applied optics

Large aperture focus stacking with max-gradient flow by anchored rolling filtering

Xuanwu Yin,[1,2] Guijin Wang,[1,*] Wentao Li,[1] and Qingmin Liao[2]

[1]Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[2]Department of Electronic Engineering, Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China
*Corresponding author: wangguijin@tsinghua.edu.cn

Received 22 April 2016; revised 5 June 2016; accepted 6 June 2016; posted 6 June 2016 (Doc. ID 263828); published 5 July 2016

Focus stacking is a computational technique to extend the depth of field through combining multiple images taken at various focus distances. However, in the large aperture case, there are always defects caused by the large blur scale, which, to the best of our knowledge, has not been well studied. In our work, we propose a max-gradient flow-based method to reduce artifacts and obtain a high-quality all-in-focus image by anchored rolling filtering. First, we define a max-gradient flow to describe the gradient propagation in the stack. The points are divided into trivial and source points with this flow. The source points are extracted as true edge points and are utilized as anchors to refine the depth map and the composited all-in-focus image iteratively. The experiments show that our method can effectively suppress the incorrect depth estimations and give a high-quality all-in-focus image.

OCIS codes: (100.3020) Image reconstruction-restoration; (110.3010) Image reconstruction techniques; (110.4155) Multiframe image processing.

http://dx.doi.org/10.1364/AO.55.005304

1. INTRODUCTION

Optical lenses can only focus on a specific scene surface (usually designed as a plane), leaving other regions of the scene subject to various scales of defocus. Due to the limitation of depth of field (DOF), it is often difficult to keep the whole content of a single image in focus. One way to extend DOF is by decreasing the aperture size; however, this yields a lower signal-to-noise ratio (SNR) and requires longer exposure time. With the development of digital imaging technology, a multifocus image fusion technique, known as focus stacking, has become popular [1,2]. This technique captures a sequence of images focusing at various planes and fuses these images into a single all-in-focus image. The stack is captured by moving the imaging sensor along the optical axis, as shown in Fig. 1. We capture a sample stack with an Imperx B4020 mono camera equipped with a SIGGMA 50 mm/F1.4 lens. This stack consists of 14 images and is utilized to describe our method in the rest of this work. Three of the 14 images are presented in Fig. 2. We also create a 3D sketch to show the optical geometry of our test scene.

The focus stacking technique has attracted a lot of attention in the last decade. Forster *et al.* [3] proposed a complex wavelet method to extend the DOF of microscopy images. A model-based 2.5D deconvolution method, which estimated the depth map and the texture simultaneously, was developed by Aguet *et al.* [4]. Tian *et al.* [5] fused multifocus images by a bilateral gradient-based criterion. Sroubek and co-workers [6–8] decomposed the images with discrete wavelet transform (DWT) and fused them in the coefficient domain. Liu *et al.* [9] utilized dense scale invariant feature transform (DSIFT) as the activity level measurement to fuse multifocus images. Haghighat *et al.* [10,11] presented a fusion approach based on variance calculated in discrete cosine transform (DCT) domain. An alternative technique, called focal sweep imaging (FSI), was proposed by Kuthirummal *et al.* [12] to extend DOF. In their work, the sensor moved along the optical axis during one exposure. The authors proved that the finally captured image was approximately a depth-invariant blurred version of the sharp image. Thus, the all-in-focus image could be restored through deconvolution with a depth-invariant kernel.

This paper focuses on focus stacking in the large aperture case. In this case, the edge will propagate the gradient in the scale of the blur kernel, making the texture behave quite different from the small aperture case. This phenomenon will yield ghost edges in the result. The frequency-domain method proposed by Alonso and co-workers [13,14] could partly overcome this problem. In their work, each frequency component was reconstructed by solving a linear equation based on a plane-wise approximation of the 3D scene. However, these frequency-domain equations will become highly ill-posed when the number of images increases; thus, the result will be unstable.

