An Efficient Anti-Occlusion Depth Estimation using Generalized EPI Representation in Light Field

Hao Zhu and Qing Wang

School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China

ABSTRACT

Light field cameras have been rapidly developed and are likely to appear in mobile devices in near future. It is essential to develop efficient and robust depth estimation algorithm for mobile applications. However, existing methods are either slow or lack of adaptability to occlusion such that they are not suitable to mobile computing platform. In this paper, we present the generalized EPI representation in light field and formulate it using two linear functions. By combining it with the light field occlusion theory, a highly efficient and anti-occlusion depth estimation algorithm is proposed. Our algorithm outperforms the previous local method, especially in occlusion areas. Experimental results on public light field datasets have demonstrated the effectiveness and efficiency of the proposed algorithm.

Keywords: Light Field, Depth Estimation, Anti-Occlusion, Generalized EPI Representation

1. INTRODUCTION

Light field cameras from Lytro[1] and Raytrix[2] have been rapidly developed and are likely to appear in mobile devices in near future. It is essential to develop efficient and robust depth estimation algorithms for mobile applications.[3–7] However, existing methods are either slow or lack of adaptability to occlusion such that they are not suitable to mobile computing platform.

Wanner *et al.*[8] applied structure tensor to analyze the horizontal and vertical Epipolar Plane Images (EPI) in light field, and measured the reliability of estimated depth using the coherence of structure tensor. However, in occlusion areas, the tensor field becomes too random to estimate, and structure tensor tends to produce wrong estimation but assigns high reliability,[9] which leads to over smooth results in occlusion boundaries.

Yu *et al.*[10] encoded the constraints of 3D lines and introduced Line Assisted Graph Cuts (LAGC) to improve depth estimation. However, the 3D lines are partitioned into small and incoherent segments in occlusion which leads to wrong estimation.

Based on the analysis of advantages and disadvantages of the defocus and correspondence cues in light field, Tao *et al.*[11] proposed to optimize depth map by combining these two cues in a Markov Random Field framework.[12] However, the light field is under sampling in angular space in occlusion areas, and the defocus and correspondence cues failed in these areas.

Based on Tao *et al.*'s work,[11] Wang *et al.*[13] analyzed the formation of the occlusion in light field, and discovered the consistency between spatial space and angular space in occluded boundaries in a local patch. They selected the un-occluded views according to the edge orientation in the spatial patch, and modified the previous algorithm using the consistency of multiple cues in the un-occluded views. Experimental results have proven that this method can achieve the best results in occlusion boundaries. However, the refocus operation[14] for constructing Disparity Space Image (DSI) is time-consuming, and inappropriate regularization parameters will lead to over smooth results in occlusion boundaries (see Fig. 5,6).

In this paper, we propose a framework of Generalized Epipolar Plane Image (GEPI) representation in light field. In this framework, the EPI is obtained by shearing a 2D slice in 4D space while traditional EPI is just a special case sheared in horizontal or vertical direction. We formulate the GEPI using two linear functions in

Further author information: (Send correspondence to Qing Wang)
Qing Wang: E-mail: qwang@nwpu.edu.cn, Telephone: 86 29 8843 1518

Optoelectronic Imaging and Multimedia Technology IV, edited by Qionghai Dai, Tsutomu Shimura, Proc. of SPIE Vol. 10020, 1002008 · © 2016 SPIE
CCC code: 0277-786X/16/$18 · doi: 10.1117/12.2246086

mathematics, and the 2D slice in any angle which satisfies the conditions can be called a GEPI. Theoretically, each GEPI can be used for depth estimation, however it does not hold due to the occlusion. By modeling the occlusion in light field, there is at least one GEPI representation which is free or nearly free of occlusion (it is called occlusion-free GEPI later). We propose to select the best occlusion-free GEPI in light field, and present an efficient local depth estimation algorithm.

