

Depth-Based Focus Stacking with Labeled-Laplacian Propagation

Wentao Li, Guijin Wang^(✉), Xuanwu Yin, Xiaowei Hu, and Huazhong Yang

Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
wangguijin@tsinghua.edu.cn

Abstract. Focus stacking is a promising technique to extend the depth of field in general photography, through fusing different images focused at various depth plane. However, existing depth propagation process in depth-based focus stacking is affected by colored texture and structure differences in guided images. In this paper, we propose a novel focus stacking method based on max-gradient flow and labeled Laplacian depth propagation. We firstly extract sparse source points with max-gradient flow to remove false edges caused in large blur kernel cases. Secondly, we present a depth-edge operator to give these sparse points 2 different labels: off-plane edges and in-plane edges. Only off-plane edges are then utilized in our proposed labeled-Laplacian propagation method to refine final dense depthmap and the all-in-focus image. Experiments show that our all-in-focus image is superior to other state-of-the-art methods.

Keywords: All-in-focus · Labeled-Laplacian · Max-gradient flow
Sparse-to-dense

1 Introduction

In general photography, optical imaging systems always have limited depth-of-field: optical lenses focus on a specific plane, while leaving other regions of the scene blurred. Although decreasing the aperture size could extend the DOF in some extent, this would lead to lower signal-to-noise ratio and longer exposure time. To overcome this limitation, focus-stacking has become more popular with the development of digital imaging technology [2, 5, 14]. It captures a sequence of images focused at various planes and fuses them into a single all-in-focus image.

The focus stacking technique has attracted a lot of attentions in the last decade, which could be divided into 2 categories: transform domain fusion approaches and depth-based approaches. For transform domain fusion approaches, source images are converted in transform domain, then corresponding transform coefficients (DWT [10], DSIFT [7], DCT [3]) are fused, finally the all-in-focus image is reconstructed by the inverse transform. These methods are usually complicated and unstable with variation of transform coefficients.

In depth-based methods [9, 11, 15], they firstly extract some sparse pixels whose depth values are the sharpest index across the stack, then propagate them

to the dense depthmap, finally the all-in-focus image is generated by fusing pixels in the stack according to the depthmap. Suwajanakorn et al. [11] used sharpness measurement and formulated the fusing problem as a multi-labeled MRF optimization problem. Moeller chose well-known modified Laplacian (MLAP) function as the measure of contrast, and propagated the resulting depth estimates in a single variational approach (VDFP [9]). Aguet et al. [1] also estimated the all-in-focus image with a model based 2.5D deconvolution method. In all methods above, the depth values of extracted sparse points are affected and noised by false edges occurred in large blur kernel case. To remove false edges, we proposed max-gradient flow [15] to extract true source points, and gave an iterative anchored rolling filter to estimate the all-in-focus image. However, in all the sparse-to-dense propagation processes in these depth-based methods, the final depthmap is affected by colored texture and structure differences in guided images.

In this paper, we propose a novel focus stacking method based on max-gradient flow and labeled Laplacian depth propagation. Firstly, we construct sparse depthmap with the max-gradient-flow proposed in our previous MGF-ARF method [15]. Then we design a depth-edge operator to give these sparse points 2 different labels: off-plane edges and in-plane edges. Here in-plane edges are image edges at the same depth plane, while off-plane edges are image edges at boundaries of different depth planes. Only off-plane edges are then utilized in the labeled-Laplacian depth propagation to generate final dense depthmap which is smoothed at textures in the same depth plane and strengthened at boundaries between different depth planes. Experiments show that our depthmap is smoothed at textures in the same depth plane and sharpened at depth boundaries, while the all-in-focus image is refined and superior to other state-of-the-art methods.

2 Sparse Depthmap with Max-Gradient Flow

In this section, we introduce the max-gradient flow [15] briefly to extract sparse depthmap. Max-gradient flow could model the propagation of gradients in the stack and remove the false edges produced in large blur kernel cases. To introduce max-gradient flow in detail, We capture a sample stack with Imperx B4020 mono camera equipped with a SIGGMA 50 mm/F1.4 lens. This stack consists of 14 images with large blur kernels, and is utilized to describe our method in the rest of our paper. Figure 1 shows 3 images focused at different depth planes from our stack.

