

Received July 1, 2021, accepted July 17, 2021, date of publication July 20, 2021, date of current version July 28, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3098753

# Focus Measurement in Color Space for Shape From Focus Systems

HUSNA MUTAHIRA<sup>1</sup>, (Student Member, IEEE), BILAL AHMAD<sup>2</sup>,  
MANNAN SAEED MUHAMMAD<sup>1</sup>, (Senior Member, IEEE),  
AND DONG RYEOL SHIN<sup>3</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, College of Information and Communication Engineering, Sungkyunkwan University, Natural Science Campus, Suwon 16419, South Korea

<sup>2</sup>Department of Computer Science, Norwegian University of Science and Technology, 2815 Gjøvik, Norway

<sup>3</sup>Department of Computer Science and Engineering, College of Computing and Informatics, Sungkyunkwan University, Natural Sciences Campus, Suwon 16419, South Korea

Corresponding authors: Dong Ryeol Shin (drshin@skku.edu) and Mannan Saeed Muhammad (mannan@skku.edu)

**ABSTRACT** Shape from Focus (SFF) has been studied extensively in computer vision for 3D shape and depth recovery. The first stage in SFF methods is to compute the focus value of every pixel by converting the colored images into gray scale and then apply the focus measure operator. Converting colored values in the images into gray scale values may lead to imprecise mapping of pixels with different colored values onto the same gray scale value, this affects the overall accuracy of the system. In a colored image, the focused pixels maintain a considerable color difference from their neighboring pixels as compared to the defocused ones, which are blended into their neighborhood. This article presents an alternative method to measure the degree of focus by directly processing colored images. The color differences of the neighbor pixels with respect to the central pixel are obtained and summed together, this is followed by calculating their spread. The sum and the spread are combined to measure the degree of focus of the pixel in consideration. The proposed focus measure is then used for shape recovery of various simulated and real objects and is compared with previous techniques. The comparison results show the proposed method has the highest correlation and smallest RMSE values confirming the effectiveness of using color images for shape recovery.

**INDEX TERMS** Color focus measure, focus measure operators, 3D shape recovery, shape from focus.

## I. INTRODUCTION

In computer vision, shape-from-X is a passive monocular technique to recover the 3D geometry of a scene or an object from a set of images. Here, X is a 2D characteristic, for example, shading, motion, stereo, focus, defocus, used as a cue to infer the 3D shape. This approach has various industrial applications in diagnostics, autonomous vehicle guidance, microscopy etc., [1]–[3]. Shape-from-Focus (SFF), deals with shape recovery using various images of the same scene that are captured after manipulating the focus settings of the imaging device, [4]. SFF measures the amount of focus in each pixel, along the optical axis, to identify the best-focused pixels, which are used to recover the depth of the scene, [5]–[7].

In SFF all the images are taken by either moving the object along the optical axis in small steps, each step of size  $\Delta_{step}$ ; or changing the focus of imaging device in small steps.

The associate editor coordinating the review of this manuscript and approving it for publication was Wenming Cao<sup>1</sup>.

An image (of dimensions  $l \times m$ ) is stored at each step, and  $n$  is the total number of images in the image stack, as shown in Fig. 1, and is given by:

$$n = \frac{U_{disp.}}{\Delta_{step}}, \quad (1)$$

where,  $U_{disp.}$  is the total displacement of the object plane.  $U_{disp.}$  is chosen in such a way that the magnification of the object remains constant, [8].

Lens law can be used to explain the focusing of every point in the image sequence. If the object distance from the lens shifts the image plane from the sensor plane by a distance  $\Delta D$ , it creates a circle of confusion (CoC), [9], which leads to a blurred image as shown in Fig. 2. However, according to geometric optics, if the image plane lies exactly on the sensor plane, no CoC is formed and a sharp image will be obtained. The object distance  $D_O$ , from the lens is given by:

$$D_O = \frac{D_I \times f}{D_I - f}, \quad (2)$$

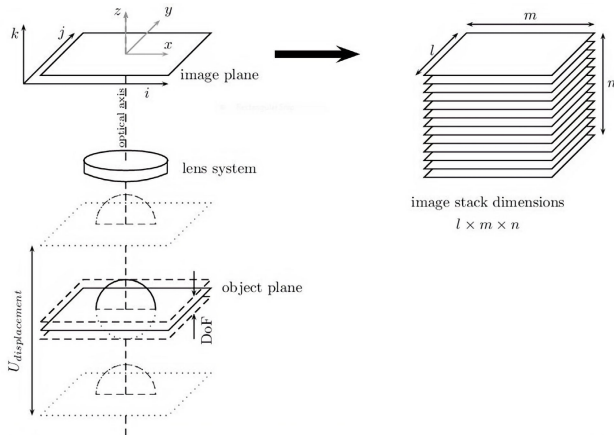


FIGURE 1. Image stack acquisition in Shape from focus.

where,  $f$  is the focal length of the lens and  $D_I$  is the image distance from the lens.

After image acquisition, a focus measure (FM) operator is applied to find the best-focused points in the image stack. The FM operator acts as a High-Pass-Filter which suppresses the out of focus points and enhances the focused points in the image stack. Ideally, the FM yields its maximum value at the best focused position. This value decreases as the object moves away from this point in both directions (towards and away from the lens). Images are taken in such a way that initially, the object is completely defocused, then images gradually become focused on every point before becoming defocused again.

Finding the best-focused pixels in SFF is a crucial step in estimating depth information. Traditionally, an FM is applied to images after converting them to grayscale. Grayscale images can be obtained using the following:

$$\Gamma(i, j) = 0.213R(i, j) + 0.715G(i, j) + 0.072B(i, j), \quad (3)$$

where,  $\Gamma(i, j)$  is the grayscale value of the  $(i, j)$  pixel in the image and  $R(i, j)$ ,  $G(i, j)$ ,  $B(i, j)$  are the RGB values of that pixel. Several other conversions have also been outlined in, [10].

However, converting a color image to grayscale leads to a loss of information. Some distinct color values map to the same gray values incorrectly during this conversion, [11]. Hence, when the FM is applied to these grayscale images, the focus values are not calculated correctly, resulting in incorrect depth extraction.

In this paper, a new focus measure,  $FM_c$ , is proposed, that uses color information in the images.  $FM_c$  receives colored images as input from which it computes the focus values, eliminating the first step of converting colored images to grayscale, as proposed by the conventional methods.  $FM_c$  is applied to different image stacks, of various objects, to recover the 3D shape of the objects. The proposed FM has shown better results when compared with previous methods.

This article is structured as follows. The next section provides a brief overview of the related work. In Section III,

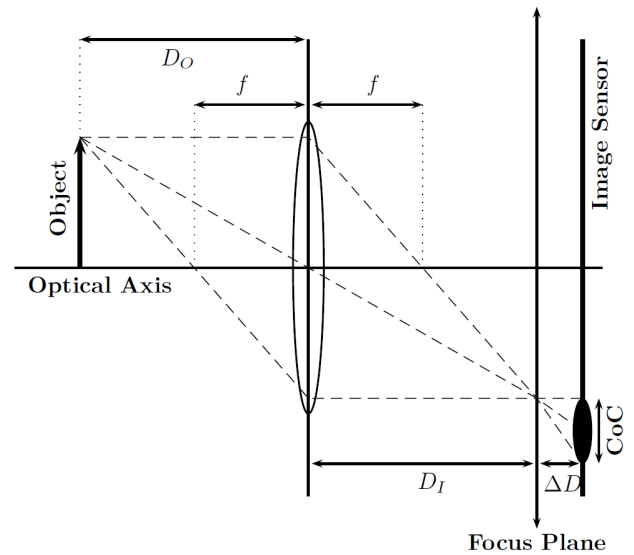


FIGURE 2. Focusing via Gaussian lens law.

the importance of color images is discussed, while the proposed method is described in Section IV. Results are presented and discussed in Section V, which is followed by Section VI that concludes the article.

## II. RELATED WORK

An FM operator acts as a High-Pass-Filter which enhances the focused regions and suppresses the defocused regions. It computes the sharpness of a selected center pixel by choosing a small local window. Since each object point has different focus values with various optical settings, these values are compared with neighboring object points to identify the position of the pixel where it is best focused, i.e., with the highest focus value.

A variety of FM operators have been proposed in the previous literature. An overview is provided in, [12]. FMs are divided into six significant groups depending upon their operating principles. Gradient-based operators take the first derivative of an image to compute the focus value, [13]–[16]. Laplacian-based operators measure the focus by calculating the second derivative of an image, [15]–[19]. Wavelet-based operators consider the wavelet transform to describe an image’s spatial and frequency contents, [15], [20], [21], whereas, statistical-based operators take into account different statistics of an image to measure the focus, [16], [19], [22], [23]. Discrete cosine transform (DCTE) based operators consider DCT coefficients from the frequency content of the image to calculate the focus level, [24], [25], and finally, many other operators, which do not lie in any of the aforementioned categories, are grouped into a category named as miscellaneous operators, [26], [27].

Laplacian operators can be applied to detect the regions of high contrast. They have been used as a focus measure in both Auto-focus, [28]–[30], and SFF, [15], [16], [18], [19], [31]. They are used to compute the second derivative for each pixel

of the given image window. However, the second derivative in the  $i$  and  $j$  directions can yield equivalent values with opposite signs, which cancel each other. Therefore, they proposed the Sum of Modified Laplacian (SML) which is the sum of absolute Laplacian in the  $i$  and  $j$  directions, and is given by:

$$FM_{SML(i_o, j_o)} = \sum_{p(i, j) \in N(i_o, j_o)} \left| \frac{\partial^2 \Gamma(i, j)}{\partial i^2} \right| + \left| \frac{\partial^2 \Gamma(i, j)}{\partial j^2} \right|, \quad (4)$$

where  $\Gamma(i, j)$  is the gray value of the pixel  $(p(i, j))$  at coordinates  $(i, j)$ , and  $N(i_o, j_o)$  is local neighborhood around  $p(i, j)$  in the window of  $\omega \times \omega$ . Modified versions of Laplacian operators are further studied in, [16], [31], [32].

Tenenbaum (TENGE) operator is discussed in, [13], [16], [33], [34]. This FM is based on the Sobel gradient operator, [31], [35]. It obtains its value by summing the squared responses of the horizontal and vertical Sobel masks. To increase this FM's robustness further, the results are summed up in the local pixel window, which is defined as:

$$FM_{TENGE(i_o, j_o)} = \sum_{p(i, j) \in N(i_o, j_o)} [\Gamma_i(i, j)^2 + \Gamma_j(i, j)^2]. \quad (5)$$

Gray Level Variance (GLVA) is one of the most popular FM operators, [16], [31], [36], [37]. It follows the assumption that regions with high gray level values are sharper than regions with low gray level values. GLV is defined as:

$$FM_{GLVA(i_o, j_o)} = \sum_{p(i, j) \in N(i_o, j_o)} [\Gamma(i, j) - \mu]^2, \quad (6)$$

where  $\mu$  is the mean gray value within a neighborhood  $N(i_o, j_o)$  around  $p(i, j)$ . This method is also discussed in, [22], [38]. The Sum of Wavelet Coefficients (WAVS) approach decomposes the image into four sub-images of wavelet transform coefficients of the image. In the first level, the image is transformed into three detail sub-bands ( $W_{LH}$ ,  $W_{HL}$ ,  $W_{HH}$ ) and one approximation sub-band ( $W_{LL}$ ). In the higher levels, the approximation sub-band is further divided into complete and approximation sub-bands, [39]. The information from both sub-bands is then used to compute the degree of focus.

$$FM_{WAVS(i_o, j_o)} = \sum_{p(i, j) \in N(i_o, j_o)} \left( |W_{LH}^{(i, j)}| + |W_{HL}^{(i, j)}| + |W_{HH}^{(i, j)}| \right). \quad (7)$$

All of these FMs can be applied on a stack of image to measure the focus values, with the aim of constructing a depth map (initial shape) of the object. Although, all these FMs have different working principles, essentially, they all take gray scale images as input and then compute the focus values of the image. The effect of window size on FM performance is discussed in, [40], and suggested that the window sizes of  $3 \times 3$  and  $5 \times 5$  give better results. The mathematical calculation for window size is provided by, [41].

These FMs can be applied to color images by processing the color channels separately, [42]. While this can be done efficiently for single or few images, processing each channel

for higher number of images is computationally inefficient and costly. This computational inefficiency hinders the application of SFF for real time applications.

### III. MOTIVATION

Aberration is a common optical problem when it comes to using lenses. It occurs when a lens fails to bring colors of different focal lengths to the same focal plane, as shown in Fig. 3. An achromatic doublet is used to reduce the amount of aberration but, it does not produce an exact correction. An accurate image distance ( $D_I$ ) can only be recovered by processing each channel (R, G, B) individually. However, since each channel has a different wavelength, they affect  $D_I$  differently. This can be shown by combining the Lens Maker equation, [43], [44], and Cauchy equation, [45], given by:

$$\frac{1}{f} = \frac{1}{D_O} + \frac{1}{D_I} = (\phi + \frac{\mu}{\lambda^2} - 1) \left( \frac{1}{r_1} - \frac{1}{r_2} \right), \quad (8)$$

where  $\phi$  and  $\mu$  are the coefficients which are combined to make the refractive index of the material of the lens, [45]. Also,  $r_1$  and  $r_2$  are the radii of the curvatures of the lens, closer to the light source, and farther from the light source, respectively. For a typical lens, (8) can be written as:

$$\frac{1}{D_I} = \frac{\psi}{\lambda^2} + \varepsilon, \quad (9)$$

where  $\psi$  and  $\varepsilon$  are given as follows:

$$\psi = \mu \left( \frac{1}{r_1} - \frac{1}{r_2} \right), \quad (10)$$

$$\varepsilon = (\phi - 1) \left( \frac{1}{r_1} - \frac{1}{r_2} \right) - \frac{1}{D_O}. \quad (11)$$

From (9), if  $\psi$  and  $\varepsilon$  are taken as constants, then the relationship between wavelength of light ( $\lambda$ ) and the image distance ( $D_I$ ) becomes directly proportional and can be written as:

$$D_I \propto \lambda^2. \quad (12)$$

Fig. 4, shows the effect of  $\lambda$  on  $D_I$ , for two different values of object distance ( $D_O$ ) when using a fused silica convex lens (LE4412). As  $\lambda$  is increased from 450 nm (Blue spectrum) to 700 nm (Red spectrum), the approximate change in  $\Delta D_I$  is almost 2.5 mm.

Equation (12) shows that the depth varies by changing the wavelength, even though all the other parameters (including  $D_O$ ) in (9) are kept constant. Therefore, accurate depth can only be recovered if colored images are considered instead of grayscale images.

