International Journal of Computer Aided Manufacturing
Vol. 1: Issue 2
www.journalspub.com

Performance Analysis of Gradient-Based Focus Measures in a Parallax Affected SFF Scenario

R. Senthilnathan[1], P. Subhasree[2], R. Sivaramakrishnan[3]*
[1]Department of Mechatronics Engineering, SRM University, Kattankulathur (Tamil Nadu), India
[2]Department of Computer Science Engineering, Sri Venkateswara College of Engineering, Sriperumbadur (Tamil Nadu), India
[3]Department of Production Technology, MIT Campus, Anna University, Chennai (Tamil Nadu), India

Abstract

For scene reconstruction, shape from focus (SFF) is a popular technique in the field of computer vision. This is the fact that the SFF technique is based the focus levels of the pixels of the image preserves depth information. Conventionally, when there is a relative motion between the camera and the scene, SFF is implemented by using telecentric lenses, which avoid parallax effects. This becomes the chief component for the limitation of the SFF technique to very small objects generally in the millimeter scale. In the current research work, a new SFF-inspired algorithm is developed that uses in a wide angle lens in place of a telecentric lens. This extends the range of object that the system can deal with, though severe magnification changes occur when a stack of images are acquired with respect to the scene. By using a variable window approach the problem is addressed, when focus measures are computed. The paper presents significant results of performance evaluation of five different focus measures based on the first order image derivative, the image gradient. The evaluation is carried out based on two different performance evaluation criteria namely root mean square error and computation time. Under various operating conditions such as different spatial resolution, window size, contrast changes, gray level saturation and camera noise, the analysis of the gradient-based measures are carried out.

Keywords: shape from focus, parallax, variable window size, focus measures, image gradient

****Corresponding Author***
E-mail: senthilnathan.r@ktr.srmuniv.ac.in

INTRODUCTION

In many fields, computer vision has been increasingly found its application, such as automated inspection, robot guidance, entertainment, and other scientific and industrial applications. Over three decades in the field of 3-D computer vision, scene reconstruction has been the topic of interest. Many techniques viz., passive, active and hybrid techniques have evolved over the years. In passive methods, 3-D reconstruction aims at estimating the 3-D surface geometry with one or more passive cameras that record the intensity information of the scene, while active vision technique reconstruct a scene by estimating the depth by deliberately releasing some form of energy into the scene which was unless not present in the scene. Hybrid methods, like the structured light technique, combine the advantages of active and passive vision techniques by throwing additional light (patterns) on the scene that is imaged by a passive camera. Generally, the techniques are grouped together under the name shape-from-X.

Although X denotes the cue used for the scene reconstruction that could be stereo, motion, shading, focus, defocus, texture etc. Apart from these methods, shape from focus (SFF) [1] and shape from defocus (SFD) used multiple images of the scene taken with different focuses. The difference between the methods come in the form of SFD,[2] generally, requires one sharp image in the foreground and background, respectively. The distance of all the points that lie between foreground and background is interpolated by a sharpness measure (actually measures the degree of unsharpness). Unlike the SFD method, SFF usaully requires more number of images along the dissimilar focal stacks of the lens.

Generally it makes that the SFF method is computationally expensive as compared to SFD, although this drawback pays in a way that the precision of SFF is better than SFD. The SFF techniques have been successfully used in various areas, such as industrial inspection, medical diagnostics involving microscopic imaging or imaging of small objects.[3] The SFF techniques demanded images of different scene with different focus levels that can obtain by translating the camera or the object or by changing the focus setting of the lens. SFF is highly sensitive to parallax, and this is one of the inherent limitations of the SFF method of reconstruction of the scene. In the recent years by means of novel image processing procedures, many authors have attempted to extend the applicability of SFF techniques.[4]

The attempts that is made for increasing the applicability of the SFF technique is purely for the complete scene reconstruction rather than using some approximate information from SFF with an imaging system prone to parallax errors. In this study, a small part of the research work that deals with the development of an algorithm inspired by conventional SFF, which may be used for the scenario that involved parallax. The reported study involved the evaluation of various focus measures based on image gradient.

IMAGE FORMATION MODEL

All types of 3-D applications, it may be 3-D pose estimation or reconstruction, rely heavily on the geometry of the imaging process. In fact image formation process dictates the information that could be extracted from one or more images. The process of image formation is itself a projective geometry engine.

In SFF, by measuring the distance of well-focused position of every object point from the camera lens, the objective is to find out the depth. The 3-D geometry of the scene can be recovered, when distances for all the points are known.

Blur Model

In some fundamental characteristics of image formation and focusing, a simple camera model consisting of thin lens and an image plane can be used to derive. A defocused pixel is surrounded by a blur circle whose diameter depends on the lens setting with respect to the scene. According to geometric optics, the intensity within the blur circle is approximately constant. However owing to aberrations, diffractions, and other undesirable effects, a two-dimensional Gaussian model has been suggested as an alternative model:

$$h(x, y) = \frac{1}{2\pi\sigma^2} e^{\frac{-1}{2}\frac{x^2+y^2}{\sigma^2}} \quad \text{Eq. (1)}$$

where σ is the spread parameter such that $\sigma = k.d$ for $k > 0$. Here, k is the proportionality constant that is a characteristic of the camera used for imaging.

The value of k can be found from calibrating the camera. The blur model is often referred as point spread function (PSF). Figure 1 illustrates the PSF behavior which varies as two-dimensional

Gaussian along the different focal planes in the focal stack.

