



Directional Ring Difference Filter for Robust Shape-from-Focus

Khurram Ashfaq  and Muhammad Tariq Mahmood * 

Future Convergence Engineering, School of Computer Science and Engineering, Korea University of Technology and Education, 1600 Chungjeolro, Byeongcheonmyeon, Cheonan 31253, Republic of Korea; khurram@koreatech.ac.kr

* Correspondence: tariq@koreatech.ac.kr; Tel.: +82-041-560-1483

Abstract: In the shape-from-focus (SFF) method, the quality of the 3D shape generated relies heavily on the focus measure operator (FM) used. Unfortunately, most FMs are sensitive to noise and provide inaccurate depth maps. Among recent FMs, the ring difference filter (RDF) has demonstrated excellent robustness against noise and reasonable performance in computing accurate depth maps. However, it also suffers from the response cancellation problem (RCP) encountered in multidimensional kernel-based FMs. To address this issue, we propose an effective and robust FM called the directional ring difference filter (DRDF). In DRDF, the focus quality is computed by aggregating responses of RDF from multiple kernels in different directions. We conducted experiments using synthetic and real image datasets and found that the proposed DRDF method outperforms traditional FMs in terms of noise handling and producing a higher quality 3D shape estimate of the object.

Keywords: focus measure; shape-from-focus; ring difference filter; depth map; 3D shape recovery

MSC: 68T45; 65D19; 65D18

1. Introduction

The depth of a scene has become an increasingly important task in the field of computer vision, with a wide range of applications in areas such as autonomous navigation, augmented and virtual reality, robot control [1], and 3D model reconstruction [2]. Among the methods used to infer the depth of a scene, the shape-from-focus (SFF) method is one of the optical methods known for its simplicity and accurate depth maps [3]. SFF operates on the principle that the depth of a scene can be inferred by utilizing the information from in-focus pixels. The main steps involved in SFF techniques are shown in Figure 1. Its pipeline commences with capturing a series of images by using a single camera with varying focus settings for each image. Such a sequence can also be obtained by translating objects toward or away from the camera in small steps and then capturing images. Next, a focus measure operator (FM) is applied to each image in the stack to determine the sharpness of each pixel. This results in an initial focus volume (FV) that provides focus information for each pixel in the image sequence. The initial FV may contain erroneous focus values, which can affect the accuracy of depth values. Therefore, an appropriate filtering technique is applied to the initial FV and an improved FV is obtained. Next, an initial depth map is obtained by locating the pixels with maximum sharpness along the optical axis. However, the resultant depth map may still contain noisy depth estimates. To address this issue, the final step involves using a cost aggregation method to refine the initially found depth map and an improved final depth map is obtained [4].

After obtaining the image sequence, the next step in SFF is to compute the focus quality for each pixel by applying an appropriate FM on the input image sequence. In the literature, a large number of FMs have been proposed, which can be grouped into various categories, such as statistical-based, first derivative-based, second derivative-based, and transformation-based [5,6]. Statistical-based FMs operate by applying statistical measures on the local pixels. One such commonly used method is the gray-level variance



Citation: Ashfaq, K.; Mahmood, M.T. Directional Ring Difference Filter for Robust Shape-from-Focus.

Mathematics **2023**, *11*, 3056. <https://doi.org/10.3390/math11143056>

Academic Editor: Samaneh Mazaheri

Received: 12 June 2023

Revised: 2 July 2023

Accepted: 7 July 2023

Published: 11 July 2023



Copyright: © 2023 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 (<https://creativecommons.org/licenses/by/4.0/>).

(GLV) [7], which works by analyzing the variation in the intensity values of neighboring pixels in a small window. Another technique is the absolute central moment (ACM) [8], which evaluates the quality of focus of an image by utilizing both the histogram and mean value of gray levels in an image. Additional techniques in this category include the eigenvalues of a local window [9], polynomial coefficients and spectral radius-based focus measure [10], and probability coefficients and modified entropy (PCME) [11]. The first derivative-based focus measures work by calculating the gradient of an image. Tenengrad focus measure (TFM) is the most commonly used; it calculates the gradient of an image using the Sobel operator in both the x and y directions, and then sums up the squared magnitude of the gradient over a small window to obtain the sharpness value. Other common methods in this category include the local edge gradient analysis [12], modulus of the gradient of the color channel (MCG) [13], and reduced Tenengrad (RT) [14], which is a slight modification to the TFM. Second derivative-based focus measures utilize the Laplacian of an image. Although they are more susceptible to noise than first derivative-based methods, second derivative-based measures can provide more accurate assessments of focus. The most widely used focus measure in this category is the sum of the modified Laplacian (ML) [3], which involves taking the absolute value of the second derivative of pixels in a small window in both the x and y directions and then summing their responses. Its other examples are squares of the partial derivatives [15] and multi-scale weighted modified Laplacian (MSWML) [16]. Transformation-based FMs use the energy of the high-frequency component or the ratio between high- and low-frequency components to measure sharpness. For instance, the energy ratio of the wavelet coefficient [17] employs the wavelet transform of the image and determines the ratio between the norms of the high-pass and low-pass bands to calculate the focus value. Several other techniques that fall into this category include the energy of coefficients in discrete curvelet transform [18], optimal discrete cosine transform coefficients [19], reorganized DCT coefficients [20,21], and Chebyshev moments [22]. Additionally, there are some other FMs that do not fall in the previously mentioned categories but produce exceptional results. For example, the sum and spread focus measure (FMSS) [23] that calculates an image's sharpness value by using basic vector operations and incorporating information from different color channels. A multi-scale morphological focus measure (MSM) [24] that uses morphological operations i.e., dilation and erosion to obtain sharpness values and integrates them on different scales. Another example is the ring difference filter (RDF) [25], which uses a unique combination of ring and disk filter styles convolved with the image to determine the focus value. Moreover, steerable filters [26], quad-tree decomposition and edge-weighted focus measure [27], and a perceptual-based robust focus measure based on the difference of Gaussian [28] are additional examples that yield impressive results.

