Depth Extraction From A Light Field Camera Using Weighted Median Filtering

Changtian Sun and Gangshan Wu*

State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China
chantesun@gmail.com,gswu@nju.edu.cn

Abstract. The invention of light field camera provides an approach of extracting depth from a collection of refocused photos. However, current depth from focus methods are suffering from the problem of inaccuracy, especially in the regions of object boundaries. In this paper, we propose a novel method to extract an accurate depth map from several refocused images. We first render a collection of images by uniformly changing the focal length. Then we deduce a rough depth map according to the amount of blur, which is measured by a multi-scale gradient operator. Finally, we apply a weighted median filter for refinement, which suppresses depth noise and supplies a well recovery of object boundaries and fine structures. The experimental results show that our method outperforms the built-in method from Lytro light field camera.

Keywords: Depth extraction, light field, weighted median filtering

1 Introduction

Depth from focus aims to recover the depth information from a set of refocused images [1]. Given a few images with different focusing parameters, we can calculate the implicit depth information. And this cast a light on the recovery of explicit depth information from the implicit information. It has been widely utilized in numerous multimedia applications, such as image-based rendering [2] and photo editing [3].

Many previous works capture different focused images sequentially, which are limited to static scenes. To illustrate, two images are acquired by varying the intrinsic parameters of the camera in [4] and two observations are obtained by two sets of lens parameters in [5]. Nowadays, the current light field camera [6] can capture 4D light field images in one shot, which enables rendering of arbitrary focused images of the same scene. Light field camera benefits various multimedia research [7,8], including depth extraction [9,10]. However, the efficiency or accuracy are still far from desired. Most previous studies involving depth recovery pay little attention to the refinement of depth maps, leaving out defects of inaccurate boundaries between different depths, wavy edges or noises in the depth maps.

In this paper, we propose a novel approach for depth from focus, which is featured for its effectiveness. We observed that depth images are full of smooth areas. Consequently, patch based depth inferring is far more robust than pixel-wise calculation. Though the patch based calculation will lead to flattening effects, it can be corrected by current joint edge-preserving filtering. Specially, we first obtain a rough depth map from a set of images refocused to different degrees, which are acquired by a light field camera in a single shot. In the camera, when the distance from the center of the lens and the image plane is viewed as a constant, the variance in focal length brings about different degree of blurring. When the blurring reaches its smallest, say, all the light rays radiated by an object point and refracted by the lens converges at a single point on the image plane, the object distance can be deduced from the focal length and the distance from the image plane to the lens center. Next, a multi-scale gradient operator and small windows are employed to work out the gradient of each pixel in images with distinct camera parameters in order to achieve higher accuracy and to get rid of the impacts of noise. However, these methods may give rise to vague boundaries or wavy edges which do not cater to the original input light field image. Therefore, at last, we need a filtering step under the guidance of the original image. Our work adapts constant time weighted median filtering [11] as a way of refining the rough depth map.

With the combination of gradient analysis of the input image and a refinement step guided by the original image, we can assure an accurate and reliable refined depth map.

Our approach outperforms the built-in depth map of the software, Lytro Desktop [12] in the way that it can recover more of the original outlines of objects. Moreover, the measure we take, constant time weighted median filtering [11], works better than the simple guided filtering [13] on the implicit depth map as it pays more attention to the actual depth of field than the outlines of objects.

2 Related Work

A number of current works devote to the recovery of depth in the images. However, nearly each of them has several defects.

In [14], a spatial constant σ needs to be evaluated. It needs two images that are identical except for the aperture size and therefore depth of field. Two σs can be derived from them and we can thereby work out the depth map. By contrast, our method works in a light field camera and images with different focal lengths can be gained. As a consequence, the parameter σ needn't be evaluated and large amount of time can be saved. Apart from this, our method does not possess the "overconstraint" problem that Alex's work has: if three or more views are obtained in the process, ambiguity of depth will be caused.

In a real-time focus range sensor [15] , a passive way of evaluating the depth was adopted: an illumination pattern is projected onto the scene via the same optical path used to image the scene. This leads to a dominant frequency in the images gained and a depth map can be acquired by two images of the same view

with different optical settings. Meanwhile, in [16], active measures are taken to get the depth map from the two images with distinct camera settings. However, the depth map derived from either of them is not precise and filters to refine the map need to be applied.