With focal stacks I_1, I_2, \dots, I_n , an all-in-focus image could be produced by selecting the sharpest pixels across the focal stack. Several different measures of pixel sharpness have been defined in some shape-from-focus literature [8–10]. In this paper, without loss of generality, magnitude of gradients is calculated as sharpness measurement, which is defined as

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



Fig. 1. Three images captured from real scene focused at 3 focusing plane respectively

where G_i is the gradient magnitude of I_i , the i -th image in the stack. Then depth value of sparse points could be calculated as:

$$D(x, y) = \operatorname{argmax} G_i(x, y) \quad (2)$$

Here $D(x, y)$ stores the depth value that gives the sharpest gradient across the stack. However, traditional methods following Eqs. (1) and (2) would produce ‘false edges’ [15]. False edges, the production of which has been explained in detail in [15], are those image edges with false depth values because of spreading of blur kernels of neighbouring strong edges in large blur kernel cases.

To remove these false edges, max-gradient flow is utilized to analysis the propagation of gradients. The max-gradient flow from [15] is defined as:

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

Here the two elements are calculated as:

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

The flow describes the propagation of gradients in the stack and is valid to divide points in the stack into 2 categories: source points and trivial points. Source points are points whose depth value calculated by Eqs. (1) and (2) is true and valid, while trivial points are points with false depth value from Eq. (2). Points whose max-gradient flow changes its direction oppositely are chosen as source points, formulated as:

$$\nabla \cdot MGF(x, y) > 0 \quad (5)$$

Otherwise, the points are defined as trivial points if

$$\nabla \cdot MGF(x, y) < 0 \quad (6)$$

We only preserve the depth value of source points to get the sparse depthmap. Figure 2 shows the comparison of performance of sparse depthmap with and without applying max-gradient flows. We could find that with max-gradient flow, the false edges are effectively suppressed and true edges are preserved as many as possible.

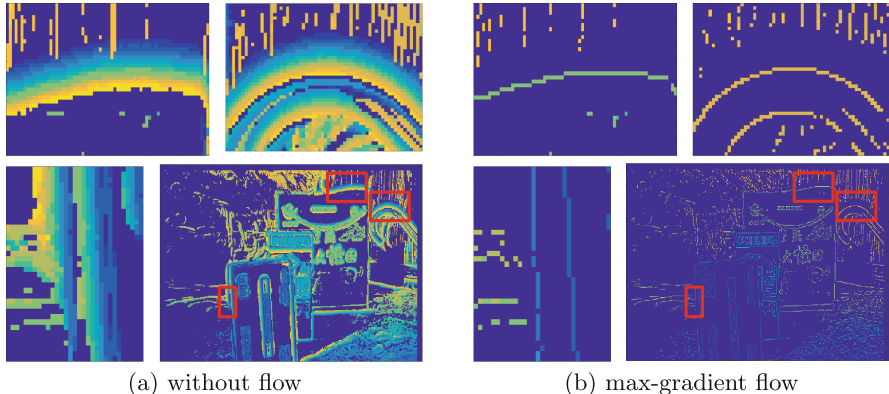


Fig. 2. (a) Depth values without applying max-gradient flow. (b) Depth value for source points extracted with max-gradient flow

3 Labeled-Laplacian Depth Propagation

Laplacian matting is a traditional sparse-to-dense propagation method. Similar with other propagation methods, it causes depth artifacts and noises at textures on the same depth plane because of color and structure differences of guided image. Therefore, to generate a refined depthmap, it is critical to differentiate image edges on the same depth plane with those points at boundaries of depth planes to propagate these 2 labels of sparse points respectively. In this section, we propose a novel two-step depth propagation process. Firstly, we construct a novel L-matrix to get a coarse dense depthmap which removes effects of colored texture and structure differences of guided images. Secondly, 2 different labels are given to sparse points by our depth-edge operators extracted from the coarse dense depthmap: off-plane edges and in-plane edges. Then, in the second propagation process, only off-plane edges are utilized to update L-matrix. In this way, two labels of points are propagated differently to refine the dense depthmap: in-plane edges are smoothed while off-plane edges are strengthened and sharpened.