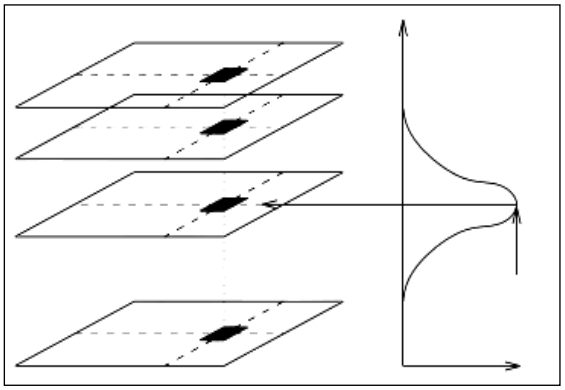


Fig. 1: *Two-Dimensional Gaussian Focus Model.*

SHAPE FROM FOCUS

In Figure 2, a schematic diagram illustrating the methodology of the SFF is shown.[4] Initially, a 3-D object was kept in the reference plane and translated along the optical axis, in a fixed and finite number of discrete steps denoted by Δd. This is a disadvantage in SFF, as the method inherently suffers from a finite determination error contributed by Δd. The optics of the camera defines a focused plane, such that all the points on it are in their best sharpness. The distance of the focused plane from the camera was ω_d.

The reference image was at a distance of d_r away from the focused plane. The distance between the translating stage and the reference plane was d_k. All the distances were known before making any depth estimate from the system. As the stage moves towards the camera, a point (x, y) on the 3-D objective increasingly comes into focus, and the best perceptiveness is attained when the distance between the reference plane and translating stage is $d_k = {}^{C}z$.

SFF basically attempts to estimate this ${}^{C}z$ for every pixel in the image. In order to achieve this, for every image in the arrangement, the focus level of every pixel is measured in a local window. The focus measure is essentially a mathematical operator, which measures the high spatial occurrence contented in the image.

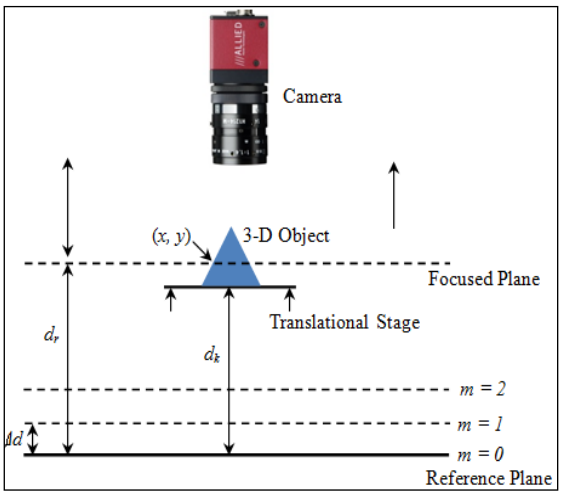


Fig. 2: *Schematic of Shape from Focus.*

At every pixel, the image frame that gives the maximum measurement of focus is identified. Accordingto the knowledge of the camera parameters, and a identified focus function, the distance of the sight point regarding to each pixel may be computed. Other features for error in the SFF comprise edge bleeding, sensor noise and loss of window registration [4]. Besides the given errors, the SFF has a few different limitations. In the case of visually standardized scenes, the depth estimate is not possible directly, since the spatial frequency content is very less for a standardized texture. This may be observed as an analogy to the stereo visualization, in the sense that rich texture content is a must to identify the corresponding points in stereo visualization.

Besides these limitations, the SFF is also a comparatively time-taking process, mainly contributed by the basic need for the acquisition of a sequence of images. Although more number of images will decrease the determination error, a minimum of 20 images is a must to have any reasonable estimate. [5]

Experimental Setup

The photograph of the experimental setup is shown in Figure 3.

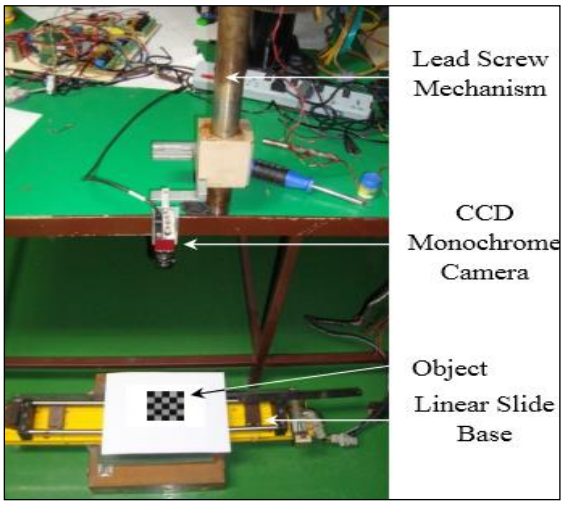


Fig. 3. *Experimental setup.*

The images were taken on the go and hence a simple DC motor was used for the purpose. The various specifications of the imaging system are shown in Table 1.

