

Article

Enhancing Focus Volume through Perceptual Focus Factor in 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: Shape From Focus (SFF) reconstructs a scene's shape using a series of images with varied focus settings. However, the effectiveness of SFF largely depends on the Focus Measure (FM) used, which is prone to noise-induced inaccuracies in focus values. To address these issues, we introduce a perception-influenced factor to refine the traditional Focus Volume (FV) derived from a traditional FM. Owing to the strong relationship between the Difference of Gaussians (DoG) and how the visual system perceives edges in a scene, we apply it to local areas of the image sequence by segmenting the image sequence into non-overlapping blocks. This process yields a new metric, the Perceptual Focus Factor (PFF), which we combine with the traditional FV to obtain an enhanced FV and, ultimately, an enhanced depth map. Intensive experiments are conducted by using fourteen synthetic and six real-world data sets. The performance of the proposed method is evaluated using quantitative measures, such as Root Mean Square Error (RMSE) and correlation. For fourteen synthetic data sets, the average RMSE measure of 6.88 and correction measure of 0.65 are obtained, which are improved through PFF from an RMSE of 7.44 and correlation of 0.56, respectively. Experimental results and comparative analysis demonstrate that the proposed approach outperforms the traditional state-of-the-art FMs in extracting depth maps.

Keywords: shape from focus; focus measure; directional ring difference filter; perceptual focus factor; depth map

MSC: 65D19



Citation: Ashfaq, K.; Mahmood, M.T. Enhancing Focus Volume through Perceptual Focus Factor in Shape-from-Focus. *Mathematics* **2024**, *12*, 102. <https://doi.org/10.3390/math12010102>

Academic Editor: Samaneh Mazaheri

Received: 5 December 2023

Revised: 23 December 2023

Accepted: 25 December 2023

Published: 27 December 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/>).

1. Introduction

In computer vision, a primary focus is on creating 3D representations of scenes using just 2D camera images, a complex but vital task. A key technique in this field is Shape From Focus (SFF) or Depth From Focus (DFF), which reconstructs 3D shapes by utilizing the focus cue of an image. This method estimates depth by analyzing various focus positions in a scene and determining the correct focus for each pixel. In contrast, other methods like depth from motion [1], shape from contour [2], shape from reflection [3], and shape from time of flight [4], also provide depth estimates, but each comes with specific limitations. For instance, shape from motion can struggle with ambiguous or degenerate motion scenarios where multiple 3D shapes can explain the observed 2D motion, shape from contour relies heavily on the presence of clear edges and contours in the image, making it less effective for textureless or featureless objects. Shape from reflection can struggle with objects that have complex or non-Lambertian reflective properties and shape from time of flight requires specialized hardware. Meanwhile, SFF stands out for its simplicity, ease of implementation, no need for special hardware, and its ability to generate a detailed depth map.

In SFF, evaluating the relative degree of focus for each pixel using an FM is considered the most crucial step. Numerous FMs have been proposed in the literature, as highlighted in [5]. These can be broadly categorized into gradient-based, statistics-based,

and transformation-based methods. Gradient-based methods involve analyzing the focus degree of each pixel by its derivative. For instance, ref. [6] utilizes a multi-scale template with varying weights and convolves it with the output of modified Laplacian applied to images, where modified Laplacian is a famous gradient operator. Another example is the edge-weighted modified Laplacian focus measure by [7], which uses the modified Laplacian weighted by the Gaussian kernel. Statistics-based FMs operate by applying statistical measures to local pixels. This includes the approach [8], which is based on probability coefficients and modified entropy, a fuzzy entropy-based image focus quality measure [9], and absolute central moment [10] that exploits histograms obtained from the grayscale version of the image. Transformation-based FMs use the frequency components of an image to evaluate sharpness. For example, [11] is based on the discrete wavelet transform, ref. [12] focuses on the energy of high-frequency components in the S-transform, and [13] involves Bayes-spectral-entropy-based focus evaluation using a discrete cosine transform. Moreover, FMs not fitting these categories include a morphology-based focus measure in a quad-tree structure [14], a new, robust to noise, directional ring difference filter [15] that exploits directionality in evaluating focus, and a perceptual-based measure [16].

Due to the diversity in the imaging content, condition, and capturing devices and the limited capabilities of FM operators, generally, the focus values in FV are erroneous and thus it results in an inaccurate depth map. To address these limitations of FM operators, several techniques have been proposed for the improvement of FV and depth maps. Filtering techniques, such as anisotropic diffusion, have been applied that compute weights adaptively from local structures [17,18]. In another work, depth reconstruction has been formulated in a total variation-based framework that includes a nonconvex data fidelity and a convex regularization term [19]. Recently, Weighted Least Squares (WLS) techniques have also been proposed to improve the initial FV [20]. A common drawback of these techniques is that these techniques do not address the problem of preserving structural edges and fine details in recovered shapes. Moreover, the performance of these techniques relies on the accuracy of the initial depth map, which in turn, depends on the quality of the initial FV. Recently, deep learning methods have been suggested that are based on an auto-encoder-style convolutional neural network [21–23]. In [22], the authors propose a deep learning-based method to estimate depth maps and all-in-focus images. In a recent work [23], authors suggested a CNN-based model to compute the deep differential focus volume (DFV) by applying the first-order derivative with the stacked features over different focal distances. Deep learning methods require large datasets of focal stacks with true depth maps for the model training. In the case of shape from focus, a limited number of training datasets are available.

Traditional focus measure operators compute focus quality for each pixel in the input sequence by using a sliding window. These pixel-based focus measures suffer from many limitations. When depth is deduced from these focus measures, these often yield incorrect depth values and representations. This inaccuracy arises from several factors. For example, the scenery and content of an image can significantly affect the evaluation. Issues such as focus measures misinterpreting shadows as objects and assigning them erroneous depths are common. The window size chosen for applying focus measures during convolution with the image sequence also impacts the accuracy of focus evaluation. Furthermore, lighting conditions in a scene can hinder the focus measures' ability to distinguish objects and their boundaries. These issues frequently result in an inaccurate depth representation of the scene's objects and contribute to substantial background noise.

In this paper, to address the above-stated issues, we propose a shape-from-focus method that enhances the traditional FV by utilizing the Perceptual Focus Factor (PFF), which is based on biological perception principles. In the first phase, a traditional focus volume is obtained by applying any traditional focus measure on the input image sequence. In the second phase of our method, the PFF is computed for each pixel of the input sequence by dividing it into small, non-overlapping blocks. The PFF has two major aspects that help in improving the traditional focus measures: (1) PFF is computed through non-

overlapping blocks whereas normal focus measures use sliding windows, and (2) the responses from DoG or LoG operator mimic the biological responses. Finally, the enhanced resultant FV is obtained by scaling the traditional FV with the PFF. Intensive experiments are conducted by using fourteen synthetic and six real-world data sets. The performance of the proposed method is evaluated using quantitative measures such as Root Mean Square Error (RMSE) and correlation. Experimental results and comparative analysis demonstrate that the proposed approach outperforms the traditional state-of-the-art FMs in extracting depth maps.

The rest of the paper is organized as follows: the SFF system, focus measures, and perceptual focus measures are explained in Section 2; the proposed method and its components are presented in Section 3; the experimental setup, analysis, results, and comparative analysis are provided in Section 4. Finally, Section 5 concludes this study.

2. Background

2.1. Shape from Focus

A general framework of Shape From Focus (SFF) is shown in Figure 1, which begins by capturing a sequence of images with varying focus settings, termed as the image sequence [24]. This is achieved by incrementally changing the focus setting of the camera system or by translating the object toward the camera with a static position. The next step involves evaluating the focus degree of each pixel by applying a Focus Measure (FM) operator to the image sequence, which results in a Focus Volume (FV). Generally, a single FM operator cannot tackle the wide range of imaging content and imaging conditions [5]. This results in inaccurate focus measurements and thus the image focus volume needs to be enhanced. In the third step, the initial focus volume is improved and outputs an enhanced image focus volume. For this, a large number of linear and non-linear approaches are suggested in the literature [25]. Finally, a depth map is constructed by selecting the image frame from the sequence that exhibits the highest focus in the optical direction.

