Analysis of the noise in backprojection light field acquisition and its optimization

NI CHEN,1,2,* ZHENBO REN,2 DAYAN LI,1 EDMUND Y. LAM,2 AND GUOHAI SITU1,3

1Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai 201800, China
2Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong, China
3e-mail: ghsitu@siom.ac.cn
*Corresponding author: nichen@siom.ac.cn

Received 21 November 2016; revised 20 January 2017; accepted 20 January 2017; posted 24 January 2017 (Doc. ID 281234); published 14 February 2017

Light field reconstruction from images captured by focal plane sweeping can achieve high lateral resolution comparable to the modern camera sensor. This is impossible for the conventional micro-lens-based light field capture systems. However, the severe defocus noise and the low depth resolution limit its applications. In this paper, we analyze the defocus noise in the focal-plane-sweeping-based light field reconstruction technique, and propose a method to reduce the defocus noise. Both numerical and experimental results verify the proposed method.

1. INTRODUCTION

A single-view image of a three-dimensional (3D) scene corresponds to the projection of a collection of light rays coming from it, and a light ray with its propagation direction is called a light field [1,2]. Unlike conventional photography, which records only the intensity distribution of the light rays, a light field camera [3] records both the intensity and direction of the light rays [4], enabling view reconstruction of the 3D properties of a scene [5]. In Fig. 1(a), we describe a five-dimensional (5D) light field function [4]. The principal plane of the light field is perpendicular to the optical axis. Many light rays with different directions go through each position on the principal plane, and every ray can be fully described by a 5D function \( I(x, y, \xi, \eta, z) \), where \((x, y)\) is the lateral position of the light field at the principal plane located at depth \(z\). \((\xi = \theta_x, \eta = \theta_y)\) are the projection angles between the light ray and the normal of the principal plane. Usually, a four-dimensional (4D) function without the \(z\) coordinate is enough to represent a light field [6], because the light field can be assumed to propagate along the optical axis. In this paper, we use the 4D light field function. Since the light field records the 3D information of a scene, and because of the storage of the direction of each light ray, it thus can be used for many applications, such as refocusing [3], and auto-stereoscopic 3D display, i.e., integral imaging [7,8]. In addition, it can also be used to synthesize a hologram to eliminate the coherence requirement in hologram recording systems [9–11].

In most light field capture techniques, a camera with a micro-lens array in front of its sensor [3–5] is used, where every micro-lens captures angular distribution of the light rays at its principal point, as Fig. 1(b) shows. The number of light rays that can be recorded depends on the lens pitch \(\Delta_x\), and the pixel pitch \(\Delta_x\) of the camera sensor. The maximum angle \(\theta_{\text{max}}\) of the light rays that can be collected depends on the specification of the micro-lens, i.e., the focal length \(f_l\) and the lens pitch \(\Delta_x\). The spatial sampling interval of the object is the same as the pitch of the lens array. This lens-array-based method enables direct capture of the light field at a single shooting. However, the spatial resolution and angular resolution of the captured light field mutually restrict each other; therefore, the achieved spatial resolution is much lower than that of the image sensor [7,8]. Although several methods have been proposed to enhance the spatial resolution, they usually require solving a computationally heavy inverse problem, sometimes with prior knowledge of the object scene [12,13]. Coded masks inserted into a camera have also been invented to obtain a higher-resolution light field. Although they achieve a better resolution than the lens-array-based techniques, they sacrifice the light transmission because of the masks [14,15].