Table 1: *System Specifications.*

Parameter	Specification
Light source	White LED area light
Illuminance	160 lux
Lighting Technique	Partially diffused bright field incident lighting
Lens type	Fixed focal length prime lens
Focal length & f#	16 mm, 1.3
Camera make and Model interface	Allied Vision Technologies Guppy F033b IEEE 1394a – 400 Mb/s, 1 port
Computer Interface	PCI – IEEE 1394a
Resolution	640×480
Aspect ratio	4:3
Sensor	Sony ICX424
Sensor type	CCD progressive, Monochrome
Operating frame Rate	45 frames per second (fps)
Mode of operation	Mono8 mode
Trigger type	Software trigger
Image acquisition Time	180 ms per image
Processor	Intel Core i5, 2.5 GHz Quad Core
Memory	4 GB, 1300 MHz primary memory

Focus Measure

The focus measure is essentially a mathematical function that gives a measure of the concentration of the image indirectly, by calculating the contrast of the image. It is usaully computed in a small square window nearby the pixels in the image. A high value for the focus measure directs a severely focused area in the image, and a low value indicates indistinct regions.

Gradient-Based Focus Measures

The first derivative of the image is termed as the gradient of the image as it denotes the difference in the grey level in the image. The algorithms presented under this category follow the assumption, that the local difference would be large for the focused regions in an image as compared to a defocused region in an image.

The Gaussian derivative was proposed by Geusebroek *et al*. [6] for autofocus in microscopy, based on the first order Gaussian derivative given in the following equation:

$$F_{GD} = \sum_{(x,y)} (I * \Gamma_x)^2 + (I * \Gamma_y)^2 \qquad \text{Eq. (2)}$$

Where, Γ_x and Γ_y are the partial derivatives of the Gaussian function $\Gamma(x, y, \sigma)$ along the *x* and *y* dimensions of the image, respectively. F_{GD} is the focus measure for a pixel $I(x, y)$ computed in a neighborhood Ω. The Gaussian function $\Gamma(x, y, \sigma)$ is given by,

$$\Gamma(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \qquad \text{Eq. (3)}$$

The value of σ is chosen, such that for a neighborhood of size $W \times W$, a total of five σs are contained along *W*. The first derivative of the image in the horizontal direction is a simple measure of its degree of focus, generally expressed as a measure, called thresholder gradient given by,

$$F_{TG} = \sum_{(i,j) \in \Omega(x,y)} |I(i, j+1) - I(i, j)| \geq T \qquad \text{Eq. (4)}$$

The performance of this measure depends on the value used for *T*. For the sake of

generality, no threshold has been considered in the current study. Many alternative definitions for this focus measure are found in the literature, which considers both the horizontal and vertical image derivatives either by addition [7] or the selection of the maximum value. [8]

The first derivative is squared in order to increase the influence of the larger value of the image gradient. This expression is used as a focus measure, which is known as the squared gradient [9] as given below:

$$F_{SG} = \sum_{(i,j)\in\Omega(x,y)} |I(i,j+1) - I(i,j)|^2 \geq T \quad \text{Eq. (5)}$$

A popular focus measure based on the gradient of the image is presented in many computer vision literature related to SFF, called the Tenengrad measure[10]. The measure is defined as,

$$F_T = \sum_{(i,j)\in\Omega(x,y)} G_x(i,j)^2 + G_y(i,j)^2 \quad \text{Eq. (6)}$$

Where, G_x and G_y are the image gradients along the x and y dimensions, respectively. The gradients are computed by convolving the image I with the Sobel operator often used in edge detection. The variance of the image gradient may be used as a focus measure, termed as Tenengrad variance[11]. This measure is popularly used for autofocus, though it may even be applied to SFF. The measure is defined as,

$$F_{TV} = \sum_{(i,j)\in\Omega(x,y)} (G(i,j) - \bar{G})^2 \quad \text{Eq. (7)}$$

Where, $\bar{G}$ is the mean value of the gradient given by $\bar{G} = \sqrt{G_x^2 + G_y^2}$ within the neighborhood Ω.

PROPOSED SFF-INSPIRED ALGORITHM

The following part of the paper presents the planned SFF-inspired algorithm, which procedures the central theme of the thesis. First a set of feature points present across the stack of images was detected using SURF feature detector. The stack suffered from combined variations in magnification and focus due to the related motion between the scene and the camera. The focus measurement of only those specific pixels was computed. This is different from the straight SFF method, in which the focus measurement is computed of all the pixels in all the images in the focal stack. Since in the current study, there is a finite SDPM, focus measures cannot be directly applied to all the images in a straight manner.

Conventional SFF uses a focus function, such as a Gaussian distribution, and incorporates the computed focus measures to find correct depth estimates. Such a model is suitable only when a tele-centric lens is used. This is true, since in the conventional SFF the depth of field is very limited, and no magnification changes happen because of the relatited motion between the scene and the camera.

In this current study, since a wide angle lens is used with a higher DOF, in order to complete a whole Gaussian distribution, large camera motion would be essential. Particularly low magnification bases the spatial resolution of the image to convert too poor for any measurements probable from the images. Due to these reasons, a coarse method of depth estimation was assumed for this study. The algorithm may be summarized as follows:

1. The initial location of the camera from the measure plane is identified a priori as s_m, where, m = 1 for the initial location of the camera.
2. Accumulate the image sequences acquired at each step m where the stand-off distance (s_m) increases in steps of Δd.
3. Measure focus, F_m for each of the SURF feature points, across the stack

at each step whose correspondences are matched using the SSD metric.
4. Find the step number *m* where the focus measurement is the maximum of a point *(x, y)*, such that $F_m = F_{max}$, where F_{max} is the maximum value of the focus measure for a particular pixel.
5. To assign the value of the distance of the camera signal as the height of the objective relating to the particular pixel, such that the height of the scene point $\bar{h} = m\,\Delta d$.
6. Once the height of a point is computed, the depth of the point ${}^{C}z$ may be computed as ${}^{C}z = s_1 - \bar{h}$.

