SCENE-ADAPTIVE IMAGE ACQUISITION FOR FOCUS STACKING

Wentao Li, Guijin Wang, Xiaowei Hu, Huazhong Yang

Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

ABSTRACT

Focus stacking is a promising technique to extend depth of field in general photography by fusing images captured at different focusing distances. In this paper, we propose a round-trip scene-adaptive image acquisition system to automatically capture focal stack and fuse a high quality all-in-focus image. Based on scene analysis, we cover entire depth range of the scene in the forward optical scanning and refine all objects' focusing positions accurately in the backward scanning. With captured images, we firstly extract depthmap and all-in-focus image with combination of max-gradient flow and blur kernel estimation. Secondly, a superpixel-level Gaussian Fitting is proposed to determine the next location to capture. Experiments on simulated data show that our method attain high quality all-in-focus image with fewer captured images.

Index Terms— all-in-focus; image acquisition; max-gradient-flow; superpixel; Gaussian-Fitting

1. INTRODUCTION

In general photography, optical lenses have limited depth of field (DOF): they usually focus on specific planes while leaving other regions of the scene blurred[1]. To extend DOF of a single image, focus stacking has become more and more popular with the development of digital imaging technology[2, 3]. It captures a sequence of images focused at different planes and fuses them into a single all-in-focus image.

Most researches focus on how to reconstruct all-in-focus images more accurately[4, 5, 6]. However, how to capture images more effectively has not received much attention. In most literature about all-in-focus image reconstruction, source images of focal stack are captured simply with equal interval[7, 8, 9]. In these methods, source images are captured moving the optical lens a uniform step-size, which might lead to images which contain nothing in focus and increase the complexity of image fusing. Hasinoff et al. [10, 11, 12] constructed a model of exposure level and DOF and discussed how to select sets of images quickly with a given depth of field, but they ignored camera overhead and the post-processing for image fusion. Vaquero at al. [13] presented an end-to-end system to select minimal sets of images for focus stacking. David Choi [14] proposed enhancement of the proposal by Vaquero and improved image selection on cameras with variable apertures and lenses with longer focal lengths. However, in these methods, image selection is based on scene distribution estimated by stream of low-resolution images densely captured beforehand. Therefore they separated image capturing and scene analysis, and led to the limitation of image capturing efficiency and increase of complexity of image fusing. Kuthirummal et al. [15] presented another image capturing technique called as Focal Sweep Imaging (FSI) to extend the DOF, where the sensor moved along the optical axis during one exposure. But the transform-domain-based method is sensitive to perturbation of transform coefficients for lack of analysis of the scene.

In this paper, a novel scene-adaptive image capturing system is proposed to improve efficiency for focus stacking. There are three main contributions in our method. Firstly, we propose a image acquisition system, for the first time, to capture focal stack at scene objects' optimal locations with only one round-trip lens scanning. We move the optical lens simply forward and backward to determine optimal capturing positions. With captured images, online scene analysis is proceeded to estimate depth distribution of the scene and determine the next location to capture. In this way, we would increase acquisition efficiency and practicability to reconstruct a high quality all-in-focus image with fewer images. Secondly, we present a novel sparse depthmap estimation method utilizing max-gradient flow and blur kernel estimation jointly with sparse focal stack. Thirdly, we present a superpixel level Gaussian Fitting method to determine the image acquisition locations. Experiments on simulated data show that our method reconstruct accurate all-in-focus image while reducing the number of captured images.

2. OUR PROPOSED CAPTURING SYSTEM

In our one round-trip capturing system, we move optical lens along two different one-way directions: forward firstly and backward secondly. We cover entire depth range of the scene efficiently in the forward scanning and refine focusing positions of all objects in the scene as accurately as possible in the backward scanning. Here we introduce framework of our image acquisition system shown as Fig.1 in detail. In the forward scanning, the optical lens moves from near to far and the image acquisition approach proceeds as follows: Initially,

