The Optimal Focus Measure for Passive Autofocusing and Depth-from-Focus

Murali Subbarao   Jenn-kwei Tyan
Department of Electrical Engineering, State University of New York
Stony Brook, NY 11794-2350
email: murali@sbee.sunysb.edu, jtyan@sbee.sunysb.edu

Abstract

The optimal focus measure for a noisy camera in passive search based autofocusing (AF) and depth-from-focus (DFF) applications depends not only the camera characteristics but also the image of the object being focused or ranged. In the early stage of this research, a new metric named Autofocusing Uncertainty Measure *(AUM) was defined which is useful in selecting the most accurate focus measure from a given set of focus measures. AUM is a metric for comparing the noise sensitivity of different focus measures. In the later stage of this research, an improved metric named* Autofocusing Root-Mean-Square Error *(ARMS error) was defined. Explicit expressions have been derived for both AUM and ARMS error, and the two metrics are shown to be related by a monotonic expression.*

AUM and ARMS error metrics are based on a theoretical noise sensitivity analysis of focus measures. In comparison, all known prior work on comparing the noise sensitivity of focus measures have been a combination of subjective judgement and experimental observations. For a given camera, the optimally accurate focus measure may change from one object to the other depending on their focused images. Therefore selecting the optimal focus measure from a given set involves computing all focus measures in the set. However, if computation needs to be minimized, then it is argued that energy of the Laplacian of the image is a good focus measure and is recommended for use in practical applications. Important properties of the Laplacian focus measure are investigated.

Keywords: *focus measure, focusing, autofocusing, depth from focus, depth from defocus, noise sensitivity of focus measure*

1 Introduction

Autofocusing of electronic cameras is an important problem in consumer video cameras, digital still cameras, digital microscopy, and machine vision. This paper deals with passive autofocusing based on searching for the lens position that gives the best focused image. In this approach, typically, a focus measure is computed for images acquired at several different lens positions, and the lens is moved to that position where the focus measure of the image is a maximum. This paper does not deal with depth-from-defocus (DFD) methods [15, 12] that require images acquired at only two or three lens positions but require knowledge of the optical transfer function (OTF) of the camera and the camera parameters. The focused lens position v (see Fig. 1) depends on the distance u of the object to be focused and the focal length f of the lens. They are related by the lens formula

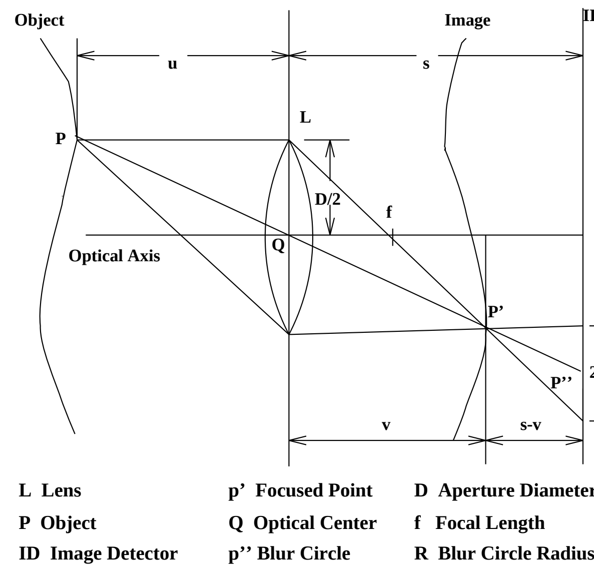

Figure 1: Image Formation in a Convex Lens

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \tag{1}$$

This same relation is used in depth-from-focus (DFF) methods to compute the object distance u from the focused lens position v.

Experimental evaluations of different focus measures have been reported in [6, 7, 16, 8, 11, 17]. So far there has not been any theoretical treatment of the noise sensitivity of focus measures. In the existing literature, all known work have been a combination of experimental observations and subjective judgement. The noise sensitivity of a focus measure depends not only on the noise characteristics but also on the image itself. The optimally accurate focus measure for a given noise characteristics may change from one object to the other depending on its image. This makes it difficult to arrive at general conclusions from experiments alone.

For a given camera and object, the most accurate focus measure can be selected from a given set through experiments as follows. For each focus measure, the object is autofocused several times, say 10, starting with an arbitrary default lens position. The mean of the 10 focused positions and their standard deviation are an estimate of the correct focused position and root-mean-square (RMS) error respectively. The focus measure with the minimum estimate of RMS error is taken to be the optimal. In practical applications such as consumer video cameras or digital still cameras, it is desirable to find the best focus measure from a given set by autofocusing only once. It is quite undesirable to repeat 10 or several trials.

If one has a detailed and accurate information on the focused image of the object to be focused and the camera characteristics such as its OTF, noise behaviour, and camera parameters, then it would be possible to estimate the RMS error theoretically with only one trial. However such information is rarely available in practical applications.

In the absence of such detailed and accurate information, we propose two new metrics named *Autofocusing Uncertainty Measure* (AUM) and *Autofocusing Root-Mean-Square Error* (ARMS error) both of which can be computed with only one trial of autofocusing. In DFF applications, AUM and ARMS error can both be easily translated into uncertainties in depth using Eq. (1).

The analysis here shows that the autofocusing noise sensitivity of a focus measure depends on the image of the object to be autofocused in addition to the camera characteristics. For an object with unknown focused image, finding the optimally accurate focus measure involves computing all the candidate focus measures at a set of lens positions and computing AUM/ARMS error for each of the lens positions. Then the lens is moved to the focused position estimated by the optimal focus measure (which has minimum AUM/ARMS error). Usually the number of candidate focus measures that should be considered for good performance is only a few (about 3). Also, almost all focus measures require only a modest amount of computing. Therefore selecting the optimal focus measure from a candidate set comes at a small computational cost. However, if it is necessary to use minimal computing in autofocusing by using the same focus measure for all objects, then we argue that the the energy of the image Laplacian is a good focus measure to use. This focus measure is shown to have some important desirable characteristics based on a spatial domain analysis.

Next section provides a brief background summary of the model of focus measures. Section 3 describes our autofocusing algorithm. AUM is defined in Section 5 and ARMS error in Section 6. It is followed by noise sensitivity analysis and a discussion of the Laplacian based focus measure in Section 7. Future research topics are described in Section 8.

2 Model of focus measures

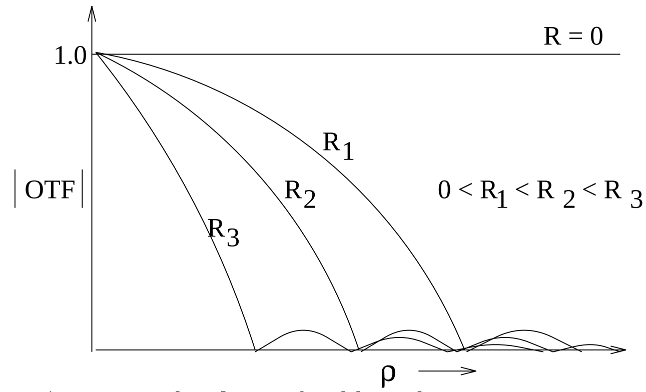

Figure 2: The OTF of a blurred camera attenuates high frequencies. The attenuation effect increases with increasing value of the blur circle radius R

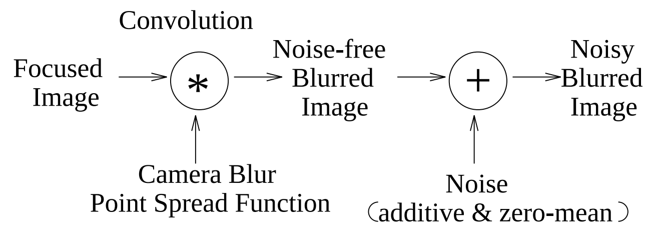

Figure 3: Model of Image Sensing

A detailed discussion of this topic can be found in several papers including [11]. Here we summarize some relevant results based on geometric optics.

When a point object P is blurred on the image detector ID (see Fig. 1) it is imaged as a blur circle P" of radius R. This image $h(x,y)$ is the point spread function (PSF) of the camera. In a small image region if the imaged object surface is (approximately) a plane normal to the optical axis, then the PSF is the same for all points on the plane. Then the blurred image $g(x,y)$ in the small image region on the image detector ID is equal to the convolution of the focused image $f(x,y)$ and the PSF $h(x,y)$. Therefore, if G, F, and H are the Fourier transforms of g, f, and h, respectively, then $G = HF$. The OTF $H(\omega, \nu)$ corresponding to $h(x,y)$ is circularly symmetric and its cross section has a form similar to the sinc function (see Fig. 2). For a focused image, the first zero crossing of the OTF is very far from the zero spatial frequency and the zero crossing moves closer to the zero spatial frequency as the blur increases. Therefore the effect of blurring is to attenuate higher frequencies. The attenuation increases monotonically within the main lobe of the sinc-like OTF. The effect of the main lobe on the computed focus measures usually dominates that of the side lobes. This model of image sensing is summarized in Fig. 3. A more detailed model of image sensing is described in [13].

