GUIDED INTEGRAL FILTER FOR LIGHT FIELD STEREO MATCHING

Hao Sheng, Shuo Zhang, Gengliang Zhu, Zhang Xiong

State Key Laboratory of Software Development Environment, School of Computer Science and Engineering
Beihang University, Beijing 100191, P. R. China

ABSTRACT

Different from the traditional multi-view images, the sampling in angular resolution of light field images is continuous in each direction. Therefore, an angular sampling image, comprising of matching points extracted from each view, can be constructed at different possible depth for each point. In this paper, we prove that the angular sampling image of an occluded point at the correct depth is similar to the scene around the point. On this basis, a guided integral filter acquired from the reference view is proposed to weight the matching points for consistency measure, which predicts the probabilities of occlusions. The discrete labeling problem is then solved by a filter-based algorithm to approximate the optimal solution. Experimental results demonstrate that the proposed method outperforms the state-of-the-art methods for light field depth estimation both in occluded and ambiguous regions, and has no requirement for angular resolution.

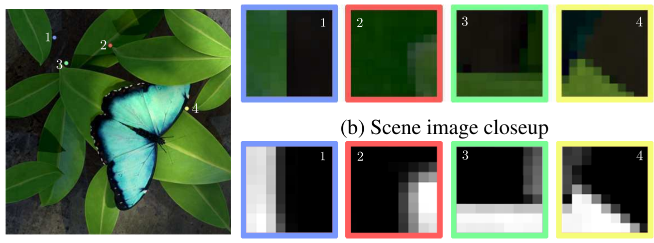
***Index Terms*—** Light Field, Stereo Matching, Depth Estimation, Guided Integral Filter, Occlusions

1. INTRODUCTION

The light field camera, *e.g.* Lytro [1] and Raytrix [2], has been developing quickly in the market recent years. Because it is able to capture the scene from each direction in one shot, the additional information in the light field reveals the structure of a scene, which allows a wide range of applications, *e.g.* super-resolution [3, 4, 5, 6], digital refocusing [7] and 3D reconstruction [8, 9, 10, 11].

As a critical step in light field applications, depth estimation from light fields has attracted much attention. However, the existing methods are always restricted by occlusions and ambiguities, which interfere with the matching process in depth calculation.

Recently, some stereo matching based methods have been proposed to handle occlusions in light field images. Yu *et al.* [12] develop a novel line-assisted graph-cut algorithm that encoded 3D line constraints into light field stereo matching. Chen *et al.* [10] introduce a bilateral consistency metric for light field stereo matching to handle significant occlusions. However, these methods usually rely on the dense sampling in angular resolution, and have matching ambiguities in textureless regions. On the other hand, taking advantage of the continuous space in angular direction, some methods based on epipolar plane image (EPI) have been developed. Wanner *et al.* [9, 6] develop a structure tensor based approach to measure each pixel's direction in EPI. These methods do not try to match pixels at different locations, hence the processing speed increases. However, the method is not effective when the occlusions occur, and the structure tensor rely on high angular resolution to guarantee the accuracy. These requirements cause limitations in light field image applications.



(a) Scene image (b) Scene image closeup (c) Angular sampling image at the right depth

Fig. 1. The Scene image and the angular sampling image of occluded points at the right depth. The angular sampling image always show the similar structure of the scene when the point is occluded by another object.

Considering the special structure in light field images, we propose that an angular sampling image (ASI) of an occluded point at the correct depth has the similar structure compared with the reference view, as shown in Fig. 1. Specifically, an occluded point must have an edge around in the reference image, and be occluded in views located in the same direction. Similar with some popular edge-preserving filters, *e.g.* bilateral filter [13] and guided filter [14], we introduce a guided integral filter to estimate the weight of each point in ASI according to the reference view. The filter is built base on the hypothesis that the points with higher contrast are more likely to occlude each other.

Therefore, we propose a novel integral filter with reference view guided for stereo matching in light field. Instead of computing bilateral metric at every possible depth based on the statistic of the angular sampling image as [10], we directly construct a weighted window according to the structure of the scene to filter the matching cost at all depth labels. In order to

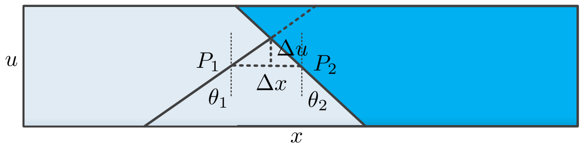


Fig. 2. Epipolar plane image when a point $P1$ is occluded by $P2$. The relative distance between the two points in the reference view is Δx, and then the point $P1$ is occluded by point $P2$ only if the perspective changes more than Δu in the same direction.

obtain a high-quality global depth map, we introduce a filter-based method which is less computationally demanding while keeps the similar performance as energy-based methods. We prove that the proposed filter kernel is able to predict the occlusions and exclude their influences in matching cost calculation. Experimental results show that our method achieves comparable performance with existing methods, both in the occlusions and ambiguous regions.