In this paper, a spatial-domain method is proposed for focus stacking in the large aperture case. We first define a

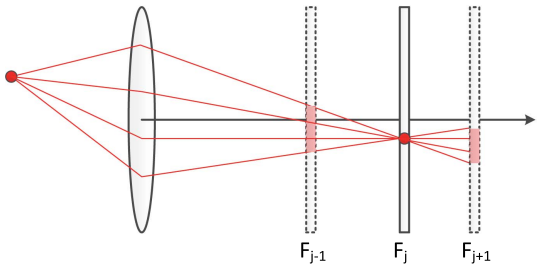

Fig. 1. Focal stack geometry; the spot is in focus at plane F_j and out of focus with various scales at plane F_{j-1} and F_{j+1}.

max-gradient flow to model the propagation of the gradient in the stack, and utilize this flow to extract the true edges from the stack. We formulate an anchored rolling filtering approach to refine the depth map and the all-in-focus image iteratively. The experiments show that our method gives a high-quality all-in-focus image and outperforms the compared state-of-the-art methods.

2. TRUE EDGE EXTRACTION WITH MAX-GRADIENT FLOW

In this section, we first define max-gradient flow to model the propagation of the gradient in the stack. With this definition, the points are divided into source points and trivial points to select the true edges in the depth map.

A. Max-Gradient Flow

In this work, the magnitude of a gradient is taken as the sharpness measure, which is defined as

$$G_i(x,y) = |\nabla I_i(x,y)| = \sqrt{\left(\frac{\partial I_i}{\partial x}\right)^2 + \left(\frac{\partial I_i}{\partial y}\right)^2}, \quad (1)$$

where I_i is the i-th image in the stack, and G_i the gradient of I_i. Once the gradients of all the images are calculated, the depth map can be initialized by

$$D(x,y) = \arg\max_i G_i(x,y), \quad (2)$$

in which the element of D stores the index that gives the maximum gradient of each pixel. This depth map is also known as a focus map. Due to the large blur kernel in the large aperture case, the edges will spread in a large scale. Thus, there may occur many ghost edges in the initial depth map.

To discuss the problem in detail, we present a sample patch in Fig. 2 (close-up views in Fig. 3). This patch contains strong and weak edges. The edges in the patch are blurry in Image 2 [Fig. 3(a)] and sharp in Image 11 [Fig. 3(b)]. The gradients of three typical pixels (red, green, blue) are presented in Figs. 3(c)–3(e), in which the vertical axis represents the gradient value of each pixel, the horizontal axis represents the index of the image in the stack. By observing Fig. 3(a), we can find that the red pixel locates just on the strong edge, the green pixel is just within the blur range of the edge, and the blue pixel is just outside the blur range. From Fig. 3(b), we can infer that the green pixel is in the same plane with the blue one. However, it can be seen in Fig. 3(d) that there exists a ghost edge in Image 2, whose gradient is even greater than that of the true edge in Image 11. This is because the edge will propagate its gradient along its normal direction when it is blurred. The gradients of these ghost edges are actually meaningless.

We define the 2D max-gradient flow as

$$\mathrm{MGF}(x,y) = [f_x(x,y), f_y(x,y)]^T, \quad (3)$$

the elements of it being

$$\begin{bmatrix} f_x(x,y) \\ f_y(x,y) \end{bmatrix} = -\begin{bmatrix} \frac{\max_j G_j(x+\Delta x,y) - \max_i G_i(x,y)}{\Delta x} \\ \frac{\max_k G_k(x,y+\Delta y) - \max_i G_i(x,y)}{\Delta y} \end{bmatrix}. \quad (4)$$

The flow describes the change of maximum gradient along each axis, i.e., the propagation of gradient in the stack. The direction along which the max-gradient propagates can be easily derived as

