A Generative Focus Measure With Application to Omnifocus Imaging

Avinash Kumar and Narendra Ahuja
Department of Electrical and Computer Engineering,
University of Illinois, Urbana-Champaign, Urbana, IL 61801 USA
{avinash,n-ahuja}@illinois.edu

Abstract

Given a stack of registered images acquired using a range of focus settings (focal stack images), we propose a new focus measure to identify the most focused image. Although, most of the paper is concerned with the new focus measure, for evaluation purposes, we will present it in the context of an application to generating omnifocus images. An omnifocus image is the composite image in which each pixel is selected form the frame in the stack in which it appears to be in best focus. Conventional focus measures usually maximize some measure of image gradient in a window. They tend to fail when one of the edges of the window lies near the boundary of an intensity edge, or the pixel is near other complex edge patterns. This leads to the misidentification of the focused frame and formation of artifacts in omnifocused image. Our proposed measure does not attempt to identify the focused frame by calculating the degree of defocus, like the gradient based methods. Rather, it hypothesizes that a specific frame is in focus and then validates or rejects this hypothesis by recreating the defocused frames in the vicinity, and comparing them with the observed defocused frames. This forward generative process leads to correct focus frame selection in regions where typical measures fail.This is because the conventional measures try to identify the focused frame from its distorted version which is the result of a complex convolution process. This involves a backward estimation for a many-to-one transformation. On the other hand, the generation of defocused frames from a hypothesized focused frame is more accurate since it involves applying an operator in the forward direction. We analytically show that under ideal imaging conditions, the proposed focus measure is unimodal in nature. This makes the search for the best focused image unambiguous. We evaluate our focus measure by generating omnifocus images from real focal stack images, and show that it performs better than all the existing focus measures.

1. Introduction

Conventional cameras have limited Depth of Field (DoF), i.e. at a time they can focus only on a fixed range of depths depending on the camera setting used for image capture. Due to this, scene depths outside the DoF limits are imaged in defocus and much of the high frequency details pertaining to edges and corners of the objects at these depths is lost. This could be critical for various low-level vision algorithms relying on pixel level image information e.g. segmentation. Additionally, partially focused images are not visually pleasing. Thus, capturing a scene with extended DoF and generating an *omnifocus* (omni=all) image of the scene has been a popular research area in computer vision [15, 18, 14, 23] and optics [8]. With the recent affordability of DSLR cameras and smart phones fitted with better cameras, there has been a renewed interest towards omnifocus imaging [17, 6, 7] and related areas of refocusing [19, 25], optimal number of images to capture subject to reduced noise and defocus [11] and exposure [10] for extended DoF capture etc. In this paper, we focus on the problem of omnifocus imaging using focal stack images [15, 25] which use a focus measure [22, 18, 23] metric to quantify the amount by which an image pixel is in focus. This metric is then used to predict the best focused pixel across the focal stack for each pixel location. In our paper, we analyze a small but critical drawback in the design of existing focus measure techniques which can lead to artifacts near *intensity edges* (not depth edges) located on locally planar surface. We thereafter propose a new focus measure to handle this problem and show results on real data. A detailed review of existing omnifocus imaging techniques including the ones employing focus measure criteria is presented next.

2. Previous Work

Most of the previous work in omnifocus imaging can be broadly divided into two categories:

(A)*Computational Cameras/Single Image*: The camera optics is modified to acquire an image with depth invariant blur and a single deconvolution is used to obtain omnifocus image [8, 7]. Levin [16] capture a single image using coded aperture and use calibrated blur kernels to simultaneously obtain depth and an omnifocus image.

(B)*Conventional Cameras/Multiple Images*: In this tech-

nique, a set of images is first captured by moving the sensor plane along the optical axis, thereby focusing on different depths in each captured image. This set of images is referred to as *focal stack* images [4] (Figure 1). To obtain the image in which a given object at a certain depth is best focused a *focus measure* is computed across the focal stack images [13, 20, 15, 22, 18, 1, 23, 14]. This measure gives a quantitative estimate of how focused an image is in any one of the given focal stack images (See Fig. 1). Thus, the extrema of the focus measure corresponds to the best focused image. Once this image is known, the pixel intensity corresponding to that object is extracted from that image and pasted on a new image. This procedure is repeated for pixels corresponding to all the objects being imaged and finally a omnifocus image is obtained. Hasinoff *et.al.* [11] have also shown that omnifocus image obtained using focal stack usually have a higher signal-to-noise(SNR) ratio as compared to single shot based techniques *i.e. Category A*. Nagahara [17] combine focal stack images captured using a computational camera and shown that the blur kernel of the integrated images is relatively constant over the set of depths imaged by all the input focal stack images. Thus, a single deconvolution of the integrated image is sufficient to yield an omnifocus image.

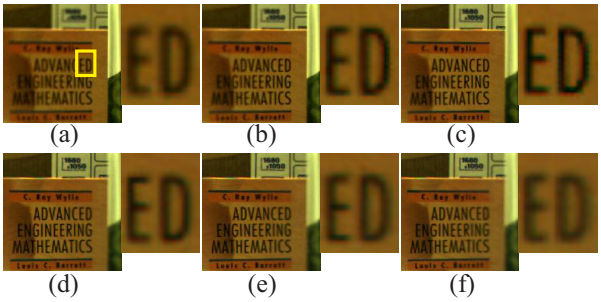

Figure 1. (a-f) Focal stack images. The goal is to design a focus measure which detects the best focused image (image (c)).

