

Article

Feature Detection of Focused Plenoptic Camera Based on Central Projection Stereo Focal Stack

Qingsong Liu ^{1,2}, Xiaofang Xie ^{1,*}, Xuanzhe Zhang ², Yu Tian ³, Yan Wang ² and Xiaojun Xu ²

¹ Longshore Defense Troop Academy, Naval Aeronautical University, Yantai 264001, China; liuqingsong17@nudt.edu.cn

² State Key Laboratory of Pulsed Power Laser Technology, College of Advanced Interdisciplinary Studies, National University of Defense Technology, Changsha 410073, China; zhangxz_zzh@163.com (X.Z.); wangyan101712@163.com (Y.W.); xuxiaojun@nudt.edu.cn (X.X.)

³ State Key Laboratory of High Performance Computing (HPCL), College of Computer, National University of Defense Technology, Changsha 410073, China; tianyu10@nudt.edu.cn

* Correspondence: Xiexf208@sina.com

Received: 14 August 2020; Accepted: 26 October 2020; Published: 29 October 2020



Abstract: Fast and accurate feature extraction can lay a solid foundation for scene reconstruction and visual odometry. However, this has been rather a difficult problem for the focused plenoptic camera. In this paper, to the best of our knowledge, we first introduce an accurate and fast feature extraction algorithm based on central projection stereo focal stack (CPSFS). Specifically, we propose a refocusing algorithm that conforms to the central projection with regard to the center of main lens, which is more accurate than traditional one. On this basis, the feature is extracted on the CPSFS without calculating dense depth maps and total focus images. We verify the precision and efficiency of the proposed algorithm through simulated and real experiments, and give an example of scene reconstruction based on the proposed method. The experimental results show that our feature extraction algorithm is able to support the feature-based scene reconstruction via focused plenoptic camera.

Keywords: focused plenoptic camera; feature detection; central projection; refocused image; CPSFS

1. Introduction

Plenoptic (or light field) cameras have become increasingly developed and widespread in recent years. In contrast with traditional pinhole cameras, they capture the light distribution both in spatial and angular dimension by the help of the inserted micro-lens array (MLA). According the MLA placement, plenoptic cameras can be classified into unfocused [1] and focused plenoptic cameras [2]. In an unfocused plenoptic camera, the main lens focuses the subject into the MLA (like Lytro). Furthermore, based on the position of MLA, focused plenoptic cameras can be classified into two types with the MLA behind (Keplerian) or in front of (Galilean) the focused main lens image. As a special case of focused plenoptic cameras, multi-focus plenoptic cameras use multiple types of micro-lenses (like Raytrix) to further extend the depth-of-field. Due to the better ability to restore depth information compared with unfocused plenoptic camera, the focused plenoptic camera is gaining more attention in fields like structure from motion (SFM) and visual odometry.

Robust and accurate feature detection methods are the foundation of many SFM algorithms [3–5]. At present, many works have been conducted with regard to feature extraction of the focused plenoptic camera. Bok et al. [6] propose a checkerboard line feature detector performed on the raw image, and Nousias et al. [7] present a checkerboard corner detector. Liu et al. [8] introduce an adaptive checkerboard corner detector, which can select the sharpest corners automatically. However, the methods mentioned [6–8] are designed for specific feature pattern, which is hard

to use in SFM. Ferreira et al. [9] use the scale-invariant feature transform (SIFT) detector [10] to extract features and calculate depth on the raw image. However, the small size of micro-image makes the results less robust and accurate. Dansereau et al. [11] propose a LIFF feature detector for the unfocused plenoptic camera based on focal stack. However, the locations of detected features are restrict to the central sub-aperture image other than the whole raw image. Kühfuß et al. [12] carry out SURF feature detector directly on total focus image. However, the computation of dense depth image and total focus image is very time-consuming. Although it is easier to extract features from the total focus image, the traditional refocusing algorithm of focused plenoptic camera has some limitations. The traditional refocusing algorithm is based on the concept of virtual depth [2], which is widely used in the calculation of depth maps and total focus image [9,13–17]. However, the theoretical imaging range of the refocused image based on traditional method [2] is limited by the sensor size and the traditional focal stack do not conform to the central projection in theory. Besides, most methods [9,13–17] ignore the difference between the center of micro-lens and micro-image when calculating depth maps and total focus images, which will result in the deviation of calculated results from the theoretical model.

In order to extract features more accurately and efficiently, we perform an innovative feature extraction on the central projection stereo focal stack (CPSFS). Specifically, we present a refocusing model that conforms to the central projection with regard to the center of main lens. The calculated refocused image is strictly consistent with the theoretical model. Then, we propose the feature detection algorithm performed on CPSFS based on the stereo focal stack proposed by Hog et al. [17]. We analyze the relationship between the layer selection of CPSFS and the detection results through simulated experiments. Furthermore, the proposed feature detection method is tested on both simulated and real data. The experiment results indicate that the proposed detection method has good performance in speed, accuracy and noise condition. A scene reconstruction example is also conducted to demonstrate that the proposed method is able to support the scene reconstruction via focused plenoptic camera.

The main innovations of this paper are as follows.

(1) An accurate refocusing model which conforms to central projection with regard to the center of main lens.

(2) A fast and accurate feature extraction algorithm based on CPSFS, which can support an efficient feature-based SFM via focused plenoptic camera without calculating dense depth map and total focus image.

2. Refocusing Model Conforms to Central Projection

For the convenience of the following discussion, we first explain the coordinate systems and relevant symbols used in this paper. The camera coordinate system $OXYZ$ is established with the origin selected as the center of the main lens and the Z -axis set as the optical axis. The establishment of the raw image pixel coordinate system $ouvw$ and the refocused image pixel coordinate system $o'st$ is shown in Figure 1. b represents the displacement from the main lens to the sensor and B stands for the displacement from the micro lens array (MLA) to the sensor. Note that both b and B are negative. Besides, f_L indicates the focal length of the main lens, which is positive.

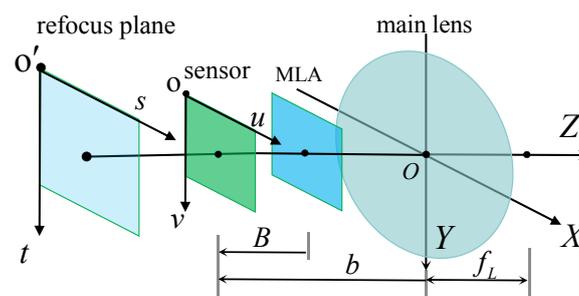


Figure 1. Illustration of coordinate system.

2.1. Problems of the Traditional Refocusing Algorithm

For the traditional refocusing algorithm [2], given the virtual depth v_F of any refocusing point F , the projection point p on the raw image can be calculated through the following equation.

$$\begin{bmatrix} 1 & 0 & (p_u - l_u) \\ 0 & 1 & (p_v - l_v) \end{bmatrix} \begin{bmatrix} F_u & F_v & v_F \end{bmatrix}^T = \begin{bmatrix} l_u \\ l_v \end{bmatrix}. \quad (1)$$

In Equation (1), (p_u, p_v) represents the pixel coordinates of p in the ouv coordinate system, (l_u, l_v) illustrates the orthographic projection coordinates of micro-lens' center related to p in ouv , and (F_u, F_v) stands for the orthographic projection coordinates of F in ouv . By weighting the pixel values of all projection points, the pixel value of F in the refocused image can be obtained. However, due to the use of orthographic projection coordinates and virtual depth, the traditional refocusing model has the following problems.