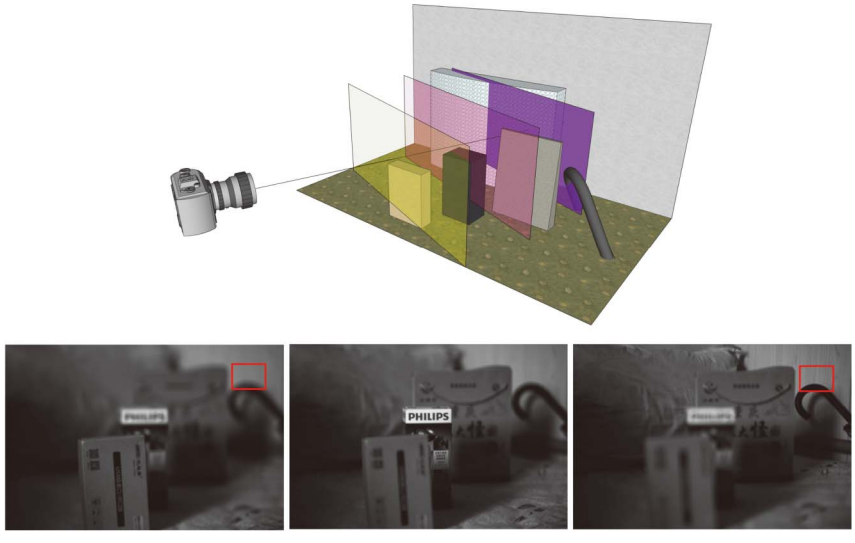

Image 2 Image 6 Image 11

Fig. 2. Top row is the 3D sketch of our test scene with three focusing planes. Bottom row shows three images focused at yellow, pink, and purple focusing planes, respectively. Red rectangle is the sample patch to demonstrate the max-gradient flow.

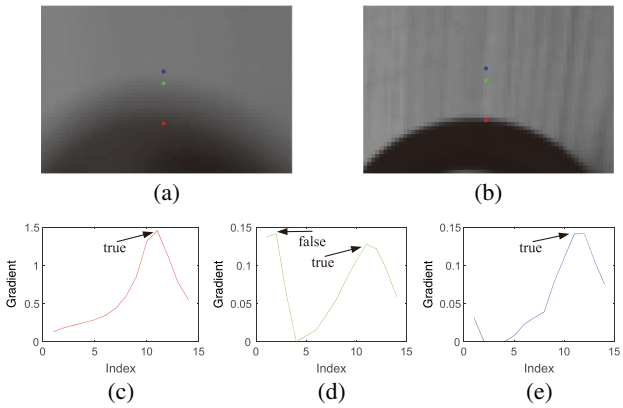

Fig. 3. (a,b) Close-up views of the sample patch of Image 2 and Image 11. Red pixel locates just on the strong edge. Green pixel locates within the blur range of the edge. Blue pixel locates outside the bur range. (c)–(e) Demonstration of the gradient changing of red, green, and blue pixel with respect to the image number, i.e., the index of the image in the stack. Note the difference of y axis limit between the plots.

$$\theta(x,y) = \arctan\left(\frac{f_y(x,y)}{f_x(x,y)}\right). \tag{5}$$

B. True Edge Extraction

With the max-gradient flow, we define two kinds of points in the focal stack: trivial point and source point. Without loss of generality, assume that the main direction of the flow at pixel (x,y) is along the x axis, the point (x,y) is defined as a trivial point if

$$f_x(x+\Delta x,y)f_x(x-\Delta x,y) > 0 \quad \text{or} \quad \max_i G_i(x,y) < G_{\text{TH}}, \tag{6}$$

which means that the direction of the flow does not change at (x,y) or that the maximum gradient at (x,y) is lower than a threshold. The threshold, G_{TH}, is introduced to remove the wrong edge detections due to noise.

Otherwise, the point (x,y) is defined as a source point if

$$f_x(x+\Delta x,y) > 0 \quad \text{and} \quad f_x(x-\Delta x,y) < 0, \tag{7}$$

which means that it is the source where the max-gradient starts propagating. The max-gradient flow of the sample patch is presented in Fig. 4. From the demonstration, we can easily derive that the red and blue pixel are source points, and the green pixel is a trivial point. All the source points form a sparse set, denoted by S, which represents the true edges of the scene. Figure 5 demonstrates the extracted true edge points of the scene. It can be seen that the ghost edges are effectively suppressed, and the true edges are preserved as much as possible.