A general focus measure is modeled as follows (see Fig. 4). First the image for which the focus measure needs to be computed is normalized for brightness by

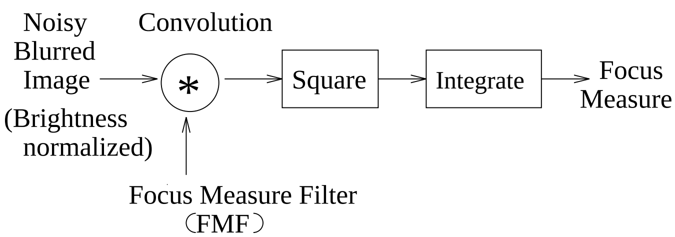

Figure 4: Model of Focus Measure. The FMF for gradient along the x and y axis are $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ respectively and the FMF for the Laplacian is $\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$

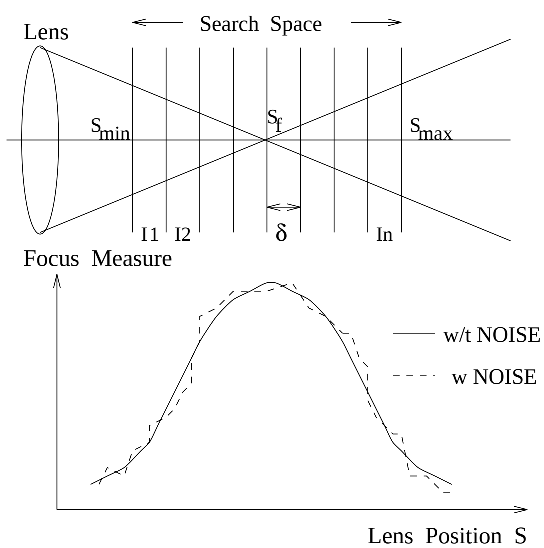

Figure 5: Autofocusing Algorithm

dividing the image by its mean brightness. Then it is convolved with a focus measure filter (FMF). Then the energy (sum of squared values) of the filtered image is computed. This energy is the focus measure. Most FMFs correspond to filters that emphasize (or amplify) high frequencies. This seems appropriate since blurring has the opposite effect, i.e. high frequency attenuation.

The focus measures considered for noise sensitivity analysis in our research are only those which have been proved to be sound [11] based on the effect of the OTF main lobe. In particular we do not consider focus measures based on summing the absolute values of the derivatives of the image [17, 6, 8] that have been used by many researchers in the past. Such filters are proved to be unsound through counter examples in Appendix B.

3 Autofocusing algorithm

In a typical passive autofocusing application such as a consumer video camera, autofocusing is done by moving the lens with respect to the image detector in a narrow range of about a few millimeters or about one tenth of an inch. Let this range be $[s_{min}, s_{max}]$ (see Fig. 5). A typical value for the limits of the range is $s_{min} = f$ and $s_{max} = 1.1f$ where f is the focal length of the lens. Within the range limits, the problem is to find the lens position s_f where the image in a specified part of the image detector is in best focus. Due to the limited depth-of-field caused by diffraction effects, the change in the best focused image is indistinguishable by the image detector when the lens is moved in a small range of size δ around the best focused position s_f. Therefore there is no benefit in achieving autofocusing accuracy better than $\pm\delta/2$. We only need to move the lens to an arbitrary position in the range $[s_f - \delta/2, s_f + \delta/2]$. Typically δ is about one part in 200 of $s_{max} - s_{min}$. Therefore the range $[s_{min}, s_{max}]$ can be divided into n intervals $I_i = [s_{min}+i*\delta, s_{min}+(i+1)*\delta)$ for $i = 0, 1, 2, \cdots, n-1$ with $s_{max} = s_{min} + n * \delta$. It is sufficient to compute the focus measures at only one point in each of these intervals during autofocusing.

In real-time autofocusing applications, the bottleneck is not the computational time but the time taken for the mechanical motion of the lens to move from one position to the other. Therefore it is important to minimize this time at the cost of some additional computation. Search algorithms such as Fibonacci search and binary search are optimal computationally but not necessarily in the time consumed in lens motion. Also, in consumer applications like hand-held video cameras, it is undesirable for the lens to oscillate between extreme focus and defocus conditions rapidly. People find it uncomfortable. It is desirable for the image to gradually come to focus with only minor overshoots near the focused condition.

Based on the above discussion we propose the following algorithm for autofocusing. First the focus measure is computed at the current lens position and the lens is moved by about 10δ to another position. The focus measure is again computed. The sign of the change in the two focus measures is used to determine the direction in which the lens should be moved. Then a sequential search is begun by moving the lens in steps of about $(n/8)\delta$ in the correct direction until the focus measure decreases for the first time. Then a binary search is initiated in the interval containing the last three lens positions until the search interval has been narrowed to about 10δ. Then a quadratic or a Gaussian is fitted to three or more points which are about 5δ apart to find the focused position. Note that, according to geometric optics, the focus measure curve will be symmetric about the focus position s_f. Also, shifting focus position s_f will shift the curve by the same amount with only small change in its shape. This algorithm combines sequential search, binary search, and interpolation, to minimize the lens motion. Additional improvement can be obtained if more information is available about the particular application.

As an alternative to the above algorithm, one may use a depth-from-defocus algorithm [15, 12] when possible to obtain an estimate of the focus position, and then refine this estimate by computing a focus measure at several points near the estimated position, fitting a curve to the points, and finding the position of the curve maximum. The initial estimate is improved by this method because depth-from-defocus methods are less accurate than the search based method above. After improving the initial estimate of the depth-from-defocus method as here, one can compute AUM/ARMS error for the focus measure used in refining the initial estimate.

4 A metric for focus measures

A metric is needed for comparing the noise sensitivity of focus measures both at the focused lens position and at an arbitrary lens position. In Section 1, an experimental method was described briefly for finding the focus measure with minimum RMS autofocusing error. The method involves repeatedly autofocusing a given object. Performance of a focus measure at an arbitrary lens position is of interest for the following reason. In practical applications it will be necessary to determine the direction in which the lens should be moved from an arbitrary initial lens position for autofocusing. The desired direction is the direction in which the best focused lens position is located (Fig. 5). This direction is found by computing the focus measure at the current lens position and at another position a small distance away. The direction in which the focus measure increases is the direction in which the lens should be moved. This method will give the correct direction for any sound focus measure in the absence of noise (because the focus measure increases monotonically), but in the presence of noise, some focus measures may be more prone to give the wrong direction than others. The best focus measure for this purpose can be once again determined experimentally. For a given camera system, object, object distance, and lens position, the sign of the finite differences of the focus measure is used to find the direction in which the lens should move in a large number of trials. The percentage of times the correct direction is found is a measure of noise sensitivity. It will be seen later that the best focus measure depends both on the camera PSF and the image of the object. For a given focus measure, the RMS error will change with the camera PSF and the image of the object.

In practical autofocusing applications, autofocusing of an object has to be done in a few seconds or less in only one trial. Trials cannot be repeated physically to determine the best focus measure. Therefore we need a theoretical metric that can be computed in only one trial of autofocusing. The metric should ideally require as little information about the camera system and the object as possible. It should also be simple and not require much computational resources. These desirable characteristics motivate the metric proposed here.

4.1 Autofocusing Uncertainty Measure (AUM)

First we introduce AUM as a metric for focus measures to illustrate some underlying concepts. Later we introduce the ARMS error which is based on weaker assumptions than AUM. At any lens position s_0 (see Fig. 6), each focus measure γ is associated with a probability density function $p(\gamma(s_0))$, an expected value (mean) $E\{\gamma(s_0)\}$, and a standard deviation $\text{std}\{\gamma(s_0)\}$. However, the focus measure with the minimum standard deviation is not necessarily the best because we are not interested in the accuracy of the focus measure itself, but in the corresponding mean lens position and its standard deviation. Estimating the standard deviation of the lens position requires a knowledge of the function that relates the expected value of the focus measure to the lens position (see Fig. 6). This function depends on the camera PSF as a function of camera parameters and the focused image of the object. In the absence of accurate information about the camera PSF and the object, the function is estimated in a desired interval through sampling and interpolation. For example, near the maximum, the focus measure may be computed at 3 to 5 nearby lens positions and a smooth function such as a quadratic polynomial or a Gaussian is fitted. The assumption is that the computed values of the focus measure are (nearly) the expected values of the focus measure. This assumption will be removed later in defining the ARMS error.

Referring to Fig. 6, the AUM at the maximum of the focus measure γ is defined as follows:

$$\overline{\gamma}_f = \overline{\gamma}(s_f) \tag{2}$$

$$\sigma = \overline{\gamma}(s_f) - \overline{\gamma}(s_1) = \overline{\gamma}(s_f) - \overline{\gamma}(s_2) \tag{3}$$

$$\text{AUM} = s_2 - s_1 \tag{4}$$

where σ is the standard deviation of the focus measure. In order to compute AUM, we need to know σ. In the next section we derive a general formula that can be used to estimate σ as a function of the image and its noise level. Further we need to know the shape of the curve $\overline{\gamma}(s)$ near the peak. As discussed earlier, the position of $\overline{\gamma}_f$ and the function $\overline{\gamma}(s)$ near $\overline{\gamma}_f$ are estimated by fitting a curve (quadratic or Gaussian) to a few points (at least 3) near the maximum. Intuitively, AUM is a measure similar to the RMS error in lens position that can be determined through repeated trials.

Fig. 7 shows a typical comparison of two focus measures. The maximum values of the two focus measures have been normalized to be the same. We see that although $\sigma_2 > \sigma_1$, $AUM_2 < AUM_1$, implying that γ_2 is better than γ_1.

Referring to figure 8, focus measure $\overline{\gamma}$ is modeled to be locally quadratic in a small interval of size 2δ with respect to lens position near the focused position:

$$\overline{\gamma}(s) = as^2 + bs + c \tag{5}$$

Let the focus measure be given at three arbitrary positions which are δ apart. Without loss of generality, let the three positions be $s_- = -\delta$, $s_0 = 0$, and $s_+ = +\delta$. Let $\overline{\Gamma}_- = \overline{\gamma}(s_-)$, $\overline{\Gamma}_0 = \overline{\gamma}(s_0)$ and $\overline{\Gamma}_+ = \overline{\gamma}(s_+)$. Near the focused position, $\overline{\Gamma}_0 > \overline{\Gamma}_-$ and $\overline{\Gamma}_0 > \overline{\Gamma}_+$.