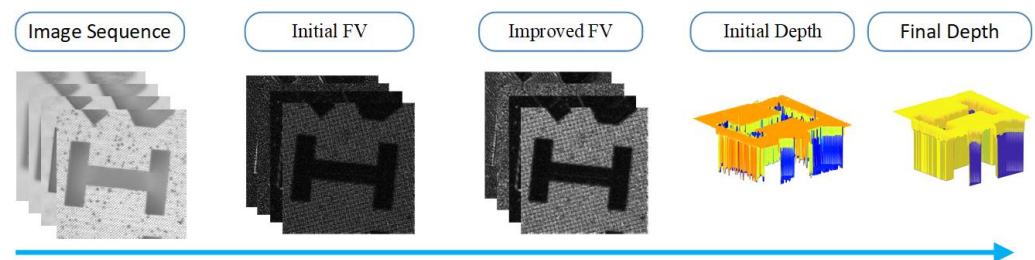


Figure 1. Main steps involved in the shape-from-focus technique.

The initial depth map obtained from the focus volume may contain outliers in the data, resulting in noisy depth maps. To address this issue, cost aggregation methods are applied, which consider neighboring pixels and their disparities within a window to improve the focus volume. Cost aggregation methods can be categorized into two types: focus volume enhancement and post-processing techniques. Focus volume enhancement involves refining the focus volume first and then using the refined focus volume to generate a depth map. A popular technique in this category involves using a Gaussian distribution to

find the peak in the image stack that represents the highly focused image [3]. Other methods include Gaussian process regression for focus curve fitting [29], weighted least squares regression for focus curve fitting [30], using the gradient of the focus measure curve with the adaptive derivative step to find the best-focused position [31], the phase correlation method that applies a discrete Fourier transform on the focus volume for peak detection [32], and optimizing the focus volume through energy minimization by exploiting the structural similarity between the image sequence and the initially obtained focus volume [33]. Post-processing techniques refine the depth map obtained from the initial focus volume. One of these techniques is bilateral filtering, which combines nearby image values based on both geometric closeness and photometric similarity while preserving edges [34]. Joint bilateral filtering is an extension of bilateral filtering that allows for the simultaneous filtering of multiple images or channels [35]. Another well-known technique is guided-image filtering, which uses a reference image to guide the filtering process of another image. It aims to preserve important edges and structures while smoothing out the rest of the pixels [36]. Due to its effectiveness, many versions of guided-image filtering have been proposed to enhance the depth map. A thorough study of these techniques is presented in [37]. The accuracy of the depth maps in SFF relies heavily on the performance of the FMs. The errors at the stage of focus computing will be propagated in the focus volume enhancement stage. Consequently, the erroneous depth maps will be extracted. There are a number of factors that can affect the performance of FM, such as scene texture, contrast, illumination, window size, noise level, and imaging device characteristics. Hence, a robust and effective focus measure is important for accurate depth maps in SFF.

In this paper, we introduce a new focus measure, the directional ring difference filter (DRDF) for SFF, which is able to handle noise more robustly. In contrast to RDF, where a 2D mask is convolved with an image sequence and the energy of the responses is collected as a focus measure, the proposed DRDF applies multiple 1D kernels in different directions. To compute the focus measure, the average response of these kernels is computed. In this way, the proposed measure helps to mitigate the response cancellation problem (RCP). Experimental results, obtained from both synthetic and real image sequences, demonstrate the effectiveness of DRDF in producing accurate and noise-robust depth maps.

The rest of the paper is organized as follows: the proposed method and motivation behind it are presented in Section 2; the experimental setup, results, and comparative analysis are provided in Section 3. Finally, Section 4 concludes this study.

2. Proposed Focus Measure

In this section, first, we provide the motivation for this work, which signifies the rationale behind the proposed method. Then, the steps in the proposed method are explained by using appropriate expressions and symbols.

2.1. Motivation

A thorough study of the focus measures is presented in [5], which reveals that the modified Laplacian focus measure (ML) is best for assessing accurate focus quality. However, ML is sensitive to noise, which restricts its usage in many real-world applications. In order to overcome this limitation, RDF, a modified version of the ML, is proposed [25]. RDF calculates the focus quality of each pixel by measuring the absolute differences with the neighboring pixels, with gaps between them; it can be expressed as follows:

$$h_{rdf} = \begin{cases} \frac{1}{\pi r_1^2}, & |p - q| < r_1 \\ -\frac{1}{\pi(r_3^2 - r_2^2)}, & r_2 \leq |p - q| \leq r_3 \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

where p is the position of the pixel of interest (central pixel in the window), q is the pixel index, r_1 is the radius of region of interest (the disk in which the pixel of interest resides),

r_2 and r_3 are inner and outer radii of the ring (the circular ring that surrounds the region of interest and gap pixels) respectively. Although RDF has shown excellent robustness against noise and reasonable performance in computing accurate depth maps, it suffers from the response cancellation problem (RCP). The RCP was pointed out in [3]. It was observed that the responses of the 2D Laplacian on an image, specifically $|\frac{\partial^2(\cdot)}{\partial x^2} + \frac{\partial^2(\cdot)}{\partial y^2}|$, canceled the effects from the opposite directions, which deteriorated the focus measure. This problem was fixed in ML $|\frac{\partial^2(\cdot)}{\partial x^2}| + |\frac{\partial^2(\cdot)}{\partial y^2}|$, where the responses from the x and y directions were calculated separately and the focus measure was taken as the sum of these responses. As RDF is also spread in multiple directions, the RCP technique significantly affects the resultant measures. To overcome this problem, we propose a directional ring difference filter (DRDF) that breaks a 2D RDF into multiple 1D kernels in various directions. The average response from these 1D kernels is then taken as the final focus measure.

Further, we investigated the effects of RDF and DRDF on focus assessment by using the datasets [38], which included 350 sharp and 350 blurred images ($350 \times 1680 \times 1180$ pixels in each dataset). We considered an RDF operator for ($r_1 = 1, r_2 = 1, r_3 = 1$) as shown in Figure 2a, and the corresponding kernels $h_i, i \in \{1, 2, 3, 4, 5, 6\}$ in the directions $\theta_c, c \in \{0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ\}$, respectively as shown in Figure 2b. These kernels in discrete form can be represented as follows:

$$h_1 = [0 \ 0 \ 0 \ 0 \ 0; \ 0 \ 0 \ 0 \ 0 \ 0; \ -1 \ 0 \ 2 \ 0 \ -1; \ 0 \ 0 \ 0 \ 0 \ 0; \ 0 \ 0 \ 0 \ 0 \ 0], \tag{2}$$

$$h_2 = [0 \ 0 \ 0 \ 0 \ 0; \ 0 \ 0 \ 0 \ 0 \ -1; \ 0 \ 0 \ 2 \ 0 \ 0; \ -1 \ 0 \ 0 \ 0 \ 0; \ 0 \ 0 \ 0 \ 0 \ 0], \tag{3}$$