After image acquisition in SFF, an FM is applied to measure the focus value. All the FM operators, in previous literature described above, use grayscale images to measure the focus value. The accuracy of an FM can be improved if RGB images are employed since they contain more information than grayscale images, [46]. There are some RGB color combinations that give the same grayscale value because isoluminant changes are not preserved when converting to grayscale

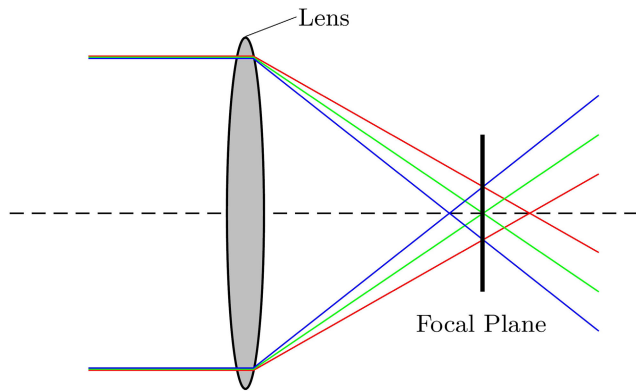


FIGURE 3. Effect of chromatic aberration on colors of different focal lengths. (This image is best viewed in color).

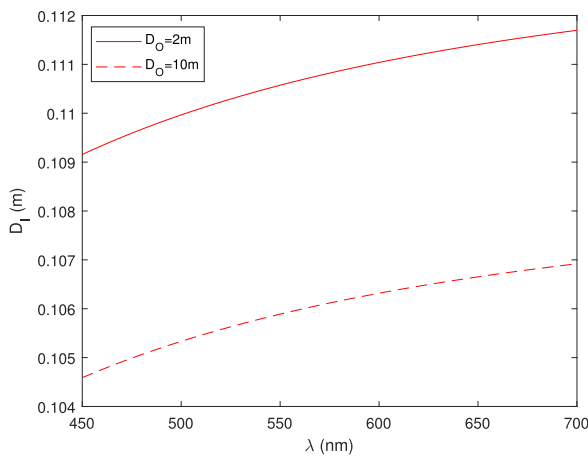


FIGURE 4. Effect on Image distance from lens by varying the wavelength.

and therefore, distinct pixels cannot be distinguished from one another with this approach, [46], as shown in Fig. 5.

In colored images, visual cues are preserved which help detect salient features, and hence are of considerable importance in various applications, [47]. Grayscale conversion of colored images suffers from information loss such as from the lightness channel in CIE L\*a\*b. Although many decolorization methods have been proposed to preserve this information, all of these methods have some limitations, [48]. In addition, grayscale mappings of color images that are constructed solely by approximating spectral uniformity, are inadequate because isoluminant visual cues signaled only by chromatic differences are lost, [49].

#### IV. PROPOSED FOCUS MEASURE

In Shape from Focus (SFF), images are acquired by moving the object in finite steps  $\Delta_{steps}$ , towards or away from the imaging device. Images can also be acquired by changing the focus settings of the imaging device, provided magnification remains constant. In every step the acquired image is stored in the image stack, [8]. Each image, in the image stack, is represented by  $I_k$ , where,  $1 \leq k \leq n$ , and  $n$  is the total number of images given in (1). The dimension of each image

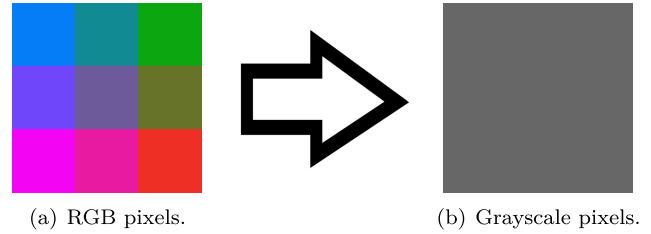


FIGURE 5. Conversion of RGB to grayscale pixels. (This image is best viewed in color).

is  $l \times m$ . Thus, each pixel in the image stack can be represented by  $P_{i,j,k}$ , where,  $1 \leq i \leq l$  and  $1 \leq j \leq m$ , as shown in Fig. 1. In any color space, the color value of a pixel  $P_{i,j,k}$ , in the image stack, is defined by:

$$\begin{aligned}
 P_{i,j,k}^{\Omega\Xi\Psi} &= \Omega(i, j, k)\hat{\omega} + \Xi(i, j, k)\hat{\xi} + \Psi(i, j, k)\hat{\psi} \\
 &= [\Omega(i, j, k) \ \Xi(i, j, k) \ \Psi(i, j, k)] \begin{bmatrix} \hat{\omega} \\ \hat{\xi} \\ \hat{\psi} \end{bmatrix} \\
 &= (C_{i,j,k}^{\Omega\Xi\Psi})^T \cdot \hat{c},
 \end{aligned} \tag{13}$$

where  $\Omega_{i,j,k}$ ,  $\Xi_{i,j,k}$ ,  $\Psi_{i,j,k}$  represent the color values of the pixel  $P_{i,j,k}$ ,  $\hat{c} = [\hat{\omega} \ \hat{\xi} \ \hat{\psi}]^T$  is the unit vector, and  $C_{i,j,k}^{\Omega\Xi\Psi}$  is the magnitude vector (in color space  $\Omega\Xi\Psi$ ).

The first step, for calculating the focus value of each colored pixel is by subtracting the neighboring color vectors from the color vector of the pixel  $P_{i,j,k}$ , and collecting them in  $\Delta_{i,j,k}$ , as:

$$\Delta_{i,j,k} = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_{\omega^2-1} \end{bmatrix}. \tag{14}$$

where each  $\delta_r$  is given by:

$$\delta_r = [C_{i,j,k}^{\Omega\Xi\Psi} - C_{\omega_i, \omega_j, k}^{\Omega\Xi\Psi}]^T [C_{i,j,k}^{\Omega\Xi\Psi} - C_{\omega_i, \omega_j, k}^{\Omega\Xi\Psi}] \tag{15}$$

where,  $1 \leq \gamma \leq \omega^2 - 1$ , and  $(\omega_i, \omega_j)$ , represents the indices of  $(\omega^2 - 1)$  neighboring pixels (around  $P_{i,j,k}$ ) present in a window of  $\omega \times \omega$ , and are given as:

$$\begin{aligned}
 (i - \frac{\omega - 1}{2}) \leq \omega_i \leq (i + \frac{\omega - 1}{2}), \\
 (j - \frac{\omega - 1}{2}) \leq \omega_j \leq (j + \frac{\omega - 1}{2}),
 \end{aligned}$$

and,  $\omega_i \neq i$  and  $\omega_j \neq j$ .

In the second step, the variance ( $\sigma_{\Delta}^2$ ) of  $\Delta_{i,j,k}$  is calculated. The final step of obtaining focus measurement  $FM_c(i, j, k)$  of  $P_{i,j,k}^{\Omega\Xi\Psi}$  is given as:

$$FM_c(i, j, k) = \sigma_{\Delta}^2 \sum_{r=1}^{\omega^2-1} \delta_r. \tag{16}$$

After calculating the focus value of each pixel, the depth map of each object-point is obtained by finding the position of the best-focused pixel, corresponding to that object-point.

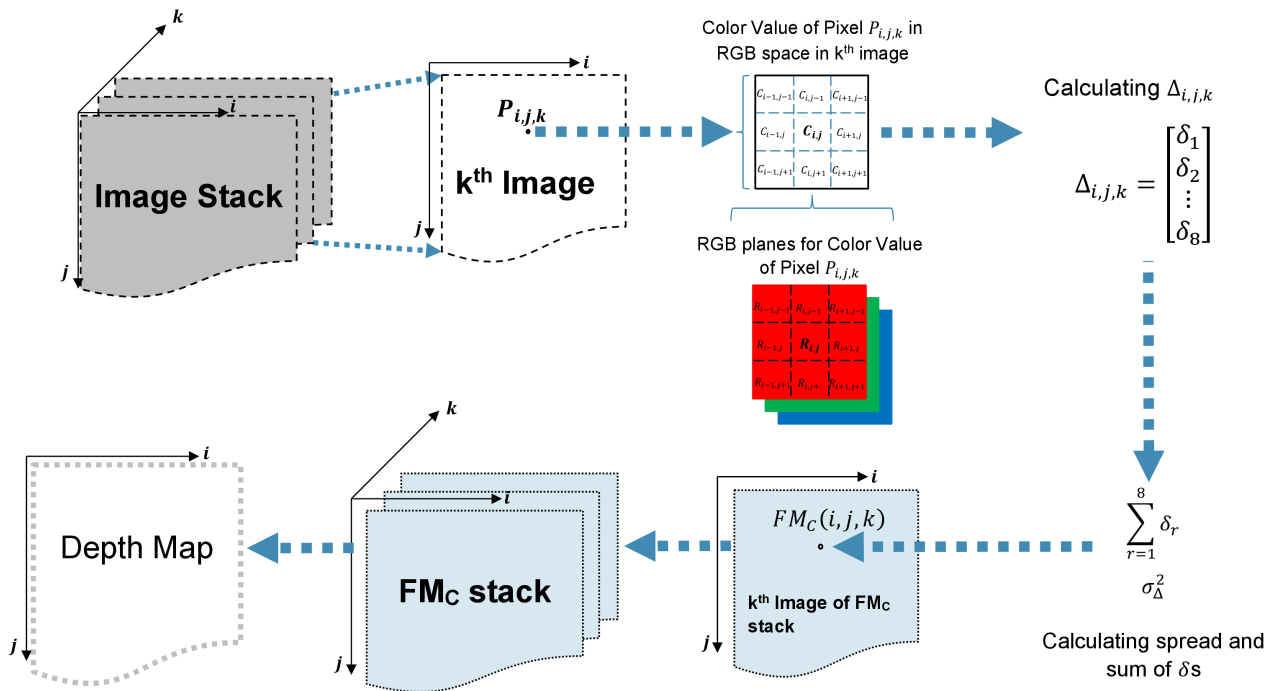


FIGURE 6. Flowchart of proposed method for focus measurement in color space (with neighboring window size of 3 × 3 and in RGB color space).

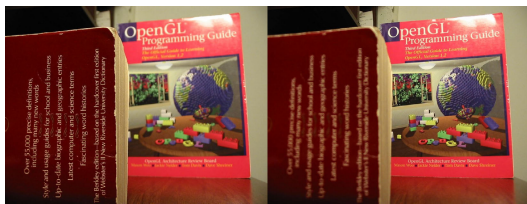


FIGURE 7. Experimented images for focus measure analysis.

This is acquired by maximizing the focus value along the optical axis using the following:

$$D(i, j) = \arg \max_k (F_{M_C}(i, j, k)), \quad (17)$$

where,  $1 \leq k \leq n$ .

When the depth map is obtained for all the object points, a 3D shape is recovered. The focus curve obtained by  $F_{M_C}$  can be further improved by applying certain fitting functions using different models, [50]. The flow of the proposed method is summarized in Fig. 6.

## V. RESULTS AND DISCUSSION

This section provides analysis of experimental results and discusses them in detail, and is divided into four subsections. First, details of the experimental setup are provided, followed by the depth map and shape assessment criteria, and the metric measures used. After the FM analysis is presented, the 3D shape results are analyzed at the end.

### A. EXPERIMENTAL SETUP

The experiments were divided into two categories: FM analysis and 3D shape reconstruction analysis. Different data-sets

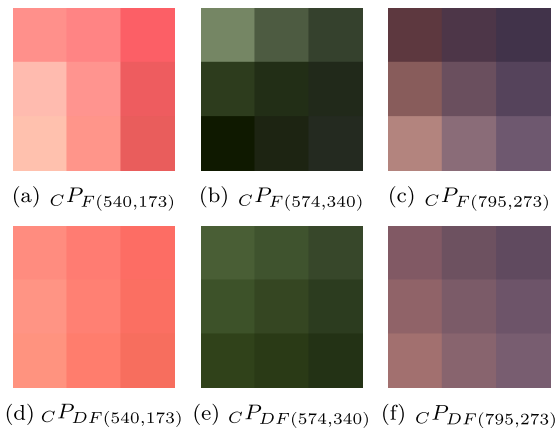
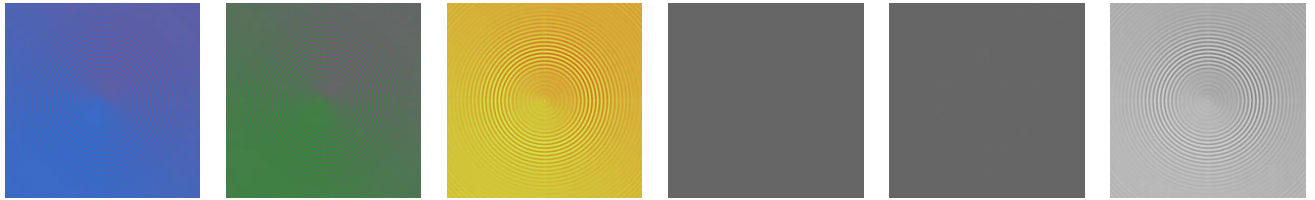


FIGURE 8. Focused and Defocused RGB pixels. (This image is best viewed in color).

were used for each analysis. The following sections explain the experimental setup.

1) EXPERIMENTAL SETUP FOR FOCUS MEASURE ANALYSIS  $F_{M_C}$  is analyzed for two images as shown in Fig. 7 (left & right). In Fig. 7 (left) the foreground is focused and the background is defocused, whereas, in Fig. 7 (right) the foreground is defocused and the background is focused. Three focused pixels are randomly selected from one image and their duplicate defocused pixels are selected from the other. The pixels are then analyzed for their focus-defocus difference. The results are presented in the later section of the manuscript. Fig. 8 provides the focus-defocus images of selected pixels along with their 3 × 3 neighbors.



(a) Simulated Cone 1 (b) Simulated Cone 2 (c) Simulated Cone 3 (d) Simulated Cone 1 (e) Simulated Cone 2 (f) Simulated Cone 3

**FIGURE 9.** 47<sup>th</sup> frame: color (top row) and grayscale (bottom row) image from simulated cone image sequences. (This image is best viewed in color).