2.1.1. Limited Theoretical Imaging Range

For any point F on a given refocus plane, the traditional method restricts (F_u, F_v) within the orthographic range of the sensor in order to keep the size of refocused images remains the same. Therefore, the imaging range of the refocus plane is limited. For example, although the raw image contains multiple imaging point of Q' , not all refocused images can image Q' , as shown in Figure 2a.

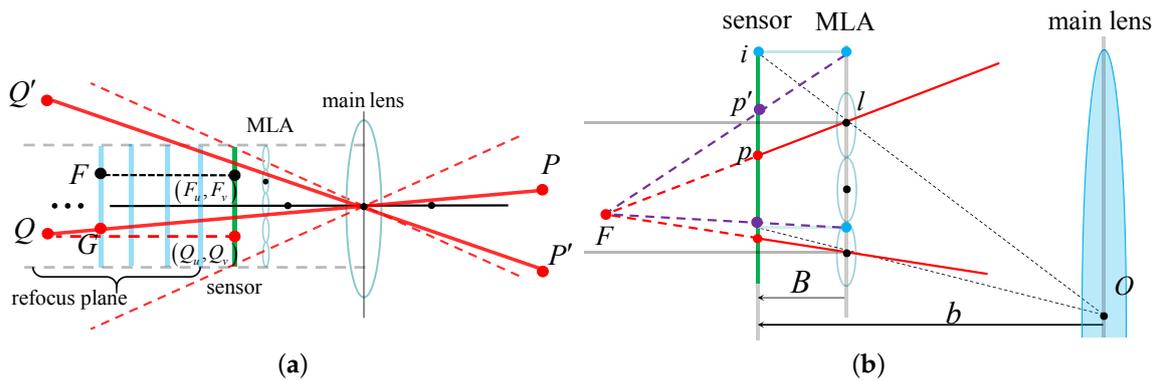


Figure 2. Illustration of traditional refocusing algorithm. (a) Traditional refocusing model. (b) Illustration of projection error.

2.1.2. Diverse Theoretical Imaging Positions of Same Object among Focal Stack

For a raw image with $N_u \times N_v$ pixels, the size of the refocused image can be determined as (dN_u, dN_v) with d representing the scale factor. Given an arbitrary imaging point Q with coordinates (Q_u, Q_v) in ouv , let G denote the center of the imaging circle of Q in certain refocus plane, as shown in Figure 2a. The pixel coordinates of G in $o'st$ is

$$(G_s, G_t) = d \frac{G_z}{Q_z} (Q_u, Q_v), \quad (2)$$

with (G_x, G_y, G_z) and (Q_x, Q_y, Q_z) standing for the coordinates of G and Q in $OXYZ$, respectively. When d is fixed, (G_s, G_t) changes with G_z . This makes the theoretical imaging positions of the same point differ in the focal stack, which makes the feature matching among traditional focal stack more difficult.

2.1.3. Deviation between Calculated Result and Theoretical Model

As the central coordinates of micro-lens cannot be obtained accurately before calibration, traditional refocusing method [2] use the central coordinates of micro-image instead during actual

process, which makes the calculated refocused (total focus) image inconsistent with the theoretical model. As shown in Figure 2b, the correct projection point corresponding to F is p , while the actual point used is p' . This will bring errors to the subsequent SFM algorithm during scene reconstruction.

2.2. Central Projection Refocusing Model Based on Plenoptic Disc Data

The plenoptic disc data is proposed by O'Brien et al. [18] in 2018, which is used to improve the calibration accuracy of the plenoptic camera. As shown in Figure 3a, (F_u^S, F_v^S) represents the central projection coordinates of F in ouv with regard to the center of the main lens. $|R|r_{mi}$ stands for the maximum pixel distance between (F_u^S, F_v^S) and the center of micro-image which contains the projection point of F , with r_{mi} denoting the pixel radius of micro-image. The vector (F_u^S, F_v^S, R_F) is called the plenoptic disc data of F , while (F_u^S, F_v^S) and R_F are called the plenoptic disc center and plenoptic disc radius, respectively. The R_F can be calculated by

$$R_F = b(b - B - F_z) / (F_z B), \tag{3}$$

with F_z stands for the Z-coordinate of F in $OXYZ$. The relation between R_F and v_F is

$$R_F = bv_F / (b - B - v_FB) \tag{4}$$

By the help of plenoptic disc data, the calculated focal stack will conform to the central projection with regard to the center of main lens, as shown in Figure 3b. Given the (F_u^S, F_v^S, R_F) of F , the projected point (p_u, p_v) on raw image can be obtained by

$$\begin{bmatrix} 1 & 0 & (p_u - i_u) \\ 0 & 1 & (p_v - i_v) \end{bmatrix} \begin{bmatrix} F_u^S & F_v^S & R_F \end{bmatrix}^T = \begin{bmatrix} i_u \\ i_v \end{bmatrix}. \tag{5}$$

In Equation (5), (i_u, i_v) denotes the central coordinates of the micro-image with regard to (p_u, p_v) in ouv . As illustrated in Figure 3b, the rendered refocused image based on Equation (5) can make full use of the light recorded by the sensor at any refocus plane. This can be done simply by restricting the central projection coordinates (F_u^S, F_v^S) other than the orthographic coordinates (F_u, F_v) within the range of raw image. Besides, given any imaging point Q , the pixel coordinates (G_s, G_t) of G in $o'st$ is

$$(G_s, G_t) = d(Q_u^S, Q_v^S), \tag{6}$$

with (Q_u^S, Q_v^S) representing the central projection coordinates of Q in ouv . When d is fixed, the coordinates (G_s, G_t) among different refocused images are strictly the same. What is more, the use of central coordinates of micro-images in Equation (5) makes the calculated results in accordance with the theoretical model without the help of camera calibration.

To sum up, the central projection refocused images have better characteristics than the traditional one, which can provide support for the accurate feature extraction.

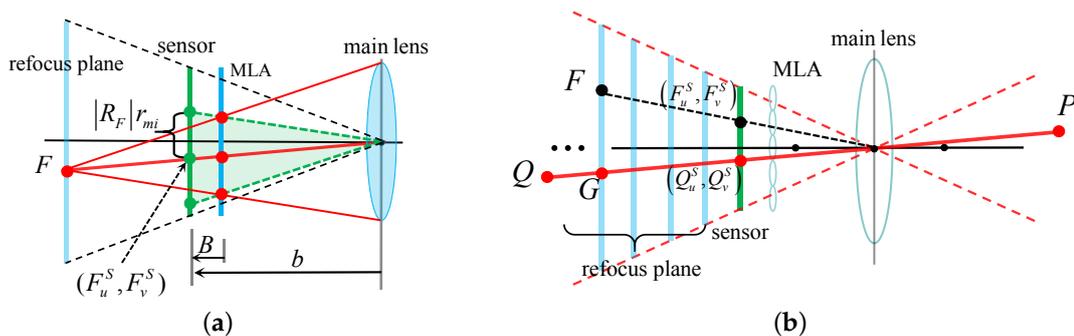


Figure 3. Illustration of proposed refocusing algorithm. (a) Illustration of plenoptic disc data. (b) Illustration of central projection refocusing model.