2. GUIDED INTEGRAL FILTER

In this paper, 2D plane is used to parametrize the 4D light field, where light through Point P intersects the main lens plane at (u, v) and intersects the micro lens plane plane at point (x, y). The light field is then expressed as $L(x, y, u, v)$. Specifically, the (u, v) can be expressed as the coordinate of the views, and (x, y) as the image coordinate captured in different views. For every point P in the scene, we can find the corresponding imaging point in different views. If we gather the point in each view according to the corresponding order, we can obtain another kind of image called angular sampling images, as well as the Surface Camera (SCam) [15] or the Surface Light Field [16].

As shown in [10], an unoccluded point on a Lambertian surface will produce an image of constant color, which means we can observe this point in every cameras. Nevertheless, if the point is occluded from some perspectives, the angular sampling image will included complicated textures due to the front occlusions. Its conceivable that an object that lies in one orientation of a point can only occlude the point when the views moves in the same orientation. In general, occlusion shape in the angular sampling image is similar to the point in the real scene. On this basis, a guided integral filter is built to predict the occlusions and further exclude the pixels from the occluders.

First, the size of the original filter is determined. We show the occlusion relationship using the Epipolar Plane Image (EPI) for clear observation. The EPI can be obtained by fixing the coordinate (y^*,v^*), and changing the (x, u). As shown in Fig. 2, it can be viewed as 2D slices of a constant angular view stack in the light field. The slope of the line in EPI, defined as θ, is able to reflect the depth information of

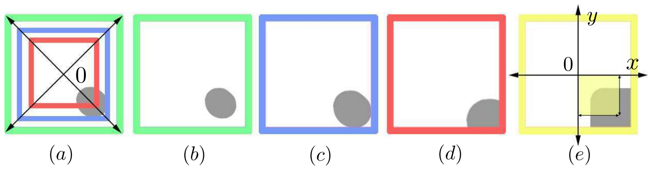


Fig. 3. The guided integral window setting. (a) The reference scene view. (b), (c), (d) The possible angular sampling images at the right depth. (e) The guided integral window. The possibilities of occlusions in the window is set to be the summarization of the region between the reference center point and the point (yellow region).

the point.

Assume that the point $P1$ is occluded by $P2$ when the perspective of the view changes horizontally, as described in EPI $I_{y,v}(x, u)$, Fig. 2. The angle of the line in EPI, indicating the depth of point P1 and P2, is θ_1 and θ_2 respectively. The distance between the two points in the center reference view is Δx. Then the distance Δu in the angular sampling plane when the points intersect can be expressed as:

$$\Delta u = \frac{\Delta x}{tan\theta_2 - tan\theta_1}. \tag{1}$$

That is to say, if point $P2$ is closer to the camera than it is to the point $P1$, and the distance between the two point in the reference view is Δx, the point $P1$ will be occluded by $P2$ when the perspective of the view moves Δu in the same direction. In other words, if Δu is fixed, the point that may occlude P1 must have the distance less than Δx. Since Δx will increase when the depth difference $tan\theta_2 - tan\theta_1$ increases, we can temporarily hypothesize that the two points are located in the farthest and nearest point separately. The maximum Δx_{max} is then equal to $\Delta u(tan\theta_{max} - tan\theta_{min})$. This means if we have no information about the true depth of the points, the point $P2$ will not occlude $P1$ if $\Delta x > \Delta x_{max}$. Under this assumption, we first build a window W_o for each point in the reference view with side length equal to $2\Delta x_{max}$. The value for pixel j in the window is set to the intensity distance with respect to the center pixel i as:

$$W^o_{i,j}(I) = |I_i - I_j|^2. \tag{2}$$

If one point in W_o has a high variance with the reference point, it will be defined as more like a occlusion. As the size of $W^o_{i,j}$ is set according to the $2\Delta x_{max}$, the actual scope will reduce if the depth distance between the two point decreases. Therefore, if we simply map the $W^o_{i,j}$ to the homologous angular sampling images, the errors still occur when the occlusion has a thin structure, as shown in Fig. 3. If the depth distance decreases, the W^o will not cover the occlusions.