**TABLE 1.** Summary of Experimented Objects for 3D shape generation.

Name	Object for Image Generation	Color Scheme	Dimensions of each image in stack	Number of Images
Simulated Cone 1	Simulated Object	RGB Color	360 × 360	97
Simulated Cone 2				
Simulated Cone 3				
Real Cone 1	Real Object	Pseudo Coloring	200 × 200	87
Real Plane				
LCD-TFT				
Real Cone 2		500 × 500	50	
Coin				
Measuring Tape		200 × 200	64	
Balls				
Bucket		360 × 640	24	
Fruits				
Keyboard				
Plants				
Kitchen				
Logi1				
Logi2		518 × 774	11	
Logi3				
Logi4				
			RGB Color	360 × 640

**TABLE 2.** Percentage difference between focus and defocus value.

Focus Measure Operator	P(540,173)	P(574,340)	P(795,273)	Average value for Whole Image
$FM_c$	194.52 %	197.84 %	174.62 %	190.99 %
CONT	133.18 %	156.53 %	118.45 %	138.05 %
CURV	34.00 %	25.40 %	13.90 %	26.43 %
DCTE	1.83 %	1.13 %	2.40 %	3.79 %
HISE	28.70 %	0.00 %	0.00 %	11.57 %
SML	38.30 %	44.50 %	54.00 %	47.60 %
LAPM	101.42 %	104.41 %	62.26 %	91.36 %
TENG	143.69 %	145.93 %	130.53 %	142.05 %
GLVA	84.65 %	85.69 %	73.96 %	83.43 %
DCTR	0.55 %	1.10 %	0.27 %	2.64 %
SFRQ	0.00 %	0.00 %	9.80 %	5.27 %
GRAT	2.60 %	27.50 %	0.00 %	12.03 %
WAVS	97.00 %	-85.00 %	84.00 %	65.67 %

2) EXPERIMENTAL SETUP FOR 3D SHAPE RECONSTRUCTION ANALYSIS

Experiments for shape reconstruction analysis were performed on publicly available *five* data-sets of total 19 objects. Table 1, provides a summary of the objects used in the 3D shape analysis, including the name of the object, object type, color scheme dimension of the image in stack and number of images in the stack. The first set contains three simulated objects, Simulated Cone 1, 2 & 3, as given in Fig. 9. Three

different colored data-sets of simulated cones were generated with different lens position using camera simulation software (AVS). The details of AVS are provided in, [51]–[53]. The Matlab code used can be downloaded from, [53]. All data-sets consists of 97 images with 360 × 360 pixels.

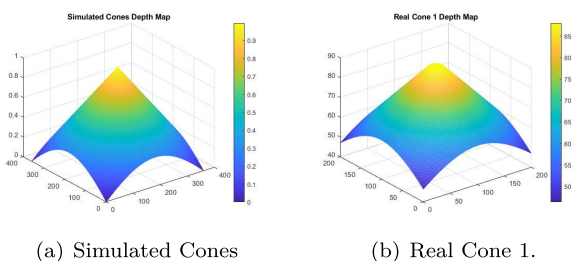
The AVS software is provided with the depth map, texture image and camera parameters. The texture map consists of concentric circles of two alternating colors. The colors are chosen in such a way that their gray level values are the

**TABLE 3.** RMSE and Correlation for depth estimation methods by different FMs for 3 data-sets of Simulated Cone and 1 data-set of Real Cone 1. The  $FM_c$  uses color image sequence, other FMs utilizes the gray scale image sequence.

Focus Measure Operator	Simulated Cone 1		Simulated Cone 2		Simulated Cone 3		Real Cone 1	
	Corr.	RMSE	Corr.	RMSE	Corr.	RMSE	Corr.	RMSE
$FM_c$	<b>0.9998</b>	<b>2.0716</b>	<b>0.9998</b>	<b>2.0713</b>	<b>0.9998</b>	<b>2.0681</b>	<b>0.9501</b>	<b>1.4929</b>
CONT	-0.3606	27.3614	0.0634	19.0177	0.8516	7.4487	0.9461	1.5349
CURV	-0.4823	28.6858	-0.0175	20.0373	0.9863	2.7112	0.3839	8.5108
DCTE	0.0946	26.9779	0.5200	20.7790	0.9984	3.1684	0.9372	1.6452
HISE	-0.0748	32.1034	-0.1342	32.1034	0.8964	32.1034	0.7414	78.3515
SML	-0.0479	34.2070	0.5502	20.9053	0.9939	3.4001	0.9409	1.6082
LAPM	-0.2121	28.4833	0.0178	18.8496	0.9985	2.4500	0.9429	1.5815
TENG	-0.0358	23.7931	0.2381	19.5518	0.9982	2.5798	0.9432	1.5749
GLVA	-0.0762	31.4500	-0.1346	34.7200	0.9835	5.1300	0.9393	3.2100
DCTR	0.1082	27.0112	0.4129	22.8013	0.9413	4.6341	0.3394	12.6873
SFRQ	-0.2397	23.3251	-0.0015	16.2861	0.9984	2.4047	0.9417	1.5999
GRAT	0.2694	17.3394	0.0181	15.7636	0.9638	3.2112	0.9416	1.6370
WAVS	-0.3888	32.1449	-0.0779	23.4656	0.9507	9.6151	0.9434	1.5800

**TABLE 4.** RMSE and Correlation for depth estimation methods by different FMs for 3 data-sets of Simulated Cone and 1 data-set of Real Cone 1. The  $FM_c$  is applied on color image sequence, the other FMs are applied on each channel of R, G and B separately, before averaging the resultant shapes.

Focus Measure Operator	Simulated Cone 1		Simulated Cone 2		Simulated Cone 3		Real Cone 1	
	Corr.	RMSE	Corr.	RMSE	Corr.	RMSE	Corr.	RMSE
$FM_c$	<b>0.9998</b>	<b>2.0716</b>	<b>0.9998</b>	<b>2.0713</b>	<b>0.9998</b>	<b>2.0681</b>	<b>0.9501</b>	<b>1.4929</b>
CONT	0.9385	5.4575	0.9498	4.8560	0.9240	5.9220	0.9303	1.8805
CURV	0.9500	3.1874	0.9222	3.4774	0.9645	3.5970	0.2627	27.8693
DCTE	0.9989	3.4341	0.9986	3.5204	0.9989	3.0550	0.9010	2.1773
HISE	-0.3178	26.1929	-0.3010	24.4407	-0.1535	27.3268	0.7219	4.4770
SML	0.9901	12.5654	0.9896	13.8673	0.9956	6.5384	0.9362	2.9610
LAPM	0.9893	2.1223	0.9833	2.1503	0.9871	2.2198	0.2875	16.7199
TENG	0.3952	10.8054	0.3720	11.4883	0.4550	8.8894	-0.2639	15.5284
GLVA	0.9989	2.8012	0.9987	2.8792	0.9989	2.4359	0.9438	1.5881
DCTR	0.9800	4.4207	0.9800	4.4624	0.9806	3.8846	0.3615	12.2318
SFRQ	0.3769	10.7838	0.2687	11.1194	0.5257	9.8676	0.3668	17.4336
GRAT	0.9292	3.4924	0.9229	3.5281	0.9253	4.5227	0.2821	21.7600
WAVS	0.9761	9.1390	0.9752	9.0620	0.9709	9.7167	0.9428	1.6030

**FIGURE 10.** Ground truth of experimented objects.

same (or close to each other) in each image sequence of simulated cone. The simulated cones used in the experiments differ from each other in color. Each image sequence has two different colors added to it. The depth maps and the texture images used for image generation via AVS for all sequences of Simulated Cone are the same.

The second data-set contains three real objects, Real Cone 1, Real Plane, and the LCD-TFT Filter. These image sequences were originally in gray scale, but for the sake of this experiment they were pseudo-colored, [54]–[56]. The colors were chosen in the same way as for data-set 1.

Fig. 11 shows the 50<sup>th</sup> frame of each image sequence. These image sequences have been widely used by many researchers, [5], [38], [57]–[59]. Fig. 10 provides the ground truth for Simulated cones and Real Cone 1 data-sets.

The third data-set consists of three real objects, a Coin, Measuring Tape, and Real Cone 2. The Coin and Measuring Tape image sequences were taken from a video source provided by, [60], using the frames from 1:15–2:23 seconds. The Real Cone 2 image sequence was generated by changing the focus of a Nikon D5300 camera, with 35–300 mm lens, the camera was mounted on the stand, using the system described in, [60]. The focal length of the lens was kept constant while the focus of the lens was changed in small steps. For this image sequence, a cone-paper-cup was used and color stripes were drawn onto it to provide texture. The two colors used were red and blue. Fig. 12 provides the 65<sup>th</sup> frame of the Real Cone 2 image sequence, and 50<sup>th</sup> frame of the Coin and Measuring Tape image sequences in both colored and gray scale.

The fourth data-set consists of six real objects, Balls, Bucket Fruits, Keyboard, Plants and Kitchen. All of these images are taken from the mobile phone. The details

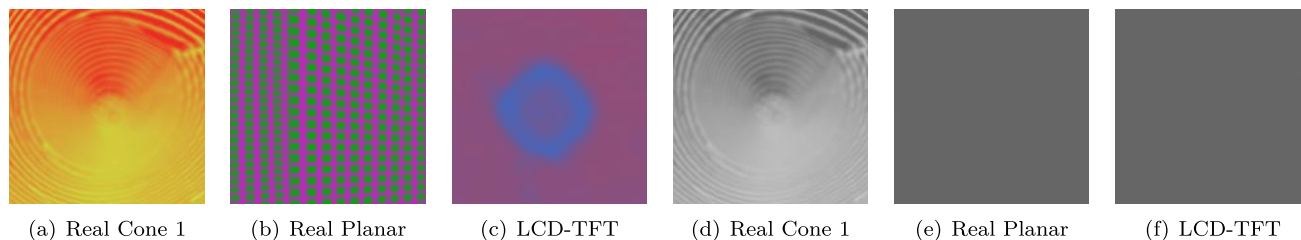


FIGURE 11. 50<sup>th</sup> frame: color (top row) and grayscale (bottom row) image from each data-set. (This image is best viewed in color).

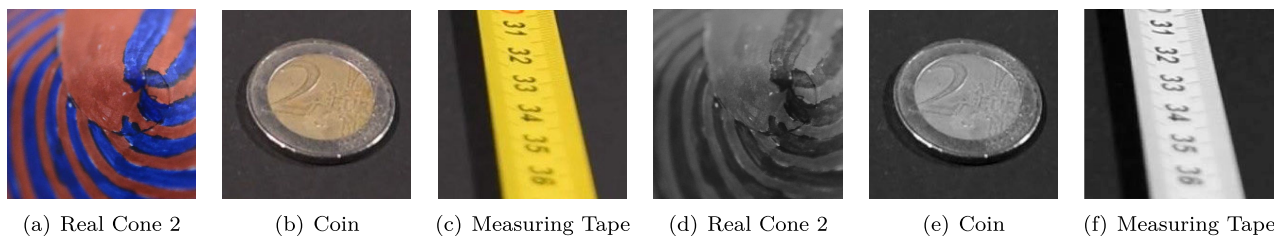


FIGURE 12. Color (a, b, c) and grayscale (d, e, f) image from each data-set. (Real Cone 2: 65<sup>th</sup> frame; Coin and Measuring Tape: 50<sup>th</sup> frame). (This image is best viewed in color).

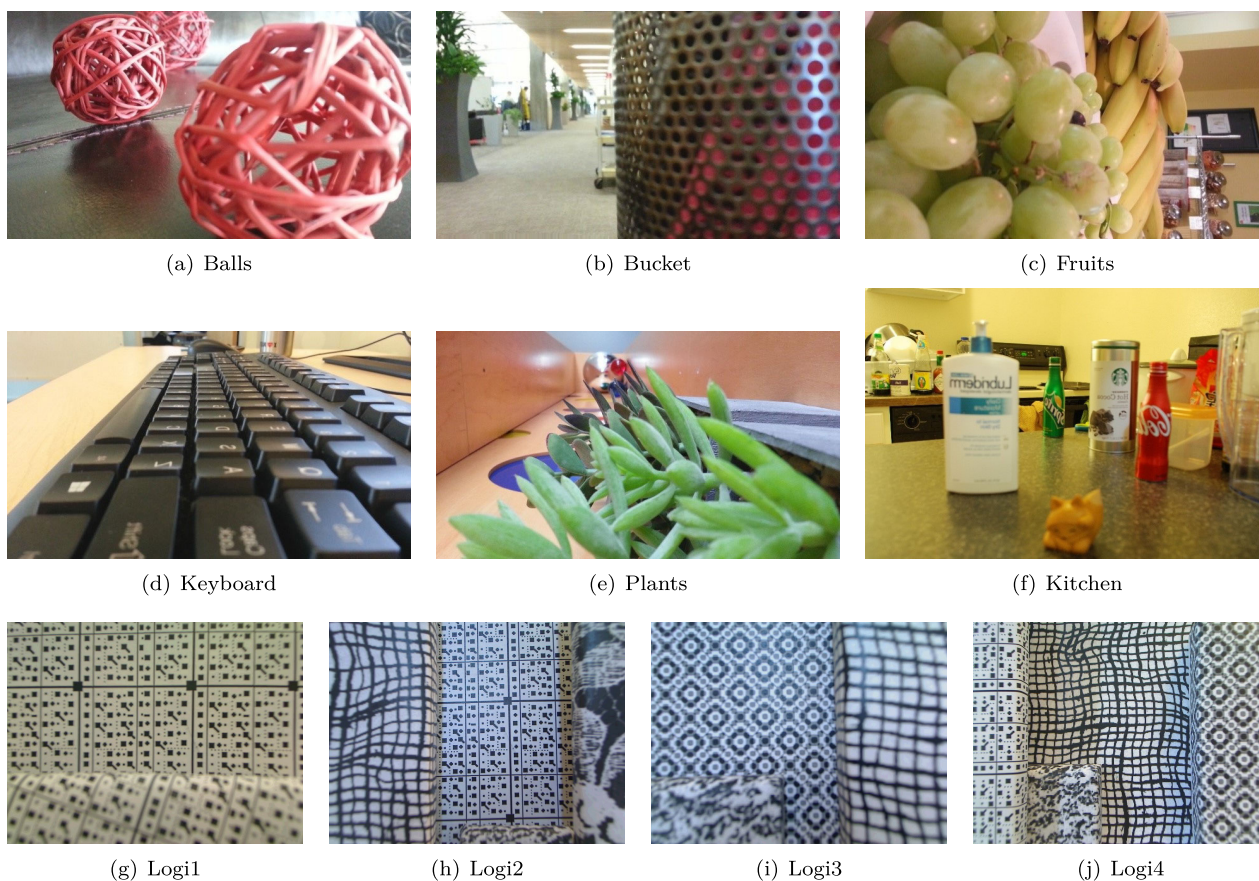


FIGURE 13. 10<sup>th</sup> frame of image sequence of other objects.

of the image sequences are provided in, [61], and are available at, [62].

The fifth data-set consists of four objects, Logi1, Logi2, Logi3 and Logi4. These image sequences are obtained from

a Logitech webcam. The details of these are provided in, [12], and are available at, [53].

Fig. 13 provides 10<sup>th</sup> image of each image sequence of fourth and fifth data-sets, respectively.



