https://doi.org/10.2352/ISSN.2470-1173.2017.17.COIMG-431

Edge-aware Light-Field Flow for Depth Estimation and Occlusion Detection

Wenhui Zhou; Hangzhou Dianzi University, Hangzhou, China;
Andrew Lumsdaine; Pacific Northwest Laboratory, Richland, WA, U.S;
Lili Lin; Zhejiang Gongshang University, Hangzhou, China, Corresponding Author;
Wei Zhang, Rong Wang; Zhejiang Gongshang University, Hangzhou, China,

Abstract

Light-field cameras capture 4-dimensional spatio-angular information of the light field. They provide more helpful multiple viewpoints or sub-apertures for visual analysis and visual understanding than traditional cameras. Optical flow is a common method to get scene structure cues from two images, however, sub-pixel displacements and occlusions are two inevitable challenges in the optical flow estimation from light-field sub-apertures.

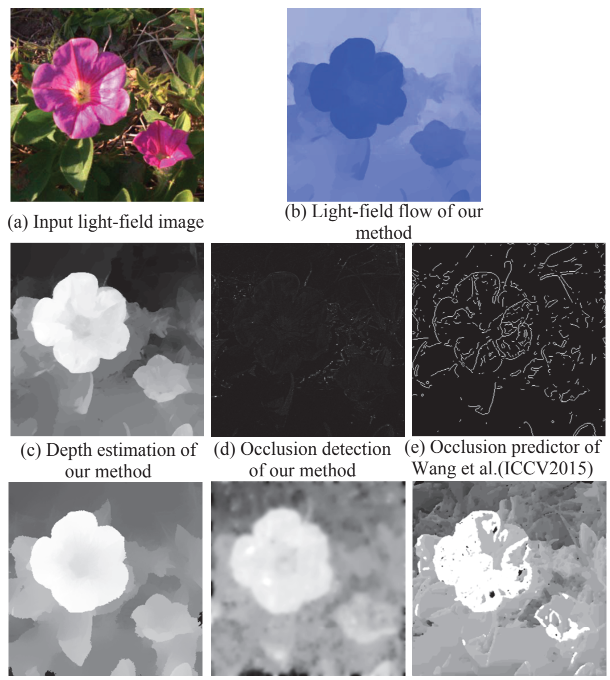
In this paper, we develop a light-field flow model, and propose an edge-aware light-field flow estimation framework for joint depth estimation and occlusion detection. It consists of three steps: i) An optical flow volume with sub-pixel accuracy is extracted from sub-apertures by edge-preserving interpolation. Then occlusion regions are detected through consistency checking. ii) Robust light-field flow and depth estimation are initialized by a winner-take-all strategy and a weighted voting mechanism. iii) Final depth map is refined by a weighted median filter based on guided filter. Experimental results demonstrate the effectiveness and robustness of our method.

Introduction

Light-field cameras capture not only 2D images, but also the angles of the incoming light rays [1, 2]. These additional light angles bring an important benefit of light-field cameras, that is multiple viewpoints or sub-apertures are available from a single light-field image. Scene structure (depth) recovery from light-field camera has been an essential and interesting task. Limited by the fundamental tradeoff between spatial and angular resolution [3], however, it still faces the challenges of robustness and accuracy, especially at regions of depth discontinuities and occlusions.

Many methods have been proposed to address this problem in recent years. Most of them are based on the focal stacks analysis and their data consistency measure [4, 5, 6, 7], such as correspondence and defocus. These methods usually face two limitations: a) their occlusion detection methods, in some cases, are sensitive to the color and texture on object surface. b) the accuracy of their depth estimation theoretically depends on the interval size of focal stacks (the pre-established α values in depth search range).

In this paper, we try to solve these problems by borrowing the idea of optical flow analysis in the traditional computer vision community. However, little work has explicitly considered the optical flow estimation for light-field images. Different from traditional optical flow applications, optical flow estimation from sub-apertures suffers from two inevitable problems: occlusion [6, 7] and sub-pixel displacement [8].