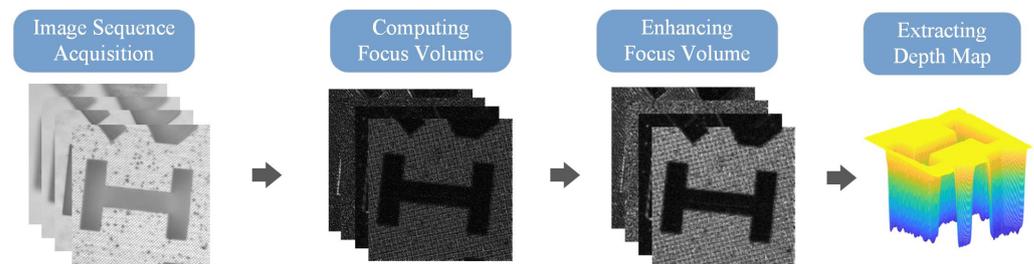


Figure 1. Shape from focus system.

2.2. Focus Measures

Let us assume that the image sequence is represented by $I_z^{(c)}(p)$ and consists of Z images. Each image has dimensions of $X \times Y$, where X represents rows, and Y represents columns. Then, the pixel intensity at coordinate (x, y) in the c -th color channel of the z -th image in $I_z^{(c)}(p)$ is denoted as $I_z^{(c)}(x, y)$, where $c \in \{r, g, b\}$ represents the different color channels of an image. To simplify the notation, we will use p as the subscript to indicate the coordinates of a pixel in the 3D volume as $I_z^{(c)}(p)$, where $p = (x, y) \in \mathbb{R}^2$. For a given input image sequence $I_z^{(c)}(p)$, an image focus volume $f_z^{(u)}(p)$ is obtained by applying a focus measure (FM) operator. The focus value or sharpness level for each pixel is computed by applying a focus measure operator (u) to $I_z^{(c)}(p)$ as:

$$f_z^{(u)}(p) = \sum_c I_z^{(c)}(p) \otimes u \tag{1}$$

where \otimes represents the two-dimensional convolution operator, and u corresponds to any suitable FM operator. In the literature, a large number of FM operators have been proposed,

and each FM operator exhibits some strengths and weaknesses. However, the most popular and commonly used FM operator is modified-Laplacian (ML) [24], which computes the focus value for a pixel as:

$$f_z^{(ML)}(p) = \sum_c \left[\left| \frac{\partial^2 I_z^{(c)}(p)}{\partial x^2} \right| + \left| \frac{\partial^2 I_z^{(c)}(p)}{\partial y^2} \right| \right], \tag{2}$$

where $\frac{\partial^2(\cdot)}{\partial x^2}$ and $\frac{\partial^2(\cdot)}{\partial y^2}$ are the second-order partial derivatives in the x - and y -directions. Thus, the first and second terms on the right-hand side are the absolute values of second-order derivatives in the x and y -dimensions, respectively. Another well-known FM operator is $f^{(GLV)}$ [26], which computes the variance of image gray levels (GLV) within a neighborhood window and is given by

$$f_z^{(GLV)}(p) = \sum_c \left[\frac{1}{|\mathcal{N}|} \sum_{(q) \in \mathcal{N}(p)} (I_z(q) - \mu)^2 \right], \tag{3}$$

where μ and $|\mathcal{N}|$ are the mean gray-level and total number of pixels in the small neighborhood window \mathcal{N} centered at (p) , respectively. One more famous FM operator is $f^{(TEN)}$ [27–29], known as Tenengrad (TEN), which computes the sharpness level as

$$f_z^{(TEN)}(p) = \sum_c \left[|G_x \otimes I(c)_z(p)| + |G_y \otimes I_z^{(c)}(p)| \right], \tag{4}$$

where G_x and G_y are the Sobel operators in the x - and y -directions, respectively. A detailed study about the focus measure can be found in [5].

2.3. Perceptual Focus Measures

In literature, a few focus measures are proposed, which are inspired by the early visual information processing mechanisms in the biological visual system, so named the Perceptual Focus Measures (PFM) [16,30]. In perceptual focus measures, the focus values are computed based on non-overlapping blocks. For instance, in [30], blocks of sizes 32×32 pixels are used to compute the perceptual focus measure. In recently proposed perceptual focus measure [16], the first image is partitioned into non-overlapping blocks of 16×16 , and then the Difference of Gaussian (DoG) operator is applied to each block. The DoG operator is defined as:

$$DoG(x, y) = G(x, y; \sigma_1) - G(x, y; \sigma_2) \tag{5}$$

where $G(x, y; \sigma)$ is defined as

$$G(x, y; \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}. \tag{6}$$

It was discovered that the responses of on-center and off-center cells in the receptive fields of the human visual system closely match the responses of the Laplacian of Gaussian (LoG) operator [31] on an objects' edges in an image. However, the direct application of LoG is inefficient computationally due to its second derivative computation requirement. Instead, we use the Difference of Gaussians (DoG), which is an effective approximation of LoG. Further, the DoG operator is a better approximation to the LoG operator if the ratio between standard deviations is $\sigma_2 : \sigma_1 = 1.6 : 1$ [32,33]. Usually, for one image, a scalar measure is computed by combining the responses from all blocks that are utilized to investigate the sharpest image in the sequence.

3. Proposed Method

The proposed scheme for depth estimation is shown in Figure 2. The proposed method consists of three main parts. In the first step, the traditional focus volume is computed, then in the second step, the perceptual focus factor for each pixel is computed, and finally, in the third step, the depth map is obtained from the combined and enhanced focus volume.

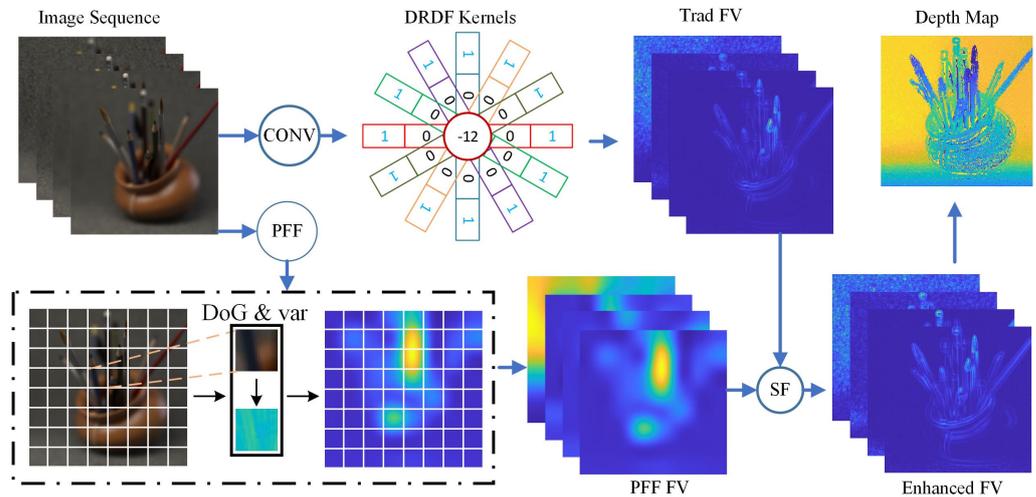


Figure 2. Our proposed method. ‘CONV’ denotes the convolution of input images with the Directional Ring Difference Filter (DRDF) kernels, ‘PFF’ denotes our Perceptual Focus Factor, and ‘SF’ denotes the Scaling Factor that is used in Equation (12) to combine the Perceptual Focus Factor (PFF), Focus Volume (FV), and Traditional (‘Trad’) FV. ‘DoG’ and ‘var’ represent the Difference of Gaussians and variance of the block after applying DoG, respectively.

