

Article

Guided, Fusion-Based, Large Depth-of-field 3D Imaging Using a Focal Stack

Yuhao Xiao ¹, Guijin Wang ^{1,*}, Xiaowei Hu ¹, Chenbo Shi ² and Long Meng ²
and Huazhong Yang ¹¹ Department of Electronic Engineering, Tsinghua University, Beijing 100084, China;

xyh18@mails.tsinghua.edu.cn (Y.X.); hu-xw15@mails.tsinghua.edu.cn (X.H.); yanghz@tsinghua.edu.cn (H.Y.)

² ShanDong MingJia Technology Co., Ltd., Taian 271021, China; shichenbo@gmail.com (C.S.);
mengl@mingjiachina.com (L.M.)

* Correspondence: wangguijin@tsinghua.edu.cn

Received: 19 September 2019; Accepted: 5 November 2019; Published: 7 November 2019

Abstract: Three dimensional (3D) imaging technology has been widely used for many applications, such as human–computer interactions, making industrial measurements, and dealing with cultural relics. However, existing active methods often require both large apertures of projector and camera to maximize light throughput, resulting in a shallow working volume in which projector and camera are simultaneously in focus. In this paper, we propose a novel method to extend the working range of the structured light 3D imaging system based on the focal stack. Specifically in the case of large depth variation scenes, we first adopted the gray code method for local, 3D shape measurement with multiple focal distance settings. Then we extracted the texture map of each focus position into a focal stack to generate a global coarse depth map. Under the guidance of the global coarse depth map, the high-quality 3D shape measurement of the overall scene was obtained by local, 3D shape-measurement fusion. To validate the method, we developed a prototype system that can perform high-quality measurements in the depth range of 400 mm with a measurement error of 0.08%.

Keywords: 3D imaging; focal stack; guided fusion

1. Introduction

Three dimensional (3D) imaging based on structured light has been widely used in many applications, such as for human–computer interactions, industrial measurements, and regarding cultural relics [1–3]. There are numerous structured-light methods which have been discussed thoroughly by Salvi et al. [4] and Zhang [5]. However, due to the limited light source power and low sensor quantum efficiency, these existing methods often require a large aperture to maximize the light throughput from the light source to the camera, which causes the depth of field (DOF) of the projector and camera to reduce. The common DOF between the projector and camera is the working volume of the structured light 3D imaging system, which is shallow as a result. That fundamental limitation makes it difficult for the system to be used to measure scenes with large depth variation. To make the structured light 3D imaging system work for general real-world scenes, it is essential to extend the working volume of the system, which is defined as the DOF of the system in this paper.

The state-of-the-art in extending DOF methods of the structured light 3D imaging system focus on the projector’s DOF extension mainly, assuming that the camera has an infinite DOF, which will not hold in the case of large aperture requirement. This paper considers the camera’s and projector’s DOF extension simultaneously and presents a novel method to extend DOF of the structured light 3D imaging system using the MEMS mirror-enabled laser projector and the liquid-lens-based imaging system. For projector DOF extension, we replaced the traditional DLP projector with the MEMS mirror-enabled laser projector whose displayed images are in focus everywhere [6]. The optics of these

projectors are fundamentally different from DLP projectors, which use laser light sources and do not suffer from focusing issues. For camera DOF extending, we adopted the liquid-lens-based imaging system to capture images to form a focal stack, which was able to generate the coarse depth map of the scene to guide local 3D measurement fusion. It is vital for 3D measurement systems to know for certain, the focal length, and keep it unchanged during calibration. Thus, extending the working range of 3D imaging systems using a liquid lens can achieve a highly stable optical performance. Specifically, we (1) captured structured light patterns with different focal settings to measure objects in the scene; (2) extracted a texture map from each focus position into the focal stack to generate the global coarse depth map; (3) fused each local measurement under the guidance of the global coarse depth map to generate the high-quality 3D measurement covering the overall scene. Our prototype system can work effectively for the depth range of 200–600 mm with the measurement error of 0.08%.

The rest of our paper is organized as follows. In Section 2, we review prior approaches that are related to our proposed method. Then, we explain the principle of the method we propose in Section 3. In Section 4, experiments on real data are presented to verify the effectiveness of the method. In Section 5, we discuss the limitations of the method. Finally, we summarize the paper in Section 6.

2. Related Work

In this section, we first review the state-of-the-art large DOF methods of the structured light 3D imaging system. Then, we introduce camera DOF extension studies, which are related to our method.

2.1. DOF Extension of the Structured Light 3D Imaging System

For the DOF extension of the structured light 3D imaging system, the main work is focused on the projector, assuming that the camera has an infinite DOF. Zhang and Nayar exploited the effect of projector defocusing as a cue to recover 3D shape in [7]. They developed a novel, temporal, defocus analysis method to recover depth at each camera pixel by estimating the projected pattern blur occurring in the frequency domain. They could recover accurate depth maps, but the effects of global light transport could not be handled well. Gupta et al. presented a method to model both projector defocus and global illumination simultaneously [8]. They illustrated that both these seemingly disparate effects could be expressed as low pass filters on the incident illumination. Gupta and Nayar proposed projecting sinusoidal patterns with frequencies limited to a narrow, high-frequency band. These patterns produce a set of images over which defocus's effects remain constant for each point in the scene. Thus, the projector lens's defocusing effect can be attenuated as much as possible [9]. However, this method's accuracy relies on the user's selection of the pattern frequency. Moreover, the information about the pattern degrades rapidly when it is faced with severe defocus; thus, it can only extend the working range to some extent. In [10], illumination defocus is exploited in a different way, which generates sinusoidal patterns by projecting binary patterns with a defocused projector. Thus, the phase shift algorithm could run in real-time and recover dynamic scenes. Achar and Narasimhan extended the working volume by projecting structured light patterns with different focus settings [11]. However, this approach is volatile due to the mechanically tunable lens. The seminal works of Brown et al. [12] were the first to improve projection quality by compensating defocus blurs using image pre-compensation. They estimated a series of spatially-varying point-spread-functions across the projector's image, which were then used to guide preprocessing based on Wiener filtering to condition the image before projection. This technique works well for small kernels but is less effective on large kernels. Zhang et al. proposed focal length and exposure adaption in [13]. They changed the illumination pattern and intensity locally based on the prior depth information. However, the accuracy of depth values measured by this method is not high. A light field projector was introduced by Kawasaki et al. [14], which is realized by attaching a coded aperture with a high-frequency mask. However, the coded aperture attenuated the light throughout and resulted in a decrease in the signal-to-noise ratio.