Recently, it has been reported that the light field can also be obtained from focal plane sweeping captured images with a conventional digital camera [16,17]. These techniques can capture a higher-resolution light field. Examples are the light field moment imaging [16,18] and the light field reconstruction with backprojection (LFBP) approach [17,19]. In these cases, the light field is calculated from several photographic images captured at different focus depths; the images are not segmented by the sublens of the
lens array, hence they can reach a higher angular and spatial image resolution comparable to that of a conventional camera sensor. Note that the angular sampling of the light field calculated from the photographic images depends on the numerical aperture (NA) and the pixel pitch of the camera sensor, rather than the number of images captured along the optical axis. As these methods do not require any special equipment like a lens array, they are easy to be implemented. And, capturing at a fixed camera location reduces the burden of the calibration greatly. Most importantly, the estimated light ray field can have high spatial resolution comparable to that of the image sensor itself. Although the capturing of a focus-stack image sacrifices the temporal resolution to reach a higher spatial resolution, a fast capturing mechanism can be achieved easily, for example, using a spatial light modulator to produce defocus instead of moving the camera [20].

Although the LFBP can reconstruct an exact light field with high angular resolution, severe defocus noise exists in the reconstruction [11,17]. In this paper, we analyze the noise in the reconstructed light field. In addition, we propose a method to suppress the noise. Numerical and experimental data are also presented to verify the proposed method.

2. DEFOCUS NOISE ANALYSIS OF THE FOCAL-PLANE-SWEEPING-BASED LIGHT FIELD RECONSTRUCTION

A. Light Field Reconstruction From Focal Plane Sweeping Captured Images

In the LFBP technique, a series of images along the optical axis are captured while the camera’s focal plane is shifted, as Fig. 2 shows. The focal plane shifting can be achieved by turning the focus ring of the camera. The image $I(x, y, z_m)$ is captured while the focal plane is located at $z = z_m$. The total number of captured images is denoted as $M$. Generally, the focal plane sweeping range should cover the depth range of the 3D scene. With these captured images, the light field with the principal plane located at $z = 0$ is calculated by using the backprojection algorithm [17]:

$$L(x, y, \xi, \eta) = \sum_{m=1}^{M} I(x + z_m \xi, y + z_m \eta, z_m).$$

Here we omit the magnification factor of the images. This is because the captured images can be aligned and resized easily with post-digital processing. The light field reconstructed with this approach has a severe noise problem [11,17]. In order to eliminate the noise, we should study its origin. The mathematical analysis is shown in the following section.

B. Analysis of the Defocus Noise in the LFBP Reconstructed Light Field

We start from considering a 3D object with its center located at the origin of the Cartesian coordinates. Since the energy traveling along a ray is considered as a constant, the light field is the integral of the object projections, as Fig. 3 shows. The light field with the principal plane located at the center of the object thus can be represented as [11]

$$L(x, y, \xi, \eta) = \int O(x + z \xi, y + z \eta, z)dz.$$  

When we capture an image of a 3D scene, suppose the camera focal plane is located at $z = z_m$, the captured images should be the convolution of the clear images of the object and the camera’s point spread function (PSF). The captured image with the camera focal plane at $z = z_m$ is marked as $I(x, y, z_m)$, and its math representation is

$$I(x, y, z_m) = \int O(x, y, z) \otimes h(x, y, z_m - z)dz$$

$$= O(x, y, z_m) + \int_{z \neq z_m} O(x, y, z) \otimes h(x, y, z_m - z)dz,$$  

where $h(x, y, z_m - z)$ is the PSF of the camera, and $\otimes$ is the two-dimensional convolution operator. The first term in the

Fig. 1. (a) Light field definition, and (b) micro-lens-array-based light field capture.

Fig. 2. Scheme for capturing focal plane sweeping images.

Fig. 3. Relation between 3D object and its light field.
last line of Eq. (3) is the clear image of the object slice at the focal plane; the second term is the blurred image contributed by the object slices, which are out of focus. Substituting Eq. (3) in Eq. (1) we obtain the equation of the LFBP reconstructed light field \( L'(x, y, \xi, \eta) \) as

\[
L'(x, y, \xi, \eta) = \sum_{m=1}^{M} O(x + x_m \xi, y + x_m \eta, z_m) + \sum_{m=1}^{M} \int_{\xi \neq x_m} O(x + x_m \xi, y + x_m \eta, z) \otimes h(x + x_m \xi, y + x_m \eta, z_m - z) \, dz.
\]