3. ANCHORED ROLLING FILTERING

In this section, we formulate an anchored rolling filtering approach to refine the depth map and the all-in-focus image with the extracted true edges.

In the focus stacking problem, the depth map and the all-in-focus image are closely related. On one hand, suppose that the depth map D is perfect; the all-in-focus image can be estimated by extracting the pixels from the focal stack by

$$\text{FI}(x,y) = I_{D(x,y)}(x,y), \tag{8}$$

where FI represents the composited all-in-focus image.

On the other hand, suppose that an estimated all-in-focus image is available; the depth map can be refined by a joint bilateral filtering approach [15]:

$$\hat{D}(p) = \frac{1}{K_p}\sum_{p'\in N_p} \exp\left(-\frac{\|p'-p\|^2}{2\sigma_s^2} - \frac{\|\text{FI}(p')-\text{FI}(p)\|^2}{2\sigma_r^2}\right) D(p'), \tag{9}$$

where $p=(x,y)$ is the pixel coordinate, K_p is the normalization factor, N_p is the neighborhood of p, and σ_s and σ_r control the spatial and range weights, respectively. Larger σ_s and σ_r will yield a smoother depth map, and smaller σ_s and σ_r will yield a sharper depth map. N_p is set as a rectangle window that covers the meaningful range of σ_s. In Eq. (9), the all-in-focus image FI is taken as the guidance image of the joint bilateral filter. This

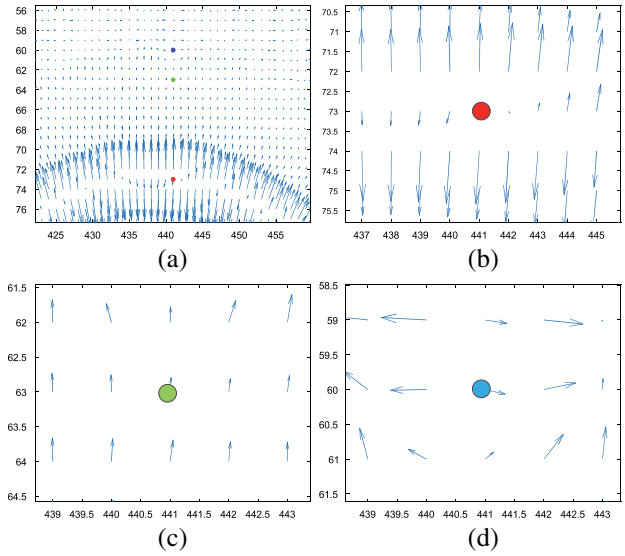

Fig. 4. (a) Max-gradient flow of the sample patch in Fig. 2. (b)–(d) Close-up view of the flow at red, green, and blue pixel (magnitude rescaled for better demonstration).

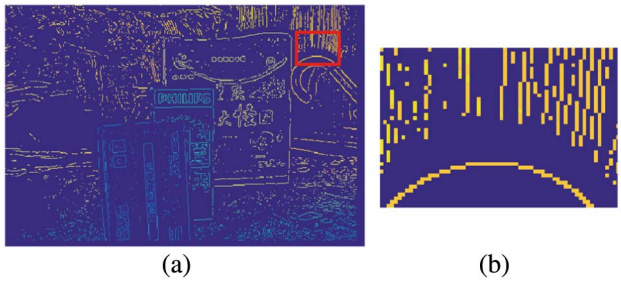

Fig. 5. True edge extraction. (a) Extracted true edges (nondark blue) and meaningless points (dark blue). Red rectangle is the same sample patch as in Fig. 2. (b) Close-up view of the sample patch.

filtering approach smooths the depth map while preserving the edges.