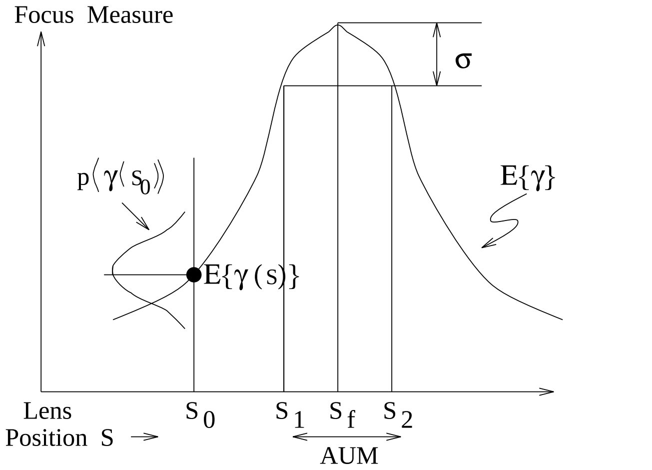

Figure 6: Definition of AUM at the focused position S_f.

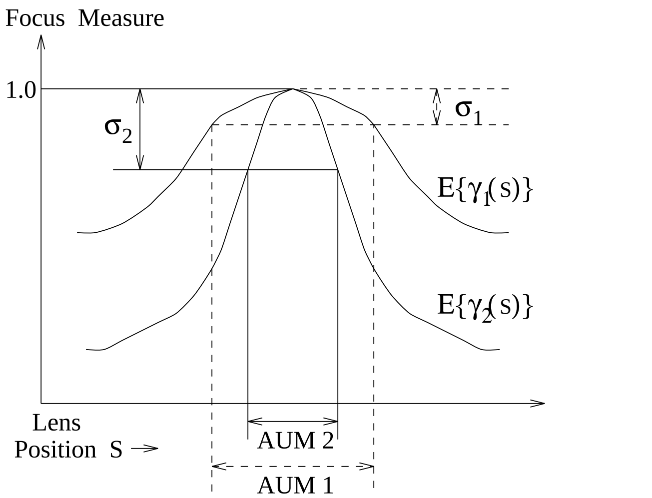

Figure 7: Comparison of two focus measures γ_1 and γ_2 at the focused position

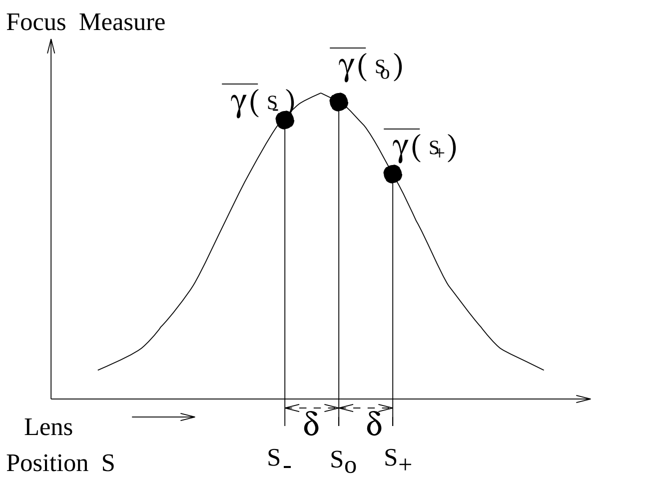

Figure 8: quadratic polynomial interpolation

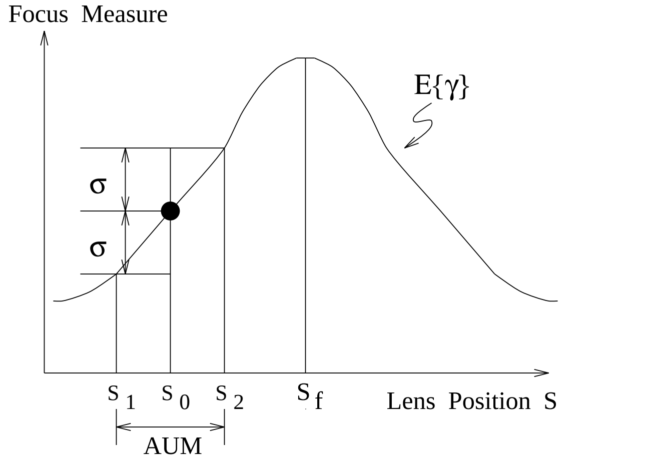

Figure 9: Definition of AUM at a position S_0 far from the focused position S_f.

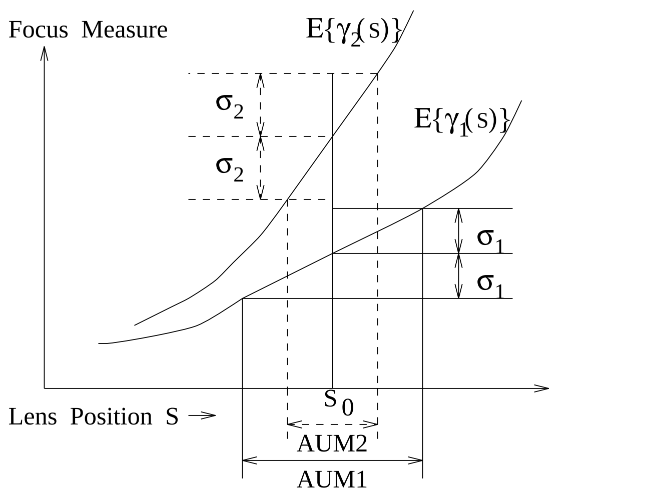

Figure 10: Comparison of two focus measures γ_1 and γ_2 at a position S_0 far from the focused position S_f. Note: $\sigma_2 > \sigma_1$ but $AUM2 < AUM1$, therefore γ_2 is better than γ_1.

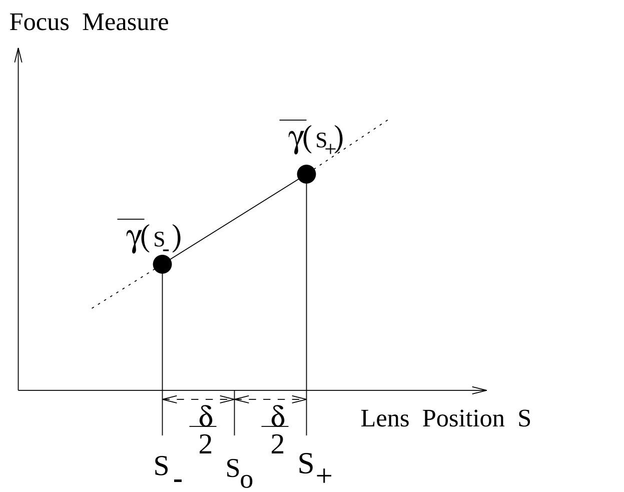

Figure 11: linear interpolation

Solving for the coefficients of the quadratic expression, we obtain

$$\begin{aligned} a &= \frac{\overline{\Gamma}_+ + \overline{\Gamma}_- - 2\overline{\Gamma}_0}{2\delta^2} \\ b &= \frac{\overline{\Gamma}_+ - \overline{\Gamma}_-}{2\delta} \\ c &= \overline{\Gamma}_0 \end{aligned} \qquad (6)$$

Let s_f be the lens position where the focus measure becomes the maximum and $\overline{\Gamma}_f = \overline{\gamma}(s_f)$. At s_f, the derivative of $\overline{\Gamma}$ vanishes. Therefore we obtain

$$\begin{aligned} s_f &= \frac{-b}{2a} \\ &= \frac{\delta}{2} \frac{(\overline{\Gamma}_+ - \overline{\Gamma}_-)}{(2\overline{\Gamma}_0 - \overline{\Gamma}_+ - \overline{\Gamma}_-)} \end{aligned} \qquad (7)$$

Substituting the above equation in (5) we obtain

$$\overline{\Gamma}_f = -\frac{b^2 - 4ac}{4a} \qquad (8)$$

Given that $AUM = s_2 - s_1$, we obtain s_1 and s_2 as the roots of the equation

$$\begin{aligned} \overline{\Gamma}(s) &= \overline{\Gamma}_f - \sigma \\ &= as^2 + bs + c \end{aligned} \qquad (9)$$

Thus solving the above equation, we obtain

$$\text{AUM} = \left| \frac{\sqrt{b^2 - 4a\left(c + \sigma - \overline{\Gamma}_f\right)}}{a} \right| \qquad (10)$$

Substituting with Eq (6) and (8) in the above equation yields

$$\text{AUM} = 2\delta \left(\frac{2\sigma}{2\overline{\Gamma}_0 - \overline{\Gamma}_+ - \overline{\Gamma}_-} \right)^{\frac{1}{2}} \qquad (11)$$

At a position far away from the focused lens position s_f, AUM is defined as in Fig. 9. This is somewhat similar to that near the peak:

$$\sigma = (\overline{\gamma}(s_2) - \overline{\gamma}(s_1))/2 \qquad (12)$$

$$\text{AUM} = s_2 - s_1 \qquad (13)$$

Once again, σ is computed from the known noise characteristics and the image. The shape of the focus measure curve is estimated by a linear (2 points) interpolation using the values of the focus measure at s_- and s_+ that are δ apart. Without loss of generality, let $s_- = -\delta/2$ and $s_+ = +\delta/2$ and the focus measures at these points be $\overline{\Gamma}_-$ and $\overline{\Gamma}_+$ respectively (see Fig. 11). The linear model yields the expression

$$\frac{s - s_-}{s_+ - s_-} = \frac{\overline{\Gamma} - \overline{\Gamma}_-}{\overline{\Gamma}_+ - \overline{\Gamma}_-} \qquad (14)$$

The above equation can be rewritten as:

$$s = \delta \left(\frac{\overline{\Gamma} - \overline{\Gamma}_-}{\overline{\Gamma}_+ - \overline{\Gamma}_-} \right) - \frac{\delta}{2} \qquad (15)$$

We obtain s_1 and s_2 by solving

$$\overline{\Gamma}(s) = \frac{\overline{\Gamma}_+ + \overline{\Gamma}_-}{2} \pm \sigma \qquad (16)$$

where σ is the standard deviation of the focus measure. Using equation (15) and solving for AUM, we obtain

$$\begin{aligned} \text{AUM} &= |s_1 - s_2| \\ &= \frac{2\delta\sigma}{|\overline{\Gamma}_+ - \overline{\Gamma}_-|} \end{aligned} \qquad (17)$$

Fig. 10 shows a comparison of two focus measures far away from the focused position. Once again We see that although $\sigma_2 > \sigma_1$, $AUM_2 < AUM_1$, implying that γ_2 is better than γ_1.