2.3. Fast Central Projection Refocused Image Rendering Based on Micro-Image

In order to increase the rendering speed, we fill out the refocused image with the micro-images extracted from the raw image. Specifically, we adopt the ray-per-pixel approximation [19] and assume each pixel in the micro-image records the intensity of a single light ray. For any micro-image, let the coordinates of any non-central point in ouv be (p_u, p_v) . Then, we can get Equation (7) according to the similar triangle theory.

$$\begin{pmatrix} \frac{(p_u-i_u)s_x}{(F_s^p-F_s^i)s_x^F} & \frac{(p_v-i_v)s_y}{(F_t^p-F_t^i)s_y^F} \end{pmatrix} = \begin{pmatrix} \frac{B}{F_z-b+B} & \frac{B}{F_z-b+B} \end{pmatrix} \quad (7)$$

In Equation (7), (s_x, s_y) and (s_x^F, s_y^F) indicate the physical size of single pixel in the raw image and the refocused image, respectively, which contains F . (F_s^p, F_t^p) and (F_s^i, F_t^i) represent the pixel coordinates with regard to point F^p and F^i in $o'st$, respectively, as demonstrated in Figure 4. Based on the similar triangle theory, the (s_x^F, s_y^F) can be obtained as $(s_x^F, s_y^F) = (F_z s_x / (db), F_z s_y / (db))$. Furthermore, we can obtain Equation (8) based on Equations (3) and (7).

$$(F_s^p, F_t^p) - (F_s^i, F_t^i) = -dR_F[(p_u, p_v) - (i_u, i_v)] \quad (8)$$

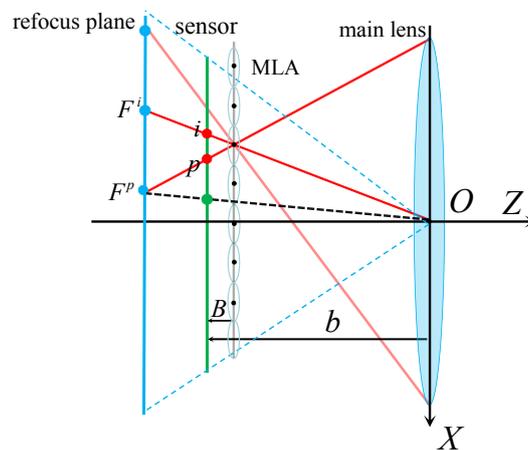


Figure 4. Illustration of relation between micro-image and refocused image.

Intuitively, the light rays recorded by a single micro-image correspond to a circular area on the refocused image. The pixel radius of this circular area is $d|R_F|$ times larger than that of the micro-image in the raw image. In this paper, we traverse all the micro-images and fill out the refocused image with the resized circle patch extracted from the micro-image. After the traversal is completed, the pixel value of the refocused image should be weighted accordingly.

3. Feature Detection Based on CPSFS

As multiple features on the raw images are associated with the same imaging point, the feature detection of the focused plenoptic camera is essentially the extraction of the imaging point, which contains the calculation of its coordinates (plenoptic disc data) and descriptor. It is easy to extract the plenoptic disc center and feature descriptor of the imaging point in the central projection focal stack, but this is not the case for plenoptic disc radius. Hog et al. [17] propose the stereo focal stack, which is formed by using only half of the micro-image during refocusing. The depth information can be obtained utilizing the parallax among the stereo focal stack. However, Hog et al. [17] still use the traditional refocusing model and ignore the difference between the center of micro-lens and micro-image, which means that the actual calculated results are not strictly consistent with the theoretical model. Considering this, we propose a fast and accurate feature detection algorithm based on CPSFS.

In this way, the plenoptic disc radius can be calculated using the parallax from central projection stereo refocused image pair. However, in practice, the calculated R_Q is error prone due to noise and mismatching. Therefore, a more robust way is to use CPSFS with two or more layers. The details are described in Sections 3.2 and 3.3.

3.2. Calculation of Plenoptic Disc Data Based on Two-Layer CPSFS

Based on Equation (10), we can get the two parallax $\epsilon_Q^1, \epsilon_Q^2$ of Q in two different refocus planes with plenoptic disc radius R_F^1 and R_F^2 . Furthermore, the R_Q can be obtained by

$$R_Q = (R_F^2 \epsilon_Q^1 - R_F^1 \epsilon_Q^2) / (\epsilon_Q^1 - \epsilon_Q^2). \quad (12)$$

In practice, we use SIFT detector (not limited to SIFT) to extract features in the central projection stereo refocused image pair and perform feature matching along the s-axis. Let the detected features be C^L, C^R, C'^L , and C'^R in the two layers, as illustrated in Figure 5b. Thus, we have $\epsilon_Q^1 = C_s^L - C_s^R$ and $\epsilon_Q^2 = C_s'^L - C_s'^R$. Due to error detection and mismatching, we should limit the ϵ_Q^1 and ϵ_Q^2 within the valid range. Therefore, we restrict $|R_Q - R_F| < \theta_R$ and have $|\epsilon_Q| < 8dr_{mi}\theta_R / (3\pi)$ based on Equation (11).

Then, the matched feature pairs C^L, C^R and C'^L, C'^R from different layers in CPSFS should be matched. This is rather simple for CPSFS. As illustrated in Figure 5b, point G and G' are located in the center of relevant matched feature pairs C^L, C^R and C'^L, C'^R . According to Equation (6), we can have

$$(Q_u^S, Q_v^S) = \frac{1}{d} \left(\frac{C_s^L + C_s^R}{2}, \frac{C_t^L + C_t^R}{2} \right) = \frac{1}{d} \left(\frac{C_s'^L + C_s'^R}{2}, \frac{C_t'^L + C_t'^R}{2} \right). \quad (13)$$

Therefore, given a fixed d , the plenoptic disc data of SIFT feature point Q can be calculated by Equations (12) and (13). The descriptor of Q is the average value of the descriptors of C^L, C^R, C'^L , and C'^R . In order to ensure the number and accuracy of the detected features, the two layer of the CPSFS should be relatively close to the actual imaging point Q .

3.3. Calculation of Plenoptic Disc Data Based on Multi-Layer CPSFS

In practice, there is usually a depth range for real scene. As detection methods based on two-layer CPSFS cannot effectively detect the features across the scene, we solve this problem by using CPSFS with multiple layers.

Specifically, there is N layers of refocus planes with plenoptic disc radius $R_m = R_0 + m\Delta R$, $m = 1, \dots, N$ ($N > 2, \Delta R > 0$). For each two layers out of N layers of refocus planes, the plenoptic disc data of SIFT features are calculated using the algorithm in Section 3.2. In order to avoid repeated detection, the detected features with close plenoptic disc center and similar descriptors are regarded as the same. Therefore, we calculate the average plenoptic disc data and descriptors of all repeated features.

The choice of parameters R_0, N , and ΔR directly affects the detected results. Generally speaking, when ΔR is about 1–2, the algorithm can get good results and speed. The choice of R_0 and N should rely on the working distance of camera. Empirically speaking, R_0 is usually about 3–5 and N is about 5–8. The detailed simulation is shown in Sections 4.1.2 and 4.1.3.

4. Experiments

In this part, simulated and real experiments are carried out to demonstrate the precision and efficiency of the proposed method. The experimental results are obtained on Windows 7 operating system with Inter Core i7-7700 CPU (3.6 GHz). All the codes are available online <https://github.com/samliu0631/Feature-Detection-using-CPSFS>.

4.1. Simulated Experiments

We simulate the raw images of multi-focused pleonptic camera based on forward ray tracing [20]. On this basis, the proposed central projection refocusing algorithm is compared with the traditional one [2] using simulated data. Besides, the performance of the proposed feature detection method with regard to accuracy, speed, and anti-noise capacity is tested on simulated raw images.

4.1.1. Comparison of Refocusing Algorithm

The simulated parameters of the multi-focused pleonptic camera are shown in Table 1. As all parameters are known in advance, including the positions of micro-lenses, the refocused images can be generated strictly in accordance with the traditional refocusing algorithm [2] and ours.