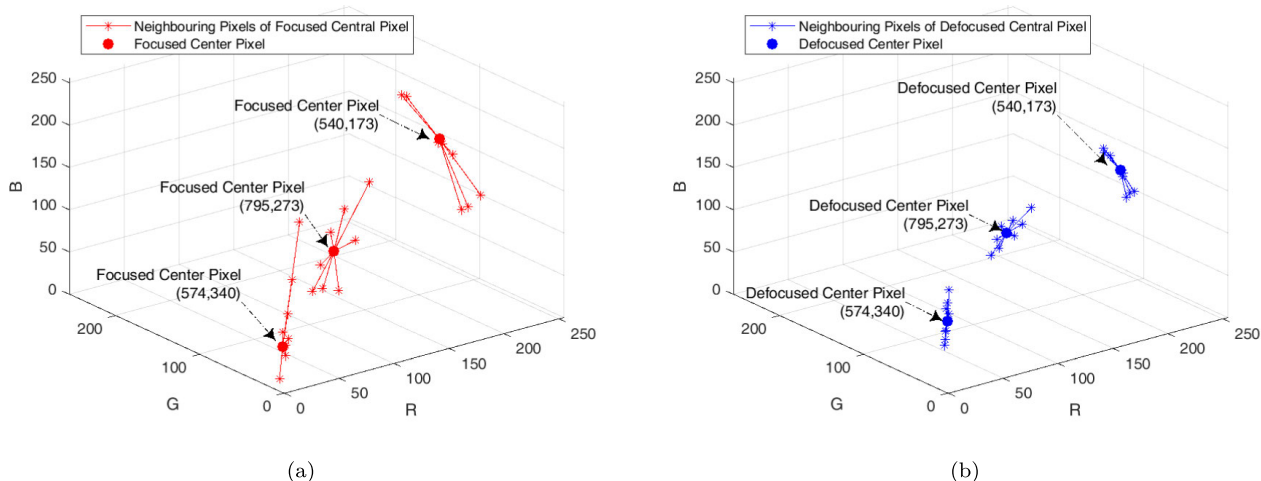


FIGURE 14. Examples of Focusing and Defocusing of a pixels in RGB color space along with its neighbors. (a) Pixels and their neighbors when central pixel is focused, (b) Pixels and their neighbors when central pixel is defocused.

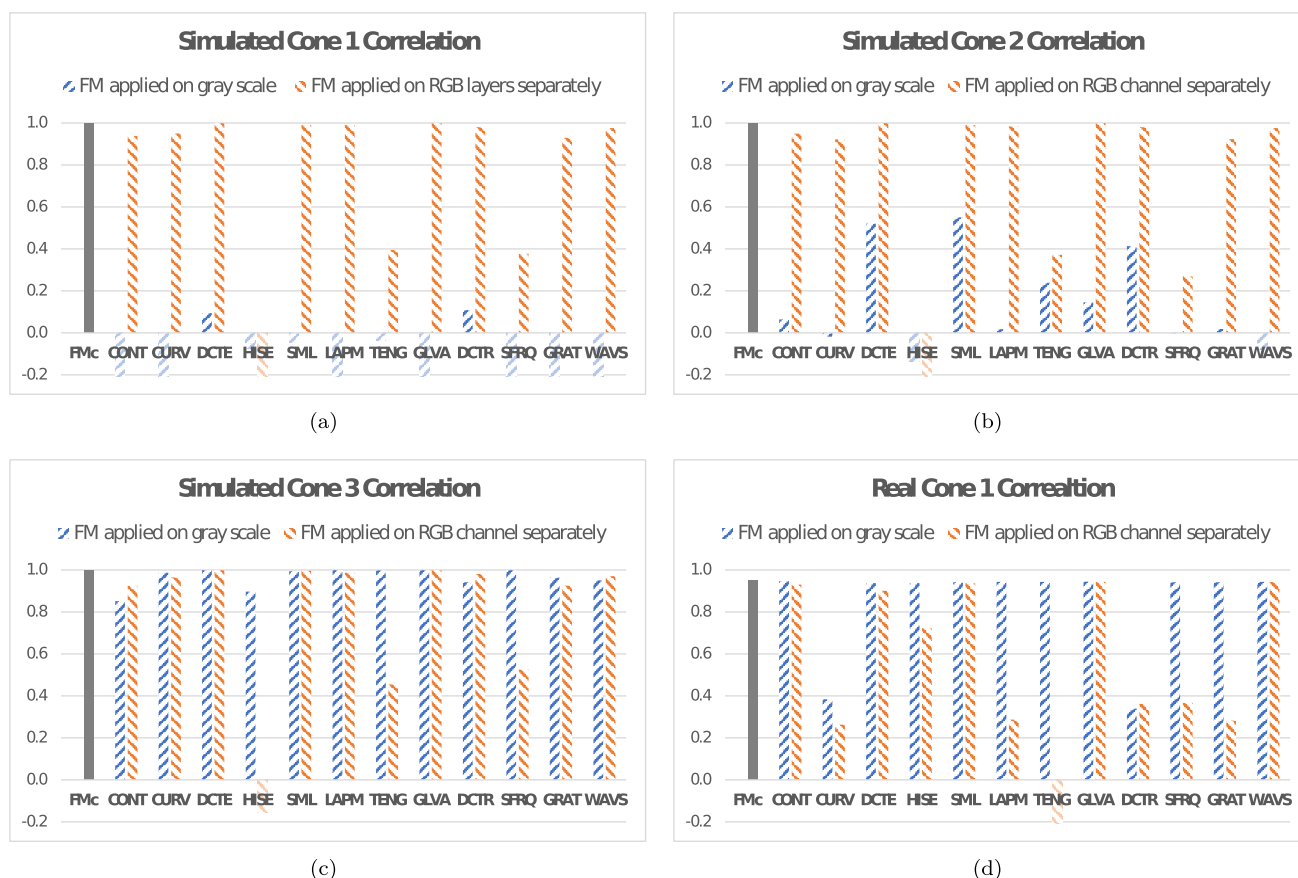
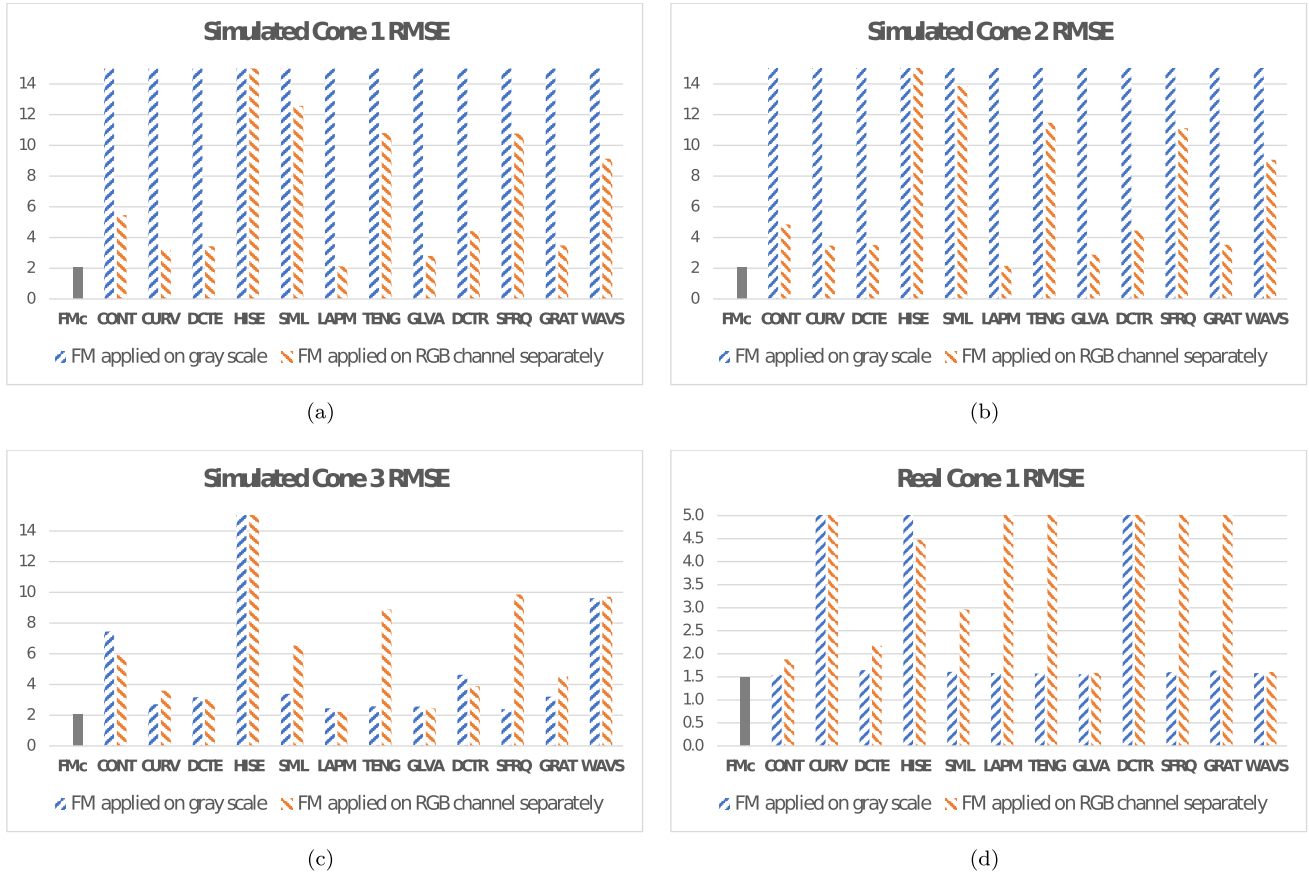


FIGURE 15. Correlation for (a) Simulated Cone 1, (b) Simulated Cone 2 & (c) Simulated Cone 3, and (d) Real Cone 1. Solid bar indicates  $FM_c$ , the shaded bars indicate the other FMs used (i) when color image sequence is converted to gray scale and (ii) when the FM is applied on each channel of R, G and B, before averaging the resultant shapes.

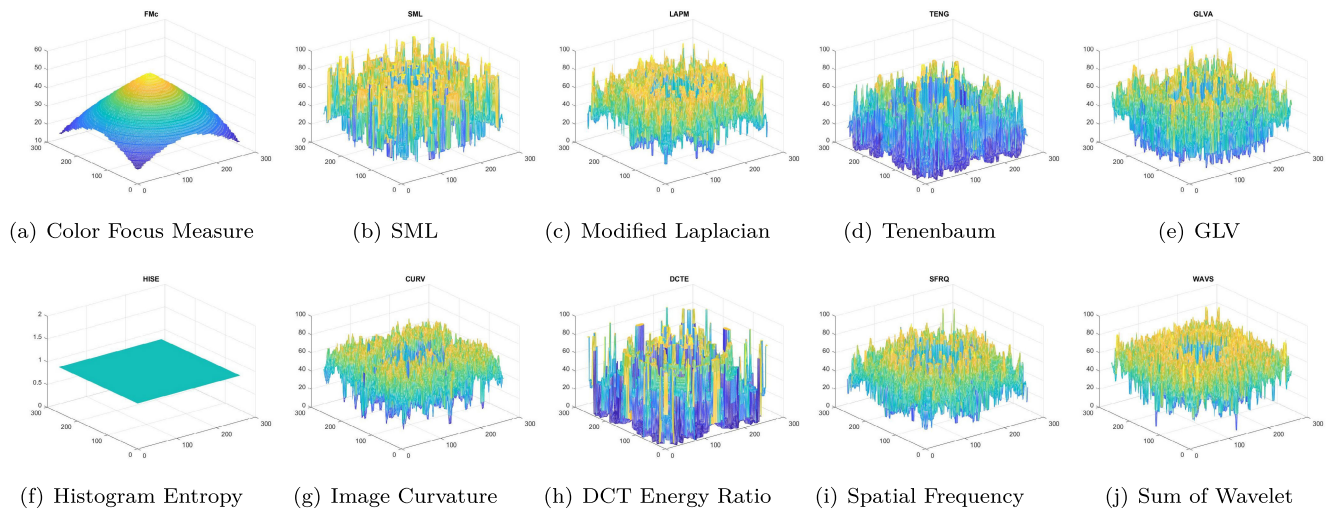
**B. DEPTH MAP/SHAPE ASSESSMENT CRITERIA**

The shape reconstruction quality is the characteristic that measures the perceived difference between the reconstructed shape and the ideal shape. The quality of shape reconstruction

reduces as the difference increases. In this article, the quality of the depth map obtained by using different focus measures is also analyzed. In the ideal case, the obtained depth map is indistinguishable from the original map and the difference is



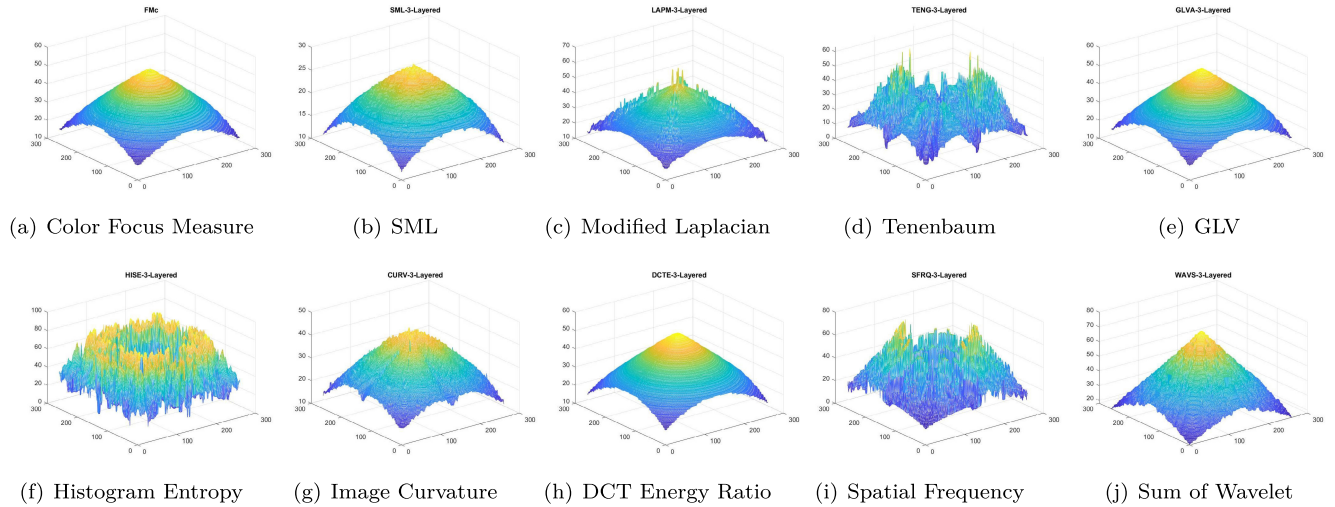
**FIGURE 16.** RMSE for (a) Simulated Cone 1, (b) Simulated Cone 2 & (c) Simulated Cone 3, and (d) Real Cone 1. Solid bar indicates  $FM_c$ , the shaded bars indicate the other FMs used (i) when color image sequence is converted to gray scale and (ii) when the FM is applied on each channel of R, G and B separately, before averaging the resultant shapes.