2.2. DOF Extension of the Camera

For the DOF extension of the camera, Kuthirummal et al. presented a novel imaging system which varies the position of the image detector during the integration time of a single photograph to extend DOF [15]. However, this system is highly dependent on the hardware device, and so it is rather complicated and power-consuming. Levin et al. proposed a modification to the conventional camera, which inserts a patterned occlusion within the aperture of the camera lens to recover all-in-focus images [16]. However, the applicable scenarios for that modification are limited, and depth recovery is not robust. Wavefront coding is a promising approach to extend DOF imaging [17,18]. An all-in-focus image can be obtained by deconvolving the captured image with a single blur kernel. Georgeiv et al. developed a prototype integral camera which samples the 4D light field of a scene to compute all in-focus images [19]. However, this camera is limited by a fundamental trade-off between spatial resolution and angular resolution. A related approach is to capture many images to form a focal stack [20,21]. This method can generate an all-in-focus image of the scene with a depth map effectively. To reduce the capture time of a focal stack, Hasinoff and Kutulakos proposed a technique which uses a combination of different apertures and focal plane locations [22].

The aforementioned camera's DOF extension methods cannot be directly employed for large depth-range 3D shape measurements, since they are primarily designed for getting focused 2D images without exactly knowing the altered physical parameters (e.g., focal length), while for any DOF extending methods to be adopted in 3D shape measurement, any changes of a physical parameter have to be precisely known. Fortunately, the focus stacking technique can compute a coarse depth map of the scene, which could be exploited to enlarge the working range of structured light 3D imaging systems effectively. The detailed content will be discussed in the following sections.

3. Computational Framework for Large DOF 3D Imaging

The framework of the method we propose is shown in Figure 1. The overall process could be divided into four modules, which are local fine 3D Imaging, texture map extraction, global depth rough estimation, and guided fusion. First, we project the structured light patterns over the scene through the MEMS mirror-enabled laser projector and then use the imaging system equipped with a liquid lens to capture a sequence of images focusing at various depth of the scene. At each focus position, we adopt the "Local Fine 3D Imaging" module to measure 3D shape of the object and extract a texture map through the "Texture Map Extraction" module. The texture maps are then combined into the focal stack to generate the global coarse depth map with the module of "Global Depth Rough Estimation." Finally, the high-quality measurement of the overall scene, which is composed of the accurate part of local measurements, is obtained through the "Guided Fusion" module. The principle of each module will be discussed in subsections that follow.

3.1. Local Depth Fine Estimation and Texture Map Extraction

The structured light technique is one of the extensively adopted methods to recover depth information of the scene. A projection device is utilized to project structured patterns over the object actively. The projected, structured patterns carry encoded information to establish the correspondence between the camera and the projector. Once the system is calibrated, 3D coordinates of the scene can be generated well, based on the ray-plane triangulation. In this paper, we choose the gray code method as the local 3D measurement due to the benefits of being easy to operate and producing accurate reconstructions of objects which are within the working range [23].

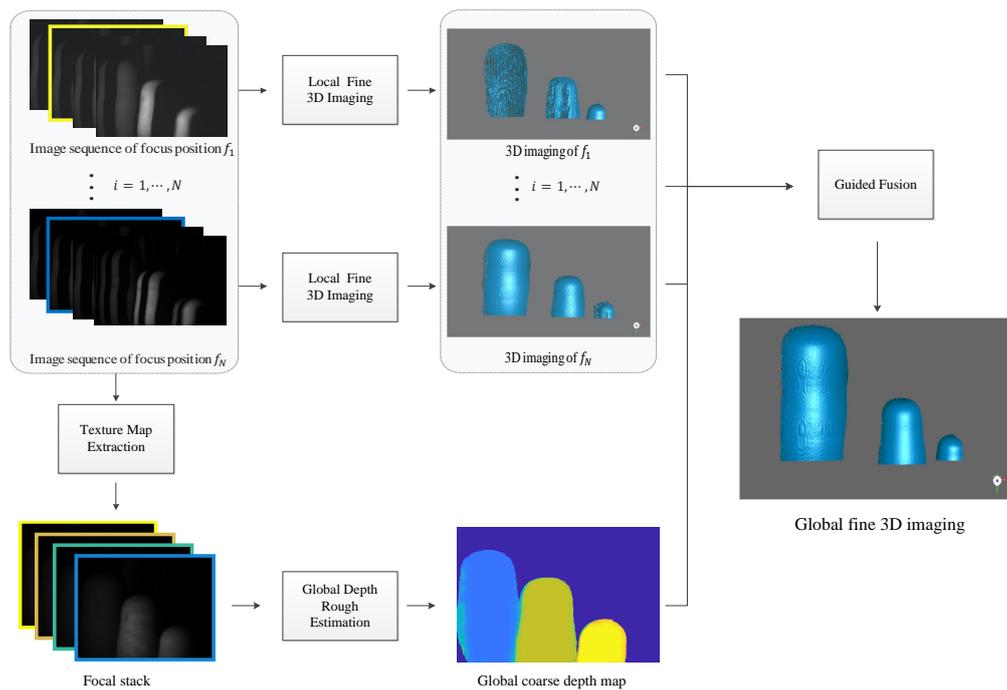


Figure 1. The framework of our proposed method. Image sequences of different focal distances are acquired by the multi-focus illumination-imaging system; then, local, fine 3D measurements and a global coarse depth map are generated by the gray-code method and focus stacking technique, respectively. These local, fine 3D measurements, along with the global coarse depth map, are finally fed into the guided fusion model to get the high-quality 3D measurement of the overall scene.

For generating a focal stack, the optimal sequence of focus positions is obtained based on [24]. Specifically, let us consider the optical system focused at distance z_k from the lens, as shown in Figure 2a. There is a depth range DOF_k around z_k for which the scene is in-focus in the imaging plane. Based on the thin lens model and geometrical considerations, the near limit, $z_k^{(1)}$, and far limit $z_k^{(2)}$ of DOF_k , satisfy:

$$\frac{1}{z_k^{(1)}} - \frac{1}{z_k} = \frac{c}{2Rd} \quad (1)$$

$$\frac{1}{z_k} - \frac{1}{z_k^{(2)}} = \frac{c}{2Rd'} \quad (2)$$

where R is the radius of the aperture, d is the fixed distance between the imaging plane and the lens, and c is the diameter of the circle of confusion. The condition to obtain an optimal sequence of focus positions is that the far limit of DOF_k coincides exactly with the near limit of DOF_{k+1} of the next focal distance z_{k+1} :

$$z_k^{(2)} = z_{k+1}^{(1)}, \quad (3)$$

and then, a recursion formula, given the foreground distance ($k = 1$) z_1 , is derived to compute the optimal set of focal distance:

$$z_{k+1} = \left(\frac{1}{z_k} - \frac{c}{Rd} \right)^{-1}. \quad (4)$$

After determining the focus positions, we need to pick a texture map from the captured images at each focus position to form the focal stack. The sharpness of the texture change is important for the focus stacking technique to distinguish which focus position is in-focus. Since there may exist texture-less objects in the general scene, we need to add texture to the surface of the objects. The gray coded patterns vary from coarse stripes to fine stripes gradually. Therefore, we regard the structured light projected on the surface of the object as its "surface texture," and extract the image which is projected by the finest stripe from each focus position to combine the focal stack, as shown in Figure 2b.