(a) Input light-field image (b) Light-field flow of our method (c) Depth estimation of our method (d) Occlusion detection of our method (e) Occlusion predictor of Wang et al.(ICCV2015) (f) Depth estimation of Jeon et al.(CVPR2015) (g) Depth estimation of Tao et al.(ICCV2013) (h) Depth estimation of Wang et al.(ICCV2015)

Figure 1. *Visual comparison of discontinuities and occlusions handling of different algorithms. The results of Tao et al. [4], Jeon et al. [8] and Wang et al. [6, 7] can be obtained by running the codes from their project webpages. It can be seen that our method has better depth-discontinuity preserving property.*

We introduce an edge-aware light-field flow model which accounts for multi-view motion field (sub-aperture images), light-field geometry (depth), translation (correspondence) and occlusion. It can estimate out a robust light-field flow together with accurate depth maps and occlusions.

Our main contributions are:

(1) We develop a light-field flow model on a single light-field image by fusing all optical flows between each sub-aperture image and the central sub-aperture image. This model is independent of the property of sub-aperture images.

(2) We introduce an edge-aware light-field flow estimation framework for joint depth recovery and occlusion detection.

As shown in Figure 1, we can obtain accurate light-field flow with good depth-discontinuity preserving property. Moreover, our depth estimation and occlusion detection results are better than those of the state-of-the-art methods.

Related Works

Depth from Light-Field Cameras: Depth recovery from a single light-field image is still an essential and challenging problem in the computational photography field. Most of recent works focus on the solutions from different depth cues, such as correspondence, defocus, shading, and occlusion, etc. Georgiev and Lumsdaine [9] estimated disparity maps by computing a normalized cross correlation between microlens images. Perwass and Wietzke [10] introduced a correspondence technique to estimate depth. Yu et al. [11] 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. [8] estimated the multi-view stereo correspondences with sub-pixel accuracy using phase shift theorem. Lin et al. [12] described a technique to recover depth from a light field image based on the focal stack symmetry analysis and data consistency measure. Tao et al. [4, 5] discussed the advantages and disadvantages of different depth cues for depth estimation. They combined shading, correspondence and defocus cues into 4D Epipolar Image (EPI) to complement the disadvantages of each other. On the basis of this work, Wang et al. [6, 7] developed a light-field occlusion model based on the physical image formation, and then Williem et al. [13] proposed two novel data costs for correspondence and defocus cues. Their light-field occlusion models, however, sometimes are difficult to distinguish between depth discontinuities and surface color/texture discontinuities. Moreover, most of existed methods preset the equal interval values of refocusing parameter α during depth search, which results in the non-equal interval displacements during correspondence points searching (refocusing ray tracing).

Optical Flow: Optical flow is an effective tool for the analysis of scene structure or camera motion, and has made significant progress in recent years. Different from traditional optical flow applications focused on large displacements, there are two main challenges for optical flow estimation from light-field images: sub-pixel displacement and ray-level occlusion. However, many excellent optical flow methods still provide the inspiration for us. Ayvaci et al. [14] formulated occlusion detection and optical flow estimation as a joint optimization problem, and presented two efficient numerical schemes to solve it. Revaud et al. [15] presented a sparse-to-dense edge-preserving interpolation of correspondences (EpicFlow) for filling occlusions. Kennedy et al. [16] proposed an optical flow framework with geometric occlusion estimation and multiple frames fusion.

Light-Field Flow Model

We develop our light-field flow model based on the light-field imaging mechanism. Different from the traditional optical flow between two images, our model is built on a single light-field image. According to the light-field imaging geometry, a light-field image includes a set of narrow baseline multi-view images (sub-aperture images). It is evident that optical flows exist in any two sub-aperture images, and they are data-dependent. We will discuss the constraints and the correlations among these optical flows and fuse them into a light-field flow model with occlusion

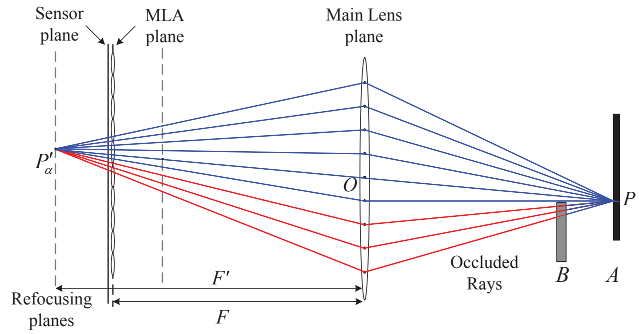


Figure 2. Ray tracking and refocusing diagram. The rays (blue rays) come from the scene point P should converge at a refocused point P'_α, and have the same radiation. Occluded rays (red ones) come from different scene points.

culling.