The rest of the paper is organized as follows: In section 2, the background of structure tensor and occlusion model are reviewed. In section 3, we define the GEPI representation in light field and introduce the method of GEPI selection and depth estimation. The experimental results are demonstrated in section 4. We summarize the paper and give the suggestions for future work in section 5.

2. BACKGROUND

In this section, we review the robust local depth estimation and the occlusion model in light field.

2.1 Robust local depth estimation

Light field has a simplified representation of radiance by a 4D function $f(x, y, u, v)$, where the dimensions (x, y), (u, v) describe the light distribution in spatial and angular space respectively.[15] When we fix one angular and spatial dimension (x^*, u^*) or (y^*, v^*), the EPI representation of light field appears. Since the slope of EPI line has a linear relationship with the depth,[16] depth estimation can be converted to the slope analysis in EPI.

We use the structure tensor to analyze the slopes of EPI lines,

$$J = \begin{bmatrix} G_\sigma(S_x S_x) & G_\sigma(S_x S_y) \\ G_\sigma(S_x S_y) & G_\sigma(S_y S_y) \end{bmatrix} = \begin{bmatrix} J_{xx} & J_{xy} \\ J_{xy} & J_{yy} \end{bmatrix}, \tag{1}$$

where G_σ represents a Gaussian smoothing operator, and S_x,S_y represent the gradient of EPI in x and y axis respectively.

The slope k of EPI line is estimated using

$$k = \frac{1}{\tan(\frac{1}{2}\arctan(\frac{J_{yy}-J_{xx}}{2J_{xy}}))} \tag{2}$$

and the reliability is measured by the coherence of the structure tensor

$$R = \frac{(J_{yy} - J_{xx})^2 + 4J_{xy}^2}{(J_{xx} + J_{yy})^2}. \tag{3}$$

2.2 Occlusion model

Wang *et al.*[13] have analyzed the formation of occlusion in light field. By assuming the single-occlusion among all views and looking at a spatial patch small enough, the occlusion edge can be approximated by a straight line.

Considering a pixel (X_0, Y_0, F) on the focal plane (the left image in Fig. 1(a)), and an occluder intersects at (X_0, Y_0, Z_0) $(0 < Z_0 < F)$. The directional vector of the occluder boundary in the plane $Z = Z_0$ is

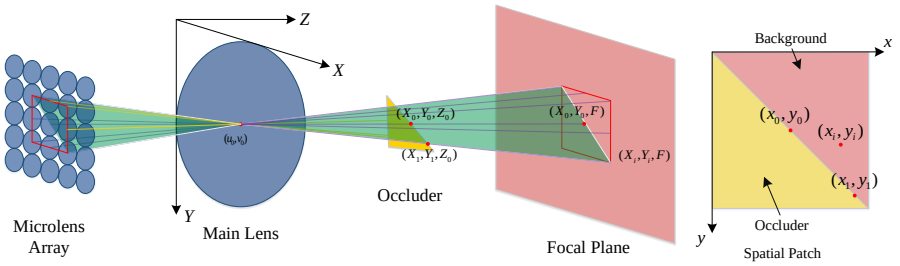
$$\vec{e}_W^1 = (k_{X_1}, k_{Y_1}) = (X_1 - X_0, Y_1 - Y_0). \tag{4}$$

The golden areas in Fig. 1 denote the occluded areas. Without loss of generality, we assume $k_{Y_1} > 0$.

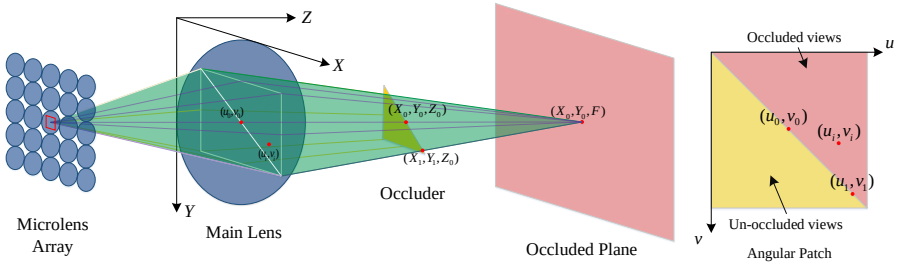
For any other pixels (X_i, Y_i, F) on the focal plane, it will be observed by the view (u_0, v_0) iff it meets the following inequality,

