LIGHT-FIELD FLOW: A SUBPIXEL-ACCURACY DEPTH FLOW ESTIMATION WITH GEOMETRIC OCCLUSION MODEL FROM A SINGLE LIGHT-FIELD IMAGE

Wenhui Zhou[⋆] *Pengfei Li*[⋆] *Andrew Lumsdaine*[†] *Lili Lin*[‡]

[⋆]School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China
[†]Pacific Northwest Laboratory, Richland, WA, U.S
[‡]School of Information and Electronic Engineering, Zhejiang Gongshang University, Hangzhou, China

ABSTRACT

Light-field cameras capture not only 2D images, but also the angles of the incoming light. These additional light angles bring the benefit of getting a sub-aperture image array from a single light-field image. Inspired by the traditional optical flow with occlusion detection, this paper focuses on the correlation analysis and the occlusion modeling for the sub-aperture array, and unifies them into a light-field flow framework. The main challenges faced are subpixel displacements and occlusion handling among the sub-aperture images. We build a light-field flow for joint depth estimation and occlusion detection, and develop a geometric occlusion model. More specifically, we firstly estimate subpixel-accuracy optical flows from each two sub-aperture images by the phase shift theorem, then a forward-backward consistency checking is adopted to detect the occluded regions. According to the geometric complementary character of occlusion in a light-field image, an occlusion filling strategy is proposed to refine depth estimation in the occluded regions. Experimental results on the synthetic scenes and Lytro Illum camera data both demonstrate the effectiveness and robustness of our method which has excellent performance in handling occlusions.

Index Terms— light-field, depth estimation, light-field flow, subpixel accuracy, geometric occlusion model

1. INTRODUCTION

Light-field cameras, known as plenoptic cameras, capture 4-dimensional spatio-angular information of the light field. They provide more helpful multiple viewpoints for visual analysis and understanding than conventional cameras [1, 2].

Scene structure (depth) recovery from light-field camera has been an important issue in recent years, and many methods have been proposed to address it [3–9]. Most of these methods are based on the sub-aperture images and data consistency measure, such as correspondence, defocus, shading, etc. Limited by the fundamental tradeoff between spatial and angular resolution [10], however, these methods still face the challenges of robustness and accuracy because of the subpixel displacements [5] and occlusions [6, 7].

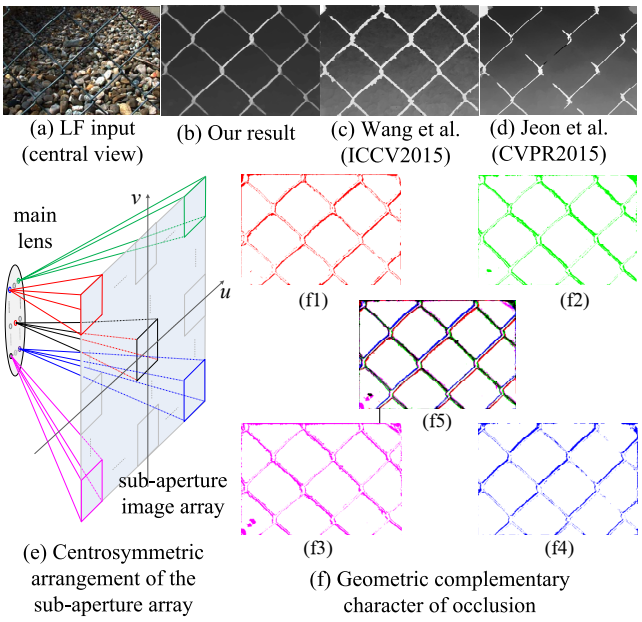

Fig. 1. Visual comparison of occlusion handling of different algorithms and the diagram of geometric complementary character of occlusion. (f1)-f(4) are the results of occlusion detected from different sub-aperture images, and we put them together in (f5) to show more clearly.