Traditional focus measures [20, 22, 18, 15] find the best focused image frame for a given pixel across the focal stack by maximizing the gradient present in a window around the pixel location. But, it has been shown in Ning [3, 23] that such methods fail to identify the best focused pixel if the window over which the gradient is being computed lies in the neighborhood of an **intensity edge** on a locally planar surface(See Fig. 2). In such a situation, the windows lying in focused image have less/no gradients, whereas the same window in a blurred image contains intensity values which have bled into it due to the defocusing of a sharp edge lying at the border of the window. A synthetic example where gradient maximizes for defocused windows near sharp intensity edges is shown in Fig. 3. In real images this leads to artifacts in omnifocused images near intensity boundaries. It can be noted that although there intensity edge could be a depth edge also, no focus measure can perform well at such boundaries as the defocus blur kernel is a complex combination of two different blur kernels [4]. As such, the proposed focus measure can not handle depth boundaries and regularization techniques based techniques need to be applied to handle it.

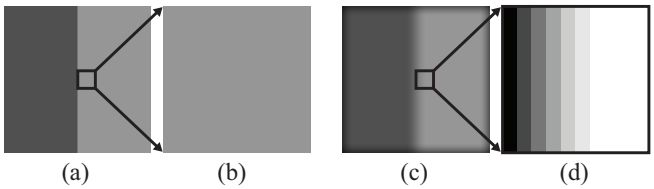

Figure 2. **Synthetic example** (a)Focused image consisting of an intensity edge. (b)Zoomed out small window near intensity edge has no gradients. (c)Artificially blurred image of (a). (d)Zoomed out patch (intensity scaled between [0 255] for better visualization) located at the same location as in (a). This window contains significant gradients causing errors in gradient based focus measure.

Figure 3. (a)**Focused** (b)**Blurred** with isotropic Gaussian blur of $\sigma = 2.47$. (c-d)**Red** indicate pixel locations where focus measures based on(c) Variance [15], (d) Energy of Gradient [22] (in a 5×5 window) is higher for image (b) compared to the corresponding location in image (a). These regions lie near intensity edges.

In order to alleviate this problem, we propose a new focus measure by modeling forward focal stack image formation (See Fig. 4), thus called as generative focus measure. Specifically, a window is first chosen around a pixel location in one of the focal stack images and it is **assumed to be in focus**. Then, based on this assumption, the amount of blurring which would be produced in windows at same location but in all other frames is predicted. If the focus assumption were indeed correct, then the predicted blur will be same or close to the actually observed blur in all the focal stack images. Thus, the L^2 norm between the predicted and observed blurred images is defined as the focus measure. This norm minimizes for the best focused image. This modeling is closest to the forward modeling of depth and omnifocus image proposed in Favaro [9] and omnifocus image estimation in Hasinoff [11]. Compared to both of these techniques, the proposed focus measure based technique is much simpler to compute and is independent of any optimization step which could be prone to local optima if not initialized properly. We show that under noiseless imaging and Gaussian optical blurring [12], the proposed focus mea-

sure at each pixel location is unimodal in nature (Section 5). Thus a focused image pixel is determined uniquely at each pixel location. Also, our technique is based on capturing all the depths in a desired range to be in focus in at-least a single image. This varies from [11], where this is not a necessary requirement and the omnifocus image is an estimation problem from an optimal set of focal stack images.

Section 3 describes our technique in detail. Also, we assume that the various camera parameters, namely focal length, F-number and the location of sensor planes used to capture focal stack are already known by camera specification or can be estimated via camera calibration [24].

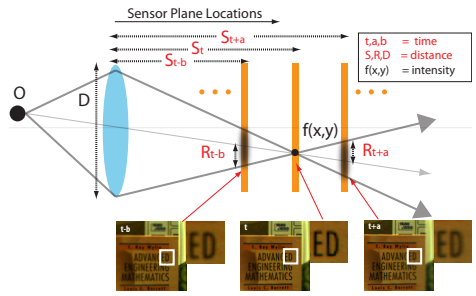

Figure 4. Multifocus imaging geometry. Focused(middle) and the defocused(left,right)

3. Focus Measure Calculation

Since the proposed focus measure depends on formation of focused and defocused images, we first discuss the focal stack geometry in Section 3.1. The captured images have the property that each image focuses on different depths. Given this focal stack, Section 3.2 explains the proposed technique for generating focus measure vector to obtain an omnifocus image.

3.1. Focal Stack Acquisition

A focal stack is acquired by first fixing all the camera parameters while setting the aperture to the maximum, thus allowing for a very small DoF to be focused in each image. The sensor plane is then sequentially moved along the optical axis in discrete amounts S_t (t denotes time) (See Fig. 4) measured from the center of the lens. At each S_t, an image is captured and stored, thus generating a set of input images. The number of acquired images and the corresponding sensor locations can be optimized [2] such that the combined sum of the DoF of all images is closest to a input larger DoF over which we want to compute the omnifocus image.