We first consider two reasonable assumptions before we start to discuss our light-field flow model.

Photo-consistency assumption: This is an important assumption and implied in most existed methods [4, 5]. It means all rays converge at a focused point should come from the same scene point and should have the same radiation, as illustrated by the blue lines in Figure 2. The radiation distribution of these rays is tight in noisy scenarios.

Ray-level occlusion assumption: When occlusion occurs, photo-consistency no longer holds, and the manifestation of occlusion at ray-level is occluded rays [17]. Moreover, these occluded rays might emit from different scene points on the occlusion objects, as shown by the red lines in Figure 2. The radiation distribution of occluded rays is sparse and random.

Light-field flow modeling

Let $\mathscr{L}(\mathbf{x},\mathbf{u})$ and $\mathscr{L}(\mathbf{x}',\mathbf{u}')$ be two sub-aperture images of light-field data, where $\mathbf{x}$ and $\mathbf{x}'$ are the spatial coordinates, and $\mathbf{u}$ and $\mathbf{u}'$ are the angular coordinates, respectively. Under the photo-consistency assumption, the correspondence between two sub-aperture images is given by

$$\mathscr{L}(\mathbf{x},\mathbf{u}) = \begin{cases} \mathscr{L}(\mathbf{w}(\mathbf{x},\mathbf{u}),\mathbf{u}+\Delta\mathbf{u}) + n(\mathbf{x},\mathbf{u}) & \mathbf{x} \notin \Omega \\ \rho(\mathbf{x},\mathbf{u}) & \mathbf{x} \in \Omega \end{cases} \tag{1}$$

where Ω is the occluded region. $\mathbf{w}(\mathbf{x},\mathbf{u})$ is the domain deformation mapping $\mathscr{L}(\mathbf{x},\mathbf{u})$ onto $\mathscr{L}(\mathbf{x}',\mathbf{u}+\Delta\mathbf{u})$ everywhere except the occluded regions. 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 $\rho(\mathbf{x},\mathbf{u})$ that is in general unrelated to $\mathscr{L}(\mathbf{w}(\mathbf{x},\mathbf{u}),\mathbf{u}+\Delta\mathbf{u})|_{\mathbf{x}\in\Omega}$.

According to the refocusing equation of light-field image [1], we have

$$\begin{cases} \mathbf{w}(\mathbf{x},\mathbf{u}) = \mathbf{x} + \beta(\mathbf{x}) \cdot \Delta\mathbf{u} \\ \beta(\mathbf{x}) = 1 - \frac{1}{\alpha(\mathbf{x})} \end{cases} \tag{2}$$

where $\alpha(\mathbf{x}) = F'/F$ is the refocusing parameter and represents the depth for each pixel $\mathbf{x}$.

Usually, optical flow denotes the incremental displacement, i.e.

$$\mathbf{v}(\mathbf{x},\mathbf{u}) = \mathbf{w}(\mathbf{x},\mathbf{u}) - \mathbf{x} = \beta(\mathbf{x}) \cdot \Delta\mathbf{u} \tag{3}$$