This work is partially supported by NSFC 61327902

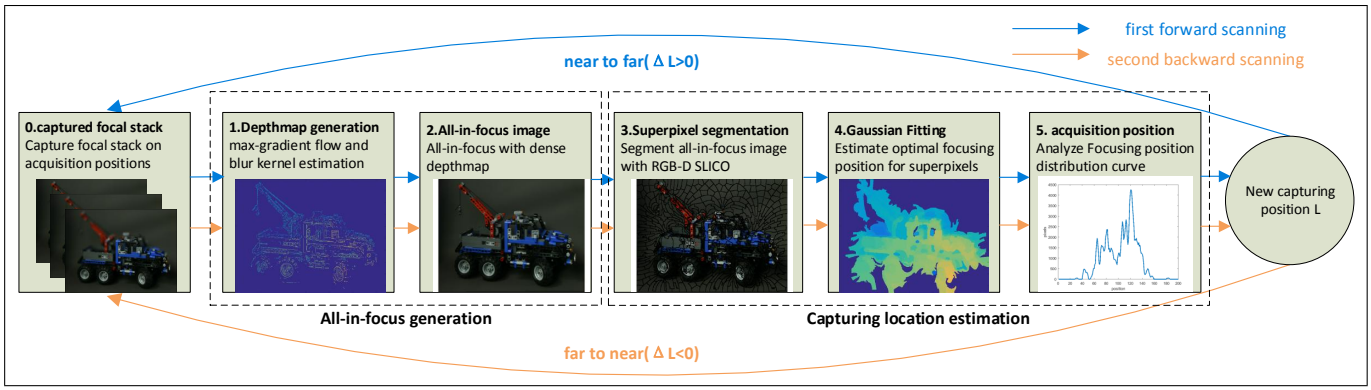

Fig. 1: Pipeline for our proposed capturing system

the image acquisition positions are set as $P_0 = [L_0, L_1, L_2]$ where L_0 represents the nearest captured position while $L_1 = L_0 + \tau_{max}$, $L_2 = L_0 + 2\tau_{max}$. Defined as maximum acquisition interval, τ_{max} is set to cover entire depth range of the scene more efficiently. We capture images at these captured positions and estimate sparse depthmap $\tilde{d}$, dense depthmap d and all-in-focus image FI with the captured sparse focal stack to understand the captured scene. Detailed calculation process would be explained in Section 3. Then we propose superpixel-level Gaussian Fitting to estimate entire scene's focusing position distribution, which would be utilized to determine the next position to capture to update image acquisition positions. After the optical lens moves to the farthest position, it moves backward from far to near and the backward scanning process begins with acquisition positions set containing captured positions P_m. The second image acquisition scanning is proceeded similarly, except for the module of capturing location decision: we introduce smaller acquisition interval τ_{min} to refine focusing positions of scene objects as accurate as possible. Specific details about capturing position decision would be explained in Section 4. In this way, we automatically capture a minimal set focal stack with the round-trip scene-adaptive by fusing the scene analysis and image acquisition.

3. ALL-IN-FOCUS GENERATION

For the i_{th} image in the captured focal stack, the pixel value $I(x, y)$ can be modeled as a convolution of a sharp point $F(x, y)$ with the point spread function (PSF) which could be approximated by a Gaussian function $G(x, y, \sigma_i(x, y))$ [16], and the blurred-point $I(x, y)$ could be given by:

$$I(x, y) = F(x, y) \otimes G(x, y, \sigma_i(x, y)) + N(x, y), \quad (1)$$

where $N(x, y)$ is the white noises of pixel (x, y).

In this section, we introduce how to estimate sparse depthmap with combination of max-gradient flow and blur kernel estimation. Then an all-in-focus image would be reconstructed with already-captured focal stack.

3.1. sparse depthmap with max-gradient flow

In our previous work [4], we utilized max-gradient flow to do all-in-focus composition. In the next few paragraphs, we mainly give additional analysis on its applicability to the scene-adaptive captured focal stack.

Max-gradient flow was proposed in [4] to model the propagation of edges in focal stack:

$$MGF(x, y) = \begin{bmatrix} \frac{\max_j G_j(x+\Delta x, y) - \max_i G_i(x, y)}{\Delta x} \\ \frac{\max_k G_k(x, y+\Delta y) - \max_i G_i(x, y)}{\Delta y} \end{bmatrix}. \quad (2)$$

Here $G_i(x, y)$ is the gradient value of pixel (x, y) in the i_{th} image in the stack. This flow describes the propagation of gradients in the stack and is valid to extract edges with true depth values as source points but has several disadvantages if the stack is captured badly. Firstly, if capturing range of stack does not cover object's actual focus position, $\max_i G_i(x, y)$ might not represent true max gradient of pixel (i, j). This would lead to missing of true source points. Secondly, if images are captured too densely, largest image gradients are difficult to contrast due to the effects of $N(x, y)$. This would increase computational complexity and bring depth noises.