From Fig. 4, given a unit source of light at O, the radius of the point spread function on frames (located at S_{t-b}, S_{t+a}) neighboring to the focus frame f (located at S_t) can be calculated as,

$$R_{t-b}(x,y) = \frac{D}{2}\left[1 - \frac{S_{t-b}}{S_t}\right]; R_{t+a}(x,y) = \frac{D}{2}\left[\frac{S_{t+a}}{S_t} - 1\right]$$

where, D is the diameter of the aperture stop. The PSF can be analytically modeled either as a *Pillbox* or a *Gaussian* distribution. The *pillbox* model assumes a perfect imaging system devoid of any noise, where the resulting light energy due to defocusing is equally spread across the area of the circular blob. It can be represented as h_p:

$$h_p(x,y) = \begin{cases} \frac{1}{\pi R^2}, & \text{for} \quad x^2+y^2 \leq R^2; \\ 0, & \text{for} \quad x^2+y^2 > R^2. \end{cases} \tag{1}$$

But, due to imaging imperfections and noise in the imaging system the spread of light energy is not uniform in the circular blob and a Gaussian distribution for the PSF seems more practical.

$$h_g(x,y) = \frac{1}{2\pi\sigma^2} \exp -\frac{x^2+y^2}{2\sigma^2}. \tag{2}$$

where $\sigma \approx \frac{R}{c}$ and $c = \sqrt{2}$ [22]. Thus, given a model of impulse response or PSF $h = \{h_p, h_g\}$, derived from known camera parameters and an unknown omnifocus image, any defocused image g captured by the camera can be obtained as the following convolution:

$$g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x-\xi, y-\eta)h(\xi,\eta)d\xi d\eta \tag{3}$$

By applying a forward convolution and assuming each focal stack image as a candidate for being an omnifocused image (locally), we derive a generative focus measure as described below.

3.2. Generative Focus Measure

This section describes a generative focus measure for omnifocus imaging. It is assumed that various camera parameters namely: intrinsic camera parameters, aperture diameter D and the sensor plane distances S_t, where t varies from 1 to N are known. As the sensor plane shifts along the optical axis in discrete steps, an image of the scene is captured. The complete set of N images are called focal stack images, as they capture the scene with different amounts of focus and defocus (described in Section 3.1).

The set of N focal stack images can be represented as a 3D matrix MI, where the index $MI(x,y,k)$ denotes the intensity value at a 2D location (x,y) in the k^{th} focal stack image. Also, let the size of each focal stack image be $X \times Y$. Since the focal stack images are registered, $MI(x,y,k)\ \forall\ k \in \{1,\cdots,N\}$ represents the complete set of focused and defocused intensities corresponding to the entire scene in the 3D world. In order to find the optimal index k^* representing the best focus frame for a given scene object, the following focus measure is applied.

- Step 1: Repeat the following for all pixel locations (x, y), where $x \in \{1, \ldots, X\}$ and $y \in \{1, \ldots, Y\}$.
- Step 2: Select a focal stack image index k, where $k \in \{1, \ldots, N\}$.
- **Assume the pixel intensity at $MI(x, y, k)$ is a focused intensity**. If the object appearing at (x, y) was indeed focused in frame k located at a distance of S_k from the lens, then the images formed on all other sensors will be defocused in accordance with the defocus procedure described in Section 3.1.
- Step 3: Set $a = 1$ and $b = 1$. Since sensor locations S_k,S_{k-b} and S_{k+b} are known, use Eq. 1 to calculate the blur radius R_{k-b} and R_{k+a}, on sensors locations S_{k-b} and S_{k+a}.
- Step 4: Given the two radius, calculate two Gaussian PSFs h_g^a and h_g^b using Eq.[2]. The choice of Gaussian PSF over Pillbox is validated in Section 4.
- Step 5: Now, assign intensities selected from a $\overline{W} = W_d \times W_d$ sized window around the pixel location (x, y) in the k^{th} focal stack image to f. Thus f is of size $\overline{W}$.
- Step 6: Apply convolution Eq. 3 to obtain artificially defocused images as : $g_b = f * h_g^b$ and $g_a = f * h_g^a$.
- Step 7: Go to Step 4, and vary b from 1 to $(k-1)$ to obtain $(k-1)$ artificially defocused images, $G_b = [g_b^1, g_b^2, \ldots, g_b^{k-1}]$. Similarly, vary a from 1 to $(N-k)$ to obtain $(N-k)$ artificially defocused images, $G_a = [g_a^1, g_a^2, \ldots, g_a^{N-k}]$.
- Calculate Focus measure $F(x, y, k)$ for (x, y) being focused in frame k as described below.