5 ARMS Error

In this section, an explicit expression for the *Autofocusing Root-Mean Square Error* (ARMS error) is derived. An exact expression for the RMS error depends on the Optical Transfer Function (OTF) of the camera and the Fourier spectrum of the focused image. Deriving such an exact expression is complicated because of the nature of the camera's OTF and the variability of the Fourier spectrum of the focused image for different objects. Further, usefulness of such an expression in practical applications is limited since all the information necessary to evaluate the expression (e.g. OTF and camera parameters) may not be available. However, an approximate expression that is very useful in practical applications can be derived under some weak assumptions. The assumption we use is that the expected value of the focus measure can be modeled to be quadratic locally with respect to the lens position. The analysis here can be extended to other models (e.g. cubic or Gaussian) but such extensions do not appear to be useful in practical applications at present.

We are interested in the RMS value of s_{max}. For this reason, the focus measure Γ_i will be expressed as the summation of their expected value $\overline{\Gamma}_i$ and their noise component n_i:

$$\Gamma_i = \overline{\Gamma}_i + n_i \qquad for \;\; i = -, 0, +. \qquad (18)$$

In this case we obtain a set of equations similar to Eqs (5) to (8) with the difference that $\overline{\Gamma}_i$ are replaced by Γ_i , therefore we obtain

$$\begin{aligned} s_{max} &= \frac{\delta}{2} \left(\frac{\Gamma_+ - \Gamma_-}{2\Gamma_0 - \Gamma_+ - \Gamma_-} \right) \\ &= \frac{\delta}{2} \left(\frac{\overline{\Gamma}_+ - \overline{\Gamma}_- + n_+ - n_-}{2\overline{\Gamma}_0 - \overline{\Gamma}_+ - \overline{\Gamma}_- + 2n_0 - n_+ - n_-} \right) \end{aligned}$$

$$= \frac{\delta}{2}\left(\frac{\overline{\Gamma}_+ - \overline{\Gamma}_-}{2\overline{\Gamma}_0 - \overline{\Gamma}_+ - \overline{\Gamma}_-}\right)\left(1 + \frac{n_+ - n_-}{\overline{\Gamma}_+ - \overline{\Gamma}_-}\right) \left(1 + \frac{2n_0 - n_+ - n_-}{2\overline{\Gamma}_0 - \overline{\Gamma}_+ - \overline{\Gamma}_-}\right)^{-1} \quad (19)$$

Near the focused position we have $\overline{\Gamma}_0 > \overline{\Gamma}_+$ and $\overline{\Gamma}_0 > \overline{\Gamma}_-$. Therefore, if the signal to noise ratio is sufficiently large, we have

$$|2\overline{\Gamma}_0 - \overline{\Gamma}_+ - \overline{\Gamma}_-| \gg |2n_0 - n_+ - n_-| \quad (20)$$

We obtain $s_{max} \approx s'_{max}$ where

$$s'_{max} = \overline{s}_{max}\left(1 + \frac{n_+ - n_-}{\overline{\Gamma}_+ - \overline{\Gamma}_-}\right) \quad (21)$$

Note: we cannot assume that $|\overline{\Gamma}_+ - \overline{\Gamma}_-| \gg |n_+ - n_-|$ because, near the focused position, $\overline{\Gamma}_+$ and $\overline{\Gamma}_-$ may be nearly equal. Simplifying the expression for s'_{max} we obtain

$$s'_{max} = \overline{s}_{max} + \frac{\delta}{2}\left(\frac{n_+ - n_-}{2\overline{\Gamma}_0 - \overline{\Gamma}_+ - \overline{\Gamma}_-}\right) \quad (22)$$

Now the ARMS error is defined as the standard deviation of s'_{max}, i.e.

$$\begin{aligned} \text{ARMS } error &= \frac{\delta}{2}\frac{1}{(2\overline{\Gamma}_0 - \overline{\Gamma}_+ - \overline{\Gamma}_-)} \cdot std\,(n_+ - n_-) \\ &= \frac{\delta}{2}\frac{1}{(2\overline{\Gamma}_0 - \overline{\Gamma}_+ - \overline{\Gamma}_-)} \cdot \left(\sigma_+^2 + \sigma_-^2\right)^{\frac{1}{2}} \end{aligned} \quad (23)$$

where σ_+ and σ_- are the standard deviations of the focus measures Γ_+ and Γ_- respectively.

For a lens position far from the maximum focused position, the above expression for ARMS error will not be valid since the assumption in Equation (20) will not be valid. In this case, the local linear model for the focus measure will be better than the local quadratic model. The ARMS error for this case is based on focus measures at only two lens positions (rather than three) that are δ apart. Without loss of generality, let the two positions be $s_- = -\delta/2$ and $s_+ = +\delta/2$ and the focus measures at these points be Γ_- and Γ_+ respectively (similar to Fig. 11). The linear model yields the expression

$$\frac{s - s_-}{s_+ - s_-} = \frac{\Gamma - \Gamma_-}{\Gamma_+ - \Gamma_-} \quad (24)$$

The above equation can be rewritten as:

$$s = \delta\left(\frac{\Gamma - \Gamma_-}{\Gamma_+ - \Gamma_-}\right) - \frac{\delta}{2} \quad (25)$$

Once again, we express Γ_+ and Γ_- as $\Gamma_+ = \overline{\Gamma}_+ + n_+$ and $\Gamma_- = \overline{\Gamma}_- + n_-$ where $\overline{\Gamma}_+$ and $\overline{\Gamma}_-$ are the expected values and n_+ and n_- are the noise components.

Now the ARMS error is defined as the standard deviation of s' where s' is the solution of $\Gamma(s) = \frac{\overline{\Gamma}_+ + \overline{\Gamma}_-}{2}$. Solving this equation we obtain

$$\begin{aligned} s' &= \frac{\delta}{2}\left(\frac{\overline{\Gamma}_+ - \overline{\Gamma}_- - 2n_-}{\overline{\Gamma}_+ - \overline{\Gamma}_- + n_+ - n_-}\right) - \frac{\delta}{2} \\ &= \frac{\delta}{2}\left[\left(1 - \frac{2n_-}{\overline{\Gamma}_+ - \overline{\Gamma}_-}\right)\left(1 + \frac{n_+ - n_-}{\overline{\Gamma}_+ - \overline{\Gamma}_-}\right)^{-1} - 1\right] \end{aligned} \quad (26)$$

Assuming $|\overline{\Gamma}_+ - \overline{\Gamma}_-| \gg |n_+ - n_-|$ and $|\overline{\Gamma}_+ - \overline{\Gamma}_-| \gg |2n_-|$, we obtain

$$\begin{aligned} s' &\approx \frac{\delta}{2}\left[1 - \frac{2n_-}{\overline{\Gamma}_+ - \overline{\Gamma}_-} - \frac{n_+ - n_-}{\overline{\Gamma}_+ - \overline{\Gamma}_-} - 1\right] \\ &\approx \frac{\delta}{2}\left(\frac{n_+ + n_-}{\overline{\Gamma}_- - \overline{\Gamma}_+}\right) \end{aligned} \quad (27)$$

Hence, the ARMS error would be

$$\text{ARMS error} = std(s') = \frac{\delta\ \left(\sigma_+^2 + \sigma_-^2\right)^{\frac{1}{2}}}{2|\overline{\Gamma}_+ - \overline{\Gamma}_-|} \quad (28)$$

5.1 Relation between AUM and ARMS error

Comparing the expressions for AUM and ARMS error from equations (11) and (23) we find

$$\begin{aligned} \frac{\text{AUM}^2}{\text{ARMS}} &= 16\delta\frac{\sigma}{\sqrt{\sigma_+^2 + \sigma_-^2}} \\ &\approx 8\sqrt{2}\delta \qquad \textbf{if} \quad \sigma_+ \approx \sigma_- \approx \sigma \end{aligned} \quad (29)$$

The ratio of the square of AUM and ARMS error is a constant. Therefore AUM and ARMS error are monotonically related. If we redefine AUM so that instead of using equation (9) we obtain s_1 and s_2 by solving

$$\Gamma(s) = \Gamma_{max} - \sigma^2 \quad (30)$$

then we find that AUM and ARMS error are linearly related for a given focus measure and focused image.

For a lens position far away from the focused position, comparing the expressions (17) and (28) for AUM and ARMS error yields

$$\begin{aligned} \frac{\text{AUM}}{\text{ARMS}} &= \frac{4\sigma}{\sqrt{\sigma_+^2 + \sigma_-^2}} \\ &\approx 2\sqrt{2} \qquad \textbf{if} \quad \sigma_+ \approx \sigma_- \approx \sigma \end{aligned} \quad (31)$$

For this case, they are linearly related.

6 Noise sensitivity analysis

In this section we derive expressions for the *expected value* (mean) and *variance* of a focus measure. These are useful in computing the *standard deviation* σ of the focus measure and its AUM/ARMS error.