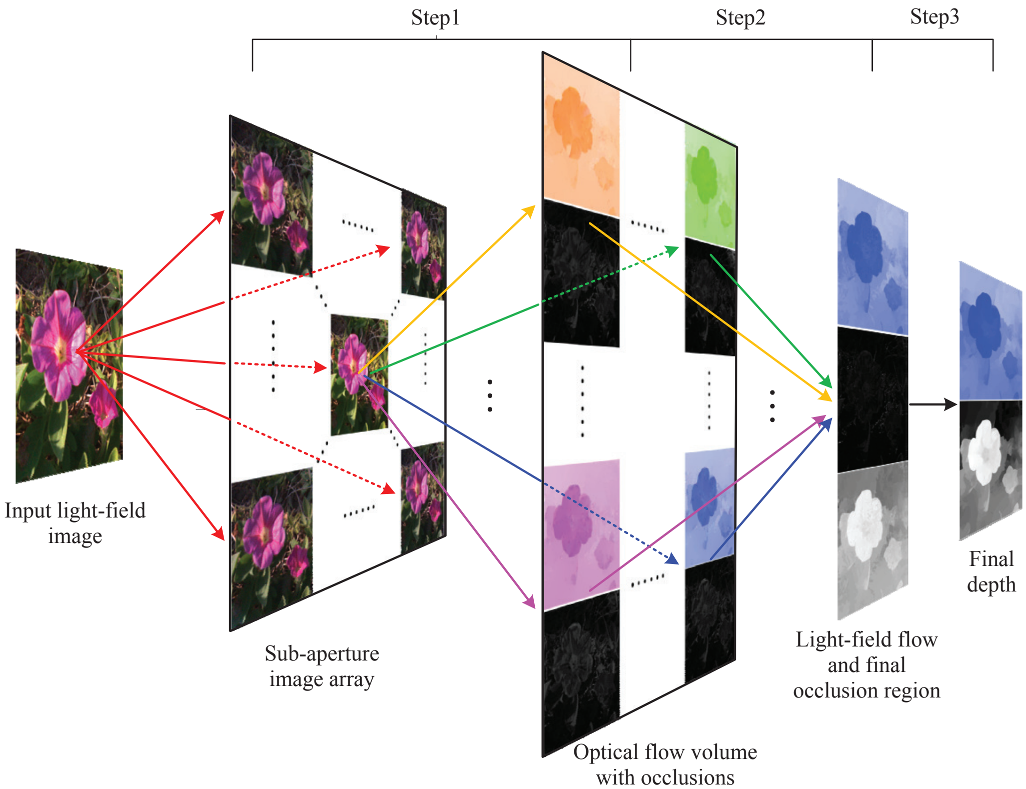


Figure 3. *The framework of our method.*

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 flow fields between any two sub-apertures have the same flow model which is independent of the property of sub-aperture images. We define it as the light-field flow model.

Considering the additive term $n(\mathbf{x},\mathbf{u})$ and the occluded region Ω, the light-field flow model is given by

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

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

According to the photo-consistency assumption and the ray-level occlusion assumption, we note that $\delta(\mathbf{x},\mathbf{u})$ is small but dense and $\varepsilon_\rho(\mathbf{x},\mathbf{u})$ is usually large but sparse, so we will use these properties as an inference criterion for $\beta(\mathbf{x})$ estimation. In the next section, we will introduce a simple but effective method to estimate it by a winner-take-all strategy and a weighted voting mechanism.

Implementation

In this section, we introduce an edge-aware light-field flow estimation framework for joint depth recovery and occlusion detection, as shown in Figure 3. It includes three main steps: a) optical flow volume extraction with sub-pixel accuracy from the sub-aperture image array; b) light-field flow estimation by a winner-take-all strategy and a weighted voting mechanism; c) Edge-aware depth regularization with depth-discontinuity preserving.

Optical flow volume extraction

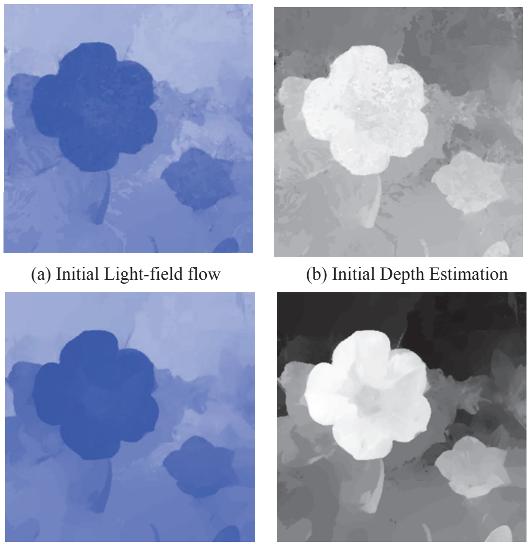
This step independently estimates optical flow from any two sub-aperture images, and then build them into an optical flow volume. Considering the sub-pixel displacements and ray-level occlusion between two sub-aperture images, we should use an optical flow with sub-pixel accuracy.