If the assumption of intensity $MI(x, y, k)$ being focused made in Step 3 of the algorithm was indeed correct, then the artificially generated images G_b and G_a will be quite similar to the originally observed intensities in a $\overline{W}$ sized window located around $MI(x, y, k-b)$ and $MI(x, y, k+a)$ where as usual $a \in \{1, \ldots, N-k\}$ and $b \in \{1, \ldots, k-1\}$. This is in accordance with focal stack imaging theory described in Section 3.1. Otherwise, if k were actually defocused, there will be considerable difference between the generated and observed. Based on this intuition, we define our focus measure $F(x, y, k)$ as the L^2 norm taken on a $\overline{W}$ sized window between artificially blurred set $\{G_b, G_a\}$ and the observations $\{MI(x, y, k-b), M(x, y, k+a)\}$. This norm is a measure of how focused an image formed at (x, y) is in the k^{th} focal stack image. Thus we have,

$$F(x, y, k) = F_b + F_a; \text{ where} \tag{4}$$

$$F_b = \sum_{i=1}^{k-1} \sum_{(x,y)\in\overline{W}} [g_b^i - MI(x, y, k-i)]^2$$

$$F_a = \sum_{j=1}^{N-k} \sum_{(x,y)\in\overline{W}} [g_a^j - MI(x, y, k+j)]^2$$

Thus, the focus measure function F is defined such that assuming that the PSF is correctly modeled, it should attain minimum magnitude only for the correctly hypothesized focus frame index. After calculating the focus measure vector, the best focused frame k^* from the set of focal stack images for a pixel location (x, y) can be easily obtained as:

$$k^* = \underset{k\in\{1,\ldots,N\}}{\operatorname{argmin}} \; F(x, y, k). \tag{5}$$

Finally, the intensities in the omnifocus image are obtained as $OF(x, y) = MI(x, y, k^*)$. This process is repeated for all pixel locations to obtain the complete omnifocus image OF. Due to the dependence of our focus measure on the existence of an accurate PSF model, the next section is devoted towards an analysis of the effect of choice PSFs on the proposed focus measure.

4. Analysis of Focus Measure : Pillbox or Gaussian PSF

The Pillbox PSF model is built on the assumption that imaging conditions are ideal, due to which the light energy is uniformly distributed in the region enclosed by the circular defocus blob formed on some sensor plane. On the other hand, a Gaussian PSF model captures the effects of diffraction and imaging imperfections [12, 20] on the intensity distribution inside the circular blob. One way of analysis each of these would be to calculate proposed focus measure curves using them and then studying the attributes of these curve to infer the characteristics of corresponding PSF models. We proceed by selecting a pixel location $Q = (x_q, y_q)$

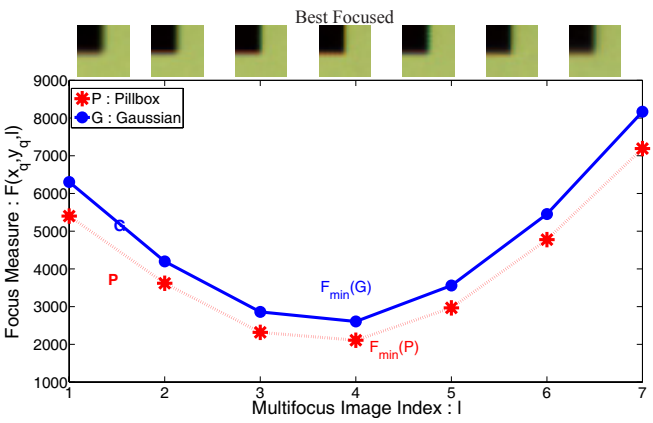

Figure 5. Plot of Focus Measure values Vs focal stack image frame index for Pillbox(P) and Gaussian PSF(G). Since, the slope of G is more than P, the focus measure curve corresponding to Gaussian PSF finds the optimal focus frame $'4'$ more unambiguously as compared to the one using Pillbox PSF.

from one of the focal stack image frames such that it lies near an intensity edge. The choice is based on the fact that

focusing/defocusing effects are more prominent near intensity edges. The focus measure $F(x_q, y_q, l)$ (Eq. 4) is then evaluated at Q in each of the input focal stack images k by first setting the PSF as $h = h_p$ (pillbox, Eq. 1) and then $h = h_g$ (Gaussian, Eq. 2 with $c = \sqrt{2}$) in Step 5 of the algorithm mentioned in Section 3.2. Thus, two focus measure curves are obtained. Fig. 5 shows two such sample curves, with the magnitude of focus measure values plotted against their corresponding focal stack frame index. Henceforth, the Pillbox PSF based focus measure curve will be referred to as P and Gaussian PSF based focus curve will be referred to as G. Additionally, the top of the Fig. 5 shows a small region from the corresponding focal stack images in which the focus measure was calculated. It is also observed that both the focus curves attain their minimum values for the same correctly focused image window in spite of using different PSF models. Each of the focus measure plots are analyzed on the basis of the following two features:

- The magnitude of the minimum of focus measure curve attained by the best focused frame,

$$\begin{aligned} F_{min} &= F(x_q, y_q, k^*) \\ \text{where } k^* &= \underset{k}{\operatorname{argmin}}\, F(x_q, y_q, k). \end{aligned} \tag{6}$$

  The F_{min} for each curve can be thought of as a measure of the accuracy with which a particular PSF model models the defocusing produced by the correctly focused frame on all other images in the focal stack; the lower the value, the better the PSF model.
- The slope of a focus measure curve given as

$$\Delta F = \frac{\partial F(x_q, y_q, k)}{\partial k} \tag{7}$$

  The slope captures the information of increase in the magnitude of $F(x_p, y_q, k)$, as one moves away from the index of the correctly focused frame k^*. The larger the increase, larger is the SNR for finding k^*.