The reality is that neither the perfect depth map nor the all-in-focus image are available. Thus, we take Eqs. (8) and (9) to form a ping-pong approach that estimates the all-in-focus image and refines the depth map in turn. Thus, starting from the initial depth map, the all-in-focus image and the depth map can be updated iteratively in a rolling way.

Furthermore, to prevent the depth map from being over-smoothed, at the beginning of each iteration, we reset the depth values in S to the initial values to control the iterative filtering process. Because the reliable points in S act as anchors, this iterative approach is called anchored rolling filtering.

To be in precise, we initialize the depth map with Eq. (2) and get D^0. The i-th iteration of our method is as follows:

Step 1: reset anchor points, $D^{i-1}(p) = D^0(p), \forall\ p \in S$;
Step 2: update all-in-focus image with Eq. (8), get FI^i;
Step 3: filter depth map with Eq. (9) on guide of FI^i, get D^i.

The anchored rolling filter continues until the depth map shows no significant change between two iterations.

4. EXPERIMENT

In this section, we present the experimental evaluation of our method and the comparisons with state-of-the-art methods.

A. Setup

The performance is tested on the focal stack as described in Section 1. This focal stack is captured with such a large aperture that it is suitable to demonstrate the effectiveness of our method. When capturing the focal stack, the movement of the focusing plane will cause the change of FOV. This FOV change is corrected with the image registration technique [16–18]. In our experiment, the parameters are set as $G_{\mathrm{TH}} = 0.05$ (with the intensity ranging from 0 to 1), $\sigma_s = 5$, $\sigma_r = 0.1$, and N_p a 30×30 window.

B. Iteration

Figure 6 presents the iterations of the anchored rolling filter. It can be seen that, in the initial depth map [Fig. 6(a)], there are many ghost edges, which are removed in the final depth map [Fig. 6(c)]. We evaluate the change of the depth map between two iterations with peak signal-to-noise ratio (PSNR). The PSNR curve is presented in Fig. 6(d). It can be seen that the depth map will not change significantly after about 60 iterations. Thus, we choose to stop the iteration after Iteration 60. The final composited all-in-focus image is presented in Fig. 6(e).

C. Comparison

We also compare our method with state-of-the-art methods. The comparison on our test large aperture data is presented in Fig. 7, which shows the whole content of the composited images of all the compared methods, along with the quantitative evaluation on the evaluation patch [red rectangle in Fig. 7(a)]. This patch contains strong and weak edges and is influenced by the large-scale blur, as described previously. We manually set the ground truth depth map of the evaluation patch and make the ground truth all-in-focus patch by extracting the content from the focal stack accordingly. The performance is evaluated with structural SIMilarity (SSIM) index [19].

From the comparison, it can be seen that our method [Fig. 7(b)] can give the highest SSIM value. Weak edges are preserved as well as strong edges. The result is free of artifacts and ghost edges. Whereas the FSI method produces artifacts near strong edges and enhances the noise [Fig. 7(c)], the 2.5D deconvolution method [Fig. 7(d)], DCT-based method [Fig. 7(e)], and DWT-based method [Fig. 7(g)] produce ghost edges in this large aperture case, especially near the strong