$$k_{Y_1}(X_i - X_0) - k_{X_1}(Y_i - Y_0) < 0, i = 1, 2, ..., n. \tag{5}$$

We then project these inequalities from the world coordinate system to image coordinate system (the right image in the Fig. 1(a)). The corresponding directional vector of $\vec{e}_W^1$ is $\vec{e}_I^1 = (k_{x_1}, k_{y_1})$ and we have $\vec{e}_I^1 = \lambda_1 \vec{e}_W^1$,



(a)



(b)

Figure 1. The light field camera model with occlusion. The left image in (a) denotes the the image captured from the view (u_0, v_0), and the right image in (a) is a local patch centered at (x_0, y_0) from the view (u_0, v_0). The left image in (b) denotes the light field is refocused in depth F, and only views upon the green plane can see the point (X_0, Y_0, F), the images formed from other views describe the occluder. The right image in (b) is the angular patch of the point (x_0, y_0).

where λ_1 is a scale factor to denote the scaling relationship between the world and image coordinate systems. For any other points (x_i, y_i) on the image, it is a background point iff

$$k_{y_1}(x_i - x_0) - k_{x_1}(y_i - y_0) < 0. \tag{6}$$

Then considering the main lens plane (the left image in Fig. 1(b). The light field is refocused to the depth F). For any other views (u_i, v_i) on the main lens plane, it can capture the pixel (X_0, Y_0, F) iff

$$k_{v_1}(u_i - u_0) - k_{u_1}(v_i - v_0) < 0, \tag{7}$$

where $\vec{e}_A^1 = (k_{u_1}, k_{v_1})$ and $\vec{e}_A^1 = \lambda_2 \vec{e}_I^1$. λ_2 is a scale factor to denote the scaling relationship between the image and angle coordinate systems.

Revisiting Eqns. 6 and 7, it is noticed that the occluded views in angular space has the same distribution with the occluder in spatial space. Furthermore, the edge orientation in the angular space can be predicted using the edge in the spatial space.

3. THE PROPOSED METHOD

3.1 Generalized EPI representations in light field

In traditional multi-view stereo, the EPI is constructed by taking a regularly spaced series of images from a linear moving camera system, and the moving direction can be arbitrary. However, existed EPI representations[8] in light field are only limited in the horizontal and vertical directions, ignoring other directions.

For a given point $p = (x^*, y^*, u^*, v^*)$ in 4D light field space, there are many 2D slices containing it and each slice can be called a Generalized EPI (GEPI) iff it satisfies the following two equalities

$$\begin{aligned} a(u - u^*) + b(v - v^*) = 0, \\ a(x - x^*) + b(y - y^*) = 0, \end{aligned} \tag{8}$$

Figure 2. The GEPI representations in light field. (a) shows how GEPIs are obtained in different directions. (b) shows 4 GEPIs in four directions. It is noticed there is no occlusion in -45° GEPI, as the occlusion boundary near p has the same orientation as GEPI direction.