3.1 Coarse Dense Depthmap

In traditional Laplacian propagation methods [6, 16], the depth propagation problem could be formulated as minimizing the following cost energy:

$$E(d) = d^T L d + \lambda (d - \hat{d})^T D (d - \hat{d}), \quad (7)$$

D is a diagonal matrix whose element $D(i, i)$ is equal to 1 if the pixel i has valid depth value. d and \hat{d} are the dense depthmap and the sparse depth map which only has valid depth values at source points. Decomposing the Eq. (7), $d^T L d$ denotes the fidelity of source points while $(d - \hat{d})^T D (d - \hat{d})$ denotes the

smoothness of depth propagation. The scalar λ controls the balance of these two parts. L is the Laplacian matrix calculated from color and structure differences of guided images, and is traditionally calculated as below:

$$L(i, j) = \sum_{k|(i,j) \in \omega_k} \left(\delta_{ij} - \frac{1}{|\omega_k|} (1 + (I_i - \mu_k)^T (\Sigma_k + \frac{\varepsilon}{|\omega_k|} U_3)^{-1}) (I_j - \mu_k) \right) \quad (8)$$

where δ_{ij} is the Kronecker delta, U_3 is identity matrix, Σ_k is the covariance matrix of the colors in patch ω_k , I_i and I_j are colors of all-in-focus image as guided image. From the equation above, differences of RGB-values of patches of guided image would affect the construction of L-matrix and the cost energy of Eq. (7). Therefore depthmap would produce depth artifacts and noises at locations of colored textures of guided images on the same depth plane.

To remove these depth noises, we assume that all pixels in each patch ω_k are constant, which makes $I_i = \mu_k$ and modify the L into:

$$L(i, j) = \sum_{k|(i,j) \in \omega_k} \left(\delta_{ij} - \frac{1}{|\omega_k|} \right) \quad (9)$$

From this equation, the construction of L matrix has nothing to do with colored textures I_i , I_j of guided image. Furthermore, the cost energy in Eq. (7) only depends on the sparse depthmap d shown in Fig. 3(a) and its distribution D . In this way, depth noises caused by colored textures of guided images in traditional propagation methods are removed and the coarse dense depthmap shown in Fig. 3(b) is produced only according to depth values of sparse source points. This dense depthmap is utilized to extract depth-edge operators in the next section.

3.2 Labeled-Laplacian Depth Propagation

From Fig. 3(a) and (b), the coarse dense depthmap, which is blurry at edges of different depth planes, is not satisfying. Therefore, we propose a labeled-Laplacian depth propagation which sharpens edges of different depth planes to refine its estimation. Firstly, we design a novel operator to give source points 2 labels: off-plane edges and in-plane edges.

For each source point as centered, we spread both N pixels along and against the rising direction of gradient in the coarse dense depthmap to construct a $(2N + 1) * 1$ pixels patch ω as our operator. In our operator, the depth value increases along with the increase of pixel index. From the definition above, we know that off-plane edges locate at boundaries of objects belonging to different depth plane, and are usually sharp when the image is focused at the nearer object. Therefore only the points with relatively small depth value and whose neighbouring depth value vary in large range should be classified as true off-plane edges. Therefore, we apply the equation below to calculate the value Ω for each depth-edge operator. The source point k is labeled as off-plane edges if

$$\Omega_k = \frac{\omega_k(2N + 1) - \omega_k(N + 1)}{\omega_k(N + 1) - \omega_k(1)} > \Omega_{TH} \quad (10)$$

Figure 3 shows the process of setting labels for source points in this section. Three different example points are displayed in Fig. 3(a). Only the red one is located at the boundaries of different depth planes, while the green point and the blue point are both at the same depth plane. Figure 3(b) shows the coarse dense depthmap generated from Eq. (12), from which we extract operators ω , and Fig. 3(c) presents the depth-edge operators of three example points. Observing Fig. 3(c), only the red point is divided as true off-plane edges with Eq. (10).