For passive depth from defocus, rational filters [17] can be adopted. In [17], median filtering was used in one example and *adaptive coefficient smoothing* in another. However, as we will point out, there are some defects in the simple median filtering. In addition, *adaptive coefficient smoothing* is only suitable for an image with background which lacks texture. Our method has overcome this disadvantage.

In our work, each of the series of images will be put to good use and a fine approach to filter the depth map will be utilized.

3 Approach

In this section, the input light field image is broken up into numbers of images and n images with different focal lengths are drawn from them uniformly, where n is a parameter given by hand according to our demand on the precision of the depth map. The gradient images are derived from the n differently refocused images using a multi-scale gradient operator. A rough depth map is acquired analyzing the gradient images. And a refined depth map is gained by filtering the rough one, as shown in Fig. 1.

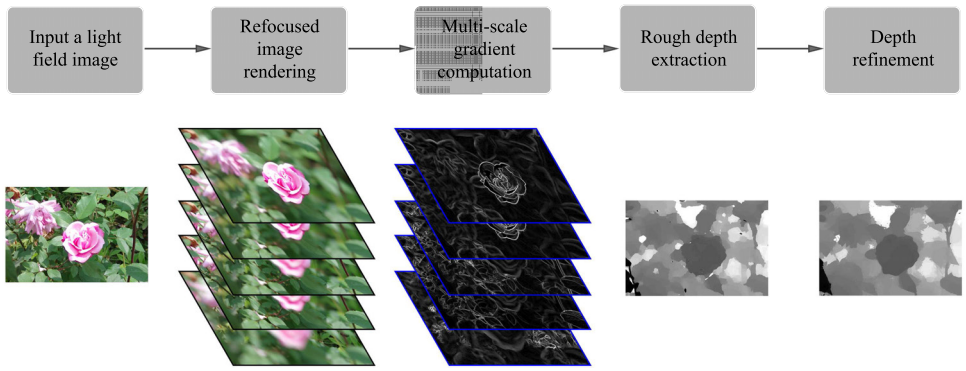


Fig. 1. A framework of our method.

3.1 Images with Different Focal Length

When the light rays radiated by an object point and refracted by the lens converge at a single point on the image plane, the following equation is satisfied:

gradient (far focus) gradient (medium focus) gradient (near focus)

Fig. 2. The gradient maps with different focal lengths.

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v}, \tag{1}$$

where f refers to the focal length, u refers to the object distance, and v represents the image distance.

Suppose an image plane is placed at a distance of v from the lens. Any object point placed behind the focus on the other side of the lens, which does not possess a distance of u from the lens, will cast a circle of confusion on the image plane. Whether the object is placed behind or in front of the intended distance u, a blurring of image can not be avoided.

In light field cameras, images with different camera parameters can be acquired from a single shot. What we need is a change in the focus to utilize the Eq. (1) to help us recover the field depth.

In practice, we need a uniform change in focal length, either from far to near or from near to far, so as to gain convenience in the processing procedure. In this paper, n refocused images can be gained from the input light field image with uniform change in focal length from far to near, with the help of the matching photo processing software of the light field camera.

In addition, an all-clear image with the largest f-number (the ratio of the focal length to the diameter of the aperture) is also available by the camera.

3.2 Acquisition of Rough Depth

A gradient map for each of the n images need to be worked out in the first place. A multi-scale gradient operator as shown in Eq. (2) and (3) will work. It is superior to a simple Sobel or Scharr operator as the former ones may ignore the effects of smooth edges and are prone to noise.

$$P_{l,k}(x, y) = A_k(2^l x, 2^l y), \tag{2}$$

$$G_k(x,y) = \sum_{l=0}^{m} Q_{l,k}(\frac{x}{2^l}, \frac{y}{2^l}), \tag{3}$$

where l refers to the lth layer of the Gaussian Pyramid, whose maximum m can be adapted according to the characteristics of the input image; k refers to the index of the image picked and A_k refers to the pixel values of the kth image; $P_{l,k}$ is an image resizing A_k to its $\frac{1}{2^l}$. Being the lth layer of the Guassian Pyramid; $Q_{l,k}$ is $P_{l,k}$ processed with the Sobel operator, and all $Q_{l,k}$s add up to the final gradient image G_k of A_k.

As shown in Fig. 2, the gradient varies as the focal length changes.