$$h_3 = [0 \ 0 \ 0 \ -1 \ 0; \ 0 \ 0 \ 0 \ 0 \ 0; \ 0 \ 0 \ 2 \ 0 \ 0; \ 0 \ 0 \ 0 \ 0 \ 0; \ 0 \ -1 \ 0 \ 0 \ 0], \tag{4}$$

$$h_4 = [0 \ 0 \ -1 \ 0 \ 0; \ 0 \ 0 \ 0 \ 0 \ 0; \ 0 \ 0 \ 2 \ 0 \ 0; \ 0 \ 0 \ 0 \ 0 \ 0; \ 0 \ 0 \ -1 \ 0 \ 0], \tag{5}$$

$$h_5 = [0 \ -1 \ 0 \ 0 \ 0; \ 0 \ 0 \ 0 \ 0 \ 0; \ 0 \ 0 \ 2 \ 0 \ 0; \ 0 \ 0 \ 0 \ 0 \ 0; \ 0 \ 0 \ 0 \ -1 \ 0], \tag{6}$$

$$h_6 = [0 \ 0 \ 0 \ 0 \ 0; \ -1 \ 0 \ 0 \ 0 \ 0; \ 0 \ 0 \ 2 \ 0 \ 0; \ 0 \ 0 \ 0 \ 0 \ -1; \ 0 \ 0 \ 0 \ 0 \ 0]. \tag{7}$$

Absolute responses for RDF and DRDF kernels were computed for all images in the datasets. Let f_0 denote the measure from RDF and m_i denote responses from DRDF kernel h_i , where the focus measure f_i can be represented as follows:

$$f_i = \sum_{i=1}^i \frac{|m_i|}{i}, \quad i \in \{1, 2, 3, 4, 5, 6\}. \tag{8}$$

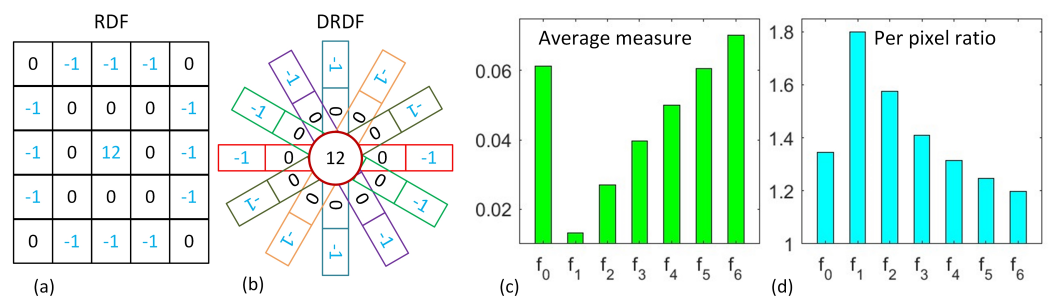


Figure 2. (a) ring difference filter, (b) directional ring difference filters, (c) average focus measure from all pixels, (d) ratio between blurred and sharp pixels.

Figure 2c shows the normalized average focus values of all pixels and Figure 2d shows the ratio of focus measures between blurred and sharp pixels for RDF and DRDF kernels respectively. It can be observed that adding responses to DRDF kernels improved the focus measure, with the final DRDF response (f_6) being better than the RDF response (f_0), as indicated by the higher average measure per pixel and lower per pixel ratio.

2.2. Method

Let the input image sequence be represented by $I_z(x, y)_c$, where x , y , and z denote the indices for the width, height, and image numbers, while $c \in \{r, g, b\}$ denotes the color channel of a vector-valued image. The ranges of x , y , and z are defined as $(1, \dots, X)$, $(1, \dots, Y)$ and $(1, \dots, Z)$, respectively, such that there are Z images, each of size $X \times Y$ pixels, and every image has three color channels.

First, the image sequence is aligned by employing a global homography-based alignment method [25]. It will fix any slight translation or magnification that may have occurred during the image-capturing process. After properly aligning the focal stack, the proposed DRDF focus measure is applied to it. A focus measure for each pixel is computed through the convolution of the images in the sequence with the directional kernels and then adding their responses. Consequently, an initial focus volume is obtained as follows:

$$F_z(x, y) = \sum_c \sum_i |I_z(x, y)_c \otimes h_i|, \quad (9)$$

where h_i denotes the i_{th} directional kernel and \otimes represents the convolution operator. The initial depth is then obtained on the basis of a 'winner takes it all' formula, where for any pixel (x, y) , the image number giving the maximum value of the focus measure is considered the initial depth for that pixel. Thus, a dense depth map is obtained as follows:

$$d(x, y) = \arg \max_z (F_z(x, y)), \quad (10)$$

Based on the initial depth map, we can extract the all-in-focus (AIF) image of the focus stack by stitching pixels from the images in the sequence corresponding to the labeled depth. The pixel value of I_{AIF} at location (x, y) is then represented as follows:

$$I_{AIF}(x, y) = I_{d(x, y)}(x, y), \quad (11)$$

where I denotes the original image in the sequence and $d(x, y)$ acts as an image index of the sequence representing the value of the initial depth at location (x, y) . Usually, the initial depth maps obtained from the initial focus volume are noisy. A cost aggregation method $\Gamma(\cdot)$ is then involved, which takes the initial focus volume $F_z(x, y)$ and $I_{AIF}(x, y)$ as input and provides an improved focus volume as output. We applied the cost aggregation method used in [25] to refine the initial volume. It uses a guided filtering operation to preserve edges while smoothing the focus volume.

$$\hat{F}_z(x, y) = \Gamma(F_z(x, y), I_{AIF}(x, y)), \quad (12)$$

$$\hat{d}(x, y) = \arg \max_z (\hat{F}_z(x, y)), \quad (13)$$

where $\hat{F}_z(x, y)$ and $\hat{d}(x, y)$ are the improved focus volume and the improved depth map, respectively.

3. Results and Discussion

In this section, first, we explain the experimental setup, which includes information about the datasets, methods for comparing results, and metrics used to evaluate the results. A comparative analysis of the results obtained through the state-of-the-art methods and the proposed methods is then presented.