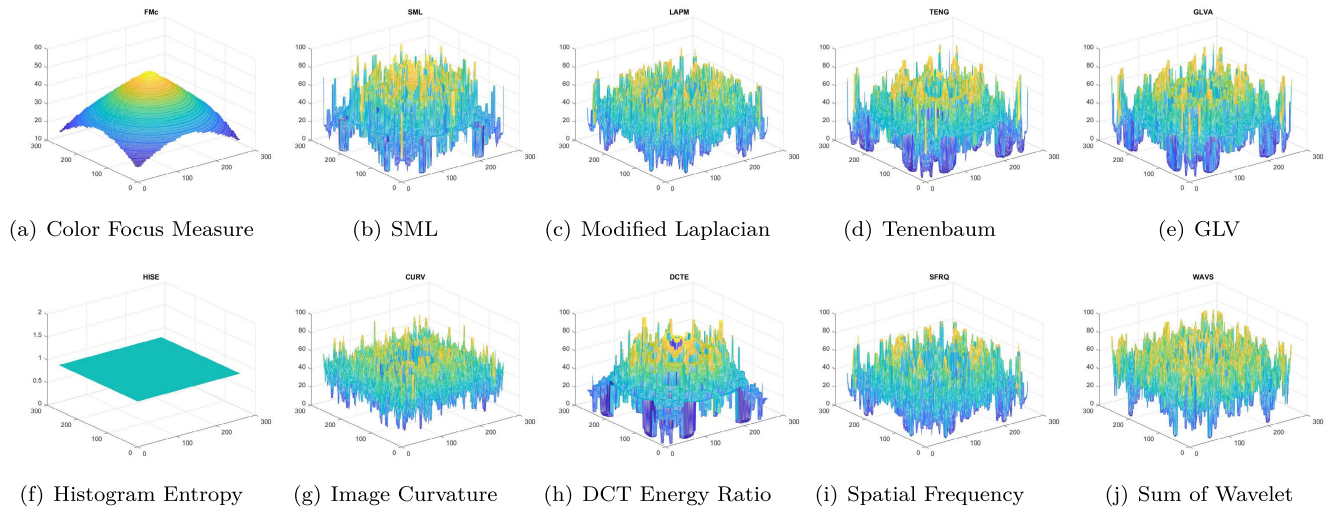
**FIGURE 17.** Shape reconstruction of Simulated Cone 1 with proposed and previous techniques. (The  $FM_c$  uses color image sequence, other FMs utilizes the gray scale image sequence).

zero, hence, the quality of the map is at its maximum. Several quality metrics have been provided in the previous literature. In this manuscript, RMSE and correlation are used to compare the proposed  $FM_c$  with other focus measures operators.

1) **ROOT MEAN SQUARE ERROR (RMSE)**  
 RMSE is the square root of the variance of the residuals of the data under observation. It indicates how close the perceived shape is from the original shape. A lower value of  $RMSE$



**FIGURE 18.** Shape reconstruction of Simulated Cone 1 with proposed and previous techniques. (The  $FM_c$  is applied on color image sequence, the other FMs are applied on each channel of R, G and B separately, before averaging the resultant shapes).



**FIGURE 19.** Shape reconstruction of Simulated Cone 2 with proposed and previous techniques. (The  $FM_c$  uses color image sequence, other FMs utilizes the gray scale image sequence).

indicates better results, [63].  $RMSE$  is given by:

$$RMSE = \sqrt{\frac{1}{XY} \sum_{j=0}^{X-1} \sum_{i=0}^{Y-1} (D_{ref}^{(i,j)} - D_p^{(i,j)})^2}, \quad (18)$$

where  $X$  and  $Y$  are the horizontal and vertical number of pixels.  $D_{ref}^{(i,j)}$  represents the reference depth map and  $D_p^{(i,j)}$  represents the processed depth map.

## 2) CORRELATION

Correlation is a the statistical relationship between two variables that are being measured. This metric is best used to demonstrate the linear relationship between two shapes, [64].

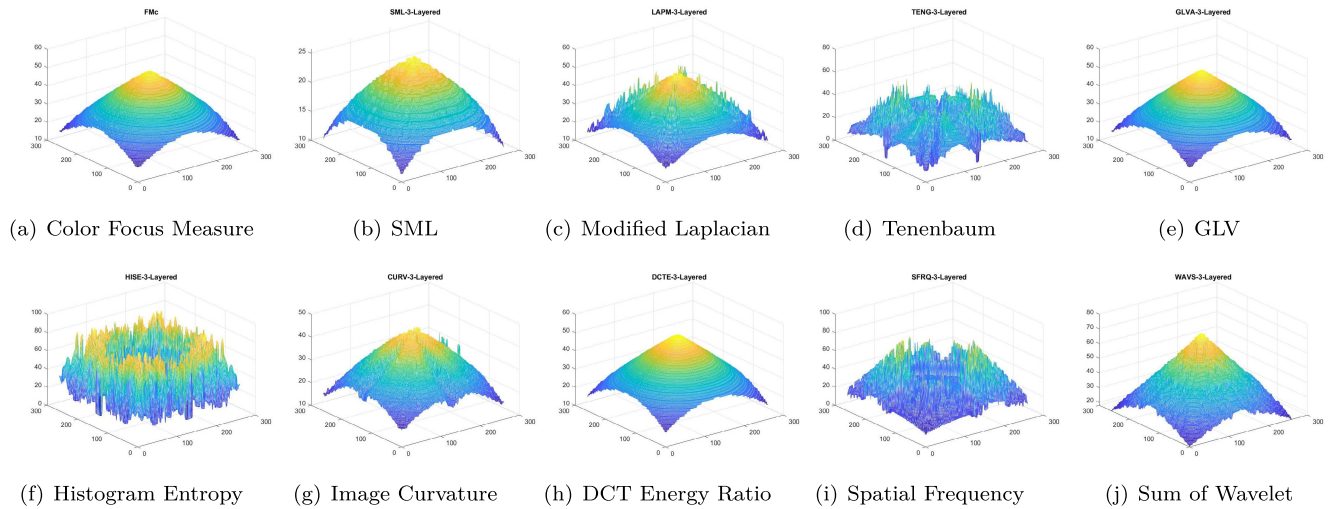
The correlation coefficient is computed as:

$$Corr. = \frac{\sum_{ij} (D_{ref}^{(i,j)} - \mu_{ref}) (D_p^{(i,j)} - \mu_p)}{\sqrt{\sum_{ij} (D_{ref}^{(i,j)} - \mu_{ref})^2 \sum_{ij} (D_p^{(i,j)} - \mu_p)^2}}, \quad (19)$$

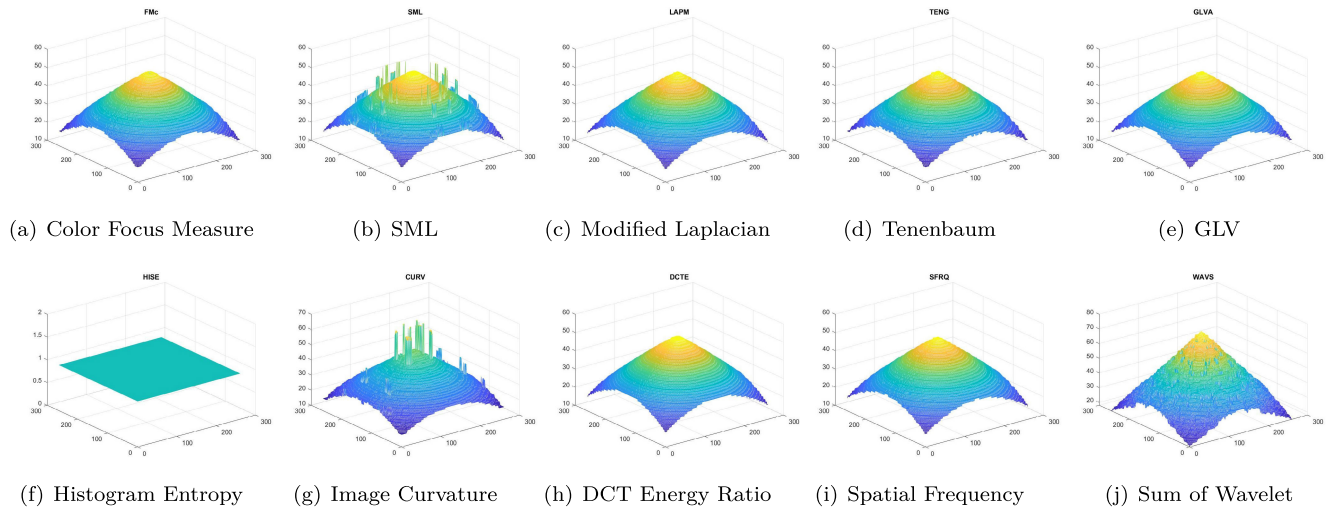
where  $D_{ref}^{(i,j)}$  and  $\mu_{ref}$  are the reference depth map and its mean value, and  $D_p^{(i,j)}$  and  $\mu_p$  represent the processed depth map and its mean value, respectively.  $Corr. = 1$  shows the maximum correlation between  $D_{ref}^{(i,j)}$  and  $D_p^{(i,j)}$ .

## C. PROPOSED FOCUS MEASURE ANALYSIS

Focused and defocused pixels are shown in Fig. 8 (top and bottom row, respectively). Their values with their neighbors' value (in the  $3 \times 3$  window) are plotted, in RGB color space



**FIGURE 20.** Shape reconstruction of Simulated Cone 2 with proposed and previous techniques. (The  $FM_c$  is applied on color image sequence, the other FMs are applied on each channel of R, G and B separately, before averaging the resultant shapes).

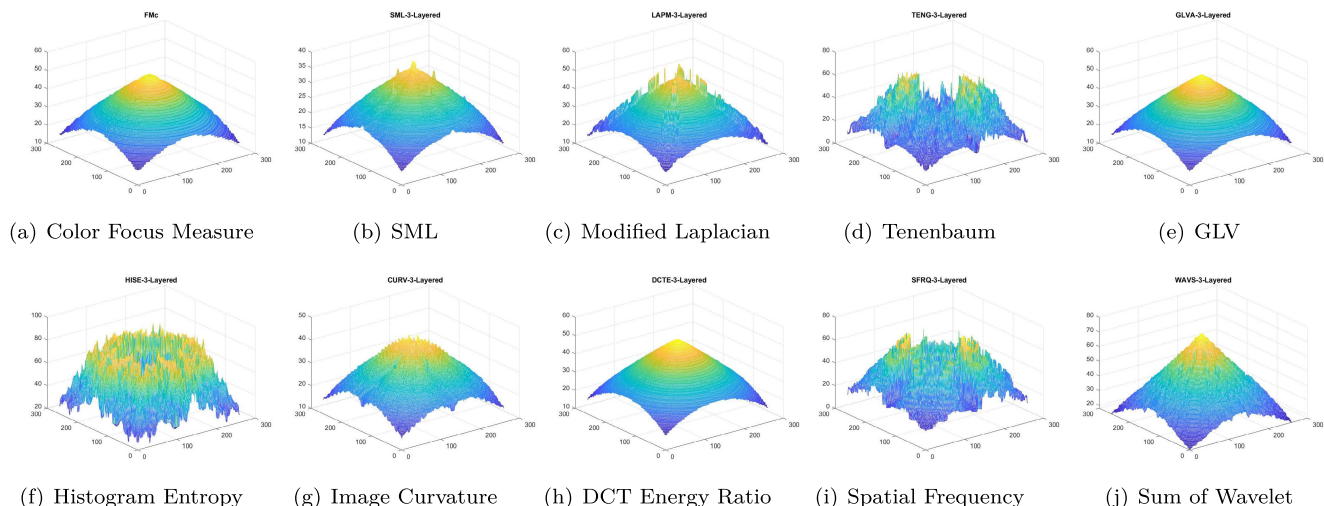


**FIGURE 21.** Shape reconstruction of simulated cone 3 with proposed and previous techniques. (The  $FM_c$  uses color image sequence, other FMs utilizes the gray scale image sequence.).

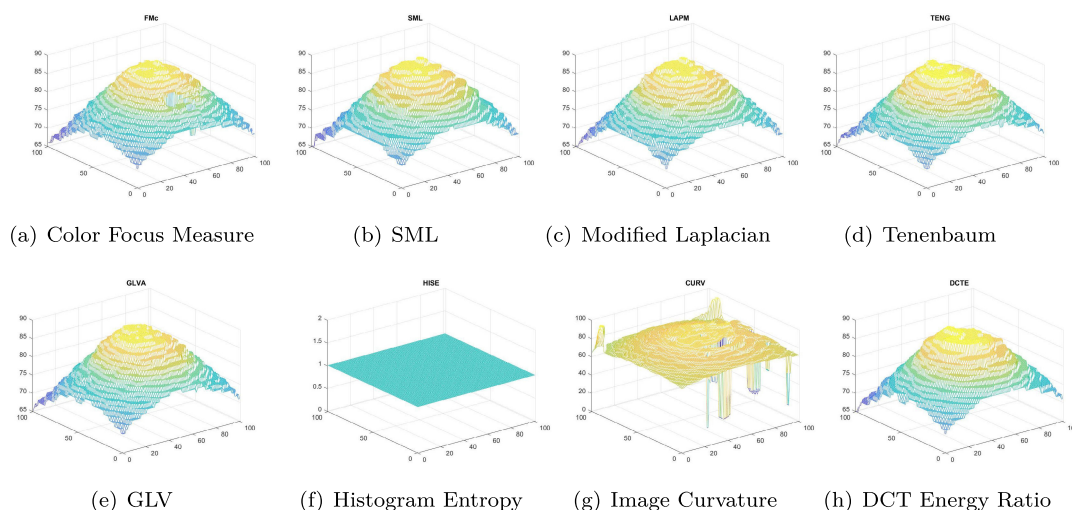
and shown in Fig. 14(a) and Fig. 14(b), respectively. It is evident from the figures that the RGB-color-distances and the spread of the distances from the *focused center pixel* to its neighbor pixels are larger as compared to the distances of the *defocused center pixel* to its neighbors. Hence, the RGB-color-distances and their spread will be higher for focused pixels. This leads to the conclusion that when the focus of the pixel is reduced, the color-distances and their spread are also reduced. This phenomenon allows us to separate the best focused pixel from other pixels.

The proposed  $FM_c$  was also compared with the algorithms of some other well-known FM methods like Image Contrast(CONT), Image Curvature (CURV), DCT Energy Ratio (DCTE), Histogram Entropy (HISE), Sum of Modified Laplacian (SML), the Modified Laplacian (LAPM), Tenenbaum (TENG) and Grey level variance (GLVA), DCT

Reduced Energy Ratio (DCTR), Spatial Frequency (SFRQ), Threshold Absolute Gradient (GRAT), and Sum of Wavelet Coefficients (WAVS). For comparison purposes, all the FMs were tested on equal terms. The colored pixels shown in Fig. 8 were converted to gray scale, since previous methods cannot process colored images. The percentage differences between the focused and defocused value of each of these selected pixels were calculated using  $FM_c$ , they were then compared with the results obtained from previous algorithms. Table 2 shows the percentage difference between the focus and defocus values of the selected pixels. It is clear that for the proposed method, there is a larger difference between the focused and defocused values, whereas the other methods give smaller difference or in some cases no difference at all. The second to fourth column of the table illustrates the percentage difference for the selected pixels, whereas, fifth



**FIGURE 22.** Shape reconstruction of Simulated Cone 2 with proposed and previous techniques. (The  $FM_c$  is applied on color image sequence, the other FMs are applied on each channel of R, G and B separately, before averaging the resultant shapes).