In this paper, we focus on the geometric complementary relationship among the sub-aperture images for occlusion handling. To the best of our knowledge, very little work has explicitly considered this relationship before. In fact, the representation of 4D light-field data [11, 12] defines a centrosymmetric arrangement of the sub-aperture array. We show that this arrangement implies a geometric complementary character of occlusion, as shown in Fig. 1 and Fig. 2. Fig. 1(f1) to (f4) show the occluded regions detected from the sub-aperture images on the upper-left, upper-right, bottom-left and bottom-right of the sub-aperture image array, respectively. As seen in Fig. 1(f5), it is obvious that the occluded regions from different sub-aperture are complementary to

each other.

Our main contributions are:

1) A light-field flow framework for joint depth estimation and occlusion handling, which is based on a subpixel-accuracy optical flow with occlusion detection.

2) A geometric occlusion model for light-field camera which improves the accuracy of depth estimation in occluded regions.

2. RELATED WORKS

Development of light-field cameras can be dated back to more than a century ago. This topic recently regained much attention when the light fields theories [13, 14] and the portable prototype [1] were developed. Light-field cameras enable various new applications beyond the reach of conventional 2D cameras, such as post-capture refocusing, focal stack extraction, depth-of-field extension, depth recovery, etc.

Along with the deep-going of the light-field researches and applications, accurate depth estimation has been an essential and challenging problem. Georgiev et al. [15] estimated disparity maps by computing a normalized cross correlation between microlens images. Yu et al. [3] analyzed the 3D geometry of lines in a light field image and computed the disparity maps through line matching between the sub-aperture images. Jeon et al. [5] estimated the multi-view stereo correspondences with sub-pixel accuracy using phase shift theorem. Tao et al. [4] estimated depth by combining multiple depth cues, such as defocus, shading and correspondence. On the basis of Tao's work, Wang et al. [6, 7] developed a light-field occlusion model based on the physical image formation, and then Williem et al. [8] proposed two novel data costs for noisy scene with occlusion. Their light-field occlusion models, however, sometimes fail to distinguish between depth discontinuities and surface color/texture discontinuities.

Many works in optical flow have also addressed the occlusion problem. Ayvaci et al. [16] formulated occlusion detection and optical flow estimation as a joint optimization problem, and presented two efficient numerical schemes to solve it. Revaud et al. [17] presented a sparse-to-dense edge-preserving interpolation of correspondences (EpicFlow) for filling occlusions. However, little work applies optical flow to light-field image. Zhou et al. [18] developed an edge-aware light-field flow. But this method depends on Ayvaci's work [16] to estimate optical flows between two sub-aperture images, so its performance is limited by that of Ayvaci's method.

3. LIGHT-FIELD FLOW WITH OCCLUSION

We build our light-field flow theory on a single light-field image. According to the light-field imaging mechanism, a light-field image includes a set of narrow baseline multi-view (sub-apertures) images in a centrosymmetric arrangement.

Let $\mathcal{I}_0(\mathbf{x}, \mathbf{u}_0)$ and $\mathcal{I}(\mathbf{x}', \mathbf{u})$ be the central sub-aperture image and another one, respectively. $\mathbf{x}$ and $\mathbf{x}'$ are spatial coordinates, and $\mathbf{u}_0$ and $\mathbf{u}$ are angular coordinates. According to the refocusing equation of light-field image [1], the correspondence between $\mathcal{I}_0(\mathbf{x}, \mathbf{u}_0)$ and $\mathcal{I}(\mathbf{x}', \mathbf{u})$ is given by

$$\mathcal{I}_0(\mathbf{x}, \mathbf{u}_0) = \mathcal{I}(\mathbf{x} + \beta(\mathbf{x}) \cdot \Delta\mathbf{u}, \mathbf{u}) \tag{1}$$

where $\Delta\mathbf{u} = \mathbf{u} - \mathbf{u}_0$, $\beta(\mathbf{x}) = 1 - \frac{1}{\alpha(\mathbf{x})}$, and α represents a virtual refocusing plane defined in reference [1].