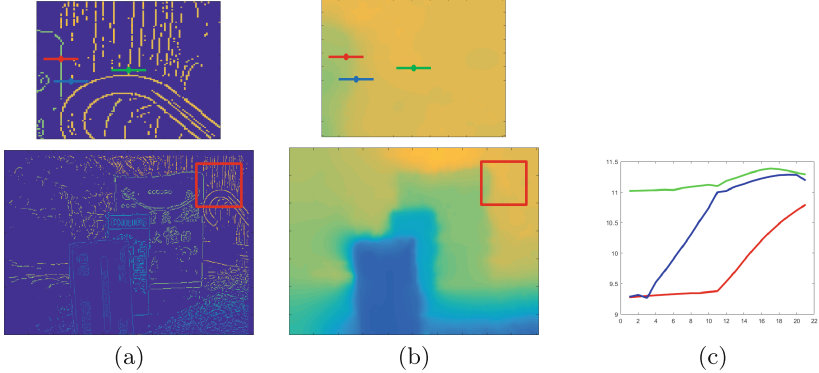


Fig. 3. (a) The close-up view and the entire source points result, (b) close-up view and entire performance of coarse dense depthmap (c) the value of feature vectors of three different points across the $(2N + 1) * 1$ patch (Color figure online)

From the labels of off-plane edges, the energy minimization equation for depth-propagation could be updated again as:

$$E(d) = d^T \hat{L}d + \lambda(d - \hat{d})^T D(d - \hat{d}), \quad (11)$$

where \hat{L} is modified as:

$$\hat{L}(i, j) = \sum_{k|(i,j) \in \omega_k} \left(\delta_{ij} - \frac{1}{|\omega_k|} (1 + (I_i - \chi(i, k))^T (\Sigma_k + \frac{\varepsilon}{|\omega_k|} U_3)^{-1} (I_j - \chi(j, k))) \right) \quad (12)$$

here

$$\chi(i, k) = (1 - \Pi_i) \mu_k + \Pi_i I_i \quad (13)$$

where $\Pi_i = 1$ when the point is classified as off-plane edge, and 0 if in-plane edges.

From the equation above, in our labeled-Laplacian propagation method, only off-plane edges' color and structure differences of guided image are utilized to update the modified L-matrix. This is because that only at off-plane edges, depth boundaries are aligned with edges of guided image. In this way, we generate the refined dense depthmap, where sparse points with different labels are propagated differently: depth differences of off-plane edges are strengthened while depth values of in-plane edges are smoothed.

4 Experiments

4.1 Setup

Performance of our method is tested on the focal stack we introduced in Sect. 2. The movement of the focusing plane when capturing the focal stack would cause the change of field of view. It is corrected with the image registration technique [4, 8, 12]. In our experiments, parameters are set as follows: $G_{TH} = 0.05$, $\Omega_{TH} = 20$, $\lambda = 0.1$, $N = 10$.

4.2 All-in-Focus Comparison

We first compare our all-in-focus performance with state-of-the-art methods. Figure 4 shows the ground truth depthmap of the 3 evaluation patches. The yellow patch and the red patch both contain off-plane edges between different planes, while the blue patch only contains on-plane edges on the same depth plane. We manually set the groundtruth depthmap to produce all-in-focus patches shown in Fig. 4 by extracting the corresponding content from the focal stack.