In our method, since the focal stack is captured based on the scene, we would find most accurate depth position $D(x, y)$ for each pixel. This position only depends on depth distribution of scene, independent of specific acquisition location of already-captured stack and would be determined based on Gaussian-Fitting in section 4.2 . Then we capture images at location $D(x, y)$ and calculate its max gradient as $G_{D_{(x,y)}}(x, y)$ instead of $\max_i G_i(x, y)$ and modify Eq.2 as:

$$MGF(x, y) == \begin{bmatrix} \frac{G_{D_{(x,y)}}(x+\Delta x, y) - G_{D_{(x,y)}}(x, y)}{\Delta x} \\ \frac{G_{D_{(x,y)}}(x, y+\Delta y) - G_{D_{(x,y)}}(x, y)}{\Delta y} \end{bmatrix}. \quad (3)$$

Therefore, on the occasion of scene-adaptive image acquisition, we cover actual focusing positions for all objects in the scene. In this way, we would extract true source points as many as possible while reducing depth noises.

3.2. sparse depthmap with blur kernel estimation

In this section, we utilize the method proposed by [16] to estimate blur kernel $\sigma_i(x, y)$ to filter our sparse depthmap

We re-blur the i_{th} captured image using a known Gaussian kernel σ_0, then the maximum ratio between the gradient magnitude of $I_i(x, y)$ and its re-blurred version would be calculated as:

$$R_i(x, y) = \frac{\nabla I_i(x, y)}{\nabla I_i'(x, y)} = \sqrt{\frac{\sigma_i(x, y)^2 + {\sigma_0}^2}{\sigma_i(x, y)^2}}. \quad (4)$$

And the blur kernel of pixel (x, y) in the i_{th} captured image would be calculated. Since the accurate focusing position of pixel (i, j) possesses the smallest blur kernel, we reserve source points with blur kernels smaller than parameter σ_{th} to filter our sparse depthmap from last section.

After generating accurate sparse depthmap $\tilde{d}$ we propagate depth values from edge locations to entire image to obtain dense depthmap d by matting Laplacian[3] with parameter λ. This method considers fidelity to the sparse depthmap as well as smoothness of propagation. Finally, we reconstruct the all-in-focus image based on dense depthmap as follows:

$$FI(x, y) = I_{d(x,y)}(x, y), \quad (5)$$

4. CAPTURING LOCATION ESTIMATION

We have discussed in the last section about how to generate dense depthmap as well as all-in-focus with several captured images. In this section, in turn, we introduce how to update capturing positions accurately. Firstly, we propose a superpixel-leveled Gaussian-Fitting method to estimate actual focusing positions of objects to analyze depth distribution of the whole scene. Secondly, we introduce how to determine new location to capture one at a time.

4.1. superpixel leveled Gaussian Fitting

In this section, we propose a RGB-D based superpixel segmentation method and design a superpixel-leveled Gaussian Fitting to estimate optimal focusing positions for each superpixel. SLICO method [17] performs satisfying in traditional superpixel segmentation method and inspired our work. Considering the dense depthmap d from last section, we modify the method and combine color distance in *lab* space of FI, spatial distance in *xy* space and depth value distance in *d* space into a single measure D' of overall proximity for superpixel segmentation as follows:

$$\begin{aligned} d_c &= \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2} \\ d_s &= \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \\ d_v &= \sqrt{(d_j - d_i)^2} \\ D' &= \sqrt{(\tfrac{d_c}{N_c})^2 + (\tfrac{d_s}{N_s})^2 + (\tfrac{d_v}{N_v})^2} \end{aligned} \quad (6)$$

After we produce the superpixel set $[S_1, S_2, ...S_M]$ containing $[n_1, n_2, ..., n_M]$ source points respectively, we calculate superpixel-leveled gradient as follows:

$$SG_j^i = \frac{1}{n_i} \sum_{(x,y)\in S_i} G_j(x, y), \quad (7)$$

where i indicates superpixel index $\in [1, ..., M]$ and j represents the capturing positions.

This equation denotes that we calculate superpixel-leveled gradients by the average gradients of all source points the superpixel contains and it records gradient values on already-captured positions. Since we consider depth value distance in our proposed RGB-D based superpixel segmentation, source points in one superpixel would have the same focusing position. Therefore it would be efficient and robust to estimate accurate capturing positions by curve fitting on the superpixel level.