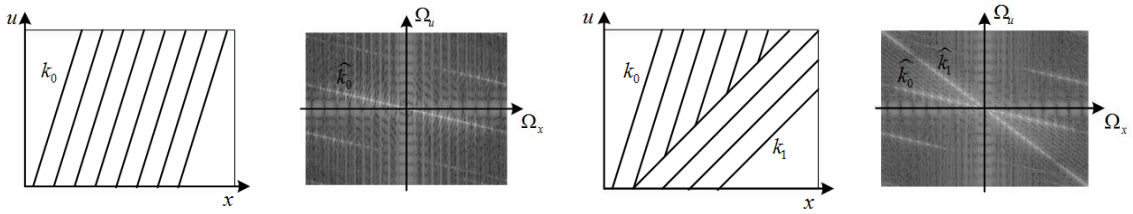


Figure 3. The property of Fourier transformation for EPI. (a) is the EPI for the scene with constant depth, all EPI lines has the same slope k_0. (b) is the fourier transform of (a), it is noticed all energy concentrate to the line $\Omega_u = \widehat{k}_0\Omega_x$. (c) is the EPI with occlusion. The slope of the foreground is k_1. (d) is the fourier transform of (c), it can be seen the energy from background concentrate to $\Omega_u = \widehat{k}_0\Omega_x$, and the energy from from foreground concentrate to $\Omega_u = \widehat{k}_1\Omega_x$.

where a and b are real numbers to describe the moving direction of camera.

It is noticed that the traditional horizontal and vertical EPI representations in 4D light field are just two special cases of GEPI representation, *i.e.*, $(a = 0, b = 1)$ and $(a = 1, b = 0)$ respectively. Furthermore, when $(a = 1, b = -1)$, diagonal (45°) EPI appears, and reverse diagonal (-45°) EPI is obtained by setting $(a = 1, b = 1)$. The GEPI obtention in different directions can be seen in Fig. 2(a), and the examples are shown in Fig. 2(b). In a word, GEPI representation can describe the moving camera system in arbitrary direction in light field.

3.2 Robust local anti-occlusion depth estimation method

As described in the Sec. 2.2, the edge in the angular space has the same orientation with the edge in the spatial space in occluded areas. A straightforward depth estimation is to select occlusion-free GEPI (-45° GEPI in Fig. 2(b)) according to the orientation of edge in the spatial space. However, it will fail in the textured occlusion boundaries since there are conflicts between the orientations of the texture edges and occlusion boundaries. Thus, it is difficult to distinguish the texture edges and occlusion boundaries without an accurate disparity map.

As a basic property of Fourier transformation, a linear symmetric image has a Fourier transform concentrated to a line passing the coordinate origin, and a line in Fourier domain is perpendicular to all lines in the spatial domain[17] (Fig. 3(a), 3(b)). Based on this feature, the structure tensor method fits a least square error line in the Fourier domain of the local EPI, and it is suitable for most situations (Fig. 3(a), 3(b)) except occlusions (Fig. 3(c), 3(d)).

In occlusion areas (Fig. 3(c) and 3(d)), the background is always occluded by the foreground. Since the foreground has a smaller depth than background, the slope of foreground k_1 is always smaller than the slope of

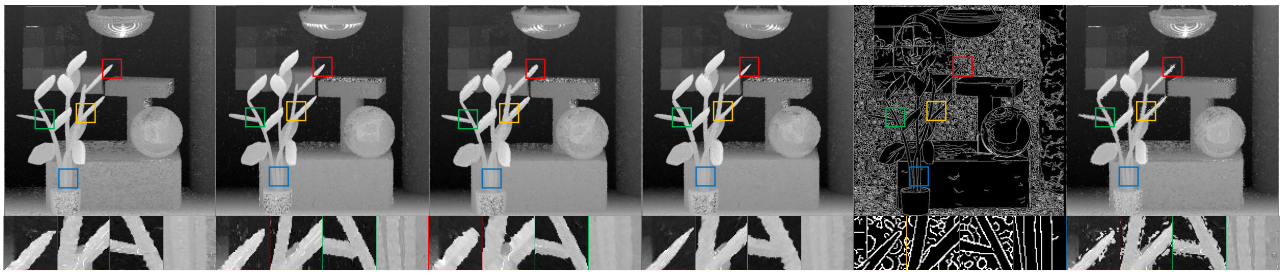


Figure 4. The conflicts between the orientation from texture edge and occlusion boundaries. (a)-(d) are the depth maps obtained from the horizontal, vertical, diagonal (45°) and reverse diagonal (−45°) GEPIs respectively. It is noticed the GEPI which has a same orientation with the occlusion boundary yield sharper occlusion boundaries. (e) is the edge map. (f) is the depth map selected from the previous 4 results according to the orientations of edges. The conflicts between the orientations from textures and occlusion boundaries lead to the poor results in occlusion boundaries.