3.1. Traditional Focus Volume

In the proposed framework, any traditional focus measure operator can be applied to obtain the traditional focus volume $f_z(p)$. In this work, we are using our own proposed focus measure, the Directional Ring Difference Filter (DRDF) [15], which is an improved and robust to noise version of the state-of-the-art Ring Difference Filter (RDF) focus measure [34]. DRDF addresses the response cancellation problem encountered in RDF. In DRDF, pixel focus quality is determined by aggregating responses from multiple kernels in different directions. DRDF’s kernels $h_i, i \in \{1, 2, 3, 4, 5, 6\}$ each of size 5×5 in six directions can be defined as:

$$\begin{cases} h_1 = [0\ 0\ 0\ 0\ 0; 0\ 0\ 0\ 0\ 0; -1\ 0\ 2\ 0\ -1; 0\ 0\ 0\ 0\ 0; 0\ 0\ 0\ 0\ 0], \\ h_2 = [0\ 0\ 0\ 0\ 0; 0\ 0\ 0\ 0\ -1; 0\ 0\ 2\ 0\ 0; -1\ 0\ 0\ 0\ 0; 0\ 0\ 0\ 0\ 0], \\ h_3 = [0\ 0\ 0\ -1\ 0; 0\ 0\ 0\ 0\ 0; 0\ 0\ 2\ 0\ 0; 0\ 0\ 0\ 0\ 0; 0\ -1\ 0\ 0\ 0], \\ h_4 = [0\ 0\ -1\ 0\ 0; 0\ 0\ 0\ 0\ 0; 0\ 0\ 2\ 0\ 0; 0\ 0\ 0\ 0\ 0; 0\ 0\ -1\ 0\ 0], \\ h_5 = [0\ -1\ 0\ 0\ 0; 0\ 0\ 0\ 0\ 0; 0\ 0\ 2\ 0\ 0; 0\ 0\ 0\ 0\ 0; 0\ 0\ 0\ -1\ 0], \\ h_6 = [0\ 0\ 0\ 0\ 0; -1\ 0\ 0\ 0\ 0; 0\ 0\ 2\ 0\ 0; 0\ 0\ 0\ 0\ -1; 0\ 0\ 0\ 0\ 0]. \end{cases} \quad (7)$$

Convolving the input image sequence $I_z^{(c)}(p)$ with these kernels $h_i, i \in \{1, 2, 3, 4, 5, 6\}$ and aggregating their responses will give us a traditional focus volume $f_z(p)$ as:

$$f_z(p) = \sum_c \sum_i |I_z^{(c)}(p) \otimes h_i| \quad (8)$$

3.2. Perceptual Focus Factor

After calculating the traditional FV, the Perceptual Focus Factor (PFF) is computed for each pixel based on the concept of Perceptual Focus Measures (PFM). In PFM, an image is divided into non-overlapping blocks, a scalar value per block is computed, and then by

aggregating responses from all blocks, a scalar measure is obtained for the entire image. In computing PFF, an image is divided into non-overlapping blocks as in PFM; however, a measure is computed for each pixel, i.e., a PFF volume of the same size as FV is obtained. Let $b_z^{(c)}(i, j) = I_z^{(c)}(i : i + m, j : j + n)$ be a non-overlapping image block having size $m \times n$ at $p(i, j)$ extracted from the original image sequence $I_z^{(c)}(p)$. Next, we apply a DoG on that block. By applying DoG, we successfully extract essential features from each block, representing them as numerical values that are driven through perceptually based logic since DoG is correlated with how our human visual system behaves. We aim to utilize it to influence the traditional focus volume, FV. The DoG responses for the block $b_z^{(c)}(i, j)$ can be computed by applying two Gaussians with different deviation parameters and computing their difference.

$$g_z(i, j) = \frac{1}{3} \sum_{c \in \{r, g, b\}} \left[G_{\sigma_1}(b_z^{(c)}(i, j)) - G_{\sigma_2}(b_z^{(c)}(i, j)) \right], \tag{9}$$

where $G_{\sigma_1}(\cdot)$ and $G_{\sigma_2}(\cdot)$ are Gaussians of standard deviations σ_1 and σ_2 , respectively, such that $\sigma_2 > \sigma_1$. To influence the traditional focus volume, we assign a scalar number to each block. Through empirical evaluations, we have determined that taking the variance of all elements within the DoG-applied block is an effective method for calculating this scalar value. The variance of the (i, j) th block is computed as:

$$s_z(i, j) = \frac{1}{m \times n} \sum_i^m \sum_j^n [g_z(i, j) - \mu]^2, \tag{10}$$

where $s_z(i, j)$ is the resultant for the z th image after computing variances for all blocks, and μ represents the mean of the block $g_z(i, j)$. As the variance of the (i, j) th block is a scalar value, the size of the resultant PFF $s_z(i, j)$ is decreased. To address this, we resize the PFF back to the size of the original image sequence. Finally, to ensure comparability and consistency, we normalize the PFF by rescaling it to a range between 0 and 1. Thus, the proposed PFF, which is also a 3D volume and denoted as $t_z(p)$, is calculated by up-sampling the $s_z(i, j)$ as:

$$t_z(p) = \text{resize}(s_z(i, j)), \tag{11}$$

where $\text{resize}(\cdot)$ represents a typical resizing operation in images, such as bilinear or bicubic interpolation; however, we have used the bicubic for resizing the $s_z(i, j)$ to its original dimension $X \times Y$. The procedure for computing perceptual focus volume $t_z(p)$ is summarized in Algorithm 1.

Algorithm 1 Computing perceptual focus factor for input image sequence

Input: Image Sequence: $I_z^{(c)}(p)$, Block Size : $m \times n$

Output: Perceptual Focus Volume : $t_z(p)$

for $z \leftarrow 1 : 1 : Z$ **do**

$[X, Y] \leftarrow \text{size}(I_z^{(c)}(p))$

 ▷ Image size

for $i \leftarrow 1 : m : X$ **do**

for $j \leftarrow 1 : n : Y$ **do**

$b_z^{(c)}(i, j) \leftarrow I_z^{(c)}(i : i + m, j : j + n)$

 ▷ Non-overlapping block

$g_z(i, j) \leftarrow \frac{1}{3} \sum_{c \in \{r, g, b\}} \left[G_{\sigma_1}(b_z^{(c)}(i, j)) - G_{\sigma_2}(b_z^{(c)}(i, j)) \right]$

 ▷ DoG using (9)

$s_z(i, j) \leftarrow \frac{1}{m \times n} \sum_i^m \sum_j^n [g_z(i, j) - \mu]^2$

 ▷ Variance of block using (10)

end for

end for

$t_z(p) \leftarrow \text{resize}(s_z(i, j), [X, Y])$

 ▷ up-sampling using (11)

end for

3.3. Depth Recovery

Next, we would combine two same-sized volumetric data sets $f_z(p)$ and $t_z(p)$ to obtain the improved focus volume $\hat{f}_z(p)$ as:

$$\hat{f}_z(p) = \exp(t_z(p)) \cdot f_z(p) \quad (12)$$

where $\exp(\cdot)$ is the Exponential function, which returns the power of input to the constant Euler's number, having a value of 2.718. Finally, the depth map is extracted from the improved focus volume using the "winner takes it all" formula. This formula gives depth for a pixel p by locating the image number that has the highest focus measure value for that pixel as:

$$d(p) = \arg \max_z (\hat{f}_z(p)) \quad (13)$$

4. Results and Discussions

4.1. Experimental Setup

To evaluate the performance of the proposed approach, experiments were conducted on both synthetic and real-world data sets. The synthetic data sets were sourced from the 4D light field benchmark [35], which includes ground truth depth maps for comparative analysis. For each synthetic data set, a collection of 30 images with varied focus settings was produced using the toolbox [36]. Additionally, since Ground Truth (GT) depth maps are available for the synthetic data sets, we conducted a quantitative comparison between the estimated depths and the GT depth maps. This was achieved by calculating the Root Mean Square Error (RMSE) and correlation (CORR) to gauge the similarity between the estimated depths and the GT depth maps. The RMSE is computed as follows:

$$RMSE = \sqrt{\frac{1}{X \times Y} \sum_p [D(p) - d(p)]^2}, \quad (14)$$

where $D(p)$ and $d(p)$ represent the GT and the estimated depth maps, respectively, and $X \times Y$ represents the total number of pixels in the map. The correlation is computed as,

$$CORR = \frac{\sum_p [D(p) - \bar{D}] [d(p) - \bar{d}]}{\sqrt{\sum_p [D(p) - \bar{D}]^2} \sqrt{\sum_p [d(p) - \bar{d}]^2}}, \quad (15)$$

where \bar{D} and \bar{d} represent the means of the GT and the estimated depth maps, respectively.