**FIGURE 23.** Shape reconstruction of Real Cone 1 with proposed and previous techniques. (The  $FM_c$  uses color image sequence, other FMs utilizes the gray scale image sequence).

column of the table provides the same analysis averaged for the whole image. Some of the methods give mixed results, such as WAVS, and therefore, they cannot be considered as reliable as  $FM_c$ . Image Contrast has the best results among all the other FMs, but the proposed FM still significantly outclasses Image Contrast FM.

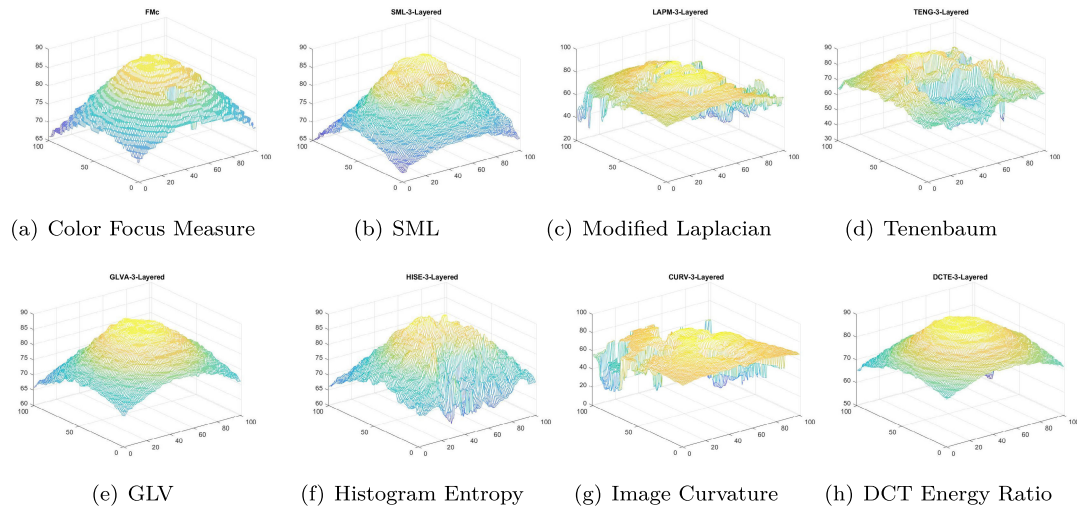
### D. 3D SHAPE RECONSTRUCTION ANALYSIS

The shape reconstruction results of  $FM_c$  are also compared with some well-known FMs (as mentioned in beginning of Section 5). Some of these operators are discussed in detail in, [12].

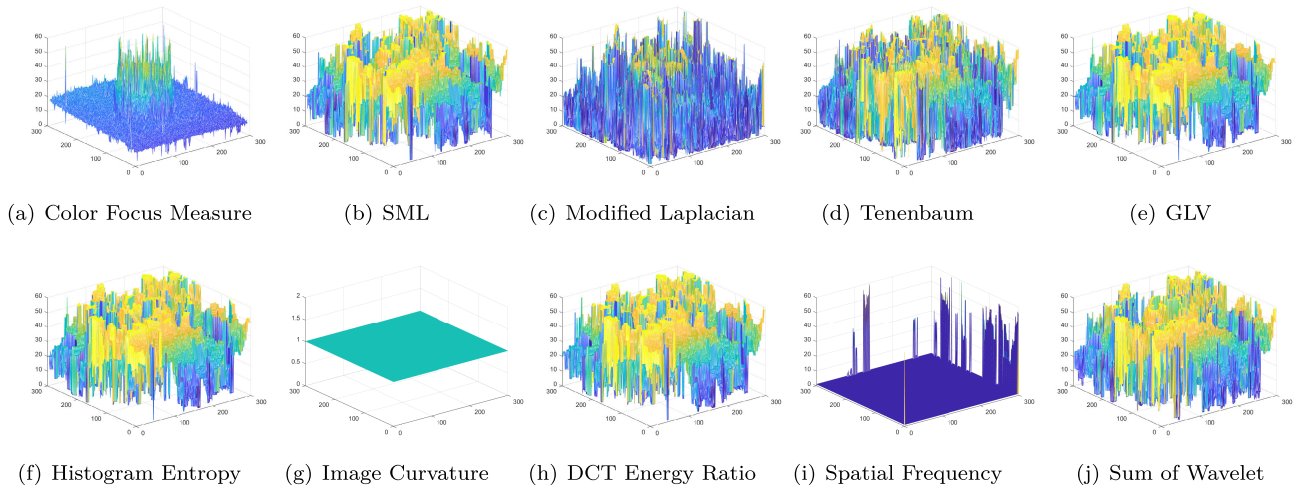
For a fair comparison, the colored data-set was converted to gray scale so that a 3D shape could be recovered using these FMs as they cannot process colored images. Some gray scale data-sets showed the same gray value, and thus some frames could not be distinguished from one another.

Fig. 9(b, d, f) shows the gray scale converted frames of colored simulated cones whereas, Fig. 11(d, e, f) represent the 50<sup>th</sup> frame converted to gray scale from the image stack of Real Cone 1, Real Plane and LCD-TFT Filter. Also, Fig. 12(d, e, f) show the example gray scale image from the stack of Real Cone 2, Coin, and Measuring Tape. Also, for the fair comparison, the FMs were applied on each channel of R, G, and B, separately. The resultant shapes acquired from each channel processing was then averaged to form the final shape. The results of these shapes reconstructed using different FMs are also compared with the shape reconstructed using  $FM_c$ .

Table 3 & 4, along Fig. 15 & 16 provide the shape analysis results in terms of correlation and RMSE for simulated cones 1, 2 & 3, and real cone 1, respectively. In the tables it is clearly visible that the values of  $FM_c$  in terms of correlation and RMSE are better than other FMs.



**FIGURE 24.** Shape reconstruction of Real Cone 1 with proposed and previous techniques. (The  $FM_c$  is applied on color image sequence, the other FMs are applied on each channel of R, G and B separately, before averaging the resultant shapes).



**FIGURE 25.** Shape reconstruction of LCD-TFT with proposed and previous techniques.

In both the tables the  $FM_c$  uses color image sequence, while for Table 3 other FMs utilizes the gray scale image sequence, and in Table 4 the other FMs are applied on each channel of R, G and B separately, before averaging the resultant shapes from each channel. Table 3 clearly demonstrates that for Simulated Cone 1 & 2, the RMSE and correlation are very poor. For Simulated Cone 3 and Real Cone 1, the results are comparatively better. The reason behind these poor results for Simulated Cone 1 & 2 for other FMs is, the fact that the conversion of color images to gray scale have lost the information of focus as many gray scale values are similar to each other. This fact is clear from the Fig. 9. For Simulated Cone 3 and Real Cone 1, the gray scale conversion have not much effect and the results are better. In Table 4 the FMs are applied on each RGB channel separately, and no gray scale conversion is applied. Therefore, the results of simulated cones 1, 2 & 3, and real cone 1, in terms of RMSE and correlation are very impressive. However,  $FM_c$  is still

outstanding among all. Fig. 15 & 16 provides the graphical representation of the data provided in the tables. These tables and figures clearly demonstrate that the  $FM_c$  has performed better as compared to other FMs even when they are applied to RGB channels separately. The application of FMs on each RGB channel not only increases the computational complexity and time consumption, but also, does not guarantee the improvement of the shape. For example, TENG for Simulated Cone 3 has performed better when applied to gray scale images sequence (when converted from color images for both RMSE and correlation) as compared to when applied on to R, G, and B channels separately to compute the shape. The similar trend for TNEG is observed in case of Real Cone 1. Other FMs like CURV, SML, SFRQ, & GRAT have shown similar behavior.

These FMs failed to process gray scale images (after converted from color sequence) for certain colors, and have also shown mixed results when applied to RGB channels

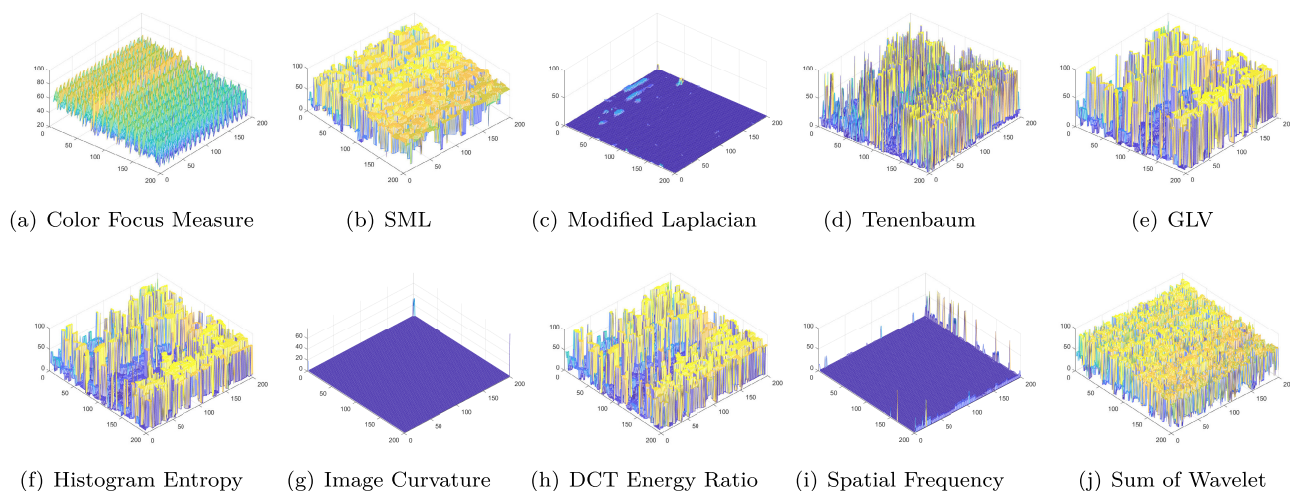


FIGURE 26. Shape reconstruction of Real Plane with proposed and previous techniques.