In order to fit all the situations, we then alter the filter kernel using the summarization from the central point like integral image. The origin of the coordinate is set to the center

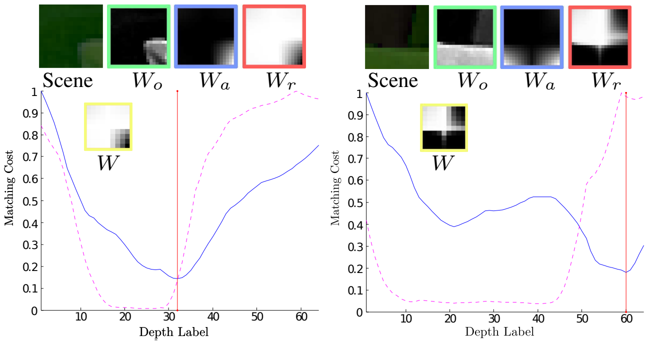


Fig. 4. The weights setting in our algorithm and the correspondence matching cost compared with the traditional cost. The matching cost can achieve the minimum value at the correct depth (solid line) compared with the traditional cost setting (dotted line).

reference point for easier understanding, and then the altered window W_a can be calculated as:

$$W_a(x, y) = \sum_{0}^{x} \sum_{0}^{y} W_o(sign(x)x, sign(y)y). \qquad (3)$$

The W_a will increase the point's occlusion possibility if the points before it (the distance between the point and the reference point is closer) is more like a occlusion. Specifically, the altered filter will extend the possible occlusion regions along the resize direction to the border as shown in Fig. 3.

The value in the filter is then be normalized and set as the possibilities of the unoccluded point in the correspondence views.

$$W_r(x, y) = e^{(-W_a(x,y)/\sigma_w^2)}, \qquad (4)$$

where σ is used to adjust the color similarity and is set to 0.1 in our experiment. The altered filter W_r is then resized to suit the angular sampling images as W, which is called guided integral filter.

3. DEPTH ESTIMATION

For depth estimation in light field image, the guided integral filter W is used in the cost volume calculation for consistency measure. Specifically, for a pixel i in the reference view I_i, and the pixel j in the angular sampling image respect to depth d, the cost volume is defined as:

$$C_{i,d} = \sum_{j} W_{i,j}(I)\rho(ASI_{j,d}, I_i), \qquad (5)$$

where ρ is a distance function, and here we compute as $\rho(a, b) = \frac{1}{K} exp(-\frac{(a-b)^2}{\sigma_c^2})$. The proposed guided integral filter and matching cost for occluded points are shown in Fig. 4. We can observe that the filters are able to detect the occlusions correctly, and make the cost achieve the minimum value at the correct depth.

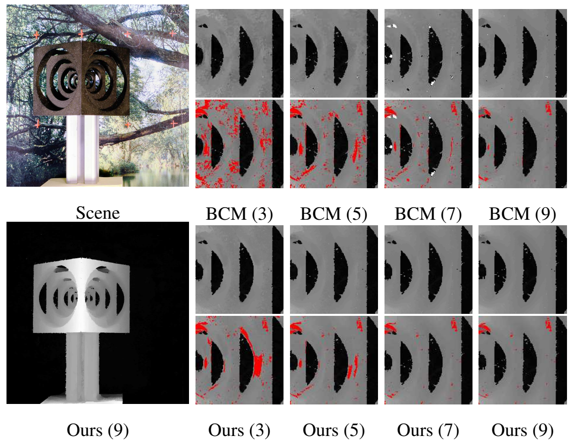


Fig. 5. Depth estimation from image "Buddha" with different angular resolution and the error pixels are labels red for distinct comparison. The proposed method achieves high accuracy even at the lower angular resolution.

After calculating the final matching cost for each point, we choose to filter the cost volume as

$$C'_{i,d} = \sum_{j} W^g_{i,j}(I)C_{i,d}, \qquad (6)$$

where W^g is the guided weight as [17]. Instead of the graph based optimazation, the filter-based method is more efficient and can be easily parallelized. Finally, we assign the global minimum value as the correct depth lable :

$$D_i = arg \max_{d} C'_{i,d}. \qquad (7)$$

4. EXPERIMENT

In this paper, we evaluate the method on light field images with available ground truth [18], and compare with stereo-based methods [10] (BCM) and tensor structure based methods [9] (GCLD), which have shown good performance on the ideal light field images recently. As a quality measurement, we use the percentage of depth value below a relative error based on the ground truth as [9]. The accuracy is calculated for the overall images (All) and occlusion regions (Occ) respectively. The optimal parameter is then found by testing a number of different parameter on different data sets. In our implementation, σ_w is set 0.1, and σ_c is set 100.