For real-world data sets, we sourced data from [37], which includes data sets of varying dimensions and image counts. Owing to these variations in dimensions and image quantities, we selected different block sizes for each real-world dataset, as detailed in Table 1. Additionally, to evaluate our method's performance under noisy conditions, we conducted a qualitative experiment using the *Buddha* dataset from [38]. This dataset comprises 29 images, each with a resolution of 768×768 pixels. The block size selected for this particular dataset was 64×64 . Lastly, the size of the Difference of Gaussians (DoG) kernel selected for all datasets and experiments was set at 4×4 .

Table 1. Description of the real dataset.

Dataset	Balls	Fruits	Keyboard	Window	Kitchen
Number of Images	25	30	32	27	11
Image Size	360×640	360×640	360×640	360×640	518×774
Block Size	20×64	20×64	20×64	20×64	74×86

4.2. Effect of Block Size

In experiments, we observed that the block size in computing PFF affects the accuracy of the resultant depth maps. The choice of block size should be chosen according to the characteristics of the data set, as various scenes necessitate different block sizes for optimal outcomes due to their dimensions and level of noise. For example, a distant landscape scene with diverse features would benefit from a larger block size, whereas a close-up image with detailed elements would require a smaller block size. In this study, we utilized the two synthetic *Pens* and *Medieval* data sets, both featuring subtle changes in the landscape. We conducted a qualitative analysis, as depicted in Figure 3, comparing the Ground Truth (GT) with depth maps generated using our method with block sizes of 8×8 , 16×16 , 32×32 , and 64×64 . The differences between these block sizes were highlighted, with red markings indicating improvements, and black markings indicating defects. It was observed that block sizes of both 32×32 and 64×64 resulted in smoother transitions of fadedness and more accurate depth maps. However, with the 64×64 block size, some blotches appeared due to their larger size, and additional irregular dots were noticed when a smaller block size of 8×8 was used. It can be concluded that the usage of a medium-sized block, such as 32×32 or 16×16 , as suggested in [16,30], for computing PFF improves the traditional focus measures for all data sets. The size of the block determines the degree of improvement. An inappropriate size will improve the traditional focus measure to a lesser degree. A small-sized block will produce the PFF closer to traditional FM and affects at a lesser degree whereas, for a larger block size, PFF will severely modify the traditional FM, so blokish effects may occur.

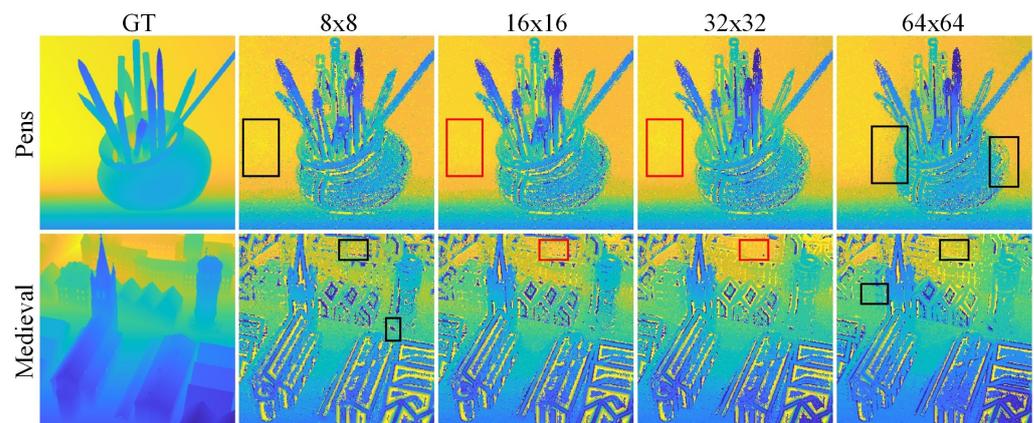


Figure 3. The depth maps of the synthetic datasets *Medieval* and *Town* after applying Perceptual Focus Factor (PFF) block sizes 8×8 , 16×16 , 32×32 , 64×64 respectively. The areas marked in black indicate areas where there was poor performance of the specific block size, while the areas marked in red represent improvements.

4.3. Comparative Analysis

The performance of the proposed method is evaluated using synthetic and real data sets. In the first experiment, fourteen synthetic data sets, including *Antinous*, *Boxes*, *Cotton*, *Dino*, *Dishes*, *Greek*, *Medieval*, *Museum*, *Pens*, *Pillows*, *Sideboard*, *Table*, *Town*, *Vinyl*, are used. Each image was 256×256 pixels in size, and there are 30 images in each data set. To compute the PFF, a medium block size 32×32 is used. First, the traditional focus volume is obtained through the DRDF, and then depth maps are computed from the enhanced focus volume through PFF. The performance of the proposed method has been thoroughly tested against five state-of-the-art methods: GLV [39], MCG [40], ML [24], FMSS [41], and RDF [34]. These five methods are applied to the fourteen synthetic data sets to obtain the focus volumes, and the depth maps are extracted by obtaining the image numbers with the best focus measures in the optical direction. As GT maps are available for synthetic data sets, RMSE and CORR measures are computed for the depth maps obtained through the

comparative methods and the proposed method. The quantitative comparison in terms of RMSE and CORR is then shown in Figure 4. In addition, All-in-Focus (AiF) images of the data sets are shown in the figure. From the figure, it can be observed that the proposed method consistently achieves lower RMSE values compared to each of the other methods across almost every data set, with a few exceptions, where it is nearly identical to the second-best, ML. A similar pattern can be seen with CORR values, where our method's CORR is the highest for almost every data set, except in a few instances where it closely matches that of the second-best.

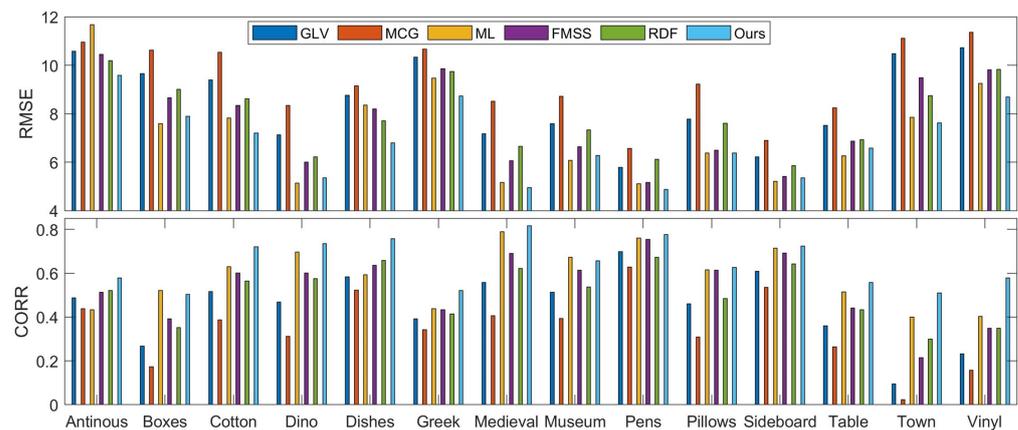


Figure 4. Root Mean Square Error (RMSE) and correlation (CORR) of synthetic datasets using Focus Measure (FM). Datasets are labeled on the x-axis of CORR graph.