background k_0 in EPI. After the Fourier transformation, the energy of the foreground and background concentrates to two lines, *i.e.*, $\Omega_u = \widehat{k}_0\Omega_x$ and $\Omega_u = \widehat{k}_1\Omega_x$. Supposing the fitting line has a slope $\widehat{k}^*$, then $\widehat{k}^*$ meets the following inequalities,

$$\widehat{k}_1 \leqslant \widehat{k}^* \leqslant \widehat{k}_0. \tag{9}$$

As the line in Fourier domain is perpendicular to lines in the spatial domain, *i.e.*, $k \cdot \hat{k} = -1$, the estimated slope k^* meets the following inequalities,

$$k_1 \leqslant k^* \leqslant k_0. \tag{10}$$

In other words, the estimated depth in occlusion areas is always less than or equal to the ground truth of the depth. That is the reason why structure tensor yields over smooth results in occlusion boundaries (see Fig. 4). With this feature, the following proposition is obtained

Proposition 1. *The depth estimated from occlusion-free EPI is always larger than or equal to non-occlusion-free EPIs in occlusion areas.*

Using this proposition, we propose a robust local anti-occlusion depth estimation method in Algo. 1. At first, we estimate N depth maps $D_i, i = 1, ..., N$ from the GEPI representations using structure tensor. Then for each possible occlusion point p [*] , the largest value $D^*(p)$ from all candidate depth $D_i(p), i = 1, ..., N$, is assigned to it final depth. For other points, the depths are selected according to the reliability of structure tensor.

Algorithm 1 The proposed robust local anti-occlusion depth estimation method

Input:
4D light field LF
Output:
The depth map D^* of the central view
Process:
(D_i, R_i)=StructureTensor$(GEPI_i)$, $i = 1, ..., N$.
$D^*(p) = \max D_i(p)$, if p is a possible occlusion point.
$D^*(p) = D_j(p)$, where $j = \arg\max_i R_i$, if p is not an occlusion point.

4. EXPERIMENTAL RESULTS

We compare the proposed method with the state-of-the-art methods in light field, which include the globally consistent depth labeling (GCDL) by Wanner *et al.*,[8] the line assisted graph cut (LAGC) by Yu *et al.*,[10] the depth

[*]The occlusion map is an dilated edge map here. Since an occlusion point is an edge point but an edge point may not be an occlusion point.

from defocus and correspondence (DFC) by Tao *et al.*,[11] and the occlusion-aware depth estimation (OADE) by Wang *et al.*[13] All results come from their published codes or executable files. It is noted that the DFC method from Tao *et al.* is only compared in the real scenes since the interface in only for the Lytro lfp files.

4.1 Depth Maps in Synthetic Scenes

The most popular light field datasets LFBD[18] are used to evaluate the quantitative performance of the proposed method. These datasets contain both synthetic and real scenes with the ground truth. Considering the light field is a sampling under a discrete 9×9 pattern, the number of GEPIs, N, is set to 4 to ensure that the angular sampling is sufficient to construct an EPI.

Fig. 5 shows qualitative comparisons in synthetic scenes. It can be seen our method yields more clear occlusion boundaries than the state-of-the-arts. Our method retains the details for small occlusion boundaries, such as the twig and leaf in the Mona, and the butterfly antennae and the leaf in the Papillon. Since the proposed method is just a local method and no smooth term is applied to regularize the final result, there are some noises in low-texture areas of the recovered depth map.

The quantitative comparisons are listed in Table 1. It is noticed the RMS error is not the best metric to measure the performance of our method in occlusion boundaries. For all pixels, the performance of the proposed method is not worse than the state-of-the-arts. However, the proposed method shows obvious advantages to the existed methods in occlusion areas. Our method achieves the minimum RMS error in two datasets, and the ranks in another 2 datasets are 2nd and 3rd.