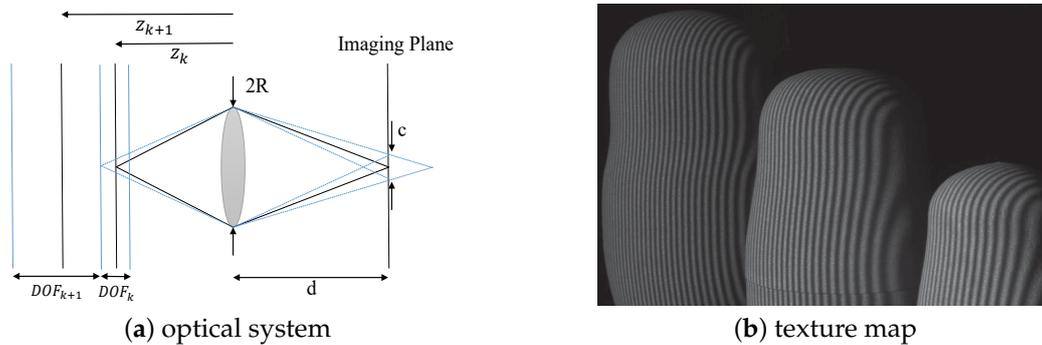


Figure 2. Optimal focus positions and texture map extraction. (a): optical system; (b): texture map. The optimal sequence of focus positions has no gaps or overlaps between depths of field (DOFs) from consecutive focal planes. The image projected by the finest stripes of the gray code sequence is extracted into the focal stack.

3.2. Global Depth Rough Estimation

The focus stacking technique is a computational approach to extend the camera DOF effectively. However, the all-in-focus image cannot be directly employed for the reason that we cannot know the altered physical parameters (e.g., focal length) exactly. Thanks to the by-product of the focus stacking technique, we could get the coarse depth map of the scene, which can guide local 3D measurements to fuse into the global 3D measurement with high quality.

As discussed before, we extract texture maps from various focus distances and combine them into the focal stack. Then, given the magnitude of gradient as the sharpness measure, we calculate the gradients of all the images in the stack by

$$G_i(x, y) = |\nabla I_i(x, y)| = \sqrt{\left(\frac{\partial I_i}{\partial x}\right)^2 + \left(\frac{\partial I_i}{\partial y}\right)^2}, \quad (5)$$

where $G_i(x, y)$ is the gradient of $I_i(x, y)$ which is the i th image in the focal stack. After that, the global coarse depth map can be obtained by

$$D(x, y) = \underset{i}{\operatorname{argmax}} G_i(x, y), \quad (6)$$

where $D(x, y)$ stores the index of image that has the maximum gradient of each pixel in the focal stack. This depth map is also known as a focus map. However, due to the large blur kernel case, the edges will spread on a large scale. Thus, many wrong values in the depth map may result. In our previous work [20], we proposed a max-gradient flow (MGF) method to remove false edges and formulated an anchored, rolling-filtering approach to refine the depth map, which effectively reduces the artifacts caused by large blur scale. Here, the main steps in MGF will be reviewed as follows:

First, we define the 2D max-gradient flow as

$$\operatorname{MGF}(x, y) = [f_x(x, y), f_y(x, y)]^T, \quad (7)$$

where

$$\begin{bmatrix} f_x(x, y) \\ f_y(x, y) \end{bmatrix} = \begin{bmatrix} \frac{\max_j G_j(x+\Delta x, y) - \max_i G_i(x, y)}{\Delta x} \\ \frac{\max_k G_k(x, y+\Delta y) - \max_i G_i(x, y)}{\Delta y} \end{bmatrix} \quad (8)$$

The flow describes the change of maximum gradient along each axis (i.e., the propagation of gradient in the stack). And the direction of propagation can be easily derived as

$$\theta(x, y) = \arctan \left(\frac{f_y(x, y)}{f_x(x, y)} \right). \quad (9)$$

Then, with the max-gradient flow, we can extract source point (x, y) if

$$f_x(x + \delta x, y) > 0 \quad \text{and} \quad f_x(x - \delta x, y) < 0, \quad (10)$$

which means that it is the source where the max-gradient starts propagating. All the source points form a sparse set, which represents the true edges in the scene. Finally, we formulate an anchored rolling filtering approach to refine the depth map with the extracted true edges. Since the depth map and the all-in-focus image are closely related, we form a ping-pong approach that estimates them in turn. Starting from the initial depth map, the all-in-focus image and the depth map are updated iteratively in a rolling way. Suppose that the depth map D is given, the all-in-focus image FI can be composited by extracting the pixel from the stack:

$$FI(x, y) = I_{D(x, y)}(x, y). \quad (11)$$

And suppose that the all-in-focus image FI is available, the depth map D can be refined by a joint bilateral filtering approach [25]:

$$\hat{D}(p) = \frac{1}{K_p} \sum_{p' \in N_p} \exp \left(-\frac{\|p' - p\|^2}{2\sigma_s^2} - \frac{\|FI(p') - FI(p)\|^2}{2\sigma_r^2} \right) D(p'), \quad (12)$$

where K_p is the normalization factor, N_p is the neighborhood of pixel $p = (x, y)$. σ_s and σ_r control the spatial and range weights. The all-in-focus image FI is taken as the guidance image of the joint bilateral filter which smooths the depth map while preserving the edges. The anchored, rolling refinement continues until the depth map is visually indifferent between two iterations.