For qualitative analysis, we present a comprehensive depth map analysis of synthetic data sets in Figure 5. In addition, All-in-Focus (AiF) images of the data sets are shown in the figure. In the *Antinous* data set, our method surpasses others in mitigating depth perception issues caused by shadows. It also excels in wide landscape images, as demonstrated in the *Medieval* data set, where our approach ensures a smooth gradient in the buildings' depth map. This results in a more intuitive depth map compared with other methods, which tend to focus on minor edges and isolate them, an approach not conducive to effective depth mapping. A similar effect is noted in the *Town* data set, where our method effectively reduces noise between buildings and maintains a consistent gradient in the depth map. In scenarios with extremely noisy backgrounds, like the *Pens* data set, our method proficiently suppresses speckle-type noise in the background, which is evident in other methods. Additionally, our method provides a clearer sense of object locations. In the *Vinyl* data set, for example, the stand and vinyl record are distinctly separated in terms of distance from the camera, unlike other methods that inaccurately portray them at similar distances.

In the second experiment, five real data sets, including *Balls*, *Fruits*, *Keyboard*, *Window*, *Kitchen*, are used. As detailed in Table 1, the number of images and the dimensions of the images vary among the data sets, so the block sizes for computing PFF are modified according to the image dimensions. In the proposed method, first, initial focus volume is obtained through the DRDF and then depth maps are computed from the enhanced focus volume through PFF. The performance of the proposed method is then thoroughly tested against five state-of-the-art methods: GLV [39], MCG [40], ML [24], FMSS [41], and RDF [34]. These five methods are applied to the five real data sets to obtain the focus volumes, and the depth maps are extracted by obtaining the image numbers having the best focus measures in the optical direction. The depth maps obtained through the comparative methods and the proposed method are shown in Figure 6. From the figure, it can be observed in the *Balls* data set that our method, alongside ML, offers superior depth perception, particularly evident in the detailed structure of the nearest ball. In the *Fruits* data set, our approach significantly improves depth perception by accurately representing darker fruits closer to the camera. Similarly, in the *Keyboard* data set, both our method and ML enhance depth perception, effectively highlighting the nearest keys with greater contrast. It is noteworthy,

however, that ML achieves the best results in this data set, providing an enhanced gradient in the depth map. In the *Plants* data set, MCG struggles due to noise, while ML and FMSS show relatively better performance. For the *Window* data set, although GLV, MCG, and FMSS focus on capturing the intricate structure of nearby objects, our method excels by creating a more effective gradual fade effect, particularly in the background, thereby enhancing overall depth perception.

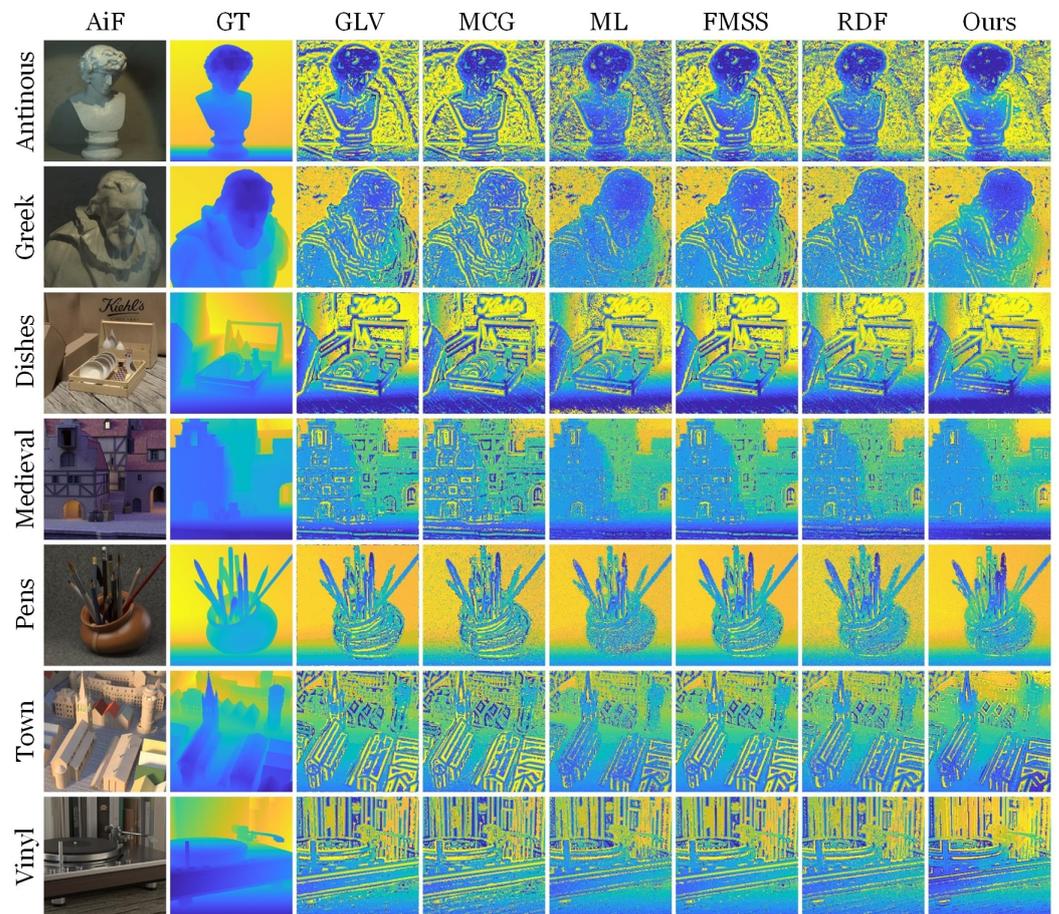


Figure 5. Depth map comparison of synthetic datasets by different methods. The first column represents an All-in-Focus (AiF) image and the second column represents the ground truth (GT) depth map.

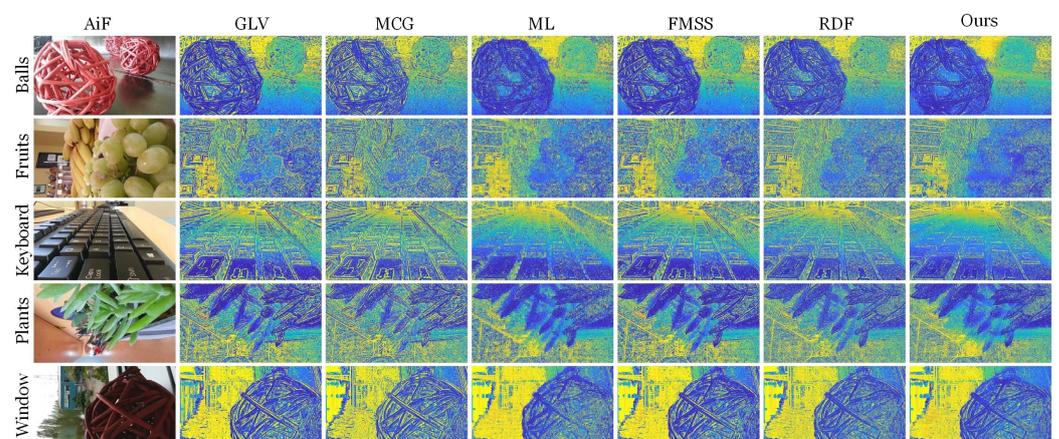


Figure 6. Depth map comparison of the real-world datasets *Balls*, *Fruits*, *Keyboard*, *Plants*, and *Window* by different methods. The first column represents an All-in-Focus (AiF) image.

The next experiment is conducted to evaluate the performance of the comparative methods and the proposed method in noisy conditions. For this, we selected the *Buddha* data set, as the images in this data set contain a significant amount of noise. Depth maps obtained from the comparative methods GLV [39], MCG [40], ML [24], FMSS [41], RDF [34], and the proposed method are shown in Figure 7. From the figure, it can be observed when the depth map was extracted; every method except ours found it challenging to counter noise. ML, in particular, performed the worst in estimating depth, and even RDF, which performed better in other comparative experiments, suffered heavily from the noise here. FMSS also struggled with noise. GLV and MCG performed better than the others, but they were not the best. Our method outperformed the rest, excelling in removing background noise and providing somewhat smooth depth values to the dice present in the scene. Hence, it can be concluded that the proposed method is effective in suppressing the noisy focus measures.