In order to raise precision and eliminate the effects of noise, a window at a certain width can be used to analyze the gradient of the very pixel at its center, as shown in Eq. (4). In this equation, $w(\boldsymbol{x})$ indicates the window at whose center is $\boldsymbol{x}$. This also helps us reduce the impact of small trembles when shooting the series of images, which leads to dislocation in some of the n images.

$$\widetilde{G}_k(\boldsymbol{x}) = \sum_{\boldsymbol{y} \in w(\boldsymbol{x})} G_k(\boldsymbol{y}). \tag{4}$$

In practice, integral image [18] , which is also known as summed area table, is used in order to speed up the processing procedure. What's more, the edges of the images need to be paid attention to, and a mirror reflection of pixels on the edge tends to be a good solution.

A 25×25 or 15×15 window is usually used according to the objects' structures which make up the image. Small windows are appropriate for images with fine structures.

In this way, an optimized gradient value $\widetilde{G}_k(\boldsymbol{x})$ is gained for each pixel according to the values of its neighbors. For the pixels with the same coordinate in the n images, the maximum optimized gradient indicates that the pixel is at its sharpest, leading to a rough estimation of the optimal focus length of the very pixel.

Suppose that a pixel reaches its sharpest at the image indexed k, where

$$k = \underset{k}{\operatorname{argmax}} \widetilde{G}_k(\boldsymbol{x}), k \in \mathbb{N}^+, k \leqslant n. \tag{5}$$

The largest focal length being $\max(f)$ and the smallest being $\min(f)$ in all n images, the optimal focal length of the pixel can be roughly represented as the following:

$$f_o = \frac{n-k+1}{n}(\max(f) - \min(f)) + \min(f). \tag{6}$$

Therefore, according to Eq. (1), the rough depth of the very pixel can be estimated as:

$$\widetilde{u} = \frac{f_o v}{v - f_o}. \tag{7}$$

where v is a known camera parameter.

input light field image rough depth map refined depth map

Fig. 3. Rough depth map refinement. The refined depth map in the third column is gained from the second column by constant time weighted median filtering [11] under the guidance of the image with largest f-number.

3.3 Attainment of Accurate Depth

Median filtering needs to be used so as to remove outlier noise. However, a simple median filter may result in the loss of sharp features of an image. Consequently, in order for suppressing the influence of nearby pixels in different colors along with preserving sharp edges and fine structures, the method of weighted median filtering is used:

$$f(\boldsymbol{x}, i) \triangleq \delta(V(\boldsymbol{x}) - i), \tag{8}$$

$$h(\boldsymbol{x}, i) = \sum_{\boldsymbol{x}' \in \mathcal{N}(\boldsymbol{x})} b(\boldsymbol{x}, \boldsymbol{x}') f(\boldsymbol{x}', i), \tag{9}$$

where $\boldsymbol{x}$ represents the pixel's coordinate; $V(\boldsymbol{x})$ represent its value in the rough depth map; V equals to $\widetilde{u}$ in Eq. (7); δ is the delta function which equals to 0 when $V(\boldsymbol{x}) = i$ and 1 otherwise.

Meanwhile, $h(\boldsymbol{x}, i)$ can be illustrated as a local histogram and $b(\boldsymbol{x}, \boldsymbol{x}')$ is the weight function. In order for the feature of the input image to be preserved, we can use guided filter weights for $b(\boldsymbol{x}, \boldsymbol{x}')$.

In a guided filter [13] , the local linear model is represented as:

$$\mathbf{a}_k = (\Sigma_k + \epsilon U)^{-1} (\frac{1}{|w|} \sum_{i \in w_k} \mathbf{I}_i p_i - \mu_k \bar{p}_k), \tag{10}$$

$$b_k = \bar{p}_k - \mathbf{a}_k^T \mu_k, \tag{11}$$

$$q_i = \bar{\mathbf{a}}_i^T \mathbf{I}_i + \bar{b}_i, \tag{12}$$

where $\mathbf{a}_k$ is a 3×1 coefficient vector, q is a linear transformation of $\mathbf{I}$ in the window w_k centered at the pixel k, Σ_k is a 3×3 covariance matrix of $\mathbf{I}$ and U is a 3×3 identity matrix.

In Fig. 3, a refined depth map is gained from a rough one under the guidance of the image with the largest f-number. Compared with the rough depth map, noises have been removed and the outlines of objects tend to be more accurate. Additionally, the depth map of the background has a tendency to be softer, which conforms to our real life perception.