We first test the proposed method on different angular resolution. The numerical results, compared with the state-of-art algorithms are illustrated in Table 1. It should be noted that the other methods are not effective if they use a small set of views (less than 9 views in each direction). Because the metric [10] used to estimate the occlusions is calculated on the angular sampling image for every possible depth label, which is not reliable if the angular resolution is too small. By contrast, our method rely on the reference view image to predict

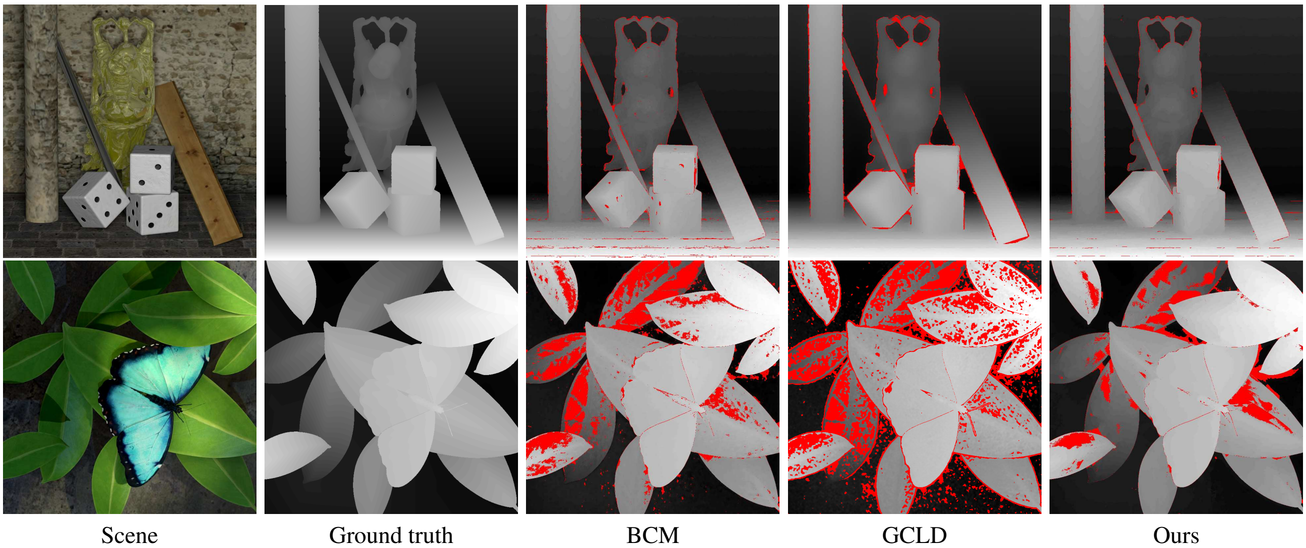


Fig. 6. Depth estimation from images “Buddha” and “Papillon” and the error pixels are labels red for distinct comparison.

Image	GCLD [9]		BCM [10]		Ours	
	All	Occ	All	Occ	All	Occ
Buddha(9)	2.41	15.01	1.72	8.31	**1.21**	**6.25**
Buddha(7)	2.58	15.68	2.14	8.34	**1.35**	**6.58**
Buddha(5)	3.72	22.72	2.54	9.11	**1.72**	**7.78**
Buddha(3)	-	-	11.37	19.43	**4.54**	**15.50**
Cube(9)	0.92	13.29	1.11	9.72	**0.83**	**8.36**
Maria(9)	**2.07**	12.07	2.67	**10.56**	2.24	11.87
StillLife(9)	4.30	8.35	**1.51**	6.53	1.62	**6.23**
Papillon(9)	21.24	34.44	12.86	12.83	**7.31**	**12.55**

Table 1. The error rate of the estimated depth compared with ground truth. (%)

the occlusions, which is independent of the angular resolution.

The visual results, shown in Fig. 5, verify that we can acquire the depth map with clear edges even with 5 views in each direction. This quality is quite important for real light field images captured by plenoptic cameras, which have fewer available views for depth estimation because of the hardware restrictions.

Then the results for light field images with 9×9 views are illustrated in Table 1. We can observe that compared with Wanner *et al.* [9], our results show better results on occlusion regions. For some images with occlusions and ambiguities, we achieve the higher accuracy than both algorithms. Chen *et al.* [10], which are designed to handle heavy occlusions in light field images, calculate the occlusion possibilities of the angular sampling images based on the statistics of the angular sampling images themselves. This method exclude much useful information when calculating the matching cost for textureless regions. By contrast, we calculate the occlusion possibilities according to the reference view, which is able to handle both ambiguities and occlusions.