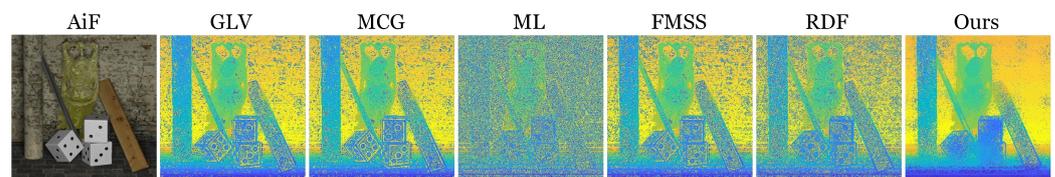


Figure 7. Depth map comparison of noisy data set *Buddha*.

4.4. Ablation Study

In this section, we analyzed the effect of the PFF. For this, we applied five traditional focus measure methods: GLV [39], MCG [40], ML [24], FMSS [41], and RDF [34] on the *Medieval* data. Then, the initial depth maps are extracted by obtaining the image numbers containing the best focus measures along the optical axis. The PFF volume for the input dataset *Medieval* is computed using a 32×32 block size, then traditional focus volumes are scaled using the PFF volume, and depth maps are extracted from the enhanced FV's. The resultant initial (without applying PFF) and improved (after applying PFF) depth maps for the *Medieval* dataset are shown in Figure 8. It is evident that PFF removes the irregularities in the depth map obtained from all the methods and provides a better depth perception, by giving a smooth fade to the depth maps. For instance, in all the methods, PFF tries to suppress irregular dots of the buildings present in the depth maps and tries to blend those dots into a more gradual fade. In other words, inaccurate depth values are replaced with closer-to-real depth values when the depth maps are extracted through the enhanced focus volumes.

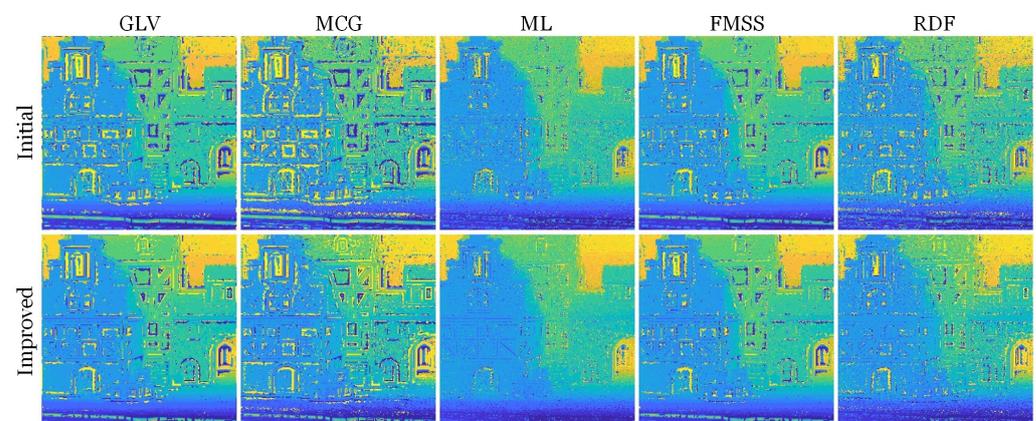


Figure 8. The depth maps of the synthetic dataset *Medieval* were obtained from different focus measures: (row-1) depth maps extracted without applying the Perceptual Focus Factor (PFF); (row-2) depth maps extracted after applying the proposed PFF.

In another experiment, the Focus Measure responses (FM responses) across all frames in an image sequence are analyzed with and without applying PFF, as shown in Figure 9. In this analysis, focus curves for specific pixels across all frames in an image sequence are studied. For this experiment, the traditional focus volume is obtained by applying the DRDF FM on the synthetic *Dino* dataset. Then, the PFF volume for the input dataset *Dino* is computed using a 32×32 block size, and an enhanced focus volume is computed. We randomly selected two pixels from the synthetic *Dino* dataset, specifically at coordinates (108,132) and (217,154), highlighted in red and green in the Figure 9a. The first pixel encompasses shadows, where most FMs typically yield inaccurate focus responses, while the second pixel involves edges. Initial focus curves obtained through traditional FM DRDF and the enhanced focus curves through PFF for the pixels (108,132) and (217,154) are shown in Figure 9b,c. For an FM response, unimodality and narrowness of the curve are preferred, as they indicate the FM's confidence in identifying the image frame with maximum focus. As observed from the figure, for the first pixel, the Perceptual Focus Factor (PFF) based on an enhanced focus measure provided a narrower indication of the image frame with the highest focus and aligned more closely with the GT value as compared to the traditional FM. For the second pixel, the enhanced focus measure through PFF was not only closer to the GT but also exhibited unimodality, unlike the traditional FM. It shows the effectiveness and contribution of the PFF toward the improvement of the traditional focus measures.

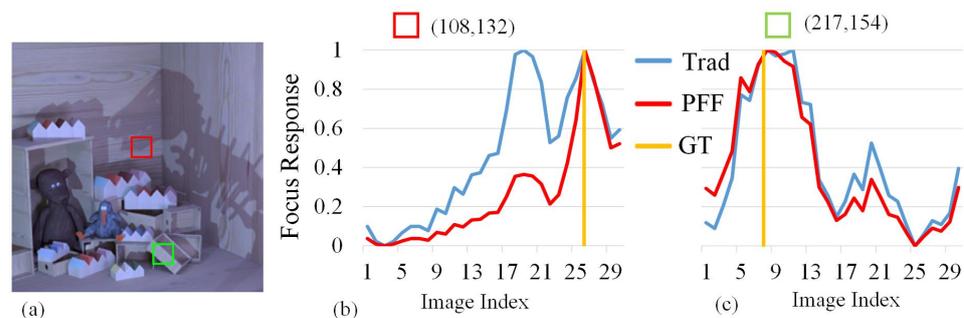


Figure 9. Analysis of the Focus Measure (FM) response using our method on synthetic data set *Dino*. (a) Locations marked by red and green boxes are our target areas that are pixels locations (108,132) and (217,154). (b) Focus curves for pixel (108,132). (c) Focus curves for pixel (217,154). The blue curve represents the FM response from the traditional FM, the red curve from the enhanced FM through Perceptual Focus Factor (PFF), and the yellow line indicates the Ground Truth (GT) value for these respective pixels.

Finally, to highlight the difference between the depth maps obtained with and without applying PFF, a quantitative analysis is provided. For this, first, we applied five traditional focus measures methods: GLV [39], MCG [40], ML [24], FMSS [41], and RDF [34] on all 14 datasets to get the traditional FVs and depth maps extracted from them. The PFF volume for each dataset is computed using block size 32×32 , and then depth maps are extracted from the enhanced FVs through PFF. RMSE and CORR measures were computed for the depth maps obtained with and without enhancement through PFF and recorded in Table 2, where it is applied to five other methods: GLV [39], MCG [40], ML [24], FMSS [41], and RDF [34]. The reference table shows the RMSE and CORR values for synthetic data sets, comparing their performance with and without the application of our PFF, with values separated by a backward slash. From the table, it is clear that the PFF enhances all methods across most data sets, underscoring its ability to improve existing techniques. For example, in the *Boxes* data set, the CORR value for GLV improved by 29.6%, MCG by 76.5%, and ML by 19.2%. In the *Town* data set, the RMSE for RDF changed by 10.7% and for ML by 13.89%. In some cases, as seen with the *Table* data set, the PFF does not degrade the final depth map, instead maintaining performance at a level comparable to the original. In these instances, performance can be further enhanced by adjusting the block size to suit the specific characteristics of the data set.

Table 2. Table of the Root Mean Square Error (RMSE) and CORR values before and after applying our proposed Perceptual Focus Factor (PFF) to different Focus Measure (FM) techniques. The initial (Ini) values before applying the PFF are listed first, followed by a backward slash '\', and then the improved (Imp) values achieved after applying the PFF.