Let $f(m,n)$ be the blurred noise free discrete image and $\eta(m,n)$ be the additive noise. The noisy blurred digital image recorded by the camera is

$$f_\eta(m,n) = f(m,n) + \eta(m,n) \tag{32}$$

The noise $\eta(m,n)$ at different pixels are assumed to be independent, identically distributed random variables with zero mean and standard deviation σ_n. This σ_n can be easily estimated for a camera by imaging a uniformly bright object and computing the standard deviation of the grey level distribution. The images are assumed to be of size $(2N+1) \times (2N+1)$ and focus measure filter (FMF) $a(i,j)$ of size $(2M+1) \times (2M+1)$. Without loss of generality, the filtering operation will be represented by the *moving weighted sum* (MWS) operator instead of the usual *convolution* operator. MWS is equivalent to convolution if, for example, the FMF is rotated by 180 degrees about its center by assigning $a(-i,-j)$ to $a(i,j)$. Denoting the MWS operator by $\star$ it is defined by

$$a(i,j) \star f_\eta(m,n) \;=\; \sum_{i=-M}^{M} \sum_{j=-M}^{M} a(i,j) f_\eta(m+i, n+j) \tag{33}$$

In the remaining part of this paper we shall use the following convention to simplify notation. A double summation will be abbreviated with a single summation as:

$$\sum_{m=-N}^{N} \sum_{n=-N}^{N} = \sum_{m,n}^{N} \quad \text{and} \quad \sum_{i=-M}^{M} \sum_{j=-M}^{M} = \sum_{i,j}^{M} \tag{34}$$

Let $g(m,n)$ be the image obtained by filtering the noisy blurred image $f_\eta(m,n)$ with the FMF $a(i,j)$:

$$\begin{aligned} g(m,n) &= a(i,j) \star f_\eta(m,n) \\ &= F(m,n) + \mathcal{N}(m,n) \end{aligned} \tag{35}$$

where

$$F(m,n) = a(i,j) \star f(m,n) \tag{36}$$

and

$$\mathcal{N}(m,n) = a(i,j) \star \eta(m,n) \tag{37}$$

The focus measure γ is defined as

$$\begin{aligned} \gamma &= \frac{1}{(2N+1)^2} \sum_{m,n}^{N} g^2(m,n) \\ &= \gamma_{signal} + \gamma_{noise} + \frac{2}{(2N+1)^2} \times \\ & \sum_{m,n}^{N} F(m,n)\mathcal{N}(m,n) \end{aligned} \tag{38}$$

where γ_{signal} and γ_{noise} are defined by:

$$\gamma_{signal} = \frac{1}{(2N+1)^2} \sum_{m,n}^{N} F^2(m,n) \tag{39}$$

$$\gamma_{noise} = \frac{1}{(2N+1)^2} \sum_{m,n}^{N} \mathcal{N}^2(m,n) \tag{40}$$

Now the expected value of the focus measure $E\{\gamma\}$ is (note that the expectation operator E is linear and commutes with summation):

$$\begin{aligned} E\{\gamma\} &= \gamma_{signal} + \gamma_\eta + \frac{2}{(2N+1)^2} \\ & \times \sum_{m,n}^{N} F(m,n) E\{\mathcal{N}(m,n)\} \end{aligned} \tag{41}$$

where

$$\gamma_\eta \;=\; E\{\gamma_{noise}\} \tag{42}$$

Since we assume $\eta(m,n)$ is zero mean, the last term of equation (41) will vanish. Now the second term can be written as

$$\begin{aligned} \gamma_\eta &= \frac{1}{(2N+1)^2} \sum_{m,n}^{N} \sum_{i_1,j_1}^{M} \sum_{i_2,j_2}^{M} a(i_1,j_1) a(i_2,j_2) \\ & \times E\{\eta(m+i_1, n+j_1)\eta(m+i_2, n+j_2)\} \end{aligned} \tag{43}$$

In the above equation, if $i_1 \neq i_2$ or $j_1 \neq j_2$, then, since noise in different pixels are independent and zero mean,

$$\begin{aligned} & E\{\eta(m+i_1, n+j_1)\eta(m+i_2, n+j_2)\} \\ =\; & E\{\eta(m+i_1, n+j_1)\} E\{\eta(m+i_2, n+j_2)\} \\ =\; & 0 \end{aligned} \tag{44}$$

However, if $i_1 = i_2$ and $j_1 = j_2$, then

$$E\{\eta^2(m+i_1, n+j_1)\} \;=\; \sigma_n^2 \tag{45}$$

Therefore, we get

$$\begin{aligned} \gamma_\eta &= \frac{1}{(2N+1)^2} \sum_{m,n}^{N} \sum_{i,j}^{M} a^2(i,j) \sigma_n^2 \\ &= A_n \sigma_n^2 \end{aligned} \tag{46}$$

where

$$A_n = \sum_{i,j}^{M} a^2(i,j) \tag{47}$$

Therefore

$$\begin{aligned} E\{\gamma\} &= \gamma_{signal} + \gamma_\eta \\ &= \gamma_{signal} + A_n \sigma_n^2 \end{aligned} \tag{48}$$

The above equation is a fundamental result. It shows that the expected value of the focus measure is a sum of two components– one due to signal alone and another due to noise alone. Therefore, if a focus measure is computed on a set of images for autofocusing, the effect of noise is to increase the computed focus measure by the same value on average for all images. The reason for this is that while the image signal changes in blur level with lens position, the noise characteristics of the camera remains the same. Therefore, the average increase in focus measure due to noise does not change the location of the focus measure peak. It is the variance of the focus measure that changes the location of the focus measure peak and therefore introduces error in autofocusing.

Now consider the variance of the focus measure:

$$Var\{\gamma\} = E\{\gamma^2\} - (E\{\gamma\})^2 \tag{49}$$

From Eq.(38), noting that $E\{\mathcal{N}(m,n)\} = 0$ and $E\{\mathcal{N}^2(m_1,n_1)\mathcal{N}(m_2,n_2)\} = 0$, we obtain

$$\begin{aligned} E\{\gamma^2\} &= \gamma_{signal}^2 + 2\gamma_{signal}\gamma_\eta + E\{\gamma_{noise}^2\} + \\ & \frac{4}{(2N+1)^4}\sum_{m_1,n_1}^{N}\sum_{m_2,n_2}^{N} F(m_1,n_1)F(m_2,n_2)\times \\ & E\{\mathcal{N}(m_1,n_1)\mathcal{N}(m_2,n_2)\} \end{aligned} \tag{50}$$

From Eqs. (48 , 49, 50) we obtain

$$\begin{aligned} Var\{\gamma\} &= E\{\gamma_{noise}^2\} - \gamma_\eta^2 + \frac{4}{(2N+1)^4}\sum_{m_1,n_1}^{N}\sum_{m_2,n_2}^{N} \\ & F(m_1,n_1)F(m_2,n_2)E\{\mathcal{N}(m_1,n_1)\mathcal{N}(m_2,n_2)\} \end{aligned} \tag{51}$$

Note that $E\{\gamma_{noise}^2\} - \gamma_\eta^2$ is the variance of γ_{noise} which is independent of signal. Therefore , equation (51) can be written as:

$$\begin{aligned} Var\{\gamma\} &= Var\{\gamma_{noise}\} + \frac{4}{(2N+1)^4}\sum_{m_1,n_1}^{N}\sum_{m_2,n_2}^{N} \\ & F(m_1,n_1)F(m_2,n_2)E\{\mathcal{N}(m_1,n_1)\mathcal{N}(m_2,n_2)\} \end{aligned} \tag{52}$$

The equation above shows that the variance of a focus measure depends on the image signal in addition to noise level. Further simplification of the above expression is presented in Appendix A. The formula presented there can be applied directly in practical applications. Now we consider 3 examples to illustrate the application of the formula. In these examples, the noise will be modeled as Gaussian. For a zero mean Gaussian random variable η with standard deviation σ_n we have [9] $E\{\eta^4\} = 3\sigma_n^4$. This result will be used in the following examples.

1. Gray Level Variance

The image is normalized by subtracting the mean grey value from the grey level of each pixel. The focus measure filter in this case is

$$a(i,j) = \begin{cases} 1 & \text{if } i=j=0 \\ 0 & \text{otherwise} \end{cases} \tag{53}$$

Using the formula (80) for variance derived in Appendix A we obtain

$$Var\{\gamma\} = \frac{2\sigma_n^4}{(2N+1)^2} + \frac{4\sigma_n^2}{(2N+1)^4}\sum_{m,n}^{N} f^2(m,n) \tag{54}$$

2. Gradient Magnitude Squared

For gradient squared along x-axis

$$a(i,j) = [-1 \quad 1] \tag{55}$$

Substituting $a(i,j)$ above in Eq. (80) for variance in Appendix A we obtain:

$$\begin{aligned} Var\{\gamma_x\} &= \frac{12\sigma_n^4}{(2N+1)^2} + \frac{4\sigma_n^2}{(2N+1)^4}\times \\ & \sum_{m,n}^{M+N} [A_x(i,j) * f(m,n)]^2 \end{aligned} \tag{56}$$

where

$$A_x(i,j) = [-1 \quad 2 \quad -1] \tag{57}$$

For gradient squared along y-axis

$$\begin{aligned} Var\{\gamma_y\} &= \frac{12\sigma_n^4}{(2N+1)^2} + \frac{4\sigma_n^2}{(2N+1)^4}\times \\ & \sum_{m,n}^{M+N} [A_y(i,j) * f(m,n)]^2 \end{aligned} \tag{58}$$

where

$$A_y(i,j) = \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix} \tag{59}$$

Now, a cross item $Var\{\gamma_{xy}\}$ is generated by the effect of noise on γ_x and γ_y which are not independent. See Appendix D for an expression of $Var\{\gamma_{xy}\}$. Therefore, combining Eq. (56), (58) and (119) we have

$$\begin{aligned} Var\{\gamma\} &= \frac{24\sigma_n^4}{(2N+1)^2} + \frac{4\sigma_n^2}{(2N+1)^4}\sum_{m,n}^{M+N} \\ & [A_x(i,j) * f(m,n) + A_y(i,j) * f(m,n)]^2 \end{aligned} \tag{60}$$