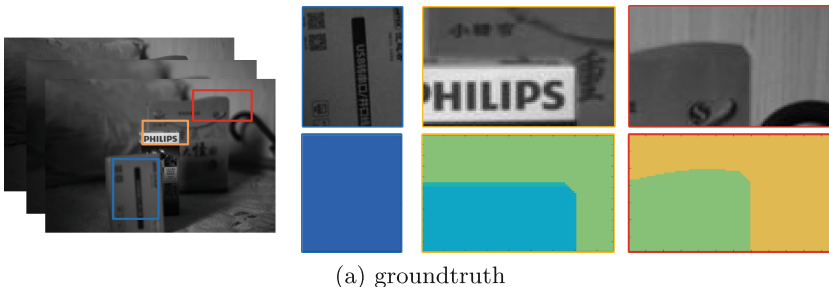


Fig. 4. Left: the focal stack with the evaluation patches (red, blue and yellow rectangle). Upper right: manually set ground truth depthmap of the patch. The embedded number represents the index of the in-focus image. Lower right: manually set ground truth of all-in-focus patch according to the groundtruth depthmap of three different patches (Color figure online)

The comparison on our test data is presented in Fig. 5. It shows quantitative evaluation on three extracted patches and whole content of the composited image of the all the compared methods (DCT, DSIFT, 2.5D deconvolution and MGF-ARF). The performance is evaluated with Structural SIMilarity (SSIM) [13] index, the higher SSIM value indicates more similarity between two images. In Fig. 5, for each method, the left part shows the composited all-in-focus image. For the right part, the upper row shows the constructed all-in-focus images for different extracted patches while the lower row presents the local SSIM value map (error map). To make the error map visualize more distinguishable, we choose



Fig. 5. Comparison of our method with state-of-the-art methods. (a) Our method (b) DCT (c) DSIFT (d) MGF-ARF (e) 2.5D deconvolution

different δ_{SSIM} for corresponding patches according to different distributions of SSIM in different patches, to map the value of $[\delta_{SSIM}, 1]$ of SSIM to $[0, 1]$ of brightness of the displayed error map.

From the comparison, we could find that our method (Fig. 5(a)) gives the highest SSIM over compared methods. Our methods preserve both off-plane edges and in-plane edges to make the strong edges and weak edges both sharpest free of artifacts and ghost edges. Whereas, the DSIFT-based method produces artifacts near both off-plane edges and in-plane edges and enhance the noise, as shown in Fig. 5(c). The DCT-based method, Fig. 5(b) and the 2.5D deconvolution method, Fig. 5(f) both produce ghost edges in the off-plane edges, which makes the strong edges blurry and the weak edges near them disappeared in the red patch and the yellow patch. The MGF-ARF method, which is presented in Fig. 5(d), although is free of artifacts of ghost edges, produces noises subject to colored texture from the guided image on the blue patch belonging to the same depth plane.

4.3 Depthmap Comparison

Figure 6 presents performance of our final dense depthmap with our labeled-Laplacian propagation method and the comparison with state-of-the-art depth propagation methods(Laplacian propagation, ARF [15] and DVFF [9]). Figure 6(d) presents our refined dense depthmap, where depth values in the same depth plane are smoothed and depth boundaries are strengthened. We also choose one patch (red) to display the advantage of our method more clearly. In Fig. 6(c), depth values are totally wrong because of false edges. In small patches of Fig. 6(a) and (b), the depth value of the farther box and the boundaries

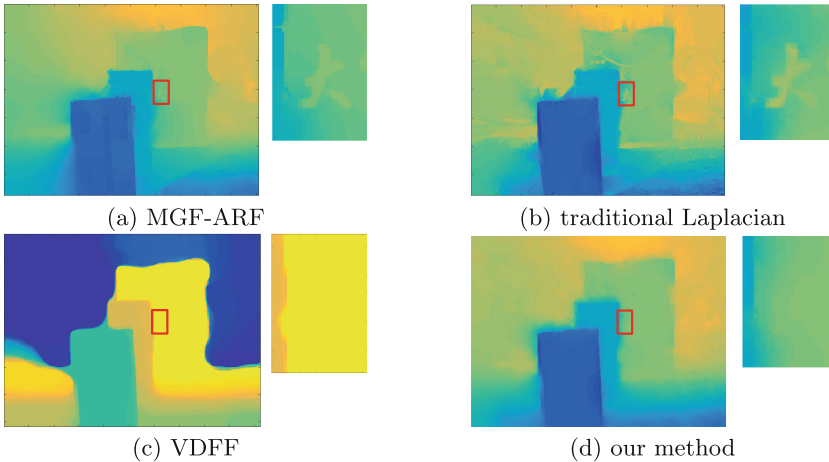


Fig. 6. Close-up views of red region and entire depthmap generated by: (a) MGF-ARF (b) traditional Laplacian optimization (c) VDFV (d) our method (Color figure online)

between two different depth plane are affected by the colored texture. In Fig. 6(d) produced by our method, however, noises in the farther box are removed and depth value are smoothed with our labeled-Laplacian propagation.