3.3. Guided Fusion

In this section, we derive a novel fusion model to generate the high-quality 3D measurement of the overall scene, which is shown in Figure 3. After the previously-described processes, we obtained multiple local 3D measurements of the scene. However, due to the working-range limitation of the traditional method, only objects in focus could be reconstructed well, while other, defocused objects could not. It is favorable to distinguish which objects are focused under the focal setting and extract the corresponding local 3D measurement. Fortunately, the global coarse depth map generated by the focus stacking technique could achieve this goal. The depth map stores the index of the image in the stack, which corresponds to the focus position of the image. Thus, we directly employ the depth map as the guidance map to extract the accurate part of each local 3D measurement. Regarding the depth map as a mask for each focus position, we define the mask function $\delta_i(x, y)$ for each focus position i as

$$\delta_i(x, y) = \begin{cases} 1, & \text{if } D(x, y) = i \\ 0, & \text{if } D(x, y) \neq i \end{cases}. \quad (13)$$

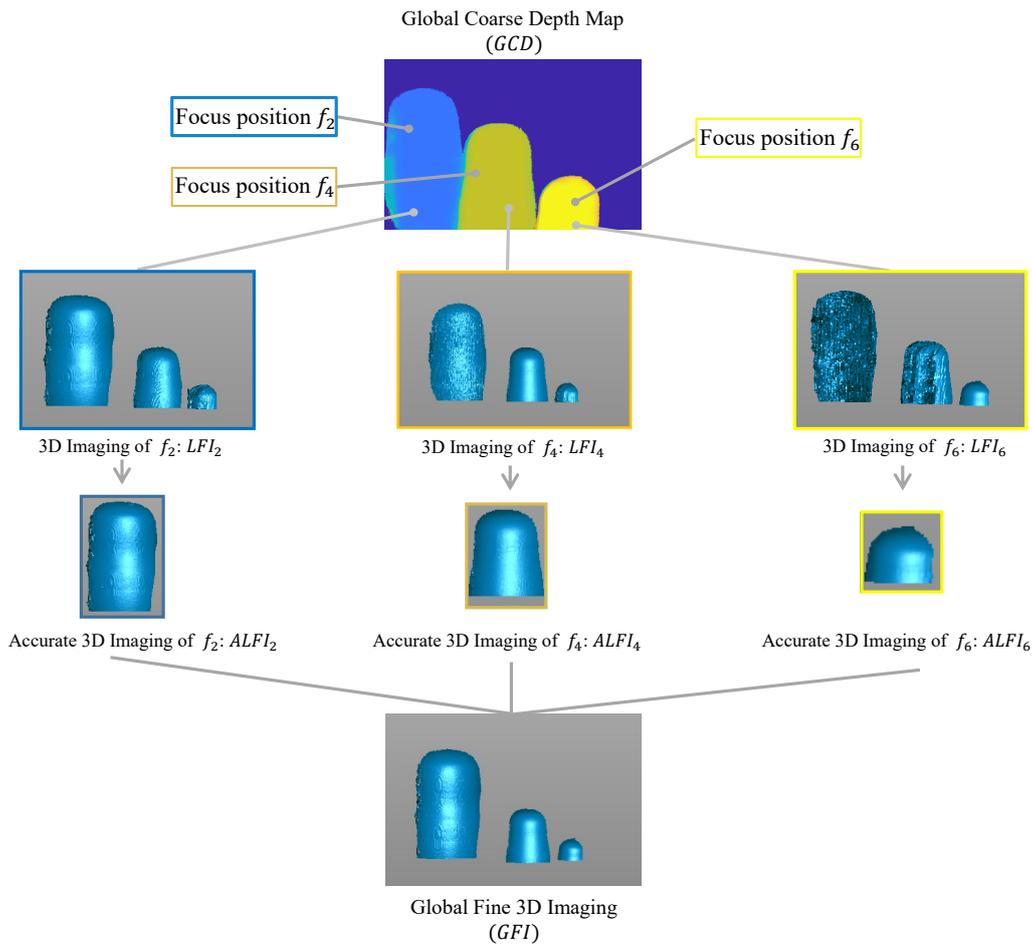


Figure 3. The guided fusion model. The global coarse depth map contains the corresponding focus position of each object. Therefore, we can extract the accurate region of each local fine 3D measurements under the guidance of the global coarse depth map to fuse into the global fine 3D measurement.

Then, the accurate part of each local 3D measurement ($ALFI_i$) is extracted under the guidance of each mask and they are fused into the global fine 3D measurement (GFI) by

$$ALFI_i(x, y) = LFI_i(x, y)\delta_i(x, y) \quad (14)$$

$$GFI(x, y) = \sum_{i=1}^N ALFI_i(x, y), \quad (15)$$

where N is the range of the index in the focal stack, and $LFI_i(x, y)$ is the 3D measurement of position (x, y) in the focus position i .

4. Experiments

In this section, we will present the experimental performance of our proposed method. In Section 4.1, we introduce the prototype system and the test scene. Section 4.2 evaluates the performance of our proposed method with the traditional method on real data. In Section 4.3, we analyze the reconstruction errors of our method with the traditional method. At last, Section 4.4 discusses the robustness of our method in the different scenes.

4.1. Setup

The prototype system and the test scene are shown in Figure 4. The system included a CMOS camera (model: Basler acA1920-155um, manufactured by the Basler AG, Ahrensburg, Germany) equipped with a 12 mm focal length lens (model: Computar m1214-mp2, manufactured by the Computar, Tokyo, Japan) and a liquid lens (model: Optotune EL-10-30-Ci, manufactured by the Optotune AG, Dietikon, Switzerland). A MEMS mirror-enabled laser projector (model: Sony MP-CL1A, manufactured by the Sony Corporation, Tokyo, Japan) was used for structured light projection. We adopted a high-precision electrical lens driver to control the liquid lens. The resolution of the camera was set to 1280×800 while the resolution of the projector was 1920×1200 . The camera aperture was always set to the maximum of $f/1.4$ to maximize light throughput, which results in a small DOF effect. Thus, we can demonstrate the effectiveness of our proposed method to extend the DOF of the structured light 3D imaging system. In order to test the effect of our method in the depth range of 200–600 mm, we placed a small doll at the distance of roughly 200 mm, a medium doll at the distance of roughly 400 mm, and a big doll at the distance of roughly 600 mm. The depth variation of the scene was 400 mm, which was large compared to the relative distance between the scene and the baseline of the prototype system described before.