	RMSE					CORR				
	GLV	MCG	ML	FMSS	RDF	GLV	MCG	ML	FMSS	RDF
	Ini\Imp									
Antinous	10.6\10.4	11.0\10.7	11.7\10.5	10.5\10.2	10.2\10.0	0.49\0.50	0.44\0.46	0.43\0.51	0.51\0.53	0.52\0.53
Boxes	9.66\9.09	10.6\9.64	7.59\6.97	8.65\8.50	9.00\8.20	0.27\0.35	0.17\0.30	0.52\0.62	0.39\0.42	0.35\0.47
Cotton	9.39\8.78	10.5\9.61	7.83\7.00	8.34\8.14	8.61\8.06	0.52\0.60	0.39\0.52	0.63\0.73	0.60\0.63	0.56\0.64
Dino	7.13\6.66	8.34\7.63	5.14\4.65	6.01\5.91	6.22\5.71	0.47\0.59	0.31\0.53	0.70\0.79	0.60\0.64	0.58\0.71
Dishes	8.76\8.28	9.14\8.34	8.36\7.22	8.19\8.05	7.71\7.03	0.58\0.63	0.52\0.61	0.59\0.71	0.64\0.65	0.66\0.73
Greek	10.3\9.81	10.7\9.73	9.46\8.81	9.86\9.64	9.74\9.04	0.39\0.43	0.34\0.42	0.44\0.50	0.43\0.45	0.41\0.48
Medieval	7.18\6.27	8.52\7.05	5.17\4.36	6.06\5.76	6.66\5.53	0.56\0.68	0.41\0.62	0.79\0.87	0.69\0.72	0.62\0.76
Museum	7.60\7.25	8.72\8.19	6.08\5.69	6.63\6.53	7.34\6.93	0.51\0.56	0.39\0.46	0.67\0.72	0.61\0.63	0.54\0.59
Pens	5.79\5.39	6.57\5.93	5.11\4.84	5.17\5.07	6.13\5.37	0.70\0.73	0.63\0.68	0.76\0.77	0.75\0.76	0.67\0.74
Pillows	7.78\6.95	9.22\8.04	6.38\5.90	6.49\6.26	7.61\7.01	0.46\0.58	0.31\0.48	0.61\0.68	0.61\0.65	0.48\0.56
Sideboard	6.22\5.96	6.90\6.49	5.21\5.02	5.41\5.36	5.85\5.67	0.61\0.65	0.54\0.62	0.71\0.76	0.69\0.71	0.64\0.69
Table	7.52\7.49	8.24\8.10	6.27\6.25	6.87\6.92	6.93\6.93	0.36\0.42	0.26\0.38	0.52\0.59	0.44\0.46	0.43\0.51
Town	10.5\9.79	11.1\9.95	7.85\6.76	9.48\9.29	8.75\7.81	0.09\0.27	0.02\0.29	0.40\0.60	0.21\0.28	0.30\0.49
Vinyl	10.7\9.97	11.4\10.2	9.26\8.40	9.82\9.58	9.84\8.84	0.23\0.42	0.16\0.45	0.40\0.59	0.35\0.41	0.35\0.57

4.5. Computational Complexity

In this section, we analyze the computational cost comparison among the comparative methods and the proposed method. The computational cost of the proposed method consists of the cost of computing PFF and the cost of computing traditional focus measures. The computational time depends on the dimensions and the number of images in the input sequence. We implemented the focus measures in MATLAB and ran them on the system having quad core 3.30 GHz CPU with 16 GB memory. The time taken by different focus measures and the PFF for synthetic and real data sets is recorded in Table 3. GLV, ML, MCG, and RDF focus measures are efficient compared to DRDF, FMSS, and PFF. The computational time mainly depends on the number of images, and the dimensions of each image. In the case of PFF, the block size also affects the computational time. In the case of real data sets, PFF has a comparable computational cost with other focus measures. In addition, PFF is more efficient than FMSS, DRDF, and GLV focus measures.

Table 3. Time computed in seconds for different focus measures and the proposed Perceptual Focus Factor (PFF) for different synthetic and real data sets.

Dataset	GLV	MCG	ML	FMSS	RDF	DRDF	PFF
Antinous	0.1100	0.0554	0.0513	0.4031	0.0425	0.1313	0.1756
Boxes	0.1062	0.0549	0.0651	0.4124	0.0361	0.1243	0.1288
Cotton	0.1044	0.0540	0.0567	0.4253	0.0355	0.1306	0.1227
Dino	0.1020	0.0561	0.0624	0.4189	0.0374	0.1308	0.1224
Dishes	0.1146	0.0676	0.0722	0.4111	0.0350	0.1181	0.1251
Greek	0.1039	0.0529	0.0531	0.4084	0.0339	0.1227	0.1306
Medieval	0.1018	0.0523	0.0538	0.4039	0.0338	0.1136	0.1271
Museum	0.1018	0.0523	0.0538	0.4039	0.0338	0.1136	0.1226
Pens	0.1015	0.0512	0.0534	0.4074	0.0355	0.1217	0.1238
Pillows	0.1035	0.0494	0.0540	0.4117	0.0333	0.1204	0.1222
Sideboard	0.1034	0.0509	0.0485	0.3945	0.0351	0.1255	0.1230
Table	0.1058	0.0553	0.0612	0.4054	0.0356	0.1215	0.1222
Town	0.1084	0.0556	0.0572	0.4153	0.0358	0.1236	0.1240
Vinyl	0.1015	0.0510	0.0491	0.3839	0.0342	0.1174	0.1240
Balls	0.7278	0.2928	0.1526	1.3000	0.1629	0.3128	0.4116
Fruits	0.8118	0.3557	0.1705	1.4771	0.1379	0.3478	0.4185

Table 3. Cont.

Dataset	GLV	MCG	ML	FMSS	RDF	DRDF	PPF
Keyboard	0.8600	0.3699	0.1730	1.5042	0.1484	0.3809	0.4679
Window	0.7342	0.3164	0.1487	1.2643	0.1309	0.3078	0.3799
Kitchen	0.5197	0.1939	0.1068	0.9294	0.1010	0.2253	0.0775
Buddha	2.1598	0.7956	0.4134	3.9632	0.3722	0.8161	0.3914

5. Conclusions

This paper introduces a perceptually driven method for depth extraction through Shape From Focus (SFF). The traditional Focus Volume (FV) obtained from applying traditional Focus Measures (FMs) on the image sequence is enhanced through the perceptual focus factor. Our approach leverages the Difference of Gaussians (DoG) operator, which reflects the human visual system's method of edge perception in scenes. By applying DoG to local areas within non-overlapping blocks of the image sequence, we create the Perceptual Focus Factor (PPF) for each pixel in the input image sequences. This new metric is then integrated with the traditional FV, leading to the extraction of a significantly improved depth map. Our method has been rigorously tested on both synthetic and real-world datasets, demonstrating a marked improvement in depth extraction accuracy over existing state-of-the-art SFF methods. Despite obtaining improved depth maps through the proposed method, there are two important issues and concerns. First, as the proposed method computes the PPF, it takes more time to compute depth maps compared to computing just focus measures. The computational time depends on the dimensions and the number of images in the input sequence. Secondly, the size of the block determines the degree of improvement. An inappropriate size will improve the traditional focus measure to a lesser degree. Determining an adaptive local block size or an optimal for the whole image will definitely improve the results. It requires a separate study, where a machine/deep learning-based model may be developed to determine the optimal or adaptive block size.

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

Funding: This work was supported by the Education and Research Promotion Program of KoreaTech (2023), by the BK-21 plus FOUR program through the National Research Foundation of Korea (NRF) under the Ministry of Education, and by the Basic research program through the NRF grant funded by the Korean government (MSIT: Ministry of Science and ICT) (2022R1F1A1071452).

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: <https://lightfield-analysis.uni-konstanz.de/> (accessed on 26 December 2023), and <https://github.com/MaximilianStaab/mDFF> (accessed on 26 December 2023).