This algorithm, as may be perceived, gives only a rough estimation of the depth. The performance of the algorithm is straight dependent on the collection of Δd. Lower values of Δd give better correctness, although there is always a nonzero determination error.

Interestingly, at times the estimated depth becomes equivalent to the real depth, dependent on the specific scene point in consideration. Apart from this, the depth error may be zero, whereas the structure as a whole suffers from a nonzero resolution error. This behavior may be observed in the results of the evaluation presented in the next section of the chapter.

Window Size

The size of the window about which the focus measurement is computed, is a dynamic parameter in the SFF method. Usually, the window size must be as small as probable to obtain accurate results. When the size of the window is large, a large neighborhood is included to compute the focus.

If the deepness of the scene relating to different points in the window differs, it may direct to averaging of dissimilar focus levels produced by dissimilar depths of the scene points. Many authors have suggested a smaller-sized window, particularly a 5×5 window, to be optimal. Window sizes lower than 5×5 may result in errors caused by random noises in grey levels. Larger mask sizes would result in averaging errors. In the current study, since the images suffered from magnification changes, a variable window size approach was developed. According to this method, the window size applied to a particular frame was scaled by the magnification factor corresponding to that frame. This means that in the current scenario, the window sizes would be reducing, starting from the first frame to the 15th image in the focal stack. Larger window sizes offer better results, but lead to averaging errors.

It is justified, since the current study which uses the SFF-inspired algorithm only to get a coarse and sparse depth estimation. This issue may be considered as a drawback of the proposed method, as it inherently suffers from slightly higher averaging errors as compared to the conventional SFF.

EVALUATION OF FOCUS MEASURES

A typical focus measure should satisfy these requirements. [12]

1) Independent of the image content;
2) Monotonic with respect to blur;
3) The focus measure must be unimodal, i.e., it must have one and only one maximum value;
4) Large variation in value with respect to the degree of blurring;
5) Minimal computation complexity;
6) Robust to noise.

Operating Conditions

The focus measures were evaluated under different operating conditions namely, different spatial resolutions, camera noise, gray level saturation and contrast. Figure 4 shows the stacks of three different spatial resolutions and the first image in each stack. The distances between the object

and the measurement plane in the three cases were 30, 50 and 84 mm, respectively. The three cases were indicated as spatial resolution 1, 2 and 3, respectively.

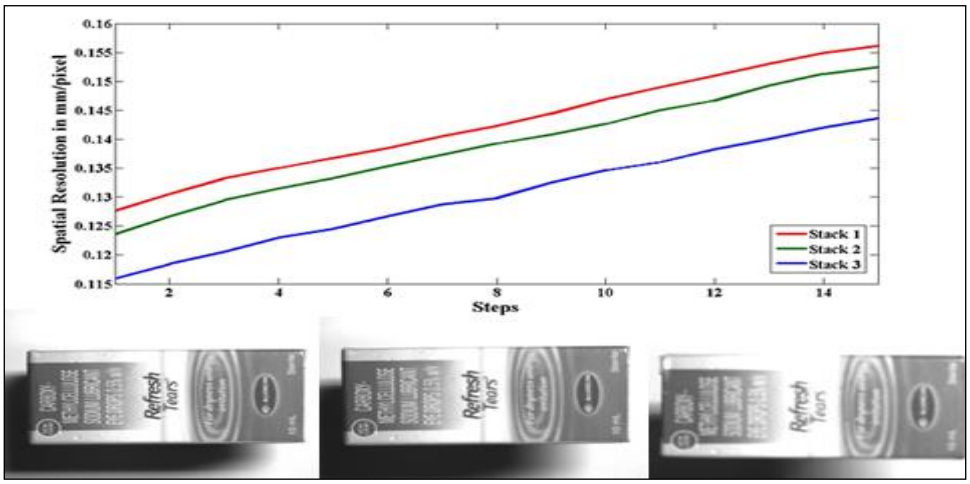


Fig. 4: *Different Spatial Resolutions Considered for Evaluation.*

Out of various camera noises that may corrupt an image acquired from a CCD sensor, the significant noise sources may be grouped as irradiance-dependent and irradiance-independent sources. [12] The noises may be modelled as follows:

$$I_{noise} = F(I + n_s + n_c) + n_q \quad \text{Eq. (8)}$$

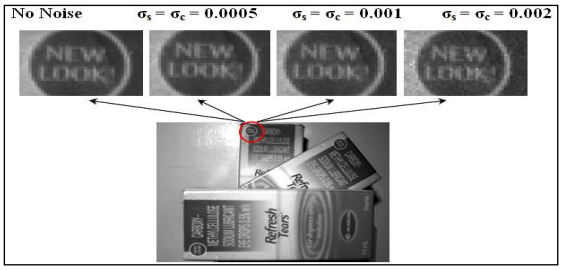


Fig. 5: *Different Noise Levels.*

Where, I_{noise} is the image that is obtained after adding the noise components to the original image I. The parameter F in the above equation is the camera response function, n_s is the irradiance-dependent noise component, n_c is the irradiance-independent noise component, and n_q is the quantization and amplification noise. The noise components are basically Gaussian white noise with zero mean, and the variances for n_s and n_c are $Var\ (n_s) = I \cdot \sigma_s^2$ and $Var\ (n_c) = \sigma_c^2$, respectively. The feature detectors and match metrics were evaluated for three different levels of noise, namely, for $\sigma_s = \sigma_c = 0.0005$, $\sigma_s = \sigma_c = 0.001$ and $\sigma_s = \sigma_c = 0.002$. Figure 5 shows a magnified view of a small portion in the image (highlighted in red color) which was subjected to different levels of noise corruption.