Table 1. Simulated parameters of multi-focus pleonptic camera. (s_x, s_y) stands for the physical size of a pixel in raw image. f_{m1} , f_{m2} , and f_{m3} represent the focal lengths of three types of micro-lenses.

f_L (mm)	b (mm)	B (mm)	s_x (μm)	s_y (μm)	f_{m1} (mm)	f_{m2} (mm)	f_{m3} (mm)
100	-104.5	-1.32	5.5	5.5	1.62	1.92	2.35

In order to display the results more clearly, the checkerboard plane is used as the imaging target, which is set at the position 0.9m away from the main lens and perpendicular to the optical axis. The simulated raw images are shown in the Figure 6a,b. During the experiment, the traditional refocusing algorithm [2] is used to render the refocused image at planes with virtual depth $v_F \in \{-6, -12\}$, as illustrated in Figure 6c,d. According to Equation (4), the corresponding value of pleonptic disc radius is $R_F \in \{-5.64, -10.54\}$. The rendered images based on our central projection refocusing algorithm are shown in the Figure 6e,f.

It can be seen that the imaging range of traditional refocused image becomes smaller with the increase of $|v_F|$, while that of the proposed method maintain the same regardless of the change of R_F . What is more, this results in the fact that the coordinates of the imaging center corresponding to the same object point in our refocused image remain the same in spite of the position of the refocus plane, while that is not the case for traditional method. As one can see, the experimental results are consistent with the analysis in Section 2.

During the actual process of calculating refocused images or total focus images, most methods [9,13–17] use the center of micro-image to approximate the center of micro-lens. Essentially, this is equivalent to refocusing based on Equation (5) other than Equation (1). If the concepts of virtual depth and orthographic coordinates are still used during reconstruction without further rectification, it will bring errors to the final reconstructed result. To verify this, the Equation (5) is used to calculate the total focus image ($R_F = -6.55$) corresponding to the simulated raw image. Then, the checkerboard corners on the total focus image are converted to real space based on the concept of traditional model and ours. The reconstruction results are shown in the Figure 7. It is obvious that the results of traditional methods deviate from the ground truth while ours are not.

The experimental results demonstrate that the central projection refocusing model in this paper is more consistent with the actual results.

4.1.2. Relation between Layer Distance in CPSFS and Detected Results

The purpose of the experiment in this part is to verify the relationship between layer distance selected in CPSFS and feature detection results. To simplify the problem, the detection method with two-layer CPSFS is used during the experiment. The simulated parameters of the camera are shown in Table 1 and we use the textured pictures to simulate the plane targets. The evaluated indicators include detection speed, and the number and depth accuracy of detection points.

During the experiment, the planar target is placed at a fixed position away from the camera and perpendicular to the optical axis, which will form an imaging plane at the position with pleonptic disc radius R_{gt} . Two refocus planes in the CPSFS are selected at $R_{gt} + \Delta R/2$ and $R_{gt} - \Delta R/2$ respectively.

We change the value of ΔR within the range $\{\Delta R|\Delta R = 0.5i, i = 1, \dots, 12\}$ and observe the change of the detection results.

For generality of the results, we carry out 20 independent feature detection experiments on 20 different simulated raw images for each ΔR . The evaluation indexes include the total number of detected features, the mean errors of the detected features' plenoptic disc radii, the root mean square errors (RMSEs) of the detected features' plenoptic disc radii, and the detection time. Besides, we repeat the above experiment 3 times with the R_{gt} set to -4 , -8 , and -12 , respectively. The other parameters used are set as $d = 1/2, \theta_R = 5$. The results are presented in Figure 8.

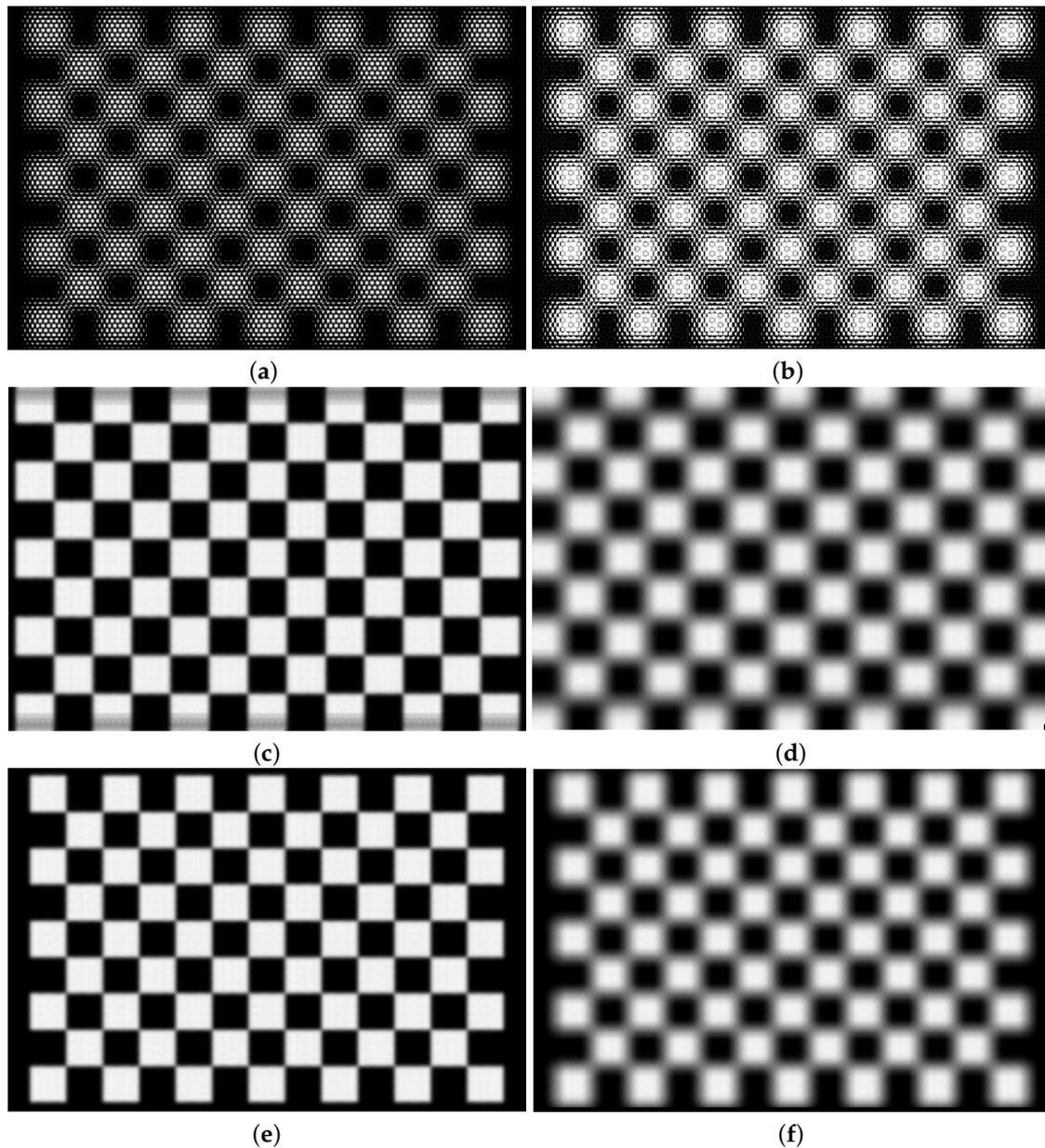


Figure 6. Simulated results of refocusing experiment. (a) Simulated raw image with vignetting. (b) Simulated raw image without vignetting. (c) Traditional refocused image at $v_F = -6$. (d) Traditional refocused image at $v_F = -12$. (e) Our refocused image at $R_F = -5.64$. (f) Our refocused image at $R_F = -10.54$.