We adopt the method of Ayvaci et.al [14] to estimate the optical flow between each sub-aperture image and the central sub-aperture image. The advantage of this method is it formulates the sub-pixel accurate optical flow estimation and sparse occlusion detection as a joint optimization problem. What need to be explained is that we do not further use EpicFlow [15] to improve the estimation accuracy of occluded regions [18] because of its limited improvement and high time-consuming.

Light-field flow estimation

According to the Equation 4, we unify the optical flow volume into a light-field flow model.

For the pixel $\mathbf{x}$ in the central sub-aperture image, the horizontal and vertical components of each optical flow in the optical flow volume can provide two potential $\beta(\mathbf{x})$ values of pixel $\mathbf{x}$. That is we can get a set of potential β values from the optical flow volume for each pixel. These potential β values undoubtedly are degraded by noise $\delta(\mathbf{x},\mathbf{u})$ and $\varepsilon_\rho(\mathbf{x},\mathbf{u})$. Our task is estimating an accurate β value for each pixel from these degraded data. Inspired



(a) Initial Light-field flow (b) Initial Depth Estimation

(c) Refined Light-field flow (d) Refined Depth Estimation

Figure 4. *The effectiveness of our regularization step.*

by the properties of $\delta(\mathbf{x},\mathbf{u})$ and $\varepsilon_\rho(\mathbf{x},\mathbf{u})$ (discussed in the previous section), we propose a simple and efficient weighted voting mechanism for β selection.

For each pixel, more specifically, we uniformly dividing the valid range of β into several levels. Each potential β value is treated as a voter, and each voter is assigned to a weight according to the occlusion probability of each pixel detected in step 1. In other words, the higher the occlusion probability, the lower the vote weights. Then we use the winner-take-all strategy to select the level with the highest number of votes as initial light-field flow (β value) for each pixel. Initial depth map can be computed by Equation 2, as shown in Figure 4(a)(b).

Edge-aware depth regularization

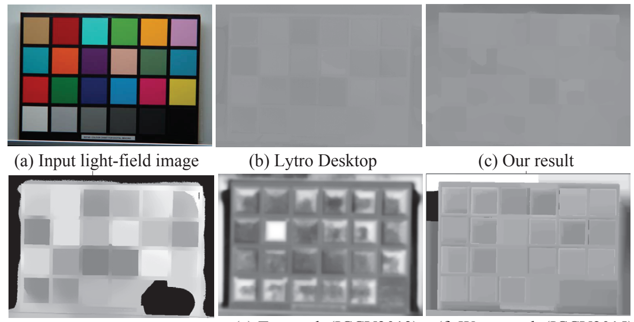
Given initial light-field flow, we regularize it with weighted median filter [19] for the final light-field flow and depth map, as shown in Figure 4(c)(d). This weighted median filter is a kind of edge-aware filter based on guided filters [20, 21]. It enforces a piecewise smooth flow with depth-discontinuity preserving for each β slice (level). The central sub-aperture image is used as the guidance image of guided filter.

Experimental Comparisons

In our implementation, the refocusing parameter α ranges from 0.2 to 2.0, so the valid range of β is from -4.0 to 0.5, 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 Jeon et al. [8], Tao et al. [4] and Wang et al. [6, 7]. The source codes can be downloaded from the authors' project webpages.

We validate our method on real-world scenes. Some are taken by Lytro 1.0 and Lytro Illum cameras, and the sub-aperture images are extracted by LFToolbox 0.4 [22, 23]. Others are from EPFL Light-Field Image Dataset [24]. We focus on the algorithm effectiveness and robustness in handling discontinuities and occlusions. 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.



