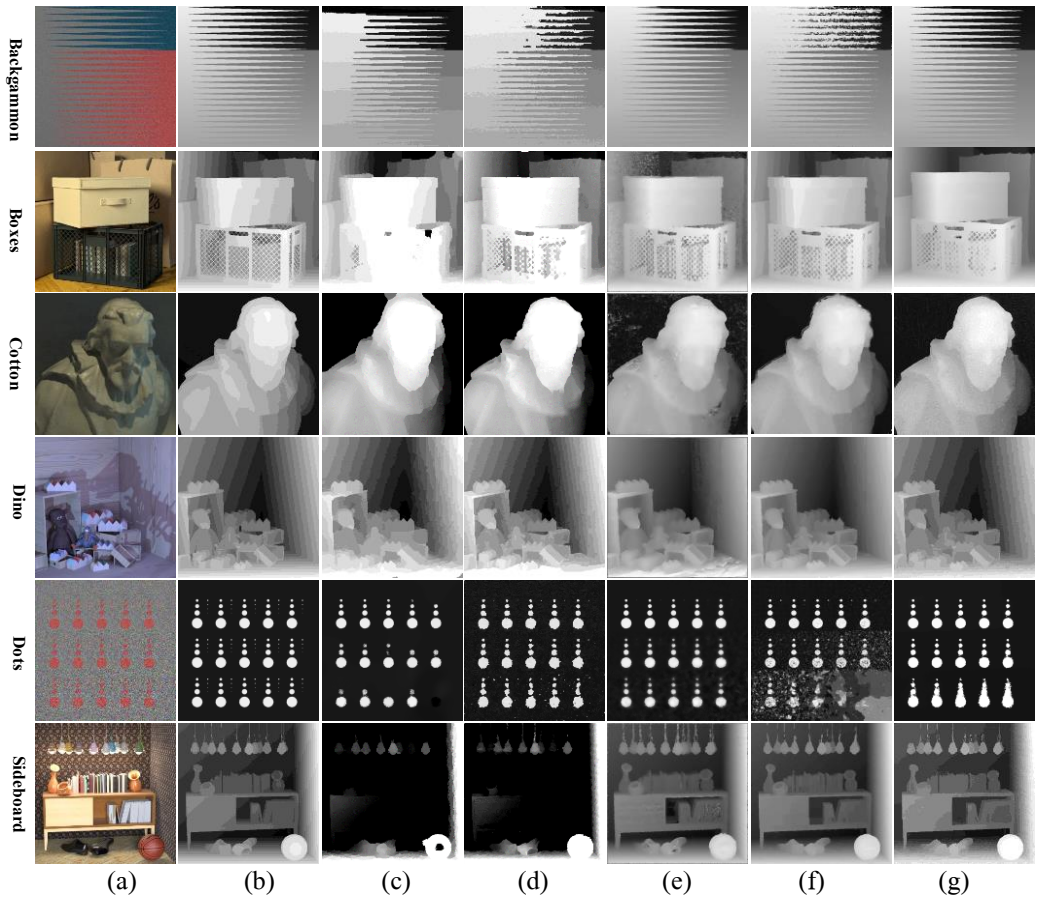


Fig. 5. Visual comparison of the disparity maps from the selected algorithms and proposed algorithm for 6 scenes of HCI datasets. (a) Central sub-aperture image. (b) Ground truth. (c) *LF* [11] (d) *LF_OCC* [15] (e) *EPI*1 [5] (f) *OMG*[16] (g) Proposed.

occlusion, it can work for different kinds of occlusions, and preserve depth transition well of complex structures, thus acquiring high accuracy.

For the quantitative evaluation, we use the BadPixel0.07 (BP) methods, indicating the percentage of erroneous pixels exceed an absolute depth error of more than 0.07 to the ground truth. The comparison is shown in Table 1. Lower value means higher quality. It can be seen that we rank the first for 4 scenes and the average result.

Table 1. BP (%) comparison for all scenes

	LF	LF_OCC	EPI1	OMG	Proposed
Boxes	23.02	26.52	24.45	24.66	**18.31**
Cotton	7.83	6.22	13.93	6.21	**5.53**
Dino	19.03	14.91	10.35	6.65	**4.81**
Sideboard	21.99	18.49	18.38	11.92	**10.03**
Backgammon	**5.52**	19.01	21.33	9.41	5.98
Dots	**2.90**	5.82	62.00	12.74	12.13
Average	13.38	15.16	25.07	11.93	**9.47**

Moreover, BP_OCC0.07 which calculates the accuracy on the occlusion pixels is also used to measure how well the method is able to handle occlusions. Table 2 lists results of the 4 scenes available for BP_OCC comparison, where we outperform others in all 4 scenes. Together with subjective evaluation in Fig.5, the proposed algorithm can provide more reliable estimation results especially at occluded regions.