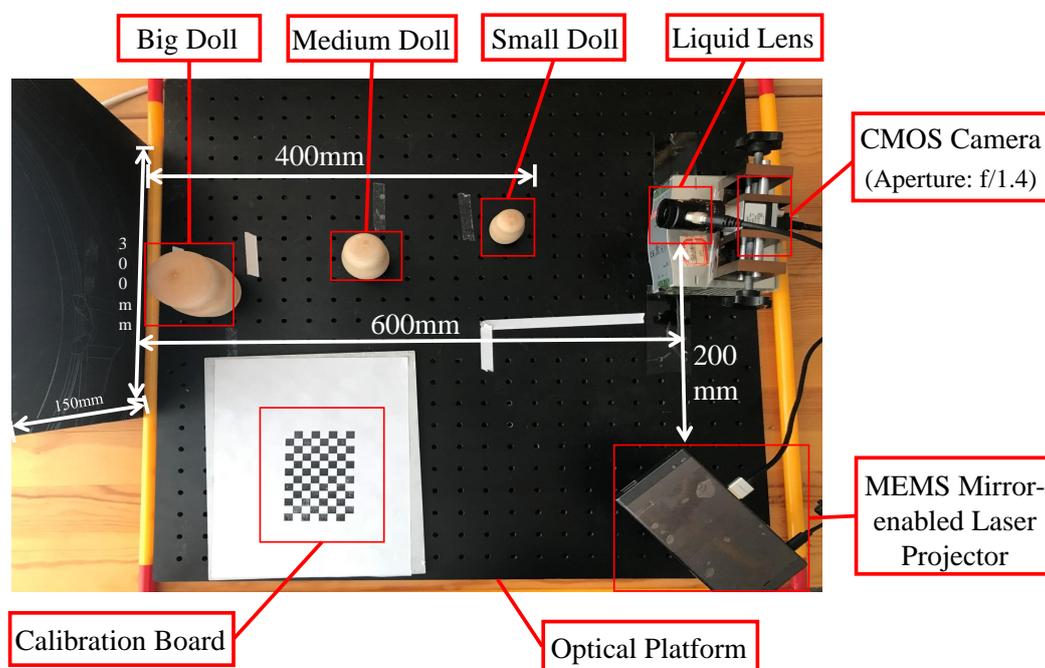


Figure 4. The prototype system and the test scene. The system includes a MEMS mirror-enabled laser projector, a CMOS camera equipped with a liquid lens, and a control platform for synchronization. Three different sizes of dolls are placed in front of the camera, from far to near. The specific scene is shown in the figure.

4.2. Overall Performance

In this experiment, we adopted six discrete focal distances for local 3D measurements based on the method in Section 3.1: $f_1 = 100$ mm, $f_2 = 200$ mm, $f_3 = 300$ mm, $f_4 = 400$ mm, $f_5 = 500$ mm, and $f_6 = 600$ mm; and their corresponding driver currents are 242.21 mA, 84.74 mA, 47.41 mA, 35.83 mA, 23.38 mA, and 16.73 mA. The union of the depth of fields of all the images in the focal stack covers the range of three dolls in the scene. After capturing the structured light patterns from f_1 to f_6 , we need to calibrate the system with the corresponding focal setting. The liquid-lens-based imaging system

allows us to accurately restore each focus position by setting driver current and performing precise calibration for 3D measurement [26].

To show the limitations of traditional methods and demonstrate the effectiveness of our method, we compare the results of local 3D measurements when $f = 200$ mm, 400 mm, and 600 mm with our proposed method in Figure 5. The parts within red rectangles correspond to objects which are not focused; thus, the measurement quality drops with the increasing amount of defocusing. Our method extracts the best quality 3D measurement from each local position and fuses them all into the global results; thus, achieving the best performance.

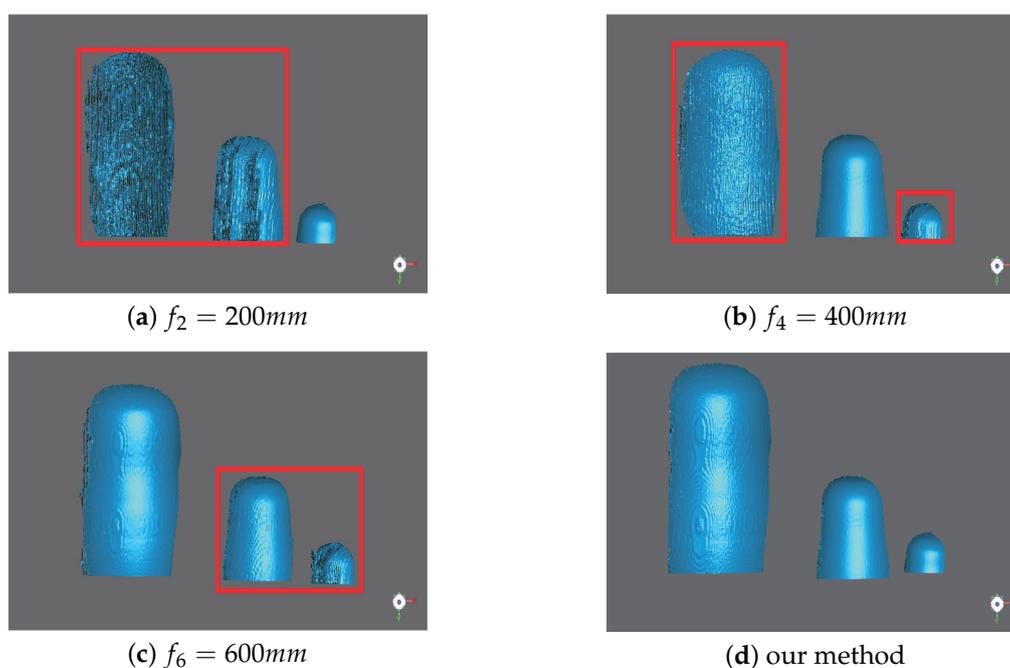


Figure 5. Comparison of measurement results. (a) The camera focal distance at $f_2 = 200$ mm; (b) the camera focal distance at $f_4 = 400$ mm; (c) the camera focal distance at $f_6 = 600$ mm; (d) our method.

As shown in Figure 5, the surface of the reconstructed object is rough when the object is not focused, while the surface is smooth when the object is focused. Therefore, we can measure the smoothness of the surface to determine the quality of measurement by Geomagic studio. The "Reduce Noise" function provides three indicators for quantifying the smoothness of the surface of the object, which are maximum distance, average distance, and standard deviation. These three indicators are for the entire 3D reconstruction that includes both the in-focus region and the out-of-focus region, and they show information regarding the distance that points are moved during the smoothing process. Maximum distance indicates the greatest distance that any point is moved. Average distance indicates the average distance that all points are moved. Standard deviation indicates the variability of the distances. The comparison results are listed in Table 1. Our proposed method is better than the traditional method in terms of the maximum distance, average distance, and standard deviation, which fully demonstrates the superiority of our method.

Table 1. Comparison of surface smoothness of reconstructed doll.