Table 1. RMS errors of recovered disparity.

	GCDL		LAGC		OADE		Ours	
	all	occ	all	occ	all	occ	all	occ
Buddha2	**0.101**	0.353	0.179	0.404	0.107	0.299	0.133	**0.263**
Mona	0.098	0.469	0.119	**0.300**	0.089	0.328	**0.081**	0.340
Papillon	0.166	0.758	0.406	0.680	**0.125**	**0.389**	0.141	0.463
Maria	**0.063**	**0.277**	0.073	0.311	0.061	0.299	0.119	**0.277**

4.2 Depth Maps in Real Scenes

Fig. 6 shows the comparisons in real scenes, captured with the Lytro light field camera. The LFtoolbox[19] is used to decode the lfp files from Lytro camera. The disparity ranges are set to $[-1, 1]$ and the depth levels are set to 100 for all methods. It is noticed the performances of all methods decrease a lot due to the heavy noise in real scenes captured by Lytro camera. However, our method performs better in occlusion boundaries, especially in the edges of the leafs in the first, third and fourth rows and the grid lines in the second row.

4.3 Running Times

All methods are evaluated in a same machine, with a 3.4GHz CPU, 24G RAM. The CUDA environment for GCDL is the GTX1080. For MRF based methods[10,11,13] and GCDL,[8] the labels are set 100 to balance the performance and the speed. The running time of different methods are listed in Table. 2. To be fair, the language environments of different methods are also listed in Table. 2. It is noticed our method has the lowest computing cost since there is no refocus operation in our algorithm.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we propose the GEPI representation in light field and develop an efficient and anti-occlusion depth estimation method. We find that there is always an occlusion-free GEPI for occlusion point in light field and the depth estimated from the occlusion-free GEPI is always larger than the depth estimated from others. Utilizing this discovery, the depth estimation is improved in two steps. First, we estimate several depth maps from different directional GEPIs. Then, the largest depth value is selected for occlusion point. We have evaluated the method

Table 2. The running time of different methods. All methods are tested in a $9 \times 9 \times 378 \times 379$ color light field.

	GCDL	LAGC	DFC	OADE	Ours
Time	65s	137min	90s	78s	120s
Language	C++/CUDA	C/C++	Matlab+C	Matlab+C	Matlab