The optical flow between $\mathcal{I}_0(\mathbf{x}, \mathbf{u}_0)$ and $\mathcal{I}(\mathbf{x}', \mathbf{u})$ denotes their incremental displacement, i.e.

$$\mathbf{v}(\mathbf{x}, \mathbf{u}) = \mathbf{x}' - \mathbf{x} = \beta(\mathbf{x}) \cdot \Delta\mathbf{u} \tag{2}$$

We can eliminate the correlation between optical flow and angular coordinates of sub-aperture images by dividing $\Delta\mathbf{u}$ on both sides. This means all optical flows between any two sub-apertures have the same flow model (depth flow) which is independent of the property of sub-aperture images. We define it as the light-field flow model.

Considering the noise and occlusion, Eq. 1 becomes

$$\mathcal{I}_0(\mathbf{x}, \mathbf{u}_0) = \begin{cases} \mathcal{I}(\mathbf{x} + \beta(\mathbf{x}) \cdot \Delta\mathbf{u}, \mathbf{u}) + n(\mathbf{x}, \mathbf{u}) & \mathbf{x} \notin \Omega \\ \rho(\mathbf{x}, \mathbf{u}) & \mathbf{x} \in \Omega \end{cases} \tag{3}$$

where Ω is the occluded region. The additive term $n(\mathbf{x}, \mathbf{u})$ is deviations resulted from illumination changes, quantization error, sensor noise, and linear interpolation error, etc. In the occluded region Ω, the image can take any value $\rho(\mathbf{x}, \mathbf{u})$. we note that $n(\mathbf{x}, \mathbf{u})$ is small but dense, while $\rho(\mathbf{x}, \mathbf{u})$ is usually large but sparse, so we will use these properties as an inference criterion for the light-field flow estimation.

The light-field flow model is given by

$$flow_{LF}(\mathbf{x}, \mathbf{u}) = \begin{cases} \beta(\mathbf{x}) + \delta(\mathbf{x}, \mathbf{u}) & \mathbf{x} \notin \Omega \\ \varepsilon_\rho(\mathbf{x}, \mathbf{u}) & \mathbf{x} \in \Omega \end{cases} \tag{4}$$

where $\delta(\mathbf{x}, \mathbf{u})$ is the deviation term caused by the additive term $n(\mathbf{x}, \mathbf{u})$. $\varepsilon_\rho(\mathbf{x}, \mathbf{u})$ can take any value when $\mathbf{x} \in \Omega$.

It is evident that optical flows exist in any two sub-aperture images, our basic idea is to estimate the truth value of light-field flow $\beta(\mathbf{x})$ from these noisy optical flows.

4. LIGHT-FIELD FLOW FRAMEWORK

According to the light-field flow theory, our light-field flow estimation framework includes the following two steps.

4.1. Subpixel-accuracy optical flow volume

This step independently estimates the forward and backward optical flows between any two sub-aperture images, and then group all the forward optical flows into an optical flow volume. Considering the subpixel displacements, we estimate optical flow using the phase shift theorem [5].

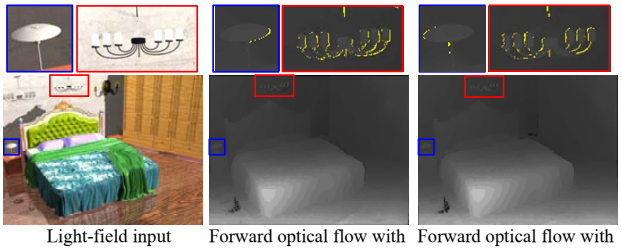

Fig. 2. Geometric complementary character of occlusion on a synthetic data. The sub-aperture images $(u_1 = 1, u_2 = 2)$ and $(u_1 = 9, u_2 = 8)$ are centrosymmetric.