3. Laplacian

The discrete Laplacian is approximated by

$$a(i,j) = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (61)$$

Substituting this $a(i,j)$ into the formula (80) for variance we obtain

$$Var\{\gamma\} = \frac{1352\sigma_n^4}{(2N+1)^2} + \frac{4\sigma_n^2}{(2N+1)^4} \times \sum_{m,n}^{M+N} [A(i,j) * f(m,n)]^2 \quad (62)$$

where

$$A(i,j) = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 2 & -8 & 2 & 0 \\ 1 & -8 & 20 & -8 & 1 \\ 0 & 2 & -8 & 2 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \quad (63)$$

7 The Laplacian filter

The Laplacian FMF has some desirable properties which makes it suitable for practical applications. It is sound, simple, rotationally symmetric, and is a linear filter (unlike the gradient magnitude). In addition it has two additional properties under the geometric optics model of image formation not shared by other filters based on image derivatives. The two properties correspond to the two cases of a noise free image: (i) when the image does not change due to blurring as is the case when the grey level variation is linear, the value of the focus measure is zero, and (ii) when the image does change as is the case when the grey level variation is a polynomial of second or higher degree, the focus measure is non-zero. When the grey level variation in a noise free image is linear, then the image does not change due to blurring under geometric optics model of image formation (see Appendix C for more details). Therefore the first property is not shared by the grey-level variance and image gradient focus measures corresponding to zeroth and first order derivatives respectively. FMFs based on third and higher order derivatives do not share the second property because when the grey level variation is only quadratic, the corresponding focus measures are zero even though the images change due to blurring. This is because the third and higher order derivatives of a quadratically varying grey level image is zero everywhere.

The first property makes it easy to test the hypothesis whether a given image region has the necessary "contrast" so that autofocusing is possible. The confidence level of the hypothesis that the "contrast signal" is indeed present in a given image region and therefore autofocusing is possible can be estimated from the quantity:

$$\beta = \frac{\gamma - \gamma_\eta}{\text{std}(\gamma_{noise})} \quad (64)$$

where γ is the computed focus measure, γ_η is the expected value of the focus measure if only noise was present, and std(γ_{noise}) is the standard deviation of the focus measure if only noise was present. If the focus measure due to noise alone γ_{noise} has a normal distribution with mean γ_η and standard deviation std(γ_{noise}), then β must be 3.0 or more for 99.9% confidence that some contrast information is present in the image region for autofocusing.

In a sense the Laplacian eliminates all unnecessary information (the constant and linear components of grey level variation in the focused image and therefore their contribution to the focus measure) and retains all necessary information (the quadratic and higher level components in the focused image and therefore their contribution to the focus measure). Therefore we recommend the Laplacian FMF for autofocusing and depth-from-focus in practical applications. In the presence of high noise levels, we recommend the Laplacian of the Gaussian (LOG) filter [2] as the FMF.

8 Appendix A

The term $E\{\gamma_{noise}^2\}$ can be shown to be equal to

$$\frac{1}{(2N+1)^4} \sum_{m_1,n_1}^{N} \sum_{m_2,n_2}^{N} \sum_{i_1,j_1}^{M} \sum_{i_2,j_2}^{M} \sum_{i_3,j_3}^{M} \sum_{i_4,j_4}^{M} \left(\prod_{k=1}^{4} a(i_k,j_k) \right) E\{\eta(m_1+i_1, n_1+j_1) \eta(m_1+i_2, n_1+j_2)\eta(m_2+i_3, n_2+j_3) \eta(m_2+i_4, n_2+j_4)\} \quad (65)$$

The above expression can be evaluated by considering the following cases.

Case 1: Among the four factors involving η, at least one of them is different from the other three. In this case the entire term evaluates to zero.

Case 2: All the four factors involving η are the same. Here the conditions C_1 and C_2 will both true where

$$C_1 : (i_1 = i_2) \ \& \ (j_1 = j_2) \ \& \ (i_3 = i_4) \ \& \ (j_3 = j_4) \quad (66)$$

$$C_2 : (m_2 = m_1 + i_1 - i_3) \ \& \ (n_2 = n_1 + j_1 - j_3) \quad (67)$$

Therefore, the term becomes

$$\begin{aligned} & \frac{E\{\eta^4\}}{(2N+1)^4} \sum_{m_1,n_1}^{N} \sum_{i_1,j_1}^{M} \sum_{i_3,j_3}^{M} a^2(i_1,j_1) a^2(i_3,j_3) \\ = & \frac{E\{\eta^4\}}{(2N+1)^4} \sum_{m_1,n_1}^{N} \left(\sum_{i_1,j_1}^{M} a^2(i_1,j_1) \right)^2 \\ = & \frac{E\{\eta^4\}}{(2N+1)^2} A_n^2 \end{aligned} \quad (68)$$

Case 3: Any two pairs among the four factors involving η are equal but not all four are equal. Here we consider two subcases.

Case (a): Condition C_1 is true and condition C_2 is false. In this case we get

$$\begin{aligned}
& \frac{\sigma_n^4}{(2N+1)^4} \sum_{m_1,n_1}^{N} \sum_{m_2,n_2}^{N} \sum_{i_1,j_1}^{M} \sum_{i_3,j_3}^{M} a^2(i_1,j_1)a^2(i_3,j_3) \\
- & \frac{\sigma_n^4}{(2N+1)^4} \sum_{m_1,n_1}^{N} \sum_{i_1,j_1}^{M} \sum_{i_3,j_3}^{M} a^2(i_1,j_1)a^2(i_3,j_3) \\
= & \sigma_n^4 A_n^2 \left[1 - \frac{1}{(2N+1)^2}\right] \qquad (69)
\end{aligned}$$

Case (b): Define conditions

$$\begin{aligned}
C_3 \ : \ & (m_1+i_1 = m_2+i_3) \ \& \ (n_1+j_1 = n_2+j_3) \\
& \& \ (m_1+i_2 = m_2+i_4) \ \& \ (n_1+j_2 = n_2+j_4)
\end{aligned} \qquad (70)$$

$$\begin{aligned}
C_4 \ : \ & (m_1+i_1 = m_2+i_4) \ \& \ (n_1+j_1 = n_2+j_4) \\
& \& \ (m_1+i_2 = m_2+i_3) \ \& \ (n_1+j_2 = n_2+j_3)
\end{aligned} \qquad (71)$$

From the above conditions we deduce respectively

$$C_5 \ : \ (i_1 - i_3 = i_2 - i_4) \ \& \ (j_1 - j_3 = j_2 - j_4) \qquad (72)$$

$$C_6 \ : \ (i_1 - i_4 = i_2 - i_3) \ \& \ (j_1 - j_4 = j_2 - j_3) \qquad (73)$$

Let Q be a boolean variable with value 1 if the following condition is true and zero otherwise:

$$Q \ :: \ (C_5 \text{ OR } C_6) \ \& \text{ NOT } C_1 \qquad (74)$$

In this case we obtain

$$\frac{\sigma_n^4}{(2N+1)^2} \sum_{i_1,j_1}^{M} \sum_{i_2,j_2}^{M} \sum_{i_3,j_3}^{M} \sum_{i_4,j_4}^{M} Q \cdot \left(\prod_{k=1}^{4} a(i_k,j_k) \right) \qquad (75)$$

Now consider

$$\begin{aligned}
& \frac{4}{(2N+1)^4} \sum_{m_1,n_1}^{N} \sum_{m_2,n_2}^{N} F(m_1,n_1)F(m_2,n_2) \\
& E\{\mathcal{N}(m_1,n_1)\mathcal{N}(m_2,n_2)\} \ = \ \frac{4}{(2N+1)^4} \times \\
& \sum_{m_1,n_1}^{N} \sum_{m_2,n_2}^{N} \sum_{i_1,j_1}^{M} \sum_{i_2,j_2}^{M} a(i_1,j_1)a(i_2,j_2) \times \\
& E\{\eta(m_1+i_1, n_1+j_1)\eta(m_2+i_2, n_2+j_2)\} \\
& \times F(m_1,n_1)F(m_2,n_2) \qquad (76)
\end{aligned}$$

The term involving η is non zero only when $m_1 + i_1 = m_2 + i_2$ **and** $n_1 + j_1 = n_2 + j_2$. Introducing the change of variables $m = m_1 + i_1 = m_2 + i_2$ and $n = n_1 + j_1 = n_2 + j_2$ for $-2(M+N) - 1 \leq m, n \leq 2(M+N) + 1$, therefore we get

$$\begin{aligned}
& \frac{4\sigma_n^2}{(2N+1)^4} \sum_{m,n}^{M+N} \left[\sum_{i_1,j_1}^{M} a(i_1,j_1)F(m-i_1, n-j_1) \right] \\
& \times \left[\sum_{i_2,j_2}^{M} a(i_2,j_2)F(m-i_2, n-j_2) \right] \\
= & \frac{4\sigma_n^2}{(2N+1)^4} \sum_{m,n}^{M+N} [a(i,j) * F(m,n)]^2 \\
= & \frac{4\sigma_n^2}{(2N+1)^4} \sum_{m,n}^{M+N} [A(i,j) * f(m,n)]^2 \\
= & \frac{4\sigma_n^2}{(2N+1)^4} \sum_{m,n}^{M+N} F'^2(m,n) \qquad (77)
\end{aligned}$$

where $*$ represents the convolution operator and

$$\begin{aligned}
A(i,j) &= a(i,j) * a(i,j) \qquad (78) \\
F'(m,n) &= A(i,j) * f(m,n) \qquad (79)
\end{aligned}$$