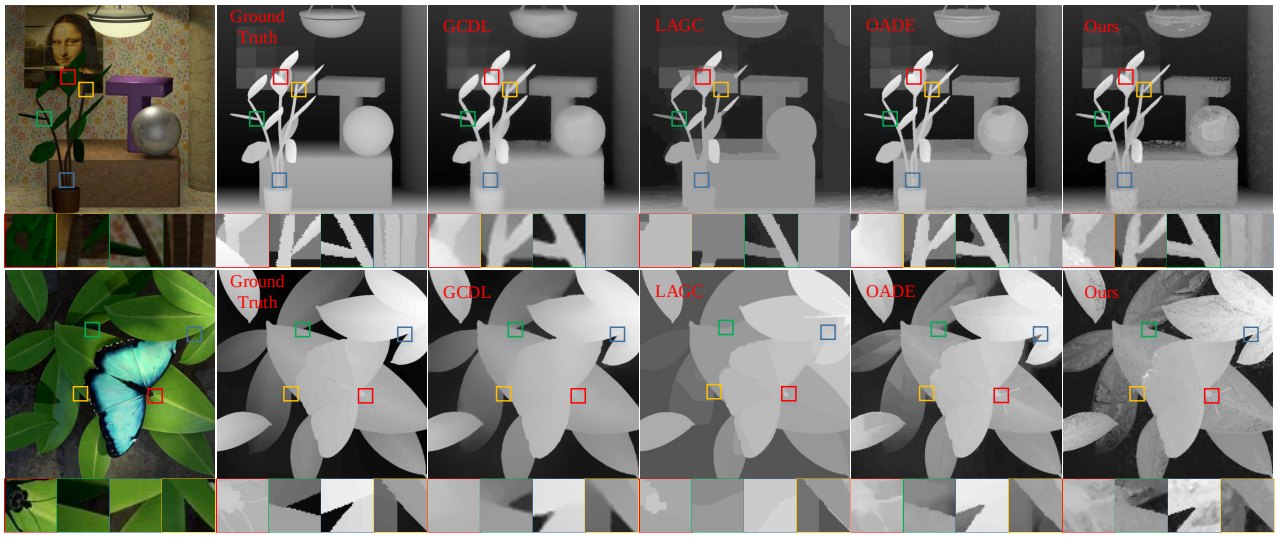


Figure 5. Recovered depth maps on synthetic scenes. The most left two columns are light field scenes and the ground truth depth. The middle three columns are the depth maps recovered from GCDL,[8] LAGC[10] and OADE[13] respectively. The most right column is the depth maps by the proposed method.

on synthetic and real scenes. Experimental results have validated our method has high performance in occlusion boundaries.

Since the proposed method follows the same occlusion model with OADE,[13] it fails in small-occlusions and multi-occluder occlusion. Apart from this, as noticed in Table 1, Fig. 5 and 6, although the method recovers the thin structures and the occlusion boundaries clearly, it performs not well in low-texture areas since no smooth term is applied. It is necessary to regularize the depth map with an appropriate smooth term to improve the performance.

Acknowledgement. The work in the paper is supported by NSFC funds (61272287, 61531014).

REFERENCES

[1] Lytro, "Lytro redefines photography with light field cameras." http://www.lytro.com (2011).
[2] raytrix, "Raytrix lightfield camera." http://www.raytrix.de (2012).
[3] Mihara, H., Funatomi, T., Tanaka, K., Kubo, H., Nagahara, H., and Mukaigawa, Y., "4d light field segmentation with spatial and angular consistencies," (2016).
[4] Xu, Y., Nagahara, H., Shimada, A., and Taniguchi, R.-i., "Transcut: Transparent object segmentation from a light-field image," in [*2015 IEEE International Conference on Computer Vision (ICCV)*], 3442–3450, IEEE (2015).
[5] Maeno, K., Nagahara, H., Shimada, A., and Taniguchi, R.-i., "Light field distortion feature for transparent object recognition," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 2786–2793 (2013).
[6] Zhang, X., Wang, Y., Zhang, J., Hu, L., and Wang, M., "Light field saliency vs. 2d saliency: A comparative study," *Neurocomputing* **166**, 389–396 (2015).