Thus, an ideal PSF model for computing the focus measure would be the one, with has lowest F_{min} and has highest ΔF among all the PSF models. For our case, P has smaller F_{min}, as well as smaller ΔF compared to G. Since, the goal is to obtain the index of best focused image with less ambiguity, the criteria of highest ΔF is favored over the criteria of least F_{min} for selecting appropriate PSF model. Thus, a Gaussian PSF model corresponding to $c = \sqrt{2}$ is used for obtaining the focus measure vector. Not surprisingly, the proposed analysis matches previous works of [20, 22] which have suggested the use of Gaussian PSF.

5. Unimodality of the Focus Measure

The search for best focus image frame for a given scene point implies minimizing Eq. 5. The solution becomes relatively unambiguous and accurate if the focus measure is *unimodal* in nature and the minimum lies at the correct focal stack image index. In the following, we prove that the proposed focus measure is indeed unimodal in nature under some mild assumptions.

Based on the observability analysis of Favaro [9] which shows that smooth regions and regions with brightness gradient do not get effected by rotationally symmetric blur (in our case Gaussian defocus blur in focal stack images), we present the current analysis for regions in the desired omnifocus image having sufficient textures. For smooth regions, any of the focal stack images would suffice as they are indistinguishable. We first assume that the sensor planes are shifted along the optical axis in such a manner that all the depths in the scene are captured in focus in at-least one of the focal stack images. Secondly, we assume that the blur can be modeled as having a Gaussian distribution with mean 0 and some variance σ^2 (as in Section 3).

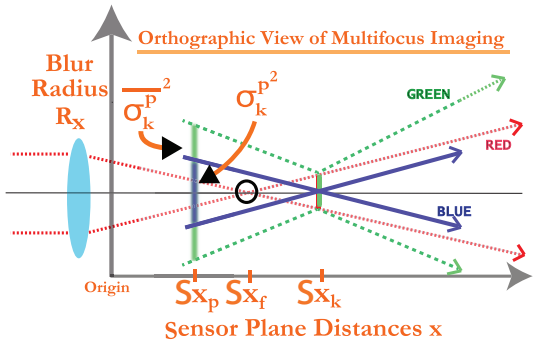

Figure 6. The RED rays correspond to the formation of a set of focal stack images on sensor plane located along the x-axis, with the focused image being formed at x_f. The BLUE rays correspond to hypothesized set of focal stack images assuming x_k is focused and blurring the observed image at x_k. But, this blurring in turn generates the GREEN set of rays, as the image at x_k was already defocused.

In order to make our analysis easy, it is assumed that sensor plane distances: S_x and corresponding blur radius formed on that sensor plane due to defocusing: R_x are continuous. Although, the sensor planes are located at discrete distances, yet unimodality of a function in continuous domain automatically holds for the discrete domain as well. Next, a coordinate system is defined with respect to which all distances would be measured. The y axis of this system coincides with the location of the sensor plane closest to the lens of the camera. The x axis coincides with the optical axis as shown in Fig. 6. Lastly, it is assumed that there is no sensor noise and the image degradation is only due to optical blur.

Now, lets suppose that focal stack is given, such that an object is imaged in best focus on the sensor plane located at a distance of S_{x_f} from the origin O (See Fig. 6). Since, we assume the PSF follows a 2D Gaussian model, the amount

of degradation due to blurring is directly proportional to σ^2, the variance of the Gaussian. Based on the algorithm proposed in Section 3.2, we select a arbitrary sensor location $S_{x_k}, S_{x_k} \neq S_{x_f}$ and assume that it is in focus. Then we calculate the blurring produced on some other sensor located at $S_{x_p}, S_{x_p} \neq S_{x_k}$ as shown in Fig. 6. If S_{x_k} was indeed focused, then the sensor located at S_{x_p} will be blurred by σ_k^p which can be given as (combining Eq. 2,

$$\sigma_k^p \approx \frac{D}{2\sqrt{2}} \left[\frac{|S_{x_k} - S_{x_p}|}{S_{x_k}} \right]. \quad (8)$$

But, since it is already known that the sensor located at S_{x_f} was in focus, blurring the sensor at location S_{x_k} with σ_k^p, would actually produce a different blur at S_{x_p} which can is parameterized by $\overline{\sigma_k^p}$, given as obtained as

$$\overline{\sigma_k^p} \approx \frac{D}{2\sqrt{2}} \sqrt{\left[\frac{|S_{x_k} - S_{x_f}|}{S_{x_f}} \right]^2 + \left[\frac{|S_{x_k} - S_{x_p}|}{S_{x_k}} \right]^2}. \quad (9)$$

where we have utilized the dependence of σ on the geometric blur radius and applied the fact that convolving two 2D Gaussians results in another 2D Gaussian, whose variance is the sum of the variance of the two original Gaussians.