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

Abbreviations

The following abbreviations are used in this manuscript:

DFF	Depth From Focus
DoG	Difference of Gaussians
DRDF	Directional Ring Difference Filter
GT	Ground Truth
FM	Focus Measure
FV	Focus Volume
LoG	Laplacian of Gaussian
PPF	Perceptual Focus Factor
SFF	Shape From Focus

References

1. Valentin, J.; Kowdle, A.; Barron, J.T.; Wadhwa, N.; Dzitsiuk, M.; Schoenberg, M.; Verma, V.; Csaszar, A.; Turner, E.; Dryanovski, I.; et al. Depth from motion for smartphone AR. *ACM Trans. Graph. ToG* **2018**, *37*, 1–19. [\[CrossRef\]](#)
2. Elder, J.H. Shape from contour: Computation and representation. *Annu. Rev. Vis. Sci.* **2018**, *4*, 423–450. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Balzer, J.; Werling, S. Principles of shape from specular reflection. *Measurement* **2010**, *43*, 1305–1317. [\[CrossRef\]](#)
4. Cui, Y.; Schuon, S.; Chan, D.; Thrun, S.; Theobalt, C. 3D shape scanning with a time-of-flight camera. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 1173–1180.
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. 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. [\[CrossRef\]](#)
7. 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. [\[CrossRef\]](#)
8. Rajevenceltha, J.; Gaidhane, V.H. A novel approach for image focus measure. *Signal Image Video Process.* **2021**, *15*, 547–555. [\[CrossRef\]](#)
9. Liu, S.; Liu, M.; Yang, Z. An image auto-focusing algorithm for industrial image measurement. *EURASIP J. Adv. Signal Process.* **2016**, *2016*, 70. [\[CrossRef\]](#)
10. 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.
11. Yang, G.; Nelson, B.J. Wavelet-based autofocusing and unsupervised segmentation of microscopic images. In Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453), Las Vegas, NV, USA, 27–31 October 2003; Volume 3, pp. 2143–2148.
12. Mahmood, M.T.; Choi, T.S. Focus measure based on the energy of high-frequency components in the S transform. *Opt. Lett.* **2010**, *35*, 1272–1274. [\[CrossRef\]](#)
13. Kristan, M.; Perš, J.; Perše, M.; Kovačič, S. A Bayes-spectral-entropy-based measure of camera focus using a discrete cosine transform. *Pattern Recognit. Lett.* **2006**, *27*, 1431–1439. [\[CrossRef\]](#)
14. De, I.; Chanda, B. Multi-focus image fusion using a morphology-based focus measure in a quad-tree structure. *Inf. Fusion* **2013**, *14*, 136–146. [\[CrossRef\]](#)
15. Ashfaq, K.; Mahmood, M.T. Directional Ring Difference Filter for Robust Shape-from-Focus. *Mathematics* **2023**, *11*, 3056. [\[CrossRef\]](#)
16. Guo, L.; Liu, L. A Perceptual-Based Robust Measure of Image Focus. *IEEE Signal Process. Lett.* **2022**, *29*, 2717–2721. [\[CrossRef\]](#)
17. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Thelen, A.; Frey, S.; Hirsch, S.; Hering, P. Improvements in shape-from-focus for holographic reconstructions with regard to focus operators, neighborhood-size, and height value interpolation. *IEEE Trans. Image Process.* **2009**, *18*, 151–157. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Moeller, M.; Benning, M.; Schönlieb, C.; Cremers, D. Variational depth from focus reconstruction. *IEEE Trans. Image Process.* **2015**, *24*, 5369–5378. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Ali, U.; Lee, I.H.; Mahmood, M.T. Incorporating structural prior for depth regularization in shape from focus. *Comput. Vis. Image Underst.* **2023**, *227*, 103619. [\[CrossRef\]](#)
21. Hazirbas, C.; Soyer, S.G.; Staab, M.C.; Leal-Taixé, L.; Cremers, D. Deep depth from focus. In Proceedings of the Computer Vision—ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, 2–6 December 2018; Springer: Cham, Switzerland, 2018; pp. 525–541.
22. Wang, N.H.; Wang, R.; Liu, Y.L.; Huang, Y.H.; Chang, Y.L.; Chen, C.P.; Jou, K. Bridging unsupervised and supervised depth from focus via all-in-focus supervision. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 12621–12631.
23. Yang, F.; Huang, X.; Zhou, Z. Deep depth from focus with differential focus volume. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 12642–12651.
24. Nayar, S.K.; Nakagawa, Y. Shape from focus. *IEEE Trans. Pattern Anal. Mach. Intell.* **1994**, *16*, 824–831. [\[CrossRef\]](#)
25. Li, Y.; Li, Z.; Zheng, C.; Wu, S. Adaptive weighted guided image filtering for depth enhancement in shape-from-focus. *Pattern Recognit.* **2022**, *131*, 108900. [\[CrossRef\]](#)
26. Krotkov, E.; Martin, J.P. Range from focus. In Proceedings of the Robotics and Automation, San Francisco, CA, USA, 7–10 April 1986; Volume 3, pp. 1093–1098.
27. Helmlí, F.S.; Scherer, S. Adaptive shape from focus with an error estimation in light microscopy. In Proceedings of the ISPA 2001, Proceedings of the 2nd International Symposium on Image and Signal Processing and Analysis. In Conjunction with 23rd International Conference on Information Technology Interfaces, Pula, Croatia, 19–21 June 2006; pp. 188–193.
28. Minhas, R.; Mohammed, A.A.; Wu, Q.J.; Sid-Ahmed, M.A. 3D shape from focus and depth map computation using steerable filters. In Proceedings of the Image Analysis and Recognition: 6th International Conference, ICIAR 2009, Halifax, NS, Canada, 6–8 July 2009; Springer: Cham, Switzerland, 2009; pp. 573–583.

29. Santos, A.; Ortiz De Solórzano, C.; Vaquero, J.J.; Pena, J.M.; Malpica, N.; del Pozo, F. Evaluation of autofocus functions in molecular cytogenetic analysis. *J. Microsc.* **1997**, *188*, 264–272. [[CrossRef](#)]
30. 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](#)]
31. Marr, D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*; MIT Press: Cambridge, MA, USA, 2010.
32. Marthon, P.; Thiesse, B.; Bruel, A. Edge detection by differences of Gaussians. In Proceedings of the Computer Vision for Robots, Cannes, France, 25–27 November 1985; SPIE: Bellingham, WA, USA, 1986; Volume 595, pp. 318–327.
33. Spring, K.R.; Russ, J.C.; Parry-Hill, M.J.; Fellers, T.J.; Davidson, M.W. Difference of Gaussians Edge Enhancement Interactive Tutorials. 2016. Available online: <https://micro.magnet.fsu.edu/primer/java/digitalimaging/processing/diffgaussians/> (accessed on 1 December 2023).
34. 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](#)] [[PubMed](#)]
35. Honauer, K.; Johannsen, O.; Kondermann, D.; Goldluecke, B. A dataset and evaluation methodology for depth estimation on 4D light fields. In Proceedings of the Computer Vision—ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, 20–24 November 2016; Springer: Cham, Switzerland, 2016; pp. 19–34.
36. Light Field Toolbox for MATLAB. 2020. Available online: <https://dgd.vision/Tools/LFToolbox/> (accessed on 1 December 2023).
37. Suwajanakorn, S.; Hernandez, C.; Seitz, S.M. Depth from focus with your mobile phone. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3497–3506.
38. Wanner, S.; Meister, S.; Goldluecke, B. Datasets and benchmarks for densely sampled 4D light fields. In Proceedings of the International Workshop on Vision, Modeling and Visualization, Lugano, Switzerland, 11–13 September 2023; Volume 13, pp. 225–226.
39. Ali, U.; Mahmood, M.T. Robust focus volume regularization in shape from focus. *IEEE Trans. Image Process.* **2021**, *30*, 7215–7227. [[CrossRef](#)] [[PubMed](#)]
40. 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](#)]
41. 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](#)]

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.