Image contrast is an important factor that affects the performance of the algorithms, related to feature detection and matching. In the current study, the contrast of the original images was reduced to different levels to analyze the robustness of the detectors in such a scenario. The contrasts of the images were reduced by compressing the histogram of the respective images. Three different set of points were chosen, to achieve 13%, 26% and 39% contrast reduction, with reference to the original images. Figure 6 shows the first image of the focal stack for different contrast levels and their corresponding grey level histogram.

In this research, image saturation has been evaluated by adding a constant offset to the original image, as presented in the equation below:

$I_{sat} = I + S$ Eq. (9)

Where, I_{sat} is the saturated image obtained by adding a constant offset S to the original image *I*. In the current study, three different offsets, namely, 25, 51, 77 were added to the images, which may be considered as 10%, 20% and 30% saturation, respectively, for an 8-bit dynamic range. Figure 7 shows the sample images subjected to various levels of saturation considered for evaluation.

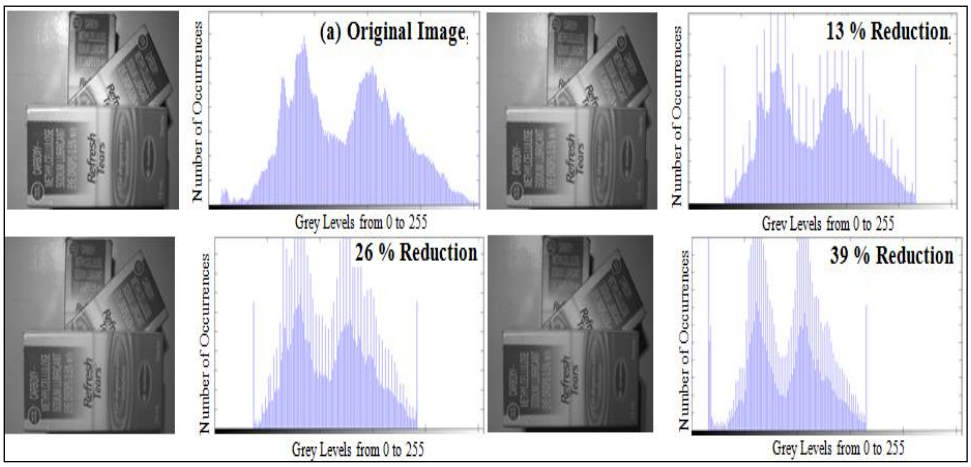


Fig. 6: *Contrast Reduction.*

Fig. 7: *Grey Level Saturation.*

Evaluation Criteria

In the current research, two criteria were used for the evaluation of the focus measures under different operating conditions, namely, the execution time of the focus measures, and the Root Mean Square Error (RMSE). The RMSE is normalized by the number of pixels in the image and the number of steps of image acquisition. The normalized RMSE is defined as follows:

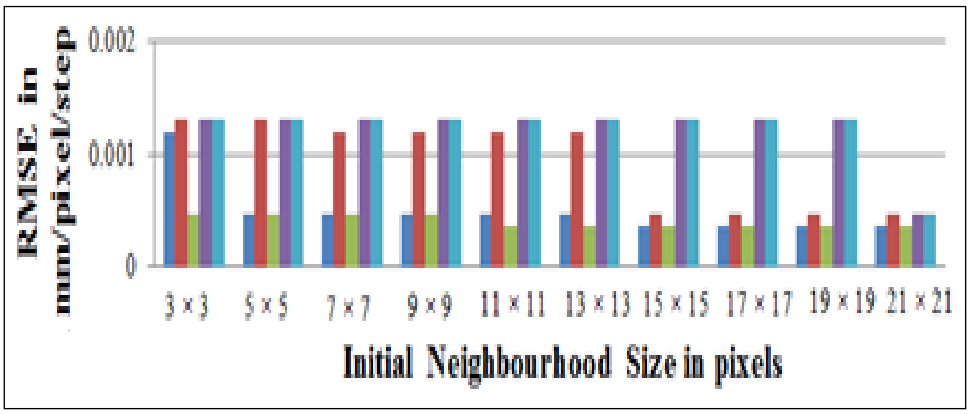
$$RMSE = \frac{\sqrt{\frac{1}{MN}\sum_{(i,j)}(G_T(i,j) - Z(i,j))^2}}{No.of\ Steps} \quad \text{Eq. (10)}$$

In the above equation, *N* and *M* are the number of pixels in the horizontal and vertical dimensions of the image. G_T *(i, j)* is the ground truth information about the actual depth of a point *(i, j)* in the scene, and *Z (i, j)* is the estimated depth of the scene obtained from a particular focus measure at that point. The values of G_T *(i, j)* are found, based on the physical measurements of the object's dimensions, with an uncertainty of 1 mm. The execution time was calculated by software means in MATLAB, which gives an approximate estimate of the time taken for the execution of a set of functions. In order to reduce error, the execution time was