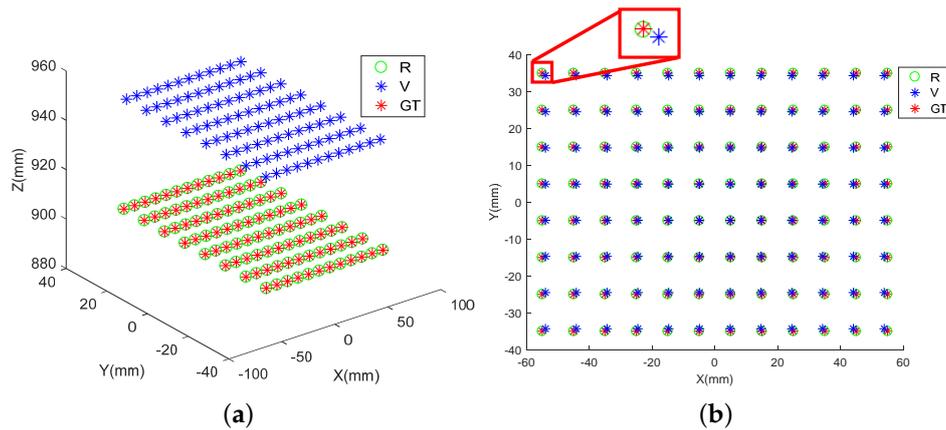


Figure 7. Reconstructed results of simulated experiment. (a) Side view of reconstructed results. (b) Top view of reconstructed results. *R* represents results using plenoptic disc radius. *V* indicates the results utilizing concept of virtual depth. *GT* stands for ground truth.

It can be seen that the trends of three curves in Figure 8 are similar, which reflects the general relationship between layer distance of CPSFS and detected results. When the positions of two refocus planes are close to the R_{gt} (ΔR is smaller), both the number of detected feature points and the RMSE of the detected features' plenoptic disc radius are large. This is because the closer the refocusing planes are to R_{gt} , the clearer the refocusing image, which makes it easier to detect more feature points. However, at the same time, the parallax among the central projection stereo refocused image pair is relatively small, which leads to the increase of calculation error in plenoptic disc radius. As can be seen from Figure 8b, the mean error is close to 0, which indicates that the proposed algorithm is unbiased. Figure 8d illustrates that the proposed algorithm is time efficient and the mean detection time of each frame is about 20 s. The example of detection result is illustrated in Figure 9.

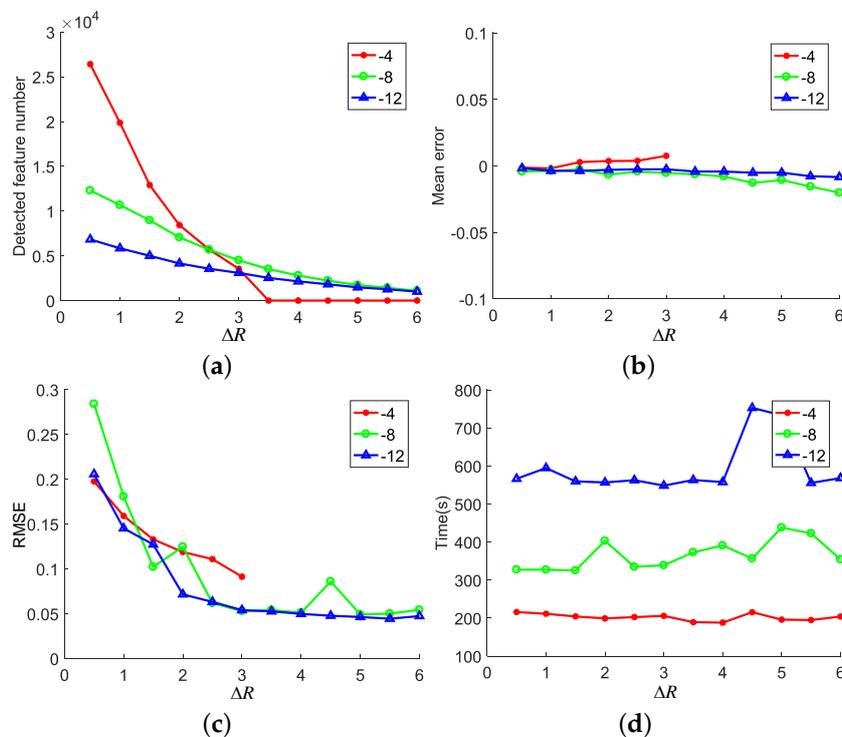


Figure 8. Detected results of proposed algorithm using 2-layer central projection stereo focal stack (CPSFS) on simulated raw images. (a) The number of the detected features. (b) Mean error of detected features' plenoptic disc radius (dimensionless). (c) Root mean square error (RMSE) of detected features' plenoptic disc radius (dimensionless). (d) Total detection time of 20 independent experiments.

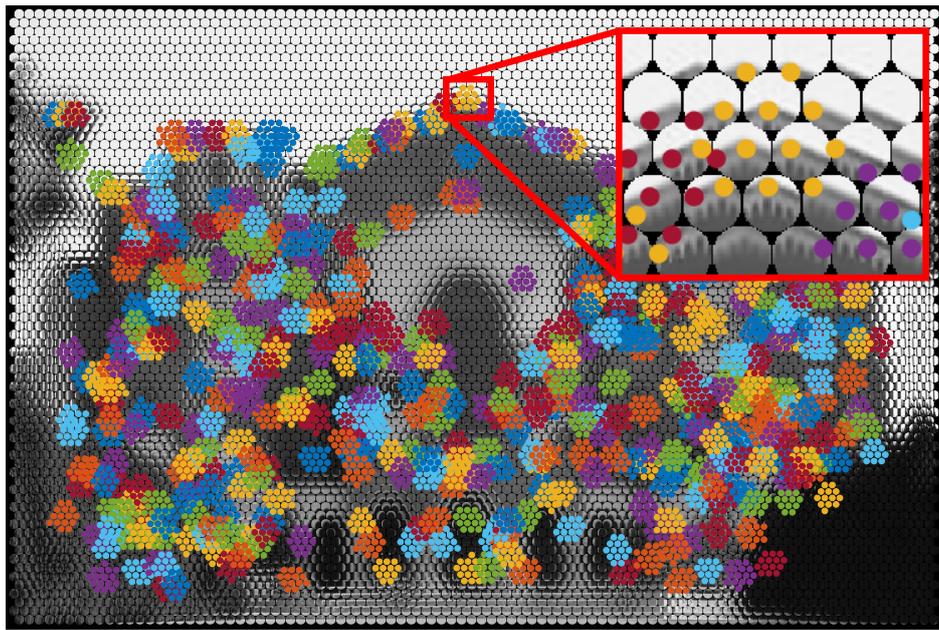


Figure 9. Example of detected results on simulated raw image using proposed algorithm based on 2-layer CPSFS with $R_{gt} = -4, \Delta R = 2$. Note that different colors in figure are used to distinguish the features projected by diverse plenoptic disc data.

Based on analysis above, we can conclude that it is necessary to keep a close layer distance in CPSFS, in order to ensure the number and accuracy of detected features. Empirically speaking, the ΔR can be chosen within $[1, 2]$.