(4)

In Eq. (4), \( O(x, y, z_m) \) is a slice of a continuous 3D object surface located at a depth of \( z = z_m \). And we know that the discrete approximate representation of a 3D object is \( O(x, y, z) \approx \sum_{m=1}^{N} O(x, y, z_m) \), and \( N \) is the slice number. This equals to sampling the object along the optical axis, and \( N \) is the sampling number; Eq. (2) becomes \( L(x, y, \xi, \eta) = \sum_{m=1}^{N} O(x + x_m \xi, y + x_m \eta, z_m) \). Therefore, when \( M \) approaches to \( N \), the first term in Eq. (4) is approximately equal to Eq. (2), which corresponds to the discrete approximation of the 3D object's light field. When \( M \) is small, it equals to axially sampling the object insufficiently; this affects the depth resolution of the reconstructed light field. The second term is the defocus noise. Obviously, it is the accumulation of the defocus noise induced by the images of the object slices which are out of focus. From this equation, we can see that there are two main parameters affecting the noise: the number of the depth images and the PSF of the camera. The PSF is related to the \( f \)-number of the camera, in other words, the NA. In order to view how the parameters affect the defocus noise, we calculate the noise with respect to the two parameters sequentially.

A 3D object with three planes while each of them has a pixel number of 256 x 256 is used in the calculation. The depth interval between two adjacent object slices is 20 mm, and the center of the object is located at the origin of the chosen Cartesian coordinate. The PSF of the camera is supposed to be a Gaussian function, and its width is determined by the NA of the camera and the distance from the camera focal plane to the object image planes [14,21].

In Fig. 4(a), the peak signal-to-noise ratio (PSNR) of the reconstructed light field is plotted in respect with the number of the captured images. In the numerical captures, the distance from the camera focal plane to the object center has changed from \( z = -25 \) mm to \( z = 25 \) mm. The distance between two photos was modified to capture a different number of images. Several results were obtained according to different camera NAs. The figure shows that the PSNR is decreasing with the increasing number of captured images, i.e., the noise becomes more and more severe with the increasing number of the captured images. This can also be observed from Eq. (4); more images lead to further noise accumulation in the second term. It may be confusing that the PSNRs stagnate when the number of the images is larger than some value. One reason is that, in the simulation, only three plane images were used as the object. A maximum of three captured images is enough for the exact light field reconstruction. Defocus noise on the extra images was introduced by the three planes, which has a upper limit. But this will be different for a continuous real 3D object, since the depth is continuous and there is no upper limit as in the simulation.

In Fig. 4(b), we plot the PSNR of the reconstructed light field with respect to the NA of the camera. Several groups of images with various numbers (NI in the figure) of photos were numerically captured and used for the light field reconstruction. It is obvious that the PSNR is decreasing with the increasing NA of the camera no matter how many photos were used to reconstruct the light field. We can also say the noise is increasing with the increasing NA of the camera. Generally, the PSF of a camera is cone-shaped and symmetrical about the focal plane. A smaller camera NA produces a slimmer cone-shaped PSF, as well as less noise accumulation in the second term of Eq. (4).

From Fig. 4, it is clear that in the LFBP technique, a smaller camera NA and fewer images produce a higher quality reconstructed light field. However, in order to maintain the depth resolution of the reconstructed light field, the number of the captured images should be large enough. This mutual constraint property makes it difficult to get a high-quality light field with the conventional LFBP technique; this can be observed from the original paper [17]. In our simulations, the objects were simple with plane images. This does not satisfy the real-world conditions. For a general object, the complexity of it may affect the results. Since higher frequencies blur faster than lower spatial frequencies, in the case of the camera settings that are given, the objects that are more complex will make the PSNR decrease faster than that of the simpler objects with respect to the number of the captured images. In the following section, we show our improved LFBP technique, which solves this problem.
3. OPTIMIZATION OF THE FOCAL-PLANE-SWEEPING-BASED LIGHT FIELD RECONSTRUCTION