Table 2. BP_OCC (%) comparison for available scenes

	LF	LF_OCC	EPI1	OMG	Proposed
Boxes	48.15	49.74	52.38	46.66	**42.96**
Cotton	40.33	29.56	40.93	35.75	**21.31**
Dino	41.72	29.71	33.61	26.34	**15.77**
Sideboard	27.06	33.79	42.52	25.04	**23.43**

4. CONCLUSION

This paper proposed a light field depth estimation method, using depth tensor KLD of the two sides of sheared EPIs for good depth quality. First, EPI is sheared and KLD of the two sides is utilized to find the optimal refocused value. Then, the confidences of results from horizontal and vertical EPIs are calculated to take advantage of both. Finally, the raw depth map is propagated by a global optimization model. Experimental results verify that the proposed method provides better results in depth discontinued area, which is beneficial to wider applications.

5. ACKNOWLEDGEMENTS

This work was supported in part by NSFC (61771275), Shenzhen Project (JCYJ20170817162658573) and Guangdong Special Support Plan for the scientific and technological innovation young talents (2016TQ03X998), China.

6. REFERENCES

[1] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Computer Science Technical Report CSTR, 2005, 2(11): 1-11.

[2] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar plane image analysis: An approach to determining structure from motion," Int. J. Comput. Vis., vol. 1, no. 1, pp. 7–55, 1987.

[3] D. Dansereau, and L. Bruton, "Gradient-based depth estimation from 4D light fields," Circuits and Systems, 2004. ISCAS'04. Proceedings of the 2004 International Symposium on. IEEE, 2004, 3: III-549.

[4] S. Wanner, and B. Goldluecke. "Globally consistent depth labeling of 4D light fields," Computer Vision and Pattern Recognition (CVPR), Conference on. IEEE, 2012:41-48.

[5] O. Johannsen, A. Sulc, B. Goldluecke. "What sparse light field coding reveals about scene structure," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016:3262-3270.

[6] S. Zhang, H. Sheng, C. Li, J . Zhang, and Z. Xiong, "Robust depth estimation for light field via spinning parallelogram operator," Computer Vision and Image Understanding, 2016, 145: 148-159.

[7] C Shin, H G Jeon, Y Yoon, et al. "Epinet: A fully-convolutional neural network using epipolar geometry for depth from light field images," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 4748-4757.

[8] H. Lin, C. Chen, S. Bing Kang, et al. "Depth recovery from light field using focal stack symmetry," Proceedings of the IEEE International Conference on Computer Vision. 2015: 3451-3459.

[9] J.Y Lee, R.H. Park. "Depth estimation from light field by accumulating binary maps based on foreground–background separation," IEEE Journal of Selected Topics in Signal Processing, 2017, 11(7): 955-964.

[10] J. Surh, H.G. Jeon, Y. Park, et al. "Noise robust depth from focus using a ring difference filter," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 6328-6337.

[11] 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," Proceedings of the IEEE conference on computer vision and pattern recognition, 2015: 1547-1555.

[12] H. G. Jeon, J. Park, G. Choe, et al. "Depth from a Light Field Image with Learning-based Matching Costs," IEEE transactions on pattern analysis and machine intelligence, 2019, 41(2): 297-310.

[13] C. Chen, H. Lin, Z. Yu, S. B. Kang, and J. Yu, "Light field stereo matching using bilateral statistics of surface cameras," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 1518-1525.

[14] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," Proceedings of the IEEE International Conference on Computer Vision. 2013: 673-680..

[15] T. C. Wang, A. A. Efros, and R. Ramamoorthi, "Occlusion-aware depth estimation using light-field cameras," Proceedings of the IEEE International Conference on Computer Vision. 2015: 3487-3495.

[16] H. Zhu, Q. Wang, J. Yu. "Occlusion-Model Guided Anti-Occlusion Depth Estimation in Light Field," IEEE Journal of Selected Topics in Signal Processing, 2017:1-1.

[17] W. Williem, I. K. Park, and K. M. Lee. "Robust light field depth estimation using occlusion-noise aware data costs," IEEE transactions on pattern analysis and machine intelligence, 2018, 40(10): 2484-2497.

[18] J Chen, J Hou, Y Ni, et al. "Accurate Light Field Depth Estimation with Superpixel Regularization over Partially Occluded Regions," IEEE Transactions on Image Processing, 2017, 27(10):4889.

[19] S. Kullback, and R.A. Leibler, "On information and sufficiency," The annals of mathematical statistics, 1951, 22(1), pp.79-86.

[20] K. Honauer, O. Johannsen, D. Kondermann, et al. "A dataset and evaluation methodology for depth estimation on 4d light fields," Asian Conference on Computer Vision. Springer, Cham, 2016: 19-34.

[21] http://hci-lightfield.iwr.uni-heidelberg.de/