Here we assume that the image gradients satisfy Gauss distribution across different capturing locations with the scene-constant standard deviation as follows:

$$SG_j^i = SG_{SD^i}^i e^{-\frac{(j-SD^i)^2}{2\sigma^2}}, \quad (8)$$

where σ indicates gradient standard deviation and determines image minimum acquisition intervals and SD^i is the depth location with max gradient of superpixel SG^i. Here we set minimum acquisition intervals τ_{min} in the backward scanning as 2σ. Therefore we apply Gaussian Fitting on superpixel-leveled discrete gradient values to update minimum acquisition intervals as well as to estimate accurate focusing positions for each superpixel. We then generate focusing positions distribution curve $C(j)$ as follows:

$$C(j) = \sum_{i=1}^{M} \delta(SD^i = j), \quad (9)$$

which records the most accurate focusing position for all superpixel and would be analyzed to estimate new positions to be captured in the following section. Here δ is the Kronecker delta while i, j indicates superpixel index and capturing locations respectively.

4.2. capturing position decision

In this section, we would introduce how to determine next position to be captured by analyzing focusing position distribution curve. For the forward scanning, if the acquisition position set $P = [L_0, L_1, ..., L_k]$, the searching approach proceeds as Algorithm 1. Here N_{sup} is the threshold to select the next acquisition position and is set as $M/100$. We choose l which is the first position farther than L_k and containing focused superpixels more than N_{sup} as the position to be captured in the forward scanning. In the forward scanning, Gaussian Fitting might produce false capturing locations due to

Algorithm 1 acquisition position searching of forward scanning

1: Initialize position to be captured as $l = L_k + 1$ and calculate acquisition distance $d = l - L_k$
2: **while** Acquisition distance is smaller than τ_{max} and number of superpixel with focusing position l is smaller than N_{sup},i.e., $d < \tau_{max}$ and $C(l) < N_{sup}$ **do**
3: Update position to be captured, $l = l + 1$
4: Update acquisition distance, $d = d + 1$
5: **end while**
6: The next capturing position l

sparsity of already-captured positions. Therefore we set maximum acquisition interval τ_{max} to avoid especially far focusing positions and ensure the entire depth range of scene would be covered in the forward scanning.

For the backward scanning, the optical lens moves from far to near. Different from the forward scanning, we add minimum acquisition interval τ_{min} to avoid capturing focal stack too densely. It means that every two adjacent acquisition positions are no less than τ_{min} to improve capturing efficiency. To smooth focusing position distribution curve, we modify $C(j)$ to:

$$C'(j) = \sum_{n=-\tau_{min}}^{\tau_{min}} C(j+n). \tag{10}$$

Since we have estimated depth distribution of all objects of the scene in the forward scanning, the Gaussian Fitting would produce more robust and accurate focusing positions. Therefore the parameter τ_{max} is eliminated in the backward scanning. If the acquisition position set $P = [L_0, L_1, ..., L_k]$, the searching approach proceeds as Algorithm 2:

Algorithm 2 acquisition position searching of backward scanning

1: Initialize position to be captured as $l = L_k - 1$
2: Set $flag = 0$
3: **while** $flag = 0$ **do**
4: Calculate minimum distance between location l and set P: $d = \min_i |L_i - l|$
5: **if** $C'(l)$ is larger than N_{sup} and distance d is larger than τ_{min},i.e. $C'(l) > N_{sup}$ and $d \geq \tau_{min}$ **then**
6: set $flag = 1$
7: **else**
8: update $l = l - 1$
9: **end if**
10: **end while**
11: The next capturing position l

5. EXPERIMENTS

5.1. Setup

To evaluate performance of our proposed method, we utilize two groups of simulated datasets: focal stack reconstructed by New Standard Light Field Archive [18] and Training set of 4D Light Field Benchmark[19]. Each group of focal stack contains 200 images which are used to reconstruct all-in-focus image for state-of-the-art methods. And the parameter of our experiments are set as belows: while parameters of others are set accordingly. $\lambda = 4$, $\tau_{max} = 20$, $\tau_{min} = 3$, $\sigma_0 = 0.9$, $\sigma_{th} = 1$, $M = 400$, $N_{sup} = 4$.

5.2. Overall Performance

In this section,we compare our method on simulated datasets with DWT-based method[7], MGF-ARF method[4], DMGF-Laplacian method[5] and FSI-based method[15] by SSIM value. SSIM values and numbers of captured images each method used are presented in Table 1. Compared with DMGF-Laplacian based method, we could find that our method achieves comparable scores with only 20% of images. Compared with other state-of-the-art methods, our method achieves higher SSIM value with only 20% images as well. Therefore our scene-adaptive image acquisition system could get high-quality all-in-focus image with fewer images.