3.1. Experimental Setup

The performance of the proposed method was evaluated through experiments using image sequences of synthetic and real objects. Synthetic image sequences of 14 objects, each consisting of 30 images, were obtained from the 4D light field benchmark [39], in which ground truth (GT) depth maps were available. Additionally, three real image sequences, i.e., *Balls*, *Kitchen*, and *Buddha*, were obtained from [25]. In the synthetic datasets, as ground

truth (GT) depth maps were provided, we quantitatively compared the estimated depth maps with the GT depth maps. To achieve this, we computed the root mean square error (RMSE) and correlation (CORR) to measure the degree of similarity between the estimated depth maps and the GT depth maps. RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{1}{|XY|} \sum_{\mathbf{x}} [D(\mathbf{x}) - \hat{D}(\mathbf{x})]^2}, \tag{14}$$

where $D(\mathbf{x})$ and $\hat{D}(\mathbf{x})$ represent the GT and the estimated depth maps, respectively, and $|XY|$ represents the total number of pixels in the depth map. A lower value of RMSE indicates a better depth map estimation. The correlation measure is computed as follows:

$$CORR = \frac{\sum_{\mathbf{x}} [D(\mathbf{x}) - \bar{D}] [\hat{D}(\mathbf{x}) - \bar{\hat{D}}]}{\sqrt{\sum_{\mathbf{x}} [D(\mathbf{x}) - \bar{D}]^2} \sqrt{\sum_{\mathbf{x}} [\hat{D}(\mathbf{x}) - \bar{\hat{D}}]^2}}, \tag{15}$$

where \bar{D} and $\bar{\hat{D}}$ represent the mean of the GT and the estimated depth maps, respectively. A higher value of the CORR measure depicts a better depth map estimation.

3.2. Comparative Analysis

First, we analyzed the performances of RDF and DRDF on all 14 synthetic datasets. RDF and DRDF, having the same filter size ($r_1 = 1, r_2 = 1, r_3 = 1$), were applied to all datasets to compute the focus volumes. The depth maps were estimated by taking the image number with the best focus measure in the optical direction. Without applying any enhancement method or post-processing of the focus volume or initial depth maps, RMSE and CORR measures were calculated with reference to the GTs. Figures 3 and 4 show the RMSE and CORR metrics, respectively. From the figures, it can be observed that the DRDF has shown a reasonable improvement for all datasets compared to the RDF. The depth maps computed for all synthetic objects from the DRDF provided lower RMSE and higher CORR in relation to their GTs. This indicates the effectiveness of the proposed focus measure.

Furthermore, in order to evaluate the performances of RDF and DRDF, with respect to different kernel sizes, we conducted experiments on the *Cotton* dataset. For the sake of simplicity, we fixed $r_1 = 1$ and changed the sizes of r_2 and r_3 . We applied kernels of different sizes on the image sequence, and RMSE and CORR measures were calculated of the estimated depth maps with respect to the GT depth map. Table 1 shows the RMSE, whereas Table 2 shows the CORR measures. It can be observed that DRDF consistently outperforms RDF across all different filter sizes. However, the results also indicate that increasing the sizes of the filters for both RDF and DRDF leads to a greater deviation from the ground truth and, hence, results in a loss of details.

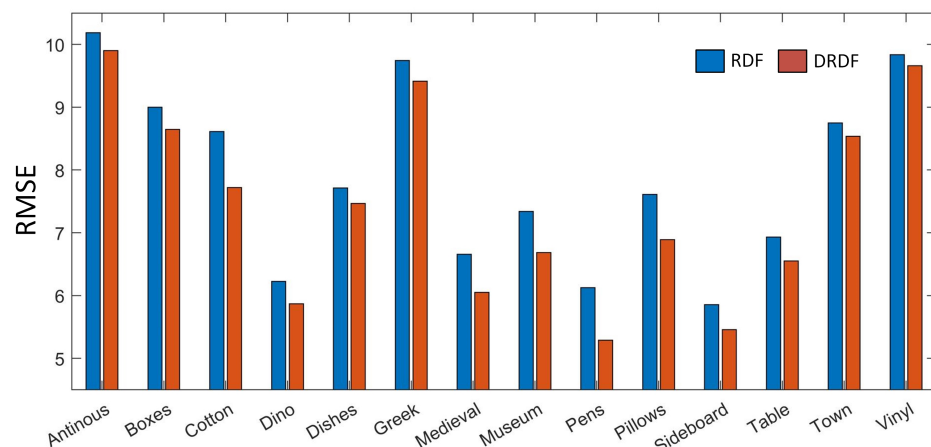


Figure 3. RMSE measures for synthetic datasets using RDF and DRDF. Datasets are labeled on the x-axis.

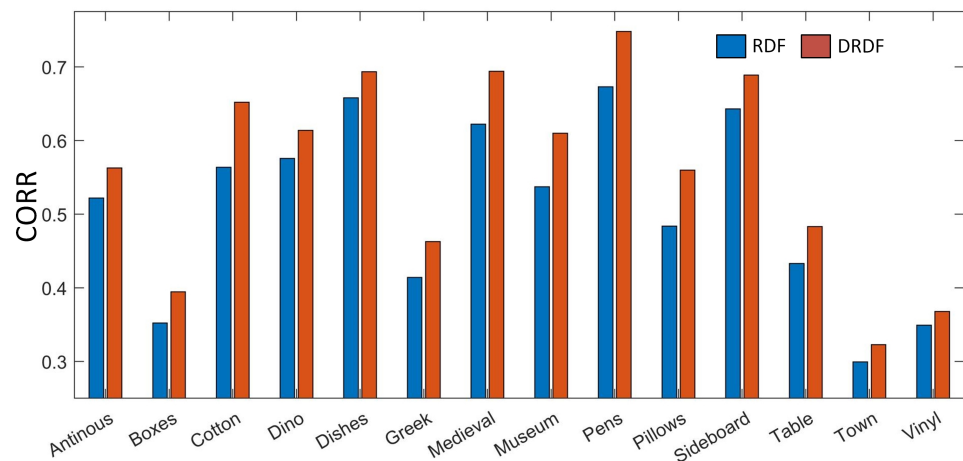


Figure 4. CORR measures for synthetic datasets using RDF and DRDF. Datasets are labeled on the x-axis.