Fig. 6 compares results for light filed image “Buddha” and “Papillon” (9×9 views) with ground truth. These images have lots of ambiguous regions along with many occlusions. For these kinds of images, the proposed method is able to handle both the occlusion and ambiguous problem, and acquire a higher accuracy.

5. CONCLUSION

Taking into account the special structure of the light field data, we propose a novel guided integral filter to handle occlusions based on the view image guided. The defined window is used to predict the occlusions in the angular sampling images, and exclude the incorrect point for matching cost calculation. The problem is formulated as a discrete labeling problem and solved using a filter-based algorithm. Compared with the state-of-the-art stereo matching methods and tensor structure based methods, the proposed method is able to handle occlusion problems as well as ambiguities. Moreover, the algorithm is less affected by the angular resolution which means it has a wide range of applications.

6. ACKNOWLEDGEMENTS

This study was partially supported by the National Natural Science Foundation of China (No. 61370122), the National High Technology Research and Development Program of China (No. 2013AA01A603), the National Aerospace Science Foundation of China (No.2013ZC51), the Programme of Introducing Talents of Discipline to Universities and the Open Fund of the State Key Laboratory of Software Development Environment under grant #SKLSDE-2015ZX-21.

7. REFERENCES

[1] Ren Ng, "Lytro redefines photography with light field cameras," http://www.lytro.com.

[2] Christian Perwa and Lennart Wietzke, "The next generation of photography," http://www.raytrix.de.

[3] Tom E Bishop, Sara Zanetti, and Paolo Favaro, "Light field superresolution," in *Computational Photography (ICCP), 2009 IEEE International Conference on*. IEEE, 2009, pp. 1–9.

[4] Tom E Bishop and Paolo Favaro, "The light field camera: Extended depth of field, aliasing, and superresolution," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 5, pp. 972–986, 2012.

[5] Sven Wanner and Bastian Goldluecke, "Spatial and angular variational super-resolution of 4d light fields," in *Computer Vision–ECCV 2012*, pp. 608–621. Springer, 2012.

[6] Sven Wanner, Christoph Straehle, and Bastian Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, 2014.

[7] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR*, vol. 2, no. 11, 2005.

[8] Edward H Adelson and John YA Wang, "Single lens stereo with a plenoptic camera," *IEEE transactions on pattern analysis and machine intelligence*, vol. 14, no. 2, pp. 99–106, 1992.

[9] Sven Wanner and Bastian Goldluecke, "Globally consistent depth labeling of 4d light fields," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012, pp. 41–48.

[10] Can Chen, Haiting Lin, Zhan Yu, Sing Bing Kang, and Jingyi Yu, "Light field stereo matching using bilateral statistics of surface cameras," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2014, pp. 1518–1525.

[11] Michael W Tao, Sunil Hadap, Jitendra Malik, and Ravi Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *IEEE International Conference on Computer Vision (ICCV)*, 2013, pp. 673–680.

[12] Zhan Yu, Xinqing Guo, Haibing Ling, Andrew Lumsdaine, and Jingyi Yu, "Line assisted light field triangulation and stereo matching," in *IEEE International Conference on Computer Vision (ICCV)*, 2013, pp. 2792–2799.

[13] Carlo Tomasi and Roberto Manduchi, "Bilateral filtering for gray and color images," in *IEEE International Conference on Computer Vision (ICCV)*. IEEE, 1998, pp. 839–846.

[14] Kaiming He, Jian Sun, and Xiaoou Tang, "Guided image filtering," in *European Conference on Computer Vision (ECCV)*, pp. 1–14. Springer, 2010.

[15] Jingyi Yu, Leonard McMillan, and Steven Gortler, "Surface camera (scam) light field rendering," *International Journal of Image and Graphics*, vol. 4, no. 04, pp. 605–625, 2004.

[16] Daniel N Wood, Daniel I Azuma, Ken Aldinger, Brian Curless, Tom Duchamp, David H Salesin, and Werner Stuetzle, "Surface light fields for 3d photography," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 2000, pp. 287–296.

[17] Christoph Rhemann, Asmaa Hosni, Michael Bleyer, Carsten Rother, and Margrit Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 3017–3024.

[18] Sven Wanner, Stephan Meister, and Bastian Goldluecke, "Datasets and benchmarks for densely sampled 4d light fields," in *Vision, Modeling & Visualization*. The Eurographics Association, 2013, pp. 225–226.