As for $\mathcal{I}_0(\mathbf{x}, \mathbf{u}_0)$ and $\mathcal{I}(\mathbf{x}', \mathbf{u})$, the forward optical flow $\mathbf{v}_f(\mathbf{x}, \mathbf{u})$ is given by

$$\begin{cases} \varepsilon_{\mathcal{I}}(\mathbf{x}, \mathbf{u}, \mathbf{v}) = \mathcal{I}_0(\mathbf{x}, \mathbf{u}_0) - \mathcal{F}^{-1}\left\{\mathcal{F}\left(\mathcal{I}(\mathbf{x}', \mathbf{u})\, e^{2\pi i \cdot \mathbf{v}(\mathbf{x}, \mathbf{u})}\right)\right\} \\ \varepsilon_{\mathcal{G}}(\mathbf{x}, \mathbf{u}, \mathbf{v}) = \mathcal{G}_0(\mathbf{x}, \mathbf{u}_0) - \mathcal{F}^{-1}\left\{\mathcal{F}\left(\mathcal{G}(\mathbf{x}', \mathbf{u})\, e^{2\pi i \cdot \mathbf{v}(\mathbf{x}, \mathbf{u})}\right)\right\} \\ \mathcal{C}_f(\mathbf{x}, \mathbf{u}, \mathbf{v}) = \gamma \left\|\varepsilon_{\mathcal{I}}(\mathbf{x}, \mathbf{u}, \mathbf{v})\right\|_2 + (1-\gamma)\left\|\varepsilon_{\mathcal{G}}(\mathbf{x}, \mathbf{u}, \mathbf{v})\right\|_2 \\ \mathbf{v}_f(\mathbf{x}, \mathbf{u}) = \underset{\mathbf{v}}{\operatorname{argmin}}\left\{\mathcal{C}_f(\mathbf{x}, \mathbf{u}, \mathbf{v})\right\} \end{cases} \tag{5}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the Fourier operator and the inverse Fourier operator. $\mathcal{G}_0(\mathbf{x}, \mathbf{u}_0)$ and $\mathcal{G}(\mathbf{x}', \mathbf{u})$ are the gradients of $\mathcal{I}_0(\mathbf{x}, \mathbf{u}_0)$ and $\mathcal{I}(\mathbf{x}', \mathbf{u})$, respectively. $\mathcal{C}_f(\mathbf{x}, \mathbf{u}, \mathbf{v})$ is the flow cost of assigning the optical flow $\mathbf{v}$ to pixel $\mathbf{x}$.

The backward optical flow can be got by the similar method, then the occluded regions are detected by the forward-backward consistency checking [16, 19].

4.2. Geometric occlusion model based light-field flow

According to the character of $n(\mathbf{x}, \mathbf{u})$,the light-field flow is estimated by finding minimum variance of the flow costs of all sub-apertures. Considering the occlusion, the flow costs of occluded pixels should not participate in the calculation of variance. Inspired by the occlusion filling [20], we propose a geometric occlusion model by using the flow cost of its centrosymmetric sub-aperture image instead, as shown in Eq. 6.

$$\begin{cases} flow_{LF}(\mathbf{x}) = \underset{\beta}{\operatorname{argmin}}\left\{\underset{\mathbf{u}}{var}\left(\mathcal{C}_f^*(\mathbf{x}, \mathbf{u}, \beta)\right)\right\} \\ \mathcal{C}_f^*(\mathbf{x}, \mathbf{u}, \beta) = \begin{cases} \mathcal{C}_f(\mathbf{x}, \mathbf{u}, \mathbf{v}) & (\mathbf{x}, \mathbf{u}) \notin \Omega \\ \mathcal{C}_f(\mathbf{x}, \mathbf{u}^*, \mathbf{v}) & (\mathbf{x}, \mathbf{u}) \in \Omega, (\mathbf{x}, \mathbf{u}^*) \notin \Omega \\ DISCARD & others \end{cases} \end{cases} \tag{6}$$

where var is the variance calculation function. If pixel $\mathbf{x}$ is still occluded in its centrosymmetric sub-aperture image, this pixel will be discarded and don't participate in the calculation.