Table 1. RMSE measure comparison: RDF vs. DRDF with varying ring sizes, r_2 and r_3 , and keeping $r_1 = 1$ for the Cotton dataset.

rsize	RDF	DRDF	RDF	DRDF	RDF	DRDF
	$r_2 = 1$	$r_2 = 1$	$r_2 = 2$	$r_2 = 2$	$r_2 = 3$	$r_2 = 3$
$r_3 = 1$	6.1262	5.2878	6.7047	5.5958	6.9941	5.9153
$r_3 = 2$	6.5750	5.5910	6.8765	5.8888	7.1658	6.1661
$r_3 = 3$	6.7978	5.8843	7.0922	6.1584	7.2921	6.3805

Table 2. CORR measure comparison: RDF vs. DRDF with varying ring sizes, r_2 and r_3 , and keeping $r_1 = 1$ for the Cotton dataset.

rsize	RDF	DRDF	RDF	DRDF	RDF	DRDF
	$r_2 = 1$	$r_2 = 1$	$r_2 = 2$	$r_2 = 2$	$r_2 = 3$	$r_2 = 3$
$r_3 = 1$	0.6728	0.7481	0.6190	0.7207	0.5885	0.6933
$r_3 = 2$	0.6313	0.7207	0.6007	0.6926	0.5717	0.6718
$r_3 = 3$	0.6087	0.6927	0.5786	0.6679	0.5608	0.6511

Next, the proposed method was compared to seven focus computation methods, i.e., gray-level variance (GLV) [7], modulus of the gradient of the color channel (MCG) [13], modified-Laplacian (ML) [3], sum and spread focus measure (FMSS) [23], reduced Tenengrad (RT) [14], multi-scale-morphological focus measure (MSM) [24], and ring difference filter (RDF) [25]. For visual comparisons, we constructed depth maps of synthetic datasets, *Antinous*, *Cotton*, and *Pens*, using different methods, as shown in Figure 5. The first column of each dataset represents the initial unaggregated depth maps, while the second column shows the cost-aggregated depth maps using the cost aggregation method proposed in [25]. The *Antinous* dataset presented a challenge to all methods in the unaggregated depth maps, except MSM, RDF, and DRDF, which correctly captured the impact of shadows on the depth map. In contrast, other methods misinterpreted the shadows as significant edges in the dataset, resulting in erroneous depth representations. When aggregated, MCG also performed well, along with RDF and DRDF, by discarding shadows and carefully detecting the edges; but, the performance of MSM deteriorated as it did not carve out the edges properly. For the *Cotton* dataset, GLV, MCG, RT, and FMSS attempted to detect the detailed features, but those detailed features appeared as white lines instead of distinct grayscale fades in the unaggregated depth maps. This became evident in aggregated depth maps, where FMSS showed an irregular depiction of the object in the dataset, GLV showed a side corner that did not exist in the initial depth map, and MCG compacted the edges, resulting in an irregular bump on the head of the object in the depth map. However, ML, MSM, RDF, and DRDF had relatively distinct grayscale fades in the unaggregated

depth maps, which became more evident in the aggregated depth maps. Lastly, in the *Pens* dataset, GLV, FMSS, and DRDF outperformed others in suppressing the noise in the background, while MCG, ML, RT, and RDF showed some speckle noise in the initial depth map, and MSM completely misrepresented the shape of the container of the pens. When aggregated, MCG and RT showed a black patch at the bottom of the pen container in the depth map, ML did not properly carve out the edges of the pens, FMSS and MSM showed completely irregular shapes that did not coincide with the objects in the dataset, and GLV inaccurately represented some pens in the depth map as being very near, which they were not. On the other hand, RDF and DRDF played better roles in capturing details and not misrepresenting items very close or far away in the aggregated depth maps. Among RDF and DRDF, DRDF had more grayscale fadedness and fewer patches in the objects than RDF. Hence, some methods performed well in one dataset or were good in either the initial depth map representation or aggregated depth map representation, but not all; DRDF proved to be one of the best-performing focus measures in all datasets and unaggregated and aggregated depth maps.