4 Experiments

4.1 Experimental Settings

In our experiments, $n = 64$ is used for drawing images at different focal lengths from the input light field image. A 25×25 window and $m = 2$ is used for calculating the optimized gradient of images with few fine structures. In the mean time, 15×15 windows and $m = 0$ is used for images with fine structures. Windows too small may lead to extra black edges or white holes for objects while windows too large may result in blocks of colors.

As we want to make full use of the 8 digits of the grayscale images for a depth map, we apply [0,1,2,...,255] for the vector of disparities. However, if the calculating process were too time-consuming on a certain machine, especially for images with big sizes, the number of disparities could be reduced. A little sacrifice on the diversity of gray scale of the depth map can save a great amount of time. In addition, the local window radius for guided filter weights is usually set to $\frac{1}{40}$ of the maximum of the input image's width and height. The regularization parameter is $\epsilon = 10^{-4}$.

4.2 Results and Discussions

Our depth map has proven to correspond more with the original image. It can be seen from Fig. 4, the built-in depth maps of Lytro Desktop tend to have wavy edges while constant time weighted median filtering [11] helps us smooth the structures. On the other hand, the outlines of objects in the built-in depth map is not so accurate as ours. Our method can show fine structures which appear in the original image.

The use of constant time weighted median filtering [11] also helps us erase some white or black blocks in the rough depth maps caused by the ambiguity of gradient change of small areas in the input light field image, just as shown in Fig. 5. The lack of texture results in this ambiguity, and a smooth surface or highlights on objects may accounts for the lack of texture.

In order for a better result, images with high variance in focal length is preferred. This contributes to the n images having greater discrimination in gradient with each other. In practice, objects which are close enough to the camera lens and objects which are far away enough in the input images are preferred so as to achieve high variance in focal length.

Compared to the results of the simple guided image filtering [13], this weighted median filtering [11] using guided filter weights pays more attention to the actual depths of the original image. As is shown in Fig. 6, the former method may produce halo effects and pays too much to the structure of the objects as well as the lighting condition on the surface of the objects. However, the latter method only reflects the depth information we need, which leads to accuracy.

There are still some limitations in our method. For objects with large textureless areas like the sky, our method will fail to predict the accurate depth for the lack of gradients. Fortunately, textureless regions can be easily detected to prevent errors in applications.

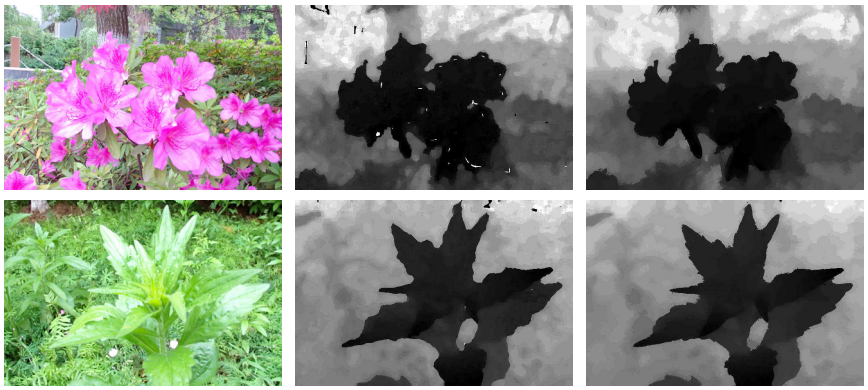
Fig. 4. A comparison between the Lytro Desktop's built-in depth map and our depth map. The first two light field images are available at http://lightfield-forum.com/.

5 Conclusion

In this paper, we have presented a novel depth from focus method from a series of images using a light field camera. By employing the multi-scale gradient operator, we achieved a robust measurement for the amount of blur. Besides, by using the constant time weighted median filtering, accurate depth maps with clear boundaries are obtained. The experiments have proven that our measure surpasses the built-in measure in the Lytro light field camera.