Table 1: SSIM of different methods on synthesized data

	ours-num	MGF 200	DMGF 200	DWT 200	FSI 200
card	**0.958**-36	0.946	0.958	0.936	0.759
truck	0.954-24	0.942	**0.956**	0.950	0.855
chess	**0.933**-42	0.898	0.933	0.905	0.825
knights	**0.866**-38	0.835	0.865	0.795	0.668
treasure	0.959-31	0.911	**0.965**	0.830	0.716
bulldozer	0.945-28	0.913	**0.949**	0.860	0.771
boxes	0.976-33	0.962	**0.977**	0.966	0.917
cotton	**0.992**-31	0.970	0.992	0.988	0.960
dino	**0.993**-34	0.973	0.993	0.979	0.938
sideboard	0.968-38	0.942	**0.969**	0.953	0.844

6. CONCLUSION

In this paper, we propose a round-trip scene-adaptive image acquisition system to automatically capture a minimal set of images focused at depth planes of all scene objects to fuse a high quality all-in-focus image. We fuse the image capturing and scene analysis to improve image acquisition efficiency. Our approach maintains better accuracy while reducing the number of captured images.

7. REFERENCES

[1] Elizabeth Allen and Sophie Triantaphillidou, *The manual of photography and digital imaging*, CRC Press, 2012.

[2] Noel T Goldsmith, "Deep focus; a digital image processing technique to produce improved focal depth in light microscopy," *Image Analysis & Stereology*, vol. 19, no. 3, pp. 163–167, 2011.

[3] Anat Levin, Dani Lischinski, and Yair Weiss, "A closed-form solution to natural image matting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 228–242, 2008.

[4] Xuanwu Yin, Guijin Wang, Wentao Li, and Qingmin Liao, "Large aperture focus stacking with max-gradient flow by anchored rolling filtering," *Applied optics*, vol. 55, no. 20, pp. 5304–5309, 2016.

[5] Guijin Wang, Wentao Li, Xuanwu Yin, and Huazhong Yang, "All-in-focus with directional-max-gradient flow and labeled iterative depth propagation," *Pattern Recognition*, vol. 77, pp. 173–187, 2018.

[6] Xuanwu Yin, Guijin Wang, Wentao Li, and Qingmin Liao, "Iteratively reconstructing 4d light fields from focal stacks," *Applied optics*, vol. 55, no. 30, pp. 8457–8463, 2016.

[7] Rafael Redondo, F Šroubek, S Fischer, and Gabriel Cristóbal, "Multifocus image fusion using the log-gabor transform and a multisize windows technique," *Information Fusion*, vol. 10, no. 2, pp. 163–171, 2009.

[8] Mohammad Bagher Akbari Haghighat, Ali Aghagolzadeh, and Hadi Seyedarabi, "Real-time fusion of multi-focus images for visual sensor networks," in *Machine Vision and Image Processing (MVIP), 2010 6th Iranian*. IEEE, 2010, pp. 1–6.

[9] Yu Liu, Shuping Liu, and Zengfu Wang, "Multi-focus image fusion with dense sift," *Information Fusion*, vol. 23, pp. 139–155, 2015.

[10] Samuel W Hasinoff and Kiriakos N Kutulakos, "Light-efficient photography," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2203–2214, 2011.

[11] Samuel W Hasinoff, Kiriakos N Kutulakos, Frédo Durand, and William T Freeman, "Time-constrained photography," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 333–340.

[12] Kyros Kutulakos and Samuel W Hasinoff, "Focal stack photography: High-performance photography with a conventional camera.," in *MVA*. Citeseer, 2009, pp. 332–337.

[13] Daniel Vaquero, Natasha Gelfand, Marius Tico, Kari Pulli, and Matthew Turk, "Generalized autofocus," in *Applications of Computer Vision (WACV), 2011 IEEE Workshop on*. IEEE, 2011, pp. 511–518.

[14] David Choi, Aliya Pazylbekova, Wuhan Zhou, and Peter van Beek, "Improved image selection for focus stacking in digital photography," in *Image Processing (ICIP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 2761–2765.

[15] Sujit Kuthirummal, Hajime Nagahara, Changyin Zhou, and Shree K Nayar, "Flexible depth of field photography," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 1, pp. 58–71, 2011.

[16] Shaojie Zhuo and Terence Sim, "Defocus map estimation from a single image," *Pattern Recognition*, vol. 44, no. 9, pp. 1852–1858, 2011.

[17] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[18] "The (new) stanford light field archive," http://lightfield.stanford.edu.

[19] Katrin Honauer, Ole Johannsen, Daniel Kondermann, and Bastian Goldluecke, "A dataset and evaluation methodology for depth estimation on 4d light fields," in *Asian Conference on Computer Vision*. Springer, 2016.