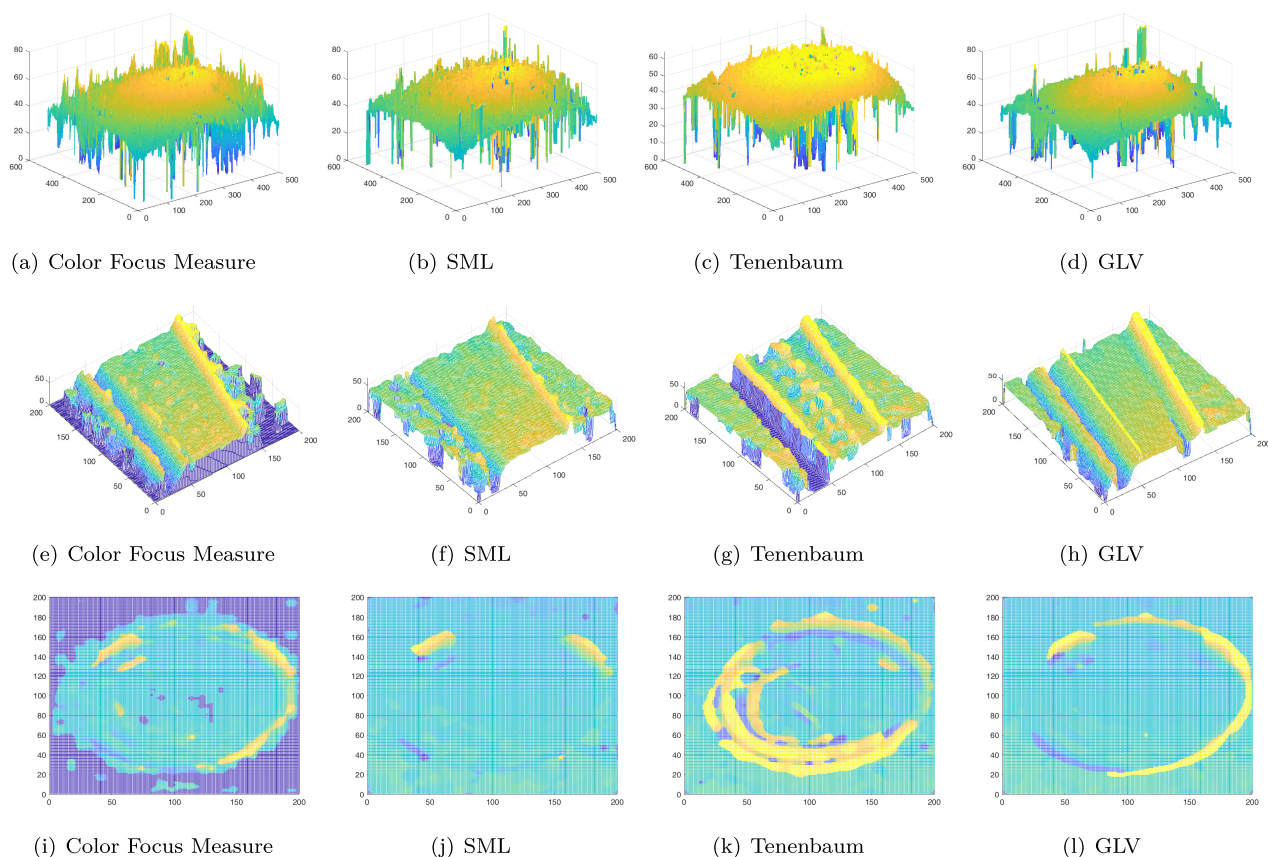


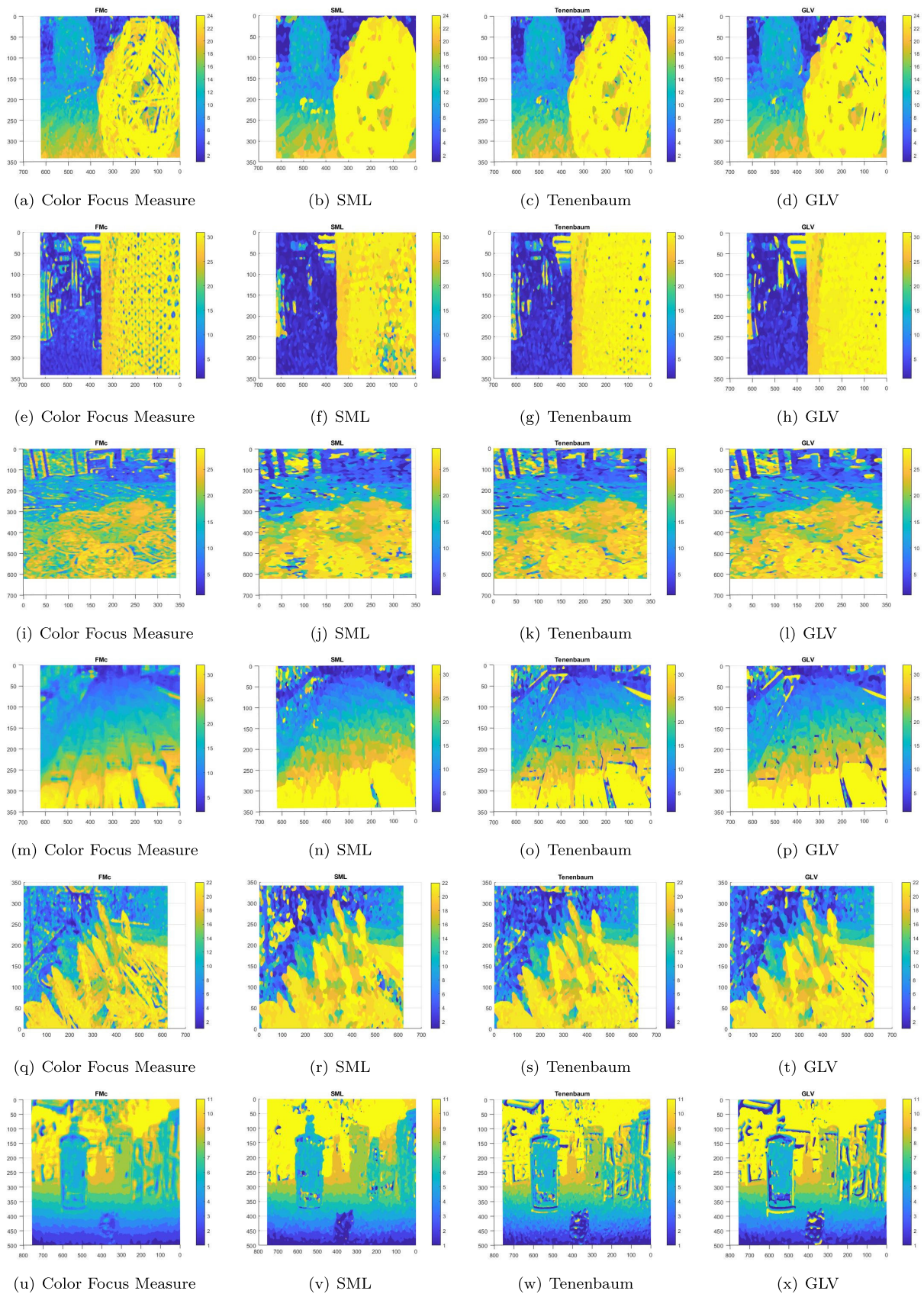
FIGURE 27. Shape reconstruction of Top: Real Cone 2, Middle: Measuring Tape and Bottom: Coin image sequence with proposed and previous techniques.

separately, whereas,  $FM_c$  showed consistent results for all four sequences of simulated cones and real cone. The graphs and tables show the effectiveness of the proposed scheme.

Fig. 17 & 18, Fig. 19 & 20, and Fig. 21 & 22 represent the shape reconstruction from three different data-sets of the Simulated Cone with  $FM_c$  and previous techniques when (i) color image sequence is converted to gray scale and (ii) the FM

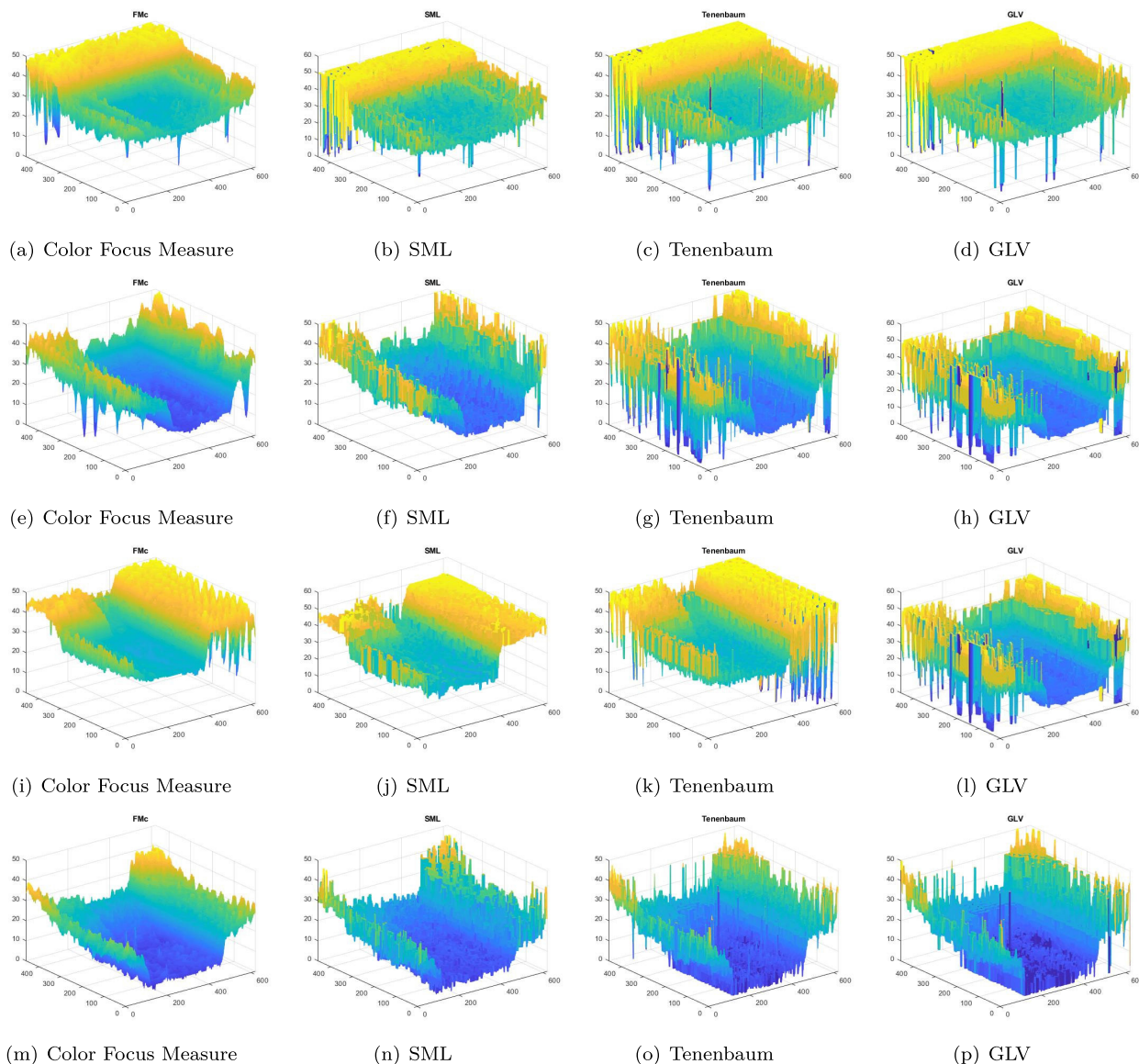
is applied on each channel of R, G and B separately, before averaging the resultant shapes. Similarly, Fig. 23 & 24, show the shape recovery of the Real Cone 1 with both scenarios.

Fig. 25 & Fig. 26 represent the shape recovery LCD-TFT Filter and Real Plane with  $FM_c$  and previous methods, respectively. It is evident from the figures that the shapes recovered using colored data-sets with  $FM_c$  are explicit whereas,



**FIGURE 28.** Shape reconstruction of different image sequence with proposed and previous techniques; Top to Bottom: Balls, Bucket, Fruits, Keyboard, Plants and Kitchen.





**FIGURE 29.** Shape reconstruction of (top to bottom) Logi1, Logi2, Logi3 and Logi4 image sequence with proposed and previous techniques.

the shapes are ambiguous when recovered with previous techniques using gray scale data-sets.

$FM_c$  was also compared with CONT, CURV, DCTE, HISE, SML, LAPM, TENG, GLVA, DCTR, SFRQ, GRAT and WAVS by measuring RMSE and correlation. Table 3 shows the RMSE and correlation for the shape reconstruction using the simulated cones and Real Cone 1 data-sets. It is clear from Table 3 that  $FM_c$  gives better results when compared with traditional FMs. The proposed  $FM_c$  yields the highest correlation and the lowest RMSE when compared with other FMs. The ground truth for the depth maps of the simulated cones and Real Cone 1 are provided in, [35], [52], [65]. The manuscript utilizes the same depth maps for the reference. Hence it can be concluded that the shapes recovered by other FMs, when using isoluminant colored (gray scale) data-sets are imprecise, therefore, they show very large

values for RMSE and very small (even negative) values for correlation.

Experiments were also performed on the third data-set, which includes Real Cone 2, Coin and Measuring Tape image sequences. Fig. 27(a-d) shows the reconstruction of Real Cone 2 using the proposed FM, SML, TENG, and GLVA FMs. Fig. 27(e-h) shows a reconstruction of the Measuring Tape results. The shape of the Measuring Tape is similar to that of a planer object with sides (in black) as background.

The proposed method not only successfully rejected the background, while preserving the planer property of the Measuring Tape, but also preserved the edges. On the other hand, other FMs were unable to do both of these together. Fig. 27(i-l) shows similar results. The Coin represents a cylindrical object. In the figure, only  $FM_c$  is able to suppress the background (black) while preserving the cylindrical shape of

the object, whereas, other FMs only preserved the edges and failed in other areas of shape reconstruction.

Fig. 28 represents the shape reconstruction of fourth data-set. Fig. 29 provides the shape construction of fifth data-set. Both figures clearly show that the shape reconstruction using proposed method (first column from left of figures) have better results than others.

Therefore, it can be concluded that the color information used by the proposed method increases the accuracy of shape reconstruction during image focus analysis. While existing focus measures cannot process color information, the proposed FM utilizes this feature and produces better results.

## VI. CONCLUSION

Measuring focus quality is an essential step in the Shape from Focus (SFF) process for depth estimation and 3D shape recovery. All previously proposed methods, use gray scale images to measure the degree of focus. Single-channel information is extracted from a three-channel color space while converting to gray scale, inevitably leading to information loss. In this article, an alternative focus measure is proposed to measure the degree of focus in color space. This color focus measure processes RGB images and hence color-critical information is preserved. The color differences of the neighbor pixels with respect to the central pixel are obtained and summed together, this is followed by calculating their spread. The sum and the spread are then combined to measure the degree of focus of the pixel in consideration.

The proposed focus measure  $FM_c$  was tested with nine colored data-sets (three Simulated Cones, Real Cone 1, Real Plane, LCD-TFT Filter, Real Cone 2, Coin, and Measuring Tape), and their 3D shapes were recovered. Previously proposed focus measures (such as the Sum of Modified Laplacian, Tenenbaum, and Gray Level Variance) were also used for depth estimation and shape recovery. For fair comparison the results are computed in two ways, (i) the colored data-sets were converted to gray scale as these FMs cannot process colored images, and (ii) the FMs were applied on each R, G, and B channels separately, before averaging the resultant shapes. The  $FM_c$  was applied on color image sequences only.

The results show better depth recovery with the  $FM_c$  as no information is lost, while for the other methods, inaccurate results were achieved. The proposed method also showed the highest correlation and lowest RMSE value when compared with previous methods when applied in both scenarios.

## APPENDIX

When working with  $RGB$  colors, color space  $\Omega \Xi \Psi$  in (13) can be replaced with  $RGB$ . Thus, (13) can be written as:

$$\begin{aligned} P_{i,j,k}^{RGB} &= (C_{i,j,k}^{RGB})^T \cdot \hat{c} \\ &= R(i, j, k)\hat{r} + G(i, j, k)\hat{g} + B(i, j, k)\hat{b} \\ &= [R(i, j, k) \ G(i, j, k) \ B(i, j, k)] \begin{bmatrix} \hat{r} \\ \hat{g} \\ \hat{b} \end{bmatrix}, \end{aligned} \quad (20)$$

where,  $0 \leq R \leq 255$ ,  $0 \leq G \leq 255$ ,  $0 \leq B \leq 255$  and  $\hat{c} = [\hat{r} \ \hat{g} \ \hat{b}]^T$  is the unit vector in  $RGB$  color space.

Similar results can be obtained by measuring the difference between the center pixel and the neighboring pixels, in the CIE  $L^*a^*b$  and HSV color spaces.

For CIE  $L^*a^*b$  space (13) is modified as:

$$\begin{aligned} P_{i,j,k}^{L^*a^*b} &= C_{i,j,k}^{L^*a^*b} \cdot \hat{c} \\ &= L(i, j, k)\hat{L} + a(i, j, k)\hat{a} + b(i, j, k)\hat{b} \\ &= [L(i, j, k) \ a(i, j, k) \ b(i, j, k)] \begin{bmatrix} \hat{L} \\ \hat{a} \\ \hat{b} \end{bmatrix}, \end{aligned} \quad (21)$$

where  $0 \leq a \leq 100$ ,  $0 \leq L \leq 256$ ,  $0 \leq b \leq 256$ , and  $\hat{c} = [\hat{L} \ \hat{a} \ \hat{b}]^T$  is the unit vector in  $L^*a^*b$  color space.

For HSV color space (13) can be written as:

$$\begin{aligned} P_{i,j,k}^{HSV} &= C_{i,j,k}^{HSV} \cdot \hat{c} \\ &= H(i, j, k)\hat{h} + S(i, j, k)\hat{s} + V(i, j, k)\hat{v} \\ &= [H(i, j, k) \ S(i, j, k) \ V(i, j, k)] \begin{bmatrix} \hat{h} \\ \hat{s} \\ \hat{v} \end{bmatrix}, \end{aligned} \quad (22)$$

where,  $0 \leq H \leq 180$ ,  $0 \leq S \leq 100$ ,  $0 \leq V \leq 100$  and  $\hat{c} = [\hat{h} \ \hat{s} \ \hat{v}]^T$  is the unit vector in  $HSV$  color space.

## ACKNOWLEDGMENT

(Husna Mutahira and Bilal Ahmad contributed equally to this work.)