Our occlusion model is based on the geometric complementary character of occlusion, that is if pixel $\mathbf{x}$ is occluded in sub-aperture image($\mathbf{u}$), it may not be occluded in its centrosymmetric one($\mathbf{u}^*$), as shown in Fig. 1 and Fig. 2.

Finally, we regularize the initial light-field flow with weighted median filter [21] for refined results. It is a kind of edge-aware filter based on guided filters [22]. It enforces a piecewise smooth flow with depth-discontinuity preserving for each β slice (level). The central sub-aperture image is used as the guidance image of guided filter.

5. EXPERIMENTAL COMPARISONS

In our implementation, the refocusing parameter α ranges from 0.2 to 5.0, so the valid range of β is from -4.0 to 0.8, and is quantized into 256 different levels. According to the experience and experiment analysis, the regularization parameter of guided filter is set to 0.0001.

We compare our method to those of Tao et al. [4], Jeon et al. [5], Wang et al. [6, 7] and Zhou et al. [18]. The source codes can be downloaded from the authors' project webpages.

We validate our method on the synthetic scenes and Lytro Illum camera data. We focus on the algorithm effectiveness and robustness in handling discontinuities and occluded regions. Therefore, the scenes with lots of depth, color and texture discontinuities, especially the objects and backgrounds are full of complex texture, are selected as our experimental data. Fig. 3 shows the comparisons on the synthetic dataset of Wang et al. [6, 7]. In Fig. 4, some real scenes with fine structures and occlusions are selected from EPFL Light-Field Image Dataset [23].

Experimental results indicate our depth estimation are more accurate and smooth at depth discontinuities. The precision of previous results at object boundaries are affected by occlusion noise, or are over-smoothed.

6. CONCLUSION AND FUTURE WORKS

This paper develops a light-field flow model for a single light-field image. The main idea is extracting subpixel-accuracy optical flows and unifying them together under a geometric occlusion model. Occlusion detection and regularization steps can smooth the depth estimation in ambiguous regions, such as textureless regions and occlusions. Moreover, optical flow volume extraction is the most time-consuming step because the optical flows estimation from sub-apertures are independent. One of our future work will focus on this issue. Quantitative evaluation will be another future work. Our method just provides visual comparison results because it is based on the plenoptic 1.0 cameras and lacks ground truth.

7. ACKNOWLEDGMENTS

This work is supported in part by the National High-tech R&D Program of China (863 Program, 2015AA015901), Key Program of Zhejiang Provincial Natural Science Foundation of China (No.LZ14F020003), and Zhejiang Provincial Natural Science Foundation of China (No.LY15F010002).