As the amount of blurring is directly proportional to variance, we use the following error measure $\Delta S_{x_k}, S_{x_p}$ to determine the difference in amount of blurring

$$\Delta(S_{x_k}, S_{x_p}) = \overline{\sigma_k^p}^2 - \sigma_k^{p2} = \frac{D^2}{8} \left[\frac{k-f}{f} \right]^2$$

On partial differentiation of $\Delta(S_{x_k}, S_{x_p})$ w.r.t S_{x_k},

$$\frac{\partial \Delta(S_{x_k}, S_{x_p})}{\partial S_{x_k}} = \frac{2}{S_{x_f}^2} [S_{x_k}^2 - S_{x_k} \times S_{x_f}] \quad (10)$$

For $S_{x_k} > S_{x_f}$, $\frac{\partial \Delta(S_{x_k}, S_{x_p})}{\partial S_{x_k}} > 0$, which means as the distance of wrongly assumed frame S_{x_k} increases beyond S_{x_f}, the difference of blurring PSFs $\Delta(S_{x_k}, S_{x_p})$ increases. In other words, the focus measure defined in Eq. 4 increases as S_{x_k} increases. Similarly, for $S_{x_k} < S_{x_f}$, $\frac{\partial \Delta(S_{x_k}, S_{x_p})}{\partial S_{x_k}} < 0$, or the error increases again as S_{x_k} moves away from the correctly focused sensor located at S_{x_f} towards origin. Thus the focus measure increases on both sides of S_{x_f} and has a minimum at S_{x_f} as $\sigma_k^p = \overline{\sigma_k^p}$. Thus the focus measure is unimodal about the correctly focused sensor location S_{x_f}.

6. Experiments and Results

The experimental setup consisted of a non-frontal imaging camera [14] with a tilted sensor plane. The aperture was set to wide open and the camera was rotated about the optic center. As shown in [14], such a system allows easy acquisition of focal stack images along with the fact that a wide field of view can be imaged. The obtained images were corrected for magnification by registering the images given the known camera calibration parameters and vignetting was removed using the technique proposed in Castano [5].

6.1. Results : Real Datasets

The accuracy of the proposed focus measure was compared to five existing methods: Energy of the gradient [22], Energy of Laplacian [22], Sum Modified Laplacian [18], Variance in a window [22] and the gray level distribution in a window [23]. All the focus measures were compared using the same window size parameter of 7×7.

Fig. 7 shows a focus chart which has many sharp edges. It is placed at a distance of 2ft from the optic center of the camera and 11 focal stack images are captured. Next, various focus measures [22, 18, 23] are applied and the final omnifocus image is obtained for each of them. A comparison of the obtained results with our generative focus measure is shown in Fig. 7. The best performer among existing methods is the energy of Laplacian [22] and the gray level distribution based measure [23].

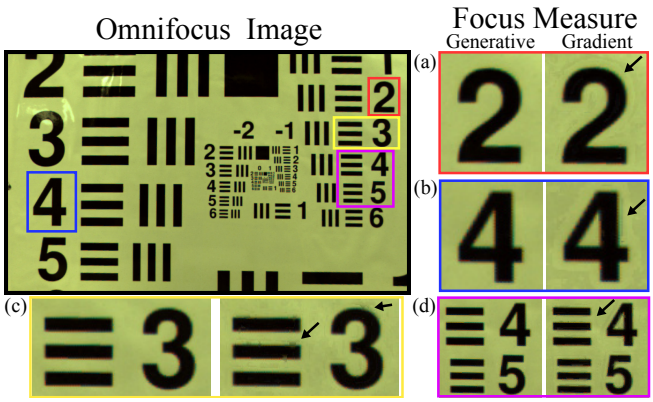

Figure 7. Omnifocus image of a planar scene with comparison on focus frame selection using generative and gradient based techniques. The artifacts due to gradient methods are indicated in zoomed out windows in black arrow.(a)energy of gradient [22]. (b)gray level distribution [23]. (c)SML [18]. (d)variance.

In Fig. 8, we apply gradient based focus measure and our proposed focus measure on a scene with non-planar objects. The artifacts due to traditional methods is clearly visible in the Fig. 8(middle). The computed initial depth map by just using the focus measure is also shown in Fig. 8(right). This depth map is refined by applying volumetric graph cuts [21]

In Fig. 9, we compare the performance of traditional focus measures with generative focus measure for a pixel located near an intensity edge. As hypothesized earlier, traditional focus measures tend to fail near such locations. As can be seen in the plot, traditional focus measures [22, 18, 23] peak at the frame index 1, which corresponds to a defocused frame. Whereas the proposed mea-