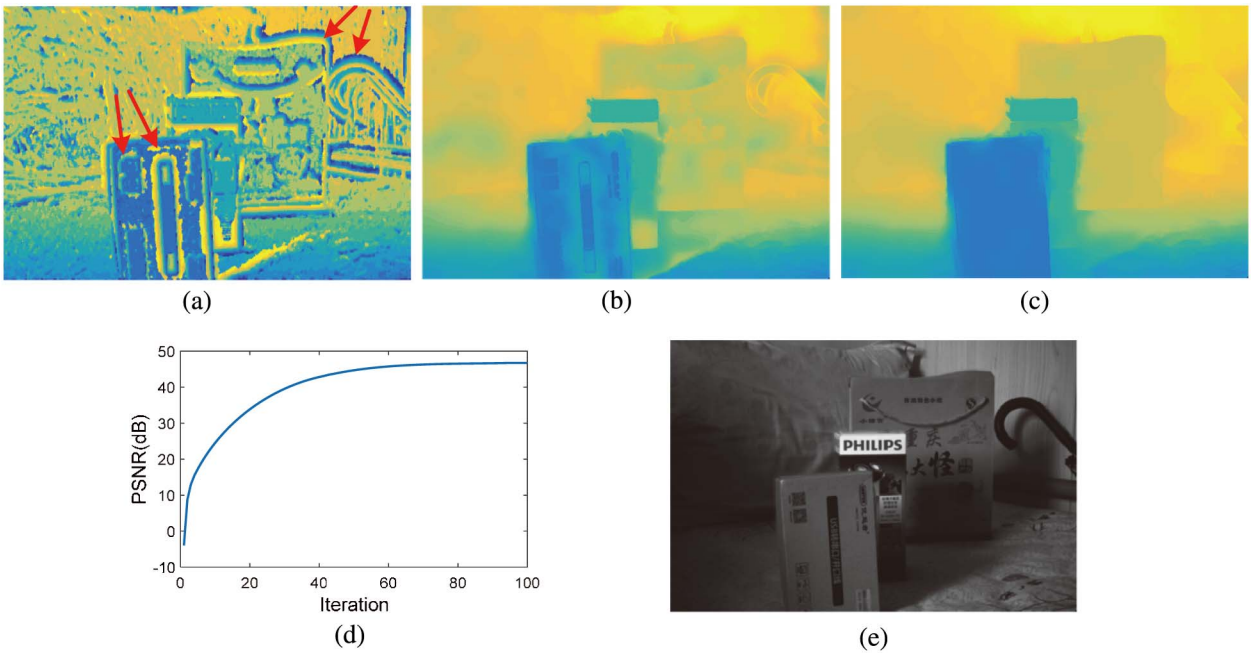

Fig. 6. Depth map refinement. (a) Iteration 0 (initial); note the ghost edges indicated by the arrows. (b) Iteration 10. (c) Iteration 60 (final). (d) PSNR between depth maps of Iteration i and Iteration $i-1$. (e) Composited all-in-focus image.