A. Principle of the Optimization Method

In the previous section, we show that the noise in the reconstructed light field comes from the accumulation of defocus noise of the captured images. In order to eliminate the defocus noise, we should detect the defocus noise first. Since we capture the images along the optical axis with the focal plane sweeping approach, the sharp image area in one image will be blurred in all the other images. The amount of changes between two adjacent images reflects the sharpness degree of an image, and the largest amount indicates the clearest image location. Therefore, detecting the maximum change of each pixel along the optical axis can help us find the focus and out-of-focus part in the captured images. In photography deblurring, the Laplace operator is an efficient approach, which detects the gradient changes of an image. Here we apply this technique to achieve our aim. The detected focus pixels in each captured image are then combined as the new images to calculate the light field. Since the redundant defocus noise is omitted before the calculation, the quality of the reconstructed light field is improved.

During the preprocessing, the captured images are treated as an image stack $I_s(x, y, z)$:

$$I_s(x, y, z) = \begin{bmatrix} I(x, y, z_1) \\ \vdots \\ I(x, y, z_M) \end{bmatrix}.$$

The Laplace operator is used to detect edges in the images. However, it is sensitive to discrete points and noise. Therefore, we filter the images with a Gaussian filter to reduce the noise; this can increase the robustness of the Laplace operator:

$$E_s(x, y, z) = \nabla^2 I_s(x, y, z) \otimes G(x, y),$$

where $G(x, y) = \frac{1}{\pi \sigma^2} \exp\left( -\frac{1}{\pi \sigma^2} (x^2 + y^2) \right)$ is the Gaussian filter to smooth the images. The Gaussian kernel size depends on what is desired. Large kernel size detects large scale edges, and small kernel size detects fine features. As the derivative is used to measure changes, the derivative having maximum magnitude is the information we are looking for:

$$z_{\text{max}}(x, y) = \arg \max_z \{E_s(x, y, z)\},$$

where $z_{\text{max}}(x, y)$ is the position where $E_s(x, y, z)$ has the maximum value. The image stack after the preprocessing thus is

$$I_s'(x, y, z') = I_s(x, y, z_{\text{max}}(x, y)).$$

The new image stack $I_s'(x, y, z')$ is used to synthesize the light field with the same method as in the conventional method.

It should be noted that the LFBP-based techniques are different from Mousnier’s method, in which a focus map, depth map, and epipolar-plane images (EPIs) were calculated sequentially from focus-stack images [19]. In the LFBP-based techniques, the known parameters are only the camera settings and the depth interval of the images, and the final reconstruction is the light field of the 3D scene. On the contrary, Mousnier’s method requires a good prior knowledge, such as what background the scene is and what the estimated projection angles of each pixel in the captured images are. Most importantly, Mousnier’s method aims at reconstructing the EPIs, and this relies on the special structure of the EPIs. The reconstructed EPIs are only 2D slices of the 4D light field. Although they can be used for approximately refocusing and perspective viewing, it is insufficient for representing a 3D scene.

4. SIMULATION VERIFICATION

A 3D object scene with three plane images is used to test our method. The lateral size of each plane is 128 mm × 128 mm, and the distance between two plane images is 20 mm. Since the center of the object is located at the origin of the coordinates, the depths of the three planes are -20, 0, and 20 mm, respectively. In the simulation, the exact light field of the object scene can be obtained by projecting all the pixels to the principal plane with Eq. (2). This can be used as the ground truth for comparison. Figure 5(b) is one of the EPI $(L(x, 0, \xi, 0))$ profiles of the light field.