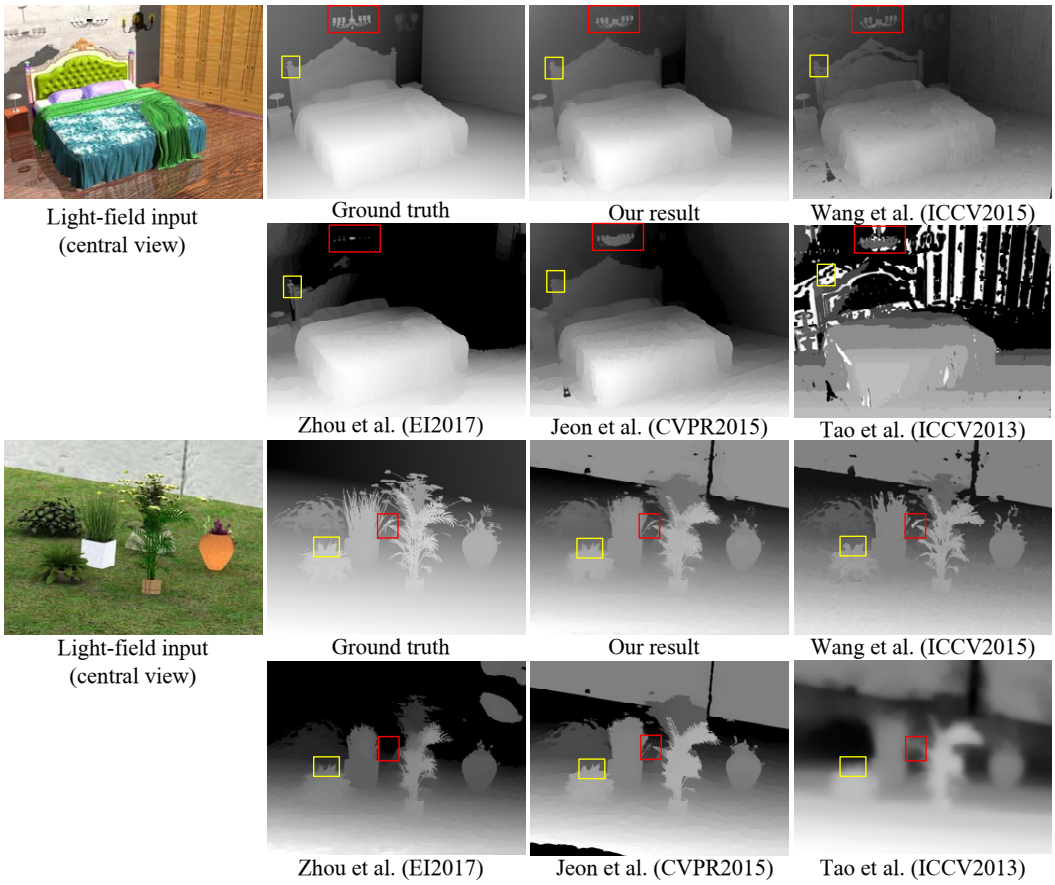

Fig. 3. Depth estimation results on the synthetic dataset of Wang et al. [6, 7].

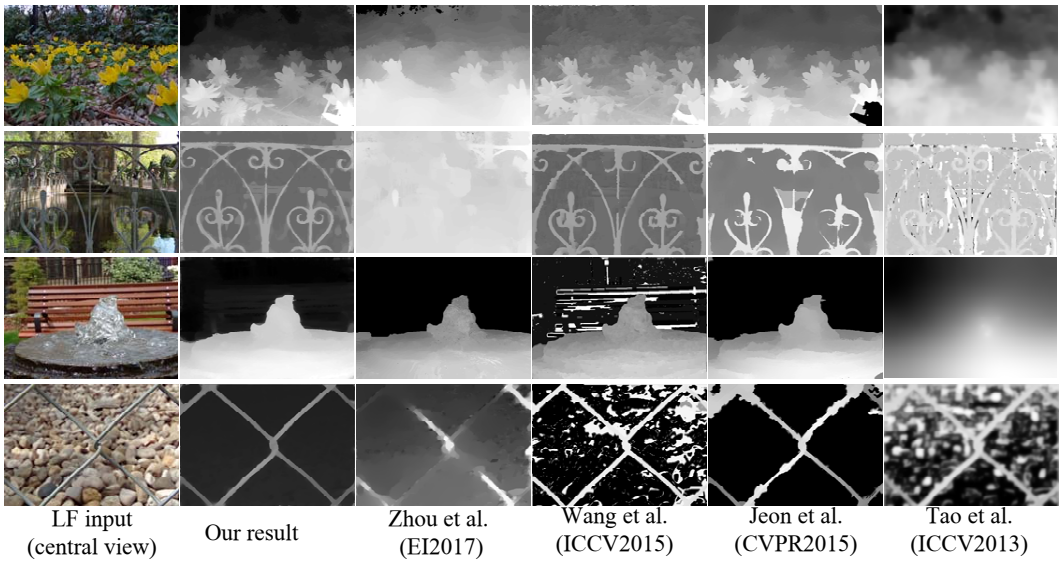

Fig. 4. Depth estimation results on the real scenes data selected from EPFL Light-Field Image Dataset [23].