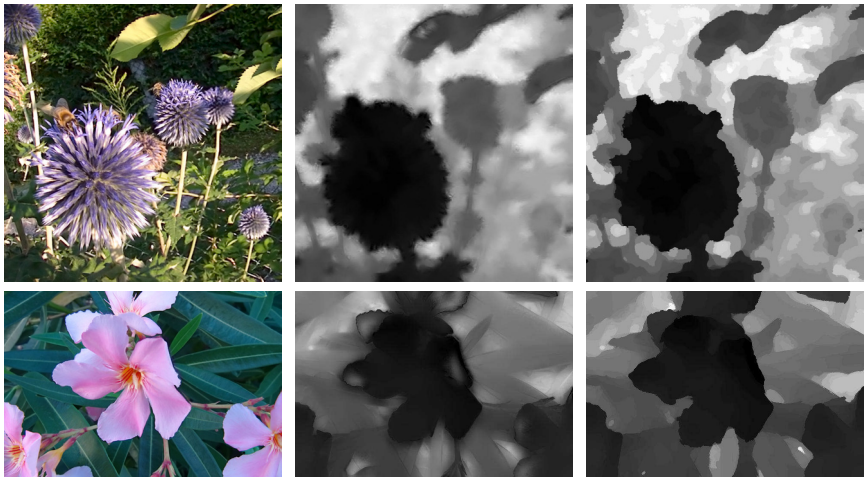
Acknowledgments

This work is supported by National Science Foundation of China (61321491) and Collaborative Innovation Center of Novel Software Technology and Industrialization.



input light field image rough depth map refined depth map

Fig. 5. Constant time weighted median filtering [11] removes black and white blocks caused by the ambiguity in the gradient change.



input light field image guided image filtering weighted median filtering

Fig. 6. A comparison between guided image filtering [13] and constant time weighted median filtering [11]. The light field images are available at http://lightfield-forum.com/.

References

1. Paul Grossmann. Depth from focus. Pattern Recognition Letters 5(1) (1987) 63-69
2. Heung-Yeung Shum, Shing-Chow Chan, and Sing Bing Kang. Image-based rendering. Springer Science & Business Media (2008)
3. Byong Mok Oh, Max Chen, Julie Dorsey, and Frdo Durand. Image-based modeling and photo editing. Annual Conference on Computer Graphics and Interactive Techniques, ACM (2001) 433-442

4. Djemel Ziou, and Francois Deschenes. Depth from defocus estimation in spatial domain. Computer Vision and Image Understanding 81(2) (2001) 143-165
5. Vinay P. Namboodiri, and Subhasis Chaudhuri. On defocus, diffusion and depth estimation. Pattern Recognition Letters 28(3) (2007) 311-319
6. Ren Ng, Marc Levoy, Mathieu Brdif, Gene Duval, Mark Horowitz, and Pat Hanrahan. Light field photography with a hand-held plenoptic camera. Computer Science Technical Report CSTR 2, 11 (2005) 1-11
7. Xudong Zhang, Yi Wang, Jun Zhang, Liangmei Hu, and Meng Wang. Light field saliency vs. 2D saliency: A comparative study. Neurocomputing 166 (2015) 389-396
8. Jun Zhang, Meng Wang, Jun Gao, Yi Wang, Xudong Zhang, and Xindong Wu. Saliency detection with a deeper investigation of light field. International Joint Conference on Artificial Intelligence (2015) 2212-2218
9. Changyin Zhou, Stephen Lin, and Shree K. Nayar. Coded aperture pairs for depth from defocus and defocus deblurring. International Journal of Computer Vision 93(1) (2011) 53-72
10. Michael Tao, Sunil Hadap, Jitendra Malik, and Ravi Ramamoorthi. Depth from combining defocus and correspondence using light-field cameras. IEEE International Conference on Computer Vision (2013) 673-680
11. Ziyang Ma, Kaiming He, Yichen Wei, Jian Sun, and Enhua Wu. Constant time weighted median filtering for stereo matching and beyond. IEEE International Conference on Computer Vision (2013) 49-56
12. Lytro Desktop `https://illum.lytro.com`
13. Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. European Conference on Computer Vision (2010) 1-14
14. Alex Paul Pentland. A new sense for depth of field. IEEE Transactions on Pattern Analysis and Machine Intelligence 4 (1987) 523-531
15. Shree K. Nayar, Masahiro Watanabe, and Minori Noguchi. Real-time focus range sensor. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(12) (1996) 1186-1198
16. Subhasis Chaudhuri, and Ambasamudram N. Rajagopalan. Depth from defocus: a real aperture imaging approach. Springer Science & Business Media (2012)
17. Masahiro Watanabe, and Shree K. Nayar. Rational filters for passive depth from defocus. International Journal of Computer Vision 27(3) (1998) 203-225.
18. John P Lewis. Fast template matching. Vision Interface, 95(120123) (1995) 15-19