	Maximum Distance (mm)	Average Distance (mm)	Standard Deviation (mm)
$f_2 = 200$ mm	1.613432	0.121577	0.155088
$f_4 = 400$ mm	1.279587	0.134344	0.130969
$f_6 = 600$ mm	1.539015	0.325372	0.276541
Our method	0.919761	0.055475	0.047435

4.3. Error Analysis

We further quantitatively evaluated the reconstruction error by measuring the sphere in this section. The dolls were replaced with the same size spheres in each location, and then we fit the reconstruction data with an ideal spherical model. We set three focal distances of the camera f_2 , f_4 , and f_6 to ensure one sphere was focused in the image, respectively. The comparison diagram is shown in Figure 6 and the comparison results are listed in Tables 2 and 3. In Figure 6, the first column shows the results of the first sphere (diameter = 30 mm at 200 mm), second sphere (diameter = 50 mm at 400 mm), and third sphere (diameter = 70 mm at 600 mm) when the camera set its focal distance at f_2 . The second column shows the corresponding results of f_4 , the third column shows the corresponding results of f_6 , and the last column shows the corresponding results of our method. The RMSEs and the mean errors are their fitting errors with the corresponding ideal sphere, not for the entire scene. The average RMSEs of each focus setting were 1.7421 mm (f_2), 1.8474 mm (f_4), and 0.7887 mm (f_6), and the average mean errors of each focus setting were 1.4298 mm (f_2), 1.4666 mm (f_4), and 0.6081 mm (f_6). Our method extracts the accurate measurement of each local position into the final composited measurement. The RMSE and mean error of our method were 0.2455 mm and 0.1679 mm, which is an achievement of approximately a 70% decrease of each local measurement. The experiments quantitatively demonstrate that the reconstruction error increases as the objects move away from the camera's focal plane, especially the region with large shape gradient. And our proposed method can effectively achieve the most accurate measurements with the guided fusion model.

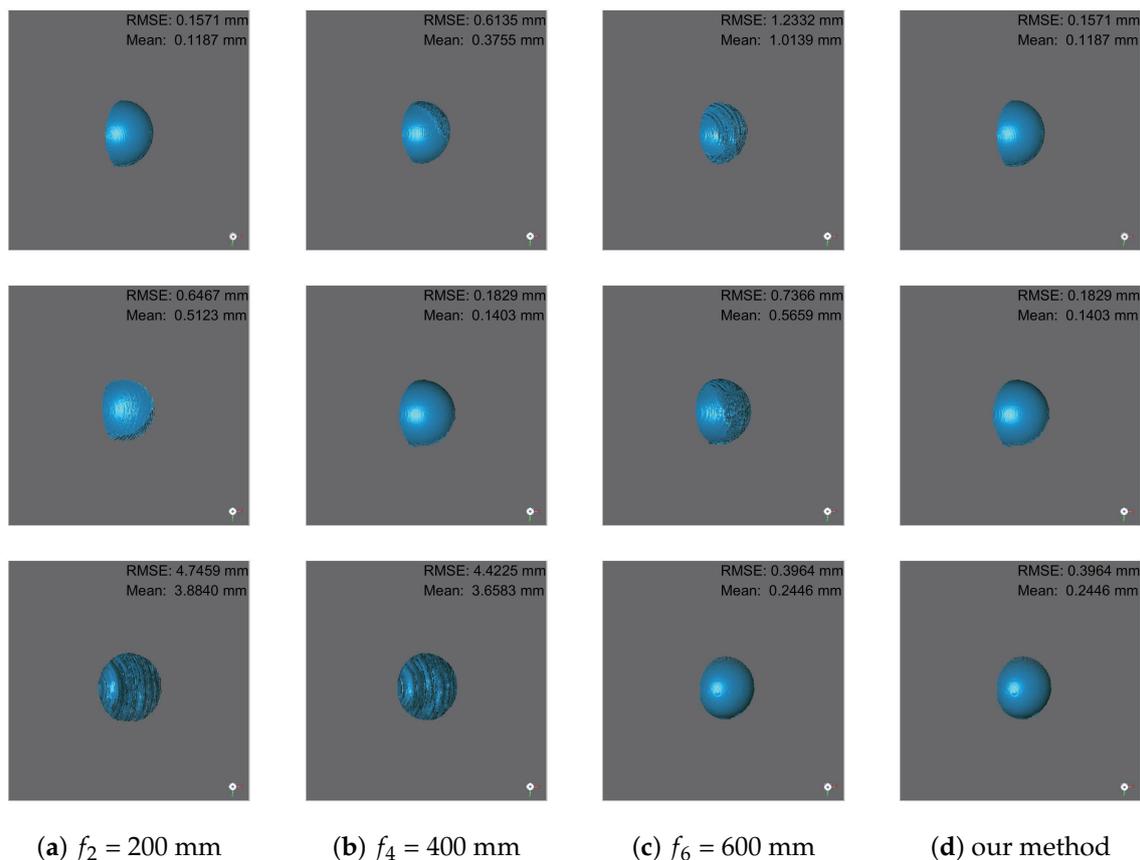


Figure 6. Comparison of reconstruction results. (a) $f_2 = 200$ mm; (b) $f_4 = 400$ mm; (c) $f_6 = 600$ mm; (d) our method. The first row represents the results of the small sphere (diameter = 30 mm at 200 mm) under different camera settings and methods. The second row represents the corresponding results of the medium sphere (diameter = 50 mm at 400 mm), and the last row represents the corresponding results of the big sphere (diameter = 70 mm at 600 mm). The reconstruction errors are shown in the upper right corner of each figure, which are RMSEs and mean errors.

Table 2. Data comparison of fitting results in terms of RMSE (mm).

	Small Sphere	Medium Sphere	Big Sphere	Average Value
$f_2 = 200$ mm	0.1571	0.6467	4.7459	1.8499
$f_4 = 400$ mm	0.6135	0.1829	4.4225	1.7396
$f_6 = 600$ mm	1.2332	0.7366	0.3964	0.7887
our method	0.1571	0.1829	0.3964	0.2455

Table 3. Data comparison of fitting results in terms of mean error (mm).

	Small Sphere	Medium Sphere	Big Sphere	Average Value
$f_2 = 200$ mm	0.1187	0.5123	3.8840	1.5050
$f_4 = 400$ mm	0.3755	0.1403	3.6583	1.3914
$f_6 = 600$ mm	1.0139	0.5659	0.2446	0.6081
our method	0.1187	0.1403	0.2446	0.1679

4.4. Robustness Analysis

In this section, we mainly analyze the robustness of our method in another two different scenes. The first scene includes various shapes and different surface reflectivity values, as shown in Figure 7a. In the process of capturing structured light patterns, we changed the focus position from near to far, and made each object in-focus at a certain focal setting. The “Global Depth Rough Estimation” module accurately extracted the focus position of each target from the focal stack and generated the corresponding depth map of the scene, which is shown in Figure 7b. We present the experiment results in Figure 8. Figure 8a shows the 3D measurement when the statue is in focus. Similarly, Figure 8b,c show the 3D measurements when the box and the doll are in focus, respectively. As discussed in Section 4.2, the best-quality 3D result is obtained when the object is in-focus and the measurement quality drops with the increasing amount of defocusing. The 3D measurement of our method is shown in Figure 8d, which perfectly preserved the shape of each target and achieved the best performance.