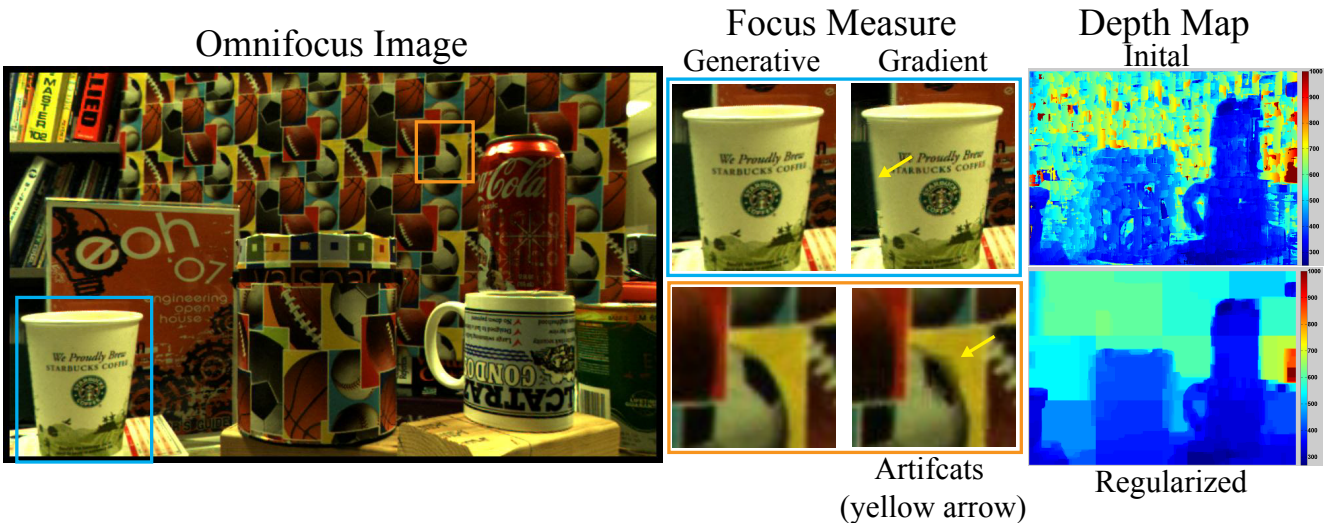

Figure 8. (left) Computed omnifocus image of a textured scene (middle) Comparison of applying focus measure using generative and gradient based technique. The artifacts due to gradient based techniques is clearly visible (yellow arrow). (right) Estimated coarse and refined (using volumetric graph cuts) depth map of the scene (blue indicates close and red indicates far, see colorbar on left).

sure and Ning [23] correctly finds the focused image index to be 3.

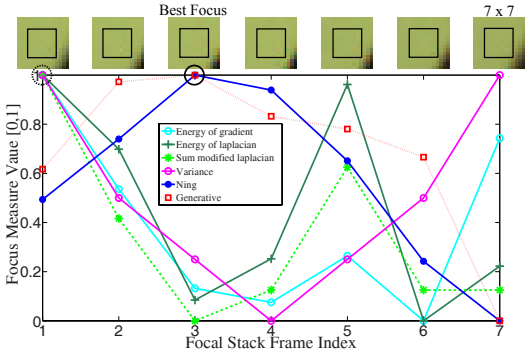

Figure 9. Normalized focus measure around an **intensity edge** using various techniques. Only the proposed technique and Ning's [23] technique succeed in finding the best focused frame 3 (encircled with solid lines). All focus measures were calculated in a 7×7 window around a pixel

In Fig. 10, where the region for calculating focus measure contains considerable gradients, all the measures find the best focused frame correctly.

Finally, in Fig. 11, we give a quantitative estimate of the accuracy of our focus measure in finding the best focused frame. A region of the scene is selected whose focus frame index is already known. The index can be correctly predicted based on the fact object depth and the sensor plane distances are known. Then, all the existing focus measures are applied at each pixel location in this patch and the best focused frame index as predicted by them is obtained. Thereafter, the absolute difference of focus measure predic-

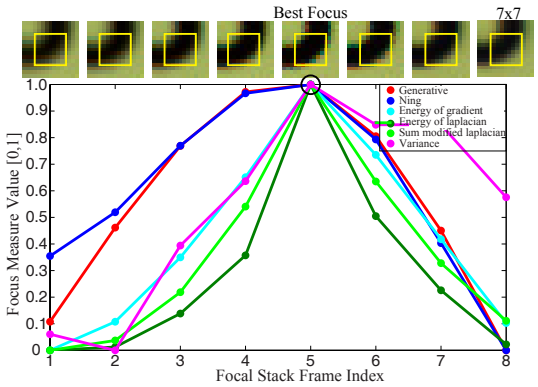

Figure 10. Normalized focus measure for various focus measure in a window containing sharp intensity edge. All the measures peak for the correct focused frame.

tion and the ground truth knowledge of focus index is calculated. The obtained differences are scaled between $[0, 255]$ and shown in Fig. 11. In the images shown, a 0 intensity indicates a perfect match with the ground truth observation, whereas any other intensity indicates false matches. As is evident, most of the focus measures find the focus frame index correctly at the boundary if intensity edge, but fail at the windows which do not contain the boundary. Compared to that, our generative measure detects all the focus frame indexes correctly. The bottom row indicates the percentage of erroneous predictions by each measure. Our measure has 3% error compared to more than 50% error in all other measures. **Computational Time:** Given N focal stack images, the computational complexity of our focus measure is $O(N^2)$ as compared to $O(N)$ of traditional focus measures.