## REFERENCES

- [1] A. Anand, I. Moon, and B. Javidi, "Automated disease identification with 3-D optical imaging: A medical diagnostic tool," *Proc. IEEE*, vol. 105, no. 5, pp. 924–946, May 2017.
- [2] M. Vázquez-Arellano, H. Griepentrog, D. Reiser, and D. Paraforos, "3-D imaging systems for agricultural applications—A review," *Sensors*, vol. 16, no. 5, p. 618, Apr. 2016.
- [3] H. S. Jang, M. S. Muhammad, and T. S. Choi, "Removal of jitter noise in 3D shape recovery from image focus by using Kalman filter," *Microsc. Res. Technique*, vol. 81, no. 2, pp. 207–213, Feb. 2018.
- [4] S.-O. Shim, M. T. Mahmood, and T.-S. Choi, "Searching surface orientation of microscopic objects for accurate 3D shape recovery," *Microsc. Res. Technique*, vol. 75, no. 5, pp. 561–565, May 2012.
- [5] M. Subbarao and T. Choi, "Accurate recovery of three-dimensional shape from image focus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 3, pp. 266–274, Mar. 1995.
- [6] S. K. Nayar, "Shape from focus," Dept. Elect. Comput. Eng., Robot. Inst., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. 15213, 1989.
- [7] S.-O. Shim and T.-S. Choi, "A novel iterative shape from focus algorithm based on combinatorial optimization," *Pattern Recognit.*, vol. 43, no. 10, pp. 3338–3347, 2010.
- [8] M. S. Muhammad and T.-S. Choi, "Sampling for shape from focus in optical microscopy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 564–573, Mar. 2012.
- [9] S. Pertuz, M. A. Garcia, D. Puig, and H. Arguello, "A closed-form focus profile model for conventional digital cameras," *Int. J. Comput. Vis.*, vol. 124, no. 3, pp. 273–286, Sep. 2017.
- [10] W. Burger and M. J. Burge, *Principles of Digital Image Processing*, vol. 111. Berlin, Germany: Springer, 2009.
- [11] A. A. Gooch, S. C. Olsen, J. Tumblin, and B. Gooch, "Color2Gray: Saliency-preserving color removal," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 634–639, 2005.
- [12] S. Pertuz, D. Puig, and M. A. Garcia, "Analysis of focus measure operators for shape-from-focus," *Pattern Recognit.*, vol. 46, no. 5, pp. 1415–1432, 2013.

- [13] A. Horii, *The Focusing Mechanism in the KTH Head Eye System*. Princeton, NJ, USA: Citeseer, 1992.
- [14] J. L. Pech-Pacheco, G. Cristóbal, J. Chamorro-Martínez, and J. Fernández-Valdivia, "Diatom autofocusing in brightfield microscopy: A comparative study," in *Proc. 15th Int. Conf. Pattern Recognit.*, vol. 3, Sep. 2000, pp. 314–317.
- [15] U. Ali and M. T. Mahmood, "3D shape recovery by aggregating 3D wavelet transform-based image focus volumes through 3D weighted least squares," *J. Math. Imag. Vis.*, vol. 62, no. 1, pp. 54–72, Jan. 2020.
- [16] H. Mutahira, M. S. Muhammad, M. Li, and D. Shin, "A simplified approach using deep neural network for fast and accurate shape from focus," *Microsc. Res. Technique*, vol. 84, no. 4, pp. 656–667, Apr. 2021.
- [17] T. Yan, Z. Hu, Y. Qian, Z. Qiao, and L. Zhang, "3D shape reconstruction from multifocus image fusion using a multidirectional modified Laplacian operator," *Pattern Recognit.*, vol. 98, Feb. 2020, Art. no. 107065.
- [18] M. B. Ahmad and T. S. Choi, "Application of three dimensional shape from image focus in LCD/TFT displays manufacturing," *IEEE Trans. Consum. Electron.*, vol. 53, no. 1, pp. 1–4, Feb. 2007.
- [19] Z. Ma, D. Kim, and Y.-G. Shin, "Shape-from-focus reconstruction using nonlocal matting Laplacian prior followed by MRF-based refinement," *Pattern Recognit.*, vol. 103, Jul. 2020, Art. no. 107302.
- [20] J.-T. Huang, C.-H. Shen, S.-M. Phoong, and H. Chen, "Robust measure of image focus in the wavelet domain," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, Dec. 2005, pp. 157–160.
- [21] H. Xie, W. Rong, and L. Sun, "Wavelet-based focus measure and 3-D surface reconstruction method for microscopy images," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2006, pp. 229–234.
- [22] R. Minhas, A. A. Mohammed, Q. M. J. Wu, and M. A. Sid-Ahmed, "3D shape from focus and depth map computation using steerable filters," in *Proc. Int. Conf. Image Anal. Recognit.* Berlin, Germany: Springer, 2009, pp. 573–583.
- [23] C.-Y. Wee and R. Paramesran, "Image sharpness measure using eigenvalues," in *Proc. 9th Int. Conf. Signal Process.*, Oct. 2008, pp. 840–843.
- [24] *Automatic Focus and Iris Control for Video Cameras*. Edison, NJ, USA: IET, 1995.
- [25] S.-Y. Lee, Y. Kumar, J.-M. Cho, S.-W. Lee, and S.-W. Kim, "Enhanced autofocus algorithm using robust focus measure and fuzzy reasoning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 9, pp. 1237–1246, Sep. 2008.
- [26] H. Nanda and R. Cutler, "Practical calibrations for a real-time digital omnidirectional camera," in *Proc. CVPR Tech. Sketch*, vol. 20, Dec. 2001, p. 2.
- [27] V. Hilsenstein, "Robust autofocusing for automated microscopy imaging of fluorescently labelled bacteria," in *Proc. Digit. Image Comput., Techn. Appl. (DICTA)*, Dec. 2005, p. 15.
- [28] S.-Y. Lee, J.-T. Yoo, Y. Kumar, and S.-W. Kim, "Reduced energy-ratio measure for robust autofocusing in digital camera," *IEEE Signal Process. Lett.*, vol. 16, no. 2, pp. 133–136, Feb. 2009.
- [29] W. Huang and Z. Jing, "Evaluation of focus measures in multi-focus image fusion," *Pattern Recognit. Lett.*, vol. 28, no. 4, pp. 493–500, 2007.
- [30] P. T. Yap and P. Raveendran, "Image focus measure based on Chebyshev moments," *IEE Proc. Vis., Image Signal Process.*, vol. 151, no. 2 pp. 128–136, Apr. 2004.
- [31] M. T. Mahmood and I. H. Lee, "Shape from focus based on 3D structure tensor using optical microscopy," *Microsc. Res. Technique*, vol. 83, no. 1, pp. 48–55, Jan. 2020.
- [32] M. Noguchi and S. K. Nayar, "Microscopic shape from focus using active illumination," in *Proc. 12th Int. Conf. Pattern Recognit.*, Oct. 1994, pp. 147–152.
- [33] Y. Xiong and S. A. Shafer, "Depth from focusing and defocusing," in *Proc. CVPR*, Jun. 1993, pp. 68–73.
- [34] J. Martinez-Baena, J. Fdez-Valdivia, and J. A. Garcia, "A multi-channel autofocusing scheme for gray-level shape scale detection," *Pattern Recognit.*, vol. 30, no. 10, pp. 1769–1786, Oct. 1997.
- [35] M. S. Muhammad and T. S. Choi, "A novel method for shape from focus in microscopy using Bezier surface approximation," *Microscopy Res. Techn.*, vol. 73, no. 2, pp. 140–151, Feb. 2010.
- [36] U. Ali, I. H. Lee, and M. T. Mahmood, "Guided image filtering in shape-from-focus: A comparative analysis," *Pattern Recognit.*, vol. 111, Mar. 2021, Art. no. 107670.
- [37] M. Moeller, M. Benning, C. Schonlieb, and D. Cremers, "Variational depth from focus reconstruction," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5369–5378, Dec. 2015.
- [38] A. S. Malik and T.-S. Choi, "A novel algorithm for estimation of depth map using image focus for 3D shape recovery in the presence of noise," *Pattern Recognit.*, vol. 41, no. 7, pp. 2200–2225, 2008.
- [39] G. Yang and B. J. Nelson, "Wavelet-based autofocusing and unsupervised segmentation of microscopic images," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, vol. 3, Oct. 2003, pp. 2143–2148.
- [40] A. S. Malik and T.-S. Choi, "Consideration of illumination effects and optimization of window size for accurate calculation of depth map for 3D shape recovery," *Pattern Recognit.*, vol. 40, no. 1, pp. 154–170, Jan. 2007.
- [41] M. S. Muhammad, H. Mutahira, K.-W. Choi, W.-Y. Kim, and Y. Ayaz, "Calculating accurate window size for shape-from-focus," in *Proc. Int. Conf. Inf. Sci. Appl. (ICISA)*, May 2014, pp. 1–4.
- [42] J. Gladines, S. Sels, J. Blom, and S. Vanlanduit, "A fast shape-from-focus-based surface topography measurement method," *Sensors*, vol. 21, no. 8, p. 2574, Apr. 2021.
- [43] R. A. Flynn, E. Fleet, G. Beadie, and J. S. Shirk, "Achromatic grin singlet lens design," *Opt. Exp.*, vol. 21, no. 4, pp. 4970–4978, 2013.
- [44] M. E. Helou, F. Dumbgen, and S. Siisstrunk, "AAM: AN assessment metric of axial chromatic aberration," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 2486–2490.
- [45] F. A. Jenkins and H. E. White, "Fundamentals of optics," *Amer. J. Phys.*, vol. 26, no. 4, p. 272, 1958.
- [46] P. R. Sanz, B. R. Mezcuca, and J. M. S. Pena, "Depth estimation-an introduction," in *Current Advancements in Stereo Vision*. London, U.K.: IntechOpen, 2012.
- [47] M. Ra, M. S. Muhammad, C. Lim, S. Han, C. Jung, and W.-Y. Kim, "Smartphone-based point-of-care urinalysis under variable illumination," *IEEE J. Transl. Eng. Health Med.*, vol. 6, pp. 1–11, 2018.
- [48] B. Cai, X. Xu, and X. Xing, "Perception preserving decolorization," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 2810–2814.
- [49] A. Ione, "A scientist's vision of art: A review of margaret Livingstone's vision and art: The biology of seeing [review of the book vision and art: The biology of seeing, by M. Livingstone]," *Psyche, Interdiscipl. J. Res. Consciousness*, vol. 9, no. 9, 2003. [Online]. Available: <https://psycnet.apa.org/record/2004-17718-001>
- [50] D.-C. Tsai and H. H. Chen, "Focus profile modeling," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 818–828, Feb. 2016.
- [51] M.-C. Lu, "Computer modeling and simulation techniques for computer vision problems," Ph.D. dissertation, Dept. Elect. Eng., State Univ. New York Stony Brook, New York, NY, USA, 1993. [Online]. Available: <http://www.ece.sunysb.edu/~murali/>
- [52] M. T. Mahmood and T.-S. Choi, "Nonlinear approach for enhancement of image focus volume in shape from focus," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2866–2873, May 2012.
- [53] S. Pertuz. *Computer Vision and Image Analysis*. Accessed: Jun. 4, 2020. [Online]. Available: <https://sites.google.com/view/cvia/home>
- [54] R. C. Contreras, O. Morandin, Jr., and M. S. Viana, "A new local search adaptive genetic algorithm for the pseudo-coloring problem," in *Proc. Int. Conf. Swarm Intell.* Cham, Switzerland: Springer, 2020, pp. 349–361.
- [55] K. Sharma, K. S. Rajpal, and R. Kant, "Study of X-ray image via pseudo coloring algorithm," *Int. J. Adv. Res., Ideas Innov. Technol.*, vol. 5, no. 1, pp. 316–319, Feb. 2019.
- [56] J. Al, H. Alhatamleh, Z. A. Alqadi, and M. Khalil, "Creating a color map to be used to convert a gray image to color image," *Int. J. Comput. Appl.*, vol. 153, no. 2, pp. 31–34, Nov. 2016.
- [57] R. Minhas, A. A. Mohammed, and Q. M. J. Wu, "An efficient algorithm for focus measure computation in constant time," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 1, pp. 152–156, Jan. 2012.
- [58] S.-O. Shim, W. Aziz, A. Banjar, A. Alamri, and M. Alqarni, "Improving depth computation from robust focus approximation," *IEEE Access*, vol. 7, pp. 20144–20149, 2019.
- [59] R. Minhas, A. A. Mohammed, and Q. M. J. Wu, "Shape from focus using fast discrete curvelet transform," *Pattern Recognit.*, vol. 44, no. 4, pp. 839–853, Apr. 2011.
- [60] PureMoCo. *Focus-zoom-unit (motorized follow focus) 1:15–2:23*. Youtube. Accessed: Jun. 11, 2020. [Online]. Available: <https://www.youtube.com/watch?v=KFryXjYbTJc>
- [61] S. Suwajanakorn, C. Hernandez, and S. M. Seitz, "Depth from focus with your mobile phone," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3497–3506.
- [62] S. or Aek. Accessed: Jun. 12, 2020. [Online]. Available: <https://www.supasorn.com/>

- [63] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Trans. Commun.*, vol. 43, no. 12, pp. 2959–2965, Dec. 1995.
- [64] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 372–387, Jan. 2016.
- [65] M. S. Muhammad, H. Mutahira, and T.-S. Choi, "Using a low pass filter to recover three-dimensional shape from focus in the presence of noise," *Int. J. Innov. Comput., Inf. Control*, vol. 8, no. 4, pp. 2777–2788, 2012.



interests include 3D shape recovery and vision-based robot navigation.

**HUSNA MUTAHIRA** (Student Member, IEEE) received the B.S. degree in computer science from Fatima Jinnah Women University, Rawalpindi, Pakistan, in 2007, and the M.S. degree in information and mechatronics engineering from Gwangju Institute of Science and Technology, Gwangju, South Korea, in 2010. She is currently pursuing the Ph.D. degree with the College of Information and Communication Engineering, Sungkyunkwan University, Suwon, South Korea. Her research



**BILAL AHMAD** received the B.S. degree in engineering sciences from the GIK Institute of Science and Technology, Pakistan, in 2014, and the M.S. degree in electrical and computer engineering from Sungkyunkwan University, South Korea, in 2019. He is currently pursuing the Ph.D. degree with the Department of Computer Science, Norwegian University of Science and Technology (NTNU), Norway. His research interests include 3D shape recovery and image processing.



include robot vision, image processing, robot navigation, and robots in health care.

**MANNAN SAEED MUHAMMAD** (Senior Member, IEEE) received the B.S. degree in mechatronics engineering from the National University of Science and Technology, Rawalpindi, Pakistan, in 2003, and the M.S. and Ph.D. degrees in mechatronics engineering from Gwangju Institute of Science and Technology, Gwangju, South Korea, in 2006 and 2012, respectively. He is currently an Assistant Professor with Sungkyunkwan University, Suwon, South Korea. His research interests



include artificial intelligence, neural networks, wireless communication and networks, ubiquitous computing and networking, sensor networks, and big data.

**DONG RYEOL SHIN** received the B.S. degree in electrical and electronic engineering from Sungkyunkwan University, in 1980, the M.S. degree in electrical and electronic engineering from Korea Advanced Institute of Science and Technology, in 1982, and the Ph.D. degree from Georgia Institute of Technology, USA, in 1992. From 1992 to 1994, he was with Samsung Data Systems, South Korea. Since 1994, he has been with Sungkyunkwan University, where he is currently a Full Professor. His research interests include artificial intelligence,

...