Combining all the results from eqs. (46, 68, 69, 75 and **??**), we get

$$\begin{aligned}
Var\{\gamma\} \ = \ & \frac{A_n^2 E\{\eta^4\}}{(2N+1)^2} - \frac{A_n^2 \sigma_n^4}{(2N+1)^2} + \frac{\sigma_n^4}{(2N+1)^2} \\
& \times \sum_{i_1,j_1}^{M} \sum_{i_2,j_2}^{M} \sum_{i_3,j_3}^{M} \sum_{i_4,j_4}^{M} Q \cdot \left(\prod_{k=1}^{4} a(i_k,j_k) \right) + \\
& \frac{4\sigma_n^2}{(2N+1)^2} \gamma'_{signal}
\end{aligned} \qquad (80)$$

where

$$\gamma'_{signal} = \frac{1}{(2N+1)^2} \sum_{m,n}^{M+N} F'^2(m,n) \qquad (81)$$

In the above equation, the first three terms do not depend on the image signal. They can be computed and prestored. Among these three terms, the first two can be computed manually, but the third term may need a small computer program to evaluate. The last term in the above equation depends on the image being processed. Exact computation of this term requires knowledge of the noise-free image which is not possible. However the value of the term can be approximated using the noisy image $g(m,n)$. The approximation is valid for high signal to noise ratio. Therefore we have

$$Var\{\gamma\} \ \approx \ \frac{A_n^2 E\{\eta^4\}}{(2N+1)^2} - \frac{A_n^2 \sigma_n^4}{(2N+1)^2} + \frac{\sigma_n^4}{(2N+1)^2}$$

$$\times \sum_{i_1,j_1}^{M} \sum_{i_2,j_2}^{M} \sum_{i_3,j_3}^{M} \sum_{i_4,j_4}^{M} Q \cdot \left(\prod_{k=1}^{4} a(i_k, j_k) \right) +$$

$$\frac{4\sigma_n^2}{(2N+1)^4} \sum_{m,n}^{M+N} (A(i,j) * g(m,n))^2 \quad (82)$$

9 Appendix B: Integral of absolute value of derivative of an image is not a sound focus measure

It is shown here that a class of images can be constructed for which the integral of absolute values of derivative of an image is not a sound focus measure

First we consider the case of a one-dimensional signal and then indicate how it can be extended to two-dimensional images.

A PSF $h(x)$ has the properties that

$$h(x) \geq 0 \text{ for all } x \quad (83)$$

and

$$\int_{-\infty}^{\infty} h(x) dx = 1 \quad (84)$$

Now consider an arbitrary function $p(x)$ with properties $p(x) \geq 0$ for all x and $\int_{-\infty}^{\infty} p(x)dx = 1$. Let γ_i be a focus measure defined as the integral of the absolute value of the i-th derivative of a blurred signal $g(x)$, i.e.

$$\gamma_i = \int_{-\infty}^{\infty} \left| \frac{d^i g(x)}{dx^i} \right| dx \quad (85)$$

Let $g(x) = h(x) \otimes f(x)$ where $f(x)$ is the focused signal corresponding to $g(x)$ and $\otimes$ denotes the convolution operator. Therefore

$$\begin{aligned} \gamma_i &= \int_{-\infty}^{\infty} \left| \frac{d^i}{dx^i} (h(x) \otimes f(x)) \right| dx \\ &= \int_{-\infty}^{\infty} \left| h(x) \otimes \frac{d^i f(x)}{dx^i} \right| dx \quad (86) \end{aligned}$$

Let $f(x)$ be the solution of the differential equation

$$\frac{d^i f(x)}{dx^i} = A\, p(x) \quad (87)$$

where A is any positive constant, i.e. $A > 0$. We have

$$\begin{aligned} \gamma_i &= A \int_{-\infty}^{\infty} |h(x) \otimes p(x)| dx \\ &= A \quad (88) \end{aligned}$$

The last step of the derivation above can be justified by the following well-known result in the probability theory. If $h(x)$ and $p(x)$ are the probability density functions of two independent random variables x_h and x_p respectively, then the probability density function of their sum, i.e. $x_h + x_p$, is $h(x) \otimes p(x)$, $h(x) \otimes p(x) > 0$ for all x, and

$$\int_{-\infty}^{\infty} h(x) \otimes p(x) dx = 1 \quad (89)$$

Therefore, for a focused signal $f(x)$ given by the solution of Eq. (87), The focus measure γ_i remains the same no matter what $h(x)$ is. Therefore γ_i is not sound in that it does not have a maximum when $h(x) = \delta(x)$ (*dirac delta function*). Two particular examples are $p(x) = \delta(x)$ for γ_1 and γ_2 where $f(x)$ will be $A \cdot u(x)$ and $A \cdot ramp(x)$ respectively. Note that $u(x)$ is the unit step function and $ramp(x)$ is the unit ramp function defined as

$$ramp(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (90)$$

The one-dimensional case above can be easily extended to two-dimensional images by assuming that the image changes along only one of the dimensions and is a constant along the other dimension. A more general counter example can be constructed as follows: Let

$$g(x,y) = 0 \quad \text{if } |x| > B \text{ or } |y| > B \quad (91)$$

$$\gamma_i = \gamma_{xi} + \gamma_{yi} \quad (92)$$

$$\gamma_{xi} = \frac{1}{(2B)^2} \int_{-B}^{B} \int_{-B}^{B} \left| \frac{\partial^i g(x,y)}{\partial x^i} \right| dx\, dy \quad (93)$$

$$\gamma_{yi} = \frac{1}{(2B)^2} \int_{-B}^{B} \int_{-B}^{B} \left| \frac{\partial^i g(x,y)}{\partial y^i} \right| dx\, dy \quad (94)$$

the PSF be separable

$$h(x,y) = h_x(x) \cdot h_y(y) \quad (95)$$

and

$$\frac{\partial^i f(x,y)}{\partial x^i} = A_x\, p_x(x) \quad (96)$$

$$\frac{\partial^i f(x,y)}{\partial y^i} = A_y\, p_y(y) \quad (97)$$

where $p_x(x)$ and $p_y(y)$ have the properties of a probability density function (i.e. they are always positive and they integrate to unity). Therefore

$$\begin{aligned} \gamma_{xi} = & \frac{1}{(2B)^2} \int_{-B}^{B} \int_{-B}^{B} \left| \frac{\partial^i}{\partial x^i} ((h_x(x) \cdot h_y(y)) \otimes f(x,y)) \right| \\ & dx\, dy \\ = & \frac{1}{(2B)^2} \int_{-B}^{B} \int_{-B}^{B} |(h_y(y) \cdot h_x(x)) \otimes (A_x \cdot p_x(x))| \\ & dx\, dy \\ = & \frac{A_x}{(2B)^2} \int_{-B}^{B} \int_{-B}^{B} \left| \{ \int_{-B}^{B} h_y(y) dy \} \{ h_x(x) \otimes p_x(x) \} \right| \end{aligned}$$

$$dx\, dy = \frac{A_x}{2B} \tag{98}$$

Similarly we can obtain $\gamma_{yi} = \frac{A_y}{2B}$. Hence the proof.

10 Appendix C: Properties of Laplacian based focus measure

The two properties of the Laplacian based focus measure discussed in section 6 are proved here. First it is shown that: (i) the image does not change due to blurring when the grey level variation is linear, and in this case, (ii) the focus measures based on grey level variance and energy of the image gradient are in general non-zero, but the focus measure based on the Laplacian is zero.

It is shown in [12] that when an image $f(x,y)$ given by a third order polynomial is blurred to obtain $g(x,y)$ by a rotationally symmetric PSF $h(x,y)$, we have

$$g(x,y) = f(x,y) + \frac{h_{2,0}}{2}\nabla^2 f(x,y) \tag{99}$$

where $h_{2,0}$ is the second moment of $h(x,y)$ defined by

$$h_{2,0} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 h(x,y)\, dx\, dy \tag{100}$$

Eq. (99) can be used to obtain an expression for $g(x,y)$ when $f(x,y)$ is linear by setting the coefficients of second and third order terms in x and y to be zero. Therefore, if

$$f(x,y) = a_0 + a_1 x + a_2 y \tag{101}$$

we have

$$g(x,y) = f(x,y) \tag{102}$$

because

$$\nabla^2 f(x,y) = 0 \tag{103}$$

Eq. (102) proves that the image does not change due to blurring when the grey level variation is linear (Eq. (101)). The brightness normalized image $g_n(x,y)$ is obtained by dividing $g(x,y)$ by its mean in an interval $-A \le x \le A$, $-A \le y \le A$. We have

$$g_n(x,y) = \frac{a_0 + a_1 x + a_2 y}{a_0} \tag{104}$$

It can be easily verified that, in general, the focus measures of $g_n(x,y)$ computed in the interval $-A \le x \le A$, $-A \le y \le A$, is non-zero for image grey level variance, and energy of image gradient along x and y directions, but the Laplacian based focus measure is zero since

$$\nabla^2 g_n(x,y) = 0 \tag{105}$$

Next it is shown that: (i) when the image is a second order polynomial (quadratic), it changes due to blurring, (ii) the Laplacian based focus measure of the image is non-zero, and (iii) focus measures based on third and higher order derivatives are zero. Let

$$f(x,y) = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 xy + a_5 y^2 \tag{106}$$

Setting the coefficients of third order terms in x and y to be zero in eq. (99), we obtain

$$\begin{aligned} g(x,y) &= a_0 + (a_3 + a_5)\, h_{2,0} + a_1 x \\ &\quad + a_2 y + a_3 x^2 + a_4 xy + a_5 y^2 \end{aligned} \tag{107}$$

After normalizing the image by dividing it by its mean we have

$$g_n(x,y) = \frac{g(x,y)}{a_0 + (a_3 + a_5)\, h_{2,0} + \frac{A^2}{3}(a_3 + a_5)} \tag{108}$$

The Laplacian of the normalized image can be expressed as

$$\nabla^2 g_n(x,y) = \frac{2\,(a_3 + a_5)}{a_0 + (a_3 + a_5)\frac{A^2}{3}\left(1 + \frac{3h_{2,0}}{A^2}\right)} \tag{109}$$

Eq. (108) proves that the image in general changes due to blurring. Eq. (109) shows that the Laplacian based focus measure is in general non-zero. In Eq. (108), we see that third and higher order derivatives of $g_n(x,y)$ are all zero. This proves that focus measures based on third and higher order derivatives are zero, even though the image has changed due to blurring. If $g(x,y)$ is not normalized for brightness, then $f(x,y)$ has to be at least a 4-th order polynomial in order for the Laplacian focus measure to change for $g(x,y)$ as compared with $f(x,y)$.