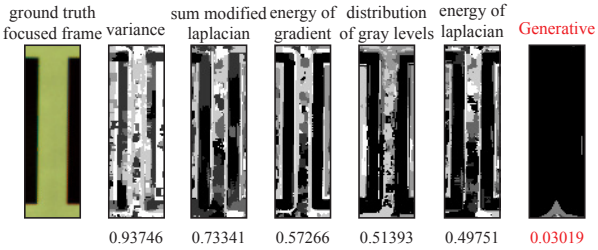

Figure 11. The left most image is the region whose ground truth focus frame index is known. The other images represent the absolute difference between the focus frame index by different focus measures and the ground truth focus frame index. The numbers in the bottom represent the percentage of erroneous predictions. The proposed generative focus measure has the best performance.

6.2. Discussion

Although, the method of [23] is specifically aimed at solving the same problem as ours, yet the initial application of that focus measure yields artifacts in smooth regions as is shown in Fig 7(d), which is also common with other conventional measures. This is because the focus measure calculations at a pixel location are based on computation of gradients within each frame, which are subject to image noise. Thus, image noise adds randomness to selection of the best focus frame. Thus, neighboring window patches in smooth regions attain different focus frame indexes. Compared to this, an estimate of focus measure for one frame by using information from multiple frames, like our generative focus measure, suppresses noise. Thus, our focus measure is locally smooth. Consequently, the omnifocus image obtained by our method are smooth in regions where there is no texture at all. Thus, the requirement of image smoothing methods like graph cuts [23] are not necessary in our case for obtaining best omnifocus image.

7. Conclusion and Future Work

In this paper, we have presented a new generative focus measure and have shown its advantages above existing focus measures via a number of analytical and empirical ways. The generative focus measure correctly handles windows near image boundaries which results in artifacts free omnifocus image. The future work would concentrate on making the algorithm faster and improving upon depth estimation.

References

[1] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen. Interactive digital photomontage. In *SIGGRAPH*, 2004.

[2] M. Aggarwal. *High dynamic range and omnifocus imaging*. PhD thesis, University of Illinois, Urbana-Champaign, 2001.

[3] M. Aggarwal and N. Ahuja. On generating seamless mosaics with large depth of field. In *ICPR*, 2000.

[4] N. Asada, H. Fujiwara, and T. Matsuyama. Edge and depth from focus. *IJCV*, 26(2):153–163, 1998.

[5] A. Castano. *Range from a nonfrontal Imaging Camera*. PhD thesis, University of Illinois, Urbana-Champaign, 1998.

[6] O. Cossairt and S. K. Nayar. Spectral Focal Sweep: Extended Depth of Field from Chromatic Aberrations. In *ICCP*, Mar 2010.

[7] O. Cossairt, C. Zhou, and S. K. Nayar. Diffusion Coding Photography for Extended Depth of Field. *SIGGRAPH*, Aug 2010.

[8] J. Edward R. Dowski and W. T. Cathey. Extended depth of field through wave-front coding. *Applied Optics*, 34(11):1859–1866, Apr 1995.

[9] P. Favaro, A. Mennucci, and S. Soatto. Observing shape from defocused images. *IJCV*, 52(1):25–43, Apr. 2003.

[10] S. W. Hasinoff and K. N. Kutulakos. Light-efficient photography. In *ECCV*, 2008.

[11] S. W. Hasinoff, K. N. Kutulakos, F. Durand, and W. T. Freeman. Time-constrained photography. In *ICCV*, 2009.

[12] B. Horn. *Robot Vision*. McGraw-Hill, 1986.

[13] R. A. Jarvis. Focus optimisation criteria for computer image processing. *Microscope*, 24(2):163–180, 1976.

[14] A. Krishnan and N. Ahuja. Range estimation from focus using a non-frontal imaging camera. *IJCV*, 20(3):169–185, 1996.

[15] E. P. Krotkov. Focusing. Technical Report MS-CIS-86-22, GRASP Lab., University of Pennsylvania, 1986.

[16] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. In *SIGGRAPH*, 2007.

[17] H. Nagahara, S. Kuthirummal, C. Zhou, and S. Nayar. Flexible Depth of Field Photography. In *ECCV*, Oct 2008.

[18] S. Nayar and Y. Nakagawa. Shape from Focus. *PAMI*, 16(8):824–831, Aug 1994.

[19] R. Ng. Fourier slice photography. *SIGGRAPH*, 2005.

[20] A. P. Pentland. A new sense for depth of field. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 9(4):523–531, 1987.

[21] S. Roy and I. Cox. A maximum-flow formulation of the n-camera stereo correspondence problem. In *ICCV*, 1998.

[22] M. Subbarao, T. Choi, and A. Nikzad. Focusing techniques. *Journal of Optical Engineering*, pages 2824–2836, 1993.

[23] N. Xu, K. Tan, H. Arora, and N. Ahuja. Generating omnifocus images using graph cuts and a new focus measure. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 4, pages 697 – 700 Vol.4, aug. 2004.

[24] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *ICCV*, 1999.

[25] C. Zhou, D. Miau, and S. Nayar. Focal Sweep Camera for Space-Time Refocusing. Technical report, Department of Computer Science, Columbia University, Nov 2012.