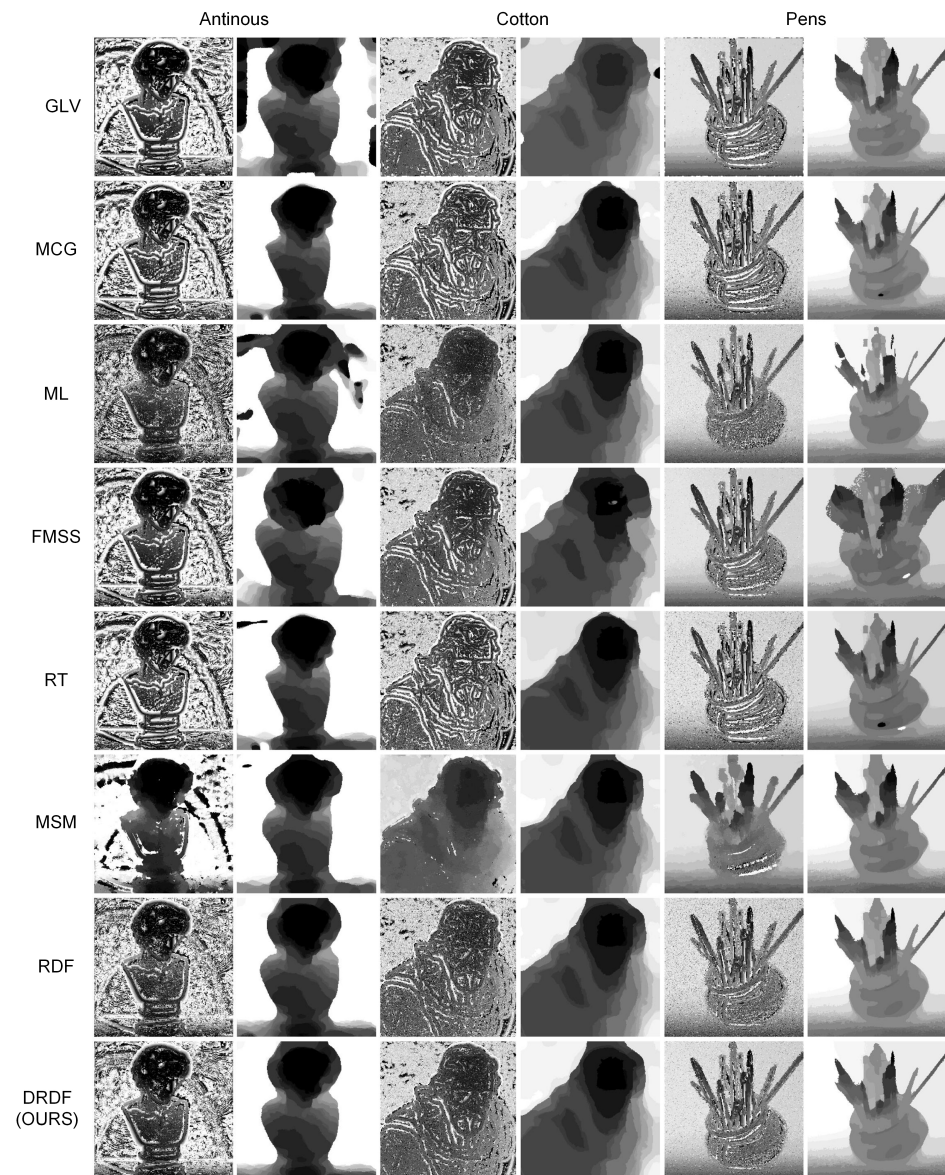


Figure 5. Depth maps of synthetic datasets, i.e., *Antinous*, *Cotton*, and *Pens*, using different methods. The first column of each dataset represents the initial depth maps before aggregation, while the second column shows the depth maps after applying the cost aggregation method.

For real datasets, we constructed depth maps of *Balls*, *Kitchen*, and *Buddha* as shown in Figure 6. The first column of each dataset represents the initial unaggregated depth maps while the second column represents the cost-aggregated depth maps. The *Balls* dataset consists of 25 images, each with a resolution of 640×360 . In the unaggregated depth maps, MSM produced the most smoothest depth map followed by ML while others tried to show detailed texture of the objects present in the dataset. However, when aggregated, ML produced a blurry depth map as it failed to capture the extra details in the objects, FMSS misrepresented the objects and their boundaries, and GLV produced some black patches. While RDF, RT, and DRDF performed better than others, MCG and MSM produced the best results with the fewest black patches and the best edge detection. The *Kitchen* dataset, comprising of 11 images with a resolution of 774×518 , did not provide much detail on the generated depth maps due to the limited number of images. In the unaggregated depth maps, MCG showed the noisiest background, followed by RDF and RT. However, after the aggregation, GLV had some unusual black patches, ML showed white patches in some objects, FMSS distorted the shapes of the objects, MSM merged two objects together, giving them the same depth values and RT produced small visible white patches in the depth map. MCG, RDF, and DRDF were among the best-performing methods, but MCG had a visible white dot in the object shape, which indicates inaccurate depth estimates. In addition, RDF showed some discontinuities in-depth maps, which also indicates imprecise depth values, while DRDF provided smoother depth maps, which indicates a better perception of depth values. Finally, the *Buddha* dataset, consisting of 29 images with varying focus settings, and a resolution of 768×768 pixels, showed that ML produced the noisiest depth map in the unaggregated state, followed by RDF and DRDF. However, after the aggregation, DRDF performed the best. FMSS distorted the shapes of the objects; RT and MCG showed white dots inside the depth map, which stipulates inaccuracies. ML provided a false perception of depth in some areas. GLV, MSM, RDF, and DRDF were among the better-performing methods for this dataset; however, GLV, MSM, and RDF incorporated a white dot inside the object in the depth map when examined closely. In contrast, DRDF reduced the white dot, which indicates a reduction in the discontinuities and an improved depth map.

3.3. Complexity Analysis

When comparing the time complexity of RDF with the proposed DRDF, it is obvious that RDF is more efficient than DRDF. In RDF, a single kernel is used, while in DRDF, 2D convolutions are used with six kernels for each image in the sequence. Considering the image size $X \times Y$ for Z images and kernel size $k \times k$, then the time complexity for RDF for point-wise operations is $O(XYZk^2)$, whereas the time complexity for DRDF is $O(6XYZk^2)$, which is six times more than RDF. In addition, there are many algorithms proposed in the literature to conduct the convolution in optimal time [40].

We implemented the proposed method in MATLAB, and for most of the comparative methods, we utilized the MATLAB code of the authors. We performed experiments on a PC with Intel(R) CPU, 16GB RAM, and the computational time was recorded. The computational times taken by all the comparative methods, including ours, for both synthetic and real datasets, are shown in Table 3. In the synthetic datasets, the size of each image was not very large, resulting in the generation of their depth maps at relatively quicker times. However, the MSM method took the longest time compared to the other methods. On the other hand, in real datasets, the image sizes were larger compared to the synthetic datasets, which led to longer computation times when generating their depth maps. It is evident from the results that the time complexity of our method is comparable to the other methods in both the synthetic and real datasets. Our proposed method (DRDF) was more efficient than the GLV, FMSS, and MSM methods in synthetic datasets as well as outperforming MCG in real datasets. However, it was still slower than ML, RT, and RDF methods in both synthetic and real datasets.