(a) Input light-field image (b) Lytro Desktop (c) Our result

(d) Jeon et al. (CVPR2015) (e) Tao et al. (ICCV2013) (f) Wang et al. (ICCV2015)

Figure 5. *Robustness comparison on a non-occlusion plane scene.*

Depth maps for plane scenes Figure 5 shows an interesting robustness comparison on a non-occlusion plane scene. The real scene is just a picture of the Color Chart. There is no occlusion on the scene except the rich color/texture discontinuity edges. Jeon's, Tao's and Wang's methods all fail to recover a meaningful depth map because of the color/texture edge effect. On the contrary, our method gives a satisfactory result and is insensitive to the surface color/texture discontinuities.

Depth maps for real scenes Figures 6 and 7 show the light-field flow and the depth estimations results on real scenes with fine structures and occlusions, captured with Lytro 1.0 and Lytro Illum light-field camera, respectively. Experimental results show both Joen's and Tao's methods lose fine structures, because the details are vulnerable to textures, edges and noises. Wang's method can recover fine details at occlusion boundaries, but it is not robust to the regions with rich textures. Clearly, our method performs better, especially at depth discontinuities and occlusions regions.

Conclusion and Future works

In this paper, we develop a light-field flow model for light-field cameras by fusing all the optical flows between each sub-aperture image and the central sub-aperture image. Then an edge-aware light-field flow estimation framework is presented, targeted at two main challenges of light-field flow: sub-pixel displacement and ray-level occlusion. We demonstrate the benefits of light-field flow by evaluating our algorithm against state-of-the-art methods.

In our method, however, optical flow volume extraction is still time-consuming because the optical flows estimation from sub-apertures are independent. Future works will focus on this issue.

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). The authors are grateful for the anonymous reviewers who made constructive comments.