8. REFERENCES

[1] R. Ng, M., M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a handheld plenoptic camera," *Computer Science Technical Report*, vol. 2, no. 11, 2005.

[2] T. Georgiev, Z. Yu, A. Lumsdaine, and S. Goma, "Lytro camera technology: Theory, algorithms, performance analysis," in *SPIE 8667, Multimedia Content and Mobile Devices*, 2013.

[3] Z. Yu, X. Guo, H. Ling, A. Lumsdaine, and J. Yu, "Line assisted light field triangulation and stereo matching," in *Proceedings of International Conference on Computer Vision*, 2013.

[4] 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.

[5] H. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y. Tai, and I. Kweon, "Accurate depth map estimation from a lenslet light field camera," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2015.

[6] T. Wang, A. Efros, and R. Ramamoorthi, "Occlusion-aware depth estimation using light-field cameras.," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015.

[7] T. Wang, A. Efros, and R. Ramamoorthi, "Depth estimation with occlusion modeling using light-field cameras," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 11, pp. 2170–2181, 2016.

[8] Williem and I. Park, "Robust light field depth estimation for noisy scene with occlusion," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2016.

[9] M. W. Tao, P. P. Srinivasan, S. Hadap, S. Rusinkiewicz, J. Malik, and R. Ramamoorthi, "Shape estimation from shading, defocus, and correspondence using light-field angular coherence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016.

[10] C. Liang and R. Ramamoorthi, "A light transport framework for lenslet light field cameras," *ACM Transactions on Graphics*, vol. 34, no. 2, 2015.

[11] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Jun 2013.

[12] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Linear volumetric focus for light field cameras," *ACM Transactions on Graphics*, vol. 34, no. 2, Feb. 2015.

[13] E. H. Adelson and J. Y. A. Wang, "Single lens stereo with a plenoptic camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 99–106, 1992.

[14] M. Levoy and P. Hanrahan, "Light field rendering," in *International Conference on Computer Graphics and Interactive Techniques*, 1996.

[15] T. Georgiev and A. Lumsdaine, "Reducing plenoptic camera artifacts," *Computer Graphics Forum*, vol. 29, no. 6, pp. 1955–1968, 2010.

[16] A. Ayvaci, M. Raptis, and S. Soatto, "Sparse occlusion detection with optical flow.," *International Journal of Computer Vision*, vol. 9, no. 3, pp. 322–338, 2012.

[17] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, "Epicflow: Edge-preserving interpolation of correspondences for optical flow," in *International Conference on Computer Vision and Pattern Recognition*, 2015.

[18] W. Zhou, A. Lumsdaine, L. Lin, W. Zhang, and R. Wang, "Edge-aware light-field flow for depth estimation and occlusion detection," in *Computational Imaging XV in 2017 IST International Symposium on Electronic Imaging*, 2017.

[19] Q. Chen and V. Koltun, "Full flow: Optical flow estimation by global optimization over regular grid," in *International Conference on Computer Vision and Pattern Recognition*, 2016.

[20] A.S. Ogale, C. Fermller, and Y. Aloimonos, "Motion segmentation using occlusions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 988–992, 2005.

[21] Z. Ma, K. He, Y. Wei, J. Sun, and E. Wu, "Constant time weighted median filtering for stereo matching and beyond," in *International Conference on Computer Vision*, 2013.

[22] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, 2013.

[23] M. Rerabek and T. Ebrahimi, "New light field image dataset," in *8th International Conference on Quality of Multimedia Experience*, 2016.