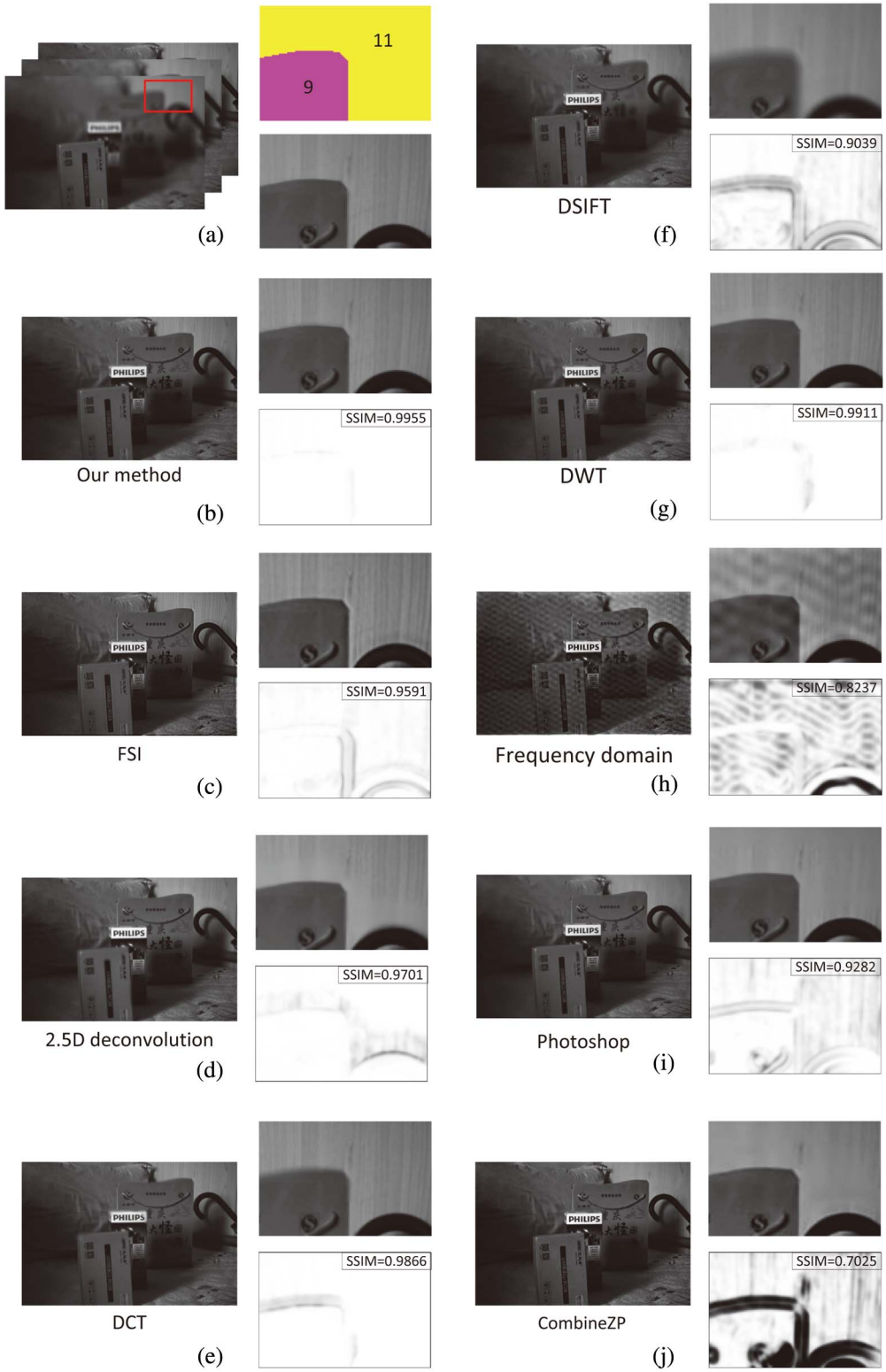

Fig. 7. Comparison of our method with state-of-the-art methods. (a) Left: Focal stack with the evaluation patch (red rectangle). Top right: Manually set ground truth depth map of the patch. The embedded number represents the index of the in-focus image. Bottom right: Manually set ground truth all-in-focus patch according to the ground truth depth map. (b)–(j) Results of our method and the compared methods. Left: Composited all-in-focus image. Top right: Close-up view of the evaluation patch. Bottom right: Local SSIM value map (error map), in which higher value represents more similarity. Overall SSIM value is also given. (b) Our method. (c) FSI method in [12]. (d) 2.5D deconvolution method in [4]. (e) DCT-based method in [11]. (f) DSIFT method in [9]. (g) DWT method in [8]. (h) Frequency domain method in [13]. (i) Photoshop. (j) CombineZP.

edges. The DSIFT-based method [Fig. 7(f)] cannot handle the large aperture images and produces blurry results. The result of frequency domain reconstruction is presented in Fig. 7(h). This method cannot give stable results when the frequency-domain equations are ill-posed, as even one inaccurate frequency component will influence the whole image. We also tested Photoshop CS6 (13.0 x 64) and CombineZP [20] on our sample data. The results [Figs. 7(i) and 7(j)] show that off-the-shelf software cannot give good results in this case, either.

5. CONCLUSION

In conclusion, our work focuses on reducing the artifacts and obtaining a high-quality all-in-focus image in a large aperture case. We define a max-gradient flow to model the gradient propagation in the stack. Based on this definition, a sparse set of true edges is extracted. We formulate an anchored rolling filtering approach to refine the depth map and the composited image iteratively in a rolling way. Experiments are made on our sample data captured with a large aperture. The results show that the proposed method can effectively reduce the artifacts caused by large blur scale and give an all-in-focus image with higher quality at strong edges and weak edges than state-of-the-art methods.