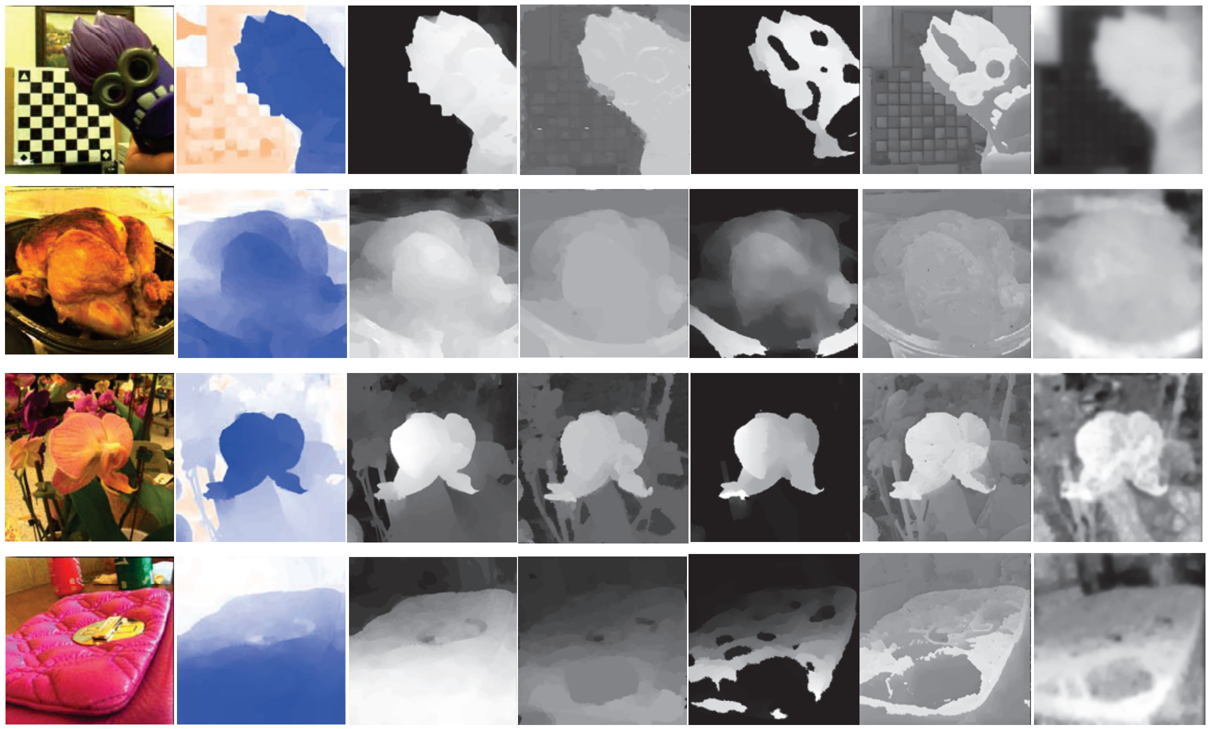


Figure 6. *Light-field flow and depth estimation results on real scenes taken by the Lytro 1.0 light-field camera. These data are selected from the work of Lin et al., and we get their depth estimation results from their paper [12].*

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 CSTR, 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] C.-K. Liang and R. Ramamoorthi, A light transport framework for lenslet light field cameras, ACM Transactions on Graphics (TOG), vol. 34, no. 2, 2015.

[4] M. 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 (ICCV), 2013.

[5] M. Tao, 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 (TPAMI), 2015.

[6] T.-C. Wang, A. Efros, and R. Ramamoorthi, Depth estimation with occlusion modeling using light-field cameras, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2016.

[7] T.-C. Wang, A. Efros, and R. Ramamoorthi, Occlusion-aware depth estimation using light-field cameras. in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015.

[8] 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 Proceedings of International Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

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

[10] C. Perwass and P. Wietzke, Single lens 3d-camera with extended depth-of-field, in SPIE Elect. Imaging, 2012.

[11] 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 (ICCV), 2013.

[12] H. Lin, C. Chen, S. B. Kang, and J. Yu, Depth recovery from light field using focal stack symmetry. in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015.

[13] Williem and In Kyu Park, Robust Light Field Depth Estimation for Noisy Scene with Occlusion, in Proceedings of International Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[14] A. Ayvaci, M. Raptis and S. Soatto, Sparse Occlusion Detection with Optical Flow. International Journal of Computer Vision (IJCV), 2012. 9(3): p. 322-338.

[15] J. Revaud, P. Weinzaepfel, Z. Harchaoui, C. Schmid, EpicFlow: Edge-Preserving Interpolation of Correspondences for Optical Flow, in Proceedings of International Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[16] R. Kennedy and C. Taylor. Optical Flow with Geometric Occlusion Estimation and Fusion of Multiple Frames. in Energy Minimization Methods in International Conference on Computer Vision and Pattern Recognition(CVPR). 2015.

[17] W. Zhou, A. Lumsdaine, and L. Lin, Depth Estimation with Cascade Occlusion Culling Filter for Light-field Cameras , in 23rd Inter-