4.1.3. Performance of Detection Algorithm Based on Multi-Layer CPSFS

In this part, we carry out feature detection using L -layer CPSFS on simulated raw images, in order to test the performance of the proposed method in scenes with different depth ranges. Specifically, different plane targets are placed perpendicular to the optical axis successively, so that the targets will form multiple imaging planes with plenoptic disc radius range $\{R_{gt} | R_{gt} = -2 - i, i = 1, 2, \dots, 8\}$. We carry out feature detection with L -layer CPSFS and observe the change of detection results with regard to the change of R_{gt} .

For each value of R_{gt} , 20 independent feature detection experiments are conducted on 20 simulated raw images with different textures. In order to reflect the relation between the number of layer in CPSFS and the detected results, we change the value of L and repeat the experiment 3 times. The specific value of L is chosen from $\{2, 4, 8\}$ and the corresponding plenoptic disc radii of the refocus planes used in CPSFS are $\{-3, -4\}$, $\{-3, -4, -5, -6\}$, and $\{-3, -4, -5, -6, -7, -8, -9, -10\}$, respectively. Note that $d = 1/2, \theta_R = 5$, during the experiment.

The detected results are shown in Figure 10. As one can see, the more layers used in CPSFS, the better the detection results, but the calculation time will increase as well. In practice, we can first determine the approximate range of the plenoptic disc radius according to the working range of the focused plenoptic camera. Then, the CPSFS can be layered accordingly. The detected results of the same raw image using 2-layer CPSFS and 8-layer CPSFS are illustrated in Figure 11.

At this point, we can conclude that the proposed feature detection algorithm based on multi-layer CPSFS is precise, effective, and time-efficient. In the actual process, the calculation efficiency can be further improved by reasonably layering of CPSFS.

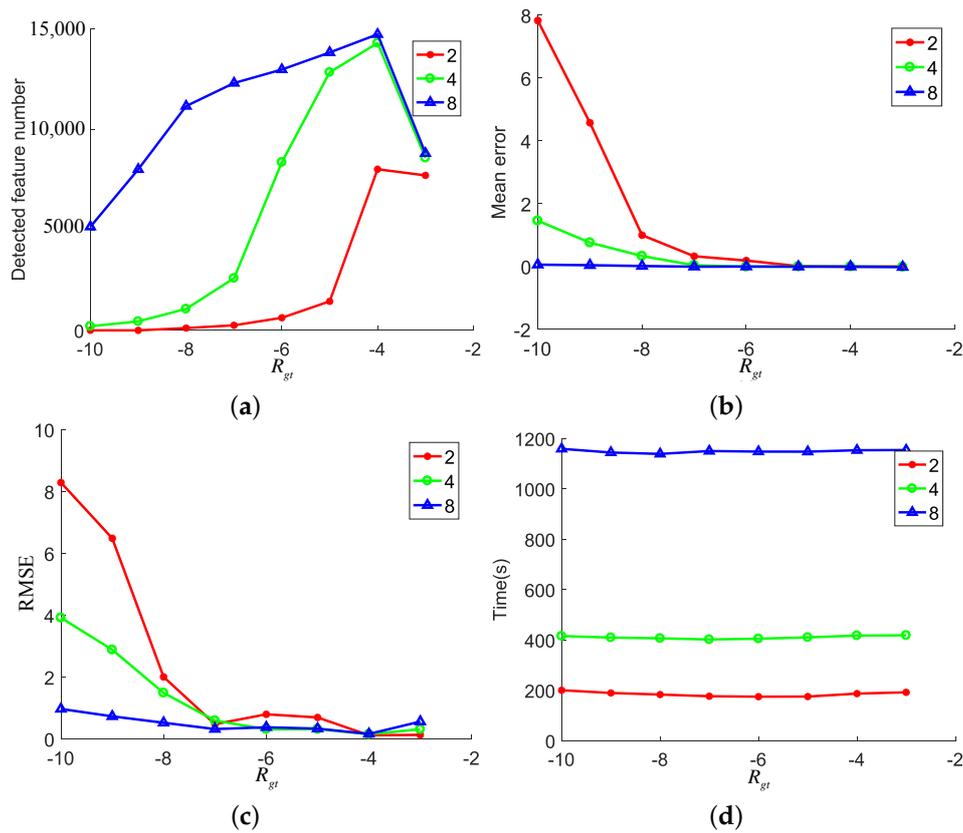


Figure 10. Detected results of proposed algorithm using multi-layer CPSFS on simulated raw images. (a) The number of detected features with regard to R_{gt} . (b) Mean error of detected features’ plenoptic disc radius (dimensionless). (c) RMSE of detected features’ plenoptic disc radius (dimensionless). (d) Total detection time of 20 independent experiments.

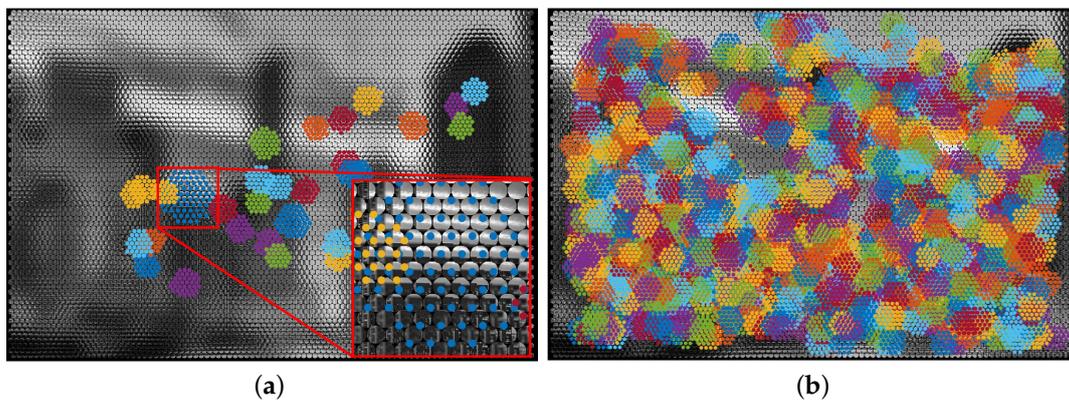


Figure 11. Examples of detected results on simulated raw images. (a) Error detected result using 2-layer CPSFS. (b) Detected result using 8-layer CPSFS. The results in panels (a,b) are based on the same raw image. Note that different colors in figures are used to distinguish the features projected by diverse plenoptic disc data.

4.1.4. Anti-Noise Capacity Test

In this part, the anti-noise capacity of the proposed method is tested on simulated raw images with noise. During the experiment, the plane object is placed at the fixed position to generate the imaging plane at $R_{gt} = -6$ and the CPSFS used has eight layers with $d = 1/4$ and $\theta_R = 5$. For generality of the results, we simulate 50 raw images with 50 different pictures. During the simulation, Gaussian noise with zero mean and a standard deviation σ is added to the simulated raw images. The detected results with regard to different σ (noise level) are shown in Figure 12.