Funding. National Natural Science Foundation of China (NSFC) (61327902); National High Technology Research and Development Program of China (863 Program) (2015AA016304).

REFERENCES

1. N. T. Goldsmith, "Deep focus; a digital image processing technique to produce improved focal depth in light microscopy," Image Anal. Stereol. **19**, 163–167 (2000).
2. W. Huang and Z. Jing, "Evaluation of focus measures in multi-focus image fusion," Pattern Recogn. Lett. **28**, 493–500 (2007).
3. B. Forster, D. Van De Ville, J. Berent, D. Sage, and M. Unser, "Complex wavelets for extended depth-of-field: a new method for the fusion of multichannel microscopy images," Microsc. Res. Tech. **65**, 33–42 (2004).
4. F. Aguet, D. Van de Ville, and M. Unser, "Model-based 2.5-D deconvolution for extended depth of field in brightfield microscopy," IEEE Trans. Image Process. **17**, 1144–1153 (2008).
5. J. Tian, L. Chen, L. Ma, and W. Yu, "Multi-focus image fusion using a bilateral gradient-based sharpness criterion," Opt. Commun. **284**, 80–87 (2011).
6. F. Sroubek, S. Gabarda, R. Redondo, S. Fischer, and G. Cristóbal, "Multifocus fusion with oriented windows," Proc. SPIE **5839**, 264–273 (2005).
7. R. Redondo, F. Sroubek, S. Fischer, and G. Cristóbal, "Multifocus fusion with multisize windows," Proc. SPIE **5909**, 59091B (2005).
8. R. Redondo, F. Sroubek, S. Fischer, and G. Cristóbal, "Multifocus image fusion using the Log-Gabor transform and a multisize windows technique," Inf. Fusion **10**, 163–171 (2009).
9. Y. Liu, S. Liu, and Z. Wang, "Multi-focus image fusion with dense SIFT," Inf. Fusion **23**, 139–155 (2015).
10. M. B. A. Haghighat, A. Aghagolzadeh, and H. Seyedarabi, "Real-time fusion of multi-focus images for visual sensor networks," in *6th Iranian Conference on Machine Vision and Image Processing (MVIP)*, Isfahan, 2010, pp. 1–6.
11. M. B. A. Haghighat, A. Aghagolzadeh, and H. Seyedarabi, "Multi-focus image fusion for visual sensor networks in DCT domain," Comput. Electr. Eng. **37**, 789–797 (2011).
12. S. Kuthirummal, H. Nagahara, C. Zhou, and S. K. Nayar, "Flexible depth of field photography," IEEE Trans. Pattern Anal. Mach. Intell. **33**, 58–71 (2011).
13. J. Alonso, A. Fernández, G. Ayubi, and J. Ferrari, "All-in-focus image reconstruction under severe defocus," Opt. Lett. **40**, 1671–1674 (2015).
14. J. Alonso, A. Fernández, and J. Ferrari, "Reconstruction of perspective shifts and refocusing of a three-dimensional scene from a multi-focus image stack," Appl. Opt. **55**, 2380–2386 (2016).
15. J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," ACM Trans. Graph. **26**, 96 (2007).
16. P. Thévenaz, U. E. Ruttimann, and M. Unser, "A pyramid approach to subpixel registration based on intensity," IEEE Trans. Image Process. **7**, 27–41 (1998).
17. B. He, G. Wang, X. Lin, C. Shi, and C. Liu, "High-accuracy sub-pixel registration for noisy images based on phase correlation," IEICE Trans. Inf. Syst. **E94-D**, 2541–2544 (2011).
18. Q. Miao, G. Wang, and X. Lin, "Kernel based image registration incorporating with both feature and intensity matching," IEICE Trans. Inf. Syst. **E93-D**, 1317–1320 (2010).
19. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Process. **13**, 600–612 (2004).
20. A. Hadley, https://groups.yahoo.com/neo/groups/combinez/info.

129.31.176.87 {ts '2016-12-18 13:20:38'}