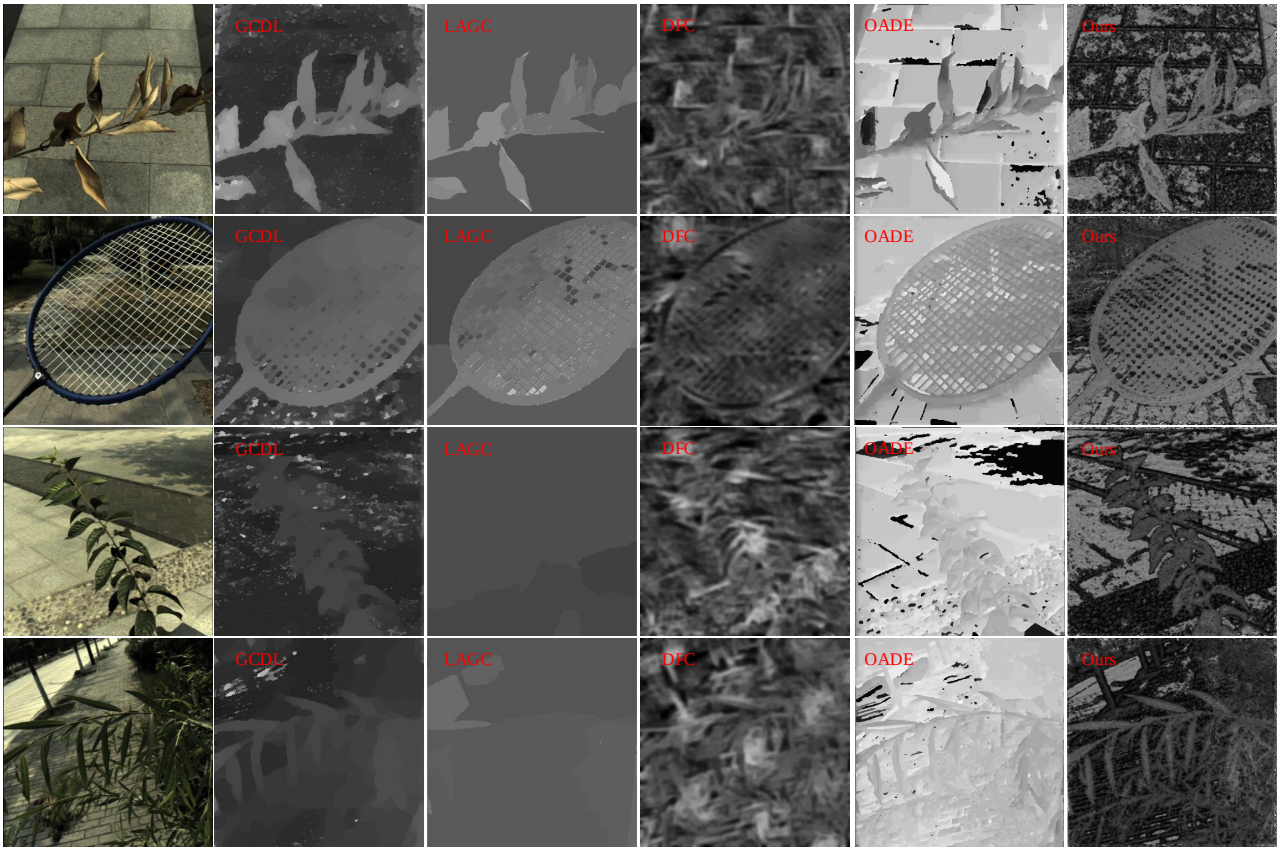


Figure 6. Recovered depth maps on real scenes. The first column is the input light field. And the rest columns are the recovered depth maps from GCDL,[8] LAGC,[10] DFC,[11] OADE[13] and ours respectively.

[7] Jarabo, A., Masia, B., Bousseau, A., Pellacini, F., and Gutierrez, D., "How do people edit light fields?," *ACM Trans. Graph.* **33**(4), 146–1 (2014).

[8] Wanner, S. and Goldluecke, B., "Variational light field analysis for disparity estimation and super-resolution," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **36**(3), 606–619 (2014).

[9] Li, J., Lu, M., and Li, Z.-N., "Continuous depth map reconstruction from light fields," *IEEE Transactions on Image Processing* **24**(11), 3257–3265 (2015).

[10] Yu, Z., Guo, X., Lin, H., Lumsdaine, A., and Yu, J., "Line assisted light field triangulation and stereo matching," in [*Proceedings of the IEEE International Conference on Computer Vision*], 2792–2799 (2013).

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

[12] Blake, A., Kohli, P., and Rother, C., [*Markov random fields for vision and image processing*], Mit Press (2011).

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

[14] Ng, R., *Digital light field photography*, PhD thesis, stanford university (2006).

[15] Levoy, M. and Hanrahan, P., "Light field rendering," in [*Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*], 31–42, ACM (1996).

[16] Hartley, R. and Zisserman, A., [*Multiple view geometry in computer vision*], Cambridge university press (2003).

[17] Bigun, J., "Optimal orientation detection of linear symmetry," (1987).

[18] Wanner, S., Meister, S., and Goldluecke, B., "Datasets and benchmarks for densely sampled 4d light fields.," in [*VMV*], 225–226, Citeseer (2013).

[19] Dansereau, D. G., Pizarro, O., and Williams, S. B., "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 1027–1034 (2013).