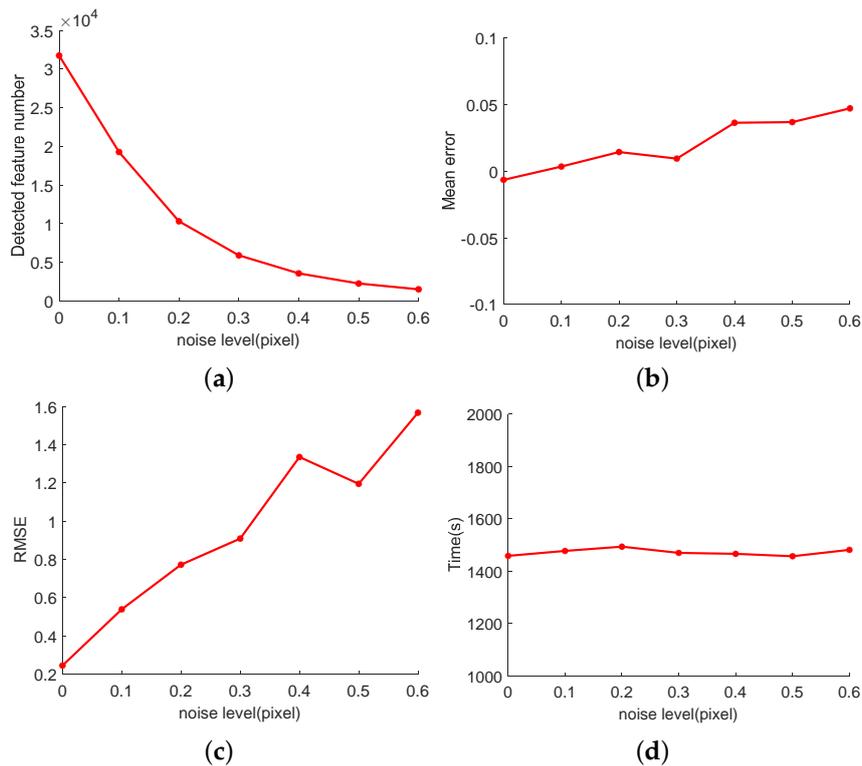


Figure 12. Detected results of proposed algorithm using 8-layer CPSFS on simulated raw images with noise. (a) Detected feature number during noise test. (b) Mean error of detected features' plenoptic disc radius during noise test (dimensionless). (c) RMSE of detected features' plenoptic disc radius during noise test (dimensionless). (d) Total detection time of 50 independent experiments during noise test.

Although the number of detection points decreases with the increase of noise level, the detection results can still maintain a high accuracy. In fact, the refocusing process is essentially equivalent to a weighted average of the corresponding pixel in the raw image, which will effectively improve the signal-to-noise ratio of the refocused image. Therefore, the proposed method based on CPSFS has good anti-noise ability. The specific detection example is presented in Figure 13. It can be seen that the raw image is in poor condition under the pollution of noise, but the proposed algorithm can still detect the SIFT [10] features in the raw image with good accuracy.

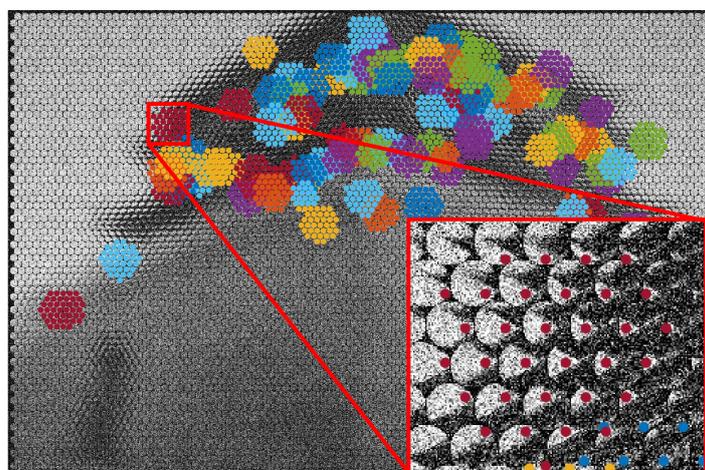


Figure 13. Detection result with noise level $\sigma = 0.3$ pixel. The size of simulated raw image is 3000 pixel \times 2000 pixel. Note that different colors in figure are used to distinguish the features projected by diverse plenoptic disc data.

4.2. Real Experiments

In order to further verify the effectiveness of the proposed algorithm, we use Raytrix's R29 camera to perform feature detection experiments based on real captured raw images. Besides, we incrementally reconstruct the detected features of 11 captured raw images to prove that the detected results of the proposed method can be used as input for SFM algorithm in scene reconstruction.

4.2.1. Feature Detection on Real Data

In this part, 5-layer CPSFS is used and the corresponding plenoptic disc radius is $R_F \in \{-4, -5.5, -7, -8.5, -10\}$. The parameters used in the algorithm are set as $d = 1/4, \theta_R = 5$. The example of generated central projection stereo refocused image pair is shown in Figure 14a,b, and the feature matching result of the central projection stereo refocused image pair is demonstrated in Figure 14c. In this paper, a total of 11 real images are tested and the total number of detected features is 7985. The average calculation time of each frame ($6576 \text{ pixel} \times 4384 \text{ pixel}$) is 86.3 s without any acceleration. The specific detection example is indicated in Figure 14d.

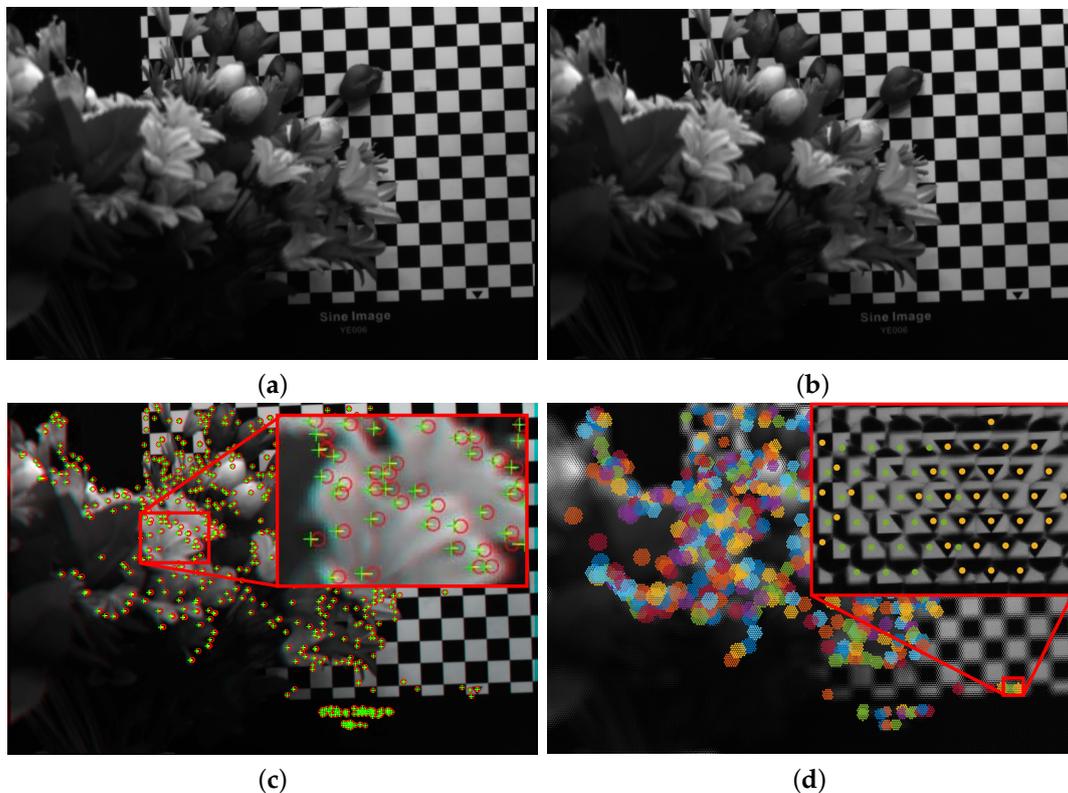


Figure 14. Example of feature detection on real data. (a) Left refocused image. (b) Right refocused image. (c) Feature matching result of central projection stereo refocused image pair. (d) Illustration of detected features on raw image.