11 Appendix D. Gradient Focus Measure

The accurate derivation for the variance of focus measure using gradient magnitude squared filter should consider a cross item generated through x direction and y direction. If we start with our definition from equation (38), the total focus measure γ would be sum of along x-axis γ_x and y axis γ_y components and it becomes:

$$\begin{aligned} \gamma &= \gamma_x + \gamma_y \\ &= \gamma_{sx} + \gamma_{nx} + \frac{2}{(2N+1)^2}\sum_{m,n}^{N} F_x(m,n)\mathcal{N}_x(m,n) \\ &+ \gamma_{sy} + \gamma_{ny} + \frac{2}{(2N+1)^2}\sum_{m,n}^{N} F_y(m,n)\mathcal{N}_y(m,n) \end{aligned} \tag{110}$$

where the suffix sx and nx mean the focus measure component contributed in signal and noise along x axis separately as so do sy and ny.

If the noise characteristics are the same in both of direction, the expected value of focus measure in the equation (48) is able to be applied here as:

$$E\{\gamma\} = \gamma_{sx} + \gamma_{sy} + 4\sigma_n^2 \tag{111}$$

However, a cross item appear as we take square to both of γ and $E\{\gamma\}$. If we define:

$$F_x\mathcal{N}_x = \frac{2}{(2N+1)^2}\sum_{m,n}^{N} F_x(m,n)\mathcal{N}_x(m,n) \qquad (112)$$

as well as

$$F_y\mathcal{N}_y = \frac{2}{(2N+1)^2}\sum_{m,n}^{N} F_y(m,n)\mathcal{N}_y(m,n) \qquad (113)$$

and substituting to eq. (110), then the variance will be written from eq. (49) as:

$$\begin{aligned} Var\{\gamma\} &= Var\{\gamma_x\} + Var\{\gamma_y\} + Var\{\gamma_{xy}\} \\ &= E\{\left(\gamma_{sx} + 2\sigma_n^2 + F_x\mathcal{N}_x\right)^2 + \left(\gamma_{sy} + 2\sigma_n^2 + F_y\mathcal{N}_y\right)^2 \\ &+2\left(\gamma_{sx} + 2\sigma_n^2 + F_x\mathcal{N}_x\right)\left(\gamma_{sy} + 2\sigma_n^2 + F_y\mathcal{N}_y\right)\} - \\ &\{\left(\gamma_{sx} + 2\sigma_n^2\right)^2 + \left(\gamma_{sy} + 2\sigma_n^2\right)^2 + 2\left(\gamma_{sx} + 2\sigma_n^2\right) \times \\ &\left(\gamma_{sy} + 2\sigma_n^2\right)\} \qquad (114) \end{aligned}$$

where

$$Var\{\gamma_x\} = E\{\left(\gamma_{sx} + 2\sigma_n^2 + F_x\mathcal{N}_x\right)^2\} - \left(\gamma_{sx} + 2\sigma_n^2\right)^2 \qquad (115)$$

and

$$Var\{\gamma_y\} = E\{\left(\gamma_{sy} + 2\sigma_n^2 + F_y\mathcal{N}_y\right)^2\} - \left(\gamma_{sy} + 2\sigma_n^2\right)^2 \qquad (116)$$

The $Var\{\gamma_x\}$ and the $Var\{\gamma_y\}$ are exactly same as we derived in eqs. (56) and (58). Thus, the $Var\{\gamma_{xy}\}$ is a cross term and equals to:

$$\begin{aligned} 2E\{\left(\gamma_{sx} + 2\sigma_n^2 + F_x\mathcal{N}_x\right)\left(\gamma_{sy} + 2\sigma_n^2 + F_y\mathcal{N}_y\right)\} \\ -2\left(\gamma_{sx} + 2\sigma_n^2\right)\left(\gamma_{sy} + 2\sigma_n^2\right) \qquad (117) \end{aligned}$$

Because of $E\{\mathcal{N}_x\} = E\{\mathcal{N}_y\} = 0$, the above equation can be simplified and expanded as:

$$\begin{aligned} Var\{\gamma_{xy}\} &= 2E\{F_x\mathcal{N}_x F_y\mathcal{N}_y\} \\ &= \frac{8}{(2N+1)^4}\sum_{m_1,n_1}^{N}\sum_{m_2,n_2}^{N} F_x(m_1,n_1)F_y(m_2,n_2) \\ &\quad E\{\mathcal{N}_x(m_1,n_1)\mathcal{N}_y(m_2,n_2)\} \\ &= \frac{8}{(2N+1)^4}\sum_{m_1,n_1}^{N}\sum_{m_2,n_2}^{N}\sum_{i_1,j_1}^{M}\sum_{i_2,j_2}^{M} \\ &\quad a_x(i1,j1)a_y(i2,j2)F_x(m_1,n_1)F_y(m_2,n_2) \\ &\quad E\{\eta(m1+i1,n1+j1)\eta(m2+i2,n2+j2)\} \qquad (118) \end{aligned}$$

The term $E\{\eta(m1+i1,n1+j1)\eta(m2+i2,n2+j2)\}$ is non zero only when $m = m_1 + i_1 = m_2 + i_2$ **and** $n = n_1 + j_1 = n_2 + j_2$ for $-2(M+N) - 1 \le m,n \le 2(M+N)+1$, therefore we have

$$\begin{aligned} Var\{\gamma_{xy}\} &= \frac{8\sigma_n^2}{(2N+1)^4} \times \sum_{m,n}^{M+N} \\ &\left[\sum_{i_1,j_1}^{M} a_x(i1,j1)F_x(m-i1,n-j1)\right] \times \\ &\left[\sum_{i_2,j_2}^{M} a_y(i2,j2)F_y(m-i2,n-j2)\right] \\ &= \frac{8\sigma_n^2}{(2N+1)^4}\sum_{m,n}^{M+N}\left[a_x(i,j) * F_x(m,n)\right] \times \\ &\quad \left[a_y(i,j) * F_y(m,n)\right] \\ &= \frac{8\sigma_n^2}{(2N+1)^4}\sum_{m,n}^{M+N}\left[A_x(i,j) * f(m,n)\right] \times \\ &\quad \left[A_y(i,j) * f(m,n)\right] \qquad (119) \end{aligned}$$

where

$$A_x(i,j) = a_x(i,j) * a_x(i,j) \qquad (120)$$

and

$$A_y(i,j) = a_y(i,j) * a_y(i,j) \qquad (121)$$

, they are the same as Eq. (57) and Eq. (59) respectively.

References

[1] J. Enns and P. Lawrence, "A Matrix Based Method for Determining Depth from Focus", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 1991.

[2] R. M. Haralick and L. G. Shapiro, "*Computer and Robot Vision*",Vol. 1, Addison-Wesley Publishing Co., Inc, 1992.

[3] B. K. P. Horn, "Focusing", Artificial Intelligence Memo No. 160, MIT, May 1968.

[4] H. C. Lee, "A review of image blur model in a photographic system using the principles of optics", *Optical Engineering*, May, 1990.

[5] H. H. Hopkins, "The frequence response of a defocused optical system" *Proceeding Royal Society of London*, A 231, 1955, pp 91-103.

[6] E. Krotkov, "Focusing", *International Journal of Computer Vision*, 1, 223-237, 1987.

[7] G. Ligthart and F. Groen, "A Comparison of Different Autofocus Algorithms", *International Conference on Pattern Recognition*, pp. 597-600,1982.

[8] Nayar, S.K., "Shape from Focus System" *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Champaign, Illinois, pp. 302-308 (June 1992).

[9] A. Papoulis, "*Probability, Random Variables, and Stochastic Processes*", McGraw-Hill Book Company, the 3rd, 1991.

[10] M. Subbarao, and A. Nikzad, "Model for image sensing and digitization in computer vision", *Proceedings of SPIE conference* , Vol 1385, pp70-84, October 1990, Boston.

[11] M. Subbarao, T. Choi, and A. Nikzad, "Focusing Techniques", *Journal of Optical Engineering*, Vol. 32 No. 11, pp. 2824-2836, November 1993.

[12] M. Subbarao and G. Surya, "Depth from Defocus: A Spatial Domain Approach", *International Journal of Computer Vision*, 13, 3, pp. 271-294 (1994).

[13] M. Subbarao, and M. C. Lu, "Computer Modeling and Simulation of Camera Defocus", *Machine Vision and Applications*, (1994) 7, pp. 277-289.

[14] M. Subbarao, and T. Wei, "Depth from Defocus and Rapid Autofocusing: A Practical Approach", Proceedings of IEEE Computer Society conference on Computer Vision and Pattern Recognition, June 1992, pp. 773-776.

[15] M. Subbarao, and J.K. Tyan, "The Optimal Focus Measure for Passive Autofocusing and Depth-from-Focus", *Proceedings of SPIE conference on Viedometrics IV* , Philadelphia, Oct 1995.

[16] Schlag, J.F., Sanderson, A.C., Neumann, C.P., and Wimberly, F.C., "Implementation of Automatic Focusing Algorithms for a Computer Vision System with Camera Control", *Tech. Report CMU-RI-TR-83-14*, Carnegie Mellon University (August 1983).

[17] J. M. Tenenbaum, "Accommodation in Computer Vision", Ph.D. Dissertation, Stanford University, November 1970.