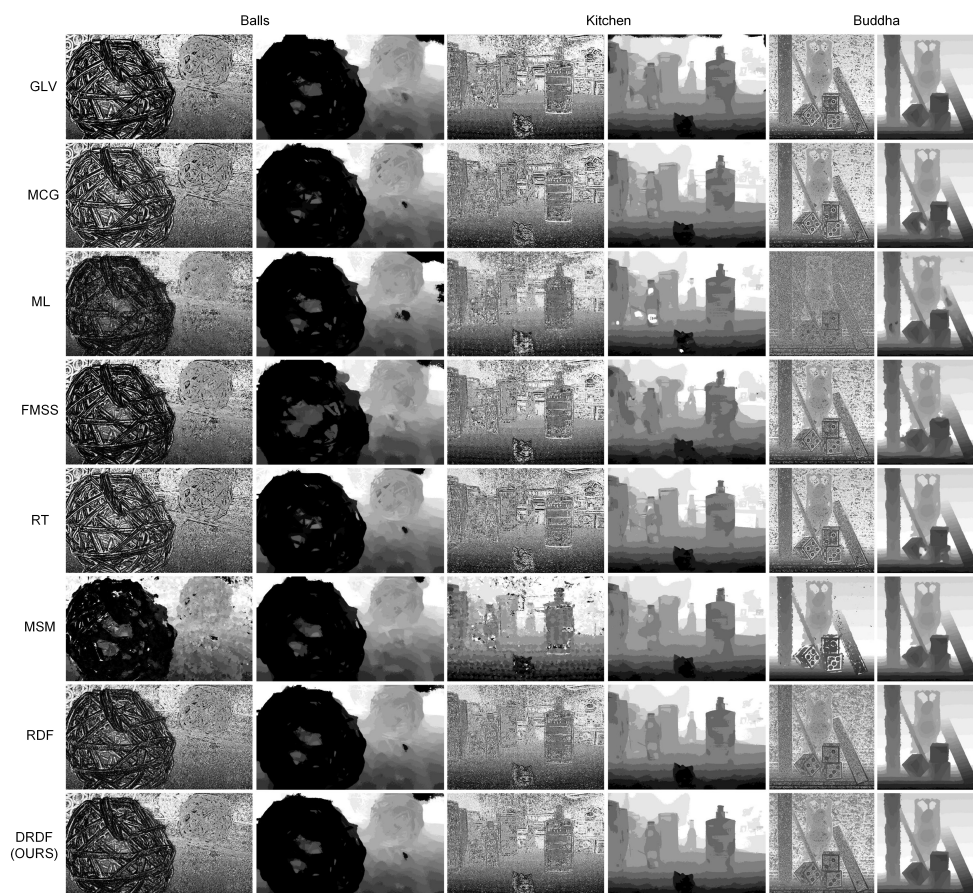


Figure 6. Depth maps of real datasets, i.e., *Balls*, *Kitchen*, and *Buddha*, using different methods. The first column of each dataset represents the initial depth maps before aggregation, while the second column shows the depth maps after applying the cost aggregation method.

Table 3. Computational times (in seconds) for the comparative methods and the proposed method for symmetric and real datasets.

Methods	Synthetic			Real		
	Antinous	Cotton	Pens	Balls	Kitchen	Buddha
GLV	0.24	0.17	0.17	0.92	0.64	2.55
MCG	0.14	0.09	0.09	0.38	0.26	0.98
ML	0.07	0.06	0.06	0.16	0.13	0.42
FMSS	0.47	0.44	0.41	1.49	1.08	4.33
RT	0.06	0.04	0.04	0.15	0.10	0.40
MSM	1.23	1.07	0.99	2.10	1.21	4.41
RDF	0.09	0.06	0.06	0.02	0.14	0.54
DRDF (OURS)	0.15	0.14	0.14	0.32	0.24	0.92

4. Conclusions

In this article, we propose a robust focus measure called the directional ring difference filter (DRDF), which improves the performance of the state-of-the-art method called the ring difference filter (RDF). In order to fix the response cancellation problem, instead of a 2D single kernel, multiple directional kernels were applied to images, and their responses were aggregated to compute the level of sharpness. Extensive experiments were conducted on both synthetic and real image datasets, and the results demonstrate that DRDF outperforms traditional focus measures in terms of noise handling and producing high-quality 3D shape estimates of objects.

Author Contributions: Conceptualization, K.A. and M.T.M.; methodology, M.T.M.; software, K.A.; validation, K.A. and M.T.M.; formal analysis, K.A.; investigation, K.A.; resources, M.T.M.; data curation, K.A.; writing—original draft preparation, K.A.; writing—review and editing, M.T.M.; visualization, K.A.; supervision, M.T.M.; project administration, M.T.M.; funding acquisition, M.T.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT: Ministry of Science and ICT) (2022R1F1A1071452).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

Abbreviations

The following abbreviations are used in this manuscript:

FM	focus measure
ML	modified Laplacian
SFF	shape-from-focus
DRDF	directional ring difference filter
RDF	ring difference filter
RCP	response cancellation problem
GT	ground truth
FV	focus volume
FMSS	focus measure sum and spread
MCG	modulus color gradient
GLV	gray-level variance
RT	reduced Tenengrad

References

1. Nourbakhsh, I.R.; Andre, D.; Tomasi, C.; Genesereth, M.R. Mobile robot obstacle avoidance via depth from focus. *Robot. Auton. Syst.* **1997**, *22*, 151–158. [[CrossRef](#)]
2. Lin, H.Y.; Subbarao, M. Vision system for fast 3-D model reconstruction. *Opt. Eng.* **2004**, *43*, 1651–1664. [[CrossRef](#)]
3. Nayar, S.K.; Nakagawa, Y. Shape from focus. *IEEE Trans. Pattern Anal. Mach. Intell.* **1994**, *16*, 824–831. [[CrossRef](#)]
4. Mahmood, M.T.; Choi, T.S. Nonlinear approach for enhancement of image focus volume in shape from focus. *IEEE Trans. Image Process.* **2012**, *21*, 2866–2873. [[CrossRef](#)]
5. Pertuz, S.; Puig, D.; Garcia, M.A. Analysis of focus measure operators for shape-from-focus. *Pattern Recognit.* **2013**, *46*, 1415–1432. [[CrossRef](#)]
6. Boshtayeva, M.; Hafner, D.; Weickert, J. A focus fusion framework with anisotropic depth map smoothing. *Pattern Recognit.* **2015**, *48*, 3310–3323. [[CrossRef](#)]
7. Ali, U.; Mahmood, M.T. Robust focus volume regularization in shape from focus. *IEEE Trans. Image Process.* **2021**, *30*, 7215–7227. [[CrossRef](#)]
8. Shirvaikar, M.V. An optimal measure for camera focus and exposure. In Proceedings of the Thirty-Sixth Southeastern Symposium on System Theory, Atlanta, GA, USA, 16 March 2004; pp. 472–475.
9. Wee, C.Y.; Paramesran, R. Measure of image sharpness using eigenvalues. *Inf. Sci.* **2007**, *177*, 2533–2552. [[CrossRef](#)]
10. Gaidhane, V.H.; Hote, Y.V.; Singh, V. Image focus measure based on polynomial coefficients and spectral radius. *Signal Image Video Process.* **2015**, *9*, 203–211. [[CrossRef](#)]
11. Rajevelceltha, J.; Gaidhane, V.H. A novel approach for image focus measure. *Signal Image Video Process.* **2021**, *15*, 547–555. [[CrossRef](#)]
12. Feichtenhofer, C.; Fassold, H.; Schallauer, P. A perceptual image sharpness metric based on local edge gradient analysis. *IEEE Signal Process. Lett.* **2013**, *20*, 379–382. [[CrossRef](#)]
13. Hurtado-Pérez, R.; Toxqui-Quitl, C.; Padilla-Vivanco, A.; Aguilar-Valdez, J.F.; Ortega-Mendoza, G. Focus measure method based on the modulus of the gradient of the color planes for digital microscopy. *Opt. Eng.* **2018**, *57*, 023106. [[CrossRef](#)]
14. Helmy, I.; Choi, W. Reduced Tenengrad focus measure for performance improvement of astronomical images. In Proceedings of the 2022 International Conference on Electronics, Information, and Communication (ICEIC), Jeju, Republic of Korea, 6–9 February 2022; pp. 1–4.
15. Subbarao, M.; Choi, T.S.; Nikzad, A. Focusing techniques. *Opt. Eng.* **1993**, *32*, 2824–2836. [[CrossRef](#)]
16. Hu, Z.; Liang, W.; Ding, D.; Wei, G. An improved multi-focus image fusion algorithm based on multi-scale weighted focus measure. *Appl. Intell.* **2021**, *51*, 4453–4469. [[CrossRef](#)]