**Figure 7.** Photograph of the test scene and the global coarse depth map. (a) Test scene; (b) depth map.

The second scene consists of one object having a large depth range; thus, one focus position can only make its part in-focus. The depth map of the scene is shown in Figure 9a, which clearly indicates the focus position of each part. We present the experiment results in Figure 9b–d. Figure 9b shows the 3D measurement when the left part is focused, and Figure 9c shows the 3D measurement when the right part is focused. Our method extracts the best quality 3D measurements under the guidance of the depth map and fuses them into the final result, which is shown in Figure 9d. The experiments clearly demonstrate that our method can effectively work in a more complicated scene and achieve high-quality measurements.

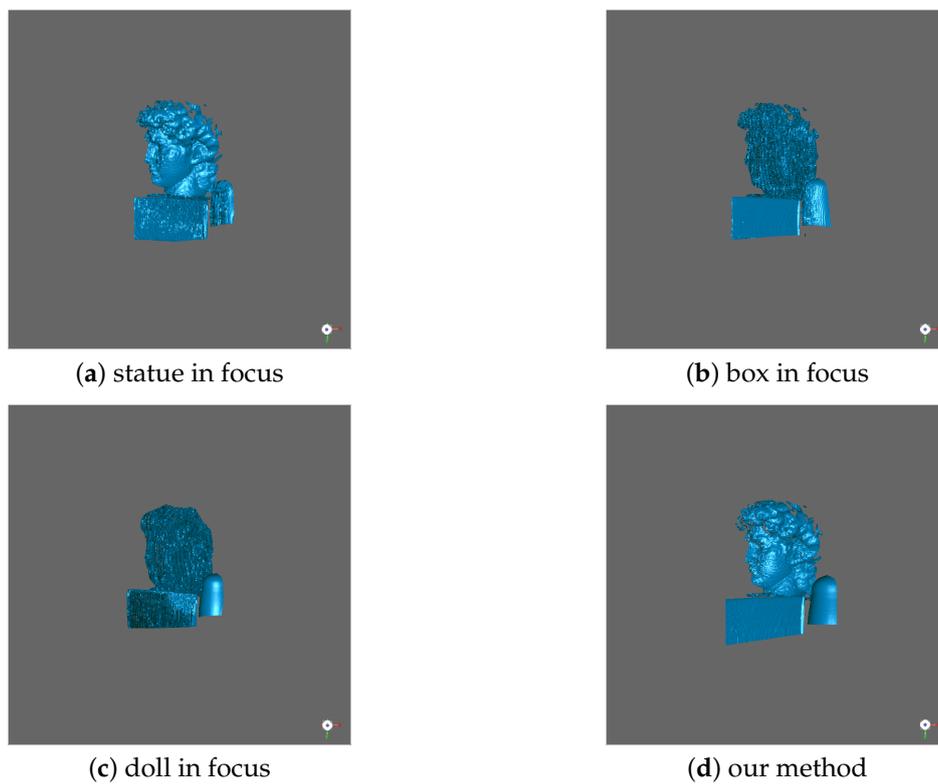


Figure 8. Comparison of measurement results. The scene includes three objects of different shapes and materials. (a) The statue in-focus; (b) the box in-focus; (c) the doll is in-focus; (d): our method.

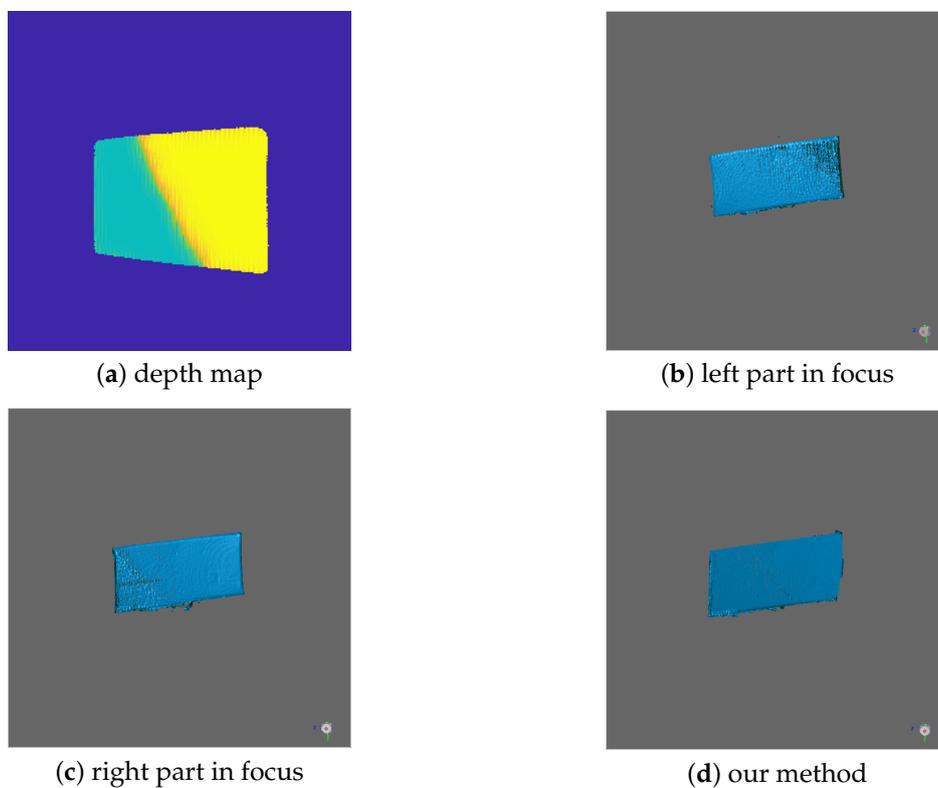


Figure 9. Comparison of measurement results. The scene includes one object having a large depth range. (a) The global coarse depth map; (b) the left part in-focus; (c) the right part in-focus; (d) our method.

5. Discussion

The proposed method achieved high-quality measurement of a large depth variation scene. Yet, it also makes some sacrifices. The major limitations associated with the proposed method are:

- *More time required:* since our method needs to capture images at different focus positions, the time taken by our method is more than that of the traditional methods.
- *High precision calibration required:* The proposed method has a strict requirement for calibration. Inaccurate calibration will introduce an artificial error, making measurement points of each focus position misaligned.

Despite these limitations, the values of the proposed methods are still substantial for high-accuracy and high-quality measurements of large-depth variation scenes. Besides, the time increase is adjustable according to the scene, and the calibration methods have been studied well in recent years.