4.2.2. Example of Scene Reconstruction

During the experiment of scene reconstruction, we use our previous work [8] to calibrate the R29 camera (<https://github.com/samliu0631/Stepwise-Calibration-for-plenoptic-camera>). The relative poses of the initial two frames are estimated using the method proposed by Li et al. [21] and the absolute poses of the remaining nine frames are calculated according to method proposed by Kneip et al. [22]. The sparse point clouds are incrementally reconstructed using method similar to COLMAP [3]. The example of captured raw image and the final reconstruction result are shown in the Figure 15. Compared with the traditional feature extraction method using raw images or total

focus images, the proposed method can improve the detection efficiency while ensuring the detection accuracy, thereby improving the efficiency of scene reconstruction.

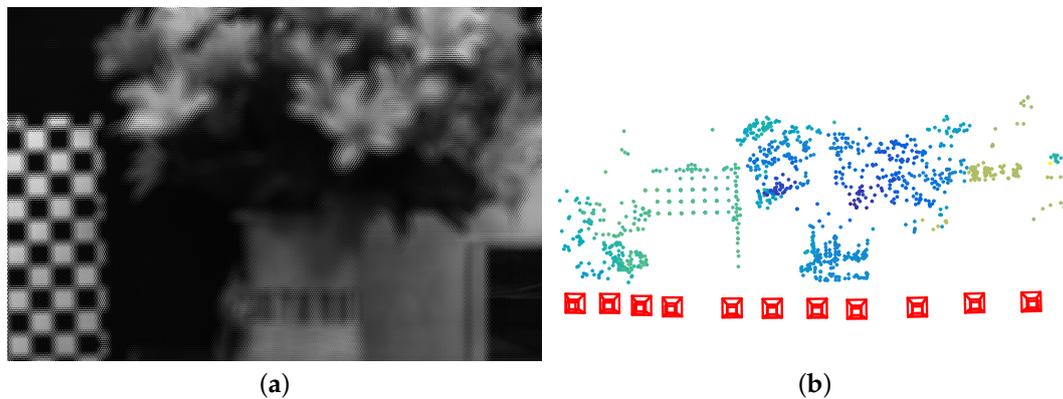


Figure 15. Sparse point cloud reconstruction results based on real data. (a) Example of captured raw image (6576 pixel \times 4384 pixel). (b) Reconstructed sparse point cloud from 11 frames.

5. Conclusions

In this paper, we propose a fast and accurate feature detection algorithm suitable for focused plenoptic cameras that can provide a reference for scene reconstruction. First, we propose a refocusing algorithm based on the concept of plenoptic disc data. The generated focal stack is consistent with the central projection with regard to the center of main lens. On this basis, we propose a feature detection algorithm using multi-layer CPSFS. The proposed method can improve the detection efficiency while ensuring the detection accuracy. Both simulated and real experiments are carried out to prove the effectiveness of our method. Besides, a specific example of scene reconstruction is conducted. The experimental results demonstrate that the feature detection method in this paper can provide a good reference for the feature-based scene reconstruction via focused plenoptic camera.

Author Contributions: All authors have contributed in some way to the concept and implementation of this paper. All authors contributed to the paper either during the writing or editing phases. All authors have read and agreed to the published version of the manuscript.

Funding: Research Grants from College of Advanced Interdisciplinary Studies, National University of Defense Technology (JC18-07).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ng, R.; Levoy, M.; Brédif, M.; Duval, G.; Horowitz, M.; Hanrahan, P. Light field photography with a hand-held plenoptic camera. *Comput. Sci. Tech. Rep. CSTR* **2005**, *2*, 1–11.
2. Perwass, C.; Wietzke, L. Single lens 3D-camera with extended depth-of-field. In Proceedings of the Human Vision and Electronic Imaging XVII, International Society for Optics and Photonics, Burlingame, CA, USA, 23–26 January 2012; Volume 8291, p. 829108.
3. Schonberger, J.L.; Frahm, J.M. Structure-from-motion revisited. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4104–4113.
4. Noursias, S.; Lourakis, M.; Bergeles, C. Large-Scale, Metric Structure From Motion for Unordered Light Fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 3292–3301.
5. Johannsen, O.; Sulc, A.; Goldluecke, B. On Linear Structure from Motion for Light Field Cameras. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 720–728.
6. Bok, Y.; Jeon, H.G.; Kweon, I.S. Geometric calibration of micro-lens-based light field cameras using line features. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 287–300. [[CrossRef](#)]

7. Nousias, S.; Chadebecq, F.; Pichat, J.; Keane, P.; Ourselin, S.; Bergeles, C. Corner-based geometric calibration of multi-focus plenoptic cameras. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 957–965.
8. Liu, Q.; Xie, X.; Zhang, X.; Tian, Y.; Li, J.; Wang, Y.; Xu, X. Stepwise calibration of plenoptic cameras based on corner features of raw images. *Appl. Opt.* **2020**, *59*, 4209–4219. [[CrossRef](#)]
9. Ferreira, R.; Goncalves, N. Fast and accurate micro lenses depth maps for multi-focus light field cameras. In *German Conference on Pattern Recognition*; Springer: Cham, Switzerland, 2016; pp. 309–319.
10. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
11. Dansereau, D.G.; Girod, B.; Wetzstein, G. LiFF: Light field features in scale and depth. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 8042–8051.
12. Kühfuß, A.; Zeller, N.; Quint, F.; Stilla, U. Feature Based RGB-D SLAM for a Plenoptic Camera. *BW-CAR | SINCOM* **2016**, *12*, 25–29.
13. Palmieri, L.; Koch, R. Optimizing the Lens Selection Process for Multi-focus Plenoptic Cameras and Numerical Evaluation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 1763–1774.
14. Palmieri, L.; Koch, R.; Veld, R.O.H. The Plenoptic 2.0 Toolbox: Benchmarking of Depth Estimation Methods for MLA-Based Focused Plenoptic Cameras. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 649–653.
15. Zeller, N.; Quint, F.; Stilla, U. Establishing a probabilistic depth map from focused plenoptic cameras. In Proceedings of the 2015 International Conference on 3D Vision (3DV), Lyon, France, 19–22 October 2015; pp. 91–99.
16. Zeller, N.; Quint, F.; Stilla, U. Filtering probabilistic depth maps received from a focused plenoptic camera. *BW-CAR | SINCOM* **2015**, *2*, 7–12.
17. Hog, M.; Sabater, N.; Vandame, B.; Drazic, V. An image rendering pipeline for focused plenoptic cameras. *IEEE Trans. Comput. Imaging* **2017**, *3*, 811–821. [[CrossRef](#)]
18. O'Brien, S.; Trumpf, J.; Ila, V.; Mahony, R. Calibrating Light-Field Cameras Using Plenoptic Disc Features. In Proceedings of the 2018 International Conference on 3D Vision (3DV), Verona, Italy, 5–8 September 2018; pp. 286–294.
19. Grossberg, M.D.; Nayar, S.K. The raxel imaging model and ray-based calibration. *Int. J. Comput. Vis.* **2005**, *61*, 119–137. [[CrossRef](#)]
20. Zhang, R.; Liu, P.; Liu, D.; Su, G. Reconstruction of refocusing and all-in-focus images based on forward simulation model of plenoptic camera. *Opt. Commun.* **2015**, *357*, 1–6. [[CrossRef](#)]
21. Li, H.; Hartley, R.; Kim, J.H. A linear approach to motion estimation using generalized camera models. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
22. Kneip, L.; Furgale, P.; Siegwart, R. Using multi-camera systems in robotics: Efficient solutions to the npnp problem. In Proceedings of the 2013 IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, 6–10 May 2013; pp. 3770–3776.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).