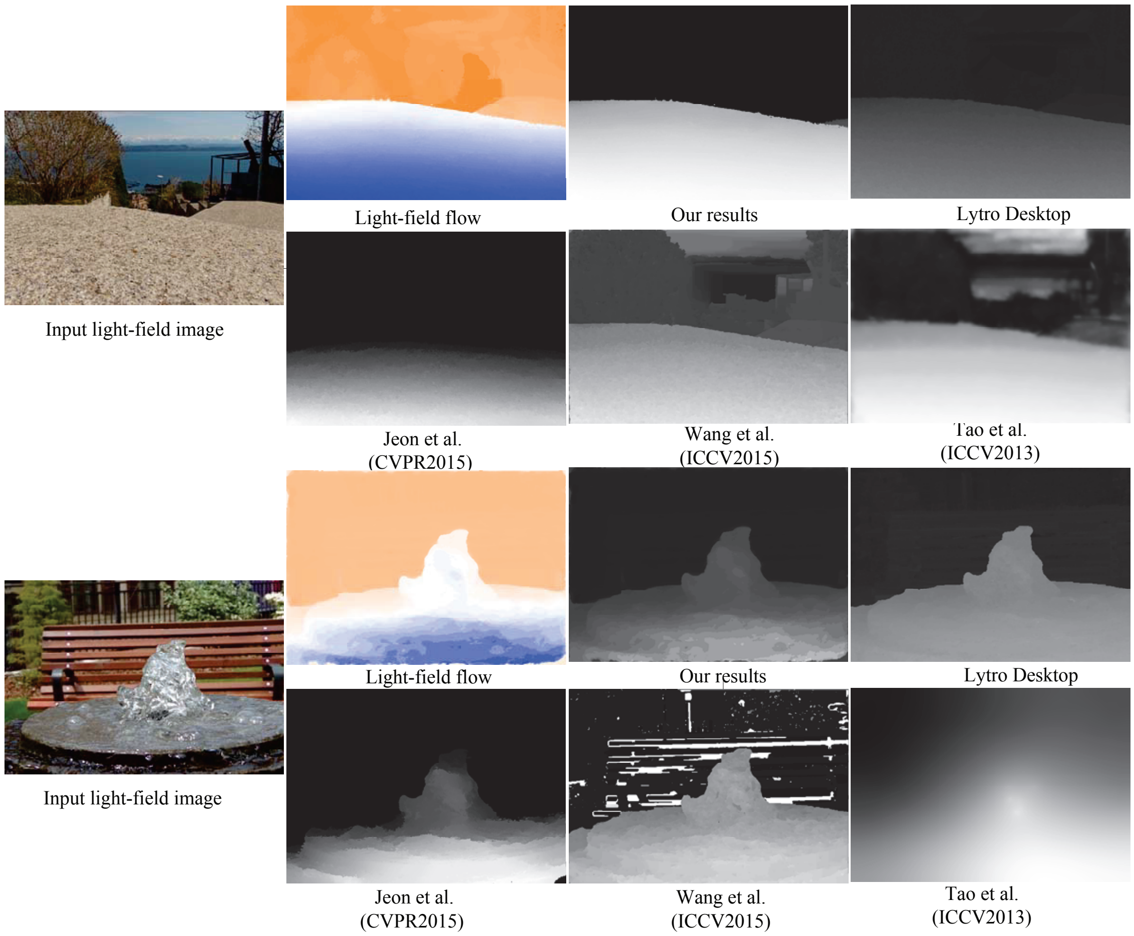


Figure 7. *Depth estimation results on real scenes taken by the Lytro Illum light-field camera. These data are selected from EPFL Light-Field Image Dataset [24].*

national Conference on Pattern Recognition (ICPR). 2016.

[18] Q. Chen and V. Koltun, Full Flow: Optical Flow Estimation By Global Optimization over Regular Grids, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016.

[19] Z. Ma, K. He, Y. Wei, J. Sun, and E. Wu. Constant time weighted median filtering for stereo matching and beyond. In Proceedings of International Conference on Computer Vision (ICCV), 2013.

[20] K. He, J. Sun, and X. Tang, Guided image filtering, in European Conference on Computer Vision (ECCV), 2010.

[21] K. He, J. Sun, and X. Tang, Guided image filtering, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), vol. 35, no. 6, pp. 1397?1409, 2013.

[22] 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 (CVPR). IEEE, Jun 2013.

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

[24] M. Rerabek and T. Ebrahimi, New light field image dataset, in 8th International Conference on Quality of Multimedia Experience (QoMEX), 2016.

Author Biography

Wenhui Zhou received his Ph.D from Zhejiang University(2005) and worked as a postdoctoral researcher in the Department of Information Science & Electronic Engineering at Zhejiang University (2005-2007). Since then he has worked in the School of Computer Science and Technology at Hangzhou Dianzi University, China. His work has focused on image processing, computer vision and computational photography.

Andrew Lumsdaine received his Ph.D. from MIT(1992). He is the Chief Scientist at the Northwest Institute for Advanced Computing (NIAC), jointly founded by the University of Washington and Pacific Northwest National Laboratory. Prior to coming to NIAC, he was Professor of Computer Science and the Associate Dean for Research in the School of Informatics and Computing and Director of the Center for Research in Extreme Scale Technologies (CREST) at Indiana University (2001-2016). His research interests include computational science and engineering, parallel and distributed computing, parallel graph algorithms, generic programming, and computational photography.

Lili Lin received her Ph.D from Zhejiang University(2005). Since then she has worked in the College of Information and Electronic Engineering at Zhejiang Gongshang University, China. Her work has focused on image processing, computer vision and computational photography.