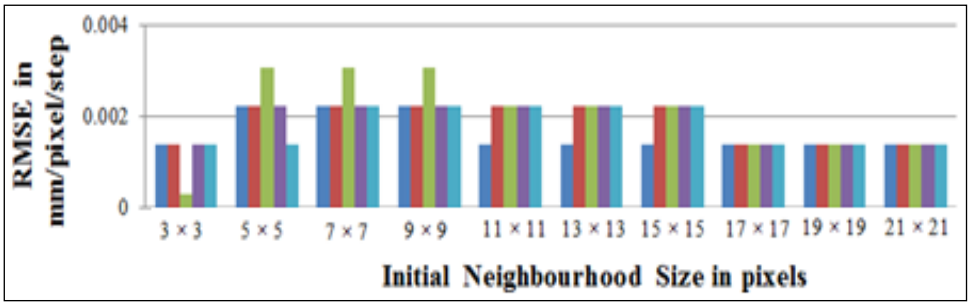
averaged from 20 trials for the same function's execution, and also all other application software was prevented from running.

RESULTS AND DISCUSSION

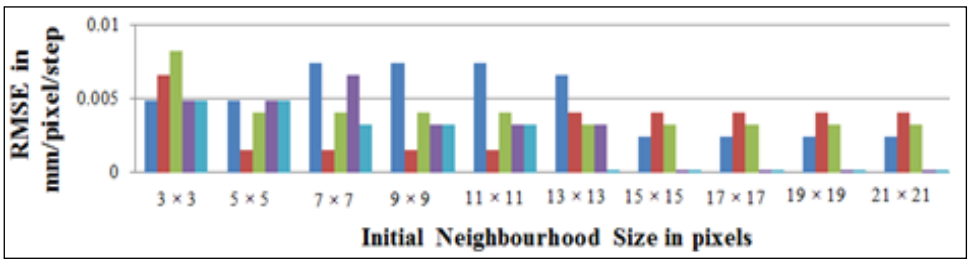
The following figures show the results of RMSE for various operating conditions as shown in Figure 8.



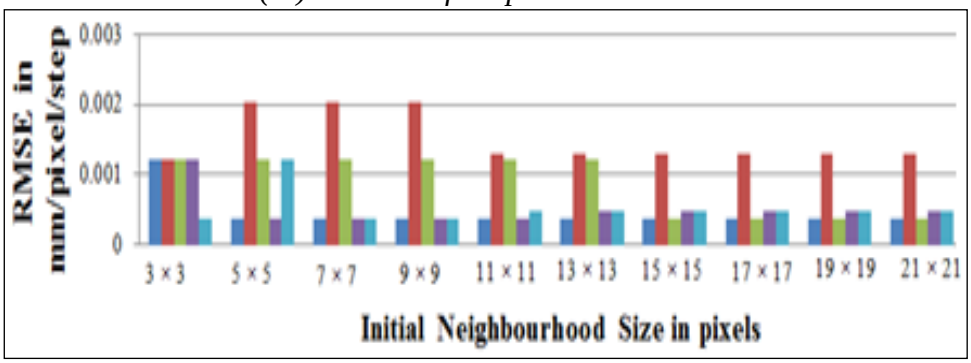
(i) RMSE for Spatial Resolution 1.



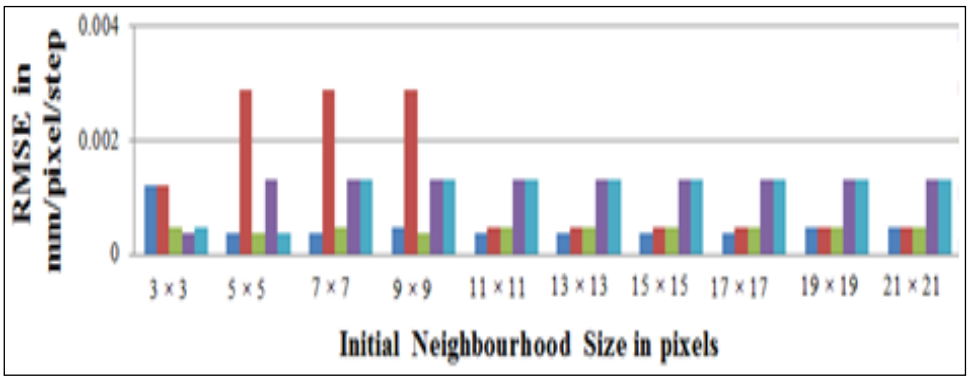
(ii) RMSE for Spatial Resolution 2.



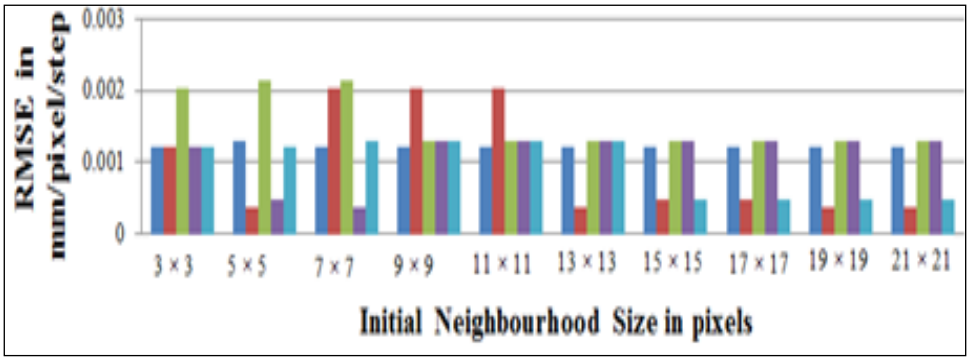
(iii) RMSE for Spatial Resolution 3.