6. Conclusions

In this paper, we proposed a novel method to extend the working range of structured light 3D imaging systems. We changed the traditional structured light 3D imaging system to the MEMS mirror-enabled laser projector and the liquid-lens-based imaging system. Based on the focus stacking technique, we developed a guided fusion model to generate a high-quality 3D measurement of the large depth variation scene. In experiments with the prototype system, we developed demonstrated that the proposed method could effectively work from a depth range of 200–600 mm with the measurement error of 0.08%.

Author Contributions: Conceptualization, Y.X. and X.H.; methodology, Y.X.; software, Y.X. and X.H.; validation, Y.X., C.S., and L.M.; formal analysis, Y.X., G.W., and H.Y.; investigation, H.Y., C.S., and L.M.; writing—original draft preparation, Y.X.; writing—review and editing, Y.X. and G.W.; visualization, Y.X. and X.H.; supervision, G.W.

Funding: This research received no external funding.

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

References

1. Liberadzki, P.; Adamczyk, M.; Witkowski, M.; Sitnik, R. Structured-Light-Based System for Shape Measurement of the Human Body in Motion. *Sensors* **2018**, *18*, 2827. [[CrossRef](#)] [[PubMed](#)]
2. Hinz, L.; Kästner, M.; Reithmeier, E. Metal Forming Tool Monitoring Based on a 3D Measuring Endoscope Using CAD Assisted Registration. *Sensors* **2019**, *19*, 2084. [[CrossRef](#)] [[PubMed](#)]
3. Song, L.; Li, X.; Yang, Y.g.; Zhu, X.; Guo, Q.; Liu, H. Structured-light based 3D reconstruction system for cultural relic packaging. *Sensors* **2018**, *18*, 2981. [[CrossRef](#)] [[PubMed](#)]
4. Salvi, J.; Fernandez, S.; Pribanic, T.; Llado, X. A state of the art in structured light patterns for surface profilometry. *Pattern Recognit.* **2010**, *43*, 2666–2680. [[CrossRef](#)]
5. Zhang, S. High-speed 3D shape measurement with structured light methods: A review. *Opt. Lasers Eng.* **2018**, *106*, 119–131. [[CrossRef](#)]
6. Mertz, C.; Koppal, S.J.; Sia, S.; Narasimhan, S. A low-power structured light sensor for outdoor scene reconstruction and dominant material identification. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Providence, RI, USA, 16–21 June 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 15–22.
7. Zhang, L.; Nayar, S. Projection defocus analysis for scene capture and image display. *ACM Trans. Gr. (TOG)* **2006**, *25*, 907–915. [[CrossRef](#)]
8. Gupta, M.; Tian, Y.; Narasimhan, S.G.; Zhang, L. A combined theory of defocused illumination and global light transport. *Int. J. Comput. Vis.* **2012**, *98*, 146–167. [[CrossRef](#)]
9. Gupta, M.; Nayar, S.K. Micro phase shifting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 813–820.

10. Lei, S.; Zhang, S. Digital sinusoidal fringe pattern generation: Defocusing binary patterns vs focusing sinusoidal patterns. *Opt. Lasers Eng.* **2010**, *48*, 561–569. [[CrossRef](#)]
11. Achar, S.; Narasimhan, S.G. Multi focus structured light for recovering scene shape and global illumination. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Cham, Switzerland, 2014; pp. 205–219.
12. Brown, M.S.; Song, P.; Cham, T.J. Image pre-conditioning for out-of-focus projector blur. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 17–22 June 2006; IEEE: Piscataway, NJ, USA, 2006; Volume 2, pp. 1956–1963.
13. Zhang, Y.; Xiong, Z.; Cong, P.; Wu, F. Robust depth sensing with adaptive structured light illumination. *J. Vis. Commun. Image Represent.* **2014**, *25*, 649–658. [[CrossRef](#)]
14. Kawasaki, H.; Ono, S.; Horita, Y.; Shiba, Y.; Furukawa, R.; Hiura, S. Active one-shot scan for wide depth range using a light field projector based on coded aperture. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 3568–3576.
15. Nagahara, H.; Kuthirummal, S.; Zhou, C.; Nayar, S.K. Flexible depth of field photography. In Proceedings of the European Conference on Computer Vision, Marseille, France, 12–18 October 2008; Springer: Berlin/Heidelberg, Germany, 2008; pp. 60–73.
16. Levin, A.; Fergus, R.; Durand, F.; Freeman, W.T. Image and depth from a conventional camera with a coded aperture. *ACM Trans. Gr. (TOG)* **2007**, *26*, 70. [[CrossRef](#)]
17. George, N.; Chi, W. Extended depth of field using a logarithmic asphere. *J. Opt. A Pure Appl. Opt.* **2003**, *5*, S157. [[CrossRef](#)]
18. Castro, A.; Ojeda-Castañeda, J. Asymmetric phase masks for extended depth of field. *Appl. Opt.* **2004**, *43*, 3474–3479. [[CrossRef](#)] [[PubMed](#)]
19. Georgiev, T.G.; Zheng, K.C.; Curless, B.; Salesin, D.; Nayar, S.K.; Intwala, C. Spatio-angular resolution tradeoffs in integral photography. *Render. Tech.* **2006**, *2006*, 21.
20. Yin, X.; Wang, G.; Li, W.; Liao, Q. Large aperture focus stacking with max-gradient flow by anchored rolling filtering. *Appl. Opt.* **2016**, *55*, 5304–5309. [[CrossRef](#)] [[PubMed](#)]
21. Wang, G.; Li, W.; Yin, X.; Yang, H. All-in-focus with directional-max-gradient flow and labeled iterative depth propagation. *Pattern Recognit.* **2018**, *77*, 173–187. [[CrossRef](#)]
22. Hasinoff, S.W.; Kutulakos, K.N. Light-efficient photography. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 2203–2214. [[CrossRef](#)]
23. Salvi, J.; Pages, J.; Batlle, J. Pattern codification strategies in structured light systems. *Pattern Recognit.* **2004**, *37*, 827–849. [[CrossRef](#)]
24. Alonso, J.R.; Fernández, A.; Ferrari, J.A. Reconstruction of perspective shifts and refocusing of a three-dimensional scene from a multi-focus image stack. *Appl. Opt.* **2016**, *55*, 2380–2386. [[CrossRef](#)] [[PubMed](#)]
25. Kopf, J.; Cohen, M.F.; Lischinski, D.; Uyttendaele, M. Joint bilateral upsampling. *ACM Trans. Gr. (ToG)* **2007**, *26*, 96. [[CrossRef](#)]
26. Moreno, D.; Taubin, G. Simple, accurate, and robust projector-camera calibration. In Proceedings of the Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission, Zurich, Switzerland, 13–15 October 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 464–471.