The captured photographic images are calculated with Eq. (3). The light field is reconstructed with the conventional and the proposed methods, respectively. In Figs. 6(a) and 6(b) are the conventional and proposed light field reconstructions with different numbers of captured images. Figures 7(a) and 7(b) are the conventional and proposed light field reconstructions with different camera NAs. As explained in Section 2, the noise in the conventional reconstruction gets worse with increasing number of the captured images and camera NAs. Conversely, the number of the captured images and camera NAs does not affect the reconstruction of the proposed method.

The visualizations in Figs. 6 and 7 are only a slice of the 4D light field; therefore, we also compare the refocused images from the whole light field. A light field with dimensions of $256 \times 256 \times 50 \times 50$ calculated from five photographic images...
are used to perform the refocusing. Figure 8 shows the refocused images located at the three depths of the original object planes, while Figs. 8(a)–8(c) are corresponding to the exact, the conventional, and the proposed method, respectively. From Figs. 8(a) and 8(b), it can be observed that the focused letters at each plane were blurred very much with the conventional method compared to the ground truth. On the contrary, the refocused images with the proposed method in Fig. 8(c) reach clear focus images comparable to the one in Fig. 8(a).

A. Experimental Verification

Our proposed method has also been verified with the real captured images. Figure 9 is the image of the objects used in the experiment. A penguin doll and a flower were separated with a depth distance of 100 mm. A Canon EOS 1100D camera with a lens of EF 50 mm f/1.2L USM was used to take the photos. The f-number of the camera is 2.5 and the sensor pixel pitch is 3.1 μm. Three groups of images were taken with 3, 5, and 11 photos in each group. All the images were cropped to a resolution of 500 × 400 pixels for reducing the computational load.

Figure 10 shows the reconstructed EPI images of the light field from a various number of photos, where Figs. 10(a) and 10(c) were performed with the conventional method and Figs. 10(b) and 10(d) with the proposed method. Figures 10(a) and 10(b) represent the reconstructed light field of \( L(x, 0.6y_{max}, ξ, 0) \) and Figs. 10(c) and 10(d) represent \( L(x, 0.2y_{max}, ξ, 0) \). In Fig. 10(a), the edges of the white areas, which reflect the slopes, were zoomed and marked with red dashed lines. From Fig. 10(a) we can observe that the edges become blurry as the number of the images increases. We even cannot observe the edge of the bottom part on the right image \( (N = 11) \) of Fig. 10(a). Conversely, the EPI images of the proposed reconstructed EPIs are affected slightly by the number of the captured images, as the red dashed lines in Fig. 10(b) indicate. The captured images and the corresponding refocused images from the light fields are shown in Fig. 11. Figure 11(a) shows the captured images, and Figs. 11(b) and 11(c) represent the conventional and proposed refocused images, respectively. Figure 12 shows several selected reconstructed parallax images with the conventional and proposed methods, respectively. The parallax can be observed from both the conventional and proposed methods, as the rectangle areas in each image show. However, the blur in the conventional method is severe, and this becomes more harmful as the view angle increases. The videos are shown by Visualization 1 and Visualization 2, respectively. As we expected, the proposed refocused and parallax images perform better quality than the conventional one.
increasing number of the captured images and the NA of the camera. These are the reasons that limit the application of the light field that is calculated with the conventional focal plane sweeping technique. Based on the analysis, we proposed a method to optimize the reconstructed light field by a previous digital deblurring process on the captured photographic images. The proposed method almost eliminates the noise in the reconstructed light field no matter how many captured images we used to calculate it. The simulation and experimental results verified our proposed method. But we should note that the occlusion is an unsolved problem in this paper, since for one spatial coordinate of the light field, only one pixel in the focal-stack has been used to recover it. But it is believed that this can be solved with a more sophisticated backprojection algorithm.

Funding. National Natural Science Foundation of China (NSFC) (61327902, 61377005); Recruitment Program of Global Youth Experts; Research Grants Council of the Hong Kong Special Administrative Region, China (HKU714/13).