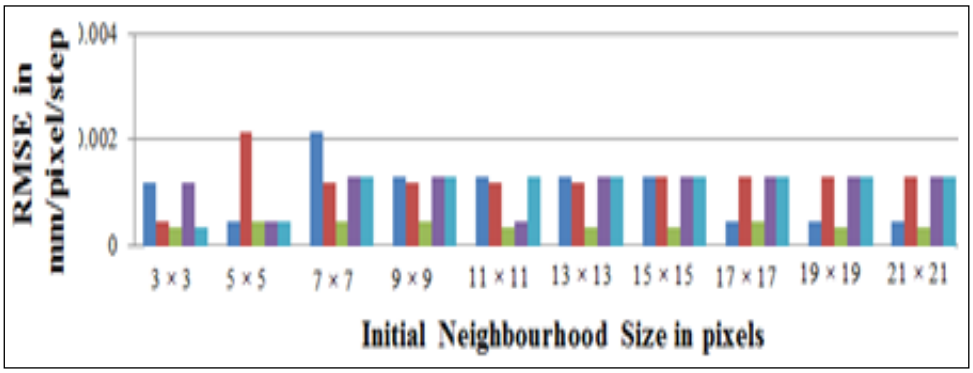
(iv) RMSE for Camera Noise Level 1.



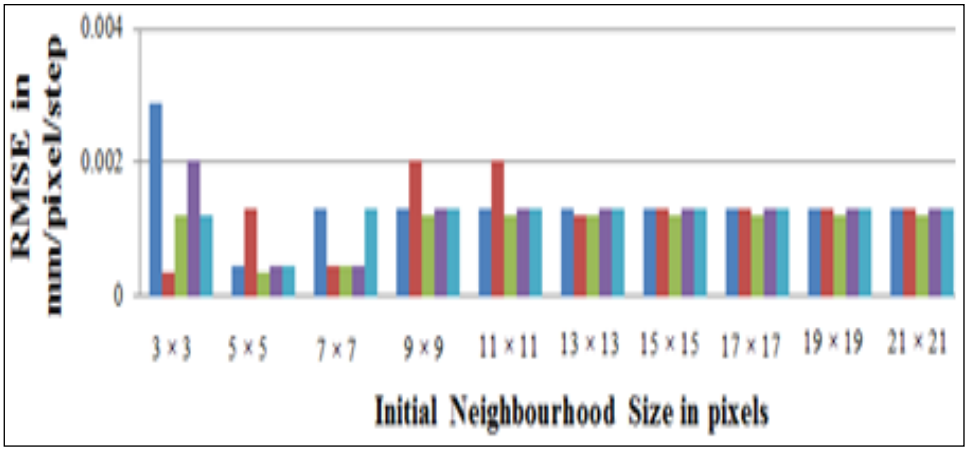
(iv) RMSE for Camera Noise Level 2.



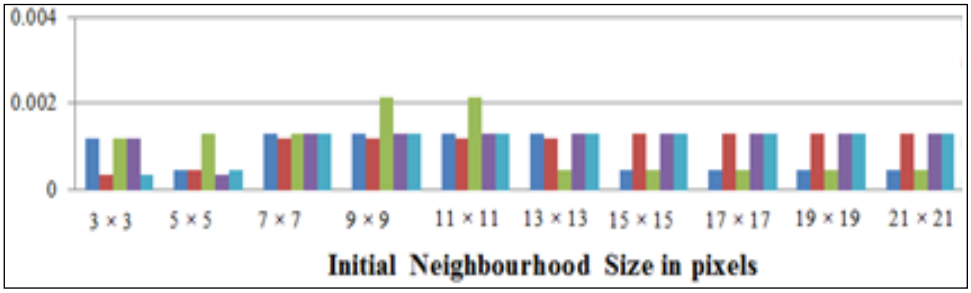
(v) RMSE for Camera Noise Level 3.



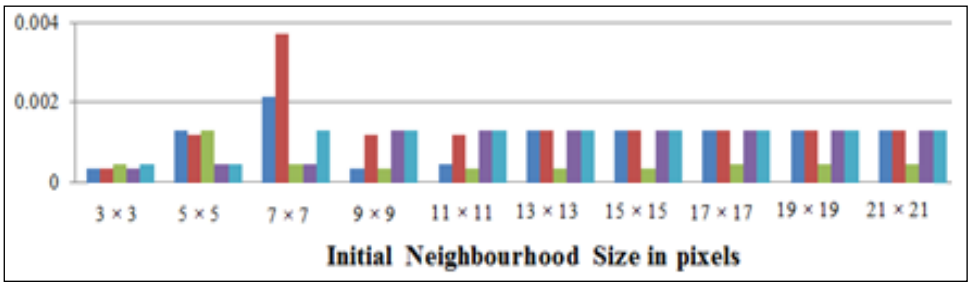
(vi) RMSE for 13% Contrast Reduction.