5 Conclusion

In conclusion, we propose a novel focus stacking method based on max-gradient flow and labeled-Laplacian depth propagation. We utilize max-gradient flow to extract true source points to generate sparse depthmap. Then we design a depth-edge operator to give these sparse points 2 different labels: off-plane edges and in-plane edges. Only off-plane edges are then utilized in the following labeled-Laplacian depth propagation to generate final dense depthmap. Experiments show that our method achieve an all-in-focus image with higher quality than state-of-the-art methods.

Acknowledgement. This work was partially supported by National Science Foundation of China (No. 61271390) and State High-Tech R&D Program of China (863 Program, No. 2015AA016304).

References

1. Aguet, F., Van De Ville, D., Unser, M.: Model-based 2.5-D deconvolution for extended depth of field in brightfield microscopy. *IEEE Trans. Image Process.* **17**(7), 1144–1153 (2008)
2. Goldsmith, N.T.: Deep focus; A digital image processing technique to produce improved focal depth in light microscopy. *Image Anal. Stereology* **19**(3), 163–167 (2011)
3. Haghghat, M.B.A., Aghagolzadeh, A., Seyedarabi, H.: Real-time fusion of multi-focus images for visual sensor networks. In: 2010 6th Iranian Machine Vision and Image Processing (MVIP), pp. 1–6. IEEE (2010)
4. He, B., Wang, G., Lin, X., Shi, C., Liu, C.: High-accuracy sub-pixel registration for noisy images based on phase correlation. *IEICE TRANS. Inform. Syst.* **94**(12), 2541–2544 (2011)
5. Huang, W., Jing, Z.: Evaluation of focus measures in multi-focus image fusion. *Pattern Recogn. Lett.* **28**(4), 493–500 (2007)
6. Levin, A., Lischinski, D., Weiss, Y.: A closed-form solution to natural image matting. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(2), 228–242 (2008)
7. Liu, Y., Liu, S., Wang, Z.: Multi-focus image fusion with dense sift. *Inform. Fusion* **23**, 139–155 (2015)
8. Miao, Q., Wang, G., Lin, X.: Kernel based image registration incorporating with both feature and intensity matching. *IEICE TRANS. Inform. Syst.* **93**(5), 1317–1320 (2010)
9. Moeller, M., Benning, M., Schönlieb, C., Cremers, D.: Variational depth from focus reconstruction. *IEEE Trans. Image Process.* **24**(12), 5369–5378 (2015)
10. Sroubeka, F., Gabardab, S., Redondob, R., Fischerb, S., Cristóbalb, G.: Multifocus fusion with oriented windows. In: *Proceedings of SPIE*, vol. 5839, p. 265 (2005)

11. Suwajanakorn, S., Hernandez, C., Seitz, S.M.: Depth from focus with your mobile phone. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3497–3506 (2015)
12. Thevenaz, P., Ruttimann, U.E., Unser, M.: A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* **7**(1), 27–41 (1998)
13. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)
14. Yin, X., Wang, G., Li, W., Liao, Q.: Iteratively reconstructing 4D light fields from focal stacks. *Appl. Opt.* **55**(30), 8457–8463 (2016)
15. Yin, X., Wang, G., Li, W., Liao, Q.: Large aperture focus stacking with max-gradient flow by anchored rolling filtering. *Appl. Opt.* **55**(20), 5304–5309 (2016)
16. Zhuo, S., Sim, T.: Defocus map estimation from a single image. *Pattern Recogn.* **44**(9), 1852–1858 (2011)