5. CONCLUSIONS

We analyzed the noise in the light field reconstruction based on the focal plane sweeping technique. From the analysis, we found that the noise in the reconstructed light field is coming from the accumulation of the defocus noise in the captured photographic images. Therefore it becomes severe with the

REFERENCES

1. A. Gershun, “The light field,” J. Math. Phys. 18, 51–151 (1936).
2. E. Y. Lam, “Computational photography with plenoptic camera and light field capture: tutorial,” J. Opt. Soc. Am. A 32, 2021–2032 (2015).
3. R. Ng, “Fourier slice photography,” ACM Trans. Graph. 24, 735–744 (2005).
4. M. Levoy, Z. Zhang, and I. McDowall, “Recording and controlling the 4D light field in a microscope using microlens arrays,” J. Microsc. 235, 144–162 (2009).
5. R. Ng, “Digital light field photography,” Ph.D. thesis (Stanford University, 2006).
6. M. Levoy and P. Hanrahan, “Light field rendering,” in Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH) (ACM, 1996), pp. 31–42.
7. J. Hong, Y. Kim, H.-J. Choi, J. Hahn, J.-H. Park, H. Kim, S.-W. Min, N. Chen, and B. Lee, “Three-dimensional display technologies of recent interest: principles, status, and issues [invited],” Appl. Opt. 50, H87–H115 (2011).
8. S.-G. Park, J. Yeom, Y. Jeong, N. Chen, J.-Y. Hong, and B. Lee, “Recent issues on integral imaging and its applications,” J. Inf. Disp. 16, 37–46 (2014).
9. J.-H. Park, M.-S. Kim, G. Baasantseren, and N. Kim, “Fresnel and Fourier hologram generation using orthographic projection images,” Opt. Express 17, 6320–6334 (2009).
10. N. Chen, J.-H. Park, J. Yeom, J. Kim, G. Li, and B. Lee, “Fourier hologram synthesis from two photographic images captured at different focal planes,” in Signal Recovery and Synthesis (Optical Society of America, 2014), paper JTu4A-6.
11. N. Chen, Z. Ren, and E. Y. Lam, “High resolution Fourier hologram synthesis from photographic images through computing the light field,” Appl. Opt. 55, 1751–1756 (2016).
12. N. Chen, J.-H. Park, and N. Kim, “Parameter analysis of integral Fourier hologram and its resolution enhancement,” Opt. Express 18, 2152–2167 (2010).
13. N. Chen, J. Yeom, J.-H. Jung, J.-H. Park, and B. Lee, “Resolution comparison between integral-imaging-based hologram synthesis methods using rectangular and hexagonal lens arrays,” Opt. Express 19, 26917–26927 (2011).
14. A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, “Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing,” ACM Trans. Graph. 26, 69 (2007).
15. K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, “Compressive light field photography using overcomplete dictionaries and optimized projections,” ACM Trans. Graph. 32, 1–12 (2013).
16. A. Orth and K. B. Crozier, “Light field moment imaging,” Opt. Lett. 38, 2666–2668 (2013).
17. J.-H. Park, S.-K. Lee, N.-Y. Jo, H.-J. Kim, Y.-S. Kim, and H.-G. Lim, “Light ray field capture using focal plane sweeping and its optical reconstruction using 3D displays,” Opt. Express 22, 25444–25454 (2014).
18. J. Liu, T. Xu, W. Yue, J. Sun, and G. Situ, “Light-field moment microscopy with noise reduction,” Opt. Express 23, 29154–29162 (2015).
19. A. Mousnier, E. Vural, and C. Guillemot, “Partial light field tomographic reconstruction from a fixed-camera focal stack,” ArXiv 1503.01903 (2015).
20. H. Wang, N. Chen, J. Liu, and G. Situ, “Light field imaging based on defocused photographic images,” in Digital Holography and 3-D Imaging Meeting (submitted).
21. J.-H. Park, S.-W. Seo, N. Chen, and N. Kim, “Fourier hologram generation from multiple incoherent defocused images,” Proc. SPIE 7690, 76900F (2010).