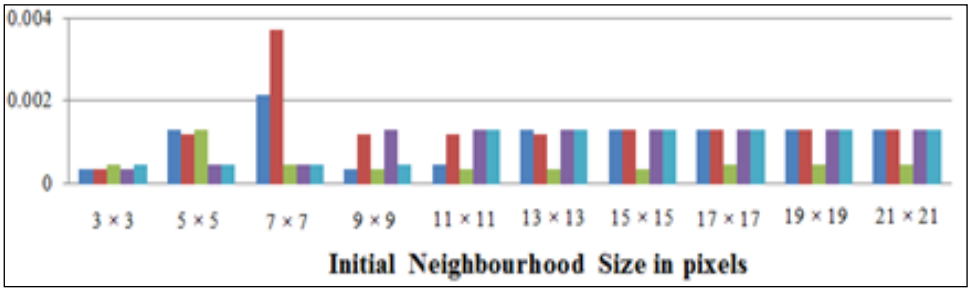
(vii) RMSE for 26% Contrast Reduction.



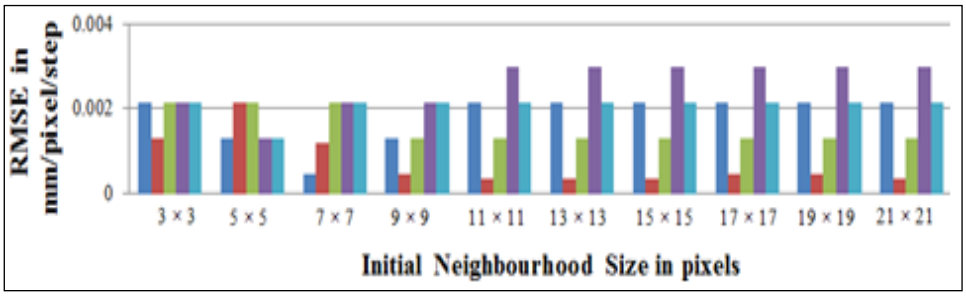
(viii) RMSE for 39% Contrast Reduction.



(ix) RMSE for 10% Saturation.



(x) RMSE for 20% Saturation.



(xii) RMSE for 30% Saturation.

Fig. 8: *Results of RMSE for Various Operating Conditions.*

Table 2 shows the average execution time for the various focus measures for a single point computed around a window, whose initial size is 21×21. Since, the execution time for a larger window takes more time, as compared to a smaller window; hence, evaluating in the worst case scenario.

Table 2: *Average Computation Time.*

Focus Measure	Average Computation Time (msec)
Gaussian derivative	1.885
Threshold gradient	0.206
Squared gradient	0.192
Tenengrad	1.355
Tenengrad variance	1.422

The RMSE increases with increase in contrast reduction and saturation. The Gaussian derivative and squared gradient gave the least RMSE for most of the cases in the category of gradient-based measures. Higher window sizes returned lower RMSE. This is mainly due to the larger neighborhood of focus measure which makes the measurement robust (at the cost of higher computational time) as compared to the noise measurements obtained from smaller masks. Selection of a unique size is subjected to the trade off and a window size of 15×15 is chosen to be the candidate window size for all the focus measurement purpose. The results presented in this paper may be used to select the right focus measure for the proposed SFF-inspired algorithm.

REFERENCES

1. Nayar S.K., Nakagawa Y. Shape from Focus: An Effective Approach for Rough Surfaces. *Proceedings of the IEEE International Conference on Robotics and Automation CRA90*; 1990 May 13–18; Cincinnati, OH. 218–25p.
2. Xiong Y., Shafer S.A. Depth from focusing and defocusing. *IEEE Computer Vision Pattern Recognition.* 1993; 7: 68–73p.
3. Helmli F.S., Scherer S. Adaptive shape from focus with an error estimation in light microscopy. *Proceedings of the 2nd International Symposium on Image and Signal Processing and Analysis*; 2001 Jun 19–21; Pula, Croatia.
4. Sahay R.R., Rajagopalan A.N. Dealing With Parallax In Shape-From-Focus. *IEEE Trans Image Processing*. 2011; 20(2): 558–69p.
5. Subbarao M., Choi T.S. Accurate recovery of three dimensional shape from image focus. *IEEE Trans Pattern Anal Mach Intell*. 1995; 17(3): 266–74p.
6. Geusebroek J., Cornelissen F., Smeilders *et al*. Robust autofocusing in microscopy. *Cytometry*. 2000; 39: 1–9p.
7. Chern N.K., Neow P.A., Ang M.H. Practical issues in pixel–based autofocusing for machine vision. *Proceedings of the International Conference on Robotics and Automation;* 2001 May 21–26; Seoul, South Korea. 2791–6p.
8. Santos A., de Solorzano C.O., Vaquero J.J., *et al*. Evaluation of autofocus functions in molecular cytogenetic analysis. *J Microsc*. 1997; 188: 264–72p.
9. Sun Y., Duthaler S., Nelson B.J. Autofocusing in computer microscopy: selecting the optimal focus algorithm. *Microsc Res Tech*. 2004; 65:139–49p.
10. Yang G., Nelson B. Wavelet-based autofocusing and unsupervised segmentation of microscopic images. *Proceedings of the International Conference on Intelligent Robots and Systems*; 2003 Oct 27–31; Las Vegas, NV, USA. 2143–8p.
11. Bay H, Tuytelaars T, Gool LV. SURF: Speeded up robust features. *Proceedings of the 9th European Conference on Computer Vision;* 2006 May 7–13; Graz, Austria. 404–17p.
12. Pertuz S., Puig D., Garcia M.A. Analysis of focus measure operators for shape-from-focus. *Pattern Recog*. 2013; 46:1415–32p.