17. Kautsky, J.; Flusser, J.; Zitova, B.; Šimberová, S. A new wavelet-based measure of image focus. *Pattern Recognit. Lett.* **2002**, *23*, 1785–1794. [[CrossRef](#)]
18. Minhas, R.; Mohammed, A.A.; Wu, Q.J. Shape from focus using fast discrete curvelet transform. *Pattern Recognit.* **2011**, *44*, 839–853. [[CrossRef](#)]
19. Jeon, J.; Lee, J.; Paik, J. Robust focus measure for unsupervised auto-focusing based on optimum discrete cosine transform coefficients. *IEEE Trans. Consum. Electron.* **2011**, *57*, 1–5. [[CrossRef](#)]
20. Zhang, Z.; Liu, Y.; Xiong, Z.; Li, J.; Zhang, M. Focus and blurriness measure using reorganized DCT coefficients for an autofocus application. *IEEE Trans. Circuits Syst. Video Technol.* **2016**, *28*, 15–30. [[CrossRef](#)]
21. Nie, X.; Xiao, B.; Bi, X.; Li, W.; Gao, X. A focus measure in discrete cosine transform domain for multi-focus image fast fusion. *Neurocomputing* **2021**, *465*, 93–102. [[CrossRef](#)]
22. Yap, P.T.; Raveendran, P. Image focus measure based on Chebyshev moments. *IEE Proc.-Vis. Image Signal Process.* **2004**, *151*, 128–136. [[CrossRef](#)]
23. Mutahira, H.; Ahmad, B.; Muhammad, M.S.; Shin, D.R. Focus measurement in color space for shape from focus systems. *IEEE Access* **2021**, *9*, 103291–103310. [[CrossRef](#)]
24. Zhang, Y.; Bai, X.; Wang, T. Boundary finding based multi-focus image fusion through multi-scale morphological focus-measure. *Inf. Fusion* **2017**, *35*, 81–101. [[CrossRef](#)]
25. Jeon, H.G.; Surh, J.; Im, S.; Kweon, I.S. Ring Difference Filter for Fast and Noise Robust Depth From Focus. *IEEE Trans. Image Process.* **2019**, *29*, 1045–1060. [[CrossRef](#)]
26. Minhas, R.; Mohammed, A.A.; Wu, Q.J.; Sid-Ahmed, M.A. 3D shape from focus and depth map computation using steerable filters. In *Image Analysis and Recognition, Proceedings of the 6th International Conference, ICIAR 2009, Halifax, Canada, 6–8 July 2009*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 573–583.
27. Wang, J.; Qu, H.; Wei, Y.; Xie, M.; Xu, J.; Zhang, Z. Multi-focus image fusion based on quad-tree decomposition and edge-weighted focus measure. *Signal Process.* **2022**, *198*, 108590. [[CrossRef](#)]
28. Guo, L.; Liu, L. A Perceptual-Based Robust Measure of Image Focus. *IEEE Signal Process. Lett.* **2022**, *29*, 2717–2721. [[CrossRef](#)]
29. Jang, H.S.; Yun, G.; Mahmood, M.T.; Kang, M.K. Optimal Sampling for Shape from Focus by Using Gaussian Process Regression. In *Proceedings of the 2020 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 4–6 January 2020*; pp. 1–4. [[CrossRef](#)]
30. Jang, H.S.; Muhammad, M.S.; Choi, T.S. Optimizing Image Focus for Shape from Focus Through Locally Weighted Non-Parametric Regression. *IEEE Access* **2019**, *7*, 74393–74400. [[CrossRef](#)]
31. Fu, B.; He, R.; Yuan, Y.; Jia, W.; Yang, S.; Liu, F. Shape from focus using gradient of focus measure curve. *Opt. Lasers Eng.* **2023**, *160*, 107320. [[CrossRef](#)]
32. Gladines, J.; Sels, S.; De Boi, I.; Vanlanduit, S. A phase correlation based peak detection method for accurate shape from focus measurements. *Measurement* **2023**, *213*, 112726. [[CrossRef](#)]
33. Ali, U.; Mahmood, M.T. Energy minimization for image focus volume in shape from focus. *Pattern Recognit.* **2022**, *126*, 108559. [[CrossRef](#)]
34. Tomasi, C.; Manduchi, R. Bilateral filtering for gray and color images. In *Proceedings of the Sixth International Conference on Computer Vision, Bombay, India, 7 January 1998*; pp. 839–846.
35. Petschnigg, G.; Szeliski, R.; Agrawala, M.; Cohen, M.; Hoppe, H.; Toyama, K. Digital photography with flash and no-flash image pairs. *ACM Trans. Graph.* **2004**, *23*, 664–672. [[CrossRef](#)]
36. He, K.; Sun, J.; Tang, X. Guided image filtering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *35*, 1397–1409. [[CrossRef](#)] [[PubMed](#)]
37. Ali, U.; Lee, I.H.; Mahmood, M.T. Guided image filtering in shape-from-focus: A comparative analysis. *Pattern Recognit.* **2021**, *111*, 107670. [[CrossRef](#)]
38. Abuolaim, A.; Brown, M.S. Defocus deblurring using dual-pixel data. In *Computer Vision—ECCV 2020, Proceedings of the 16th European Conference, Glasgow, UK, 23–28 August 2020*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 111–126.
39. Honauer, K.; Johannsen, O.; Kondermann, D.; Goldluecke, B. A dataset and evaluation methodology for depth estimation on 4D light fields. In *Computer Vision—ACCV 2016, Proceedings of the 13th Asian Conference on Computer Vision, Taipei, Taiwan, 20–24 November 2016*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 19–34.
40. Seznec, M.; Gac, N.; Orioux, F.; Sashala Naik, A. Computing large 2D convolutions on GPU efficiently with the im2tensor algorithm. *J. Real-Time Image Process.* **2022**, *19*, 1035–1047. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.